

International Journal of Image Processing (IJIP)

ISSN : 1985-2304

VOLUME 4, ISSUE 3

PUBLICATION FREQUENCY: 6 ISSUES PER YEAR

International Journal of Image Processing (IJIP)

Volume 4, Issue 3, 2010

Edited By
Computer Science Journals
www.cscjournals.org

Editor in Chief Professor Hu, Yu-Chen

International Journal of Image Processing (IJIP)

Book: 2010 Volume 4 Issue 3

Publishing Date: 31-07-2010

Proceedings

ISSN (Online): 1985-2304

This work is subjected to copyright. All rights are reserved whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication of parts thereof is permitted only under the provision of the copyright law 1965, in its current version, and permission of use must always be obtained from CSC Publishers. Violations are liable to prosecution under the copyright law.

IJIP Journal is a part of CSC Publishers

<http://www.cscjournals.org>

©IJIP Journal

Published in Malaysia

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

CSC Publishers

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 four 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.

IJIP give 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

Editor-in-Chief (EiC)

Professor Hu, Yu-Chen
Providence University (Taiwan)

Associate Editors (AEiCs)

Professor. Khan M. Iftekharuddin
University of Memphis ()

Dr. Jane(Jia) You
The Hong Kong Polytechnic University (China)

Professor. Davide La Torre
University of Milan (Italy)

Professor. Ryszard S. Choras
University of Technology & Life Sciences ()

Dr. Huiyu Zhou
Queen's University Belfast (United Kindom)

Editorial Board Members (EBMs)

Professor. Herb Kunze
University of Guelph (Canada)

Assistant Professor. Yufang Tracy Bao
Fayetteville State University ()

Dr. C. Saravanan
(India)

Dr. Ghassan Adnan Hamid Al-Kindi
Sohar University (Oman)

Dr. Cho Siu Yeung David
Nanyang Technological University (Singapore)

Dr. E. Sreenivasa Reddy
(India)

Dr. Khalid Mohamed Hosny
Zagazig University (Egypt)

Dr. Gerald Schaefer
(United Kingdom)

Dr. Chin-Feng Lee
Chaoyang University of Technology (Taiwan)

Associate Professor. Wang, Xiao-Nian
Tong Ji University (China)

Professor. Yongping Zhang
Ningbo University of Technology (China)

Table of Contents

Volume 4, Issue 3, July 2010.

Pages

- 192 – 204 Semantic Gap in CBIR: Automatic Objects Spatial relationships
Semantic Extraction and Representation
Hui Hui Wang, Dzulkifli Mohamad, N. A. Ismail
- 205 - 217 Performance Comparison of Density Distribution and Sector mean
of sal and cal functions in Walsh Transform Sectors as Feature
Vectors for Image Retrieval
H. B. Kekre, Dharendra Mishra
- 218 - 231 A Fuzzy Watermarking Approach Based on Human Visual System
Oueslati Sameh, Adnane Cherif, Basel Solaiman
- 232 - 239 Invariant Recognition of Rectangular Biscuits with Fuzzy Moment
Descriptors, Flawed Pieces Detection
**Pulivarthi Srinivasa Rao, Sheli Sinha Chaudhuri, Romesh
Laishram**
- 240 - 250 Genetic Algorithm Processor for Image Noise Filtering Using
Evolvable Hardware
**A. Guruva Reddy, K. Sri Rama Krishna, M.N.GIRI PRASAD,
K.Chandrabushan Rao, M. Madhavi**

- 251 - 262 Developing 3D Viewing Model from 2D Stereo Pair with its
Occlusion Ratio
Himanshu Johari, Vishal Kaushik, pawan kumar upadhyay
- 263 - 286 Optical Approach Based Omnidirectional Thermal Visualization
Wai Kit Wong, Chu Kiong Loo , Way Soong Lim

Semantic Gap in CBIR: Automatic Objects Spatial Relationships Semantic Extraction and Representation

Hui Hui Wang

*Department of Computing and Software Engineering,
Faculty of Computer Science and Information Technology,
Universiti Malaysia Sarawak,
94300 Kota Samarahan, Sarawak, Malaysia*

hhwang@fit.unimas.my

Dzulkifli Mohamad

*Department of Computer Graphics and Multimedia,
Faculty of Computer Science and Information Technology,
Universiti Teknologi Malaysia,
81300 Skudai, Johor Bahru, Malaysia*

dzulkifli@utm.my

N. A. Ismail

*Department of Computer Graphics and Multimedia,
Faculty of Computer Science and Information Technology,
Universiti Teknologi Malaysia,
81300 Skudai, Johor Bahru, Malaysia*

azman@utm.my

Abstracts

The explosive growth of image data leads to the need of research and development of Image retrieval. Image retrieval researches are moving from keyword, to low level features and to semantic features. Drive towards semantic features is due to the problem of the keywords which can be very subjective and time consuming while low level features cannot always describe high level concepts in the users' mind. This paper is proposed a novel technique for objects spatial relationships semantics extraction and representation among objects exists in images. All objects are identified based on low level features extraction integrated with proposed line detection techniques. Objects are represented using a Minimum Bound Region (MBR) with a reference coordinate. The reference coordinate is used to compute the spatial relation among objects. There are 8 spatial relationship concepts are determined: "Front", "Back", "Right", "Left", "Right-Front", "Left-Front", "Right-Back", "Left-Back" concept. The user query in text form is automatically translated to semantic meaning and representation. Besides, the image similarity of objects spatial relationships semantic has been proposed.

Keywords : Semantic Gap, Objects Spatial Relationships Semantic, Automatic Image Semantic Extraction, Image Retrieval

1. INTRODUCTION

Image retrieval is the field of study concerned with searching and browsing digital images from database collection. This area of research is very active research since the 1970s [1, 2]. Due to more and more images have been generated in digital form around the world, image retrieval attracts interest among researchers in the fields of image processing, multimedia, digital libraries,

remote sensing, astronomy, database applications and other related area. Effective and fast retrieval of digital images has not always been easy, especially when the collections grow into thousands. An effective image retrieval system needs to operate on the collection of images to retrieve the relevant images based on the query image which conforms as closely as possible to human perception.

1.1. Evolution of Image Retrieval

The purpose of an image database is to store and retrieve an image or image sequences that are relevant to a query. There are a variety of domains such as information retrieval, computer graphics, database management and user behavior which have evolved separately but are interrelated and provide a valuable contribution to this research subject. As more and more visual information is available in digital archives, the need for effective image retrieval has become clear [3,4]. In image retrieval research, researchers are moving from keyword based, to content based then towards semantic based image retrieval and the main problem encountered in the content-based image retrieval research is the semantic gap between the low-level feature representing and high-level semantics in the images.

1.1.1. Keyword Based Image Retrieval

In 1970s, the conventional image retrieval system used keyword as descriptors to index an image however the content of an image is much richer than what any set of keywords can express.

Text-based image retrieval techniques employ text to describe the content of the image which often causes ambiguity and inadequacy in performing an image database search and query processing. This problem is due to the difficulty in specifying exact terms and phrases in describing the content of images as the content of an image is much richer than what any set of keywords can express. Since the textual annotations are based on language, variations in annotation will pose challenges to image retrieval.

1.1.2. Content Based Image Retrieval

In 1990s, Content-based image retrieval (CBIR) then has been used as an alternative to text based image retrieval. Unlike keywords-based system, visual features for contents-based system are extracted from the image itself. CBIR can be categorized based on the type of features used for retrieval which could be either low level or high level features. At early years, low level features include colour, texture, shape and spatial relations were used. The summary of CBIR researches done in retrieving the image based on their visual content can be found in our paper, ref [5]

Although there are many sophisticated algorithms to describe color, shape and texture features approaches, these algorithms do not satisfied and comfort to human perception This is mainly due to the unavailability of low level image features in describing high level concepts in the users' mind. For an example finding an image of a little boy is playing a ball in the garden. The only way a machine is able to perform automatic extraction is by extracting the low level features that represented by the color, texture, shape and spatial from images with a good degree of efficiency.

1.1.3. Semantic Based Image retrieval

In 2000s, semantic based image retrieval has been introduced. This is due to neither a single features nor a combination of multiple visual features could fully capture high level concept of images. Besides, the performance of image retrieval system based on low level features are not satisfactory, there is a need for the mainstream of the research converges to retrieve based on semantic meaning by trying to extract the cognitive concept of a human to map the low level image features to high level concept (semantic gap). In addition, representing the image content with semantic terms allows users to access images through text query which is more intuitive, easier and preferred by the front end users to express their mind compare with using images. The review and general framework of semantic based image retrieval can be found our paper in ref [6]

1.2. Semantic Gap

Bridging the semantic gap for image retrieval is a very challenging problem yet to be solved [7,8]. Describing images in semantic terms is an important and challenging task that needed to carry out to fulfill human satisfaction besides to have more intelligent image retrieval system.

Human beings are able to interpret images at different levels, both in low level features (colour, shape, texture and object detection) and high level semantics (abstract objects, an event). However, a machine is only able to interpret images based on low level image features. Besides, users prefer to articulate high-level queries [9,10], but CBIR systems index images using low-level features. Hence, introducing an interpretation inconsistency between image descriptors and high-level semantics that is known as the semantic gap [3,10]. The semantic gap is the lack of correlation between the semantic categories that a user requires and the low-level features that CBIR systems offer. The semantic gap between the low-level visual features (color, shape, texture, etc.) and semantic concepts identified by the user remains a major problem in content based image retrieval [8].

Semantic content representation has been identified as an important issue to bridge the semantic gap in visual information access. It has been addressed as a good description and representation of an image, it able to capture meaningful contents of the image. Current researches often represent images in terms of labeled regions or images, but pay little attention to the spatial positions or relationships between those regions or objects [11]. Spatial relationship is needed in order to further increase the confidence in image understanding. Besides, users preferred to express their information needs at the semantic level instead of the level of preliminary image features. Moreover textual queries usually provide more accurate description of users' information needs.

The attempt to overcome the gap between high level semantic and low level features by representing images at the object level is needed [7] as well as the spatial relationship of objects to further increase the image understanding

2. RELATED WORKS

In general, there is no direct link between such high-level semantic concepts and the automatically extracted, low-level image features. Therefore, to support query by semantic concept, there is a compelling need for CBIR systems to provide maximum support towards bridging the 'semantic gap' between the low-level visual features [3,12] and it is a very challenging task to extract and manage meaningful semantics and to make use of them to achieve more intelligent and user friendly retrieval [13].

2.1. Manual Image Semantic Extraction

One conventional and common ways to describe the image in high level is using the manual annotation. Manual annotation needs to annotate every image by human where users enter some descriptive keywords when the images are loaded/registered/browsed. Existing applications that support manual image annotation include Wikipedia image collection [14], lonely planet [15], photoblog [16], fotopages [17], flickr and etc. They allow human to annotate images with some keywords. It is based on whole images and cannot annotated based on the objects or regions of the images. Inotes [18] and facebook [19] are most popular manual image annotation approaches where user can annotate various objects or regions based on selected regions in an image instead of just annotate whole images.

Although manual annotation of image content is considered a "best case" in terms of accuracy, since keywords are selected based on human determination of the semantic content of images, as well as able to support user queries in text. However it is a labor intensive and tedious process. In addition, manual annotation may also introduce retrieval errors due to users forgetting what descriptors they used when annotating their images after a lengthy period of time.

Researchers have explored techniques for improving the process of manual annotation. So, researchers are moving toward automatically automatic extraction of the image semantic content.

2.2. Semi/Automatic Image Semantic Extraction

Reference [20] was the one of the early proposed automatic solutions, where a probabilistic framework based on estimating class likelihoods of local areas, labeled as either man made or natural, inside or outside objects scenes. Zhao and Grosky [13] proposed a method integrating the LSI, normalization and term weighting to obtain the meaningful features mapped to semantic landscapes.

All the methods discussed are only able to retrieve similar images which have the whole semantics and does not indicate which part of the image gives rise to which words, so it is not explicitly object recognition. They are lacking of the ability to find the object semantics in images

Various methods have been proposed to automatically capture region semantic of images instead of image semantic only. Reference [21] introduced the region-based co-occurrence model to assign the word to the region. Reference [22] was proposed a model of object recognition as machine translation. In this model, the mapping between regions and keywords is learnt using a method based on the EM algorithm. Reference [23] implement a cross-media relevance model and identify 70 object concepts. The model in learns the joint probability of associating words to image features from training sets and uses it to generate the probability of associating a word to a given query image. Reference [24] then improved the reference [38] approach by using continuous probability density functions to estimate the probability of observing a region given an image, it is proven that it can improve on the results obtained in [23]. In [25], the authors propose a framework to link signal-based features to semantic data by mapping the multidimensional space defined by the extracted low-level features to predefined semantic classes through a Bayesian network.

Even though the object/region semantics can be captured but then the extraction of spatial relational semantic descriptors is often neglected. They do not take into account the relational spatial semantics among objects in the images which affects the quality of the retrieval results. In other words, they only able to recognize images which contain a cat and a dog. However it fails to tell the actual relative direction (spatial relationship) between the dog and the cat. Representation of spatial relations semantics among objects are important as it can convey important information about the image and to further increase the confidence in image understanding contribute to richer querying and retrieval facilities.

2.3. Objects Spatial Relationships Semantic Extraction

Some methods has been introduced to capture the spatial relationship semantic. Some researches [11, 26] use ontology method to capture the relation semantics. Reference [26] adding the spatial semantic in image annotation by adopting the spatial concepts from the Suggested Upper Merged Ontology (SUMO). This approach is having high dependency on the semi-annotation process. Reference [26] extending it in both the granularity of the absolute positions, the extraction of combined relations (like above and to the left of) and through the use of object properties in the ontology to infer more complex spatial relations. Reference [42] proposed a unified multi-faceted framework unifying visual semantics and relational spatial characterization for automatic image retrieval that enforces expressivity and computational efficiency through fast lattice processing. The spatial relation of objects in image is represented using conceptual structures. The image is index using conceptual graph. The combined relation also can be captured using the mentioned method.

Even thought simple relation and combined relation has been captured. However there are still having some false objects spatial relationships extraction concept (Example in Figure 1, the object B and C suppose to have Front/Back spatial relationship however ref [26, 27] extracted it

as Left/right relation concept and ref [11] extracted it as above-right relation).The combined relation is limited to 2 objects only for query images such as Object A is Left to and in Front object B and also lack of abstract query such as traffic jam situation. The multiple objects with combined relation should be considered such as Object A is Left to Object B and Right-Front Object C. And also, there is none of spatial relation semantic similarity for the spatial semantics description. Besides, It should be fully automatic image and spatial relation semantic extraction without involving any user or relevance feedback during the retrieval process.

3. THE RESEARCH FRAMEWORK

The research framework consists of four main components, which include low level features extraction, object identification and object semantic extraction and representation.

- 1) The Low Level Features Extraction Component – the low level features of images are extracted and grouped based on their common characteristics to get the regions/segment of images.
- 2) Object Identification Component- Regions/segment of images are integrated with line detection technique to extract the object of interest in images.
- 3) Object Semantic Extraction Component - The object identified is indicated using Minimum Bound Region (MBR) and the properties of objects are extracted (eg. Color of car). Each object is represented by a reference coordinate.
- 4) Objects Spatial Relationship Semantic Extraction Component -This component automatically extracts and identifies spatial information. It captured the spatial relationship among objects in the images.

In this paper, the discussion and concentration is mainly on the objects and their spatial relationship semantics extraction and representation in the image. Traffic images are used as the domain of study.

3.1. Objects Spatial Relationships Semantic extraction

The objects spatial relationships semantic extraction approach has 6 main stages

3.1.1 Determine the Road Slope as Z-axis

The road slope is needed to determine and it is used as the reference slope for getting spatial relationship for all car objects in image. It is actually act as a Z-axis due to the image view is slanted.

$$\text{Road Slope, } m(R) = \frac{y_2 - y_1}{x_2 - x_1}$$

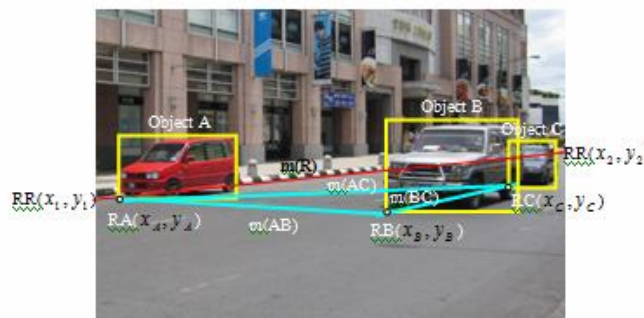


FIGURE 1 : Traffic images

3.1.2 Calculate the Slope of Each Possible Pair of Objects in Image

Each of the objects is represented using Minimum Bound Region (MBR) that indicated using a box as show above. Each object has a reference coordinate (indicated by a small circle, left bottom of MBR)

From Figure 1, Image consists of multiple car objects, $I = \{\text{Object A, Object B, Object C}\}$
The reference coordinate of objects as below

$$\text{Object A} = \text{RA}(x_A, y_A)$$

$$\text{Object B} = \text{RB}(x_B, y_B)$$

$$\text{Object C} = \text{RC}(x_C, y_C)$$

The slope of each possible pair of objects is calculated based on their object's reference coordinate as below.

$$\text{Object A and Object B} = m(AB) = \frac{y_B - y_A}{x_B - x_A}$$

$$\text{Object A and Object C} = m(AC) = \frac{y_C - y_A}{x_C - x_A}$$

$$\text{Object B and Object C} = m(BC) = \frac{y_C - y_B}{x_C - x_B}$$

The spatial relationship of two objects is defined by computing and comparing the slope of two relative objects.

3.1.3. Determine the Relational/Directional Relationship

The basis of interpreting positions in reality is assumed as follows: The positions of left and right when viewing an image is inversed from the positions in real life. This means that when interpreting the image in Figure. 1, the red car is on the left in the image but in actual fact, it is on the right.

There are four basic relative relations of "Left to", "Right to", "Front to" and "Back to" as well as the composite spatial relations of "Right-Front", "Left-Front", "Right-Back", "Left-Back" will be determined and discussed. Those are the most important relationship in the domain of traffic images.

a) Front / Back Relationship

The positive and negative value of slope between pair of objects will be used as indicator for Front/Back relation concept to ensure object is front or back of another object.

