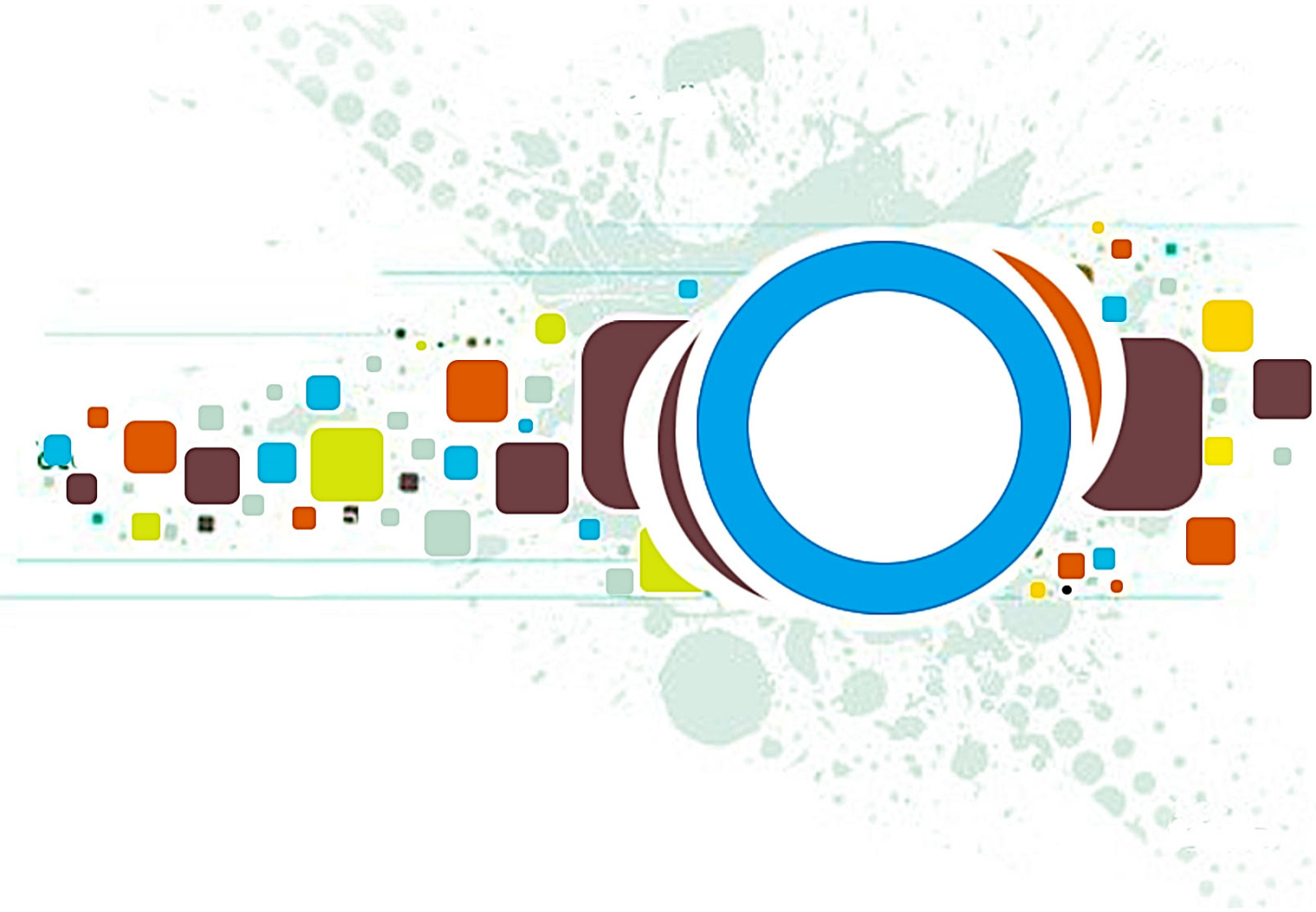
Volume 8 • Issue 3 • June 2014

INTERNATIONAL JOURNAL OF

IMAGE PROCESSING (IJIP)

ISSN : 1985-2304

Publication Frequency: 6 Issues Per Year



CSC PUBLISHERS
http://www.cscjournals.org

INTERNATIONAL JOURNAL OF IMAGE PROCESSING (IJIP)

VOLUME 8, ISSUE 3, 2014

EDITED BY
DR. NABEEL TAHIR

ISSN (Online): 1985-2304

International Journal of Image Processing (IJIP) is published both in traditional paper form and in Internet. This journal is published at the website http://www.cscjournals.org, maintained by Computer Science Journals (CSC Journals), Malaysia.

IJIP Journal is a part of CSC Publishers

Computer Science Journals

http://www.cscjournals.org

INTERNATIONAL JOURNAL OF IMAGE PROCESSING (IJIP)

Book: Volume 8, Issue 3, June 2014
Publishing Date: 01-06-2014
ISSN (Online): 1985-2304

IJIP Journal is a part of CSC Publishers
http://www.cscjournals.org

Published in Malaysia

Typesetting: Camera-ready by author, data conversation by CSC Publishing Services – CSC Journals, Malaysia

CSC Publishers, 2014

EDITORIAL PREFACE

The International Journal of Image Processing (IJIP) is an effective medium for interchange of high quality theoretical and applied research in the Image Processing domain from theoretical research to application development. This is the *Third* Issue of Volume *Eight* of IJIP. The Journal is published bi-monthly, with papers being peer reviewed to high international standards. IJIP emphasizes on efficient and effective image technologies, and provides a central for a deeper understanding in the discipline by encouraging the quantitative comparison and performance evaluation of the emerging components of image processing. IJIP comprehensively cover the system, processing and application aspects of image processing. Some of the important topics are architecture of imaging and vision systems, chemical and spectral sensitization, coding and transmission, generation and display, image processing: coding analysis and recognition, photopolymers, visual inspection etc.

The initial efforts helped to shape the editorial policy and to sharpen the focus of the journal. Starting with Volume 8, 2014, IJIP appears in more focused issues. Besides normal publications, IJIP intends to organize special issues on more focused topics. Each special issue will have a designated editor (editors) – either member of the editorial board or another recognized specialist in the respective field.

IJIP gives an opportunity to scientists, researchers, engineers and vendors from different disciplines of image processing to share the ideas, identify problems, investigate relevant issues, share common interests, explore new approaches, and initiate possible collaborative research and system development. This journal is helpful for the researchers and R&D engineers, scientists all those persons who are involve in image processing in any shape.

Highly professional scholars give their efforts, valuable time, expertise and motivation to IJIP as Editorial board members. All submissions are evaluated by the International Editorial Board. The International Editorial Board ensures that significant developments in image processing from around the world are reflected in the IJIP publications.

IJIP editors understand that how much it is important for authors and researchers to have their work published with a minimum delay after submission of their papers. They also strongly believe that the direct communication between the editors and authors are important for the welfare, quality and wellbeing of the Journal and its readers. Therefore, all activities from paper submission to paper publication are controlled through electronic systems that include electronic submission, editorial panel and review system that ensures rapid decision with least delays in the publication processes.

To build its international reputation, we are disseminating the publication information through Google Books, Google Scholar, Directory of Open Access Journals (DOAJ), Open J Gate, ScientificCommons, Docstoc and many more. Our International Editors are working on establishing ISI listing and a good impact factor for IJIP. We would like to remind you that the success of our journal depends directly on the number of quality articles submitted for review. Accordingly, we would like to request your participation by submitting quality manuscripts for review and encouraging your colleagues to submit quality manuscripts for review. One of the great benefits we can provide to our prospective authors is the mentoring nature of our review process. IJIP provides authors with high quality, helpful reviews that are shaped to assist authors in improving their manuscripts.

Editorial Board Members
International Journal of Image Processing (IJIP)

EDITORIAL BOARD

TABLE OF CONTENTS

Volume 8, Issue 3, June 2014

Pages

Image Processing Application on Graphics processors

Chouchene Marwa *ch.marwa.84@gmail.com*
Laboratory of Electronics and Microelectronics (EμE)
Faculty of Sciences Monastir
Monastir, 5000, Tunisia

Bahri Haythem *bahri.haythem@hotmail.com*
Laboratory of Electronics and Microelectronics (EμE)
Faculty of Sciences Monastir
Monastir, 5000, Tunisia

Sayadi Fatma Ezahra *sayadi_fatma@yahoo.fr*
Laboratory of Electronics and Microelectronics (EμE)
Faculty of Sciences Monastir
Monastir, 5000, Tunisia

Atri Mohamed *mohamed.atri@fsm.rnu.tn*
Laboratory of Electronics and Microelectronics (EμE)
Faculty of Sciences Monastir
Monastir, 5000, Tunisia

Abstract

In this work, we introduce real time image processing techniques using modern programmable Graphic Processing Units GPU. GPU are SIMD (Single Instruction, Multiple Data) device that is inherently data-parallel. By utilizing NVIDIA new GPU programming framework, “Compute Unified Device Architecture” CUDA as a computational resource, we realize significant acceleration in image processing algorithm computations. We show that a range of computer vision algorithms map readily to CUDA with significant performance gains. Specifically, we demonstrate the efficiency of our approach by a parallelization and optimization of image processing, Morphology applications and image integral.

Keywords: Image Processing, GPU, CUDA.

1. INTRODUCTION

The graphics processing unit (GPU) has become an integral part of today’s mainstream computing systems. The modern GPU is not only a powerful graphics engine but also a highly parallel programmable processor featuring peak arithmetic and memory bandwidth that substantially outpaces its CPU counterpart.

The speed of GPU increases the programming capacity so the resolution and processing complex problems. This effort in computing has positioned the GPU as an alternative to traditional microprocessors in high-performance computing system.

CUDA (Compute Unified Device Architecture) is a general architecture for parallel computing introduced by NVidia in November 2007[1]. It includes a new programming model, new architecture and an different set instruction.

In other words, generally the image processing algorithms treat all pixels in any way. But with new technology it is possible to treat all pixels in parallel. For this reason, we can use the

graphics cards because they are capable of performing a large number of times the same function in parallel. In this paper, we will perform some image processing algorithms on the GPU. The paper is organized as follows: In the first part we present a GPU overview. Subsequently a second part deals with the image processing on the GPU. Finally, we conclude the paper.

2. GPU OVERVIEW

There are about ten years of the birth of a hardware dedicated to graphics processing offloading the CPU, called GPU, Graphics Processing Unit. GPUs are supported by a map annex, taking charge of their image processing and 3D data, and display. With its architecture is designed for a massively parallel processing of data, they should become a co-processor used by any application.

Graphics processors are now used to accelerate graphics and some general applications with high data parallelism (GPGPU). For good performance, the parallel data is processed by a large number of processing units that are integrated within the GPU. These are organized according to a single instruction, multiple data (SIMD) or single program, multiple data (SPMD). There are arithmetic units, loading units and interpolation, units of data pre-processing and finally evaluation units of the basic functions.

Indeed, these architectures have evolved to move towards systems more programmable: first through the shaders, but remains very specific 2D or 3D graphics, and more recently through the kernels total. In addition, it is possible that a cluster of GPUs is both more efficient and less energy than a cluster of CPU [2], which makes GPU clusters particularly attractive in the field of high performance computing.

3. NVIDIA CUDA PROGRAMMING

Traditionally, general-purpose GPU programming was accomplished by using a shader-based framework [3]. This framework has a steep learning curve that requires in-depth knowledge of specific rendering pipelines and graphics programming. Algorithms have to be mapped into vertex transformations or pixel illuminations. Data have to be cast into texture maps and operated on like they are texture data. Because shader-based programming was originally intended for graphics processing, there is little programming support for control over data flow; and, unlike a CPU program, a shader-based program cannot have random memory access for writing data. There are limitations on the number of branches and loops a program can have. All of these limitations hindered the use of the GPU for general-purpose computing. NVIDIA released CUDA, a new GPU programming model, to assist developers in general-purpose computing in 2007 [4]. In the CUDA programming framework, the GPU is viewed as a compute device that is a co-processor to the CPU. The GPU has its own DRAM, referred to as device memory, and execute a very high number of threads in parallel. More precisely, data-parallel portions of an application are executed on the device as kernels which run in parallel on many threads. In order to organize threads running in parallel on the GPU, CUDA organizes them into logical blocks. Each block is mapped onto a multiprocessor in the GPU. All the threads in one block can be synchronized together and communicate with each other. Because there is a limited number of threads that a block can contain, these blocks are further organized into grids allowing for a larger number of threads to run concurrently as illustrated in Figure 1. Threads in different blocks cannot be synchronized, nor can they communicate even if they are in the same grid. All the threads in the same grid run the same GPU code.

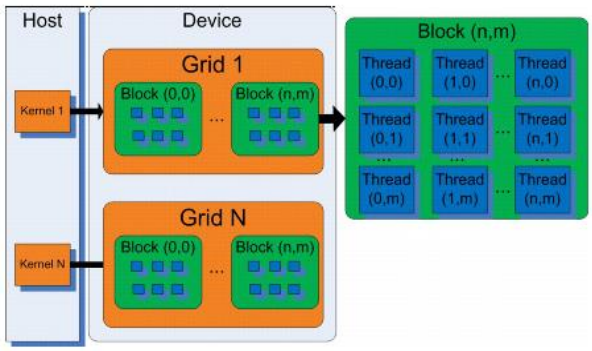


FIGURE 1: Thread and Block Structure of CUDA.

CUDA has several advantages over the shader-based model. Because CUDA is an extension of C, there is no longer a need to understand shader-based graphics APIs. This reduces the learning curve for most of C/C++ programmers. CUDA also supports the use of memory pointers, which enables random memory-read and write-access ability. In addition, the CUDA framework provides a controllable memory hierarchy which allows the program to access the cache (shared memory) between GPU processing cores and GPU global memory. As an example, the architecture of the GeForce 8 Series, the eighth generation of NVIDIA's graphics cards, based on CUDA is shown in Figure 2.

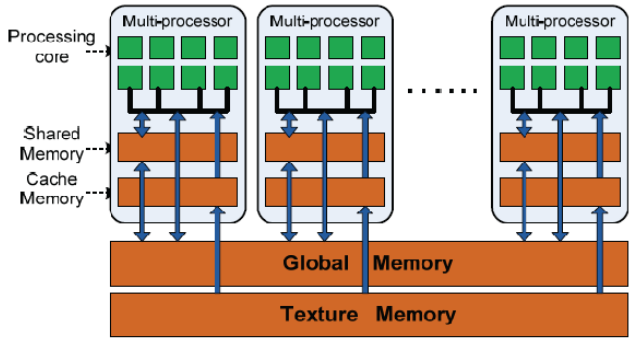


FIGURE 2: GeForce 8 Series GPU Architecture.

The GeForce 8 GPU is a collection of multiprocessors, each of which has 16 SIMD (Single Instruction, Multiple Data) processing cores. The SIMD processor architecture allows each processor in a multiprocessor to run the same instruction on different data, making it ideal for data-parallel computing. Each multiprocessor has a set of 32-bit registers per processors, 16KB of shared memory, 8KB of read-only constant cache, and 8KB of read-only texture cache. As depicted in Figure 2, shared memory and cache memory are on-chip. The global memory and texture memory that can be read from or written to by the CPU are also in the regions of device memory. The global and texture memory spaces are persistent across all the multiprocessors.

4. GPU COMPUTATION IN IMAGE PROCESSING

In recent years, manufacturers of graphics cards highlight the capacity of their GPU to do anything but the game use include implementation we have the image processing applications.

In this part we have implemented many application of image processing using GPU.

4.1 Transformation RGB to Gray

Our application is to display an image, tests are performed on an RGB color image, and we will load this image and send it directly as a texture in memory and then process it for image grayscale first time on CPU and GPU on the second time.

In the remainder of our testing of graphics processors, we ran our program by changing the used image size. We measured the time taken to carry out this treatment using the CPU and GPU, the results are as follows:

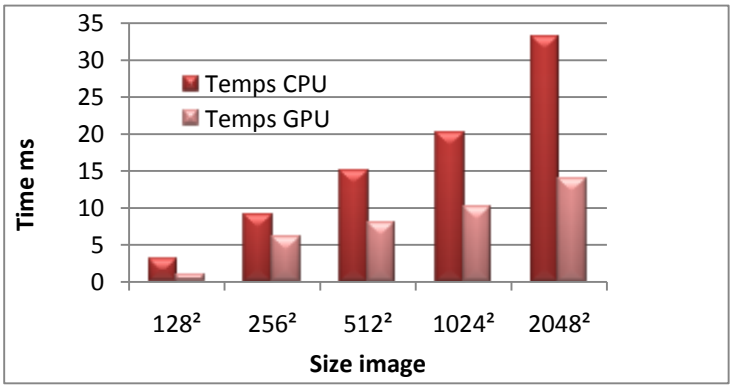


FIGURE 3: Image Processing of Various Sizes.

It should be noted that the execution time on GPU still lower than on the CPU. By increasing the size of the image The GPU time becomes slower and it is due to the data transfer time.

4.2 Morphology Applications

In the following, we will present results of applying morphological (Sobel, Erode) on different images, performed using library GPUCV, and compared with that of the OpenCV library.

We compared the execution time of native CPU operators with their GPU counterpart. GPUCV is up to times faster than native CPU, as shown in figure 4.

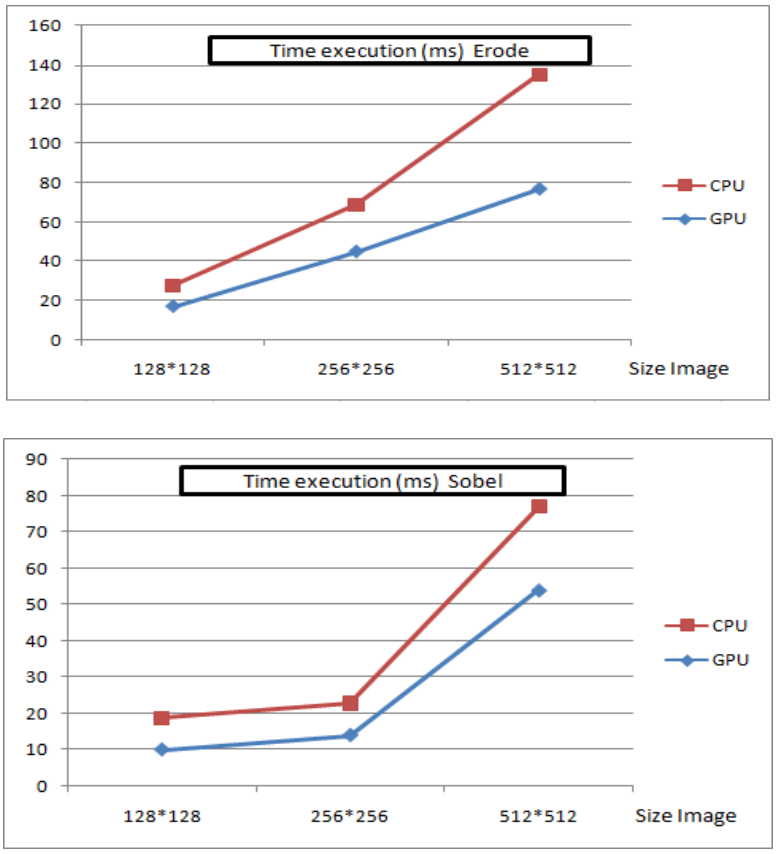


FIGURE 4: Sobel and Erode Transformation of Various Sizes.

4.3 Calculation For Integral Image

The integral image was proposed by Paul Viola and Michael Jones in 2001 and has been used in the real-time object detection [5]. The integral image used is constructed from an original image grayscale.

In [6], the integral image has been extended; it is then used to compute the Haar-like features, and characteristics of centersurround. In the SURF algorithm [7] [8], it accelerates the calculation first and second order of Gaussian derivative that uses the integral image.

In the algorithm CenSurE [9], and its improved version SUSur [10], it uses two levels of filter to approximate the Laplace operator using the integral image.

We present our algorithm to calculate the integral image. The implementation steps are shown in figure 5:

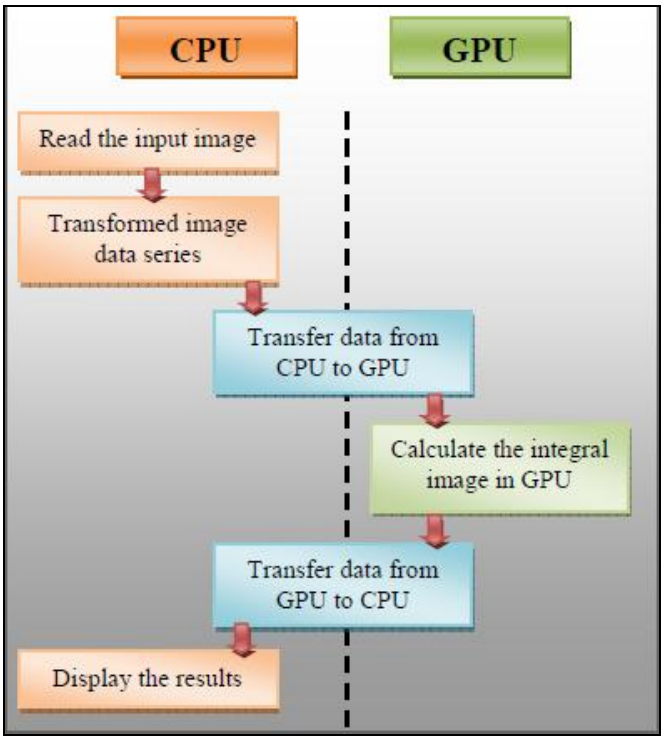


FIGURE 5: Algorithm as Implemented on Hardware.

In order to evaluate our parallel integral image implementation, we executed our algorithm for different image sizes in both CPU and GPU (figure 6).

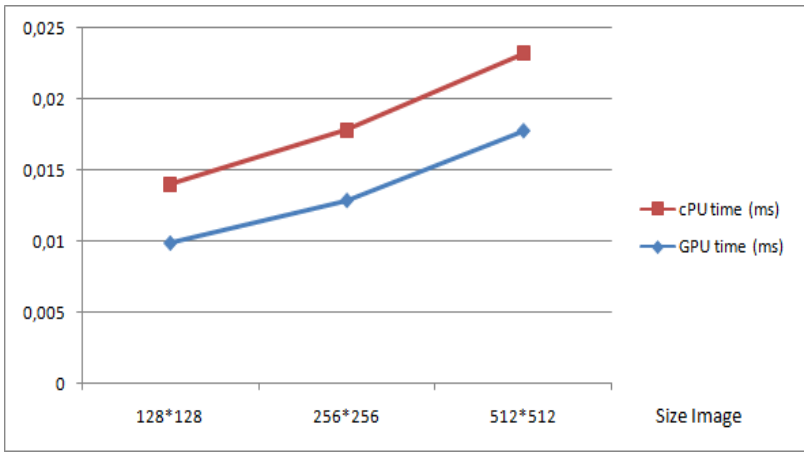


FIGURE 6: Integral Image For Various Sizes.

As can be seen from the table 1, compared to the corresponding CPU-based serial algorithm, our algorithm has a relatively reduction in time-consuming. With the increasing of the image size used in the experiment, speedup increases. However, due to the memory limitations in the video chip, the processed image size at one time cannot be increased unlimitedly.

5. CONCLUSIONS

In this work, we focused on the use of generic graphics cards (GPU) as a parallel computing machine, by explaining their structure and characteristics of their programming. Their increasing use in this context was exposed through a wide range of diverse application areas.

With the GPU, we responded to the needs generated by different computational applications. First, we have accelerated the simulation of a simple image processing. In a second step, we proposed an adaptation GPU, a morphological algorithm (Sobel, Erode). We got time gains evidence from a CPU implementation.

Finally, a parallel integral image algorithm, is presented and implemented on GPU, and compared with the sequential implementations based on CPU. Performance results indicate that significant speedup can be achieved.

6. REFERENCES

[1] NVIDIA Corporation, "NVIDIA CUDA Compute Unified Device Architecture Programming Guide", Version 3, 2013.

[2] L. Abbas-Turki, S. Vialle, B. Lapeyre, and P. Mercier. "High Dimensional Pricing of Exotic European Contracts on a GPU Cluster, and Comparison to a CPU Cluster". In Second InternationalWorkshop on Parallel and Distributed Computing in Finance, May 2009.

[3] Allard, J. and Raffin, B., "A shader-based parallel rendering framework. in Visualization", 2005,VIS 05. IEEE, pp 127-134.

[4] NVIDIA, CUDA Programming Guide Version 1.1. 2007, NVIDIA Corporation: Santa Clara, California.

[5] P Viola, M Jones, "Rapid object detection using a boosted cascade of simple features", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition.2001.

[6] R Lienhart, A Kuranov, V Pisarevsky, "Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection", Pattern Recognition. 2003, vol. 2781, pp. 297-304.

[7] H Bay, A Ess, T Tuytelaars, L Van Gool, "Speededup robust features (SURF) ", International Journal on Computer Vision and Image Understanding, 2008, vol. 110, no. 3, pp. 346–359.

[8] H Bay, T Tuytelaars, L Van Gool, "SURF: Speeded up robust features", Proceedings of the European Conference on Computer Vision, 2006, Springer LNCS volume 3951, part 1, pp 404–417.

[9] M Agrawal, K Konolige, M R Blas "Censure: Center surround extremas for realtime feature detection and matching", 2008, ECCV (4), volume 5305 of Lecture Notes in Computer Science, Springer. pp 102– 115.

[10] M Ebrahimi, W W Mayol-Cuevas, "SUSurE: Speeded Up Surround Extrema feature detector and descriptor for realtime applications", IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2009, CVPRW, pp.9-14.

Multimodal Approach for Face Recognition using 3D-2D Face Feature Fusion

Naveen S *nsnair11176@gmail.com*
Assistant Professor, Dept. of ECE
LBS Institute of Technology for Women
Trivandrum, 695012, Kerala, India

Dr R.S Moni *moni2006rs@gmail.com*
Professor, Dept. of ECE
Marian Engineering College,
Trivandrum, Kerala, India

Abstract

3D Face recognition has been an area of interest among researchers for the past few decades especially in pattern recognition. The main advantage of 3D Face recognition is the availability of geometrical information of the face structure which is more or less unique for a subject. This paper focuses on the problems of person identification using 3D Face data. Use of unregistered 3D Face data for feature extraction significantly increases the operational speed of the system with huge database enrollment. In this work, unregistered 3D Face data is fed to a classifier in multiple spectral representations of the same data. Discrete Fourier Transform (DFT) and Discrete Cosine Transform (DCT) are used for the spectral representations. The face recognition accuracy obtained when the feature extractors are used individually is evaluated. The use of depth information alone in different spectral representation was not sufficient to increase the recognition rate. So a fusion of texture and depth information of face is proposed. Fusion of the matching scores proves that the recognition accuracy can be improved significantly by fusion of scores of multiple representations. FRAV3D database is used for testing the algorithm.

Keywords: Point Cloud, Rotation Invariance, Pose Correction, Depth Map, Spectral Transformations, Texture Map and Principal Component Analysis.

1. INTRODUCTION

3D Face recognition has been an active area of research in the past decades. The complications encountered in the enrollment phase and the huge computational requirements in the implementation phase have been the major hindrance in this area of research. The scenario has improved tremendously due to the latest innovations in 3D imaging devices and has made 3D Face recognition system a reliable option in security systems based on Biometrics. Though poor resolution is a major drawback encountered in 3D Face images the geometrical information present in 3D facial database can be exploited to overcome the challenges in 2D face recognition systems like pose variations, bad illumination, ageing etc.

In this work, focus is made on an identification problem based on 3D Face data using fusion schemes. Identification corresponds to the person recognition without the user providing any information other than the 3D facial scan. The system arrives at an identity from among the enrolled faces in the database. Use of texture information along with the geometrical information of the face seems to improve the recognition accuracy of face recognition system when pose correction is not done as a preprocessing step.

Alexander M. Bronstein, Michael M. Bronstein and Ron Kimmel [2] proposed an idea of face recognition using geometric invariants using Geodesic distances. C. Beumier [3] utilized parallel

planar cuts of the facial surfaces for comparison. Gang Pan, Shi Han, Zhaohui Wu and Yueming Wang [4] extracted ROI of facial surface by considering bilateral symmetry of facial plane. Xue Yuan, Jianming Lu and Takashi Yahagi [5] proposed a face recognition system using PCA, Fuzzy clustering and Parallel Neural networks. Trina Russ et al [6] proposed a method in which correspondence of facial points is obtained by registering a 3D Face to a scaled generic 3D reference face. Ajmal Mian, Mohammed Bennamoun and Robyn Owens [7] used Spherical Face Representation for identification. Ondrej Smirg, Jan Mikulka, Marcos Faundez-Zanuy, Marco Grassi and Jiri Mekyska [8] used DCT for gender classification since the DCT best describes the features after de-correlation. Hua Gao, Hazım Kemal Ekenel, and Rainer Stiefelhagen[9] used Active Appearance model for fitting faces with pose variations. Mohammad Naser, Moghaddasi Yashar Taghizadegan and Hassan Ghassemian [10], used 2D-PCA for getting the feature matrix vectors and used Euclidean distance for classification. , Omid Gervei, Ahmad Ayatollahi, and Navid Gervei[11] proposed an approach for 3D Face recognition based on extracting principal components of range images by utilizing modified PCA methods namely 2D-PCA and bidirectional 2D-PCA. Wang et al.,[14] described point signatures in a 3D domain and facial feature points by using Gabor filter responses in a 2D image. Chang *et al.*, [15] tested the recognition algorithm using fusion of 3D and 2D information and was found effective in improving face recognition rate(FRR). C. McCool *et al.*, [16] used Log-Gabor Templates of depth and texture data was used along with Mahalanobis Cosine metric as the distance measure and in [17] PCA difference vectors are modeled using Gaussian Mixture Models (GMMs) and is compared with Mahalanobis Cosine metric. Pamplona Segundo *et al.*, [18] used boosted cascade classifiers using range images as input for scale invariant face detection for the automation of face recognition systems. Jahanbin S *et al.*, [19] 2-D and 3-D Gabor coefficients and the anthropometric distances are calculated and three parallel classifiers are fused at the match score level to form a face recognition system.