The rules for inferring Front/Back relations are defined and illustrated in Figure 2:

$$m(AB) > 0, \text{ A is Front to B}$$

$$m(AB) < 0, \text{ A is Back to B}$$

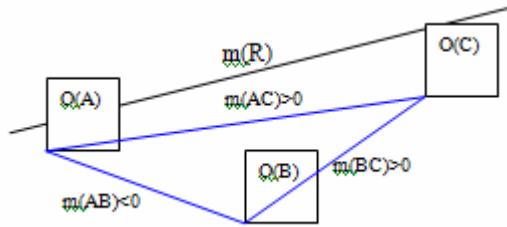


FIGURE 2 : Front/Back relationship determination

b) Absolute Front/Back and Right/Left Relationship

Two objects are considered absolutely Front/Back relation when their slope value always infinity and absolutely Right/Left relation when their slope value is 0.

The rules for inferring absolutely Front/Back and Right/Left relations are defined and illustrated in figure 3 :

$m(AB) = \infty$, A is Front/Back to B

$m(AB) = 0$, A is Right/Left to B

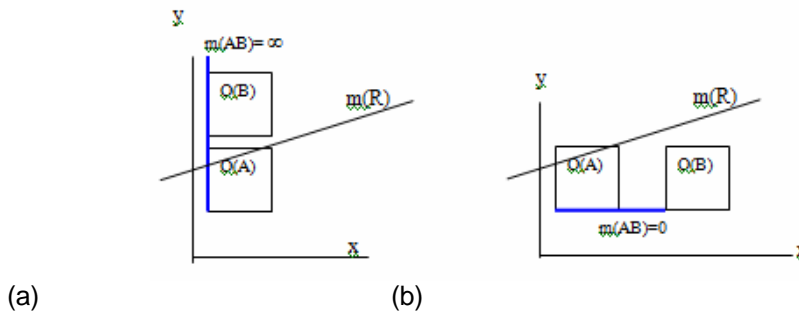


FIGURE 3 : Absolute Front/Back relationship(a) and Absolute Right/Left relationship (b)

From the Figure 3, it shows that 2 objects only have 1 relation. It's either Front/Back relation for object A and B in Figure 3(a) or Right/Left relation for object A and B in Figure 3(b). However, there are some composite relation exists between objects such as Left-Front, Left-Back, Right-Front and Right-Back.

c) Composite Relationship

In traffic image, the road slope, $m(R)$ value is used as reference line (axis-z) instead of axis-x and axis-y. Given the figure below,

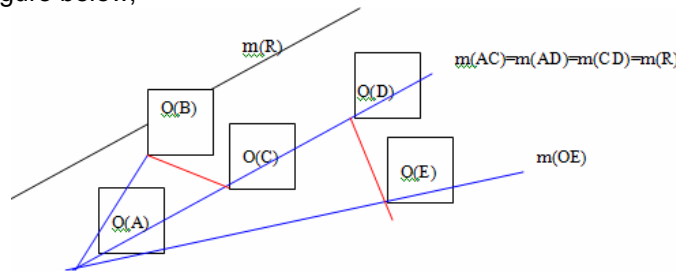


FIGURE 4 : Composite Relation

If the slope of 2 objects is same with slope of road, it means there is only Front/Back relation. The rules for inferring Absolute Front/back relationship are defined as below

$m(R) = m(O_i, O_j | i \neq j)$, Object I is absolute Front to Object j

If the slope of 2 objects is same with 0, its mean there is only Right/Left relation.

$m(O_i, O_j | i \neq j) = 0$, Object i is absolute Right to Object j

The composite relation is observed when the slope of 2 objects has value either greater or smaller than slope of road,

$m(O_i, O_j | i \neq j) < m(R)$, Object i is Right-Front to Object j, $m > 0$

$m(O_i, O_j | i \neq j) < m(R)$, Object i is Right-Back to Object j, $m < 0$

$m(O_i, O_j | i \neq j) > m(R)$, Object i is Left-Front to Object j, $m > 0$

3.1.4. Distance of Spatial Relationship Semantic

The distance of the spatial relationship among 2 objects are calculated by,

Spatial relationship Distance, d , $d(O_i, O_j | i \neq j) = |m(O_i, O_j | i \neq j) - m(R)|$

where $m(O_i, O_j | i \neq j)$ is the slope of object O_i and object O_j , $m(R)$ is the slope of the road.

3.1.5. Spatial Relationship Semantics Representation

Spatial relation semantic concept of image can be represented as, Sp

$Sp = \{m(R), [O: (P_{ij}(O_i, O_j), m_{ij}(O_i, O_j), d_{ij}(O_i, O_j))] | i \neq j, \forall i, j \in O, O \in I\}$

where $m(R)$ is the slope of road, $P_{ij}(O_i, O_j)$ is pair of objects O_i and O_j , $m(O_i, O_j)$ is the slope between object O_i and O_j and $d(O_i, O_j)$ is the distance of the spatial relation between object O_i and O_j , O is total number of object in image I .

4. USER QUERY

The user query is used to express the user's information need to retrieve images in collection of database that conform to human perception. According to Ref [28], to define a semantic meaning and representation of the input query that can precisely understand and distinguish the intent of the input query are the major challenges. It is difficult and often requires many human efforts to meet all these challenges by the statistical machine learning approaches.

Querying by visual example is a paradigm, particularly suited to express perceptual aspects of low/intermediate features of visual content [29]. Visual content refer to color, shape and texture features of images. Although promising progresses have been made in image retrieval techniques based on visual features, formulating a query such as submitting an example image or a sketch is sometimes not convenient for users.

Text-based queries are the most popular query method. User usually prefers using keywords to indicate what they want. [9,30]. Textual queries usually provide more accurate description of users' information needs as it allow users to express their information needs at the semantic level and high level abstractions instead of limited to the level of preliminary image features. However, the textual words need to be translated automatically to semantic meaning and representation that are matched in the images semantic representation in database in order to have fully and precisely understand the user input.

4.1 Semantic Extraction and Representation for User Query

The user query in text forms that mainly focus on Object Spatial Relationship is automatically translate to semantics meaning and representation. The object spatial relationship semantic translation from user query has 3 main stages

4.1.1. Determine Rules for Objects Spatial Relationship

Right/Left, Rule 1: $(P_{ij}(O_i, O_j), m_{ij}(O_i, O_j) = 0, d_{ij}(O_i, O_j) < 0)$, i Right j, j Left i

Front/Back Rule 2: $(P_{ij}(O_i, O_j), m_{ij}(O_i, O_j) = \infty, d_{ij}(O_i, O_j) > 0)$, i Front j, j Back i, j left-Back i

Left-Front Rule 3: $(P_{ij}(O_i, O_j), m_{ij}(O_i, O_j) > 0, d_{ij}(O_i, O_j) > 0)$, i Left-Front j

Right-Front Rule 4: $(P_{ij}(O_i, O_j), m_{ij}(O_i, O_j) > 0, d_{ij}(O_i, O_j) < 0)$, i Right-Front j

Right-Back Rule 5: $(P_{ij}(O_i, O_j), m_{ij}(O_i, O_j) < 0, d_{ij}(O_i, O_j) < 0)$, i Right-Back j

Logical operation L: AND, OR, NOT

4.1.2. Sub Divided User Query to Sub User Query (if there is any).

Given example user query:

Object A Right Object B AND Right-Front Object C

$Q = (Q C_1 \dots Q C_k)$,

$Q C_1$: Sub Query 1 : Object A Right Object B

$Q C_2$: Sub Query 2 : Object A Right-Front Object C

4.1.3. Assign the Sub Query to Rules for Objects Spatial Relationship

The conversion of the above user query to spatial relationship semantics representation rules is as below,

$Q = (Q C_1 \dots Q C_k)$,

Where $Q C_1 \dots Q C_k$ is sub user query and k is the number of sub query for the user query.

$Q C_1 = (P_{AB}(O_A, O_B), m_{AB}(O_A, O_B) = 0, d_{AB}(O_A, O_B) < 0)$ Operator

$Q C_2 = (P_{AC}(O_A, O_C), m_{AC}(O_A, O_C) > 0, d_{AC}(O_A, O_C) < 0)$

There is a operator between sub query, the operator can be logical operation AND, OR or NOT.

5. IMAGE SPATIAL SEMANTICS SIMILARITY MEASUREMENT

Image Spatial semantic similarity is used to define the Spatial similarity between the image in database and user query.

Database Images $D_I = I_{i=0}, I_1, \dots, I_p$

$I_i = (P_o, P_1, \dots, P_n)$

Where database images consists of k number of images, each image i consists of the n number of object pair. The n number is different for the image from i=0 to i=p.

User Query, $Q = (Q C_1, \dots, Q C_k)$,

Where $Q C_1, \dots, Q C_k$ is sub user query and k is the number of sub query for the user query.

The steps of the image semantics similarity is discuss as below

5.1. Query Matching

Query matching is the process to determine total number of matched element in sample image I to the Query Q. It is evaluated in terms of the matching based on m and d characteristic

As an example of Query Q, object A Right Object B is defined as below,

$$Q = (P_{AB} (O_A, O_B), m_{AB} (O_A, O_B = 0), d_{AB} (O_A, O_B) < 0),$$

And image I,

$$I = (P_{ij} (O_i, O_j), m_{ij} (O_i, O_j), d_{ij} (O_i, O_j)) \mid i \neq j, \forall i, j \in I$$

Where I consists of a list of pair object i and Object j

The matching of Q to image I is defined as a total number of matched pairs in I as show below

Match(Q,I) = number of element for $(Q \cap I)$

5.2. Image Similarity

The Image similarity is used to define the similarity between user query and Image. The image similarity based on Single condition of User Query (without logical operation) and Multiple condition of User query (Logical operation involved) are then determined

5.2.1. Single Condition of User Query (without Logical Operation)

In this single condition of user query, there is no logical operation involved in the user query.

The similarity of image I to query Q is defined as the ratio of number of matched pair, over the total number of pairs in image as show below,

$$S_{sim} (Q C_1, I_i) = \frac{\text{number of element } (Q, I)}{n},$$

Where number of element (Q,I) is the number of matched pair in query matching while n is total number of pairs in images

5.2.2. Multiple Condition of User Query (Logical Operation Involved)

For multiple condition of user query, image Semantics Similarity is determined based on the involvement of logical operation AND, OR and NOT.

5.2.2.1. Logical AND Operation

If there is a or more logical AND operation involved in user query, all of their sub query must be combined to get the object relationship semantic similarity, S_{AND}

$$S_{sim(AND)} = \frac{\sum_i^k (S_{sim}(QC_i, I_i) + S_{sim}(QC_j, I_i))}{Total(k)} \quad i \neq j$$

Where $S_{sim}(QC_i)$ is semantic similarity of user sub query i and $S_{sim}(QC_j, I_i)$ is semantic similarity of user sub query j respectively to the image i while k is total number of sub user query.

5.2.2.2. Logical OR Operation

If there is a logical OR operation involved in user query, the maximum value of the object spatial relationship semantics similarity of sub query is chosen as it represents the closest match between database images and user query.

$$S_{sim(OR)} = \text{Max} \sum_i^k (S_{sim}(QC_i, I_i), S_{sim}(QC_j, I_i))$$

Where $S_{sim}(QC_i, I_i)$ is the semantics similarity for sub query i and k is total number of sub query.

5.2.2.3 Logical NOT Operation

For logical NOT operation, the list of images are the images that are not listed in the candidate objects in Candidate object filters.

$$O_{NOT} = \overline{O_{QC_i}}$$

where $\overline{O_{QC_i}}$ is list of objects that do not found in the O_{QC_i}

5.3. Range of Image Spatial Semantic Similarity

The degree of image spatial similarity ranges from 0 to 1. The value 1 indicates that there is a perfect match of the database images with user query while value 0 indicates that there are dissimilar.

6. CONCLUSION

This paper provides a study of image retrieval work towards narrowing down the 'semantic gap'. Recent works are mostly lack of semantic features extraction especially in objects spatial relationships semantic and user behavior consideration. Therefore, a new method and approach for extracting objects spatial relationships semantic and representation automatically from images has been proposed in order to bridge the semantic gap besides enhance the high level semantic

search and retrieval. There are 8 of main important spatial relationship concept has been introduced. Besides, the objects spatial relationship semantic similarity is also introduced. This work will be enhanced and expanded further to include the object characteristics as well as more abstract high level queries with spatial relationship semantic representation. There is a need of image retrieval system that is capable to interpret the user query and automatically extract the semantic feature that can make the retrieval more efficient and accurate to bridge the semantic gap problem in image retrieval.

7. REFERENCES

1. A. Rosenfeld. "Picture processing by computer". ACM Computing Surveys, 1(3):147-176, 1969.
2. H. Tamura and S. Mori. "A data management system for manipulating large images". In Proceedings of Workshop on Picture Data Description and Management, pages 45-54, Chicago, Illinois, USA, April 1977
3. Ying Liu, Dengsheng Zhang, Guojun Lu, Wei-Ying M. "A survey of content-based image retrieval with high-level semantics". Pattern Recognition, 40(1):262 – 282, 2007
4. R.Datta, D.Joshi, J.Li, J.Z.Wang. "Image Retrieval: Ideas, Influences, and Trends of the New Age". ACM Transactions on Computing Surveys, 40(2), 2008
5. Hui Hui Wang, Dzulkifli Mohamad, N.A.Ismail. "Image Retrieval: Techniques, challenge and Trend". In International conference on Machine Vision, Image processing and Pattern Analysis, Bangkok, 2009
6. Hui Hui Wang, Dzulkifli Mohamad, N.A.Ismail. "Towards Semantic Based Image Retrieval : A Review". In The 2nd International Conference on Digital Image Processing (ICDIP 2010), Singapore, 2010. Proceedings of SPIE, Vol 7546, 2010
7. C.Y.Chang, H.Y.Wang, C.F.Li. "Semantic analysis of real world images using support vector machine". Expert systems with application : An international journal, 36(7):10560-10569, 2009
8. Najlae Idrissi. "Bridging the Semantic Gap for Texture-based Image Retrieval and Navigation", Journal Of Multimedia, 4(5):277-283, October 2009
9. M. L. Kherfi, D. Ziou, and A. Benard. "Image retrieval from the World Wide Web: Issues, techniques and systems". ACM Computing Surveys, 36(1):35-67, 2004
10. A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. "Content-based image retrieval at the end of the early years". In IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(12):1349-1380, December 2000.
11. Zurina Muda, Paul H. Lewis, Terry R. Payne, Mark J. Weal. "Enhanced Image Annotations Based on Spatial Information Extraction and Ontologies". In Proceedings of the IEEE International Conference on Signal and Image Processing Application, ICSIPA 2009
12. Haiying Guan, Sameer Antani, L. Rodney Long, and George R. Thomas. "Bridging the semantic gap using making SVM for image retrieval". In Proceedings of the sixth IEEE international Conference on Symposium on Biomedical Imaging: From Nano to Macro, USA 2009.
13. Zhao and Grosky "Bridging the semantic gap in Image retrieval". Idea Group Publishing Series, Book Chapter 2, pp 14-36 (2002)
14. Wikipedia, <http://wikipedia.org>. Accessed at June 2010
15. Travel Photography and Professional Stock Photography-Lonely Planet Images, <http://www.lonelyplanetimages.com>. Accessed at June 2010
16. Photoblog-Your Life in Photos, <http://www.photoblog.com>. Accessed at June 2010
17. Photo Blogging made Easy-fotopages, <http://www.fotopages.com>. Accessed at June 2010
18. Inote : Image Annotation in Java, <http://www.iath.virginia.edu/inote/>. Accessed at June 2010
19. Facebook, <http://www.facebook.com>. Accessed at June 2010
20. Bradshaw B. "Semantic based image retrieval: A probabilistic approach". In Proceedings of the Eight ACM International Conference on Multimedia, United States, 2000

21. Y. Mori, H. Takahashi, and R. Oka. "Image-to-word transformation based on dividing and vector quantizing images with words". In Proceedings of the First International Workshop on Multimedia Intelligent Storage and Retrieval Management, 1999
22. P Duygulu, K Barnard, N de Fretias, and D Forsyth. "Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary". In Proceedings of the European Conference on Computer Vision, pages 97–112, 2002
23. Jeon,J.,Laverenko,V.,Manmatha, R " Automatic image annotation and retrieval using cross-media relevance models". In:Proceedings of international ACM Conference on Research and Development in Information Retrieval, 2003
24. Lavrenko,V., Manmatha,R.,Jeon,J:"A model for learning the semantic of pictures". In proceeding of Advances in Neural Information Processing Systems, NIPS, 2003
25. Mittal A, Cheong LF. "Framework for synthesizing semantic-level indexes". Multimedia Tools and Applications 20(2):135–158.2003
26. L. Hollink, G. Nguyen, G. Schreiber, J. Wielemaker, and B. Wielinga, "Adding spatial semantics to image annotations," In 4th International Workshop on Knowledge Markup and Semantic Annotation, 2004
27. Belkhatir M. "An operational model based on knowledge representation for querying the image content with concepts and relations". Multimed Tools Application, 43:1–23. Springerlink Netherlands, 2009
28. Jian Hu, Gang Wang, Fred Lochovsky, Jian-Tao Sun, Zheng Chen. "Understanding User's Query Intent with Wikipedia". Proceedings of the 18th international conference on World wide web. ACM. 2009
29. E. Chang, K. Goh, G. Sychay, G. Wu, CBSA: CBSA: "Content-based Soft Annotation for Multimodal Image Retrieval using Bayes Point Machines", In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 26-38, 2003.
30. Hong Fu. "Attention driven image interpretation, annotation and retrieval". PhD thesis. The Hong Kong Polytechnic University. 2006

Performance Comparison of Density Distribution and Sector mean of sal and cal functions in Walsh Transform Sectors as Feature Vectors for Image Retrieval

H.B.Kekre

*Senior Professor, Computer Engineering Department
SVKM's NMIMS (Deemed-to-be University)
Mumbai, 400056, INDIA*

hbkekre@yahoo.com

Dhirendra Mishra

*Assistant Professor and PhD. Research Scholar
SVKM's NMIMS (Deemed-to-be University)
Mumbai, 400056,INDIA*

dhirendra.mishra@gmail.com

Abstract

In this paper we have proposed two different approaches for feature vector generation with absolute difference as similarity measuring parameter. Sal-cal vectors density distribution and Individual sector mean of complex Walsh transform. The cross over point performance of overall average of precision and recall for both approaches on all applicable sectors sizes are compared. The complex Walsh transform is conceived by multiplying sal components by $j = \sqrt{-1}$. The density distribution of real (cal) and imaginary (sal) values and individual mean of Walsh sectors in all three color planes are considered to design the feature vector. The algorithm proposed here is worked over database of 270 images spread over 11 different classes. Overall Average precision and recall is calculated for the performance evaluation and comparison of 4, 8, 12 & 16 Walsh sectors. The overall average of cross over points of precision and recall is of all methods for both approaches are compared. The use of Absolute difference as similarity measure always gives lesser computational complexity and Individual sector mean approach of feature vector has the best retrieval.