A typical 3D Face is shown in Figure.1. Figure.2 represents its axis level representation. Figure 3 and 4 represents the texture map with two different orientations.

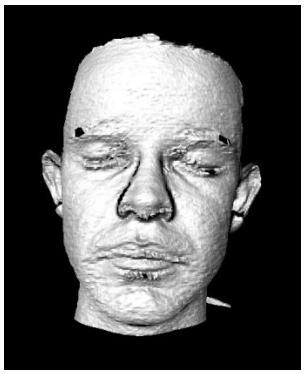


FIGURE 1: 3D Face Model.

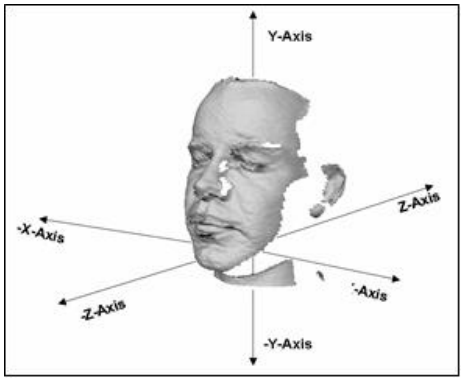


FIGURE 2: 3D Face in Space.

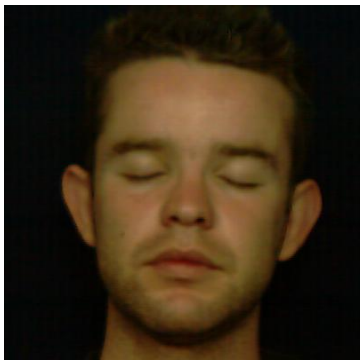


FIGURE 3: 2D Texture Map (Frontal View).

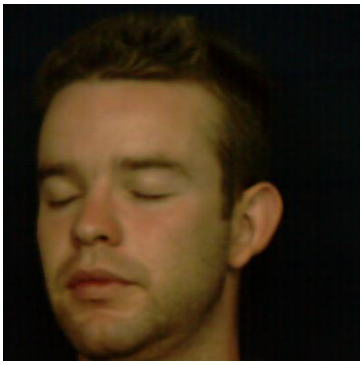


FIGURE 4: 2D Texture Map (Left Turned Head).

First the fusion of the representative transformations from 3D to 1D space and 3D to 2D space is considered. Since only a sparse set of points in the 3D point cloud are available, it is necessary to increase the data density by using multiple data representations generated from same raw data. For this the data is transformed into spectral domain using DFT and DCT. This sparse set of data with occlusion can be effectively countered by invoking multiple score fusion schemes which can effectively improve the feature data density. Use of Depth information alone is not sufficient for an efficient recognition system since pose correction is not done. So texture information is also incorporated with the fusion scheme.

2. PROPOSED IDEA FOR FACE RECOGNITION

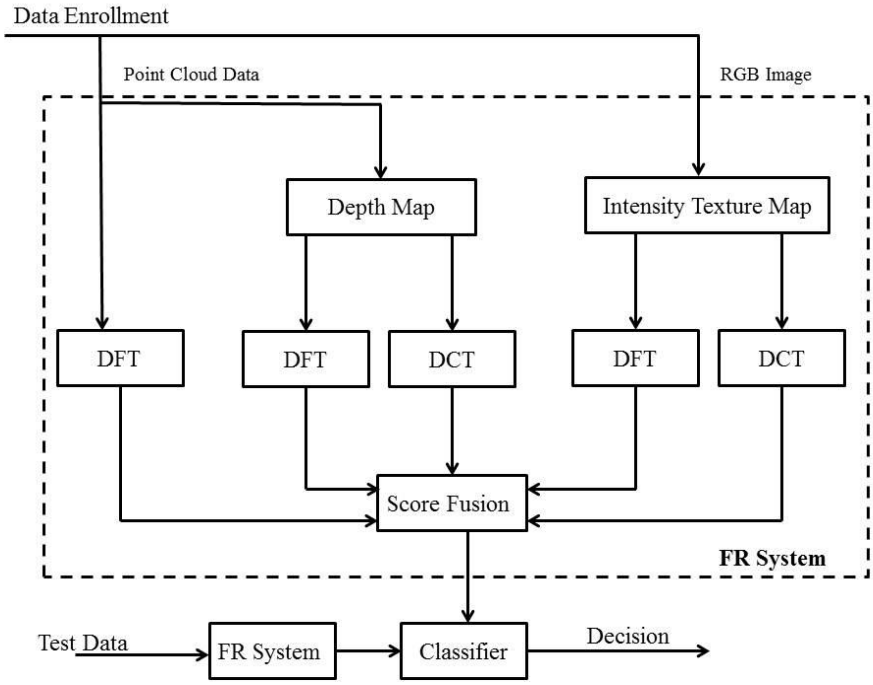


FIGURE 5: Proposed Method **.**

The system aims at extracting the feature from the input data through multiple feature extraction tools and fuses the scores to get a system with better recognition accuracy. The main feature extraction principle used in this system is the spectral transformation. The spectral transformation tools used here are 1D-DFT and 2D-DFT along with 2D-DCT. These spectral transformations transform the data to a better representation which increases the accuracy of recognition system.

The most important part of this work lies in the pattern classification problem. A pattern of data points is available. This pattern is not sufficient for the recognition system to work since the data will be highly occluded due to pose variations in the X, Y and Z axis or in any complex plane.

The 3D Face recognition scheme is affected by pose variations of the subject (person under consideration). There are methods available in which the correction to this effect of pose variations also is included. One such method is the Iterative Closest Point (ICP) algorithm. But the main disadvantage of these methods is that a reference face is to be used as a model for other rotated faces to be corrected. Also the processing time taken is very high. Further, the reliability of this result depends on the accuracy in selection of the reference face model used. Therefore, in this work done, this correction to the effect of pose variation is not considered. The method aims at recognizing the subject without much computationally complex mathematical procedure. Also the results prove that the efficiency of system is comparable with a system with pose correction. The idea behind spectral representation of data is that, when data is in spatial

domain, comparison will be done as one to one pixel level or voxel level. So the rotation and translation of data will highly affect the result. Moreover the accuracy of the system will go down to even 5% under severe pose variations in X, Y and Z axis. When spectral transformation is done the distributed data will be concentrated or it may be represented in a more uniform way. I.e. the input data will be concentrated and represented uniformly in spectral domain. The translation and rotation invariance properties of the transformations used will aid to improve the accuracy of system significantly.

Here FRAV3D database is considered. It contains the facial data with different face orientations and expressions. When depth information alone was considered the Face recognition accuracy (FRA) was not high. So texture data of face is also considered which significantly improves the FRA.

The data available for the analysis and testing will be in Point Cloud format which is a matrix array of size Mx3. The value of M denotes the number of points in the 3D space. For each data input the number of points used will be different for representation. An optimum number of points are selected for Point Cloud data vector. Texture information is available as RGB image of size 400x400. This image is converted to intensity image and is down sampled by a factor 2 to reduce the computation time. Down sampling doesn't reduce the FRA, it is verified.

The proposed method involves the following steps given below in sequence. As a preliminary step Z component of the point cloud data alone is taken and DFT is applied on this data to get spectral representation. Second step is to map the 3D points to 2D grid without pose correction to get the depth map. From this 2D depth map nose tip is detected using Maximum Intensity Method and the area around the nose (ROI-Region of Interest) is extracted (Figure 7 and Figure 8). On this ROI data 2D-DFT and 2D-DCT is applied. Simultaneously spectral representation of Texture map is also taken using 2D-DCT and 2D-DFT. Once spectral representations are obtained, Principal Component Analysis (PCA) is applied on that data to get the corresponding weight vectors. This weight vectors are fed to a classifier which uses Euclidean distance for classification. Here 2D PCA is used for depth and texture features and 1D PCA for Z component of point Cloud.

2.1 Point Cloud Representation as a One Dimensional Vector

We have the point cloud data as an M x 3 matrix. Of this the third dimension, which is the depth information alone is taken. This reduces the number of points under consideration to M while the original being 3M. This also aids the real time implementation of the system faster. We call this data as Metric.

The Metric data will be distributed in spatial domain as shown in Figure 6. It relate to the data of Z axis with some unknown orientation and scale. The data distribution is different for different axis at different orientation. But when DFT is computed, it turns out that the data distribution becomes similar. This significantly will improve the recognition accuracy. The reason for choosing DFT over DCT which has a better energy compaction property is that with DCT, the recognition accuracy with X axis and Y axis rotation has been found to be almost half of that obtained using DFT. Transformation of Metric data to spectral domain can be done using 1D-DFT, using equation (2).

$$F(K)=\sum_{n=0}^{N-1} f(n)e^{-j2\pi Kn/N} \text{ , for a vector of size Nx1} \quad (2)$$

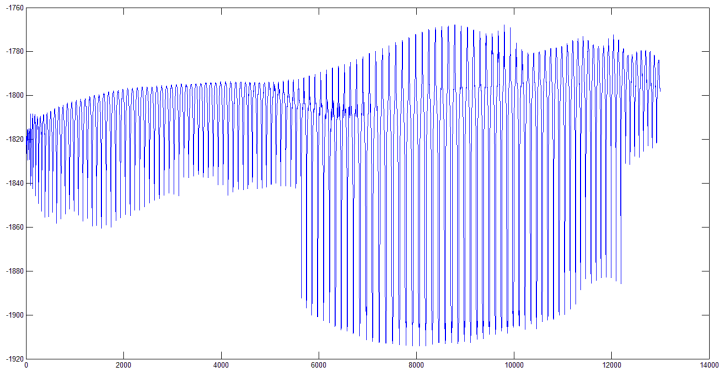


FIGURE 6: Metric in Spatial Domain.

2.2 Nose Tip Localization and Face Area Extraction

For localizing the nose tip, maximum intensity method is used. In this method assumption is made that the nose tip will be the point with maximum pixel intensity. Once the nose tip is found the circular area (ROI) around the nose tip is extracted using an optimum radius. Now the depth map will contain the face area only, all other unwanted portions are cropped away. Next face area is centralized by making the nose tip as the center pixel of the image. Otherwise the matching process will result in a lower accuracy. The face area is also normalized by the maximum intensity. The centralized face image is as shown in Figure 7 and Figure 8. Figure 9 represents the depth map obtained from point cloud data oriented in different axis.

FIGURE 7: Depth Map.

FIGURE 8: ROI from Depth Map.

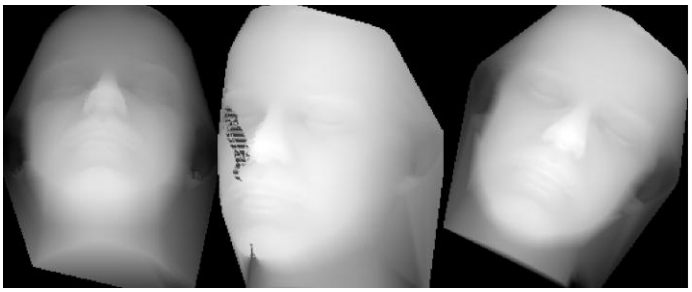


FIGURE 9: Depth Map with different pose variations along X, Y and Z axis**.**

2.3 Use of 2D DFT and 2D DCT on Depth Data and Texture Map

The point cloud data in 3D space is projected to an X-Y grid to get 2.5D (2.5D image is the depth map itself) image using standard projection formula. This depth data will be having the pixel value as the Z- Coordinate of Point Cloud data. Face images have higher redundancy and pixel level correlation which is a major hindrance in face recognition systems. Transforming face images to

spectral domain will reduce the redundancy. Here only the magnitude of spectral data is taken alone since it is not transformed back to spatial domain in any of the processing stages.

Now, the Depth image is transformed to spectral domain using 2D-DCT. The energy compaction will take place and the result will be again an M x N matrix. The result is shown in Figure 10. Here the global feature extraction by DCT is made used of. DCT has the property of de-correlation which enables the data structure to loose spatial pixel dependency. The low frequency components which mainly form the facial features will be prominent in the transformed space which makes the pattern classification more reliable, since the human eyes are more sensitive to information in low frequency spectrum.

Transformation to spectral domain using 2D-DCT can be done using equation (3) and the transformed image is shown in Figure.10.

$$F(u,v)=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\cos\left(\frac{(2x+1)\pi u}{2M}\right)\cos\left(\frac{(2y+1)\pi v}{2N}\right), \text{ for a M x N depth image.} \quad (3)$$

Again the depth image is transformed using 2D-DFT so that rotation effects are reduced. DFT is a rotation invariant transformation. So that the distributed pixel values (normalized) are properly aligned, this enables the pattern matching more efficient. DFT spectrum of face image will appear as shown in Figure 11. Transformation to spectral domain using 2D Discrete Fourier Transformation can be done using equation (4).

$$F(U,V)=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)e^{-j2\pi(Ux/M+Vy/N)}, \text{ for a M x N depth image} \quad (4)$$

Now the error score is estimated using all the multiple representations separately. For processing metric, 1D-PCA is used and for processing 2D DCT and 2D DFT representation 2D-PCA is used. 1D-PCA was also checked for the 2D representations but 2D-PCA gave better result for 2D representations.

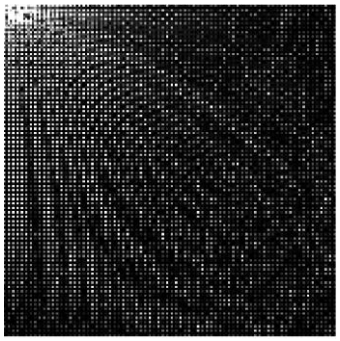


FIGURE 10: DCT Representation of ROI Depth Map. **FIGURE 11:** DFT Representation of ROI Depth Map.

The same procedure is repeated for the texture map also to get the spectral representation of gray scale intensity image using 2D-DCT as shown in Figure 12 and using 2D-DFT as in Figure 13.

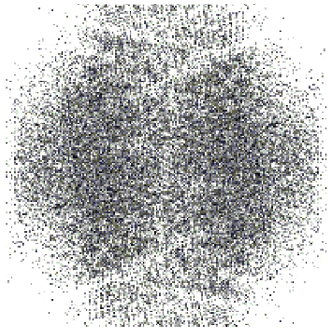


FIGURE 12: DFT Representation of Texture Map.

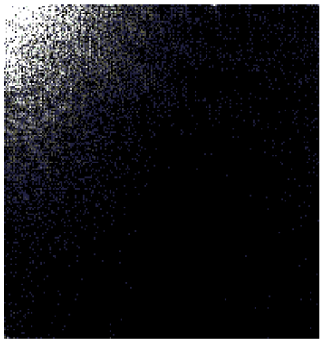


FIGURE 13: DCT Representation of Texture Map.

2.4 Principal Component Analysis

Use of spectral transformations will make the data samples almost uncorrelated. Even then, some spatial dependency may exist. So Principal Component Analysis (PCA)[12] is used for the correlation process as it uses orthogonal transformations to get linear uncorrelated data sets called principal components. Conventional covariance method for the calculation of principal components is used here. Feature extraction using 1D-PCA is done as follows.

Let X_i be the spectral transformed 1D Euclidean Metric which representations i^{th} person, it is grouped as a M x N matrix $X=[X_1\ X_2\ \ldots X_N]$, where N is the number of face samples under consideration.

Mean vector is calculated as follows

$$X_m=\frac{1}{N}\sum_{i=1}^{N} X_i \tag{5}$$

Standard deviation will be calculated

$$X_{SD}=\frac{1}{N}\sum_{i=1}^{N}(X_i - X_m) \tag{6}$$

Covariance matrix is calculated

$$X_{COV}= X_{SD^*}\, X^T_{SD} \tag{7}$$

This is a matrix of size M x M, which is of very large dimension. Also it gives M Eigen values and M Eigen vectors which are very large in number to process. The base idea of dimensional reduction by changing the construction of covariance matrix can be now used.

$$X_{COV}= X^T_{SD^*}\, X_{SD} \tag{8}$$

The result is a matrix of size N x N, where N is the number of subjects under consideration. It gives N Eigen values and N Eigen vectors. The Eigen values are sorted in descending order and will select the first N' largest Eigen values and corresponding Eigen vectors. Eigen vectors in N' dimension is transformed to the higher dimension of M by multiplying with Standard deviation Matrix. Now the test data is projected to this lower dimension space to get the corresponding weight vectors.

In 2D-PCA 2D spectral representation of Depth map is considered. The only difference in calculating the Covariance matrix is that here a 2D matrix is used when compared to 1D Matrix in 1D-PCA. After determining the Eigen values and Eigen vectors a 2D weight vector matrix is obtained which is then converted to a column matrix.

2.5 Score Fusion

Error values are calculated for each data representations and all this error values are combined as a single error value using the linear expression as given in equation9.

W= [W_DMDCT W_DMDFT W_PCDFT W_TMDCT W_TMDFT] (9)

E= [Error_DMDCT Error_DMDFT Error_PCDFT Error_TMDCT Error_TMDFT] (10)

$Error = W * E^{T}$ (11)

On the equation 11 W_DMDCT is weight for error value obtained using 2D DCT on depth data and is its value is taken as 1 , W_DMDFT and W_PCDFT are weights for the error value obtained using 2D DFT and 1D DFT on depth map and metric respectively. This weight values are selected in such way that the error values of 2D DFT and 1D Euclidean metric are in same scale as that of error value due to 2D DCT. Similarly texture score weights W_TMDCT, W_TMDFT are the weights of error score obtained using 2D DCT and 2D-DFT on intensity image. Weight value can be approximated using equation 12, 13, 14 and 15.

$$\frac{1}{W\_DMDFT} = \frac{Error\_DFT}{Error\_DCT}, \text{ rounded to } 10^{2}\text{'s.} \quad (12)$$

$$\frac{1}{W\_PCDFT} = \frac{Error\_PCDFT}{Error\_DCT}, \text{ rounded to } 10^{9}\text{'s.} \quad (13)$$

$$\frac{1}{W\_TMDCT} = \frac{Error\_TMDCT}{Error\_DCT} \approx 1 \quad (14)$$

$$\frac{1}{W\_TMDFT} = \frac{Error\_TMDFT}{Error\_DCT}, \text{ rounded to } 10^{2}\text{'s} \quad (15)$$

These values give optimum recognition accuracy. Finding an optimum weight for this error function which can minimize it can be another optimization problem.

3. RESULTS

3.1 Results of Individual Representation Accuracy Analysis

For analysis and testing FRAV3D database is used here. It contains 106 subjects. Of these 100 subjects were taken into consideration. Testing was done on the input data with pose and orientations as Frontal, Right turn 25º (respect to Y axis), Left turn 5º (respect to Y axis), Severe right turn (respect to Z axis), Soft left turn (respect to Z axis), Smiling face, Open mouth, Looking upwards (turn respect to X axis), Looking downwards (turn respect to X axis), Frontal view with lighting changes etc. A substantial change in the recognition accuracy is observed with fusion scheme.

First recognition accuracy obtained with Depth data and Point cloud metric will be considered. Accuracy is checked by varying the number of faces from 10 to 100, with enrollment increment of

2 faces and a summary is shown in Table 1. When the DCT of Depth Map is analyzed the accuracy variance is about 9.12% with maximum of 54.44% and minimum of 45%, for DFT it is 14.40% with maximum of 57.29 % and minimum of 46.67% and for Point Cloud metric it is 17.13 % with maximum of 60 % and minimum of 43.33%.

No of Subjects	Total Face Samples Tested	Depth Map DCT	Depth Map DFT	Metric DFT
		Accuracy	Accuracy	Accuracy
60	960	48.33	48.33	43.33
90	1440	54.44	56.67	45.56
100	1600	53.00	56.00	46.00

TABLE 1: Comparison of Accuracy obtained using Depth Map Spectral Transformation and Point Cloud Metric.

When the reliability of a Face recognition system is considered, the system should correctly classify the users based on the current pose and posture of the user. A user can be enrolled in the data base of the FR system mostly in limited number of pose/ expression variations. As the number of enrollment sessions increases the system will be trained in such a way that the user can be correctly identified in a latter testing session. But it does have a limitation unless the training sessions given are sufficient which covers all the pose, translation and expression variations.

Here such an analysis is done to check the reliability of the system. The summary results are shown in Table 2.

When Table 2 is analyzed, it can be observed that all 16 sessions are not identified by any of the feature extraction scheme. For example, 7 sessions are successfully identified only by 39 enrolled faces when DCT is used, 46 when DFT is used and 43 when point cloud metric is used out of the total 100 enrolled faces. Obviously this brings down the overall recognition accuracy of the system and hence the reliability. In other words only 273, 322 and 301 samples out of 1600 samples were successfully identified while others are rejected.

	Depth Map DCT	Depth Map DFT	Point Cloud DFT
Sessions	Subjects	Subjects	Subjects
1	100	100	100
2	96	96	96
3	87	91	85
4	82	81	69
5	69	73	55
6	53	61	49
7	39	46	43
8	28	30	33
9	19	22	24
10	14	13	14
11	7	3	10
12	4	1	5
13	2	1	1
14	0	1	1
15	0	0	0
16	0	0	0

TABLE 2: Comparison of number of enrolled faces successfully identified in 16 different sessions.

3.2 Results of Fusion Scheme Analysis

As explained earlier, when Texture Map Score obtained by the spectral transformation of RGB data using 2D-DCT and 2D-DFT is also fused with the score obtained using depth data notable improvement in recognition accuracy is observed. Those observations are shown in Table 3. It can be also noted that the accuracy deviation with final fusion scheme is nearly 4.17%, i.e. the system is reliable even when more number of faces are enrolled and face recognition accuracy won't drop too much.

No of Subjects	Total Face Samples Tested	Fusion DM	FUSION of DM and TM
		Accuracy	Accuracy
60	960	50.31	74.58
90	1440	52.36	76.88
100	1600	52.44	78

TABLE 3: Comparison of FRR of Texture fusion scheme with FRR of Depth Fusion scheme.

When Table 3 is analyzed the accuracy variance for Depth Score Fusion is also around 4.06%. In short fusion of individual score of different representations reduces the accuracy variance and hence increases the reliability of the FR systems.

Table 4 shows the difference in number of subjects identified with fusion of Texture and Depth spectral representation scores and depth alone.

	Fusion DM	Fusion of DM and TM
Sessions	Subjects	Subjects
1	100	100
2	100	100
3	99	100
4	97	100
5	92	100
6	83	100
7	80	95
8	67	87
9	61	82
10	50	73
11	41	66
12	30	55
13	20	37
14	13	25
15	5	14
16	1	3

TABLE 4: Comparison of number of enrolled faces successfully identified in each sessions.

All testing are done only with a single image enrolled in data base as the reference Image. When the images in the database itself are fed as the test image the accuracy is 100% itself. All other samples are tested as the class that is not present in the database, which ensures that the accuracy obtained in the proposed system is insensitive to the session variations and least sensitive to pose and expression variations. The insensitivity is very clear from the accuracy obtained using fusion schemes. Even when the number of samples is doubled, say from 800 to 1600, the FRR change is near to 2% only. Testing with 160 samples which contains complex pose variations and expressions gave 80% accuracy and 1600 samples gave an accuracy of 78% which ensure that the system is reliable.

The main reason for the reduction in accuracy is due to the loss of laser signals while scanning the face and the occlusions due to pose variations [13]. The advantage of proposed work is that face alignment or face registration is not done here. The registration process takes sufficient computation time and it is not acceptable in real time systems. This system requires only a single image as training image for achieving this much accuracy. By increasing the number of training images FRR can be further improved. This method seems to perform better when compared to [10] where the recognition rate is only 70% with a single training image. While 2D PCA [11] method gives 76.2% when tested over 540 samples.

3.3 Improvement In Accuracy Rates by Fusion Scheme

When the accuracy obtained in the fusion scheme is analyzed, when depth data is fused with texture information a significant improvement is observed as shown in Table 5.

No of Subjects	Correct samples additionally recognized	Fusion DM and TM
		Improvement in FRR
60	464	48.23
70	507	45.21
90	675	46.83
100	769	48.04

TABLE 5: Improvement in FRR with fusion of Depth and Texture information.

Almost an average improvement of nearly 48% can be obtained when the fusion scheme is used.

3.4 Face Recognition Rate-Graphical Comparison

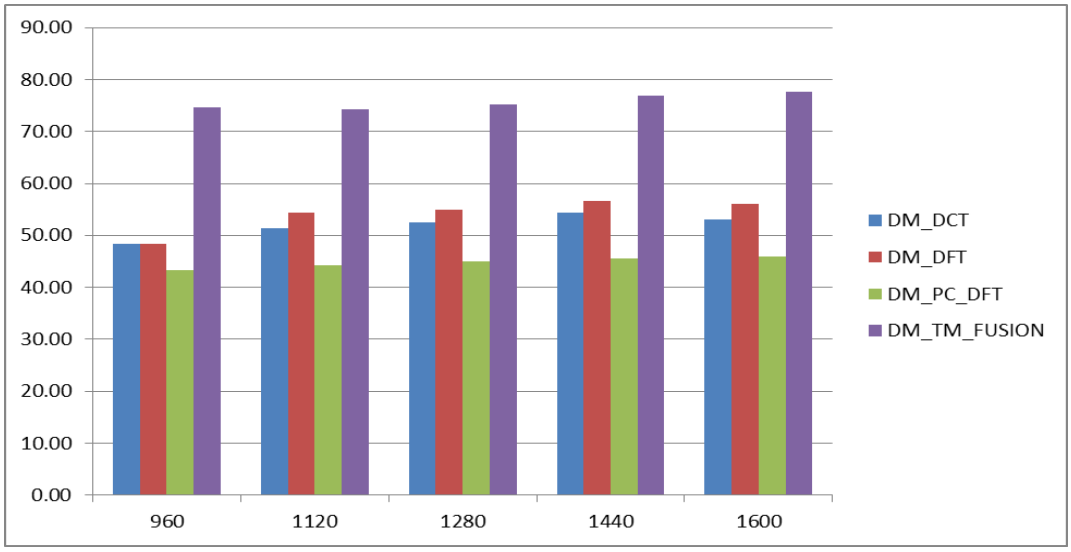


FIGURE 14: FRR Comparison of Individual and Fusion Scheme.

3.5 Computation Time

Testing of algorithm is done on 3GHz, Core I-5 processor; the average identification time per sample is 400ms.This will again increase as the number of subjects increase. By using down sampling and abstracting the data representations computational time can be further reduced. But with a dedicated system, the testing time can be further reduced to microseconds.

4. CONCLUSIONS

The fusion algorithm is tested on unregistered 3D Faces with different pose orientations and expressions. The algorithm gives an accuracy of 78% even with unregistered face data. Experiments are conducted on various representation types and feature extraction methods for 3D Face recognition. The experimental results show that the features can be effectively extracted from depth data, point cloud and texture map using spectral transformation like DFT and DCT. Fusion experiments were conducted at score level. The surface representation of 3D Face data in terms of metric with its spectral transformation, in addition to the spectral representation of projected depth information to the XY plane and Intensity image as texture Map using 2D-DFT and 2D-DCT are also used. When the numbers of subjects in the database are increased, the recognition rate remained more or less same which shows that the further enrollment of faces in

the database won't affect the true recognition rate. There is ample scope for further improvement using more fusion schemes at the representation level and at spectral level. This method can be implemented in real time systems since the processing time required is lesser on a dedicated system. Dimensional reduction method can also improve the time performance of the system. Advancement of technology in 3D Face capturing and faster processing systems can make the system more efficient in all aspects.

Acknowledgment
Thanks are due to Dr. Enrique Cabello, Universidad Rey Juan Carlos, Spain, for providing us with the FRAV3D database.

5. REFERENCES

[1] http://www.frav.es/databases/FRAV3D

[2] Alexander M. Bronstein; Michael M. Bronstein and Ron Kimmel. "Expression-invariant 3D face recognition." In Proc. International Conference on Audio- and Video-based Biometric Person Authentication, volume 2688 of Lecture Notes in Computer Science Guildford, UK, 2003, pp:62-70.

[3] C. Beumier, "3D face recognition" In IEEE Int. Conf. on Computational Intelligence for Homeland Security and Personal Safety (CIHSPS2004), Venice, Italy, Jul 2004, pp:21-22.

[4] Gang Pan; Shi Han Zhaohui Wu and Yueming Wang. "3D Face Recognition using Mapped Depth Images." Proceedings of the IEEEComputer Society Conference on CVPR (CVPR'05) Workshops- Volume- 03, 2005 p:175.

[5] Xue Yuan;Jianming Lu and Takashi Yahagi. "A Method of 3D Face Recognition Based on Principal Component Analysis Algorithm." IEEE International Symposium on Circuits and Systems, Vol. 4 May 2005. pp: 3211 - 3214.

[6] Trina Russ; Chris Boehnen and Tanya Peters. "3D Face Recognition Using 3D Alignment for PCA", Proceedings of the 2006 IEEE Computer Society Conference on CVPR (CVPR'06) Volume 2, 2006 pp: 1391 – 1398.

[7] Ajmal Mian; Mohammed Bennamoun and Robyn Owens. "Automatic 3D Face Detection, Normalization and Recognition." Proceedings of the Third International Symposium on 3DPVT (3DPVT'06) Jun 2006, pp: 735-742.

[8] Ondrej Smirg, Jan Mikulka, Marcos Faundez-Zanuy, Marco Grassi and Jiri Mekyska. "Gender Recognition Using PCA and DCT of Face Images." Advances in Computational Intelligence , Lecture Notes in Computer Science Volume 6692, 2011, pp: 220-227.

[9] Hua Gao, Hazım Kemal Ekenel and Rainer Stiefelhagen. "Pose Normalization for Local Appearance-Based Face Recognition." Advances in Biometrics, Lecture Notes in Computer Science Volume 5558, 2009, pp:32-41.

[10] Mohammad Naser-Moghaddasi Yashar Taghizadegan and Hassan Ghassemian.(2012, Feb), ."3D Face Recognition Method Using 2DPCA-Euclidean Distance Classification", ACEEE International Journal on Control System and Instrumentation(Vol 3), Available: http://hal.archives-ouvertes.fr/docs/00/74/16/40/PDF/70.pdf.