Keywords: CBIR, Walsh Transform, Euclidian Distance, Precision, Recall, Kekre's Algorithm

1. INTRODUCTION

The explosive growth in the digital information in the form of videos and images has given the birth to issue of its storage and management [1]. Digital information in the form of images are very widely used on web and in various applications like medical, biometric, security etc. Management of such large image databases requires search and retrieval of relevant images to use it in time. The earlier methods of search and retrieval of relevant images from the database were based on texts which used to be ambiguous and involved lot of human intervention. The need of better method of search and retrieval of images has given the birth to the content based Image retrieval approach. Here the image itself is used as a query and images having similar contents are retrieved from the large database. The earliest use of the term content-based image retrieval in

the literature seems to have been by Kato [3] to describe his experiments into automatic retrieval of images from a database by using color, texture and shape as features. The term has since been widely used to describe the process of retrieving desired images from a large collection on the basis of features (such as colors, texture and shape) that can be automatically extracted from the images themselves. The typical CBIR system [2-6] performs two major tasks. The first one is feature extraction (FE), where a set of features, called image signature or feature vector, is generated to closely represent the content of each image in the database. A feature vector is much smaller in size than the original image, typically of the order of hundreds of elements (rather than millions). The second task is similarity measurement (SM), where a distance between the query image and each image in the database using their signatures is computed so that the top—closest images can be retrieved.[8-10]. Various methods of CBIR like Bit truncation coding (BTC)[14], use of Transforms[15],[16],[18],[20],[21], Vector quantization [12],[19] has already been proposed for feature vector generation. The basic features of images like color [12][13], shape, texture are taken into consideration. Here in this paper we have proposed the innovative idea of sectorization of complex Walsh transform and use of density distribution of sal and cal elements in sectors of complex Walsh transform and individual mean of sal and cal distributions in each sector as two different methods to generate the feature vector. We have also proposed the use of absolute difference which is faster as similarity measuring parameter for CBIR instead of Euclidian distance which requires large computations .

2. Walsh Transform

These Walsh transform [16] matrix is defined as a set of N rows, denoted W_j , for $j = 0, 1, \dots, N - 1$, which have the following properties:

- W_j takes on the values $+1$ and -1 .
- $W_j[0] = 1$ for all j .
- $W_j \times W_k^T = 0$, for $j \neq k$ and $W_j \times W_j^T = N$, for $j=k$.
- W_j has exactly j zero crossings, for $j = 0, 1, \dots, N-1$.
- Each row W_j is either even or odd with respect to its midpoint.

Walsh transform matrix is defined using a Hadamard matrix of order N . The Walsh transform matrix row is the row of the Hadamard matrix specified by the Walsh code index, which must be an integer in the range $[0, \dots, N - 1]$. For the Walsh code index equal to an integer j , the respective Hadamard output code has exactly j zero crossings, for $j = 0, 1, \dots, N - 1$.

Kekre's Algorithm to generate Walsh Transform from Hadamard matrix [17]:

Step 1:

Arrange the 'n' coefficients in a row and then split the row in 'n/2', the other part is written below the upper row but in reverse order as follows:

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
15 14 13 12 11 10 9 8

Step 2:

We get two rows, each of this row is again split in 'n/2' and other part is written in reverse order below the upper rows as shown below.

0 1 2 3
15 14 13 12
7 6 5 4
8 9 10 11

This step is repeated until we get a single column which gives the ordering of the Hadamard rows according to sequency as given below:

0, 15, 7, 8, 3, 12, 4, 11, 1, 14, 6, 9, 2, 13, 5, 10

Step 3:

According to this sequence the Hadamard rows are arranged to get Walsh transform matrix. Now a product of Walsh matrix and the image matrix is calculated. This matrix contains Walsh transform of all the columns of the given image. Since Walsh matrix has the entries either +1 or -1 there is no multiplication involved in computing this matrix. Since only additions are involved computation complexity is very low.

3. Feature Vector Generation and Similarity Measure

The proposed algorithm makes novel use of Walsh transform to design the sectors to generate the feature vectors for the purpose of search and retrieval of database images. The complex Walsh transform is conceived by multiplying all sal functions by $j = \sqrt{-1}$ and combining them with real cal functions of the same sequency. Thus it is possible to calculate the angle by taking \tan^{-1} of sal/cal. However the values of tan are periodic with the period π radians hence it can resolve these values in only two sectors. To get the angle in the range of 0-360 degrees we divide these points in four sectors as explained below. These four sectors are further divided into 8, 12 and 16 sectors. We have proposed two different approaches for feature vector generation with absolute difference as similarity measure which is compared with Euclidean distance [8-10], [12-15] which is widely used as similarity measuring parameter. Sa-cal density distribution in complex transform plane and Individual sector mean are two different approaches to generate feature vectors. In the first approach the density distribution of sal and cal in each sector is considered for feature vector generation. Each Walsh sector is represented by single percentage value of sal cal distribution in particular sector w.r.t. all sectors for feature vector generation; calculated as follows:

$$\frac{\text{(Total number of sal in particular sector)}}{\text{(Total number of sal in all sectors)}} \times 100 \quad (1)$$

Thus for 8, 12 and 16 Walsh sectors with 8, 12 and 16 feature components for each color planes i.e. R, G and B are generated. The feature vector is of dimension 24, 36 and 48 components. In the second approach mean of the vectors which has sal and cal as components in each sector is calculated to represent each sector for all three color planes i.e. R, G and B. Thus forming the feature vector of dimension 12, 24, 36 and 48 for 4, 8, 12 and 16 complex Walsh transform sectors.

3.1 Four Walsh Transform Sectors:

To get the angle in the range of 0-360 degrees, the steps as given in Table 1 are followed to separate these points into four quadrants of the complex plane. The Walsh transform of the color image is calculated in all three R, G and B planes. The complex rows representing sal components of the image and the real rows representing cal components are checked for positive and negative signs. The sal and cal Walsh values are assigned to each quadrant. as follows:

Sign of Sal	Sign of Cal	Quadrant Assigned
+	+	I (0 – 90 Degrees)
+	-	II (90 – 180 Degrees)
-	-	III(180- 270 Degrees)
-	+	IV(270–360 Degrees)

TABLE 1. Four Walsh Sector formation

The equation (1) is used to generate individual components to generate the feature vector of dimension 12 considering three R, G and B Planes. The sal and cal density distribution in all sectors is used for feature vector generation. However, it is observed that the density variation in

4 quadrants is very small for all the images. Thus the feature vectors have poor discretionary power and hence higher number of sectors such as 8,12 and 16 were tried. In the case of second approach of feature vector generation i.e. individual sector mean has better discretionary power in all sectors. Absolute difference measure is used to check the closeness of the query image from the database image and precision recall are calculated to measure the overall performance of the algorithm. These results are compared with Euclidian distance as a similarity measure.

3.2 Eight Walsh Transform Sectors:

Each quadrants formed in the previous section obtained 4 sectors. Each of these sectors are individually divided into 2 sectors using angle of 45 degree as the partitioning boundary. In all we form 8 sectors for R,G and B planes separately as shown in the Table 2. The percentage density distribution of sal and cal in all 8 sectors are determined using equation (1) to generate the feature vector.

Quadrant of 4 Walsh sectors	Condition	New sectors Formed
I (0 – 90 ⁰)	Cal >= Sal	I (0-45 Degrees)
	Sal > Cal	II (45-90 Degrees)
II (90 – 180 ⁰)	Sal > Cal	III(90-135 Degrees)
	Cal >= Sal	IV(135-180 Degrees)
III (180- 270 ⁰)	Cal >= Sal	V (180-225 Degrees)
	Sal > Cal	VI (225-270 Degrees)
IV (270 – 360 ⁰)	Sal > Cal	VII (270-315 Degrees)
	Cal >= Sal	VIII (315-360 Degrees)

TABLE 2. Eight Walsh Sector formation

3.3 Twelve Walsh Transform Sectors:

Each quadrants formed in the previous section of 4 sectors are individually divided into 3 sectors each considering the angle of 30 degree. In all we form 12 sectors for R,G and B planes separately as shown in the Table 3. The percentage density distribution and mean value of sal and cal in all 12 sectors are determined to generate the feature vector

4 Quadrants	Condition	New sectors
I (0 – 90 ⁰)	Cal >= √3 * Sal	I (0-30 ⁰)
	1/√3 cal <=sal<= √3 cal	II (30-60 ⁰)
	Otherwise	III (60-90 ⁰)
II (90 – 180 ⁰)	Cal >= √3 * Sal	IV (90-120 ⁰)
	1/√3 cal <= sal <= √3 cal	V (120-150 ⁰)
	Otherwise	VI (150-180 ⁰)
III (180- 270 ⁰)	Cal >= √3 * Sal	VII (180-210 ⁰)
	1/√3 cal <= sal <= √3 cal	VIII(210- 240 ⁰)
	Otherwise	IX (240-270 ⁰)
IV (270 – 360 ⁰)	Cal >= √3 * Sal	X (270-300 ⁰)
	1/√3 cal <= sal <= √3 cal	XI (300-330 ⁰)
	Otherwise	XII (330-360 ⁰)

TABLE3. Twelve Walsh Sector formation

4. Results and Discussion



FIGURE 2: Query Image

The image of the class Bus is taken as sample query image as shown in the FIGURE. 2 for both approaches of sal cal density distribution and individual sector mean. The first 21 images retrieved in the case of sector mean in 12 Walsh sector used for feature vectors and Absolute difference as similarity measure are shown in the FIGURE, 3. It is seen that only 7 images of irrelevant class are retrieved among first 21 images and rest are of query image class i.e. Bus. Whereas in the case of sal cal density in 12 Walsh Sectors with Absolute Difference as similarity measures there are only 3 images of irrelevant class and 18 images of the query class i.e. Bus is retrieved as shown in the FIGURE.4.





FIGURE 3: First 21 Retrieved Images based on individual sector mean of 12 Walsh Sectors with Absolute Difference as similarity measures for the query image shown in the FIGURE 2.



FIGURE 4: First 21 Retrieved Images based on individual sector sal cal density in 12 Walsh Sectors with Absolute Difference as similarity measures for the query image shown in the FIGURE 2.

Once the feature vector is generated for all images in the database a feature database is created. A query image of each class is produced to search the database. The image with exact match gives minimum absolute difference. To check the effectiveness of the work and its performance with respect to retrieval of the images we have calculated the precision and recall as given in Equations (2) & (3) below:

$$\text{Precision} = \frac{\text{Number of relevant images retrieved}}{\text{Total Number of images retrieved}} \quad (2)$$

$$\text{Recall} = \frac{\text{Number of Relevant images retrieved}}{\text{Total number of relevant images in database}} \quad (3)$$

The FIGURE.5 - FIGURE.7 shows Overall Average Precision and Recall performance of sal cal density distribution in 8, 12 and 16 complex Walsh Transform sectors with Euclidian Distance (ED) and Absolute difference (AD) as similarity measures. The bar chart Comparison of Overall Precision and Recall cross over points based on individual sector mean of Walsh 4, 8,12 and 16 sectors and sal and cal density distribution with Euclidian distance (ED) and Absolute difference (AD) as similarity measure are shown in FIGURE 12 and FIGURE 13 respectively. It is observed that the Individual sector mean approach of feature vector generation gives the better retrieval compared to sal cal density distribution with both methods of similarity measures i.e. ED and AD.

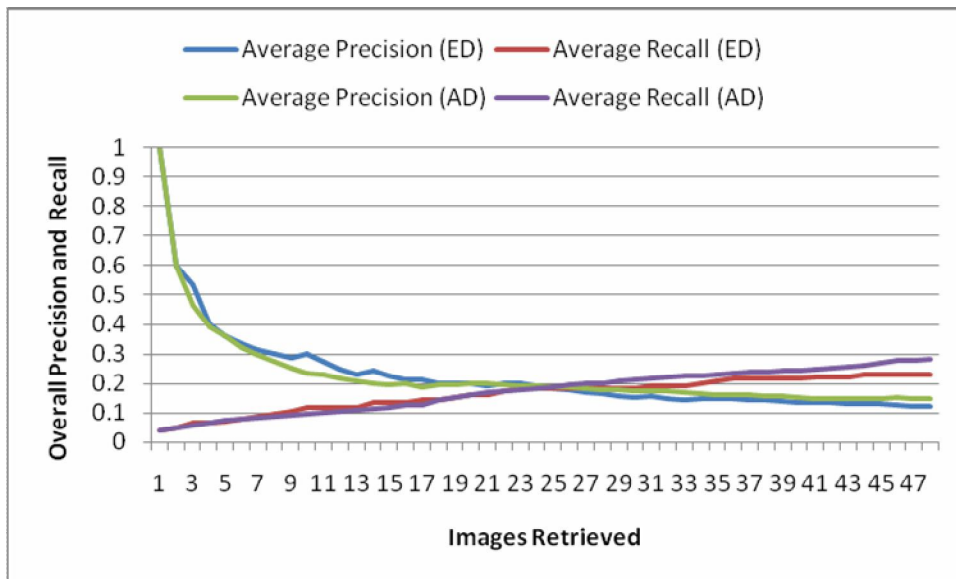


FIGURE. 5: Overall Average Precision and Recall performance of sal cal density distribution in 8 complex Walsh Transform sectors with Euclidian Distance (ED) and Absolute difference (AD) as similarity measures.

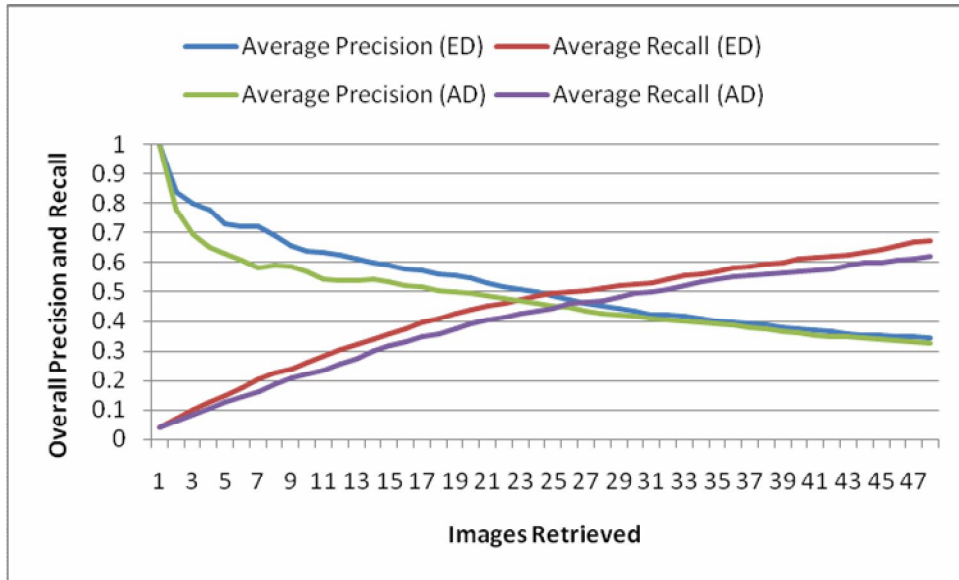


FIGURE. 6: Overall Average Precision and Recall performance of sal cal density distribution in 12 complex Walsh Transform sectors with Euclidian Distance (ED) and Absolute difference (AD) as similarity measures.

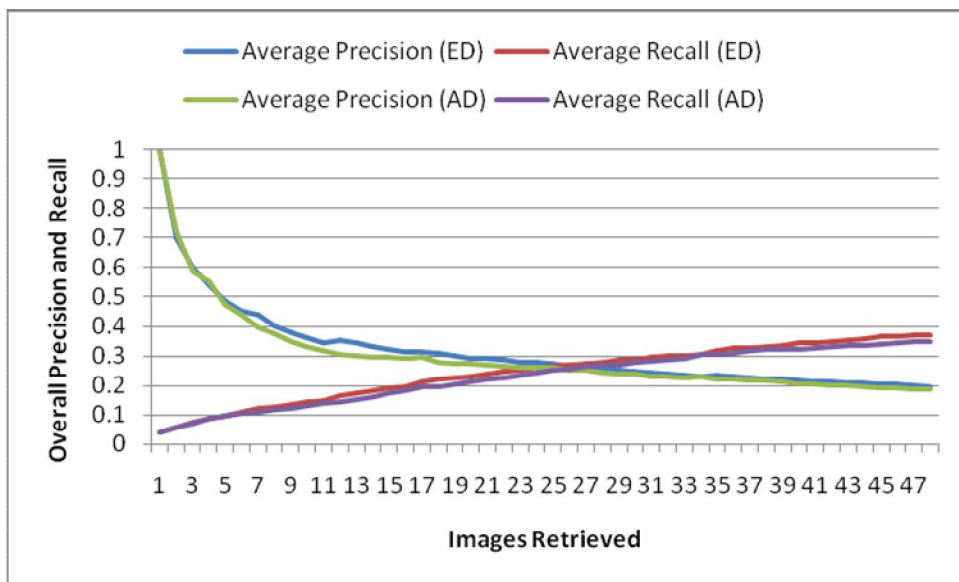


FIGURE 7: Overall Average Precision and Recall performance of sal cal density distribution in 16 complex Walsh Transform sectors with Euclidian Distance (ED) and Absolute difference (AD) as similarity measures.

The FIGURE.8 - FIGURE.11 shows the Overall Average Precision and Recall performance of individual sector mean of 4, 8, 12 and 16 Walsh Transform sectors with Euclidian Distance(ED) and Absolute Difference (AD) respectively. The performance of 4, 8, and 12 sectors give better retrieval rate of 0.59,0.568,0.512 in case of ED and 0.54,0.54 and 0.50 in case of AD .