[11] Omid Gervei, Ahmad Ayatollahi and Navid Gervei."3D Face Recognition Using Modified PCA Methods" World Academy of Science, Engineering & Technology; Mar 2010, Issue 39, p264

[12] M. Turk and A. Pentland, "Eigenfaces for recognition", J. Cognitive Neuroscience , 1991, 3(1), pp. 71 – 86.

[13] Image Analysis and Recognition, Third International Conference, ICIAR 2006, Póvoa de Varzim, Portugal, September 18-20, 2006, Proceedings, Part II, Lecture Notes in Computer Science , Volume 4142 2006.

[14] Y. Wang, C. Chua and Y. Ho, “Facial feature detection and face recognition from 2D and 3D images”, Pattern Recognition Letters, vol. 23, 2002, pp. 1191-1202.

[15] K. I. Chang, K. W. Bowyer and P. J. Flynn, “Multiple nose region matching for 3D face recognition under varying facial expression”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 10, 2006, pp. 1695-1700.

[16] C. McCool, J. Cook, V. Chandran and S. Sridharan, “Combined 2D / 3D Face Recognition using Log-Gabor Templates”, Proceedings of the IEEE International Conference on Video and Signal Based Surveillance (AVSS'06), 2006, pp. 83.

[17] C. McCool, J. Cook, V. Chandran and S. Sridharan, “Feature Modelling of PCA Difference Vectors for 2D and 3D Face Recognition”, in Proceedings of IEEE International Conference on Video and Signal Based Surveillance, 2006, pp. 57.

[18] Pamplona Segundo, M , Silva, L and Bellon, O.R.P, “ Real-time scale-invariant face detection on range images” , in IEEE International Conference on Systems, Man and Cybernetics (SMC), 2011 , pp:914 - 919.

[19] Jahanbin, S, Hyohoon Cho, Bovik, A.C, “Passive Multimodal 2-D+3-D Face Recognition Using Gabor Features and Landmark Distances”, in IEEE Transactions on Information Forensics and Security(Volume:6,Issue: 4), 2011, pp: 1287 - 1304.

K2 Algorithm-based Text Detection with An Adaptive Classifier Threshold

Khalid Iqbal *kik.ustb@gmail.com*
Dept. of Computer Science and Technology,
School of Computer and Communication Engineering,
University of Science and Technology Beijing,
Beijing 100083, China.

Xu-Cheng Yin *xuchengyin@ustb.edu.cn*
Dept. of Computer Science and Technology,
School of Computer and Communication Engineering,
University of Science and Technology Beijing,
Beijing 100083, China.

Hong-Wei Hao *hongwei.hao@ia.ac.cn*
Institute of Automation
Chinese Academy of Sciences,
Beijing 100190, China.

Sohail Asghar *sohail.asghar@uaar.edu.pk*
University Institute of Information,
PMAS-Arid Agriculture University,
Rawalpindi, Pakistan

Hazrat Ali *engr.hazratali@yahoo.com*
Department of Computing,
City University London,
United Kingdom.

Abstract

In natural scene images, text detection is a challenging study area for dissimilar content-based image analysis tasks. In this paper, a Bayesian network scores are used to classify candidate character regions by computing posterior probabilities. The posterior probabilities are used to define an adaptive threshold to detect text in scene images with accuracy. Therefore, candidate character regions are extracted through maximally stable extremal region. K2 algorithm-based Bayesian network scores are learned by evaluating dependencies amongst features of a given candidate character region. Bayesian logistic regression classifier is trained to compute posterior probabilities to define an adaptive classifier threshold. The candidate character regions below from adaptive classifier threshold are discarded as non-character regions. Finally, text regions are detected with the use of effective text localization scheme based on geometric features. The entire system is evaluated on the ICDAR 2013 competition database. Experimental results produce competitive performance (precision, recall and harmonic mean) with the recently published algorithms.

Keywords: Bayesian Network, Adaptive Threshold, Bayesian Logistic Regression, Scene Image.

1. INTRODUCTION

With the advent of digital cameras at consumer-end, new challenges are opened to process and analyze the content of images. The image content captured with a camera may have several degradations such as distorted view, uneven illumination, curved or shadow effects. However,

image content can have useful information about vehicle license plates, landmarks recognition, gas/electricity meters and product recognition. The exploitation of image content constitutes a challenging research area to detect and recognize scene text. The extracted information from scene images has to be robustly located. The determination of locating text in scene images, with complex background, different illumination and variable fonts and size, and text orientation, refers to text localization.

Text localization techniques can be grouped into region-based, connected component (CC)-based [1] and hybrid methods [2]. Region-based techniques employ a sliding window to look for image text with the use of machine learning techniques for text identification. Sliding window based methods tend to be slow due to multi scale processing of images. A new text detection algorithm extracts six dissimilar classes of text features. Modest AdaBoost classifier is used to recognize text regions based on text features [3]. CC-based methods group extracted candidate characters into text regions with similar geometric features. CC-based methods are demanding to apply additional checks for eliminating false positives. To find CCs, stroke width for every pixel is computed to group neighboring pixels. These CCs were screened and grouped into text regions [4]. Pan et al. [2] proposed hybrid method that exploits image regions to detect text candidates and extracts CCs as candidate characters by local binarization. False positive components are eliminated efficiently with the use of conditional random field (CRFs) model. Finally, character components are grouped into lines/words. Recently, Yin et al. [5] extracted maximally stable extremal regions (MSERs) as letter candidates. Non-letter candidates are eliminated using geometric information. Candidate text regions are constructed by grouping similar letter candidates using disjoint set. For each candidate text region, vertical and horizontal variances, color, geometry and stroke width are extracted to identify text regions using Adaboost classifier. Besides, MSER based method is the winner of ICDAR 2011 Robust Reading Competition [6] with promising performance.

In this paper, we proposed a new robust and accurate MSER based approach to localize text in scene images. Firstly, MSER is extracted as candidate characters. Some of the non-character candidates are pruned based on total number of pixels. Second, Bayesian network scores for each candidate characters feature are obtained using K2 algorithm [7]. Third, Bayesian Logistic Regression Classifier (BLRC) [8, 9] is built from these score based features of candidate characters and used to identify text regions. Also, a posterior probability based adaptive threshold is used to filter out non-character candidates. In order to use ICDAR 2013 [10] competition dataset, candidate characters are grouped into words by using effective text localization approach [5]. The output of each step is shown in Figure 1. Our method is tested on the ICDAR 2013 test dataset. The experimental results show the remarkable performance in terms of accuracy.

The remainder structure of this paper is organized as follows. Section II explains the text localization method in perspective of candidate character extraction using MSER, Bayesian network score learning, Bayesian logistic regression classifier with an adaptive threshold in subsequent subsection and grouping of candidate character regions. Section III briefly describes the comparative performance of different text localization algorithm experimental results. In last section, the paper has been concluded.

2. K2 SCORES AND ADAPTIVE THRESHOLD BASED TEXT DETECTION METHOD

Text localization method is divided into four parts: candidate character extraction using MSER, Bayesian network scores for each candidate character, classification of candidate characters using Bayesian network scores with the use of adaptive threshold and group candidate characters by eliminating false positives.

2.1 Candidate Character Extraction with MSER

The digital camera-based images may have distortions [11] with motivation by [6, 12] to localize text from scene images. Our method extracts MSER from scene image as candidate characters region. MSER was proposed by Matas et al. [13] to catch matching between different viewpoints of images. MSER is a well-known and best reported robust method against scale and light variation [14]. Each extracted MSER is resized (*f X f*) to perform binarization using an adaptive threshold. Consider an image *I*, and a set of resized extracted MSERs E_{rc} referred as candidate character region, with *f* number of features. Binarization is performed to eliminate extra pixels of candidate characters using an adaptive threshold as given in equation 1. Additional check is applied on the number of pixels of all candidate character regions to filter out as non-character regions. The purpose of additional constraint is applied on candidate binary character region to minimize the processing time of learning Bayesian network scores using K2 algorithm [7].

$$\tau = \frac{1}{k}\sum_{r=1}^{k}\left(\sum_{c=1}^{k} CCR_{rc} \Big/ k\right) \rightarrow (1)$$

Where, CCR_{rc} represent candidate character region before application of binarization, and *k=25* is the number of features.

$$CBR = CCR > \tau \rightarrow (2)$$

Where, CBR represent the candidate binary character region. A candidate binary character region is discarded as non-character if it has less than or equal to 30 pixels or greater than or equal to 600 pixels.

2.2 K2 Algorithm-based Bayesian Network Scores

Bayesian network is an extensively used model for data analysis. Bayesian networks [7] exploit a set of features of candidate character binary region of an image to measure probabilistic relationship graphically among features. The optimal determination of Bayesian network structure of a given data is NP-hard problem [15]. K2 algorithm [7] is one of the widely used efficient heuristic solutions to learn structural relationship dependencies among features according to a specified order. The initial feature in a given order has no parents. K2 then adds incrementally parents for a current feature by increasing the score in resulting structure. When no predecessor features addition for a current feature as parent can increase the score, K2 stops adding parents to feature. As ordering of features are known beforehand, so search space is minimized make it efficient as well as parents of features can be chosen independently [16]. Consider $F = \{f_1, f_2, f_3 \ldots f_k\}$ are the features of candidate binary character region of scene image. All the candidate binary character regions features are ordered from left to right for simplicity. To find scores of each feature, scoring function of K2 algorithm for i^{th} features is presented by equation 3. The final scores of all features in a network structure are obtained by multiplying the individual score of features of candidate binary character.

$$g(i,\pi_i) = \prod_{i=1}^{k}\prod_{j=1}^{q_i}\frac{(r_i-1)!}{(N_{ij}+r_i-1)!}\prod_{z=1}^{r_i}\alpha_{ijk}\,! \rightarrow (3)$$

Where, $g(i,\pi_i)$ represent the Bayesian network score using K2 algorithm for each feature of candidate binary character region of an image. However $g(i,\pi_i)$ (*where, i=1)*, has no parents and zero score. Therefore, we omit the first feature of candidate binary character region in all the test set ICDAR 2013 images. The purpose of omitting initial feature of candidates' binary characters is to produce point to point product of all Bayesian network K2 scores as given in equation 4.

$$K2F_{score} = [g(i-1, \pi_{i-1})] \circ [g(i-1, \pi_{i-1})]^T \rightarrow (4)$$

2.3 Classification of Bayesian Network Scores

Bayesian Logistic Regression is an approach that models a relationship between a Bayesian network scores and automatically labeled candidate character regions, in which a statistical analysis is taken under Bayes rule. The application of Bayesian Logistic Regression can be applied in pattern recognition to classify Bayesian network scores of candidate character regions for two or more classes. The most important advantage of Bayesian Logistic Regression is its dominating classification accuracy while estimating a probabilistic relationship between Bayesian network scores and labeled candidate character regions.

Suppose *L = {0, 1}* be the labels of Bayesian network scores of candidate character regions *R*. The log-likelihood ratio $\ln\left(P(L=1/R,\omega)/P(L=0/R,\omega)\right)$ is assumed to be linear in R, such that conditional likelihood for *L=1* is given by the sigmoid function. $P(L=1/R,\omega) = 1/\left(1+e^{-\omega^T R}\right) = \delta(-\omega^T R)$. Similarly, $P(L=0/R,\omega) = 1-P(L=1/R,\omega) = 1/\left(1+e^{\omega^T R}\right)$ such that $P(L/R,\omega) = \delta(L\omega^T R)$.

Let training set τ composed of Bayesian network scores of an image candidate character regions *R* and their labels *L, i.e.* $\tau = \{\mathrm{L}, \mathrm{R}\}$ are input parameters to Bayesian Logistic Regression Classifier. Conditional probability $P(\omega/\tau)$ and priori probability $P(\omega)$ are used to compute posterior probability. As sigmoid likelihood data does not permit a conjugate-exponential prior to find posterior analytic expression. The quadratic approximation ω is used exponentially, such that conjugate-Gaussian prior with parameterization of α as hyper-parameter. The prior of this approximation and modeling of conjugate Gamma distribution can be represented as given by equations (5) and (6).

$$P(\omega/a) = \left(\frac{\alpha}{2\pi}\right)^{1/2} e^{\frac{\alpha}{2}\omega^T\omega} \rightarrow (5)$$

$$P(a) = \left(\frac{1}{\Gamma a_0}\right) b_0^{a_0} \alpha^{a_0-1} e^{-b_0\alpha} \rightarrow (6)$$

FIGURE 1: Text Detected in ICDAR 2013 Scene Images.

2.4 Posterior Probability-based Adaptive Threshold

To filter non-character regions, an adaptive threshold is computed using posterior probabilities computed by Bayesian Logistic regression in previous subsection. For simplicity, we use geometric mean and standard deviation of $P(\omega/a)$ to filter non-character candidate regions. An adaptive threshold is computed by equation 7.

$$T = \sqrt[n]{\prod_{i=1}^{n} P_i(\omega/a)} - 3\left(\frac{n\sum_{i=1}^{n} P_i(\omega/a)^2 - (\sum_{i=1}^{n} P_i(\omega/a))^2}{n^2}\right) \rightarrow (7)$$

The entire candidate character region of an image classified through Bayesian network scores are eliminated if their posterior probability is less than or equal to adaptive threshold value $\mathbb{T}$.

2.5 Candidate Character Regions Filtering and Grouping

Aspect ratio [12] is a cascade of filter that can be used to eliminate too wide or too narrow candidate regions. For example, W_i is a width and H_i is the height of candidate region. In this way, candidate character regions are further reduced after adaptive threshold.

$$l_{min} < \frac{W_i}{H_i} < l_{max} \rightarrow (8)$$

In all our experiments, adaptive thresholds are adapted empirically. However, aspect ratio threshold (l_{min}=*0.1* and l_{max}=*10)* are fixed empirically throughout the experiment. Now, candidate character regions are grouped using effective text localization method proposed by Yin et al. [5]. This technique constructs candidate regions with the use of geometric information and group similar candidate character region using disjoint sets. Also, vertical and horizontal variances, color, geometry and stroke width features of each candidate character regions are extracted to build Adaboost classifier for identification of text regions. In contrary to [5], we used Bayesian network based scores and Bayesian logistic regression classifier with additional adaptable and fixed constraints. The output text regions after grouping are presented in Figure 1.

3. EXPERIMENTAL RESULTS

The Robust Reading Competition was organized by ICDAR 2013 [6, 10, 17] with an objective to deliver performances of different algorithms as a benchmark. The evaluation system used by ICDAR 2013 demonstrated object level precision and recall of detecting algorithms according to their superiority of detection. Our text localization method used test dataset of ICDAR 2013 Robust Reading Competition Challenge 2 Task 1 [10]. The whole dataset consisted of 233 images containing text of different fonts and colors with different backgrounds.

3.1 Performance Evaluation of Text Localization System

ICDAR 2013 dataset was used in most recent text detection competition as a benchmark. To provide a baseline comparison, we tested our method on the publicly available dataset in [10]. The output of an algorithm is a set of rectangles designating bounding boxes for detected words in scene images and referred as *estimate* as given by Figure 1. A set of ground truth boxes is provided in the dataset which referred to *targets*. The match of an estimate and target by a text detecting algorithm can be defined as the area of intersection between two rectangles divided by the area of minimum bounding box containing both rectangles. The identical rectangles have value one and zero for rectangles with no intersection. Therefore, a closest match was found between the estimated and targeted rectangles to measure the performance. For this purpose, precision, recall and harmonic mean are used to recognize the matching of intersecting rectangles. Precision is the ratio of the number of matched detected rectangles to the total number of matched and unmatched detected rectangles. For example, M represents the matched detected rectangles and U represents the unmatched detected rectangles, Precision can be defined as given by equation 9.

$$Precision = \frac{M}{M+U} \rightarrow (9)$$

Similarly, recall can be defined as the matched detected rectangles to the total number targeted rectangles. For example, M are the matched and T are the targeted rectangles, recall can be defined as given by equation 10.

$$Recall = \frac{M}{M+T} \rightarrow (10)$$

Finally, F-measure can be defined as the harmonic mean of precision and recall and can represented as given by equation 11.

$$F-Measure = \frac{2}{\frac{1}{Precision}+\frac{1}{Recall}} \rightarrow (11)$$

Consequently, performances of our localization method with some of the published algorithms in ICDAR 2013 competition are presented in Table 1. Our text localization method shows a competitive recall *62.37*, precision *84.97* and a *71.94* harmonic mean, which is competitive with the leading methods reported by [10]. However, our text localization method performs better than the ICDAR 2011 Robust Reading Competition methods reported by [6].

Method	Recall	Precision	HMean
USTB_TexStar [27]	66.45	88.47	75.89
Text Spotter [20], [21], [22]	64.84	87.51	74.49
CASIA_NLPR [23], [24]	68.24	78.89	73.18
Text_Detector_CASIA [25], [26]	62.85	84.70	72.16
Our Method (LocaTeXT)	**62.37**	**84.97**	**71.94**
I2R NUS FAR	69.00	75.08	71.91
I2R NUS	66.17	72.54	69.21
TH-TextLoc	65.19	69.96	67.49
Text Detection [18], [19]	53.42	74.15	62.10
Baseline	34.74	60.76	44.21
Inkam	35.27	31.20	33.11

TABLE 1: Performance (%) Comparison of Text Detection Methods on ICDAR 2013 Dataset.

4. REFERENCES

[1] Y.-F. Pan, X. Hou, and C.-L. Liu. A hybrid approach to detect and localize texts in natural scene images. IEEE Transactions on Image Processing, 20(3):800 –813, March 2011.

[2] Y.-F. Pan,, X. Hou, and C.-L. Liu. "A hybrid approach to detect and localize texts in natural scene images." Image Processing, IEEE Transactions on 20, no. 3 (2011): 800-813.

[3] J.-J. Lee, P.-H. Lee, S.-W. Lee, A. Yuille, and C. Koch. Adaboost for text detection in natural scene. In ICDAR 2011, pages 429 –434, September 2011.

[4] B. Epshtein, E. Ofek, and Y. Wexler. Detecting text in natural scenes with stroke width transform. In CVPR 2010, pages 2963 –2970, June 2010.

[5] X. Yin, X-C Yin, H-W. Hao, and K. Iqbal. "Effective text localization in natural scene images with MSER, geometry-based grouping and AdaBoost." In Pattern Recognition (ICPR), 2012 21st International Conference on, pp. 725-728. IEEE, 2012.

[6] A. Shahab, F. Shafait, and A. Dengel. ICDAR 2011 robust reading competition challenge 2: Reading text in scene images. In ICDAR 2011, pages 1491 –1496, September 2011.

[7] G. F. Cooper and E. Herskovits, A Bayesian method for the induction of probabilistic networks from data.Machine Learning, vol. 9, no.4, pp. 309-347, 1992.

[8] K. Iqbal, X.-C. Yin, H.-W. Hao, X. Yin, H. Ali, (2013), Classifier Comparison for MSER-Based Text Classification in Scene Images, In Neural Networks (IJCNN), The 2013 International Joint Conference on (pp. 1-6). IEEE.

[9] Hosmer, David W.; Lemeshow, Stanley (2000). “Applied Logistic Regression” (2nd ed.). Wiley. ISBN 0-471-35632-8.

[10] D. Karatzas, F. Shafait, S. Uchida, M. Iwamura, S. R. Mestre, J. Mas, D. F. Mota, J. A. Almazan, and L. P. de las Heras. "ICDAR 2013 Robust Reading Competition." In Document Analysis and Recognition (ICDAR), 2013 12th International Conference on, pp. 1484-1493. IEEE, 2013.

[11] X.-C. Yin, H.-W. Hao, J. Sun, and S. Naoi. Robustvanishing point detection for MobileCam-based documents. In ICDAR 2011, pages 136 – 140, September 2011.

[12] C. Merino-Gracia, K. Lenc, M. Mirmehdi, A head-mounted device for recognizing text in natural scenes, Camera-Based Document Analysis and Recognition, pages 29--41, 2012.

[13] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In British Machine Vision Conference 2002, volume 1, pages 384–393, 2002.

[14] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool. A Comparison of Affine Region Detectors. International Journal of Computer Vision, 65(1):43–72, November 2005.

[15] D. M. Chickering, Learning Bayesian networks is NP-complete. In D. Fisher and H.J. Lenz, editors, Learning from Data: Articial Intelligence and Statistics V, Springer-Verlag, pp. 121-130, 1996.

[16] N. Friedman, D. Koller, Being Bayesian About Network Structure: A Bayesian Approach to Structure Discovery in Bayesian Networks, Machine Learning 50 (1-2) (2003) 95-125.

[17] D. Karatzas, S. Robles Mestre, J. Mas, F. Nourbakhsh, P. Pratim Roy , "ICDAR 2011 Robust Reading Competition - Challenge 1: Reading Text in Born-Digital Images (Web and Email)", In Proc. 11th *International Conference of Document Analysis and Recognition*, 2011, IEEE CPS, pp. 1485-1490.

[18] J. Fabrizio, B. Marcotegui, and M. Cord, “Text segmentation in natural scenes using toggle-mapping,” in Proc. Int. Conf. on Image Processing,2009.

[19] J. Fabrizio, B. Marcotegui, M. Cord, “Text detection in street level images,” Pattern Analysis and Applications, 2013, Volume 16, Issue 4, pp 519-533.

[20] L. Neumann and J. Matas, “A method for text localization and recognition in real-world images,” in Proc. Asian Conf. on Computer Vision, 2010, pp. 2067–2078.

[21] L. Neumann, and J. Matas. Real-time scene text localization and recognition. in Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. 2012., pp. 3538–3545.

[22] L Neumann, J Matas, “On combining multiple segmentations in scene text recognition,” in Proc. Int. Conf. on Document Analysis and Recognition, 2013. International Conference on Document Analysis and Recognition (ICDAR 2013), Washington D.C., 2013.

[23] Y.-M. Zhang, K.-Z. Huang, and C.-L. Liu, “Fast and robust graph-based transductive learning via minimum tree cut,” in Proc. Int. Conf. on Data Mining, 2011.

[24] C.-L. L. B. Bai, F. Yin, “Scene text localization using gradient local correlation,” in Proc. Int. Conf. on Document Analysis and Recognition, 2013.

[25] C. Shi, C. Wang, B. Xiao, Y. Zhang, S. Gao, and Z. Zhang, “Scene text recognition using part-based tree-structured character detections,” in Proc. Int. Conf. on Computer Vision and Pattern Recognition, 2013.

[26] C. Shi, C. Wang, B. Xiao, Y. Zhang, and S. Gao, “Scene text detection using graph model built upon maximally stable extremal regions,” Pattern Recognition Letters, vol. 34, no. 2, pp. 107–116, 2013.

[27] X.-C Yin, X. Yin, K. Huang, and H.-W. Hao Robust Text Detection in Natural Scene Images., IEEE Transactions on Pattern Analysis and Machine Intelligence, preprint, 2013.

Segmentation of Tumor Region in MRI Images of Brain using Mathematical Morphology

Ashwini Gade *ashwinigade036@gmail.com*
Department of Electronics and Telecommunication
MPSTME, NMIMS
Mumbai, India.

Rekha Vig *rekha_vig@yahoo.co.in*
Department of Electronics and Telecommunication
MPSTME, NMIMS
Mumbai, India.

Vaishali Kulkarni *vaishalikulkarni6@yahoo.com*
Department of Electronics and Telecommunication
MPSTME, NMIMS
Mumbai, India.

Abstract

This paper introduces an efficient detection of brain tumor from cerebral MRI images. The methodology consists of two steps: enhancement and segmentation. To improve the quality of images and limit the risk of distinct regions fusion in the segmentation phase an enhancement process is applied. We applied mathematical morphology to increase the contrast in MRI images and to segment MRI images. Some of experimental results on brain images show the feasibility and the performance of the proposed approach.

Keywords: Cerebral MRI Images, Mathematical Morphology, Tumor.

1. INTRODUCTION

Many research work centres are focused on the dynamic growth in the region of cerebral cancer diagnosis because of the fact that cerebral cancer is spreading among world population [1]. We have observed in the US, nearly 3000 children are diagnosed with brain tumors. Almost 50% will die within few years, making it the most fatal cancer among children [2]. It's associated with neurological disabilities, retardation and psychological problems and increased risk of death. Despite over all increases in incidences and death from cerebral cancer in the general world population; Africans are more likely than other patients to die of the disease. In Tunisia, for instance, the cancers mortality is responsible for 14.8% of deaths among the elderly. They represent the second leading cause of death after cardiovascular diseases [3]. Due to its negative effects on affected people, the cancer diseases constitutes a high burden on national economy and a source of suffering for the family as well as the society [3].

Imaging plays a central role in the diagnosis of brain tumors. Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) are the high resolution techniques locate brain tumor. The information obtained will influence the treatment a patient will receive. The most widely used clinical diagnostic and research technique is MRI. It's an efficient medical imaging technique that has different methods (T1,T2, ARM, …) having each particular property and an effective way that enables to clarify the various tissues and to obtain a 2D, 3D and even 4D sight (3D+T) of a part of the body, in particular of the brain. It's based on the principal of nuclear magnetic resonance (NMR). Due to various sequences various tissues with high contrast can be observed [4].

It is difficult in medical imaging diagnosis systems to separate cells and their nuclei from the rest of the image content [5]. As the process of separation is very important, much attention of the expert diagnosis system has to be paid to the segmentation stage. Segmentation is a crucial step in image processing tasks. In literature, there are different definitions of segmentation. Haralick, Zhang and Freixenet [6] summarize the segmentation definitions found in literature. From general point of view segmentation is the partitioning of an image into a set of homogeneous and significant regions having a single label and common or similar properties. Many algorithms were thus proposed during the last decades. They are based on various approaches: contour, region and texture.

In the image processing analysis, enhancement process improves the quality of images since the majority of images dealt with have low contrast and is a pre-treatment of segmentation process.

The paper is organized into IV Sections. Section -I is about Brain Tumor. Section-II presents the proposed method of tumor detection from cerebral a MRI image which consists of two steps: Enhancement and Segmentation. Enhancement Process is applied by mathematical morphology and segmentation process is also applied by mathematical morphology with increased size of structuring element. Experimental results are reported in section-III. Finally concluding remarks are drawn in section- IV.

2. PROPOSED METHOD

The proposed method of detection and segmentation of the tumor from the cerebral MRI Images is summarized in Figure: 1

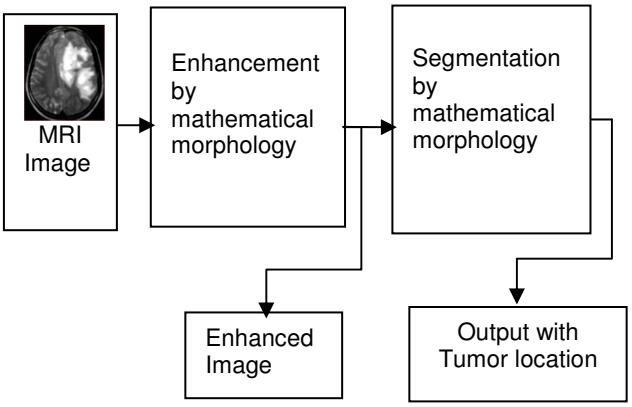


FIGURE 1: Steps of Proposed Method

2.1 Enhancement

Enhancement in medical imaging is to make image clearer, smoother and improves the quality of given image. It can be accomplished by removing noise, enhancing contrast, emphasizing edges and modifying shapes. Figure: 2 illustrate the enhancement process.

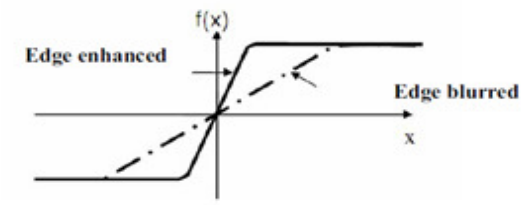


FIGURE 2: The Enhancement Process.

In blurred images, the intensity values of pixels on both sides of edge changes linearly. After performing image enhancement, there is drastic change in the intensity level on both sides of edge.

The enhancement techniques have widely applied in the field of radiology, where the subjective quality of images is important for diagnosis. Many general purpose tools for enhancement have developed and applied to medical images [8]. They include histogram modification, mean filters, Gaussian filters, linear shift invariant filters and morphological filters. For contrast enhancement based on mathematical morphology theory there are two methods: the first is Top-Hat which deals with images after the segmentation process. This algorithm enhances the edges of the segmented region of interest. The second algorithm deals with the contrast of the original image to enhance the segmentation process.

Here, we consider the Mathematical morphology for contrast enhancement. A detailed theory of mathematical morphology is provided in [9]. Experimentally, we found that structuring element as a disk of radius 7 yields the best performance of the morphological algorithm.

Image enhancement has applied successfully to different fields such as medical, industry and military fields [7].

2.1.1 Mathematical Morphology

Mathematical morphology [10, 11] is a relatively new approach to image processing and analysis. This approach is based on set-theoretic concepts. In morphology objects present in an image are treated as sets. As the identification of objects and object features directly depend on their shape, mathematical morphology is becoming an obvious approach for various machine vision and recognition processes.

In morphology, the objects in an image are considered as set of points and operations are defined between two sets: the object and the structuring element (SE). The shape and the size of SE are defined according to the purpose of the associated application. Basic morphological operations are erosion and dilation. Other operation like opening (closing) is sequential combination of erosion (dilation) and dilation (erosion). Golay logic processor [12], Leitz Texture Analysis System TAS [13], the CLIP processor arrays [14] and the Delft Image Processor DIP [15] have started including morphological processor. Initially morphology dealt with binary images only and basic operation was dilation and erosion also known as Minkowski addition and subtraction [16] respectively.

2.1.1.1 Tophat Transformation

The *tophat transformation* [17] provides an excellent tool for extracting bright (respectively dark) features smaller than a given size from an uneven background. Gray-scale opening helps to remove the brighter areas from an image that is features that cannot hold the structuring element. Subtracting the opened image from the original one yields an image where the features that have been removed by opening clearly stand out. Similar thing is true for closing operation also. That means using a *closing* instead of an *opening* and subtracting the original image from the closed one helps us extract dark features from a brighter background. Let us call it a *black tophat* transformation as opposed to *white tophat* transformation in case of opening. Suppose the structuring element used in both opening and closing is a disk, or, more specifically, a discrete approximation of disk. Therefore, the bright tophat transformation decomposes an image into two parts.

An ordered sequence of morphological tophat filtering through opening (closing) of the image with a disk structuring element with different size extracts bright (dark) features from the image. These features resulting from the tophat transformation of the image can be amplified selectively to achieve local contrast stretching. Based on this we propose an enhancement method in the following section.

2.1.1.1.2 Enhancement using Mathematical Morphology

The bright Tophat tranformation decomposes an image into two parts. This may be expressed in terms of gray scale morphological operators as

$$g(r,c) = \underbrace{(goB)(r,c)}_{\text{Part-1}} + \underbrace{[g(r,c) - (goB)(r,c)]}_{\text{Part-2}} \quad (1)$$

Where *B* is a disk in discrete domain. Let us call part 1 of "(1)" the *base image* with respect to *B*(*r*, *c*). And let us call part 2 of "(1)" the *feature image* at size of *B* as it contains all the bright features of *g*(*r*, *c*).The feature image [i.e. part 2 of "(1)"] gives a measure of local contrast in the original image due to presence of bright features. *k* is a global amplification factor and is greater than one. So this transformation makes bright features brighter and, thus, improves the quality of the image. And $k1 > k2 > k3$…, since we know that smaller the size of a bright feature, more should be its intensity for detectibilty. It not give the better result therefore after that some morphological operations are applied on the image. The basic purpose of the operations is to show only that part of the image which has the tumor that is the part of the image having more intensity.

The proposed method is summarized in figure 3.

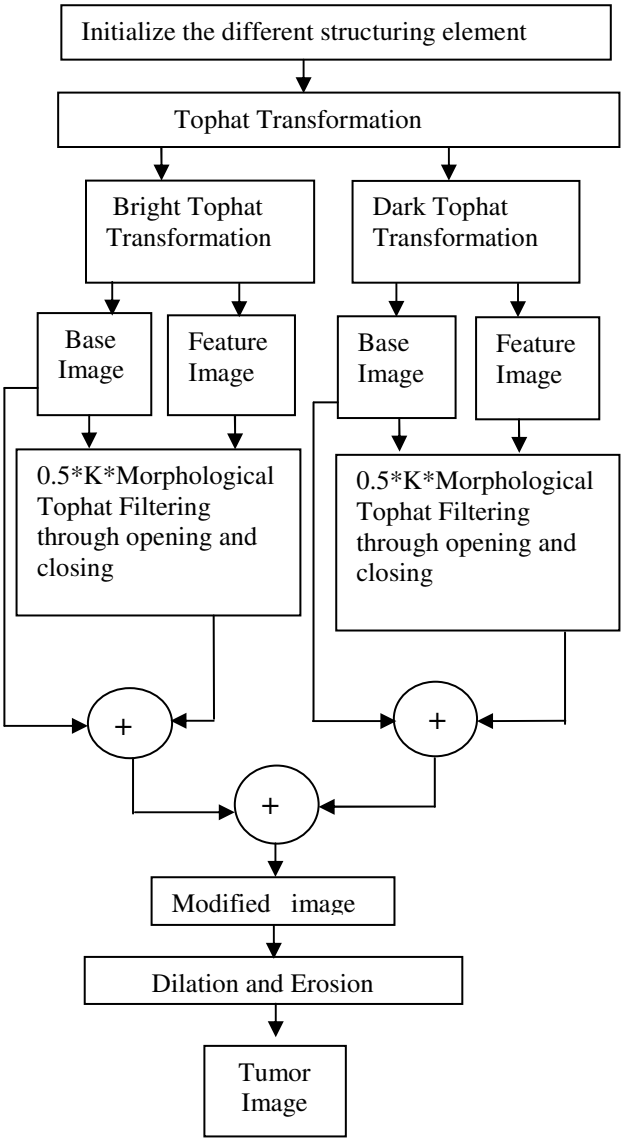


FIGURE 3: Steps of The Proposed Algorithm.

The basic commands used in this step are strel, imerode and imdilate, Imerode: It is used to erode an image. Imdilate: It is used to dilate an image.

2.2 Segmentation

After pre-processing phase, we adopt a segmentation algorithm. Segmentation is the partitioning of an image into homogeneous regions (Spatially connected groups of pixels called classes, or subsets) with respect to one or more characteristics or features. The organ's anatomy, the patient's position of catching image and other medical images characteristics add more difficulties to the problem of image segmentation. This explains the variety of segmentation methods. In literature there exist two types of segmentation techniques: edge based segmentation and region based segmentation.

Edge based segmentation aims at finding object boundaries and segmenting regions enclosed by the contours. Edge based techniques include Roberts, Prewitt, Robinson, Kirsch, Laplacian and Frein-Chen filters [18]. They prove to be computationally fast and don't require prior information about the image content. However a drawback of the edge approach is that the edges do not enclose the object completely.

In region-based techniques, segmentation is applied by identifying all pixels that belong to the object based on the intensity of pixels. They include region growing, watershed algorithm and thresholding [21]. In the medical image segmentation field, region growing technique can be applied in kidney segmentation, cardiac images, extraction of brain surface etc. The capability of generating joined regions and appropriately segmenting regions having matching property are the benefits of this segmentation method. One of the drawbacks of this method is that dissimilar starting points may not result growing into identical regions. In addition to this, since outcome of region growing is dependent on homogeneity criterion, failure in correctly choosing criterion may result in adjacent areas or regions not belonging to the object of interest [22].

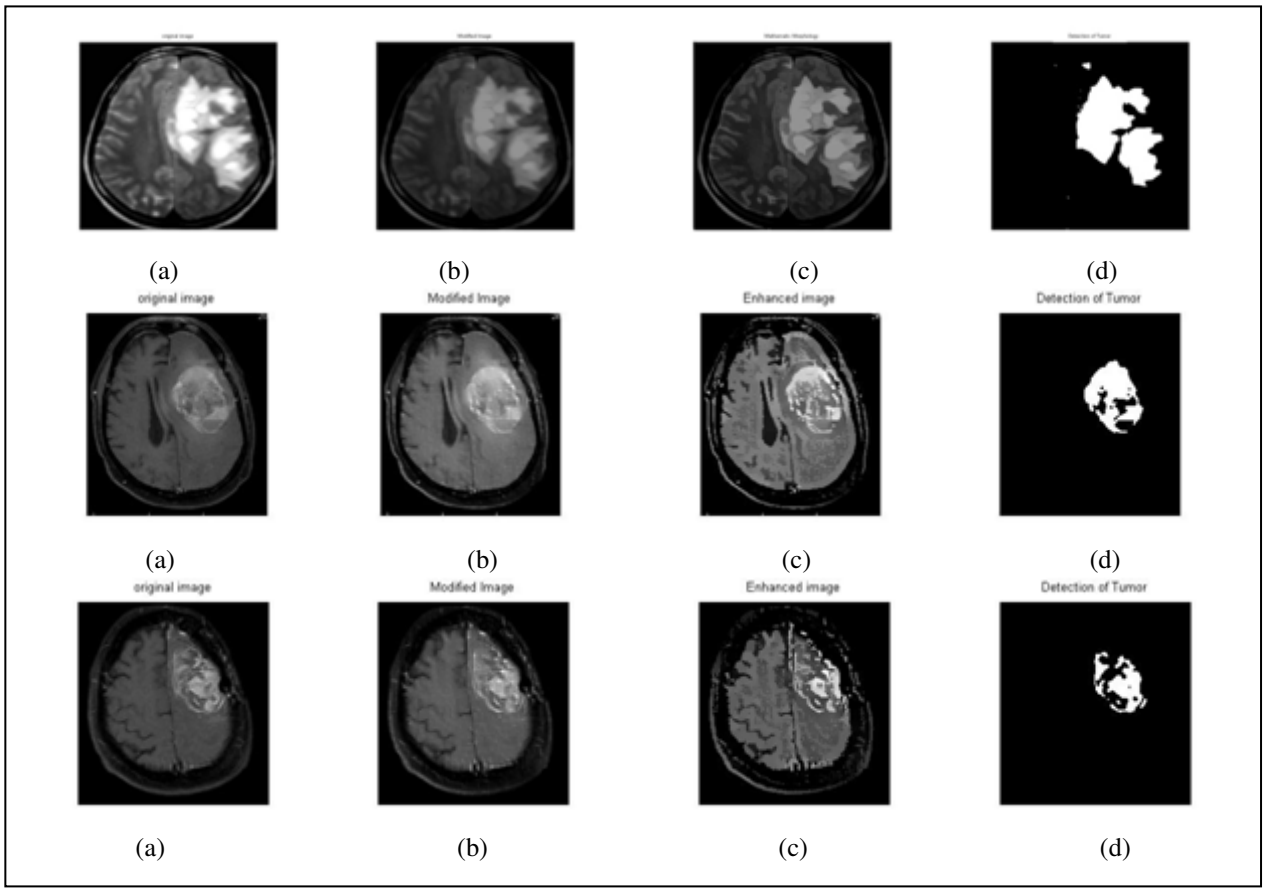


FIGURE 4: Segmented Tumor.

Here we consider mathematical morphology same as implemented in image enhancement for segmentation of tumors. We applied several tests with different type and size of structuring element and come to conclude that Disk of radius 160 is the best type in image application.

3. EXPERIMENTAL RESULTS

The proposed algorithm has been tested on a set of biomedical images and the results have been compared with that of other methods. Here we present detailed study with example images containing MRI scan of human brain. The application of proposed method is illustrated in the

figure 4. figure (a) is the original MRI scan images; figure (b) Shows the modified image where the dark and bright Features are balanced; figure (c) is the output after mathematical morphology using Disk of size 7 and then figure (d) is the resultant segmented image after mathematical morphology using Disk with increased size of SE of figure (c)

The results of segmented tumor using mathematical morphology preserves the qualitative and quantitative tumor region better as compared to edge and region based segmentation.

4. CONCLUSION AND FUTURE SCOPE

In this paper, an efficient detection of brain tumor has been introduced. It is based on mathematical morphology. The algorithm reduces the extraction steps and improves the quality of images. Due to the enhancement process applied the algorithm limits the risk of distinct region fusion in the segmentation phase. The enhancement and segmentation are performed by applying the mathematical morphology. The results show that this method can improve image quality and have good visual effect. The tumor area is effectively detected by using this algorithm which paves the way for the expert to decide the degree of malignancy or aggressiveness of a brain tumor. To classify a brain tumor as “benign” or “malignant” will be the subject of future research.

5. REFERENCES

[1] M. Hrebién, P. Stéc, T. Nieczkowski and A. Obuchowicz. "Segmentation of breast cancer Fine needle biopsy cytological images". International Journal of Applied Mathematics and Computer Science, 2008, Vol.18, No.2, 159–170, 2008. http://www.aacr.org/home/public media/for-the-media/factsheets/organ-site-fact/sheets /brain-cancer.asp.

[2] http://www.lapresse.tn/archives/archives270205/ societe/grandes.html.

[3] S. MIRI, "Segmentation des structures Cérébralesen IRM: intégration decontrainte topologiques". Master, Universite Louis Pasteur Strasbourg, 2007.

[4] M. Lee and W. Street "Dynamic learning of shapes for automatic object recognition", Proceedings of the17th Workshop Machine Learning of Spatial Knowledge, Stanford, CA, pp. 44-49, 2000.

[5] S. Chabrier, H. Laurent, Ch. Rosenberger and A. rankotomamonjy. "Fusion de critères pour l’évaluation de résultats de segmentation d’images". Actes du 20ème Colloque GRETSI, 2005.

[6] C. Wright, E.J. Delp, N. Gallagher, “Morphological based target enhancement algorithms,” Multidimensional Signal Processing Workshop, 1989.

[7] H.D. Cheng, Xiaopeng Cai, Xiaowei Chen, Liming Hu, Xueling Lou, "Computer-aided detection and classification of microcalcfications in mammograms: a survey," Pattern Recognition. vol. 36, pp. 2967 – 2991, 2003.

[8] P. Soille, "Morphological Image Analysis, Principals and Applications", Springer, 1999.

[9] G.M. Matheron, Random Sets and Integral in Geometry, Wiley, New York, 1975.

[10] J. Serra, Image Analysis Using Mathematical Morphology, Academic Press, London, 1982.

[11] M.J. Golay, Hexagonal parallel pattern transformations, IEEE Trans. Comput. C-18 (1969) 733-740.

[12] J.C. Klein, J. Serra, The texure analyzer, J. Microscopy 95(1977) 349-356.

[13] M. Duff Parallel processors for digital image processing, in: P. Stucki (Ed.), Advances in Digital Image Processing, Plenum, New York 1979.

[14] P.F. Leonard, Pipeline architechture for real Time Machine vision, Proceedings of the IEEE Computer Society Workshop on Computer Architecture for Pattern Analysis and Image Database Management, 1985, pp. 502-505.

[15] H. Minkowski, Volume and oberflache, Math. Ann. 57 (1903) 447-495.

[16] F. Meyer, Contrast feature extraction, in: J.L Chermant (Ed.), Quantitative Analysis of Microstructures in Material Sciences, Biology and Medicine, Riederer Verlag, Stuttgart, Germany, 1978.

[17] R.M. Haralick, L.G. Shapiro, " Survey: Image Segmentation techniques," Comp. Vision Graph Image Proc., Vol. 29, pp. 100-132, 1985.

[18] J.S.Weszka, "A Survey of threshold selection techniques", Computer Graphics and Image Proc., vol. 7, pp. 259-265, 1978.

[19] G Mallat, “A theory for multiresolution signal decomposition: The wavelet representation”, IEEE transactions on pattern analysis and machine intelligence, VOL II. No 7 11(7):674-693 Juillet 1989.

[20] Dzung L. Pham, Chenyang Xu, Jerry L. Prince;"A Survey of Current Methods in MedicalMedical Image Segmentation" Technical Report JHU / ECE 99-01,Department of Electrical and Computer Engineering. The Johns Hopkins University, Baltimore MD 21218, 1998.

[21] Issac N. Bankman, “Handbook of medical image processing and analysis”, Second ediion, Academic press, USA, 2008.

Research on Image Classification Model of Probability Fusion Spectrum-Spatial Characteristics Based on Support Vector Machine

Zheng Zhang *bestonezheng@hotmail.com*
Geomatics
Xi'an University of Science and Technology
Xi'an, 710054, China

Xiaobing Huang *happyxiaobing@hotmail.com*
Geomatics
Xi'an University of Science and Technology
Xi'an, 710054, China

Hui Li *free.h_li@hotmail.com*
Aerial remote sensing department
The Third Surveying and Mapping Institute of GuiZhou Province
Guiyang, 550004, China

Abstract

For insufficient information of imaging spectrum with high spatial resolution, detailed imaging information, reduction of mixed pixels, increase of pure pixels and problems of image characteristic extraction and model classification produced from this, we provide a classifier model of a united spectrum-spatial multi-characteristic based on SVM, and use this model to finish the image classification. The model completely uses the multi-characteristic information, and overcomes the over-fitting problems produced by accumulating high-dimensional characteristics. The model includes three classifications of spectrum-spatial characteristics, namely spectral characteristics-spectral characteristic of multi-scale morphology, spectral characteristics-physical characteristics of underlaying surfaces of multi-scale morphology and spectral characteristics-features spatial extension characteristics of multi-scale morphology. Firstly the three classifications of spectrum-spatial characteristics are classified through SVM, then carries out the probability fusion for the classification results based on the pixels to obtain the final image classification results. This article respectively uses WorldView-2 image and ROSIS image to experiment, and the results show that the model has better classification effect compared with VS-SVM algorithm.

Keywords: High Spatial Solution, Spectral - Space Characteristics, Multi-Scale Morphological Sequence, SVM.

1. INTRODUCTION

In recent years, with the continuous improvement of high spatial resolution of images, applications of image spatial characteristics in classification are promoted. For the classification of urban hyperspectral data with high spatial resolution, Benediktsson provides extended morphology sequence（EMP）, which is to adopt the images of which principal components of the hyperspectral images are changed as the basic images for creating morphological sequences, and then apply the principal components of the spectrum and their EMP to a neural network classifier[3]. Because original methods do not completely consider the spectral information in the data, Fauvel changes those methods by combining the hyperspectral information with EMP to form feature vectors[4]. Huang and Zhang carries out comparative study on space approach for urban surveying and mapping by using the hyperspectral images with high

spatial resolution of Pavia in the north of Italy, extracts and classifies the features by using different spectrum-spatial characteristics， including morphological sequences (MPs), gray level co-occurrence matrix (GLCM), pixel shape index (PSI), feature-oriented classification method based on fractal network and multi-scale mean-shift algorithm; the results show that precision of pure spectral classification can be effectively improved by combining the spectrum-spatial characteristics[1][5][9].

In methods for classifying and processing the high-resolution data by combining various spatial information, the method which uses the vector superposition of each characteristic and classifies through the support vector machine has better result, but it also has following shortcomings: firstly,during the process for forming the spatial characteristics, different parameters, such as sizes, scales and directions, are needed to be set, which may form a high-dimensional characteristic space, so that the calculation complexity is increased; secondly, with the continues increase of characteristic space, when the characteristic space dominates all the characteristic spaces, the spectral characteristics are submerged to some extent, which causes the identification error, so in order to improve the classification precision by using the spectrum-spatial characteristics and overcome the shortcomings of VS-SVM classification, we provide the image classification model of probability fusion spectrum-spatial characteristics of SVM, and apply the model to the high-resolution image classification. The model completely uses the spectrum-spatial characteristics and perfectly prevents the over-fitting problems produced by accumulating high-dimensional characteristics.

2. RELATED WORK

In the images with high spatial resolution, detail ground information can be completely exhibited; the inner parts of the same classifications of the features have variability, even the homogeneous areas have obvious spectrum differences, so that the interclass variance enlarges, and the separability of spectrum areas are reduced. When processing the images, we not only reduce the inner spectrum changes of the homogeneous areas, but also protect its edge and detail information, namely needing to consider the multi-scale characteristics of the features.

This article adopts the multi-scale morphological sequence which can open and close creation hybrid operation to process the spectral information. The model has three classifications of the spectrum-spatial characteristics, namely spectral characteristics-spectral characteristic of multi-scale morphology, spectral characteristics-physical characteristics of underlaying surfaces of multi-scale morphologyand spectral characteristics-features spatial extension characteristics of multi-scale morphology. The methods presented in this paper firstly respectively classifies the three classifications of the spectrum-spatial characteristics of the model by using the SVM classifier, and then carries out the probability fusion for the classification results based on the pixels to obtain the final image classification results.

3. OVERVIEW OF THE METHODS

3.1 Difference Multi-scale Morphological Sequence Based on CFO Operator

The bright details which are smaller than the structural bodies in the images can be smoothed and the stability of the overall characteristics can be maintained by opening the recreation function of morphology, and the dark details which are smaller than the structural bodies in the images can be smoothed and the stability of the overall characteristics can be maintained by closing the recreation function of morphology, therefore this article uses the multi-scale morphological characteristics of the morphological operators opening by reconstruction followed by closing by reconstruction (CFO) to calculate. The features of which spectrums in images are similar are easy to mix, such as the roads and the buildings with low illumination, but anisotropy and isotropy exist among the features, such that the roads have anisotropy but the buildings have isotropy, so that this article adopts the linear structural element (SE), sets as $SE = strel('line', d, s)$, wherein d and s respectively represent the direction and size of SE. So the CFO operators definition is as follows:

$$CFO^{g}(d,s) = \gamma_R^g(\phi_R^g(d,s)) \tag{1}$$

The white-hat conversion can extract the bright structures and remove the dark structures in the images, so that the areas which are smaller than the defined structural elements and brighter than the adjacent areas around can be obtained by subtracting the two images. So the White-hat conversion of CFO operators definition is as follows:

$$W-TH_{CFO}(d,s) = \phi_R^g(d,s) - CFO^g(d,s) \tag{2}$$

$\phi_R^g(d,s)$ uses the images obtained by closing the recreation for image g through SE, and $CFO^{g}(d,s)$ uses the images obtained by opening the recreation for image $\phi_R^g(d,s)$ through SE.

We do not confirm the specific sizes of the targets of interest in advance, and some specific targets are shown as multi-scale morphology in the images, so that, this article adopts the multi-scale Morphological sequence method proposed by Benediktsson[12], establishes Multi-scale morphological sequences based on CFO operator, can be known as the generalization of the characteristics on different scales，which is defined as follows：

$$\begin{cases} MP_{W-TH_{CFO}}(d,s) = W-TH_{CFO}(d,s) \\ MP_{W-TH_{CFO}}(d,o) = g \end{cases} \tag{3}$$

Combined with multi-scale morphological difference principle, we establish difference multi-scale morphological sequence based on CFO operator, which is defined as follows：

$$DMP_{W\ \ TH_{CFO}}(d,s) = \left| MP_{W\ \ TH_{CFO}}(d,s+\Delta s) \quad MP_{W\ \ TH_{CFO}}(d,s) \right| \tag{4}$$

wherein Δs is the distances of the morphological sequences, $s \in (s_{\min}, s_{\max})$.

3.2 Spectral Characteristics-spectral Characteristic of Multi-scale Morphology

The spectral characteristics of the images are the basis of the image classification, but the classification results cannot achieve the satisfying effect if the images are classified by directly according to spectral characteristics, so that this article adds the spatial characteristics based on the spectral characteristics, namely multi-scale morphological operations are carried out for the spectral characteristics, recorded as SDspe, and then use the SVM classifier to classify the images. The is defined as follows:

$$\overset{\text{spec}}{DMP}_{W-TH_{CFO}}(d,S) = \{\overset{spec}{DMP}_{W-TH_{CFO}}(d_1,S_1), \overset{spec}{DMP}_{W-TH_{CFO}}(d_2,S_2), \cdots, \overset{spec}{DMP}_{W-TH_{CFO}}(d_n,S_{\mathrm{n}})\} \tag{5}$$

wherein （d,S）expresses the different scales of the Linear SE sequences.

3.3 Spectral Characteristics-features Spatial Extension Characteristics of Multi-scale Morphology

In remote-sensing images, the similarity level of the pixels of the same feature is high on the spectral characteristic, so it provides the base for extracting the pixels of the same feature. We

can use this characteristic of the features to extract the spectral-spatial extension characteristics of the features, then carry out the morphological operations for the obtained extension characteristics, and then use the SVM classifier to classify the images.

The spatial extension characteristics of the features in the images can be expressed as extension of the pixels, namely the spatial extension characteristics have succession of the pixels of which attributes are similar in the special direction. Therefore this article creates a similarity function sim (t,θ,W) of the pixels, wherein t is responsible for the threshold values of the similarity of the pixels, θ is responsible for the extension direction, and W is responsible for the size of the window. This article evaluates the similarity of the adjacent pixels by using the spectral angle method; for the vectors x and y of the two spectrums, the similarity level of the spectrum is determined according to the size of the included angle. If the included angle is smaller, the two spectrums are more similar. The calculation method of the spectral angle is as follows:

$$\cos(x, y) = \frac{x^T y}{|x||y|} = \frac{x^T y}{[(x^T x)(y^T y)]^{1/2}} \tag{6}$$

This article defines the four extension directions of the centre pixel which extend to outside θ: 0°, 45°, 90° and 135° shown in Figure1, and the direction of the arrow is the main direction of the extension direction. Take θ=0 for example: the size of the window in the figure is set as w×w, the pixel o(i,j) is set as the centre pixel and its initial extension value is 1=0, when the spectral angle between the pixel o(i,j) and the adjacent pixel o(i,j+1) in the main direction is not larger than some per-set threshold value t, the spectral properties of the two pixels are similar, then the extension value adds 1, and then o(i,j+1) is adopted as the new starting point to continuously solve the extension; or the similarity of the pixels o(i-1,j+1), o(i+1,j+1) and o(i,j) is respectively calculated, if the spectral angle between one of the points and the starting point is not larger than some pre-set threshold value, this point is adopted as the starting point to continuously calculate until to reach the border of the window; or the calculation in this direction is ended; later the calculation is repeated in the opposite direction of 0° to obtain the extension value 1-, later the extension value 1 on the direction0° and the value 1- on the opposite direction are added, and the result is adopted as the extension characteristic, for example, the red line in Figure 2 is the extension example of θ=0°, for the similar directions, the extension values on the directions 45°, 90°and 135° can be calculated. If the extension values on the four directions0°,45°,90°and135° are recorded as {10,145,190,1135}, the the formulas to calculate of extension characteristic are such as Formula (7) - Formula(9).

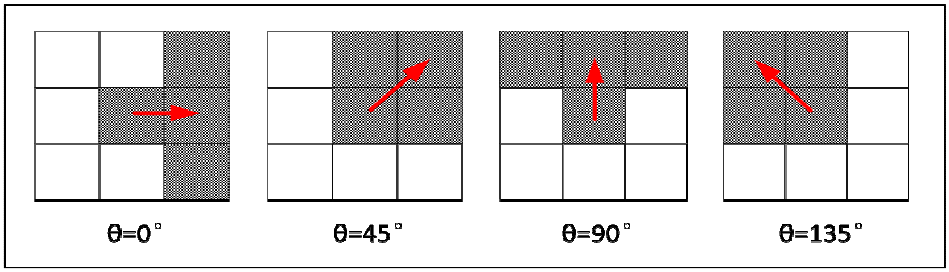


FIGURE 1： Extension of Four Directions.

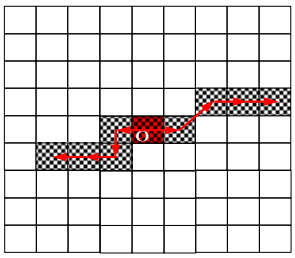


FIGURE 2: Extension Example of θ=0°.

$$\mathrm{eLong}_1(i,j) = mean\ (\{l^0, l^{45}, l^{90}, l^{135}\}) \quad (7)$$

$$\mathrm{eLong}_2(i,j) = \max(\{l^0, l^{45}, l^{90}, l^{135}\}) \quad (8)$$

$$\mathrm{eLong}_3(i,j) = \min(\{l^0, l^{45}, l^{90}, l^{135}\}) \quad (9)$$

So each pixel has the spatial extension characteristics of three features. For simplifying the analysis, the principal component analysis (PCA) method is used to synthesize the three extension characteristics, and the first principal component is used for expressing its extension characteristics. During the process of calculating the spatial extension characteristics of the features, this article adopts the adaptive control method based on cluster, for simplifying the setup of the threshold value t, which is to firstly cluster the mean value K for the image, then replace the similarity calculation with the pixel accumulation of the same clusters on the extension direction, and then replace the value t with the clusters number k.

Multi-scale morphological operations are carried out for the extension characteristics of the features, and get spectral characteristics-features spatial extension characteristics of multi-scale morphology($\mathrm{Spectral}_{\mathrm{NMF}}$ - $\overset{\mathrm{eLong}}{DMP}_{W\ \ TH_{CFO}}(d,s)$), record them as $\mathrm{SD}_{\mathrm{elo}}$, and then classifies the support vector machines. The $\overset{\mathrm{eLong}}{DMP}_{W\ \ TH_{CFO}}(d,s)$ is defined as:

$$\overset{\mathrm{eLong}}{DMP}_{W\ \ TH_{CFO}}(d,S) = \{\overset{\mathrm{eLong}}{DMP}_{W-TH_{CFO}}(d_1,S_1), \overset{\mathrm{eLong}}{DMP}_{W-TH_{CFO}}(d_2,S_2), \cdots, \overset{\mathrm{eLong}}{DMP}_{W-TH_{CFO}}(d_n,S_n)\} \quad (10)$$

3.4 Spectral Characteristics-physical Characteristics of Underlaying Surfaces of Multi-scale Morphology

In urban land coverage pattern, the underlaying surfaces of the urban ground are considered to be the linear combination of the vegetation, the impermeable stratum and soil[16]. The impermeable stratum mainly includes the features with high reflectance and the features with low reflectance. The features with high reflectance are responsible for the land coverage pattern with high spectral reflectance, such as bright buildings and dry bare lands; the features with low reflectance are responsible for the land coverage pattern with low spectral reflectance, such as the surface of the dark impermeable stratum, shadow, water bodies and wet lands. The earth's surface corresponded to each pixel in the image includes different coverage patterns and its spectral response characteristics are different, so this article defines the type of the end member in the multi-spectral remote sensing image as the features with high reflectance, the features with low reflectance, the vegetation and the soil, and uses the linear spectral mixing model to extract the physical characteristics of the underlaying surfaces of the ground. This method only applies the spectral characteristics of the single feature, the spatial characteristics of the feature are ignored，so in this paper, multi-scale morphological operations are carried out for the underlying surface physical characteristics, get spectral characteristics-physical characteristics of

underlaying surfaces of multi-scale morphology(Spectral$_{NMF}$ - $\overset{Fraction}{DMP}_{W-TH_{CFO}}(d,s)$), record as SD$_{Fra}$, and classifies the support vector machines. The $\overset{Fraction}{DMP}_{W-TH_{CFO}}(d,s)$ is defined as:

$$\overset{Fraction}{DMP}_{W\ \ TH_{CFO}}(d,S) = \{\overset{Fraction}{DMP}_{W-TH_{CFO}}(d_1,S_1), \overset{Fraction}{DMP}_{W-TH_{CFO}}(d_2,S_2),\cdots, \overset{Fraction}{DMP}_{W-TH_{CFO}}(d_n,S_{\mathrm{n}})\} \quad (11)$$

3.5 Probability Fusion Based on Pixels

The output results of the three support vector machine classifiers are integrated, and the final classification result shall be confirmed after passing through the maximum posterior probability, which calculation formula is as followed:

$$C(x) = \arg\max_{k=\{1,\ldots,K\}} \left\{ \frac{1}{F} \sum_{f=1}^{F} S_f(x) \cdot p_f^k(x) \right\} \quad (12)$$

F is responsible for the sum of the spectrum-spatial characteristics, $p_f^k(x)$ is responsible for the probability value of the pixel x in the probability output result of the SVM classification category k corresponded to the spectrum-spatial characteristic f, $S_f(x)$ is responsible for the category certainty of the pixel x of the SVM classification corresponded to the spectrum-spatial characteristic f, $S_f(x)$ is called the weight of the probability value $p_f^k(x)$, purpose for applying $S_f(x)$ is to reduce the influence of the classification results caused by the uncertain information and increase the weight of the reliable information.