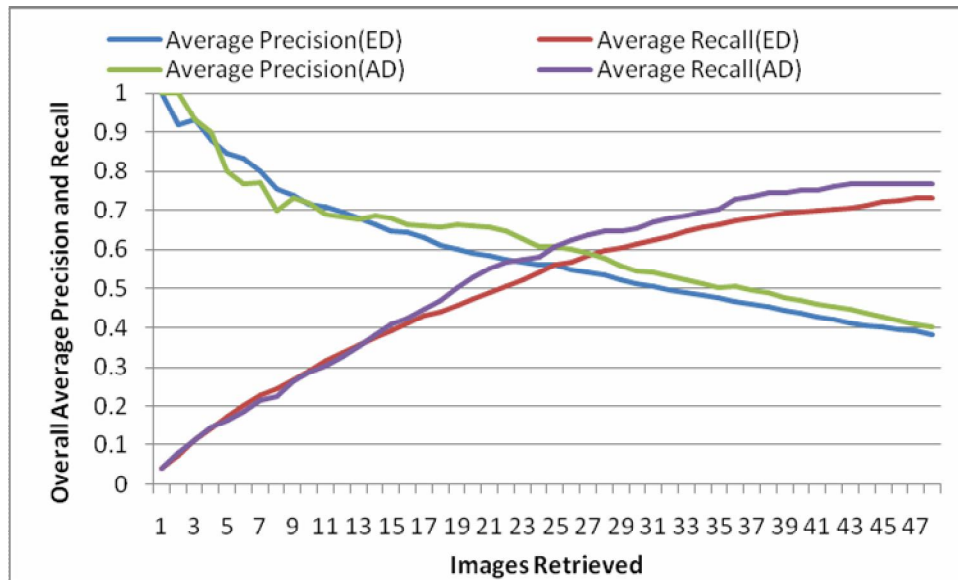


FIGURE 8: Overall Average Precision and Recall performance of individual sector mean in 4 complex Walsh Transform sectors with Euclidian Distance (ED) and Absolute difference (AD) as similarity measures.

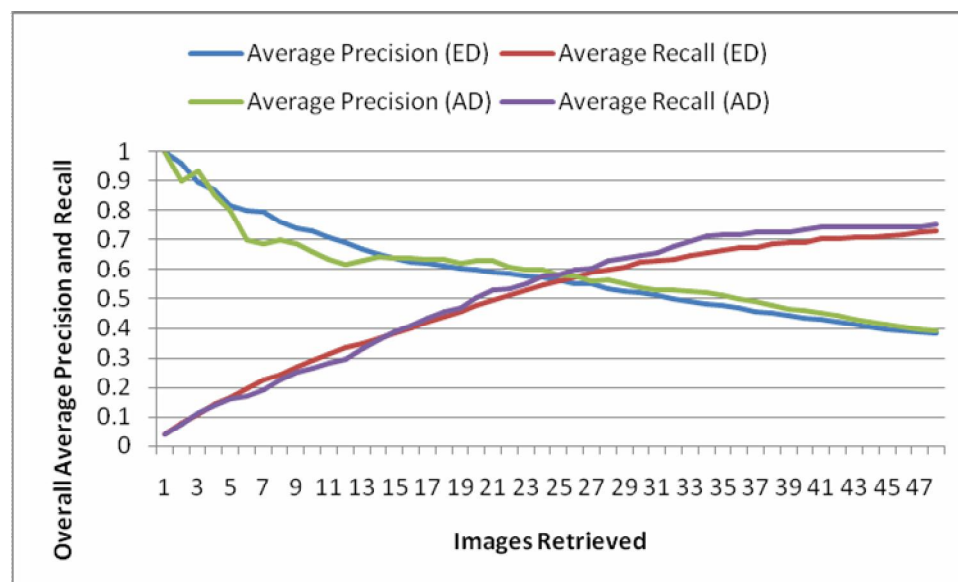


FIGURE 9: Overall Average Precision and Recall performance of individual sector mean in 8 complex Walsh Transform sectors with Euclidian Distance (ED) and Absolute difference (AD) as similarity measures.

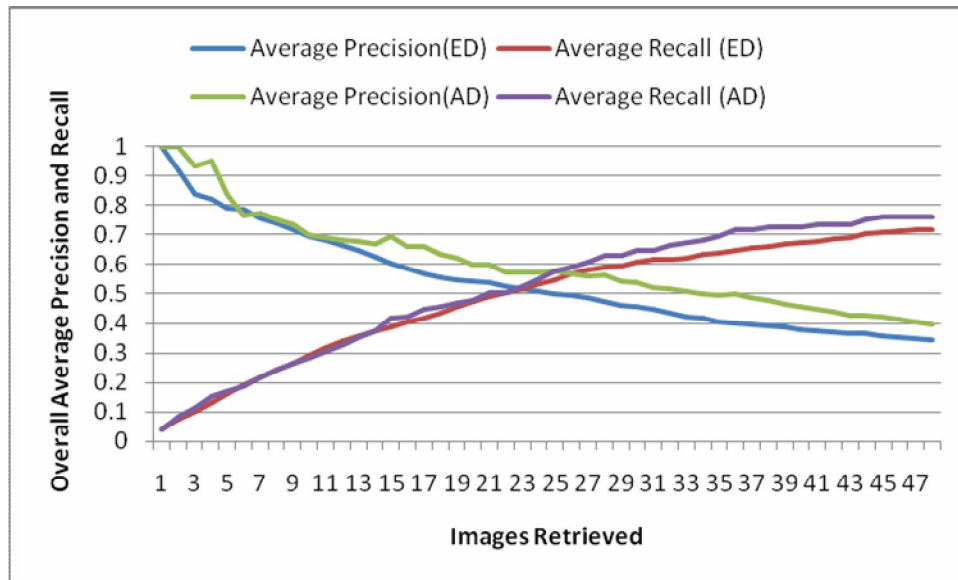


FIGURE 10: Overall Average Precision and Recall performance of individual sector mean in 12 complex Walsh Transform sectors with Euclidian Distance (ED) and Absolute difference (AD) as similarity measures.

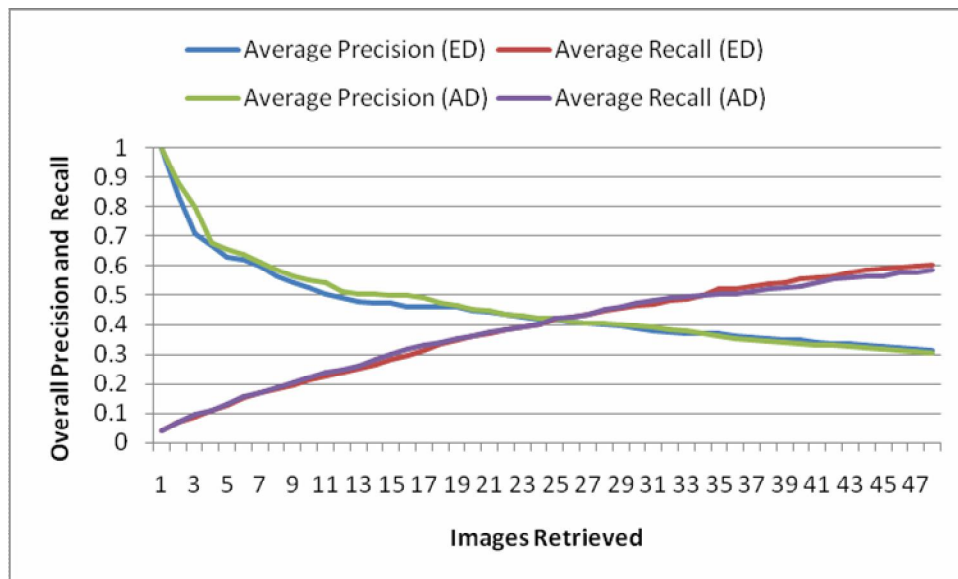


FIGURE 11: Overall Average Precision and Recall performance of individual sector mean in 16 complex Walsh Transform sectors with Euclidian Distance (ED) and Absolute difference (AD) as similarity measures.

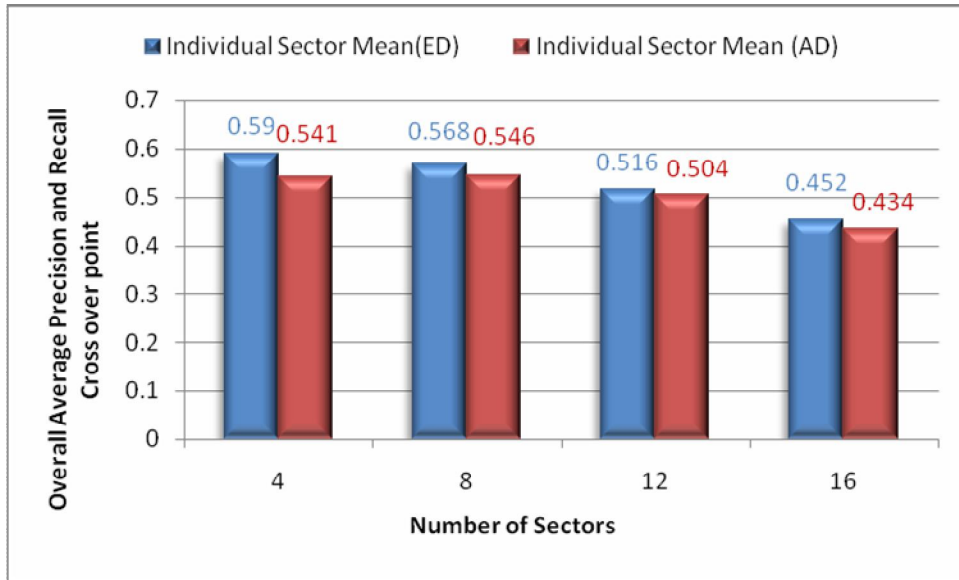


FIGURE 12: Comparison of Overall Precision and Recall cross over points based on individual sector mean of Walsh 4, 8,12 and 16 sectors with Euclidian distance (ED) and Absolute difference (AD) as similarity measure.

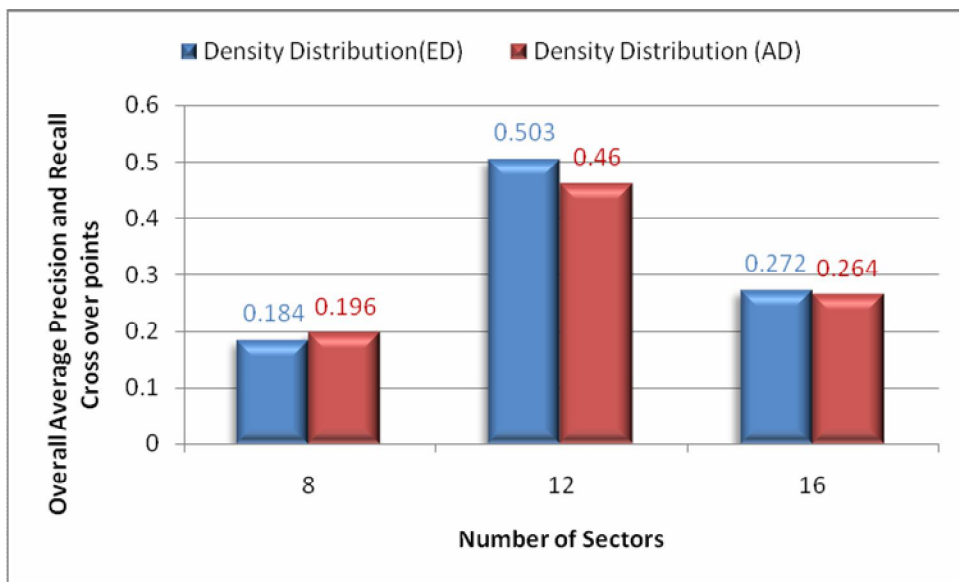


FIGURE 13: Comparison of Overall Precision and Recall cross over points based on sal and cal density distribution in Walsh 8,12 and 16 sectors with Euclidian distance (ED) and Absolute difference (AD) as similarity measure.

5. CONCLUSION

The Innovative idea of using complex Walsh transform 4, 8, 12 and 16 sectors of the images to generate the feature vectors for content based image retrieval is proposed. We have proposed two different approaches for feature vector generation namely density distribution and mean value of the vectors having sal and cal components of same sequency. In addition we have proposed a new similarity measure namely sum of absolute difference of feature vector

components and its results are compared with commonly used Euclidian distance. The cross over point performance of overall average of precision and recall for both approaches on all applicable sectors are compared. We have used the database of 270 images having 11 different classes. The results are summarized below:

Using Walsh transform and absolute difference as similarity measuring parameter which requires no multiplications reduces the computational complexity reducing the search time and calculations by a factor of 8.

The performance of 8, 12 and 16 sectors is compared for density distribution. It is found that 12 sectors give the best result followed by 16 and 8. 12 sectors give the best outcome of overall average precision and recall cross over point at 0.503 compared to 0.184 and 0.272 for 8 and 16 sectors respectively as shown in the FIGURE. 12.

For mean value the best results are obtained for sector 4. Here also the Euclidian distance measure gives cross over point as 0.59 compared to 0.541 in case of Absolute difference. It is also observed that the mean value approach has better discretion as compared to density distribution. Finally it seems that absolute difference similarity measure is much faster than Euclidian distance measure but this happens at slight degradation of the performance.

5. REFERENCES

1. John Gantz and David Reinsel, "The digital universe decare- are you ready", IDC iVew, May 2010, EMC Corporation
2. Anil Jain, Arun Ross, Salil Prabhakar, "Fingerprint matching using minutiae and texture features," *Int'l conference on Image Processing (ICIP)*, pp. 282-285, Oct. 2001.
3. Kato, T., "Database architecture for content-based image retrieval in Image Storage and Retrieval Systems" (Jambardino A and Niblack W eds), *Proc SPIE 2185*, pp 112-123, 1992.
4. John Berry and David A. Stoney "The history and development of fingerprinting," in *Advances in Fingerprint Technology*, Henry C. Lee and R. E. Gaensslen, Eds., pp. 1-40. CRC Press Florida, 2nd edition, 2001.
5. Emma Newham, "The biometric report," SJB Services, 1995.
6. H. B. Kekre, Dharendra Mishra, "Digital Image Search & Retrieval using FFT Sectors" published in proceedings of National/Asia pacific conference on Information communication and technology (NCICT 10) 5TH & 6TH March 2010. SVKM'S NMIMS MUMBAI
7. H.B.Kekre, Dharendra Mishra, "Content Based Image Retrieval using Weighted Hamming Distance Image hash Value" published in the proceedings of international conference on contours of computing technology pp. 305-309 (Thinkquest2010) 13th & 14th March 2010.
8. H.B.Kekre, Dharendra Mishra, "Digital Image Search & Retrieval using FFT Sectors of Color Images" published in International Journal of Computer Science and Engineering (IJCSE) Vol. 02, No. 02, 2010, pp. 368-372 ISSN 0975-3397 available online at <http://www.enggjournals.com/ijcse/doc/IJCSE10-02-02-46.pdf>
9. H.B.Kekre, Dharendra Mishra, "CBIR using upper six FFT Sectors of Color Images for feature vector generation" published in International Journal of Engineering and Technology (IJET) Vol. 02, No. 02, 2010, 49-54 ISSN 0975-4024 available online at <http://www.enggjournals.com/ijet/doc/IJET10-02-02-06.pdf>
10. Arun Ross, Anil Jain, James Reisman, "A hybrid fingerprint matcher," *Int'l conference on Pattern Recognition (ICPR)*, Aug 2002.
11. A. M. Bazen, G. T. B. Verwaaijen, S. H. Gerez, L. P. J. Veelenturf, and B. J. van der Zwaag, "A correlation-based fingerprint verification system," *Proceedings of the ProRISC2000 Workshop on Circuits, Systems and Signal Processing*, Veldhoven, Netherlands, Nov 2000.
12. H.B.Kekre, Tanuja K. Sarode, Sudeep D. Thepade, "Image Retrieval using Color-Texture Features from DCT on VQ Codevectors obtained by Kekre's Fast Codebook Generation",

- ICGST International Journal on Graphics, Vision and Image Processing (GVIP), Available online at <http://www.icgst.com/gvip>
13. H.B.Kekre, Sudeep D. Thepade, "Using YUV Color Space to Hoist the Performance of Block Truncation Coding for Image Retrieval", IEEE International Advanced Computing Conference 2009 (IACC'09), Thapar University, Patiala, INDIA, 6-7 March 2009.
 14. H.B.Kekre, Sudeep D. Thepade, "Image Retrieval using Augmented Block Truncation Coding Techniques", ACM International Conference on Advances in Computing, Communication and Control (ICAC3-2009), pp.: 384-390, 23-24 Jan 2009, Fr. Conceicao Rodrigous College of Engg., Mumbai. Available online at ACM portal.
 15. H.B.Kekre, Tanuja K. Sarode, Sudeep D. Thepade, "DCT Applied to Column mean and Row Mean Vectors of Image for Fingerprint Identification", International Conference on Computer Networks and Security, ICCNS-2008, 27-28 Sept 2008, Vishwakarma Institute of Technology, Pune.
 16. H.B.Kekre, Sudeep Thepade, Archana Athawale, Anant Shah, Prathmesh Velekar, Suraj Shirke, "Walsh transform over row mean column mean using image fragmentation and energy compaction for image retrieval", International journal of computer science and engineering (IJCSE), Vol.2.No.1, S2010, 47-54.
 17. H.B.Kekre, Vinayak Bharadi, "Walsh Coefficients of the Horizontal & Vertical Pixel Distribution of Signature Template", In Proc. of Int. Conference ICIP-07, Bangalore University, Bangalore. 10-12 Aug 2007
 18. H.B.Kekre, Kavita sonawane, "DCT Over Color Distribution of Rows and Columns of Images for CBIR" SANSKODHAN, SFIT Technical Magazine No 5, pp. 11-17, March 2010.
 19. H.B.Kekre, Tanuja Sarode, "Two Level Vector Quantization Method for Codebook Generation using Kekre's Proportionate Error Algorithm" International Journal of Image Processing, Vol.4, Issue 1, pp.1-10, January-February 2010 .
 20. Dr. H. B. Kekre, Sudeep D. Thepade, Akshay Maloo, " Performance Comparison of Image Retrieval Using Fractional Coefficients of Transformed Image Using DCT, Walsh, Haar and Kekre's Transform", International Journal of Image Processing, Vol.4, Issue 2, pp.142-155, March-April 2010.
 21. Hiremath P.S, Jagadeesh Pujari, "Content Based Image Retrieval using Color Boosted Salient Points and Shape features of an image", International Journal of Image Processing (IJIP) Vol.2, Issue 1, pp 10-17, January-February 2008.