4. EXPERIMENTS & RESULTS

We respectively apply the model, which is proposed in this paper namely P-fusion-SVM, to the experiments of ROSIS images and WorldView-2 images. Compare the overall classification accuracy and Kappa coefficient of different classification methods, such as classification results for simply considering the spectral characteristics and each result of three classification methods which mentioned in this paper and so on, to verify the feasibility of the Classification model proposed in this paper.

4.1 ROSIS Image Experiment

We test the images of Pavia University's School of Management shot by ROSIS aerial optical sensor on July 8, 2002. Figure3 shows the data of the shot experimental area of Pavia University: (a) the RGB color images of the area; (b) the multispectral images of which the number of the wave bands is 3 after NMF conversion; (c) ground reference data after locally surveying and mapping. The images and the ground reference data are provided by professor Gamba of Pavia University. The number of the training samples and the test samples of the images of Pavia University refers to Table 1.

For the hyperspectral images with high resolution, we usually reduce their dimensions, and then extract the characteristic of the images. The usual dimension method is to converse the principal analysis method PCA and the non-negative matrix factorization NMF, and reduce three spectral bands.

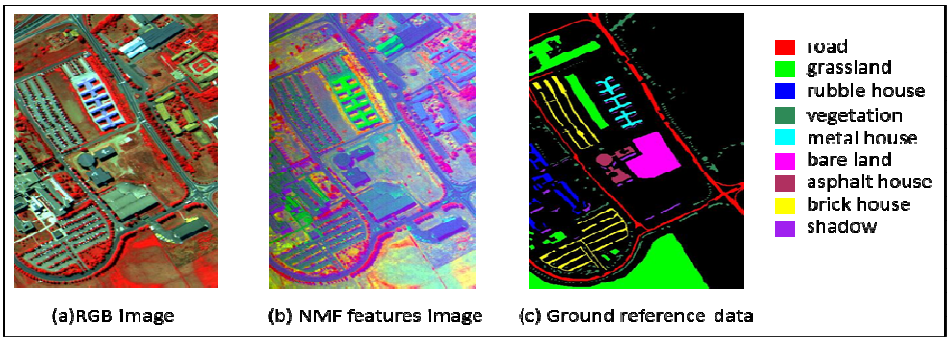


FIGURE 3: Data of Pavia University.

Categories	Training Samples	Test Samples
Road	530	6206
Grassland	507	16123
Rubble house	288	1880
Vegetation	251	2933
Metal house	343	1345
Bare land	367	5029
Asphalt house	221	1330
Brick house	410	3682
Shadow	254	947
Total samples	3171	39475

TABLE 1: Training Samples and Test Samples of the Images of Pavia University.

Classification / Categories	Hyperspectral 103-D	NMF 3-D	PCA 3-D
Road	80.86	81.41	76.86
Grassland	87.82	85.70	73.39
Rubble house	61.10	61.43	78.40
Vegetation	62.60	92.09	89.94
Metal house	99.04	98.88	99.78
Bare land	70.18	55.18	83.10
Asphalt house	47.73	40.23	65.49
Brick house	80.17	78.33	73.85
Shadow	99.68	99.89	99.26
OA	79.12	78.36	72.93
Kappa	73.34	72.98	66.06

TABLE 2: Feature Classification Precision Table Based on Different Spectral Data.

Table 2 is a feature classification precision table based on different spectral data, the spectral data includes the 103-dimensional hyperspectral data, the spectral data of which number of the wave bands is 3 after PCA conversion, and the spectral data of which number of the wave bands is 3 after NMF conversion. From the Table, the classification precision of NMF （OA=78.36%）is close to the classification precision of the 103-dimensional hyperspectral data（OA=79.12%）, which is higher than the classification precision of PCA（OA=72.93%）, so that NMF is rational and feasible to be adopted as the spectral characteristics of the experimental area.

Figure 4 shows the spectral curves of various features in the images of Pavia University in the hyperspectral data, each spectral curve is responsible for a classification of the special features, the spectral values of each classification of the features are the mean value of the spectral reflectance of all the samples of the features. From the Figure, the metal houses have high spectral reflectance within the whole wave band, but the reflectance of the shadow is low within the whole wave band, so that the two classifications of the features are easy to distinguish; the spectral curves of the asphalt houses, the roads, the brick houses, the rubble houses and the bare land are very similar, so we are difficult to simply distinguish them according to the spectral characteristics.

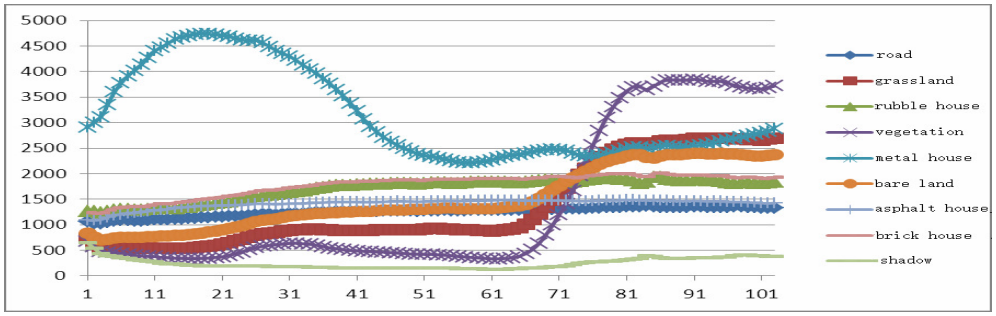


FIGURE 4: Spectral Curves of Various Features in the Images of Pavia University.

For this experiment, the specific parameters are set as follows: the number of the clusters of the spatial extension characteristics of the features K=20, the size of the window 17*17. The structural elements of the multi-scale morphological characteristics of the spatial extension characteristics of the features and the physical characteristics of the multi-scale morphological characteristics of the underlaying surfaces of the ground, set the sizes as s={2 4 6 8 10}, and the directions as d={45°, 90°, 135°, 180°}.

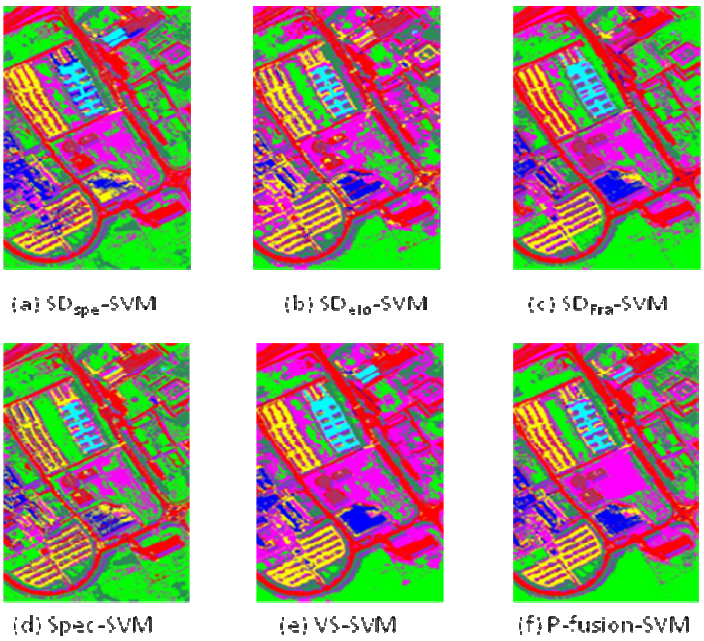


FIGURE 5: The Results of the University of Pavia Image Classification.

In Figure 5, the classification results of the features by using different methods include: (a) the SD_{spe}-SVM classification results; (b) the SD_{Fra}-SVM classification results ;(c) the SD_{Fra}-SVM classification results; (d) the SVM classification results for simply considering the spectral characteristics; (e) the VS-SVM classification results for overlaying the three classifications of the spatial characteristics and the spectral characteristics; (f) the P-fusion-SVM classification results of the feature classification model for using the probability fusion spectrum-spatial characteristics of the support vector machines. For verifying the accuracy of the model, we evaluate the precisions of the classification results by using six methods, and the evaluation results refer to Table 3.

Methods / Categories	SD_{spe}-SVM	SD_{elo}-SVM	SD_{Fra}-SVM	Spec-SVM	VS-SVM	P-fusion-SVM
	27-D	23-D	35-D	3-D	79-D	
Road	82.19	81.87	99.27	80.86	99.05	94.68
Grassland	79.19	89.49	88.41	87.82	86.31	91.57
Rubble house	87.29	64.79	80.69	61.10	88.67	88.81
Vegetation	91.75	95.19	91.68	62.60	95.40	95.12
Metal house	99.63	99.41	99.85	99.04	99.85	99.85
Bare land	76.22	74.05	80.23	70.18	84.63	80.17
Asphalt house	50.23	50.68	94.81	47.73	93.98	95.68
Brick house	82.75	85.96	92.50	80.17	95.00	96.30
Shadow	99.47	99.58	99.47	99.68	99.58	99.79
OA	81.14	84.51	87.20	79.12	90.73	93.35
Kappa	76.13	80.05	87.39	73.34	88.16	90.84

TABLE 3: Different Methods of Classification Accuracy Evaluation Form.

From Table3, the overall precision of the Spec-SVM classification is lower than the overall classification precision of other united spectrum-spatial characteristics, which proves that the spatial characteristics of the images can obviously improve the classification precision. The overall precisions of classifications SD_{spe}-SVM，SD_{elo}-SVM and SD_{Fra}-SVM are respectively 81.14%，84.51% and 87.20%, if the overall classification precision of the VS-SVM model is 90.73%, the traditional VS-SVM method can improve the accuracy of classification, the extraction precision of the features, such as the asphalt houses, the roads, the brick houses, the rubble house, the bare land and so on, of which spectrums are similar to distinguish, are obviously improved. If the overall classification precisions of the feature classification model which uses the probability fusion spectrum-spatial characteristics of the support vector machines(P-fusion-SVM) are 93.35%, compared with the VS-SVM method, the precision is improved for 2.62%, and the extract precisions of the special features---the buildings are also improved: the extraction precision of the rubble house is improved for 0.14%, the precision of the metal house remains the same, the precision of the asphalt house is improved for 1.7%, and the precision of the brick house is improved for 1.3%, so it proves that the model has good classification precision, feasibility and rationality.

4.2 WorldView-2 Image Experiments

We test the WorldView-2 images of Sanya, Hainan shot on January 18, 2010. The size of the experimental area is 400 rows 400 columns, the types of the features include seven classifications of the features: the roads, the shadows, the buildings, the vegetation, the water bodies, the bare land and the grassland. Figure 6 shows the data of the experimental area of the WorldView-2 images: (a) the RGB color images of the area; (b) the on-site reference data of the ground. The number of the training samples and the test samples of the images of World View-2 refers to Table 4.

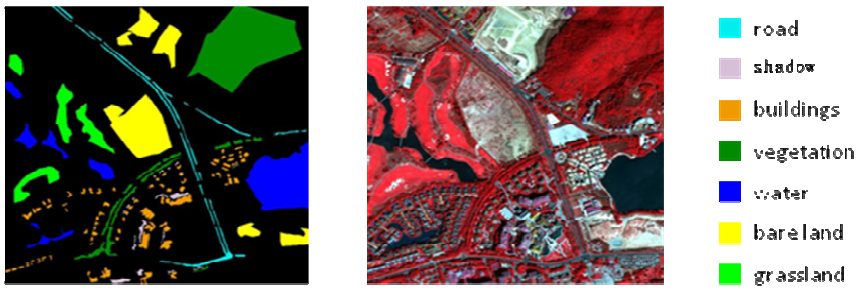


FIGURE 6: Data of the WorldView-2 Images.

Categories	Training Samples	Test Samples
Road	157	4058
Shadow	250	5655
Buildings	221	16624
Vegetation	109	1270
Water	210	25793
Bare Land	256	7435
Grass Land	102	17430

TABLE 4: Training Samples and Test Samples of the Images of WorldView-2.

Figure 7 shows the spectral curves of each classification of the features in the Sanya images, each spectral curve is responsible for a classification, and the spectral values of each classifications of the features are the mean value of the spectral reflectance of the reference samples. From the Figure, the reflectance of the two groups of the features: the buildings and the bare land, the water bodies and the shadow, is highly similar, so that the features are difficult to simply distinguish according to the spectral characteristics.

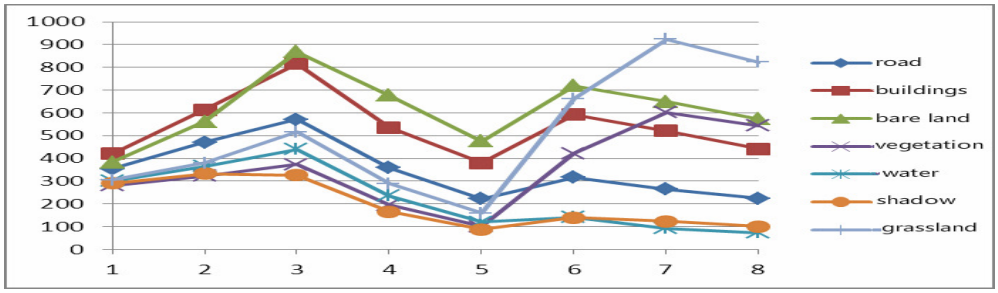


FIGURE 7: The Spectral Curves of Classifications in the Sanya Image.

Based on the description above, we apply the feature classification model of the probability fusion spectrum-spatial characteristics based on the support vector machine to this experiment. For the experiment, the specific parameters are set as follows: the number of the clusters of the spatial extension characteristics of the features K=20, the size of the window 11*11.The structural elements of the multi-scale morphological characteristics of the spatial extension characteristics of the features and the physical characteristics of the multi-scale morphological characteristics of the underlaying surfaces of the ground, set the sizes as *s*={1 2 3 4 5} and the directions as d={45°, 90°, 135°, 180°}.

In order to compare conveniently, we introduce SD_{spe}-SVM, SD_{elo}-SVM, SD_{Fra}-SVM, Spec-SVM and VS-SVM methods, and them classification results refer to Figure 8.

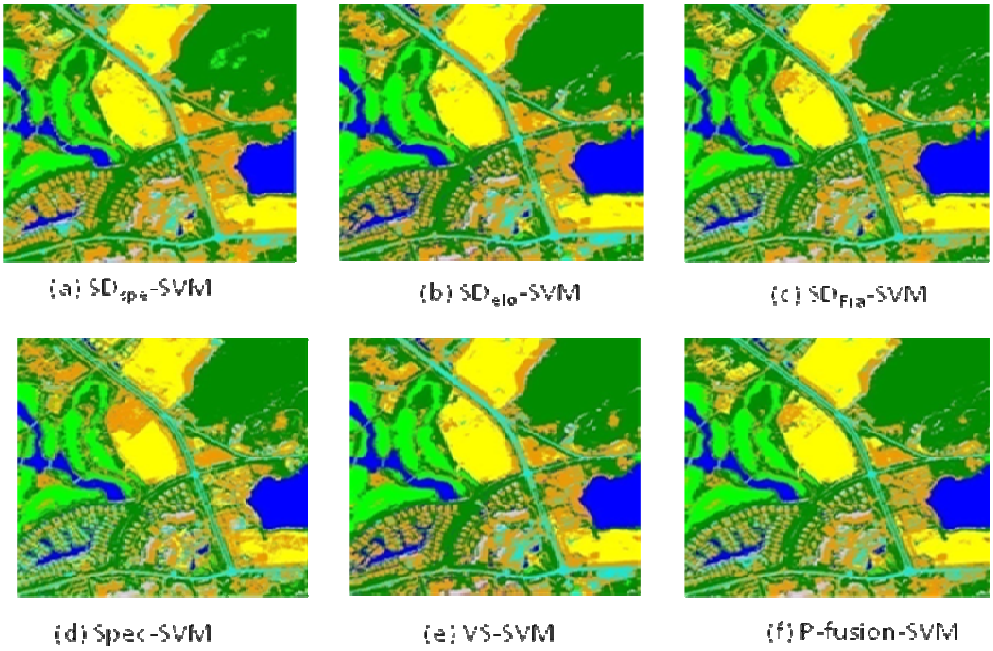


FIGURE 8: Classification Results of Sanya Image.

In order to analyze the classification of the images and the extraction results of the buildings, we evaluate the precision of the six classification methods, and the evaluation result refers to Table 5. From the Table, relative to the five methods: SD_{spe}-SVM, SD_{elo}-SVM, SD_{Fra}-SVM, Spec-SVM and VS-SVM, the overall precisions of P-fusion-SVM are respectively improved for 1.59%, 9.2%, 2.37%, 11.11% and 1.31%. The Kappa coefficients of P-SVM are respectively improved for1.43%, 11.12%, 2.46%, 14.23% and 1.66%. The precision analysis above proves that the model improves the classification precision of the images.

Methods / Categories	SD_{spe}-SVM	SD_{elo}-SVM	SD_{Fra}-SVM	Spec-SVM	VS-SVM	P-fusion-SVM
	27-D	23-D	35-D	3-D	79-D	
road	97.81	89.33	89.60	91.01	96.80	96.16
shadow	94.72	96.22	81.89	94.88	94.41	95.43
buildings	92.93	82.94	94.27	80.58	92.60	97.50
vegetation	96.98	96.83	94.08	97.15	95.04	99.48
water	100	99.96	97.47	99.92	100	99.98
bare land	92.77	67.52	97.72	57.88	95.72	97.05
grassland	99.91	91.58	98.99	94.20	99.76	99.13
OA	96.68	89.07	95.90	87.16	96.96	98.27
Kappa	96.35	86.66	95.32	83.55	96.12	97.78

TABLE 5: Different Methods of Classification Accuracy Evaluation Form.

5. CONCLUSIONS & FUTURE WORK

The image classification model of the probability fusion spectrum-spatial characteristics based on the support vector machine uses three classifications of the spectrum-spatial characteristics（SDspe-SVM, SDelo-SVM, SDFra-SVM）to classify images ,then carries out the

probability fusion for the classification results based on the pixels to obtain the final image classification results. For testing the feasibility of the model, we verify the model through the images of two different sensors and compare the classifications with the qualified VS-SVM model , the result shows that the model can improve the precisions of classifications.

The Classification model proposed in this paper is only aimed at the probability of a pixel fusion, does not take into account the pixel neighborhood influence on its category judgment, therefore how to reasonably consider other pixels within the scope of pixel neighborhood impact category of the classification results will be study in the future. And the various classification methods of the Classification model l proposed in this paper have the same fusion probability weighting, it may also impacts on pixel category judgment. Therefore, how to distribute the weight of each method in the model is also need further research. The Classification model proposed in this paper is combination the results of three classification methods, if there add more methods or use other methods would get better result will be study in future.

6. REFERENCES

[1] Huang X., Zhang L., Li P..“An adaptive multiscale information fusion approach for feature extraction and classification of IKONOS multispectral imagery over urban areas.” Geoscience and Remote Sensing Letters, vol. 4, pp. 654-658, Oct. 2007.

[2] Dell'Acqua F., Gamba P., Ferrari A., et al.“Exploiting spectral and spatial information in hyperspectral urban data with high resolution.” Geoscience and Remote Sensing Letters, vol. 1, pp. 322-326, Oct. 2004.

[3] Benediktsson J. A., Palmason J. A., Sveinsson J. R..”Classification of hyperspectral data from urban areas based on extended morphological profiles.” Geoscience and Remote Sensing, vol. 43, pp. 480-491, Mar. 2005.

[4] Fauvel M., Benediktsson J. A., Chanussot J., et al.”Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles.” Geoscience and Remote Sensing, vol. 46, pp. 3804-3814, Nov. 2008.

[5] Huang X., Zhang L..” A comparative study of spatial approaches for urban mapping using hyperspectral ROSIS images over Pavia City, northern Italy.” International Journal of Remote Sensing, vol. 30, pp. 3205-3221, Jun. 2009.

[6] Waske B., Benediktsson J. A.. ”Fusion of support vector machines for classification of multisensor data.” Geoscience and Remote Sensing, vol. 45, pp. 3858-3866, Dec. 2007.

[7] Pal M., Foody G. M.. “Feature selection for classification of hyperspectral data by SVM.” Geoscience and Remote Sensing, vol. 48, pp. 2297-2307, May 2010.

[8] Waske B, van der Linden S, Benediktsson J A, et al. “Sensitivity of support vector machines to random feature selection in classification of hyperspectral data.” Geoscience and Remote Sensing, vol. 48, pp. 2880-2889, Jul. 2010.

[9] Huang X., Zhang L..”Morphological Building/Shadow Index for Building Extraction From High-Resolution Imagery Over Urban Areas.” Selected Topics in Applied Earth Observations and Remote Sensing, vol. 5, pp. 161–172, Feb. 2012.

[10] Huang X., Zhang L., Li P..”Classification and Extraction of Spatial Features in Urban Areas Using High Resolution Multispectral Imagery.” Geoscience and Remote Sensing, vol. 4, pp.260–264, Apr. 2007.

[11] Ouma Y. O., Ngigi T. G., Tateishi R..”On the Optimization and Selection of Wavelet Texture for Feature Extraction from High-resolution Satellite Imagery with Application towards Urban-tree Delineation.” International Journal of Remote Sensing, vol. 27, pp. 73–104, Feb. 2007.

[12] Changshan Wu, Alan T. Murray."Estimating impervious surface distribution by spectral mixture analysis." Remote Sensing of Environment, vol. 84, pp. 493 – 505, Apr. 2003.

[13] Soille P., Pesaresi M.."Advances in Mathematical Morphology Applied to Geoscience and Remote Sensing." Geoscience and Remote Sensing, vol. 40, pp. 2042–2055, Sep. 2002.

[14] X. Jin, Davis C. H.."Automated Building Extraction from High-resolution Satellite Imagery in Urban Areas Using Structural, Contextual, and Spectral Information" EURASIP Journal on Applied Signal Processing, vol. 2005, pp. 2196–2206, Jan. 2005.

[15] Gamba P., F. Dell'Acqua, G. Lisini, and G. Trianni. "Improved VHR urban mapping exploiting feature boundaries." Geoscience and Remote Sensing, vol. 45, pp. 2676–2682, Aug. 2007.

[16] Ridd, M. K."Exploring a V – I – S (vegetation – impervious surface – soil) model for urban ecosystem analysis through remote sensing: comparative anatomy for cities." International Journal of Remote Sensing, vol. 16, pp. 2165 – 2185, Jan. 1995.

[17] Tian J., Chen D. M.."Optimization in Multi-scale Segmentation of High-resolution Satellite Images for Artificial Feature Recognition." International Journal of Remote Sensing, vol. 28, pp. 4625–4644, Sep. 2007.

[18] Pesaresi M., Benediktsson J. A.."A New Approach for The Morphological Segmentation of High-resolution Satellite Imagery." Geoscience and Remote Sensing, vol. 39, pp. 309–320, Feb 2001.

Usage of Shape From Focus Method For 3D Shape Recovery And Identification of 3D Object Position

Joanna Florczak *florczak@agh.edu.pl*
Faculty of Mechanical Engineering and Robotics
Department of Robotics and Mechatronics
AGH University of Science and Technology
Cracow, 30-059, Poland

Maciej Petko *petko@agh.edu.pl*
Faculty of Mechanical Engineering and Robotics
Department of Robotics and Mechatronics
AGH University of Science and Technology
Cracow, 30-059, Poland

Abstract

Shape from focus is a method of 3D shape and depth estimation of an object from a sequence of pictures with changing focus settings. In this paper we propose a novel method of shape recovery, which was originally created for shape and position identification of glass pipette in medical hybrid robot. In the proposed algorithm, Sum of Modified Laplacian is used as a focus operator. Each step of the algorithm is tested in order to select operators with the best results. Reconstruction allows not only to determine the shape but also precisely define the position of the object. The results of proposed method, obtained on real objects, have shown the efficiency of this scheme.

Keywords: Shape From Focus, Image Processing, 3D Shape Recovery.

1. INTRODUCTION

Recovery of a three-dimensional shape from two-dimensional images has been actively studied for a long time. In the literature, there are many proposed techniques like Shape From Shading[1] or Shape From Motion[2]. One of them is also Shape From Focus (SFF)[3]. In this method, the shape of the objects is estimated from series of images of the same scene with changing focus level. In SFF, after image acquisition, a focus measure operator is applied to specify points with the best focus for every image. In the next step, every pixel is assigned to the frame where the best focus has been obtained. Thanks to that, it is possible to reconstruct the shape of the object[3]. In this paper, a novel Shape From Focus algorithm, which was originally developed for the shape and the position identification of a glass pipette in a medical hybrid robot, is presented. This element is hard to examine because of it's partial transparency, so the algorithm needs to be adapted to the type of a reconstructed object. The proposed algorithm reconstructs shape even when object is difficult to determine or is very complex. Different focus measure operators have been studied and the one with the best results has been chosen. After focus measure, obtained pictures have been filtered with a Butterworth first order filter in order to get rid of high frequencies, occurring on the reconstruction as picks. Method for elimination of noise and determination, not only the shape, but also the position of the object was developed. Every pixel was assigned to the height were it obtained the maximum focus measure. The beginning of the pipette was the critical area, which was crucial to defined in this experiment, this is why it was used as a threshold in the reconstruction process. According to this threshold, the decision was made if the focus value of the pixel is enough to be considered during the reconstruction. In the end, the shape was filtered with the average filter, in order to make edges smoother and also to eliminate remained noises.

2. PREVIOUS WORKS

There are numerous solutions of SFF method examined in the literature. Each of them deals with different aspects of this algorithm. Many of works consider focus measure operator used in the algorithm. Mostly common are : Modified Laplacian (ML), Sum of Modified Laplacian (SML), Tenenbaum gradient (TEN) and gray level variance (GLV)[3].

SML is based on Laplace filter. In this filter, second derivatives in x and y directions can cancel each other. This is why Modified Laplacian model has been proposed [3]. It is defined as:

$$ML(x,y) = \left|\frac{\partial^2 g(x,y)}{\partial x^2}\right| + \left|\frac{\partial^2 g(x,y)}{\partial y^2}\right|$$

,where g(x,y) is the gray value for the coordinates (x,y). To improve the results of reconstruction focus measure in coordinates x,y are presented as a sum of ML in local window.

$$SML(x_0,y_0) = \sum_{p(x,y)\in \mathrm{U}(x_0,y_0)} ML(x,y)$$

Tennenbaum method is based on Sobel gradient operator. This focus measure method uses sum of squares of horizontal and vertical Sobel masks, also summed in a local window[4].

$$TEN(x_o,y_o) = \sum_{p(x,y)\in \mathrm{U}(x_0,y_0)} [G_x(x,y)^2 + G_y(x,y)^2$$

The last one, GLV focus measure is based on variations in a gray level values. On sharp image this variation is bigger that on image that is blurred. We get information about focus of the object by detecting variations between gray level variance in local window[3].

$$GLV(x_0,y_0) = \frac{1}{N-1} \sum_{p(x,y)\in \mathrm{U}(x_0,y_0)} [g(x,y) - \mu_{\mathrm{U}(x_0,y_0)}]^2$$

where $\mu_{\mathrm{U}(x_0,y_0)}$ is an average gray level variance of neighborhood pixels $\mathrm{U}(x_0,y_0)$.

There are also some innovative methods of focus measure proposed in the literature. One of them measures focus with fast discrete curvelent transform – FDCT. [5] Another one uses Anisotropic Nonlinear Diffusion Filtering – ANDF [6].

Effectiveness of focus measure not only depends on used algorithm. Very important is also the type of window used in the calculations. In one of proposed algorithm instead of 2D, 3D window was used[7].

Some of works focus on the initial state of the algorithm which is acquisition of the pictures. An example is the algorithm based on combinatorial optimization[6]. This algorithm tries to find combination of frames for which the maximum focus level will be achieved. A similar problem is taken into consideration in the paper about the minimum number of images required to obtain sufficient shape reconstruction [8].

Another issue often analyzed in order to improve SFF is a method of approximation of edges after focus measure process. An examples of such improvement can be usage of Lorentzian-Cauchy function[9] or kernel regression [10].

3. PROPOSED ALGORITHM

In this chapter, a new approach for improvement of the SFF method, is shown. The proposed algorithm is stated below. Basic process contains steps like: image acquisition, normalization, focus measure, filtering, creating initial and final depth map, as shown in FIGURE 1.

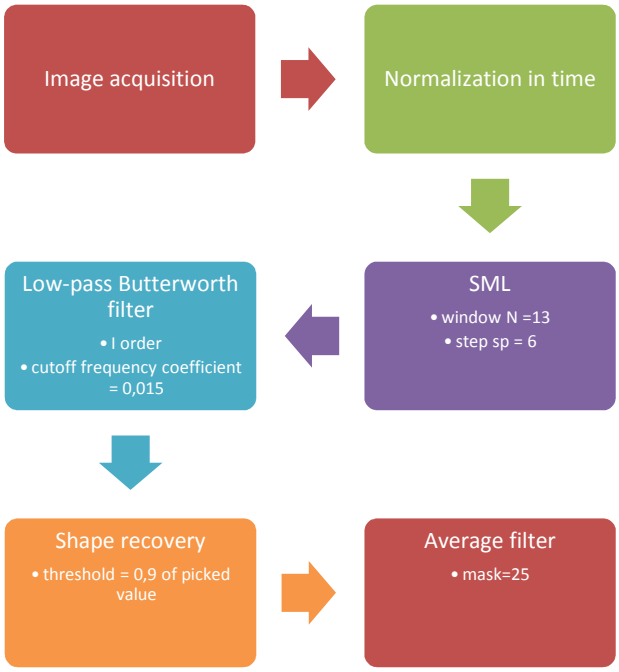


FIGURE 1: Proposed Algorithm.