A Fuzzy Watermarking Approach Based on the Human Visual System

Sameh Oueslati

sameh.oueslati@telecom-bretagne.eu

*Faculty of Sciences of Tunis,
Department of Physics,
Laboratory of Signal Processing,
University Tunis El Manar, TUNIS, 1060, TUNISIA.*

Adnane Cherif

adnane.cher@fst.rnu.tn

*Faculty of Sciences of Tunis,
Department of Physics,
Laboratory of Signal Processing,
University Tunis El Manar, TUNIS, 1060, TUNISIA.*

Basel Solaiman

Basel.Solaiman@telecom-bretagne.eu

*Higher National School
of Telecommunication of Bretagne,
Department: Image and Information Processing,
Technopole of Brest Iroise, 29285 Brest, FRANCE.*

Abstract

The implementation of our watermarking system is based on a hybrid system combining the human visual system (HVS) and the fuzzy inference system (FIS), which always passes through the transcription of human expertise in the form of fuzzy rules expressed in natural language, which allows our watermarking system remain understandable for non expert and become more friendly. The technique discussed in this paper is the use of an advanced approach to the technique of watermark that is the multi-watermark or the watermarking multiple of medical images in the frequency domain. In this approach, the emphasis will be on the safe side and the invisibility while maintaining robustness against a certain target range of attacks. Furthermore, this approach is based on a technique totally blind as we will detail later.

Keywords: Digital watermarking, Fuzzy inference system (FIS), insertion force, The human visual system (HVS), Robustness.

1. INTRODUCTION

The new technologies of information and communication media are likely to make doctors and patients decisive help in finding a better quality of care. But the ease of access afforded by these new technologies poses the problem of the security of information particularly of medical images [10]. The role of security imposed on the medical image is attributed to watermarking [17]. However, its use in the domain of medical image is rare and few studies have been devoted. Even if the watermarking images are generally translated by imperceptible signals, the very fact that it modifies the host image hinders its spread in the hospital [25] [15]. However, in this article a major constraint and a question arose: How should provide a robust watermarking system applied in the medical field?

It is clear that we want it to be tamperproof and must therefore be protected by keys ensuring this security. It is not necessary that it has to be robust to a wide range of attacks as the most robust watermarking that protect the copyrighted images. Indeed, if the patient dares to degrade the watermark, he probably will alter the image also and the authenticity will be suspect at the same time it would degrade the diagnosis.

The work presented in this paper proposes a new watermarking technique based on a fuzzy inference system to extract the knowledge of sensitivity of human eye using the HVS model. The FIS and the HVS handsets are used in this work to automatically determine the insertion force in certain frequency components without being visually perceptible. Like most watermarking algorithms proposed, the invisibility of the watermark is obtained by using different properties of human visual system (HVS).

2. The Human visual system model

To hide the watermark in the image, it is useful to exploit the weaknesses of the human visual system [19, 13, 20, and 4]. However, the study of human perception is not confined only to the understanding of the optical device that represents the eye: Between the visions of the image that is printed on the retina and its interpretation in the human brain. The road is long and involves complex processes that are, todate, far from being fully mastered.

We will try in what follows to present the essential characteristics of human visual system that can be exploited to improve the imperceptible signature and to watermark the image with an adjusted coefficient by its internal dynamics.

We focus here on the sensitivity of brightness, texture and frequency. The HVS model used in this work has been suggested in [1, 2]. This model is also used in many insertion algorithm and detection of the watermark.

2.1 Luminance sensitivity (L_k)

Firstly, the eye is sensitive to contrast, and although the human visual system is capable of detecting small differences of luminance, there is a limit below which the differences are no longer collected [21].

This limit depends on the luminance L_0 of the bottom on which are the components of the image. Indeed, we notice hardly the changes of the image due to watermarking in the areas where the contrast is very important. The brighter the background, the lower the visibility of the embedded signal. Therefore a longer and a stronger embedded signal can be used. The luminance sensitivity is estimated by the following formula:

$$L_k = \left(\frac{V_{DC,k}}{V_{DC}} \right)^\gamma \quad (1)$$

Where $V_{DC,k}$ is the DC coefficient of the DCT of the k^{th} block, $\overline{V_{DC}}$ is the mean value of all $V_{DC,k}$ coefficients of a specific image, and γ is set to 0.649 to control the degree of luminance sensitivity.

2.2 Texture sensitivity (T_k)

The degradations caused by the signature will be very low in the homogeneous areas of the image [3, 23], but may be more intense in highly textured areas for which the eye will not be able to differentiate between the signal from the image and the signal from the signature [8, 5, and 14]. The stronger the texture is less visible is the watermark. The texture sensitivity can be estimated by quantizing the DCT coefficients of an image using the JPEG quantization table. The result is then rounded to nearest integers. The number of non-zero coefficients is then computed. This method can be calculated by the following formula:

$$T_k = \sum_{x,y=1}^N \text{cond}\left(\left[\frac{V_k(x,y)}{Q(x,y)}\right]\right) \quad (2)$$

Where (x, y) represents the location in the k^{th} block. And $\text{cond}(R)$ takes the rounded value of R and returns '1' if the value is not equal to zero, '0' otherwise.

2.3 Frequency sensitivity (F_k)

The insertion in the low frequency band of DCT blocks correspond to homogeneous zones in the image features a marking robust but distorts apparently the watermarked image [6, 12, and 16]. Contrariwise, an insertion in the high frequency band corresponding to the contours and abrupt changes of intensity of gray in the image, will characterize an imperceptible marking but fragile. Hence the needs to choose a medium frequency band that provide a compromise between the degradation of image quality and very visible and a maximum resistance to attacks. The sensitivity of frequency is represented by the quantization table (luminance).

3. Fuzzy Inference System (FIS)

In our fuzzy logic approach was a process of insertion well known by a human operator. The goal however is to automate. This is called knowledge by an expert who knows what to do in all cases; Increase the insertion force in less sensitive areas (areas textured) Decrease the strength of integration in sensitive areas (edges and homogeneous areas) ... ect. The fuzzy logic is used in our work, given the uncertainty and the changes in the unknown parameters (insertion force, pixels carrying the signature ...) and the structure of the watermarking system. When human experts are available to provide subjective and qualitative descriptions of behavior of system watermark with words in natural language. The expertise of the operator, often consists of among other simple rules, allows him to manage the system more properly a classical algorithm.

The principle of fuzzy logic approaches the human approach in the sense that the variables treated are not logical variables (within the meaning of the logic for example) but of variable linguistic relatives of human language as « a little, a lot more clearly ... very textured, very homogeneous etc » Moreover, these linguistic variables are processed with rules that refer to some knowledge of system behavior, most often obtained empirically. A fuzzy inference system consists of three blocks as shown in Figure 1. The first floor of fuzzification that converts numeric values to degrees of membership of different fuzzy sets. The second block is the inference engine, consisting of all rules. Finally, a defuzzification stage allows, if necessary, to infer a net worth, from the result of the aggregation rules.

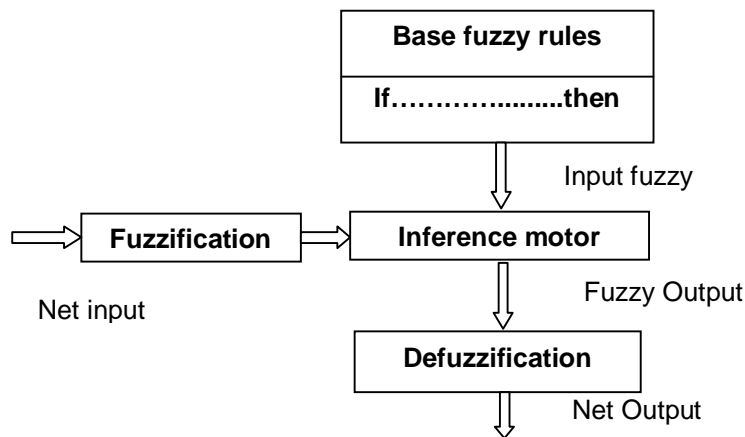


FIGURE 1: A Fuzzy inference system.

In what follows, we will retain only the essential elements to understanding the principle of adjustment of force of insertion in a watermarking system by fuzzy logic; these elements are fuzzy variables, rules of inference, and the membership functions.

3.1 The fuzzy variables

The binary logic has the advantage of simplicity but is rather distant from the logic used naturally by humans. If we represent the same problem using fuzzy logic, variables are not binary but have an infinite number of possible values between the « true » and the « false ».

Note that this representation is much closer to how human's reason, since it can involve concepts such as « smallish, a little, clearer ... highly textured, very homogeneous etc ». This advantage is, of course, at the expense of simplicity of representation. The fuzzy variables have a graduation between these two values.

3.2 Rules of inference

The role of the expert is here because he will fix the rules of the command that will cover only the linguistic values of variables.

We call rules of inference, all fuzzy rules connecting the different variables of a fuzzy system input variables and fuzzy output of this system. These rules come in the form:

- If (condition 1) and / or condition (N) then (action on the outputs).

In terms of artificial intelligence, summarize these rules makes the experience of the expert and they are usually not uniquely definable as each individual creates his own rules. It is necessary to convert quantities into variables crowds. To do this we define two notions.

- The membership functions that define the degree of truth of fuzzy variable depending on the input.

- The fuzzy intervals which determine the number of fuzzy variables.

3.3 The membership functions

It is a relationship between the degree of truth of fuzzy variable and the input correspondent. This is « fuzzification » this is to specify the domain of variation of variables: the universe of discourse, which is divided into intervals (in fuzzy sets or linguistic values). This distribution, which is to determine the number of such securities and distribute them on the domain, is made based on knowledge of the system according to the desired accuracy. The input and output membership functions exploited in this scheme are shown in Fig. 2. It is important to realize that this approach enables the luminance sensitivity (or texture) membership functions to be adjusted in such a manner to best fit the image's properties. In consequence, the approximations of the inferred values are optimized and are used to generate the adaptive strength of the watermark. Although the membership functions used in our algorithm consists of triangular and trapezoidal functions, they are not limited to these forms. We can obviously choose any form for membership functions.

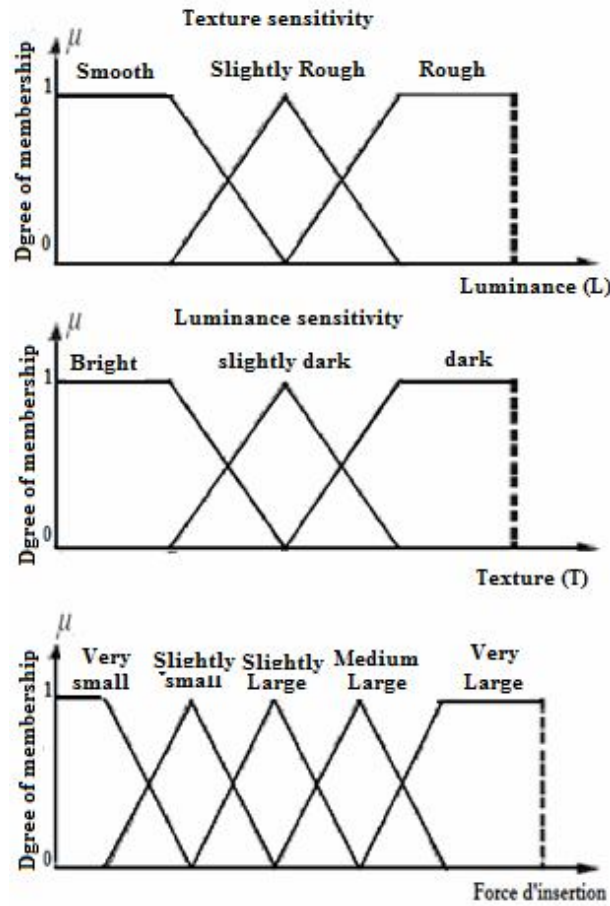


FIGURE 2: Dynamic Membership Functions and mapping of their input/output variables to fuzzy sets

3.4 Defuzzification

On output, the fuzzy system can not provide fuzzy values that can only operate. It is therefore necessary to provide accurate values, the role of the Defuzzification. This step is always done using the membership functions. From the degrees of membership functions, we obtain as many values it has States. It is therefore necessary to provide accurate values, the role of the Defuzzification. This step is always done using the membership functions. From the degrees of membership functions, we obtain as many values it has States. To determine the exact value to use, you can either keep the maximum, either calculates the weighted average, or to determine center of gravity values. In this work, the inference results are subsequently computed by means of the centroid defuzzification method, where the inferred value i_k of a specific block k of an image is calculated as in Equ. (3), where μ_c is the aggregated resultant membership function of the output fuzzy sets and i_n is the universe of discourse corresponding to the centroid of μ_c .

$$i_k = \frac{\sum_{n=1}^N \mu_c(i_n) i_n}{\sum_{n=1}^N \mu_c(i_n)} \quad (3)$$

In order to compute the adaptive watermark strength, the inferred value i_k is multiplied by the frequency sensitivity as it is shown in the following formula:

$$\alpha_{x,y,k} = F_{x,y} \cdot i_k \quad (4)$$

Where $\alpha_{x,y,k}$ corresponds to the adaptive strength of a watermark at index (x, y) of the k^{th} block of an image. Also, $F_{x,y}$ corresponds to the frequency sensitivity at index (x, y) .

4. The Proposed Approach

The FIS and the combined HVS are used in this work as shown in Figure (3), to adjust and determine a way to force automatic insertion. When the watermarking image, it must reflect compromise robustness invisibility. To enhance the robustness against various image distortions that can undergo using an insertion force α which must be below a threshold of visual perceptibility.

This force is not always consistent on all components of the watermark inserted but depends on the characteristics of the zones of insertion (Textured, uniform), because the human eye does not detect changes in low luminance, even better if they are hidden under sudden changes such as edges, contours, textures, etc.. This force depends not only on the human visual system (HVS), but also characteristics of the zones of insertion.

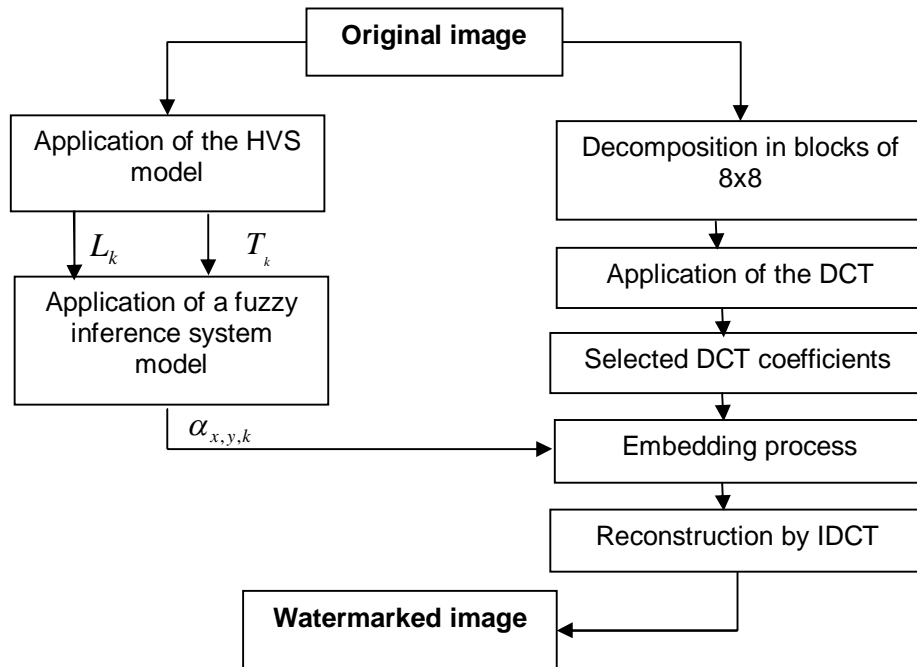


FIGURE 3: Embedding process based on the HVS and FIS

4.1 The Watermark Embedding Process

Among the work of watermarking proposed, algorithm [22], this consists of coding on a pair of frequency values $\{0, 1\}$. The use of frequency domain DCT can fulfill not only the invisibility through the study of optimization of the insertion force used, but also security by providing a blind algorithm which use the original image is not essential and the extraction of the mark is through a secret key [24,18,11]. In addition we will target a robust con of specific distortions such as lossy JPEG compression, the approach we propose, makes the coding of several marks in binary on DCT coefficients selected from each block of the image [23]. It proceeds in the following illustration of the stages of the insertion procedure. We proceed by determining the number of

coefficients to be used for coding the mark setting their Cartesian coordinates in the mid-band frequency of each DCT block size (8×8).

The number of coefficients chosen depends on the size of mark, the level of redundancy of the mark which is reflected by the number of coefficients carry a single mark.

We proceed to a linearization of the mark $M = [0,1,1,0,1,1,1,0,0,0..]$ of size $n = N \times N$. In each block we chose two coefficients of the same value of quantification by JPEG quantization table.

A scan on the blocks selected through the key positions is essential for ordering coefficients as follows:

For i varying from $1 : n$,

Denotes C_1 and C_2 DCT coefficients and mark holders.

If $M(i) = 0$ then we check:

If $C_1 < C_2$ then $C_1 = C_2$ and $C_2 = C_1$.

So the order becomes $C_1 > C_2$.

If $M(i) = 1$ then we check:

If $C_1 \geq C_2$ then $C_1 = C_2$ and $C_2 = C_1$.

So the order becomes $C_1 < C_2$.

Thus the entire sequence of binary vector represents a scheduling coefficient holder C_1 and C_2 is applied to characterize the value of the bit up the message blocks identified by the key. The insertion step is then applied as follows:

If $C_1 - C_2 \leq \alpha$ and if $C_1 > C_2$ then:

$$\begin{cases} C_1 = C_1 + \frac{\alpha}{2} \\ C_2 = C_2 - \frac{\alpha}{2} \end{cases} \quad \text{Otherwise} \quad \begin{cases} C_1 = C_1 - \frac{\alpha}{2} \\ C_2 = C_2 + \frac{\alpha}{2} \end{cases} \quad (5)$$

4.2 The Watermark Detection Process

In our approach we have two key insertion markings to secure the site where the mark was introduced. The first key indicates the positions of the two selected coefficients with values of quantification. While the second key relates to the position of the blocks which will carry the marks among the total of all constituent blocks the transformed image. The extraction step is as follows: Compare the values of DCT coefficients to determine if the bit concerned the message was a "0" or "1".

If $C_1 > C_2$ and $C_1 - C_2 \geq \alpha$ then $M(i) = 0$

Otherwise $M(i) = 1$.

Thus we come to remove the marks included in all blocks without using the original image.