After image acquisition, normalization process was performed. Every frame was divided by the maximum value of this frame. Thanks to that every frame was in range 0÷1.

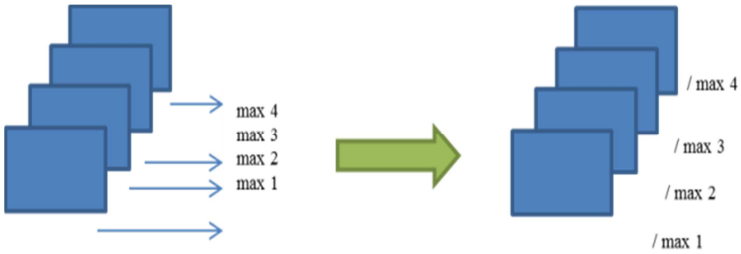


FIGURE 2: Normalization In Time.

In the second step, it was decided which focus operator should be used in the algorithm. Many operators have been proposed to measure the degree of focus level of the picture. The most commonly used are: Laplacian (Lap), Modified Laplacian (ML), Sum of Modified Laplacian (SML) and Tenenbaum gradient (TEN).[3] Focus measure values in this paper have been obtained by using the Sum of Modified Laplacian, because it gives the best results in experiments. Results of one of this experiments are shown below. In the FIGURE 3 one frame, from series that was taken into analysis, is presented. Focus measure values obtained with different focus operators for pixel marked in red are shown on FIGURE 4.Knowledge of the series, indicates that maximum focus measure should be obtained on the last – 18 frame.

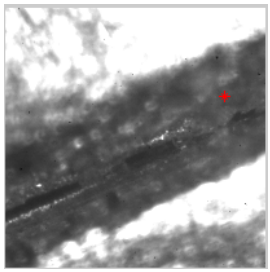


FIGURE 3: Frame from series that was taken into analysis.

As it can be seen only SML was able to detect the maximum focus correctly. ML found the maximum focus on 18th frame too, but it also detected high focus on 14th frame, what was incorrect. This is an example of experiments conducted to find the best focus measure operator. According to the results of those experiments SML will be operator used in this algorithm.

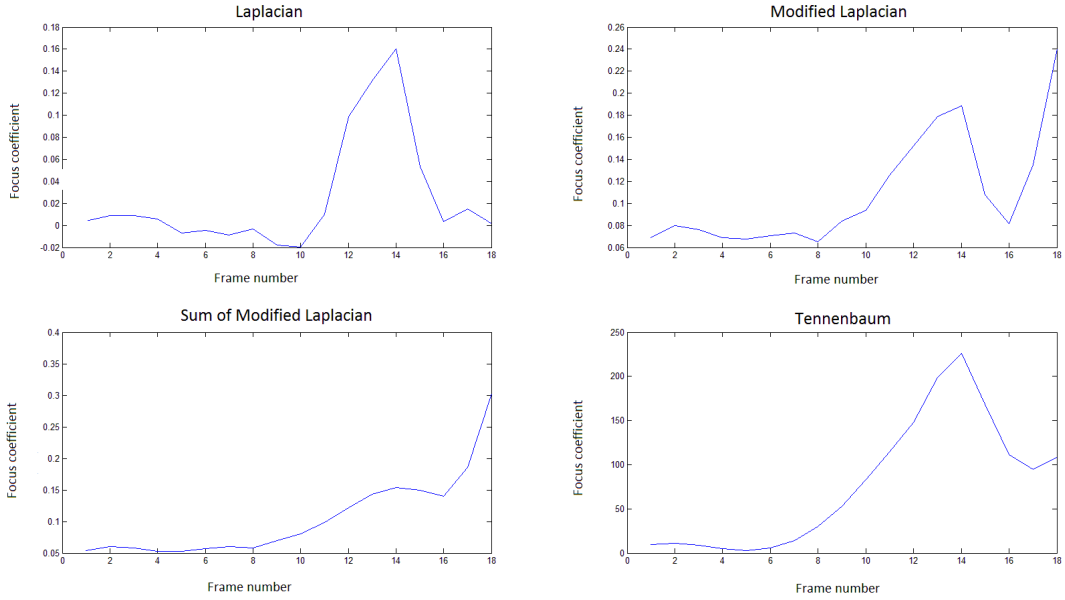


FIGURE 4: Focus measure obtained with different focus operators.

Afterwards, frames were filtered with low pass Butterworth filter. This allows to get rid of undesirable high frequencies, occurring on reconstructed image as a peaks. It is also the method of approximation of edges after focus measure process. Butterworth filter has flat frequency response, so it will smooth the edges without significant distortion. After experiments, first-order Butterworth filter with cutoff coefficient equal 0.015, shown in FIGURE 5, was used.

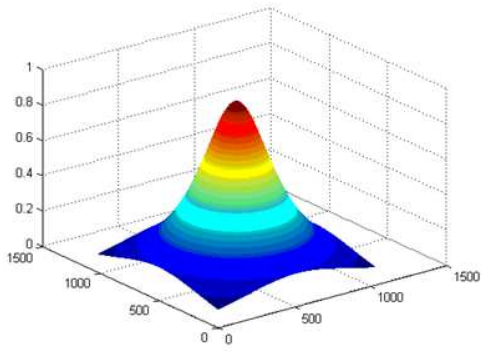


FIGURE 5: First-order Butterworth Filter.

Prepared images were ready for the shape recovery process. For every pixel, frame where the highest focus measure level was obtained, was found. Knowing picture number and high of that picture made possible to assign every pixel to high where it obtained maximum focus measure. It was necessary to set a threshold. It was a the minimal value of pixel focus measure that should be reached to allow the assignment process. To do this one point was picked, for this reconstruction, it was beginning of the pipette because correct estimation for this area was the most crucial. Experimentally, the threshold was set as a 90% of picked value. This process allowed to create initial depth map. To enhance results of reconstruction, average filtering was applied. Thanks to that final reconstructed shape was smoother and did not have a high frequency noises.

4. RESULTS

In this section we show the improved performance of proposed scheme by testing it on a various images. Experiments were conducted on four different types of objects shown on FIGURE 6 (a-d).

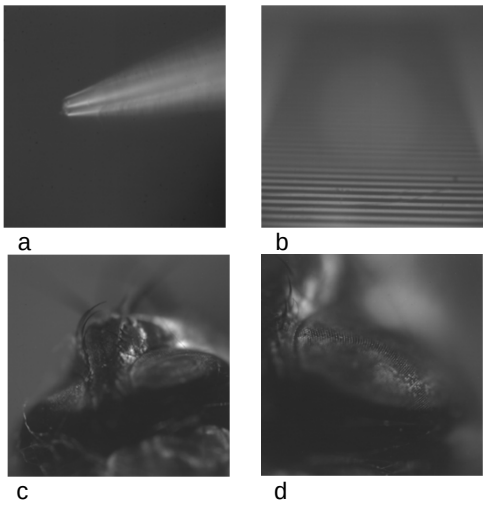


FIGURE 6: Objects on which experiments were proceeded (a-d).

All series of pictures have been made with specially designed optical system. It contains fully apochromatic zoom system, where zoom, iris diaphragm and focus are motorized. It provides high resolution images without chromatic distortions. The system also has objective LEICA PLAN APO 5x/0.5 LWD, what makes 95% of an area flat during the observation. In addition, system has illumination module that provides a coaxial illumination of the test object. First object was a partially transparent pipette (FIGURE 6a), for which this algorithm was originally dedicated. This object is highly difficult to reconstruct because of it's transparent structure. The shape was reconstructed from a sequence of 22 pictures.

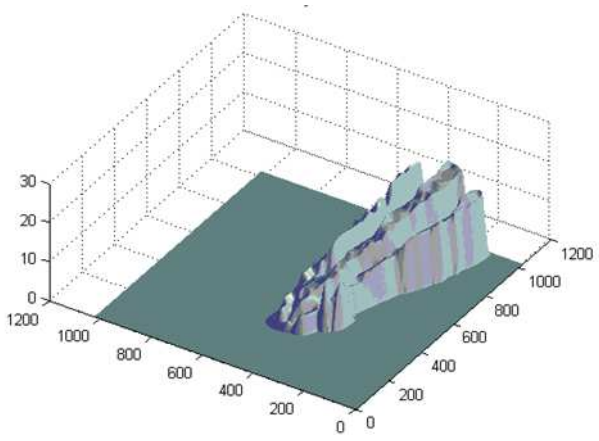


FIGURE 7: Reconstruction of the Pipette.

As it can be seen on FIGURE 7, there are areas for which reconstruction was impossible because of transparency of those places. However, the general shape was successfully reconstructed. Also the pick of pipette, which was crucial in this experiment was exactly detected.

Second object was *DOF 5-5 Depth of Field Target, Edumnd Optics* (FIGURE 6b). The shape was reconstructed from a sequence of 22 pictures. This experiment was mostly performed to show that it is possible not only to reconstruct shape but also keep proportions and exact depth map of the object. Depth of Field Target was placed so that it would be possible to observe the surface at an angle of 45 degrees. As it can be seen in the FIGURE 8 this angle was preserved while reconstruction process. Steps that are visible on reconstructed object are a result of established step of scanning and iris settings. To rectify this effect it is need to decrease the scanning step or change depth of filed by opening the iris. In this experiment it was important to verify the correct angle of the reconstructed surface, exact shape was not crucial, that is why big step has been used.

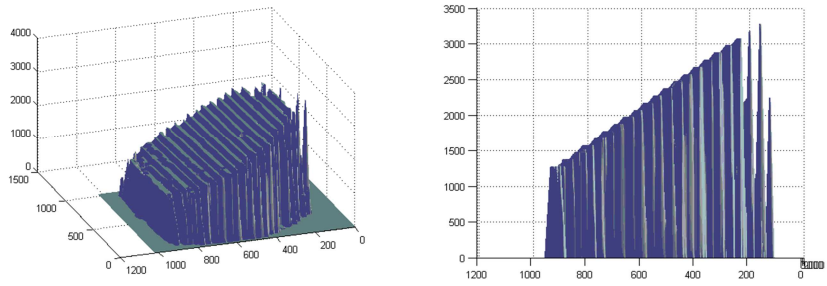


FIGURE 8: Reconstructions of DOF 5-5 Depth of Field Target, Edumnd Optics.

In the FIGURE 9, it is shown how the reconstruction changed when depth of field was higher. Algorithm with the same number of pictures was able to reconstruct shape properly.

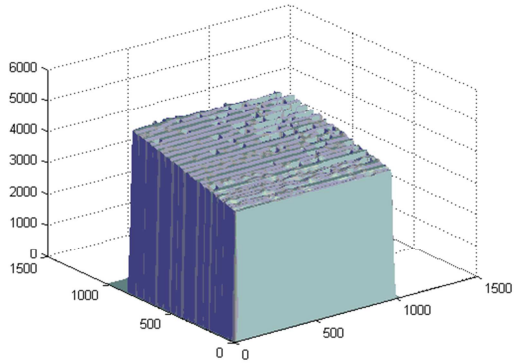


FIGURE 9: Reconstructions of DOF 5-5 Depth of Field Target, Edumnd Optics with high depth of field.

In third experiment a biological object was examined. A fly was used for this purpose (FIGURE 6c). This shape is very complicated so it was possible to fully prove the possibilities of the algorithm. Of course reconstruction was not exact, because only view from one perspective was available. However, object has been reconstructed accurately despite from it's high complexity.

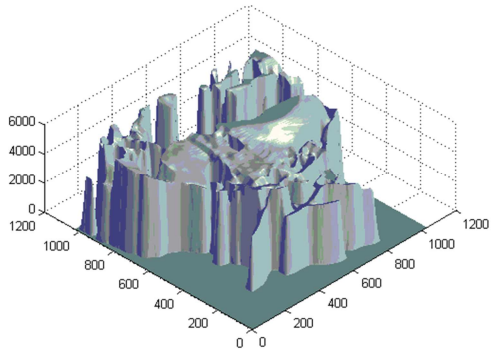


FIGURE 10: Fly's Head Reconstruction.

The last experiment was conducted also with the fly, but attention was focused on smaller part (FIGURE 6d). Only the eye of the fly was reconstructed. Pictures were made with six time zoom and iris open with 100%, what gives very small depth of field. Distance of 1275 um was scanned with step of 25 um, what gives 53 pictures. Those parameters were needed because of small size of the object and small variation of surface. Results of the experiment shown on FIGURE 11 are very satisfying.

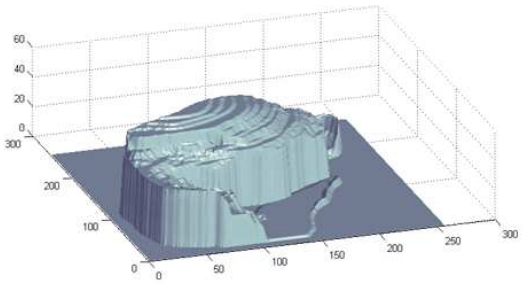
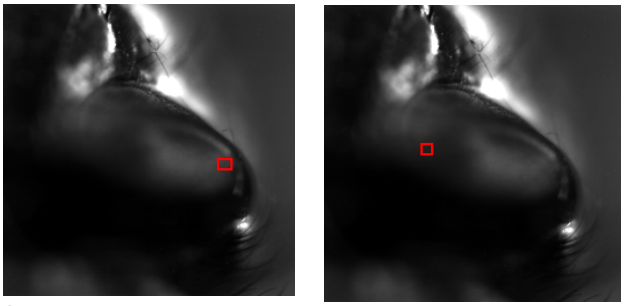


FIGURE 11: Fly's Eye Reconstruction.

Shape of the eye has been carefully restored, even small changes on the surface were detected by the algorithm.

One more experiment was proceed to shows possibilities of algorithm to define the position of the object and distance between two parts of reconstructed shape. Small areas on fly's eye were selected, what is shown of FIGURE 12 (a,b).



a B

FIGURE 12: Selected Areas (a,b).

In FIGURE 13, it is shown how value of focus measure depends on high in Z axis. Plot blue is for one selected area and plot green is for the other. As it can be seen, maximum focus measure for both is strictly defined, so it is possible to find the high of them and consequently distance between them. The results confirm that it is possible to obtain accurate information about individual elements, even when they are really small or complicated.

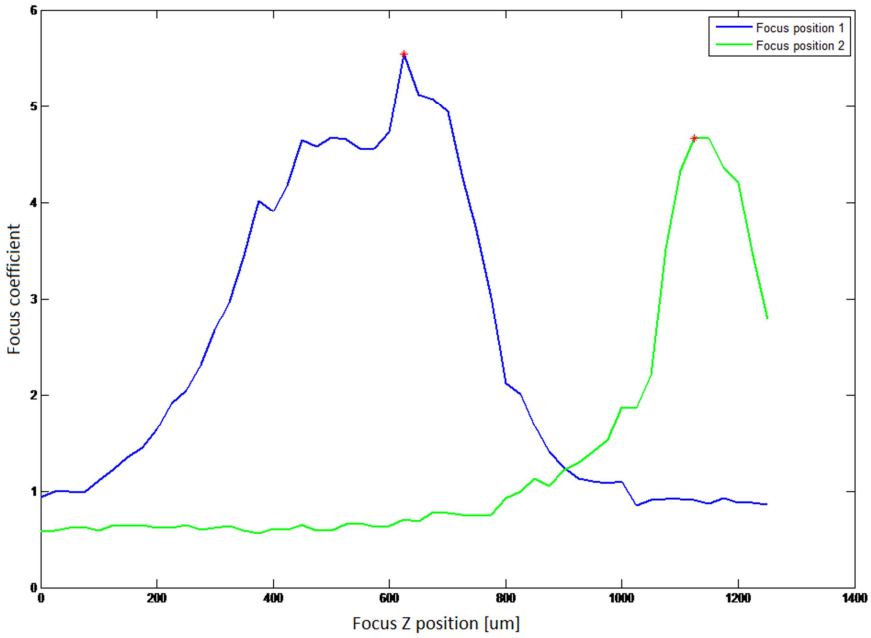


FIGURE 13: Value of focus measure depending on high in Z axis.

5. DISCUSSION

Proposed algorithm is a substantial improvement in the field of shape recovery methods. While creating the processing scheme, every aspect of reconstruction process was carefully analyzed and the most adequate method was selected. For each step the extensive literature background was checked, the best solutions were found, tested, and in the end, parameters for those methods were chosen to fulfill the requirements as close as it was possible[11][12][13]. The novel Shape from Focus method was able to reconstruct objects with high accuracy and preserve proportions even with a large complexity of the object. Also the height of a single point and the distance between two points were possible to obtain. Thr proposed methodology was dedicated for a glass pipette, which is difficult to reconstruct, but the algorithm managed to gain this goal. There is anxiety that the algorithm created for one specific case would not be efficient for different objects. To limit this problem, there are steps like normalization in time, or picking the threshold, which make the proposed method useful for diverse cases. In most of researched papers authors were focused on one aspect of the algorithm, like the image acquisition[6][8],the focus measure operator[5][6][14], the size of the window[7] or approximation methods [9][10]. In the proposed scheme all aspects of the algorithm were analyzed to create the most efficient solution. Comparative evaluation with other authors wasn't fully possible, because object that were used in other experiments were not available for tests. The algorithm proposed in this paper, in comparison to other works, was able to reconstruct even really complex and difficult to determine structures.

6. CONCLUSION

In this paper, a novel SFF method was presented for 3D Shape recovery and identification of a 3D object position. The algorithm was originally created for the shape and the position identification of a glass pipette in a medical hybrid robot. In the proposed algorithm, after image acquisition, there is a normalization process. In the next step Sum of Modified Laplacian is applied in order to find areas with the maximum focus. After this, images are filtered with a first order low-pass Butterworth filter. The initial depth map is obtained by assigning each pixel to the height, where it reached the maximum focus value. The threshold in the reconstruction process depends on the value of a pixel, for which proper reconstruction is the most important. In this case, it is beginning of a pipette. Thanks to that, it is sure that the most important area will be reconstructed correctly. In the end, average filtering is applied to smooth the edges of the reconstructed shape. The proposed scheme was tested on four different objects and reconstruction gave really good results. Moreover, the proposed algorithm is able to define the position of the object and distance between two selected areas.

7. Acknowledgements

The project was funded by the National Science Centre.

8. REFERENCES

[1] M. Breuß, O. Vogel and A. Tankus. " Modern shape from shading and beyond". Image Processing (ICIP), 2011 18th IEEE International Conference on. IEEE, pp. 1-4, 2011.

[2] M. Marques and J. Costeira. "3D face recognition from multiple images: A shape-from-motion approach", Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on. IEEE, pp. 1-6, 2008.

[3] S.K. Nayar and Y.Nakagawa. "Shape from focus", Pattern analysis and machine intelligence, IEEE Transactions on, 16.8, pp. 824-831,1994.

[4] J.M. Tenenbaum. "Accommodation in computer vision". Stanford Univ Ca Dept of Computer Science, 1970.

[5] R. Minhas , A. Adeel Mohammed and Q.M. Jonathan Wu. "Shape from focus using fast discrete curvelet transform". Pattern Recognition, 44.4, pp. 839-853. 2011.

[6] M.T. Mahmood and T.S. Choi. "Nonlinear Approach for Enhancement of Image Focus Volume in Shape From Focus", 21.5, pp. 2866-2873, 2012.

[7] S.O. Shim and T.S. Choi. "A novel iterative shape from focus algorithm based on combinatorial optimization". Pattern Recognition, 43.10, pp. 3338-3347, 2010.

[8] M.S. Muhammad and T.S. Choi "Sampling for Shape from Focus in Optical Microscopy". Pattern Analysis and Machine Intelligence, IEEE Transactions on, 34.3, pp. 564-573, 2012.

[9] M.S. Muhammad and T.S. Choi "3D shape recovery by image focus using Lorentzian-Cauchy function". Image Processing (ICIP), 2010 17th IEEE International Conference on. IEEE, pp. 4065-4068, 2010.

[10] M. Tariq Mahmood and T.S. Choi "Shape from focus using kernel regression". Image Processing (ICIP), 2009 16th IEEE International Conference on. IEEE, pp. 4293-4296, 2009.

[11] P. Kamencay, M. Breznan, R. Jarina and M. Zachariasova. "Estimation and Image Segmentation of a Sparse Disparity Map for 3D Reconstruction."

[12] N.N. Kachouie. "Anisotropic Diffusion for Medical Image Enhancement". International Journal Of Image Processing (IJIP), 4.4, pp.436, 2008.

[13] R. Maini and H. Aggarwal. "Study and Comparison of Various Image Edge Detection Techniques". International Journal of Image Processing (IJIP), 3.1, pp. 1-11, 2009.

[14] F.S. Helmli and S. Scherer. "Adaptive Shape from Focus with an Error Estimation in Light Microscopy". Image and Signal Processing and Analysis, 2001. ISPA 2001. Proceedings of the 2nd International Symposium on. IEEE, pp. 188-193, 2001.

Comparative Study of Image Registration Methods

Supriya S. Kothalkar *supriya.kothalkar@yahoo.com*
Saraswati College of Engineering
Mumbai University
Kharghar,Mumbai,India

Dr. Manjusha Deshmukh *manju0810@yahoo.com*
Saraswati College of Engineering
Mumbai University
Kharghar,Mumbai,India

Abstract

The main objective of image registration is to match two or more images captured at different times by different sensors or by different angles or from different viewpoints. Image registration has become a crucial step in most of the image processing tasks used in various areas. It is a key technology which is applied for computer vision, remote sensing, image processing, medical image analysis and other fields. Medical image registration is used to find a spatial transformation to match all the anatomical points and diagnostic points on the image. In general, the majority of registration methods consist of the following four steps: feature extraction, feature matching, transform modeling, and finally image resampling. The accuracy of a registration process is highly dependent to the feature extraction and matching methods. Cross Correlation is the basic statistical approach to image registration. It is used for template matching or pattern recognition. Template is considered as a sub-image from the reference image, and the image is considered as a sensed image. The objective is to establish the correspondence between the reference image and sensed image. It gives the measure of the degree of similarity between an image and template, which can be used for image registration. Normalized Cross Correlation (NCC) method is improved by using Feature Based Method. Image are effectively represented by any of its feature such as edges, points, curves etc. and these features are effectively used for image registration. Images are applied with the filters to extract edges. Post that NCC is done to find the sharp point on NCC plot. This method restricts us with only monomodal images. For multimodal images we have used Mutual Information as a measure of similarity. A widely used measure is Mutual Information (MI). This method requires estimating joint histogram of the two images. Experiments are presented that demonstrate the approach. The technique is intensity-based rather than feature-based. Mutual Information is effectively used as similarity measure between two images which can be monomodal or multimodal. Mutual information is consider as a measure of how well one image explains the other; it is maximized at the optimal alignment. To make this more effective Contourlet transform is used. Contourlet is a recent development of transform theory as an improvement of wavelets. It is a multiscale and multidirectional, two dimensional transform. It is a combination of Laplacian pyramid and directional filter bank. The discrete contourlet transform has a fast iterated filter bank algorithm that requires order N operations for N-pixel images. the contourlet transform effectively captures smooth contours that are the dominant feature in natural images. This leads us to more efficient image registration.

Keywords: Image Registration, Normalized Cross Correlation, Steerable Pyramid, Contourlet Transform.

1. INTRODUCTION

Image Registration is the determination of a geometrical transformation that aligns points in one view of an object with corresponding points in another view of that object or another object. Relationship between type of distortion and the type of image registration is the most important

task. There are two types of distortions can be distinguished. First type is those which are type of misregistration i.e. they are the cause of misalignment between two images. The second type of distortions is usually effect of intensity values. The basic need of image registration is for integrating information taken from different sources, finding changes in images taken at different time or at different conditions. Image registration is the process of overlaying images (two or more named reference and sensed images) captured from the same scene but at different times and view points, or even by using different sensors. Therefore, it is a crucial step of most image processing tasks in which the final information is obtained from a combination of various data sources and images are not designed. These include image fusion, change detection, robotic vision, archeology, medical imaging, and multichannel image restoration. Typically, image registration is required in remote sensing applications such as change detection, multispectral classification, environmental monitoring, weather forecasting, super resolution images, and integrating information into Geographic Information Systems (GIS). It is also used in biomedical image processing applications for combining computer tomography (CT) and magnetic resonance imaging (MRI) data to obtain more complete information about the patient, monitoring tumor growth, treatment verification, and comparison of patients' data with anatomical atlases. It is also used in cartography (for map updating), computer vision (for target localization), automatic quality control, motion analysis, and target tracking.

The most fundamental characteristics of image registration technique is the type of transformation used to properly overlay two images. The primary general transformations are affine, projective, perspective and polynomial. These are all defined well mapping of one image into another. Basic image registration can be categorized in two modalities, Monomodal Image Registration and Multimodal Image Registration. Modalities refers to the means by which the images to be registered are acquired. When image registration is done with the two images of object with same sensor, it can be treated as Monomodal image registration whereas with the images from different sensors it is known as Multimodal image registration. Multimodality registration methods are often used in medical imaging as images are obtained from different scanners. It includes CT, MRI or whole body Positron Emission Tomography (PET). These images are used for tumor localization, segmentation of specific part of body and registration of ultrasound and CT images for prostate localization in radiotherapy [1,24,25,28].

1.1 Steps of Image Registration

In general, the majority of automated registration methods consist of the following four steps: feature extraction, feature matching, transform modeling, and image resampling. In the feature extraction step, manually or preferably automatically, salient and distinctive features are extracted. In the feature matching step, the correspondence between the features detected in the sensed image and those detected in the reference image is established. In order to match the extracted features, a similarity measures (such as iterative closest points (ICP), Bessel algorithm, geometric ICP, least squares error (LSE), normalized cross correlation (NCC), average magnitude difference function (AMDF), and maximization of mutual information (MMI)) are used. In the transform modeling step, the parameters of the mapping function are computed by means of the established feature correspondence. Finally, in the image resampling step, the sensed image is registered by means of the mapping function. Due to the saliency of edge features in panchromatic images and also their stability against environmental and illumination changes image edges can be extracted as primary features. Then extracted edge corners will be required control points (CPs). Consequently, more accurately extraction of edges leads to better CP detection which in turn improves the registration results. The accuracy of extracted edges and the required computational cost of the algorithm form the measures to evaluate different methods.[1, 25, 28]

1.1.1 Feature Extraction

To manually or preferably automatically extract salient and distinctive features such as closed boundary regions, edges and points. For further processes, these features can be represented by their point representatives (centers of gravity, line endings, distinctive points, moment or differential descriptors for curves), which are called control points. The proposed

method can use contourlet transform to extract strong, smooth, and connected edges from input images. There are many edge extraction methods reported. These could be grouped in two main categories: point-wise and region wise. In point wise methods, only isolated pixel values take part in the edge extraction process. They include highpass and bandpass filtering as well as Robert, Sobel, Prewitt, and Canny edge detectors. These methods have low computational costs but they cause ringing effects on extracted edges and also amplify high frequency noise (see Table 3). These methods result in disconnected edges, too. Only the Canny edge detector of this group leads to connected extracted edges, however in a blind manner, and thus may lead to wrong edges and corners. In region-wise methods, the edges are extracted using a small neighborhood of pixels. Rank-based filters, statistical methods, Fourier transform, Spline-interpolation, Laplacian based and wavelet-based are samples of region wise methods. All of these methods could extract salient and acceptable edges. However, they need heavy preprocessing leading to high computational costs. For example, in the rank-based method, instead of raw pixel values, the value near median of a neighborhood around the edge pixels is selected. In statistical methods, the distribution function of two neighboring objects is used to determine the edge pixels. Fourier transform-based edge extraction methods use the frequency responses of edge pixels. In Spline-interpolation-based methods, first the edge pixels are interpolated along a Spline and then the edge pixels are extracted. In Laplacian and wavelet based methods, the edge extraction and its verification is done in a multi resolution manner. In present work we propose contourlet transform for feature extraction.

1.1.2 Feature Matching

In this step, the correspondence between the features detected in the reference and sensed images are established. Various feature descriptors (along with spatial relationships among them) and similarity measurements are reported for this purpose. Feature descriptors include: B-Splines, chain code, snake-model, length code algorithms. Similarity measures include: iterative closest points (ICP), Bessel algorithm, geometric ICP, least squares error (LSE), normalized cross correlation (NCC), average magnitude difference function (AMDF), and maximization of mutual information (MMI).

1.1.3 Transform Modeling

In this step, the type and parameters of mapping function which aligns the sensed image to the reference image are estimated. The parameters of the mapping function are computed by means of the established feature correspondence.

1.1.4 Image Resampling

In this step, the sensed image is registered by means of the mapping function. Image values in non-integer coordinates are computed by the appropriate interpolation technique.Various survey papers for Image registration n research papers are studied as described in chapter 2. Different method of Image registration and its applications are discussed. Chapter 3 overviews various methods used for Image Registration. It includes the use of Normalizes Cross Correlation, Mutual information and Contourlet Transform for Image Registration. Also it explains about algorithm used.Based on the algorithm, various experiments are done on different images. Some are done for monomodal an some for multimodal and results obtained and discussed in chapter 4. According to the experiments done we reached to some conclusion and future scope explained in chapter 5.Chapter 6 introduces us with various references used for the study.

2. LITERATURE SURVEY

Image Registration is presented efficiently by Lisa Gottesfeld Brown and Image Registration deals with the matching of two images. Registration is the determination of a geometrical transformation that aligns points in one view of an object with corresponding points in another view of that object or another object [1, 2, 5].