5. Validation of the New Approach

For medical images the signature must be imperceptible: The watermarked image should be broadly similar to the original image so as not to lead to a misdiagnosis. The validity of any technical watermark can become more important than testing it against various types of attacks. For this, we chose to subject the medical image watermarked a series of attacks and test the sensitivity of the watermark and its ability to detect any change in the image. After application of each attack, all the marks included are extracted and compared through a study of similarities

with the original marks (M' and M) to be sure that these marks are not damaged by the attacks applied on the image.

5.1 The correlation coefficient

It measures the ratio of similarity between two images to show the close resemblance between the original mark inserted M and the extracted M' .

More it is close to 1 more likeness is perfect, and vice versa. This similarity measure is given by the following equation:

$$corr = \frac{\sum_1^N MM'}{\sqrt{\sum_1^N M^2 \sum_1^N M'^2}} \tag{6}$$

5.2 The PSNR (Peak Signal to Noise Ratio)

PSNR, which is presented in equation (7), is an assessment of the decibel difference between the original image and one that is processed. In fact, a PSNR below 30 dB image can be considered untreated, inexploitable.

$$PSNR = 10\log_{10}\left(\frac{X_{max}^2}{MSE}\right) = PSNR = 10\log_{10}\left(\frac{255^2}{MSE}\right) \tag{7}$$

Where X_{max} : The maximum luminance. Where MSE is the mean-square error between the original image and the distorted one.

MSE is defined as:

$$MSE = \frac{\sum_{i=1}^N \sum_{j=1}^M (I_{ij} - I_{ij}^*)^2}{NM} \tag{8}$$

Where M , N is the size of the image and contains $N \times M$ pixels I is the host image and I^* is the watermarked image.

6. Simulation and Discussion

Regardless of the domain of insertion of the watermark, have a good PSNR is an important requirement especially for medical images, where image quality plays a major role for better use of the image.

This work has been applied to several images of size 256x256 pixels resolution and 8 bits / pixel different. We begin our evaluation by a preliminary study, that is to say, a study without the application of attack on the tagged images.

Figure 4 shows the original medical image and tattooed. We note that the human visual system does not distinguish the difference caused by the marking.

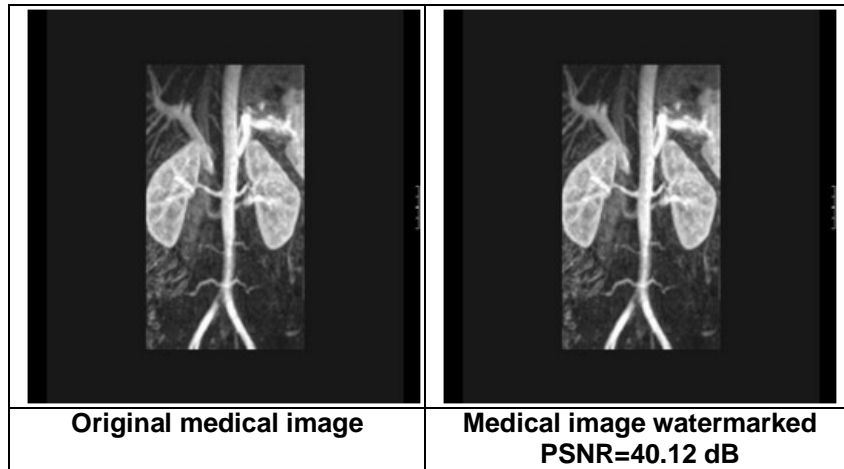


FIGURE 4: Original image and image watermarked.

The first test of robustness for an attack innocent which is compression. The compression algorithms are particularly dangerous for watermarking process. It is through the use of these algorithms do not keep the image that the components essential to their understanding (an invisible mark is obviously not essential). Therefore in the proposed approach, we inserted the signature in places fairly significant image, and it was based on a studied selection of DCT coefficients. We have chosen to apply the watermarked images at different compression ratios as shown in the table below which presents the results of simulations showing the rate of extraction of the similarities after contested mark with the original face of attacks applied.

TABLE 1: Presentation of the correlation values after attack by JPEG compression between the marks extracted and the original.

Index	JPEG compression attack	Correlation values
1	20	0.9245
2	40	0.9482
3	60	0.9691
4	70	0.9825
5	80	1
6	90	1

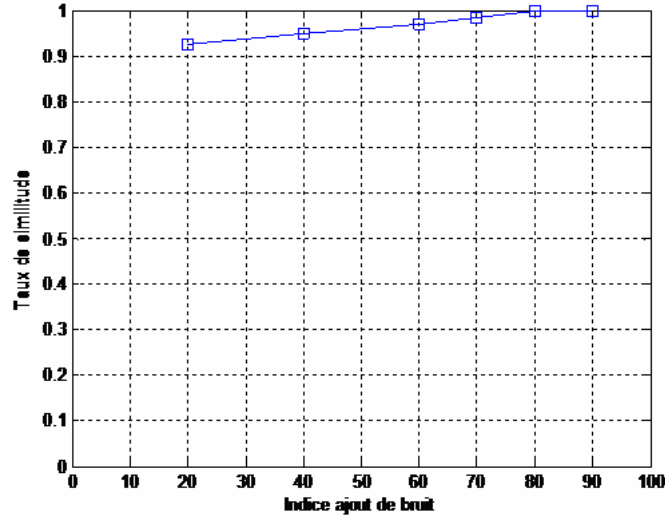


Figure 5: Original image and image watermarked.

The second test of robustness is the median filter that is most used in image processing. This filter replaces the pixel value by the median of its neighbors that exist in the analysis window centered on that pixel. The size of the window set the filter type; the types of filters that are treated in this section are listed in the table below.

TABLE 2: Correlation values after attack by median filter.

Median filter size	3x3	5x5	7x7
Correlation values	0.9158	0.8687	0.7459

The third attack is the application of a noise. This test has three types of noise: additive noise, gaussian noise and speckle noise. Following the application of an additive noise presents results similar rates in Table 3. This noise is measured by the SNR (Signal to Noise Ratio) to measure the quality of the noisy image compared to the original. This parameter is given by:

$$SNR = 10 \times \log_{10} [Var(I_B) / Var(I_B - I_0)] \tag{9}$$

With I_B the image is noisy and I_0 is the original image. The variance of noise added is equal to $X\%$ the variance of the non-noisy image.

TABLE 3: Recognition rates between the extracted watermark and the original after attack by noise.

Index	Index of adding noise	Correlation values
1	20	0.8542
2	30	0.8093
3	40	0.7845
4	50	0.7285
5	60	0.6523

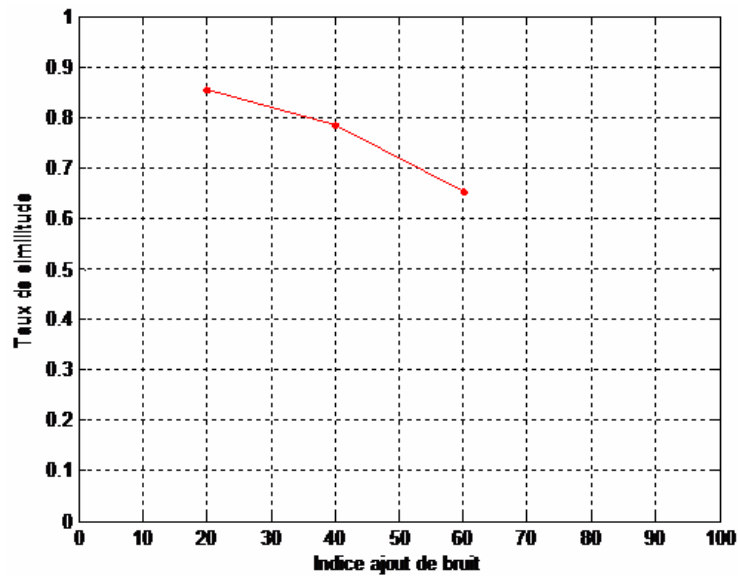





Figure 6: Presentation of correlation values according in the rate of noise.

We tested the robustness of the proposed approach compared to several generations as possible noise by varying each time the type of noise or variance. The PSNR values have their maxima for different variances of the multiplicative noise (speckle) that retains the best quality of the watermarked image. The Gaussian noise gives the lowest values of PSNR (relative to the other two types of noise tested). However, PSNR values are always above 30 dB, which allows the image to be usable as well. Noise "salt and pepper" does not alter too much the image quality digitally, but visually, the images seem to be degraded.

Table 4 summarizes the results established for the watermarked images after applying the three types of noise.

TABLE 4: Image watermarked attacked by different types of noise.

salt and paper noise	Gaussian noise	speckle noise
		
Var=0.03, PSNR=37.44dB	Var=0.03, PSNR=32.13dB	Var=0.03, PSNR=34.78dB

At the end, we evaluate our approach to asynchronous attacks. In what follows we focus on the rotation causes a change in the position of pixels causing a change in the information inserted in the image represented by the mark. The rotation angles used vary between 1° and 5°.

From Figure 7 the large angles of rotation were the most degraded quality digital image. They close the lower rate of similarity values. This degradation affects the holders of the signature involving loss of information.

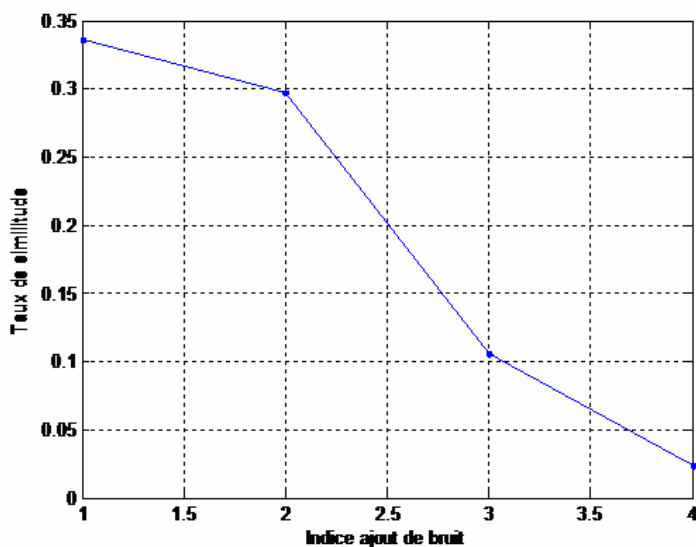


Figure 7: Presentation of correlation values according to the angles of rotation.

By scanning all blocks insertion of the image watermarked we succeeded in extracting the marks inserted in all the blocks used without recourse to the original image.

Note that all tests of medical images in our database, the transposition of the two coefficients by inverting their positions will not affect the image quality and does not introduce changes or visual distortions.

The whole procedure of insertion and extraction applied to the DCT.

This technique improves the safety of blind media tattooed cons of specific distortions such as adding noise or filtering or compression that can occur during transmission or storage.

More, only marked the image used in the process of extraction of patient data that are secured through two key positions indicating the location of coefficients holders and blocks insertion.

Independently of the area of insertion of watermarking, have a good PSNR is an important requirement especially for medical images, where the image quality plays a major role for better use of the image.

The study shows the variation of PSNR and rate of similarity based on different levels of attacks lossy compression, filtering, Median, noise and rotation are presented in Tables (1, 2, 3 and 4) and figures (5, 6 and 7). The marks obtained after the attack type rotation showed a slight similarity with the original image showing that the mark has been inserted damaged by this type of attack used.

7. CONSLUSION & FUTURE WORK

This paper presents a new blind watermarking scheme in the frequency domain based on the HVS and the FIS, which can increase the robustness of watermarking images in terms of quality and confidentiality. This approach of watermarking replacement has positive effects on the robustness of watermarking as it has allowed increasing the coefficient of insertion in certain frequency components without it being visually perceptible. We decided to validate the algorithm against such attacks; filtering, adding noise, rotation and focus here in particular to the JPEG compression. It was proved that the ability of this method is interesting. It resists attacks JPEG acceptable. In contrast, the method proves fragile compared to geometrical attacks. It is not sufficiently reliable to the robustness and security of the watermarking scheme is essential. Although some domains are robust against attacks synchronous, they may submit a weakness

against specific attacks or asynchronous. None of these domains can provide robustness against a wide range of synchronous and asynchronous attacks except through combinations of multiple insertion domains.

8. REFERENCES

1. A.B. Watson, "DCT: A technique for visual optimization of DCT quantization matrices for individual images". Society for Information Display Digest of Technical Papers, pp. 946-949, 1993
2. N.S. Jayant, J.D. Johnson and R. Safranek. "Signal compression based on models of human perception". Proceedings of IEEE, Volume 81, No. 10, pp. 1385-1422, 1993
3. B. Miladi, M. Sayadi and F. Fnaiech. "Textures synthesis methods". International Conference on Electrical Systems and Automatic control, Tunisia, 2010
4. S. Zaboli, A. Tabibiazar and R. Safabakhsh. "Entropy-Based image watermarking using DWT and HVS". 3rd International Conference of Sciences of Electronic, Technologies of Information and Telecommunications, Tunisia, 2005
5. F.Davoine and S.Pateux. "Tatouage de documents audiovisuels numériques". Livre édition Hermès Lavoisier, traité IC2 Série Traitement du signal et de l'image, (2004).
6. Patrick Bas and François Cayre, Natural Watermarking: a secure spread spectrum technique for WOA Information Hiding", pp.1-14, 4437, 2006
7. Coatrieux G.,Puentes J.,Lecornu L., Roux. C., "Compliant secured specialized electronic patient record platform". In D2H2'00, Proceedings of D2H2, Washington, Etats-Unis, 2006
8. P. Bas, F. Cayre, B. Mathon. "Techniques sûres de tatouage pour l'image". Compression et Représentation des Signaux Audiovisuels, Montpellier, France, 2007
9. N. Palluat. "Méthodologie de surveillance dynamique à l'aide des réseaux neuro-fous temporels". PhD Thesis, University of Franche-Comte, 2006
10. O. Cadet. "Méthodes d'ondelettes pour la segmentation d'images : Applications à l'imagerie médicale et au tatouage d'images". PhD Thesis at the Polytechnic Institute of Grenoble, 2004
11. S. Pereira. "Robust Digital Image Watermarking". PhD Thesis, University of Geneva, 2000
12. B. Mathon, P. Bas, F. Cayre. "Practical performance analysis of secure modulations for WOA spread-spectrum based image watermarking". Multimedia and Security Workshop, Dallas, Texas, USA, 2007
13. L. Lin, I. J. Cox and G. Doerr. "An efficient algorithm for informed embedding of dirty-paper trellis codes for watermarking". IEEE Int. Conference on Image Processing, 2005
14. M. Vincent. "Contribution des filtres LPTV et des techniques d'interpolation au tatouage numérique". PhD Thesis presented in Toulouse, 2006
15. A. Westfeld. "A Regression-Based Restoration Technique for Automated Watermark Removal". Multimedia & Security ACM Workshop MMSEC, Oxford, United Kingdom 2008
16. C. REY. "Tatouage d'image: Gain en robustesse et intégrité des images". PhD Thesis, University of Avignon and Pays de Vaucluse, 2003
17. G. Coatrieux M. Lamard, W.J. Puente, and C. Roux "A low distortion and reversible watermark: Application to angiographic images of the retina". In EMBC'05, Proceedings of Int. Conference of the IEEE-EMBS, p. 2224-2227, Shangai, Chine, 2005
18. A. Zaid. "Compression et tatouage d'images à des fins d'archivage et de transmission : application aux images médicale". Habilitation University El Manar, Tunisia, April 2009
19. J.Reichel and M.Kunt. "Performance comparison of masking models based on a new psychovisual test method with natural scenery stimuli". Signal processing: Image communication, 17:807-823, 2002
20. M.Kutter and S.Winkler. "A Vision-based Masking Model for Spread-Spectrum Image Watermarking". IEEE Transactions on Image processing, 1(2): 244-250, 2002
21. N.sakr et al. "Copyright protection of learning using wavelet objects using wavelet-based watermarking and fuzzy logic, 3rd annual e-learning conference on Intelligent Interactive learning Object". Repositories Montreal. Quebec, Canada, 2006
22. J.Zhao, E.Koch. "Towards robust and hidden image copyright labelling". IEEE Workshop on Nonlinear Signal and Image Processing, 1995

23. S. Geetha, Siva S. Sivatha Sindhu, N. Kamaraj. "Close Color Pair Signature Ensemble Adaptive Threshold based Steganalysis for LSB Embedding in Digital Images". Transactions on Data Privacy, pp. 140 – 161, 2008
24. M. Barni, F. Bartolini, V. Cappellini and A. Piva, "A DCT domain system for robust image watermarking". Signal Processing, vol. 66, no. 3, pp. 357-372, 1998
25. G. Coatrieux, H. Maitre. "Images médicales, sécurité et tatouage". Annales des Télécommunications, Numéro Spécial Santé, vol. 58, p. 782-800, 2003

Invariant Recognition of Rectangular Biscuits with Fuzzy Moment Descriptors, Flawed Pieces Detection

Pulivarthi Srinivasa Rao

*Electronics & Communication Department
Andhra Polytechnic
Kaninada, 533003, India*

srinivasp08@gmail.com

Sheli Sinha Chaudhuri

*Electronics & Telecommunication Department
Jadavpur University
Kolkata, 700032, India*

shelisc@yahoo.co.in

Romesh Laishram

*Electronics & Communication Department
Manipur Institute of Technology
Takyelpat, 795001, India*

romeshlaishram@gmail.com

Abstract

In this paper a new approach for invariant recognition of broken rectangular biscuits is proposed using fuzzy membership-distance products, called fuzzy moment descriptors. The existing methods for recognition of flawed rectangular biscuits are mostly based on Hough transform. However these methods are prone to error due to noise and/or variation in illumination. Fuzzy moment descriptors are less sensitive to noise thus making it an effective approach and invariant to the above stray external disturbances. Further, the normalization and sorting of the moment vectors make it a size and rotation invariant recognition process. In earlier studies fuzzy moment descriptors has successfully been applied in image matching problem. In this paper the algorithm is applied in recognition of flawed and non-flawed rectangular biscuits. In general the proposed algorithm has potential applications in industrial quality control.

Keywords: Fuzzy moment descriptors, Euclidean distance, Edge detection, Flawed biscuits detection.

1. INTRODUCTION

The problem of pattern recognition has been studied extensively in different literatures in different domain. Invariant pattern recognition in Hough space [1]-[4] has already been addressed by Krishnapuram et al [5], and by Sinha et al [6], none of these two works includes the treatment of size-scaling. Authors in [5] and [6] use convolution in 9-space to achieve the rotational registration between sample objects and the templates, an additional processing is then necessary to determine the translational correspondence. The use of a pre-defined standard position in Hough space along with Distance-Discriminator Neural Neurons to achieve invariant pattern recognition has been addressed by Montenegro et al [7]. In that the distance between a template and a sample vector gives the corresponding likeliness degree. If several objects appear simultaneously in image space a preprocessing is necessary to single out individual objects by some of the broadly known labeling [8] technique. The authors in [9] have applied Hough transform in contour identification for object identification. In [10]-[11], the author proposed the

geometric transformation invariant pattern recognition in Hough space for polygonal shaped objects. In [12] the author proposed the algorithm for flawed biscuits detection in Hough space. In their method the recognition process is done entirely in the Hough space. The same algorithm has been applied in recognition of rectangular chocolates [13] and metallic corners-fasteners [14]. In [15] the authors proposed a technique for extracting rectangular shape objects from real color images using a combination of global contour based line segment detection and a Markov random field (MRF) model.

In this paper we proposed rectangular broken biscuit detection problem using Fuzzy moment descriptor. Fuzzy logic [16] which has proved itself successful in matching of inexact data can equally be used for inexact matching of close image attributes. In [17] the authors introduced the concept of fuzzy moment descriptor technique in invariant recognition of gray image and successfully applied in face recognition problem. The authors used the Euclidean distance of a reference image and the test images as the metric for recognition. The test image that has least Euclidean distance is taken as the best matched image. The authors in [18] proposed a new method for image registration based on Nonsubsampled Contourlet Transform (NCST) for feature extraction and the matching of a test image with a reference image is achieved using Zernike moments. This approach can be extended for shape matching and object classification. However our present work is mainly focused on the used Fuzzy moment descriptor proposed in [17].

In this technique an input gray image has been partitioned into n^2 non overlapped blocks of equal dimensions. Blocks containing region of edge is then identified. The degree of membership of a given block to contain edges is measured subsequently with the help of a few pre-estimated image parameters like average gradient, variance and the difference of the maximum and the minimum of gradients. Fuzzy moment which informally means the membership-distance product of a block $b[i, w]$ with respect to a block $b[j, k]$, is computed for all $1 \leq i, w, j, k \leq n$. A feature called 'sum of fuzzy moments' that keeps track of the image characteristics and their relative distances is used as image descriptor. We used Euclidean distance measure to determine the distance between the image descriptors of a reference image and a test image. Then we set a threshold value of the Euclidean distance for separating flawed and non-flawed rectangular biscuits.

The remainder of the paper is outlined as follows. Section 2 describes the estimation of fuzzy membership distribution of a given block to contain edge. A method for computing the fuzzy moments and a method for constructing the image descriptors are presented in section 3. The algorithm for the invariant recognition of rectangular biscuits is described in section 4. The simulation result is presented in section 5 and section 6 concludes the paper.

2. IMAGE FEATURES AND THEIR MEMBERSHIP DISTRIBUTIONS

This section describes briefly the image features used in the proposed technique and their fuzzy membership distributions.

Definition 2.1 An edge is a contour of pixels of large gradient with respect to its neighbors in an image.

Definition 2.2 Fuzzy membership distribution $\mu_Y(x)$ denotes the degree of membership of a variable x to belong to Y , where Y is a subset of a universal set U .

Definition 2.3 The gradient (Gonzalez and Wintz, 1993) at a pixel (x, y) in an image is estimated by taking the square root of the sum of difference of gray levels of the neighboring pixels with respect to pixel (x, y) .

Definition 2.4 The gradient difference (G_{diff}) within a partitioned block is defined as the difference of maximum and minimum gradient values in that block.

Definition 2.5 The gradient average (G_{avg}) within a block is defined as the average of the gradient of all pixels within that block.

Definition 2.6 The variance (σ^2) of gradient is defined as the arithmetic mean of square of deviation from mean. It is expressed formally as

$$\sigma^2 = \sum (G - G_{avg})^2 P(G)$$

Where G denotes the gradient values at pixels, and P(G) represents the probability of the particular gradient G in that block.

2.1 Fuzzy Membership Distributions

Once the features of the partitioned blocks in an image are estimated following the above definitions, the same features may be used for the estimation of membership value of a block containing Edge.

Consider when a block contains edge, the gradient in such blocks will have non-zero values only on the edges. So there must be a small positive average gradient of the pixels in these blocks. Consequently, σ^2 should be a small positive number. We may thus generalize that when σ^2 is close to zero but positive, the membership of a block b_{ij} to contain edge is high, and low otherwise, based on these intuitive assumptions, we presumed the membership curves $\mu(b_{jk}) = 1 - \exp(-bx^2)$, $b > 0$. The exact value of b can be evaluated by employing genetic algorithms (Man et al., 1999). It may be noted that $1 - \exp(-bx^2)$ has a zero value at $x = \sigma^2 = 0$ and approaches unity as $x = \sigma^2 \rightarrow \infty$. The square of x, (x^2) represents the faster rate of growth of the membership curve $1 - \exp(-bx^2)$.

Analogously, the average gradient G_{avg} of a block b_{ij} containing only edge is large. Thus, the membership of a block b_{ij} to contain edge will be high, when G_{avg} is large. This phenomena can be represented by the membership function $1 - \exp(-\alpha x)$, $\alpha > 0$. The selection of the membership function for edge w.r.t. G_{diff} directly follows from the previous discussion.

TABLE 1: Membership functions for features

Parameter	Edge membership
G_{avg}	$1 - e^{-bx}$
G_{diff}	$1 - \exp(-bx^2)$ $b > 0$
σ^2	$1 - \exp(-bx^2)$ $b > 0$

The membership values of a block $b [j,k]$ containing edge can be easily estimated if the parameters and the membership curves are known. The fuzzy production rules, described below, are subsequently used to estimate the degree of membership of a block $b [j, k]$ to contain edge by taking into account the effect of the parameters.

2.2 Fuzzy production rules

A fuzzy production rule is an if-then relationship representing a piece of knowledge in a given problem domain. For the estimation of fuzzy memberships of a block $b [j, k]$ to contain, say, edge, we need to obtain the composite membership value from their individual parametric values. The if-then rules represent logical mapping functions from the individual parametric memberships to the composite membership of a block containing edge. The production rule PR1 is a typical example of the above concept.

PR1: IF $(G_{avg} > 0.142)$ AND
 $(G_{diff} > 0.707)$ AND
 $(\sigma^2 \approx 1.0)$
 THEN (the block contains edges).

Let us assume that the G_{avg} , G_{diff} and σ^2 for a given partitioned block have found to be 0.149, 0.8 and 0.9, respectively. The $\mu_{edge}(b_{jk})$ now can be estimated first by obtaining the membership values $\mu_{edge}(b_{jk})$ w.r.t. G_{avg} , G_{diff} and σ^2 , respectively by consulting the membership curves and then by applying the fuzzy AND (minimum) operator over these membership values. The single valued membership, thus obtained, describes the degree of membership of the block $b[j,k]$ to contain edge.

3. FUZZY MOMENT DESCRIPTORS

In this section we define fuzzy moments and evaluate image descriptors based on those moments. A few definitions, which will be required to understand the concept, are defined below.

Definition 3.1. Fuzzy edge moment $\left[M_{iw}^{jk} \right]_{edge}$ is estimated by taking the product of the

membership value $\mu_{edge}(b_{jk})$ (of containing edge in the block $b[j,k]$) and normalized Euclidean distance $d_{iw,jk}$ of the block $b[j,k]$ w.r.t. $b[i,w]$. Formally,

$$\left[M_{iw}^{jk} \right]_{edge} = d_{iw,jk} \times \mu_{edge}(b_{jk}) \quad (3.1)$$

Definition 3.2. The fuzzy sum of moments (FSM), for edge S_{iw} , w.r.t. block $b[i,w]$ is defined as the sum of edge moments of the blocks where edge membership is the highest among all other membership values.

Formally,

$$S_{iw} = \sum_{\exists jk} d_{iw,jk} \times \mu_{edge}(b_{jk}) \quad (3.2)$$

Where $\mu_{edge}(b_{jk}) \geq \text{Max} \left[\mu_x(b_{jk}) \right]$, $x \in \text{set of features}$

After the estimation of the fuzzy membership values for edges, the predominant membership value for each block and the predominant feature are saved. The FSM with respect to the predominant features are evaluated for each block in the image. Each set of FSM is stored in a one-dimensional array and is sorted in a descending order. These sorted vectors are used as descriptors for the image.

For matching a reference image with a set of input images, one has to estimate the image descriptors for the reference and the input images. For our recognition process we used a non broken biscuit as the reference image. The matching of images requires estimation of Euclidean distance between the reference image with respect to all other input images. The measure of the distance between descriptors of two images is evident from Definition 3.3.

Definition 3.3 The Euclidean distance $[E_{i,j}]_k$, between the corresponding two k th sorted FSM descriptor vectors V_i and V_j of two images I and J of respective dimensions $(n \times 1)$ and $(m \times 1)$ is estimated first by ignoring the last $(n - m)$ elements of the first array, where $n > m$ and then taking the sum of square of the elemental differences of the second array with respect to the modified first array having m elements.

Definition 3.4 The measure of distance between two images, hereafter called image distance, is estimated by taking exhaustively the Euclidean distance between each of the two similar descriptor vectors of the two images and then by taking the weighted sum of these Euclidean distances.

Formally, the distance D_{ry} between a reference image r and an image y is defined as

$$D_{ry} = \sum \beta_k \times [E_{ij}]_k \quad (3.3)$$

Where the suffix i and j in $[E_{ij}]_k$ correspond to the set of vectors V_i for image r and V_j for image y .

In our technique of flawed pieces detection of rectangular biscuits we estimate the image distance D_{ry} where $y \in$ the set of input images and r denotes the reference image. The image for which the image distance D_{ry} is larger than a predefined threshold belongs to flawed pieces otherwise the biscuit is non-broken.

4. ALGORITHM

In this section the algorithm used in invariant recognition of rectangular biscuits is presented. The algorithm requires a reference image which is a non-broken biscuit image and a set of input images for recognition process. Here all the images to the algorithm are supplied after the images pass through an edge detector. For recognition of rectangular biscuits only the edge information of the image is sufficient. The edge detection may be performed using Sobel edge detector or Canny edge detector. For our problem we have used a simple Sobel edge detector.

The major steps for the process is as depicted in figure 4.1

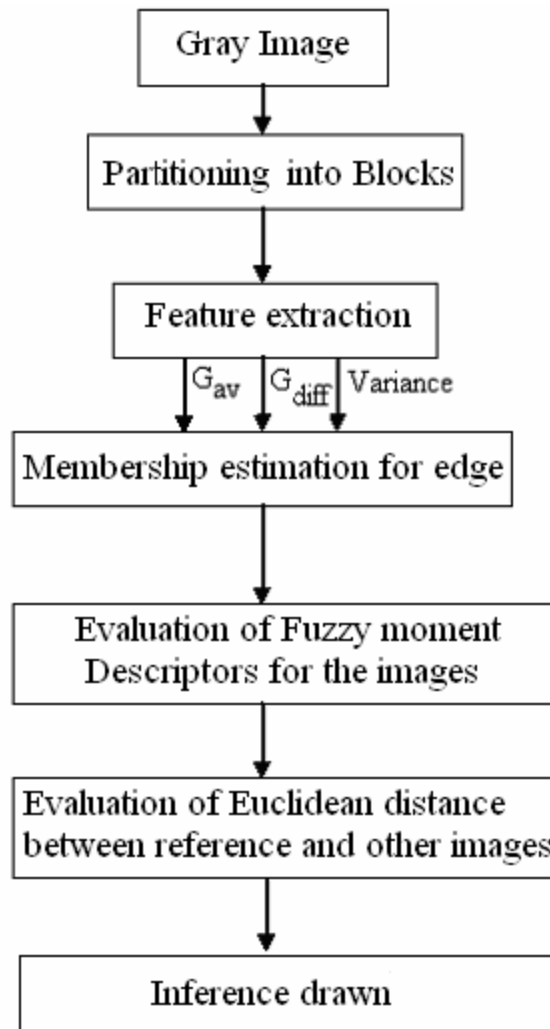


Figure 1: Basic steps involved in the recognition process

In order to keep the matching process free from size and rotational variance, the following strategies have been adopted.

- I. Euclidean distances between each two blocks of an image used for estimation of fuzzy moments are normalized with respect to the image itself.
- II. the descriptor vectors are sorted so as to keep the blocks with most predominant features at the beginning of the array, which only participate subsequently in the matching process is free from rotational variance of the reference image

5. SIMULATION RESULTS

To study the effectiveness of the proposed algorithm we consider computer generated images as shown in figure 2. These computer synthesized images directly follows from [12]. In the figure 2, the image ref is the reference image which is a non-broken rectangular biscuit and the images r2 – r12 are taken as the input test images

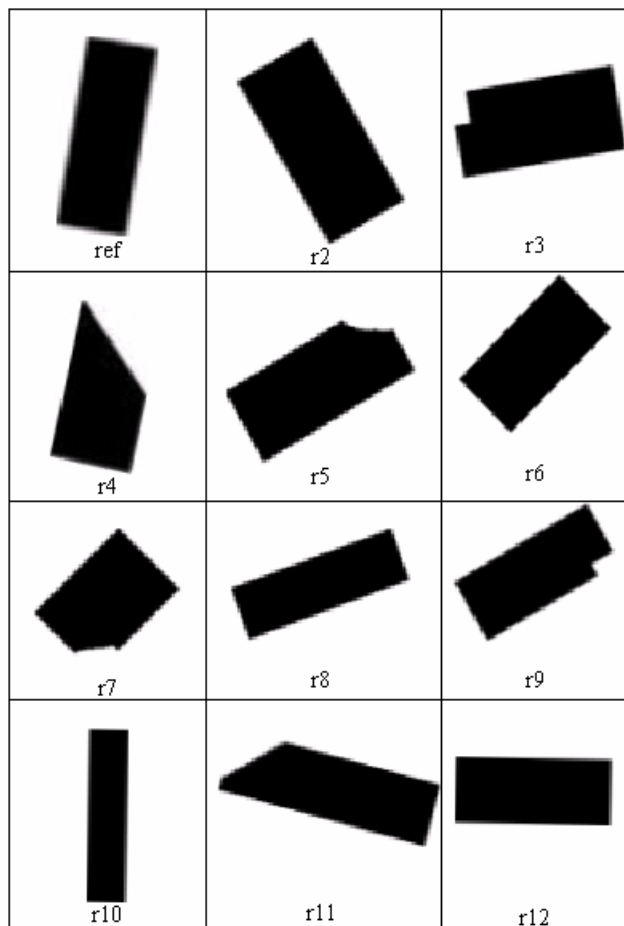


Figure 2: Computer generated biscuit images

Table 2: Euclidean Distance between the reference & the test input images

Test images	Euclidean distance
r2	0.244295
r3	1.850987
r4	3.645934
r5	1.711132
r6	0.462434
r7	2.635301
r8	0.254587
r9	1.538170
r10	0.430162
r11	1.517850
r12	0.142401

From table 2 we can observe that if the test input images are unbroken biscuits then the Euclidean distance between the input images and the reference image are small whereas for broken biscuits (flawed pieces) the Euclidean distance is high. Therefore using the Euclidean distance as the metric we can separate broken rectangular biscuits from unbroken rectangular biscuits. We can draw an inference from the Euclidean metric that, if the Euclidean distance is greater than a predefined threshold the biscuit is a broken biscuits otherwise unbroken. The choice of the threshold is very critical in the successful detection of broken biscuits. There may be many methods for designing the threshold value. Presently we choose the threshold intuitively based on the database in performing the simulations. The test input images r2, r6, r8, r10 and r12 are unbroken biscuits of different size and orientations (figure 2) while the remaining biscuits are flawed pieces. It can be observed from table 2 that for unbroken biscuits the Euclidean distance is below 1 and for flawed pieces it is greater than 1. For simulation using the above set of images we choose the value of the Euclidean distance threshold to be equal to 1. For set the of test input images this threshold perfectly works. However this may not be an ideal way of designing the threshold. But if the database is very large then the intuitive way of choosing threshold may produce result with less error. The results may be less satisfactory compared to the method proposed in [12] however their algorithm will fail to recognize the image r10 and r12 in figure 2 as it cannot be applied to Horizontally and vertically align images. Our proposed algorithm is more robust compared to algorithm based on Hough transform[12].

6. CONCLUSION

In this paper we have shown that fuzzy moment descriptor can be used in invariant recognition of rectangular biscuits with detection of flawed biscuits pieces detection. The proposed technique is not limited to only rectangular objects; it can be applied to appropriate pattern recognition problem also. The proposed technique is less sensitive to external noise as compared to the technique using Hough space. But the intuitive way of choosing the threshold may not be an ideal. The use of neural networks or any optimization technique in choosing the threshold is considered as future aspect in continuation of this work.