Jignesh N Sarvaiya, Dr. Suprava Patnaik & Salman Bombaywala illustrates efficient use of Normalized cross correlation for image registration. Template matching is done using NCC. In

this paper, correspondence between main image and template image is established which gives the degree of similarity between them. Then, the minimum distortion, or maximum correlation, position is taken to locate the template into the examined main image [16]. Luigi Di Stefano, Stefano Mattoccia & Martino Mola used cross correlation as a method for feature matching. This method is based on the rotation and scale invariant normalized cross-correlation. Both the size and the orientation of the correlation windows are determined according to the characteristic scale and the dominant direction of the interest points [4]. There are various methods of Image Registration. Some are discussed by Lisa Gottesfeld Brown and Barbara Zitova, Jan Flusser [1, 5]. Methods are Correlation, Fourier Method, Point mapping etc. Correlation can be used effectively. Point mapping method is less sensitive to local distortions as they use control points and local similarity, they use information from special relationship between control points and they are able to consider possible matches based only on supporting evidence. It can be efficient to use method point mapping for image registration. J. Flusser used moment based approach to correct affine distortion, he has also done degraded image analysis to locate invariants in images [10, 11].

P. Ramprasad, H.C. Nagaraj and M.K. Parasuram presented a new wavelet based algorithm for registering noisy and poor contrast dental x-rays. Proposed algorithm has two stages, first stage is preprocessing stage, which removes the noise from x-ray images, Gaussian Filter has been used. Second stage is a geometric transformation stage [12]. Sangit Mitra and B.S. Manjunath explained various contour based approaches for multispectral image registration in their different papers [13]. Sh. Mahmoudi-Barmas and Sh. Kasaei have explained the edge extraction for Image Registration using Contourlet Transform. Control points are used to detect edges [14]. For multimodal images registration is explained by P.Pradeepa and Dr. Ila Vennila .Proposed method said that for two images which are multimodal registration can be done with the mutual information between two images [15]. Also Yonggang Shi have explained about multimodal image registration. Proposed method said that Mean and variance of Joint Intensity Distribution can be efficiently used for the multimodal images [16]. Nemir Ahmed Al-Azzawi with Harsa Amylia and Wan Ahmed K. Abdullah have explained about the monomodal image restration. Nonsubsampled Contourlet Transform with the help of Mutual information can be effectively used for monomodal image registration [17]. Manjusha Deshmukh and Udhav Bhosale explained about image registration and use of Mutual Information for image registration [24,25,30].

3. PROPOSED METHOD

Image registration is widely used in remote sensing, medical imaging, computer vision etc. In general, its applications can be divided into main groups according to the manner of the image acquisition. It can be different viewpoints (multiview analysis), different times (multitemporal analysis), or from different sensors (multimodal analysis) etc. And accordingly registration method can be used.

3.1 Cross Correlation Based Template Matching

Correlation methods are useful for registration of monomodal images, for comparison of several images of the same object. One of the fundamental task in image registration is Template matching. Template matching is the process of finding the location of a sub image, called a template, inside an image. Template matching involves comparing a given template with windows of the same size in an image and identifying the window that is most similar to the template. The basic template matching algorithm consists in measuring the degree of similarity between the template image and the reference image.

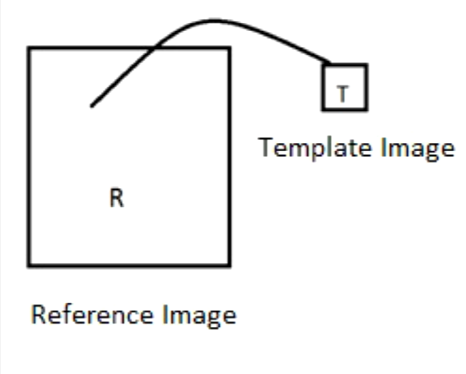


FIGURE 3.1: Template Matching T < R.

Then, the maximum correlation position is taken to locate the template into the examined image. Normalized Cross Correlation (NCC) is often the adopted for similarity measure due to its better robustness [4, 5]. Cross-correlation is the basic statistical approach of registration. It is often used for template matching or pattern recognition in which the location and orientation of a template or pattern is found in picture. Cross correlation is a similarity measure or match metric. For template T and image I, where T is small compared to I, the two dimensional normalized cross-correlation function measures the similarity for each translation.

$$\Gamma(u,v) = \frac{\sum[f(x,y) - fuv][t(x-u,y-v) - t]}{\left\{[f(x,y) - fuv]^2 [t(x-u,y-v) - t]^2\right\}^{0.5}} \quad \text{..... (3.1)}$$

Normalized cross-correlation can be used as basic approach for image registration. NCC is the simplest but effective method foe similarity measure. Here we are trying to find out the location of the template image in main image.Normalized cross correlation is done for the template and the main image. Then the maximum correlated point is determined which indicates the proper position of the template in main image. The offset for the point of maximum correlation will be calculated. And accordingly template is highlighted in main image. Consider template image T which we want to match with the reference image R, as shown in Figure 3.1. Generally T is very small image compare to given reference image R. Normalized Cross Correlation is done between two and the point of maximum correlation is obtained.

Algorithm

- Calculate cross-correlation in the spatial or the frequency domain, depending on size of images.
- Calculate local sums by precomputing running sums.
- Use local sums to normalize the cross-correlation to get correlation coefficients.
- Find location of point at which maximum normalized cross correlation is obtained.
- Locate the template at that location on reference image.

Above NCC method can be improved by using Feature Based Method. Image can be effectively represented by any of its feature such as edges, points, curves etc. and these featues can be effectively used for image registration. In above method, before two images be cross correlated, they are applied with the filters to extract edges.

3.2 Mutual Information

Image registration methods based on mutual information criteria have been widely used in multimodal medical image registration and have shown hopeful results. Although the information content of the images being registered is constant, the information content of the portion of each image that overlaps with other image will change with each change estimated registration

transformation. Therefore a suitable technique for measuring joint entropy is to measure with respect to marginal entropy. This measure is known as Mutual Information *I* (*A,B*) .It can be independently and simultaneously proposed for multimodal medical image registration.

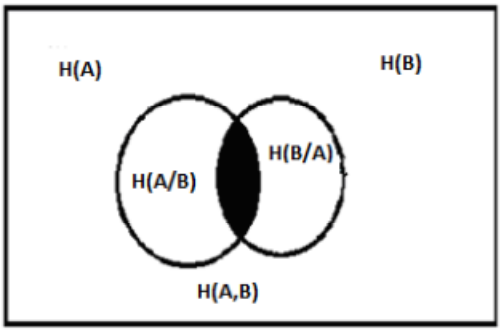


FIGURE 3.2: Mutual Information.

mutual information is employed as a similarity measure to be optimized. Since MI made its entrance into the field of medical image registration, it has been adopted by a large number of researchers [15-17, 30].

$$I(A,B) = H(A) + H(B) - H(A,B) \qquad \ldots\ldots(3.2)$$

Where $H(A,B)$ is joint entropy and $H(A/B)$ & $H(B/A)$ are conditional entropy.

$$H(A) = -\sum P_A(a)\, log P_A(a) \qquad \ldots. (3.3)$$

$$H(A/B) = -\sum P_{AB}(a,b)\, log P_{AB} \qquad \ldots\ldots (3.4)$$

$$H(A/B) = -\sum P_{AB}(a)\, log P_{A/B}(a/b) \qquad \ldots.. (3.5)$$

Mutual information is a direct measure of the amount of information common between the two images as shown in Figure 3.2. During image registration, however, different transformation estimates are evaluated, and these transformation estimates will result in varying degree of overlap between images, though it is better than joint entropy. The problem has been addressed by proposing various normalized form of mutual information that are more overlap independent. Mutual information can be consider as a measure of how well one image explains the other; it is maximized at the optimal alignment.

Algorithm

- Read Reference main image and Image to be registered.
- Use Mutual information as simillarity criteria.
- Calculate the angle of rotation and scale.
- Rotate the image with calculated angle of rotation and scale image accordingly.

3.3 Steerable Pyramid

In many early vision and image processing tasks we require oriented filters. It needs to apply the same filter, rotated to different angles under adaptive control, or wishes to calculate the filter response at various orientations. Here is an efficient architecture to synthesize filters of arbitrary orientations from linear combinations of basis filters, al-lowing one to adaptively “steer" a filter to any orientation, and to determine analytically the filter output as a function of orientation [23]. A new registration approach based on just two resolution scales is proposed. Firstly, the image datasets will be decomposed into multi-scale, multi-orientation subimages by the steerable pyramid. To improve registration accuracy and efficiency, only the lowest resolution scale and the highest-resolution scale will be used for the registration purpose. In each of these two scales, a new “magnitude sub-band” will be constructed from the multi-orientation band-pass coefficients to

extract edge information. Then, to improve registration accuracy, both raw coefficients in low-pass sub-bands and the edge features in magnitude sub-band will be used as registration feature space. The steerable pyramid can be used to construct the multiscale, multi-orientation image representation with the property of rotation- and translation-invariance. In this pyramid representation, the image is decomposed into subbands by the basis functions which are directional derivative operators with different sizes and orientations.

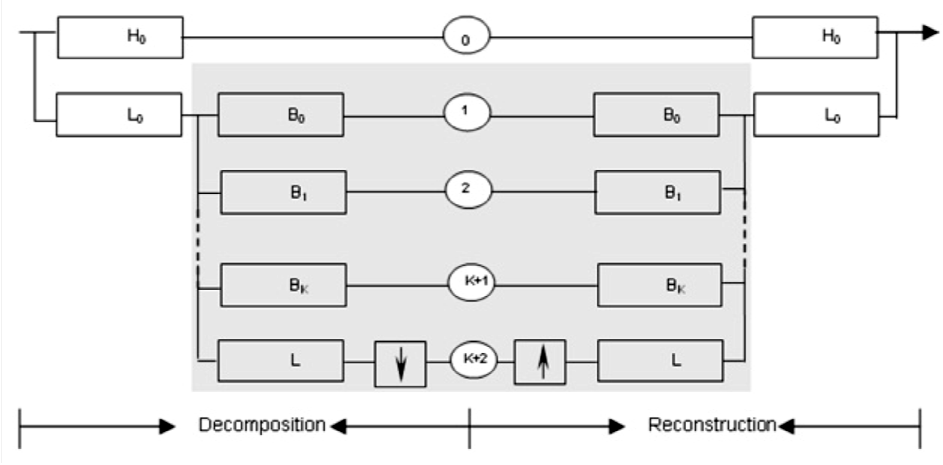


FIGURE 3.3: System diagram for the steerable pyramid.

As all the basis functions are derived by translations, dilations, as well as rotations of a single function, steerable pyramid can be thought as a "wavelet". Similar to conventional orthogonal wavelet decompositions, the steerable pyramid is implemented by recursively dividing an image into a set of oriented subbands and a low-pass sub-image . The input image is firstly decomposed into a lowpass subband L_0 and a highpass subband H_0 . The lowpass band is then decomposed into N directional bandpass subbands B_k where k = 0,..., N –1 and a lowpass subband L . The subbandsB_k describe image directional features and are designed to be polarseparable in Fourier domain. The subbands decimated by factor 2 will be decomposed into a lower-resolution level. Finally, the multi-directional and multi-scale representation of the image can be derived from the recursive procedure. The Steerable Pyramid is a linear multi-scale, multi-orientation image decomposition that provides a useful front-end for image-processing and computer vision applications. Filters with orientation tuning are often used in the detection of lines and edges. Canny's edge operator can be used efficiently, which is optimize to detect step edges; Canny's system can also be used with different filter choices to detect features other than step edges. Filter that is optimized for use with an edge will give spurious responses when applied to features other than edge. For example, when the Canny edge filter is applied to a line rather than an edge, it produces two extrema in its output rather than one, and each is displaced to the side of the actual line position. On the other hand, if a filter is optimized for detecting lines, it will give spurious responses with edges. Since natural images contain a mixture of lines, edges, and other contours, it is often desirable to find a contour detector that responds appropriately to the various contour types.

Algorithm

- Read Reference main image and Image to be registered.
- Apply Steerable filter on both images.
- Use Mutual information as simillarity criteria.
- Calculate the angle of rotation and scale.
- Rotate the image with calculated angle of rotation and scale image accordingly.

3.4 Contourlet Transform

For Many signal processing tasks such as compression, denoising, feature extraction and enhancement image can be efficiently represented by Contourlet Transform (CT), which is one of several transforms aimed at improving the representation sparsity of images over the Wavelet Transform (WT). The main feature of these transforms is the potential to efficiently handle 2-D singularities, i.e. edges, unlike wavelets which can deal with point singularities exclusively. This difference is caused by two main properties that the CT possess: 1) the directionality property, i.e. having basis functions at many directions, as opposed to only 3 directions of wavelets 2) the anisotropy property, meaning that the basis functions appear at various aspect ratios (depending on the scale), whereas wavelets are separable functions and thus their aspect ratio equals to 1. The main advantage of the CT over other geometrically representations, is its relatively simple and efficient wavelet-like implementation using iterative filter banks. Due to its structural resemblance with the wavelet transform, many image processing tasks applied on wavelets can be easily adapted to contourlets.

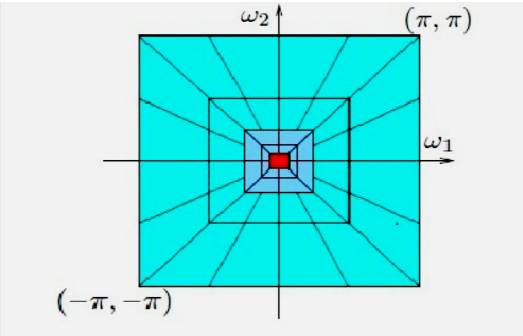


FIGURE 3.4: Frequency decompositions by the contourlet transform.

The contourlet transform was proposed as a directional multiresolution image representation that can efficiently capture and represent singularities along smooth object boundaries in natural images. Its efficient filter bank construction as well as low redundancy makes it an attractive computational framework for various image processing applications. The contourlet transform is implemented via a twodimensional filter bank that decomposes an image into several directional subbands at multiple scales. This is accomplished by combining the Laplacian pyramid with a directional filter bank at each scale. Due to this cascade structure, multiscale and directional decomposition stages in the contourlet transform are independent of each other. One can decompose each scale into any arbitrary power of two's number of directions, and different scales can be decomposed into different numbers of directions. This feature makes contourlets a unique transform that can achieve a high level of flexibility in decomposition while being close to critically sampled. Figure above shows an example frequency partition of the contourlet transform where the four scales are divided into four, four,eight, and eight directional subbands from coarse to fine scales, respectively. Figure 3.5 shows examples of possible frequency decompositions by the contourlet transform.

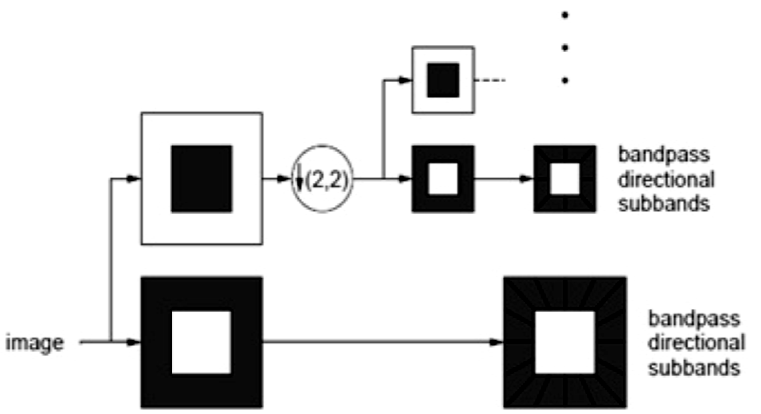


FIGURE 3.5: Contourlet transform.

In particular, by altering the depth of the DFB decomposition tree at different scales (and even at different orientations in a contourlet packets transform), we obtain a rich set of contourlets with variety of support sizes and aspect ratios. This flexibility allows the contourlet transform to fit smooth contours of various curvatures well. The major drawback of DWT in two dimensions is their limited ability in capturing directional information. In light of this, Do and Vetterli [24] developed the Contourlet Transform, based on an efficient two-dimensional multiscale and directional filter bank (DBF). Contourlet Transform not only possess the main features of DWT, but also offer a high degree of directionality and anisotropy. It allows for different and flexible number of directions at each scale, while achieving nearly critical sampling. In addition, Contourlaet Transform uses iterated filter banks, which makes it computationally efficient (O(N) operations for an Npixels image) [24].

Contourlet Transformgives a multiresolution, local and directional expansion of image using Pyramidal Directional Filter Bank (PDFB). The PDFB combines Laplacian Pyramid (LP) which captures the point discontinuities, with a DFB which links these discontinuities into linear structures. Figure 3.7, shows the flowchart of Contourlet Transform for a 512 × 512 image. As shown in Fig. 1, first stage of Contourlet Transform is LP decomposition and DFB is the second stage. The Contourlet Transform is a multiscale, multi resolution filter that comprised of Pyramidal filter and Directional filter. The Proposed algorithm used Laplacian Pyramidal filter as Pyramidal filter and Steerable filter as directional filter. The Contourlet Transform enhances the image with its property of decomposition and reconstruction. Laplacian Pyramid Filter The Laplacian Filter highlights regions of rapid intensity changes.

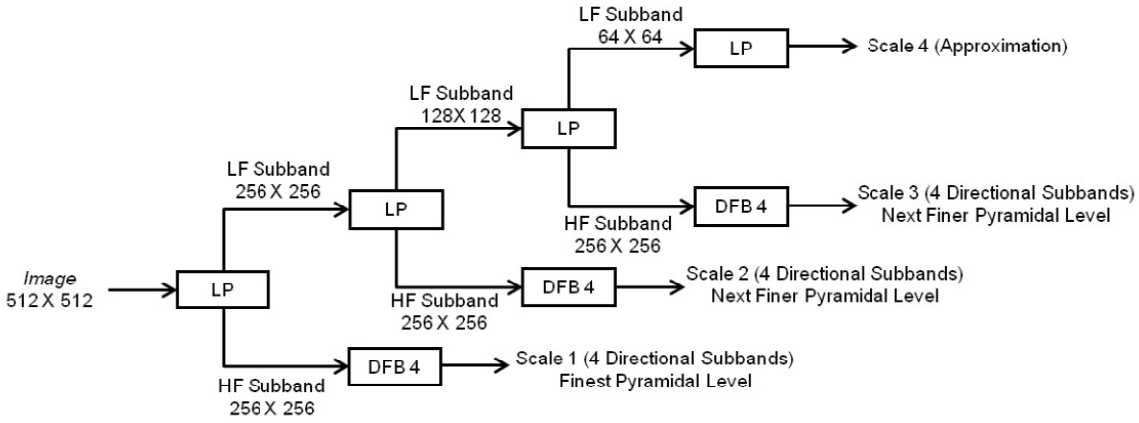


FIGURE 3.6: Flowchart of Contourlrt Transform for a 512 X 512 image.

The Laplacian Filters smooth the input image using a Gaussian smoothing filter in order to reduce its sensitivity to noise.Laplacian pyramid filter is used to capture the edge point. Directional Filter Bank is used to link the discontinuities point in linear structures.

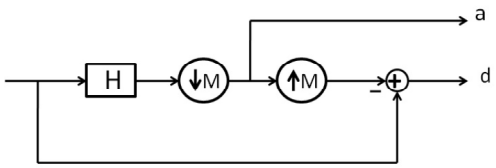


FIGURE 3.7: Construction of LP.

The Laplacian Filter decomposes the images into information at multiple scales. This filter extracts features of interest and to attenuate noise that present in the image. The applications of this filter can be image enhancement, restoration and image analysis. The efficient image coding for the image modelling is achieved by the property of the redundancy reduction of this filter. The Laplacian filter consists of Low pass filter and High pass filter. The decomposition is based on the difference between the two filters. The image is recursively decomposed into low-pass and high-pass bands. In each decomposition level the LP creates a down sampled low-pass version of the original image and a band-pass image.In Laplacian Pyramid method each input image is decomposed into a subband of low frequency of original image and a bandpass high frequency subbands. The same process is repeated for low frequency subband for the specified Contourlet decomposition level. The Laplacian Pyramid decomposition process is shown in Figure 3.6. The input image is applied to a LP filter H and then down sampled to derive a coarse approximation a (Lowpass Subband). Then the image is up sampled . The resultant highpass subbands are derived from subtracting the output of the synthesis filter with the input image. The output of Laplacian filter is follwed by Directional Pyramidal Filters leads us to the contourlet transform.Here Steerable Pyramidal filter is used.

Algorithm

- Read Reference main image and Image to be registered.
- Obtain contourlet transformed of both images.
- Use Mutual information as simillarity criteria.
- Calculate the angle of rotation and scale.
- Rotate the image with calculated angle of rotation and scale image accordingly.

4. RESULTS AND DISCUSSION

The proposed method was applied to register various sets of images. The following figures present the reference image, target image and registered image.

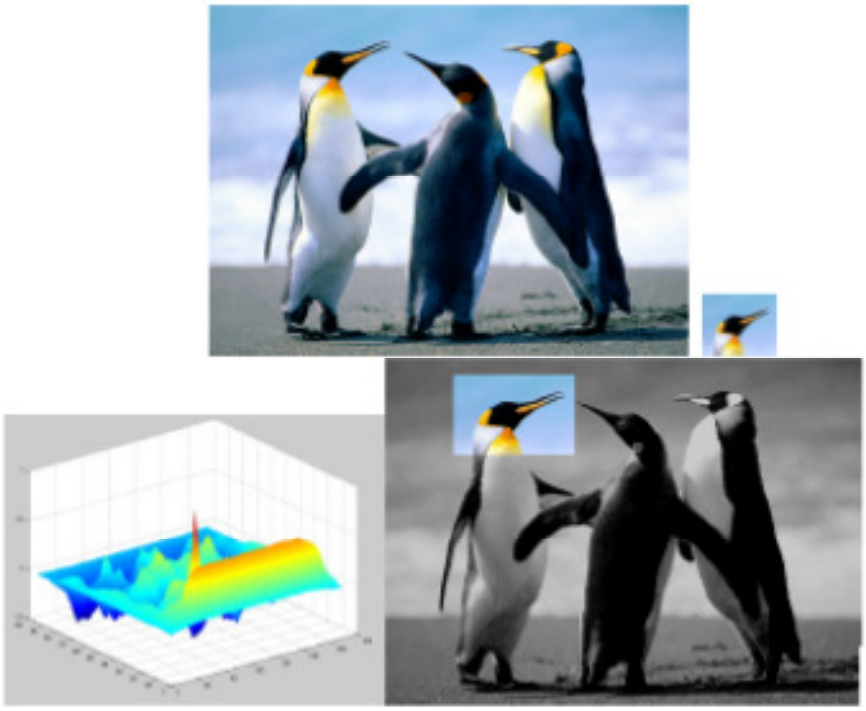
4.1 Cross Correlation Based Template Matching of Monomodal Images

Penguins' image is considered as original image for first dataset. Some part of this image is cropped which is considered as Template image. Figure 4.1 (a) show the original image and the template is extracted from original image is shown in Figure 4.1 (b). Using Normalized cross-correlation the NCC plot is obtained as shown in Figure 4.1 (c) and offset points are obtained using maximum peak in the NCC plot. Finally generated image for registration is shown in Figure 4.1 (d). Some results are obtained for various datasets as shown in Table I, which shows the offset for the point of maximum cross correlation. This offset helps to find out the place of template image in original image. Similarly for another image which is a real time image of trees, same process is done, Normalized Cross Correlation is done and plot obtained and results are obtained as shown in Table 1.

TABLE 1: Cross Correlation Coefficient.

	Actual offset	Observed offset		
	Left	Top	Left	Top
Dataset 1	151	73	151	73
Dataset 2	87	69	87	69

Dataset 1

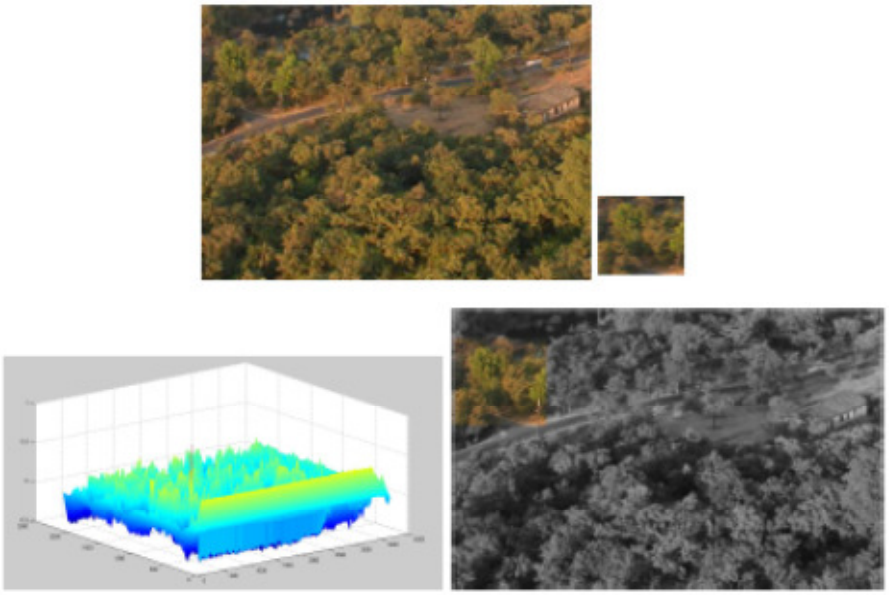


a	**b**
c	**d**

FIGURE 4.1: (a) Reference Image (b) Template Iimage (c) NCC Plot (d) Registered Image.

Figure 4.2 shows the offset obtained for dataset II. Figure 4.2 (a) show the original image and the template is extracted from original image is shown in Figure 4.2 (b). Using Normalized cross-correlation the NCC plot is obtained as shown in Figure 4.2 (c) and offset points are obtained using maximum peak in the NCC plot. Finally generated image for registration is shown in Figure 4.2 (d).

Dataset 2



a	b
c	d

FIGURE 4.2: (a) Reference Image (b) Template (c) NCC Plot (d) Registered Image.

Observation

NCC is effective tool for similarity measure between two images. Two main drawbacks of the correlation-like methods are the flatness of the similarity measure maxima (due to the self-similarity of the images) and high computational complexity. The maximum can be sharpened by preprocessing or by using the edge or vector correlation. It is used in next section.

4.2 Feature Based Template Matching of Monomodal Images

Various feature of image can be extracted and effectively used for registration.This is the basic step in image Registration. Featue can be edges or points or curves in image. Edge detection or extraction is a preliminary step in various image processing applications. There are various methods for edge extraction such as Laplacian operator, gradient operator and various filters, out of which canny edge detector can be used very effectively. The Canny method finds edges by looking for local maxima of the gradient of I. The gradient is calculated using the derivative of a Gaussian filter. The method uses two thresholds, to detect strong and weak edges, and includes the weak edges in the output only if they are connected to strong edges. This method is therefore less likely than the others to be fooled by noise, and more likely to detect true weak edges. Canny edge filter is applied on images. Using this filter edges are extracted and that images can be used for further processing. Correlation can be carried out on these images which results in better Normalised Cross Correlation graph and hence better matching.

Dataset 1

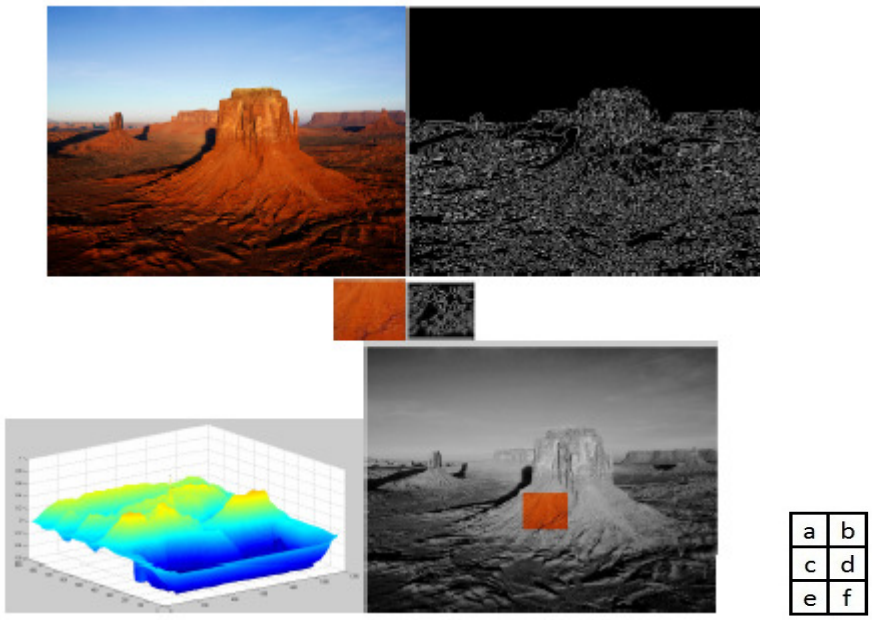


FIGURE 4.3: (a) Reference Image (b) Filered Reference image (c)Template Image (d)Filtered Template Image (e) NCC Plot (f) Registered Image.

Dataset2

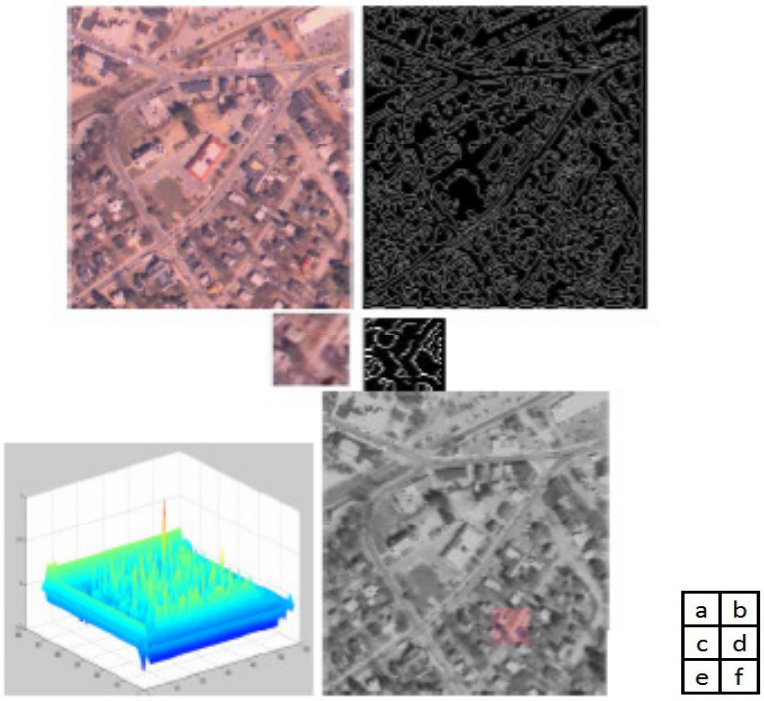


FIGURE 4.4: (a) Reference Image (b) Filtered Reference image (c)Template Image (d) Filtered Template Image (e) NCC Plot (f) Registered Image.

Observation
The maximum is sharpened by using the edge detection filter. This method fails for the multimodal images. We can use another similarity measure for such images. It is used in next section.

4.3 Mutual Information Based Monomodal Image Registration to Correct Rotation And Scale

Two monomodal images are registerd using Mutual Information. Mutual information is used here as a similarity measure between two images.

Dataset

FIGURE 4.5: (a) Reference Input Image (210x210) (b) Image 2: Rotated Image (234x234) (c) Registered Image (210x210).

Observation
Mutual Information is effective tool for similarity measure between two monomodal images.
Above image is roteted by angle 7 degrees. After applying Mutual information algorithm,Image is registerd back but with 1 degree of difference.

4.4 Mutual Information Based Multimodal Image Registration to Correct Rotation And Scale

Mutual Information can be used as similarity measure for monomodal as well as multimodal images. Multimodal here refers to images taken by different source. Two images of same brain are used for the experiment. One is MRI of brain and other one is PET image of same brain.Mutual information is calculated for them and roteted and scales image is registered. Initialy experiment is done for roteted image later followed by roteted and scaled image.

Dataset 1

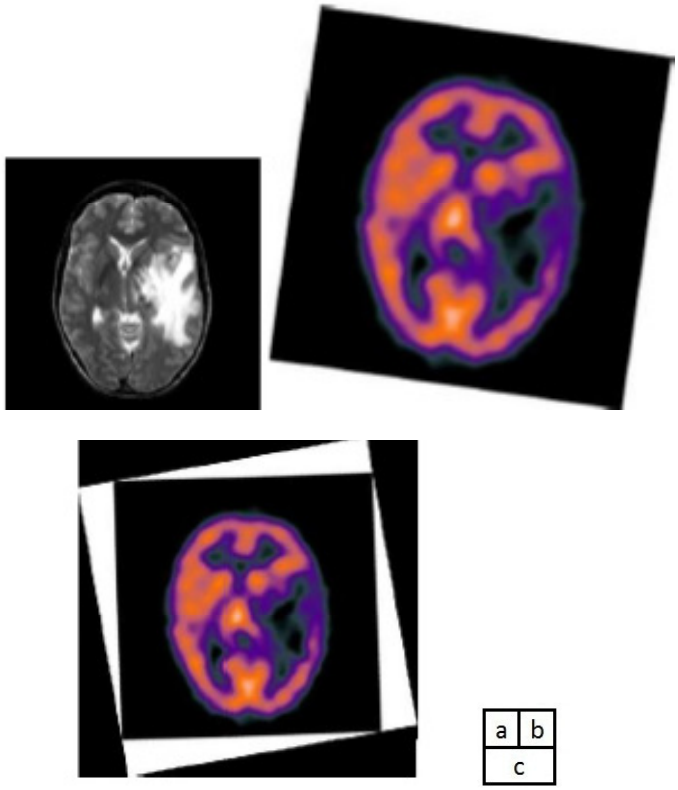


FIGURE 4.6: (a) Reference Input Image (210x210) (b) Image 2: Rotated Image (234x234, Rotation 8 degree) (c) Registered image (304x303).

Observation

Mutual Information is effective tool for similarity measure between two images, for monomodal as well as multimodal images. It has limitation of speed, when time is an important constraint one cannot use this method. Although it has some limitations entropy and mutual information are best approaches for multimodal image registration. The maximum can be sharpened by preprocessing. It is used in next section.

4.5 Steerable Transform Based Image Registration to Correct Rotation and Scale

For multimodal images mutual information can be effectively used as similarity measue and registration can be done.This registration can become more correct n effective if we use feature of images. Here features are extracted using steerable transform and then it is followed by mutual information as mapping function leads to increase in accuracy.

Dataset 1

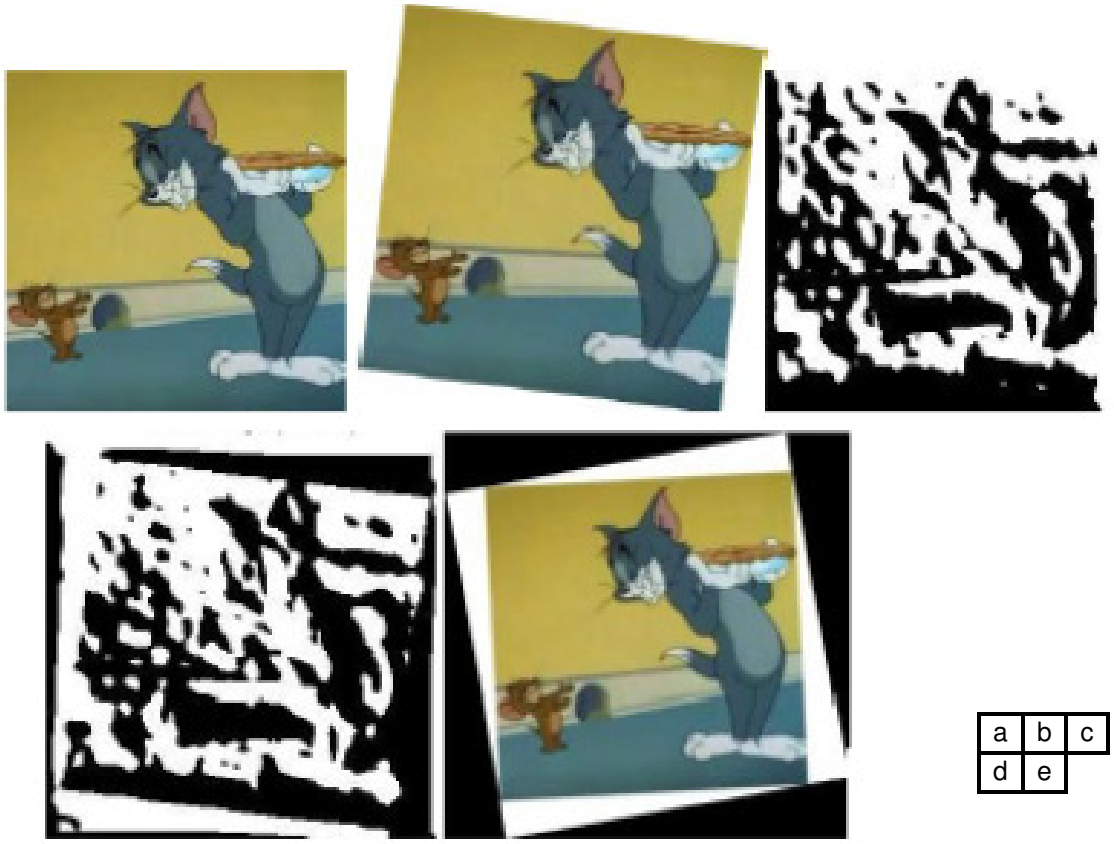
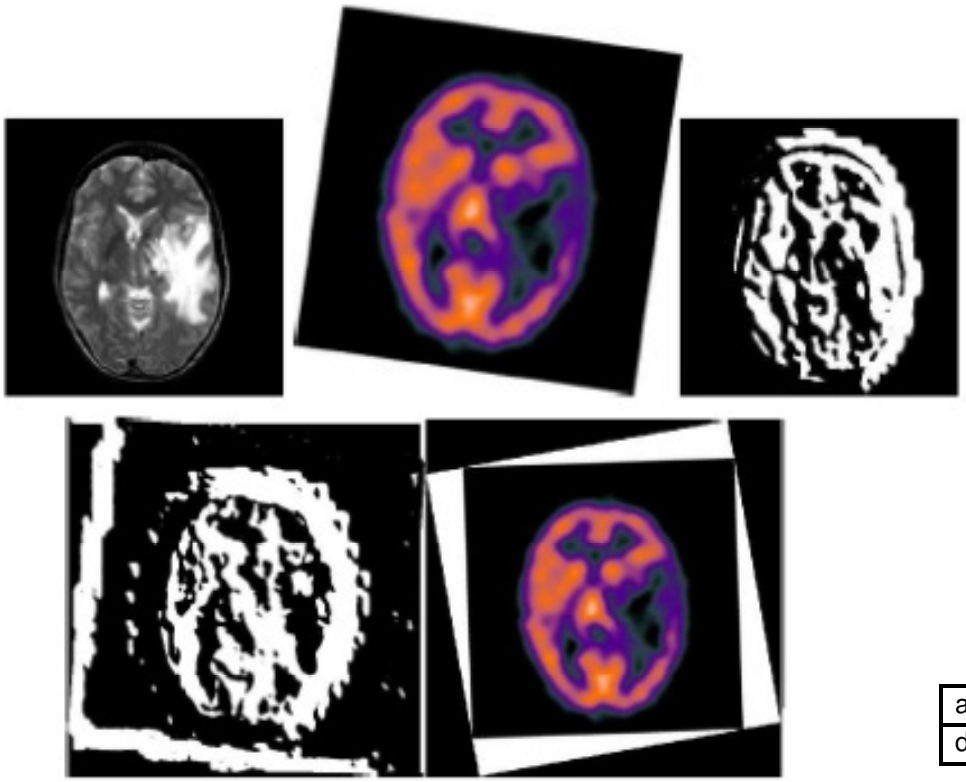


FIGURE 4.7: (a) Reference Input Image (210x210) (b) Image 2: Rotated Image (234x234) (c) Transformed Reference Input Image (210x210) (d) Transformed Image 2: Rotated Image (234x234) (e) Registered image (210x210).

Dataset 2



a	b	c
d	e	

FIGURE 4.8: (a) Reference Input Image (Image size 210x210 pixels) (b) Image 2: Rotated Image (Image size 500x500 pixels, Rotation 8degree, scaled by 2) (c) Transformed Reference Input Image (Image size 374x296 pixels) (d) Transformed Image 2: Rotated Image (Image size 398x320 pixels) (e) Registered image (Image size 551x473 pixels).

Observations

It is observed that combinational approach of Steerable transform and mutual information gives better results. This combination can be used in case of multimodal image registration. To improve accuracy of registration we can use contourlet transform which is explained in next section.

4.6 Contourlet Transform Based Image Registration to Correct Rotation and Scale

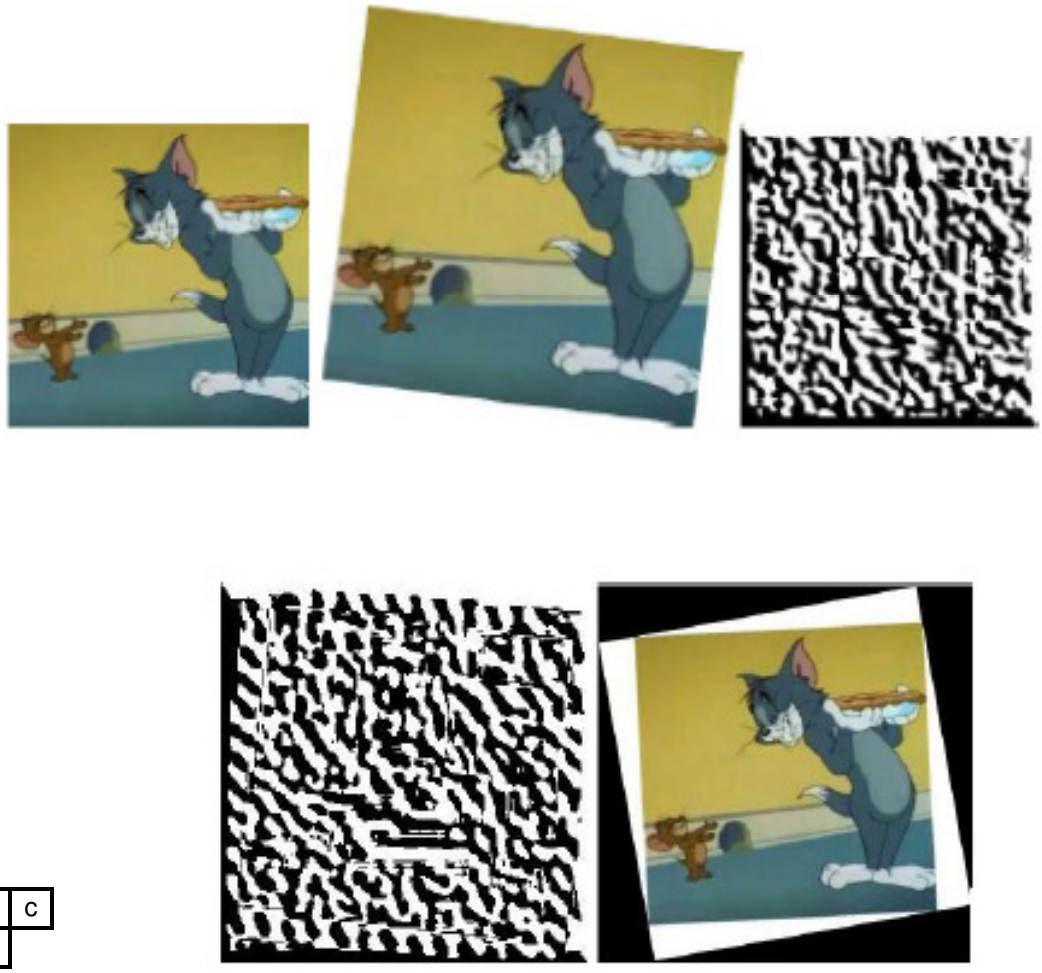
For multimodal images mutual information can be effectively used as similarity measure and registration can be done. This registration can become more correct n effective if we use feature of images. Here features are extracted and then it is followed by mutual information as mapping function leads to increase in accuracy.

Dataset 1

a	b	c
d	e	

FIGURE 4.9: (a) Reference Input Image (256X256) (b) Image 2: Rotated Image (296X296) (c) Transformed Reference Input Image (256X256) (d) Transformed Image 2: Rotated Image (296X296) (e) Registered image (256X256).

Dataset 2



a	b	c
d	e	

FIGURE 4.10: (a) Reference Input Image (210x210) (b) Image 2: Rotated Image (234x234) (c) Transformed Reference Input Image (210x210) (d) Transformed Image 2: Rotated Image (234x234) (e) Registered image (210x210).

Dataset 3

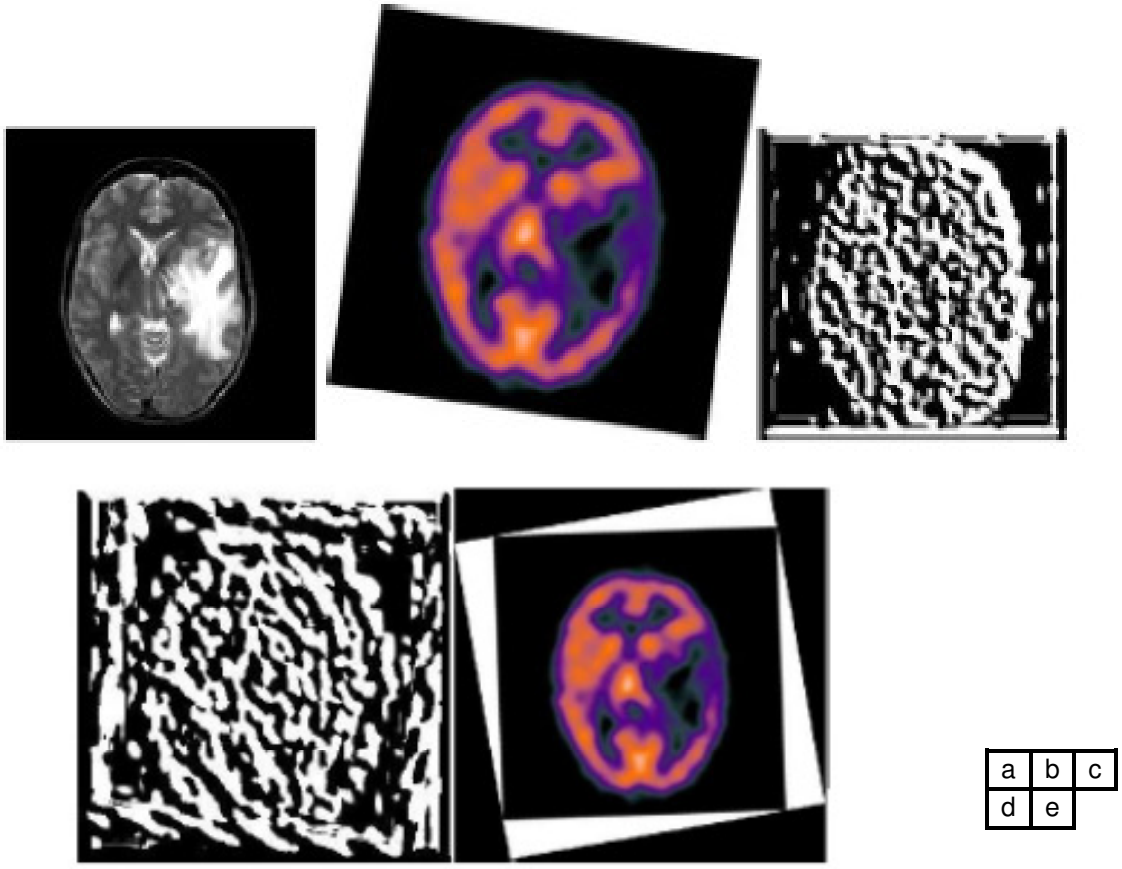


FIGURE 4.11: (a) Reference Input Image (Image size 210x210) (b) Image 2: Rotated Image (Image size 500x500, Rotation 8degree, scaled by 2) (c) Transformed Reference Input Image (Image size 374x296) (d) Transformed Image 2: Rotated Image (Image size 398x320) (e) Registered image (Image size 551x473).

Observations

It is observed that combinational approach of contourlet transform and mutual information gives better results. Even contourlet and mutual information combination can be used in case of multimodal image registration. If we involve mutual information as similarity criteria, speed decreases and hence we can not use this combination in applications where speed is required.

5. CONCLUSION AND FUTURE SCOPE

We have described a template-matching algorithm, for image registration based on NCC on similarity measures. It is clear that Normalized Cross-Correlation (NCC) is the good approach to image registration by template matching. It gives perfect template matching in the given reference image. The Maximum Cross-correlation coefficient values indicate the perfect template matching. This approach cannot be efficient for multimodal image registration. We get number of points with maximum correlation coefficient, instead only one matched point. So it will be insufficient to obtain exact offset and hence, for multimodal image registration we can go for registration using feature extraction or we can use any transform for the proper threshold value or mapping functions can also be used. Feature based method also can be used for image registration. Feature can be extracted from an image, which can help us to match two images. This feature can be points or edges in that image which can help us to match two images. When registering images with non-linear, locally dependent geometric distortions, we are faced with two basic problems—how to

match the CPs and what mapping functions to use for registration. In multimodal registration, MI technique has become a standard reference, mainly in medical imaging. Some authors combined the MI with other, preferably feature-based, methods to gain higher robustness and reliability. Mutual information can be considered as a measure of how well one image explains the other; it is maximized at the optimal alignment. This method can effectively register image but little lagging in accuracy. To increase accuracy, features of image can be extracted followed by Mutual Information as similarity measure. The dependencies across scales, space, and directions were quantitatively compared using mutual information measures. Contourlet Transform can be efficiently used for image registration. For a computational image representation to be efficient, it should based on a local, directional and multiresolution expansion. The need for image registration is to capture fine curves in image with multiresolution, which can be efficiently done with Contourlet Transform. With this method scale is corrected and angle of rotation is corrected with the difference of 1 degree. The Contourlet Transform is capable of resolving two dimensional singularities and representing image edges more efficiently. The future development on this field could pay more attention to the feature-based methods, where appropriate invariant features can provide good platform for the registration.

6. REFERENCES

[1] Lisa Gottesfeld Brown, ”A Survey of Image Registration Techniques”, Dept. of Computer Science, Columbia University, New York. Jan 12,1992 W. Förstner, “A Feature-Based Correspondence Algorithm For Image Matching”, *International Archives of Photogrammetry and Remote Sensing*, vol.26, no.3, pp.150-166, 1986.

[2] Feng Zhao, Qingming Huang, Wen Gao “Image Matching By Normalized Cross-Correlation”, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China Graduate School of the Chinese Academy of Sciences, Beijing, China

[3] Luigi Di Stefano, Stefano Mattoccia, Martino Mola DEIS-ARCES, “An Efficient Algorithm for Exhaustive Template Matching Based on Normalized Cross Correlation”, *IEEE Computer society*, Proceedings of the 12th International Conference on Image Analysis and Processing, 2003A.

[4] Zitova B., and Flusser J., “Image Registration Methods: A Survey,” *Journal of Image and Vision ELSEVIER*, Vol. 21, pp. 977-1000, June. 2003.

[5] Radke R. J., Andra S., Al-Kofahi O. and Royasm B., “Image Change Detection Algorithms: a Systematic Survey,” *IEEE Transaction on Image Processing*, Vol. 14, No. 3, pp. 294-307, March 2005.

[6] Makela T., Clarysse P., Katila T. and Magnin I. E., “A Review of Cardiac Image Registration Methods,” *IEEE Transaction on Medical Imaging*, Vol. 21, No. 9, September 2002.

[7] Zhengzhou, China Eugenio F., Marques F. and Marcello J., “A Contour-Based Approach to Automatic and Accurate Registration of Mmulti-Temporal and Mmultisensory Satellite Imagery,” *IEEE International Geoscience and Remote Sensing Symposium*, 2002. IGARSS ’02. 2002

[8] Guo Xiao-jun, Han Li-li, Ning Yi “Feature Points Based Image Registration between Endoscope Image and the CT Image”, Department of Mechanical Engineering HUANGHE S&T COLLEGE Zhengzhou, China.

[9] J. Flusser, T. Suk, “Degraded Image Analysis: an Invariant Approach”, *IEEE transaction on Pattern Analysis and Machine Intelligence* 20(1998)590-603.

[10] J. Flusser, T. Suk, ”A Moment Based Approach to Registration of Images with Affine Geometric Distortion”, *IEEE transactions on Geoscience and remote sensing* 32(1994)382-387.

[11] H. Manjunath, B. S. Mitra, 1995. “A Contour–Based Approach to Multisensor Image Registration”, *IEEE Transactions on image processing*, 4(3), pp. 320-334.

[12] P. Ramprasad, H.C. Nagaraj and M.K. Parasuram, ”Wavelet based Image Registration Technique for matching Dental x-rays”, *International Journal of Electrical and Computer Engineering* 4:2 2009

[13] Sh. Mahmoudi-Barmas and Sh. Kasaei,” Contourlet-Based Edge Extraction for Image Registration” , *Iranian Journal of Electrical & Electronic Engineering*, Vol. 4, Nos. 1 & 2, Jan. 2008

[14] P.Pradeepa and Dr. Ila Vennila, “Multimodal Image Registration Using Mutual Information”, *IEEE International conference On Advances in Engineering*, Science And Manegement (ICAESM-2012) March 30,31,2012

[15] Yoggang Shi, “Multimodal Image Registration Using Mean and Variance of Joint Intensity Distribution”, School of information and electronics, Beijing Institute of technology,Beijing 100081, China ICSP2010 Proceedings

[16] Nemir Ahmed Al-Azzawi, Harsa Amylia Mat Sakim and Wan Ahmed K. Wan Abdullah, “MR Image Monomodal Registration Based on thr Nonsubsampled Contourlet Transform and Mutual Information”, 2010 *International conference on Computer Applications and Industrial Electronics* (ICCAIE),December 5-7,2010,Kuala Linpur, Malaysia.

[17] Issac N.Bankman, "HandBook of Medical Imaging Processing & Analysis", *Academic Press*, 2000.

[18] J. V. Chapnick, M. E. Noz, G.Q. Maguire, E. L. Kramer, J. J. Sanger, B. A. Birnbaum, A. J. Megibow, “Techniques of Multimodality Image Registration” , *Proceedings of the IEEE nineteenth Annual North East Bioengineering conference*, 18-19 March 1993, 221-222.

[19] Jignesh N Sarvaiya Dr. Suprava Patnaik Salman Bombaywala, ”Image Registration By Template Matching”, 2009 *International Conference on Advances in Computing, Control, and Telecommunication Technologies.* 16

[20] Duncan D.,Y. Po and Minh N Do. Member IEEE, ” Directional Multiscale Modeling of Images using the Contourlet Transform”, *IEEE Transactions on Image Processing*.

[21] S.Anand and R.Aynesh Vijaya Rathna, “ Architectural Distortion Detection in Mammogram using Contourlet Transform and Texture Features”, *International Journal of Computer Applications* (0975 – 8887) Volume 74– No.5, July 2013

[22] William T. Freeman and Edward H. Adelson, “The Design and Use of Steerable Filters”, The Media Laboratory and yDept. of Brain and Cognitive Sciences Massachusetts Institute of Technology Cambridge, Massachusetts 02139 *IEEE Trans. Patt. Anal. and Machine Intell.*, Vol. 13, No. 9, pp. 891-906, Sept., 1991.

[23] M. N. Do and M. Vetterli, “The contourlet transform: An efficient directional multiresolution image representation,” *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2091–2106, 2005.

[24] Dr. Manjusha Deshmukh & Udhav Bhosale,’’ *Image fusion and image quality assessment of fused images*’’, Saraswati College of Engg. Navi Mumbai, India, International Journal of Image Processing (IJIP), Volume (4): Issue (5),pp. 484-508.

[25] Dr. Manjusha P. Deshmukh & Udhav Bhosle ,” *A Survey of Image Registration* “, International Journal of Image Processing (IJIP), Volume (5) : Issue (3) , 2011.

[26] Dr. Manjusha Deshmukh and Udhav Bhosle, " *Image Registration in Time and Frequency Domain*", In Proceedings of International Conference on Machine Vision, Image Processing and Pattern Analysis(MVIPPA 2008), World Academy of Science, Engineering and Technology, Bangkok, Thailand, Dec 2008, pages 121-127.

[27] Dr. Manjusha Deshmukh and Udhav Bhosle, "*Image Registration for Image Fusion*" In Proceedings International Conference on Emerging Technologies and Applications in Engineering, Technology (ICETAETS 2008), Rajkot, Gujarat, January 2008, pages 1856-1860.

[28] Dr. Manjusha Deshmukh and Udhav Bhosle, " *Mutual Information Based Image Registration* ", In Proceedings of International Conference on Signal Processing, World Academy of Science and Technology, Bangkok, December 2009 pages 113-118.17

INSTRUCTIONS TO CONTRIBUTORS

The *International Journal of Image Processing (IJIP)* aims to be an effective forum for interchange of high quality theoretical and applied research in the Image Processing domain from basic research to application development. It emphasizes on efficient and effective image technologies, and provides a central forum for a deeper understanding in the discipline by encouraging the quantitative comparison and performance evaluation of the emerging components of image processing.

We welcome scientists, researchers, engineers and vendors from different disciplines to exchange ideas, identify problems, investigate relevant issues, share common interests, explore new approaches, and initiate possible collaborative research and system development.

To build its International reputation, we are disseminating the publication information through Google Books, Google Scholar, Directory of Open Access Journals (DOAJ), Open J Gate, ScientificCommons, Docstoc and many more. Our International Editors are working on establishing ISI listing and a good impact factor for IJIP.

The initial efforts helped to shape the editorial policy and to sharpen the focus of the journal. Starting with Volume 9, 2015, IJIP will be appearing with more focused issues. Besides normal publications, IJIP intends to organize special issues on more focused topics. Each special issue will have a designated editor (editors) – either member of the editorial board or another recognized specialist in the respective field.

We are open to contributions, proposals for any topic as well as for editors and reviewers. We understand that it is through the effort of volunteers that CSC Journals continues to grow and flourish.

LIST OF TOPICS

The realm of International Journal of Image Processing (IJIP) extends, but not limited, to the following:

- Architecture of imaging and vision systems
- Character and handwritten text recognition
- Chemistry of photosensitive materials
- Coding and transmission
- Color imaging
- Data fusion from multiple sensor inputs
- Document image understanding
- Holography
- Image capturing, databases
- Image processing applications
- Image representation, sensing
- Implementation and architectures
- Materials for electro-photography
- New visual services over ATM/packet network
- Object modeling and knowledge acquisition
- Photographic emulsions
- Autonomous vehicles
- Chemical and spectral sensitization
- Coating technologies
- Cognitive aspects of image understanding
- Communication of visual data
- Display and printing
- Generation and display
- Image analysis and interpretation
- Image generation, manipulation, permanence
- Image processing: coding analysis and recognition
- Imaging systems and image scanning
- Latent image
- Network architecture for real-time video transport
- Non-impact printing technologies
- Photoconductors
- Photopolymers

- Prepress and printing technologies
- Remote image sensing
- Storage and transmission
- Protocols for packet video
- Retrieval and multimedia
- Video coding algorithms and technologies for ATM/p

CALL FOR PAPERS

Volume: 9 - **Issue: 1**

i. Submission Deadline : November 30, 2014 **ii. Author Notification:** December 31, 2014

iii. Issue Publication: January 2015

CONTACT INFORMATION

Computer Science Journals Sdn BhD

B-5-8 Plaza Mont Kiara, Mont Kiara

50480, Kuala Lumpur, MALAYSIA

Phone: 006 03 6204 5627

Fax: 006 03 6204 5628

Email: cscpress@cscjournals.org

CSC PUBLISHERS © 2014
COMPUTER SCIENCE JOURNALS SDN BHD
B-5-8 PLAZA MONT KIARA
MONT KIARA
50480, KUALA LUMPUR
MALAYSIA

PHONE: 006 03 6204 5627
FAX: 006 03 6204 5628
EMAIL: cscpress@cscjournals.org