7. REFERENCES

1. J. Illingworth and J. Kittler. "A survey of the Hough transform", CVGIP, vol. 44, pages 87-116.1988
2. P.V.C Hough. "A method and means for recognizing complex patterns", U.S. patent 3,069,654. (1962)
3. R. O. Duda, and P.E. Han. "Use of the Hough transform to detect lines and curves in pictures", Communications of the ACM. Vol. 15, No 1, 11-15, 1972

4. R Gonzalez. and R. E. Woods, "Digital image processing", Addison-Wesley Publishing, (1992)
5. R. Krishnapuram. and D. Casasent, "Hough space transformation for discrimination and distortion estimation", CVGIP, vol. 38, PP 299-316.1987
6. P.K. Sinha, F.Y. Chen. R.E.N. Horne, "Recognition and location of shapes in the Hough pattern space", IEE Elect. Div. Colloq. on Hough transforms 19931106. Savoy place, London, 1993
7. J. Montenegro Joo, L. da F. Coata and R. KBberle, "Geometric transformation invariant pattern recognition with Hough transforms and distance discriminator neural networks", presented at the Workshop sobre Computapio de Alto Descmpenho pam Processamenfo de Sinars, Si0 Carlos, SP, Brazil, 1993.
8. R. Scha1koff. "Digital image processing and computer vision", John Wiley & Sons Inc. (1989)
9. G. Gesig G. and F. Klein., "Fast contour identification through efficient Hough transform and simplified interpretation strategy", UCPR-8. Paris. 498.500. (1986).
10. J. Montenegro Joo. "A Polar-Hough-Transform Based Algorithm for the Translation, Orientation and Size-Scale Invariant Pattern Recognition of Polygonal Objects", UMI Dissertations I.D03769, 1998.
11. J. Montenegro Joo. "Geometric-Transformations Invariant Pattern Recognition in the Hough Space", Doctoral Degree Project. Cybernetic Vision Research Group, Instituto de Ffsica tic Sao Carlos (IFSC), Dpto. de Fisica c Informtica. Universidade de San Paulo (USP). San Carles, SP. Brazil. (August 1994).
12. J. Montenegro Joo. "Invariant recognition of rectangular biscuits through an algorithm operating exclusively in Hough Space. Flawed pieces detection". Revista de Investigacion de Fisica, Vol 5, No 1,2 ,2002
13. . Montenegro Joo. "Hough Transform based Algorithm for the automatic Invariant Recognition of Rectangular Chocolates, Detection of defective pieces", Industrial Data, Lima, Peru.Vol. 9, No 2, 2006
14. J. Montenegro Joo, (2007). "Hough-transform based automatic invariant recognition of metallic corner-fasteners", SISTEMAS E INFO RMA TICA. Industrial Data, Lima, Peru. Vol 10, No 1, 2007
15. Yangxing Liu, Ikenaga. T, Goto.S. "A Novel Approach of Rectangular Shape Object Detection in Color Images Based on an MRF Model" In Proceedings of 5th IEEE International Conference on Cognitive Informatics, 2006
16. Amit Konar. "Artificial Intelligence and soft computing: Behavioral and Cognitive modeling of the Human brain", CRC press (2000)
17. Biswas B., Konar A., Mukherjee A.K. "Image matching with fuzzy moment descriptors". Engineering Applications of Artificial Intelligence, 14 (1), pp. 43-49, 2001
18. Jignesh Sarvaiya, Suprava Patnaik & Hemant Goklani." Image Registration using NSCT and Invariant Moment", International Journal of Image Processing ISSN (1985-2304) Vol. (4),issue (2),pp 119-130,May 2010.

Genetic Algorithm Processor for Image Noise Filtering Using Evolvable Hardware

Dr. K. Sri Rama Krishna

*Professor & HOD /ECE
V R Siddartha Engg College
Vijayawada, Krishna (dist), A.P-520007, India*

srk_kalva@yahoo.com

A. Guruva Reddy

*Professor / ECE
NRI institute of Technology
Pothavarappadu , Agiripalli, Krishna (D), A.P-521 212, INDIA*

guruvarreddy78@mail.com.com

Dr. M.N.GIRI PRASAD

*Principal,
JNTU college of Engineering,
Pulivendala, Kadapa (dist). A.P – 516390, India*

mahendran_gp@rediffmail.com

Dr. K.Chandrabushan Rao

*Professor/ECE
MVGR College of Engineering,
Vijayanagarm, A.P, India*

cbraokota@yahoo.com

M. Madhavi

*Associatet Prof / ECE
NRI institute of Technology
Pothavarappadu , Agiripalli, Krishna (D), A.P-521 212, INDIA*

madhavi.418@gmail.com

Abstract

General-purpose image filters lack the flexibility and adaptability of un-modeled noise types. On the contrary, evolutionary algorithm based filter architectures seem to be very promising due to their capability of providing solutions to hard design problems. Through this novel approach, it is made possible to have an image filter that can employ a completely different design style that is performed by an evolutionary algorithm. In this context, an evolutionary algorithm based filter is designed in this paper with the kernel or the whole circuit for automatically evolved.

The Evolvable Hard Ware architecture proposed in this paper can evolve filters without a priori information. The proposed filter architecture considers spatial domain approach and uses the overlapping window to filter the signal. The approach that is chosen in this work is based on functional level evolution whose architecture includes nonlinear functions and uses genetic algorithm for finding the best filter configuration.

Keywords: Reconfigurable hardware, Processing elements, Genetic algorithm, Virtual Reconfigurable Circuit

1. INTRODUCTION

Non-linear Image processing [13] with good flexibility and adaptability is highly desirable in several applications such as image transformation, correction of distortion effects, noise removal, histogram equalization etc. Conventional adaptive filter lacks the flexibility for adapting to changes in architecture and is therefore suitable for compensating non-uniform variations in Signal-to-Noise Ratio (SNR). It is also reported that conventional adaptive filter performs well, only when, the spatial density of the noise is less. In this paper, a reconfigurable computing FPGA architecture is designed for adapting to the changes in task requirements or changes in environment, through its ability to reconfigure its own hardware structure dynamically and autonomously with design objectives such as high performance, specialization and adaptability.

In this paper, minimally sufficient hardware resources are dynamically allocated based on noise levels at specific time intervals. To enable automatic recovery of a device after damage, an autonomous restructuring algorithm to handle internal processing element fault is implemented and tested successfully. The novelty of this work is the implementation of a reconfigurable system [1] and an associated design methodology that has both flexibility and autonomous restructuring [3] of Processing Elements. The reconfiguration process is achieved using an evolutionary algorithm [2] such as Genetic Algorithm (GA).

2. IMAGE ENHANCEMENT USING EVOLUTIONARY DESIGN

The Evolvable Hard Ware (EHW) architecture [6] proposed in this work for filtering the noise present in the image that is subsequently realized on an Field Programmable Gate Array (FPGA) based image processing board, consists of the GA processor [11] and a virtual reconfigurable circuit and is shown in Figure 1. This type of implementation integrates a hardware realization of genetic algorithm and a reconfigurable device. These two modules of the EHW [5] are described in the following sections:

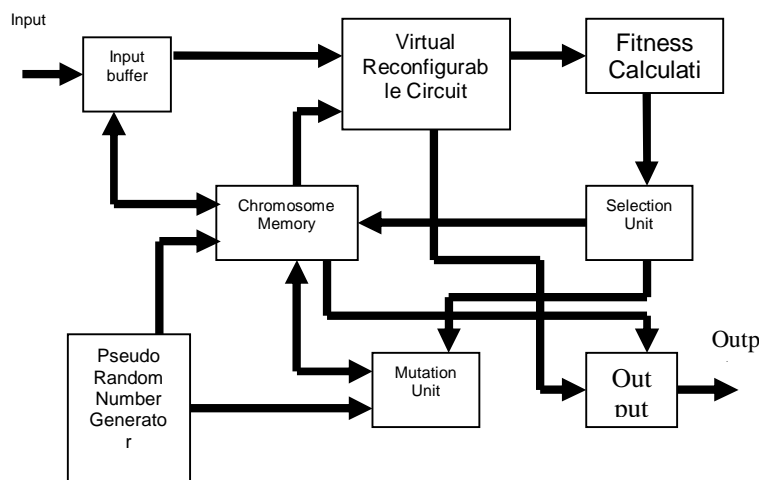


FIGURE 1: EHW chip with the VRC and GA Processor

2.1 Implementation of the GA Processor

The implementation of simple GA [12] is composed of following basic modules: pseudo random number generator, population memory, selection unit, mutation unit, fitness evaluator and output buffer. These modules have already been discussed in chapter II of this thesis.

2.2 Implementing the Virtual Reconfigurable Circuit

The Virtual Reconfigurable Circuit (VRC) is implemented [10] as a combinational circuit using the concepts of pipelining. It consists of processing elements arranged in rows and columns. In this work, a total of 25 PE's are selected and are arranged in six rows and four columns with the 25th PE representing the final output.

2.3 Architecture of The VRC

The architecture of the VRC is shown in Figure 2. I_4 represents the filtered output from the VRC. The nature of operation performed on the input depends on the configuration bits downloaded into the configurable memory from the genetic unit.

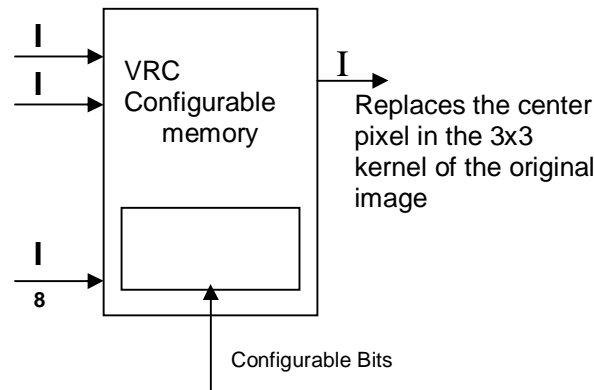


FIGURE 2: Architecture of the VRC

3. GENETIC ALGORITHM FOR EVOLVABLE HARDWARE

The configuration bits are obtained using genetic algorithm and are downloaded into the reconfiguration circuit which results in relocation of hardware modules inside VRC. The flow diagram of reconfiguration process is shown in figure 3.

The class of evolutionary algorithm most commonly used in evolvable hardware [4] is the genetic algorithm. Most commonly these operate on a fixed size population of fixed length binary strings called chromosomes. Each chromosome encodes a common set of parameters that describe a collection of electronic components and their interconnections, thus each set of parameter values represents an electronic circuit. The set of all possible combinations of parameter values defines the search space of the algorithm, and the circuits that they represent define the solution space of the algorithm.

The Evolvable hardware with Genetic algorithm [7] begins by initialising the bits of each chromosome with random values. The chromosomes are then evaluated in turn by creating a circuit based on the parameter values, either as a simulated model of the circuit or as a concrete circuit embodied in reconfigurable hardware. The circuit's fitness for performing the target task is then measured by passing it a set of test values and evaluating the veracity of the circuit's output. The selection operator then probabilistically populates the next generation of chromosomes such that chromosomes with high fitness are more likely to be selected. The operator selects two individuals at random and compares their fitness. Only the individual with the highest fitness is inserted into the next generation. If they have equal fitness the individual to be inserted is chosen at random. Once the new population has been selected, it is varied. Common variation operators are one-point crossover and point mutation. One point crossover recombines two chromosomes by choosing a position at random along the chromosome and swapping every bit beyond this point between the strings. It is stochastically applied according to a fixed probability. Point mutation independently inverts each bit in the chromosome according to a fixed probability. These operators are applied to all members of the new population. Often in addition to these operators, the best member of the original population is copied into the new population unchanged, The new population is now complete and the algorithm then iterates the steps of

evaluation, selection and variation until a circuit that functions adequately is found, or a pre-specified number of generations is completed.

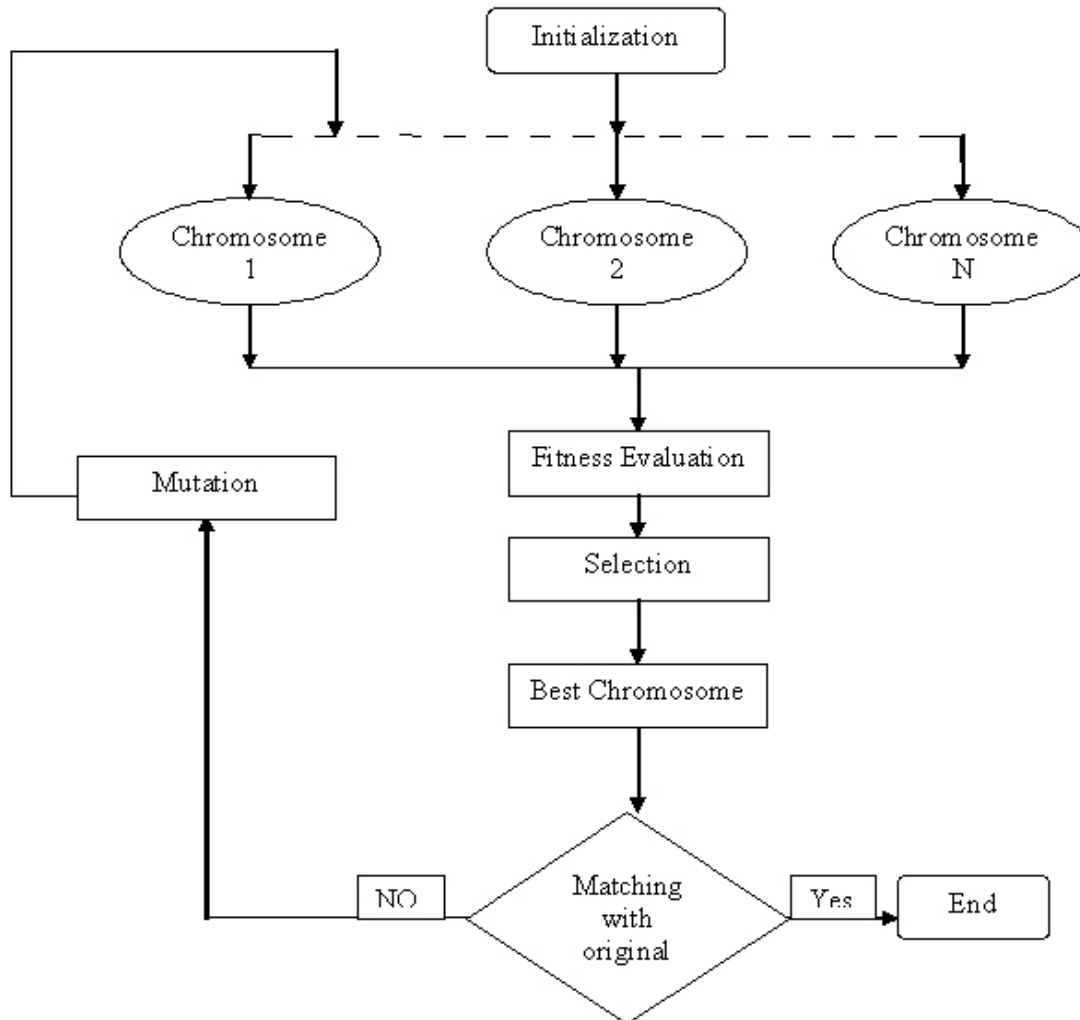


FIGURE 3: Flow diagram of VRC using Genetic Algorithm.

The configuration bits are obtained using genetic algorithm [12] and are downloaded into the reconfiguration circuit which results in relocation of hardware modules inside VRC.

4. COMPONENTS OF VIRTUAL RECONFIGURABLE CIRCUIT

In reconfigurable computing sequence of configurations is not known at the design time. It involves the usage of software and hardware components normally referred to as IP cores. This concept is very much useful in effective designing of the complex systems. The reconfigurable device is part of the evolvable component. The genetic unit in the evolvable component generates the configuration bits (Chromosomes). The fitness calculation is usually performed outside the evolvable component. Internal reconfiguration implies that a CLB can configure other CLB's.

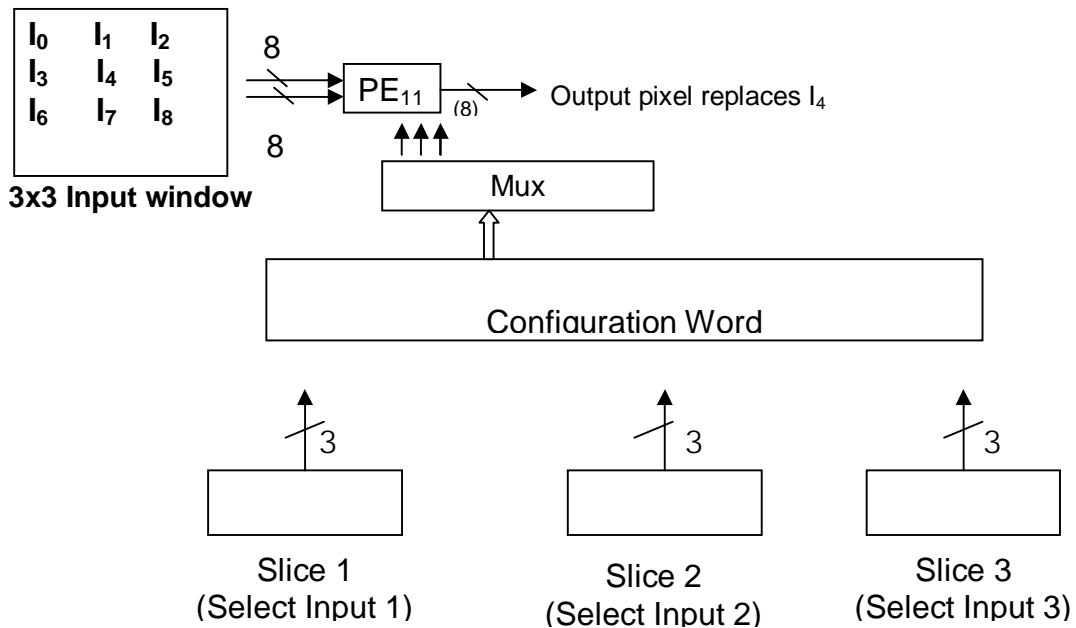


FIGURE 4: VRC with the internal MUX for selecting inputs and functions Helvetica

In figure 4. Slice 1 is a 'm' bit vector and selects any one of many inputs and assigns it as first input X

Slice 2 is a 'm' bit vector and selects any one of many inputs and assigns it as second input Y

Slice 3 is a 'n' bit vector and selects any one of the 16 functions to be performed on X and Y

The X and Y are both 8-bit vectors and the processed output is also an 8 - bit vector.

5 IMAGE ENHANCEMENT ALGORITHM

5.1 Fitness Function

Popular measures of performance for evaluating the difference between the original and filtered images includes

- i. Peak Signal to Noise Ratio (PSNR) and
- ii. Mean Difference per Pixel (MDPP)

In many applications the error is expressed in terms of a signal-to-noise ratio (SNR), and is given in equation 1

$$SNR = 10 \log_{10} \frac{\sigma^2}{MSE} \text{ dB} \quad (1)$$

where σ^2 is the variance of the desired or original image. The peak signal-to-noise ratio (PSNR) is expressed as equation 2,

$$PSNR = 10 \log_{10} \frac{255^2}{MSE} \text{ dB} \quad (2)$$

The fitness function using MDPP is given by

$$MDPP = \frac{1}{N \times N} \sum_{i,j=1}^N |orig(i, j) - filt(i, j)| \quad (3)$$

where $|orig(i,j) - filt(i,j)|$ is the absolute difference between the original and filtered images. In this work, the Mean Difference per Pixel (MDPP) is used as a performance measure by the fitness evaluator module in the GA processor for the reason that MDPP fitness function is computationally easier for hardware implementation in comparison to PSNR. The EHW architecture that has the best MDPP (minimum MDPP) after a specified number of generations is chosen as the evolved architecture.

6 NOISE MODELS

6.1 Multiplicative Noise

In this model, the noise magnitude depends on the signal magnitude itself. An example of multiplicative noise is the degradation of film material caused by the finite size of silver grains used in photosensitive emulsion.

6.2 Quantization Noise

This occurs when insufficient quantization levels are used, for example, only 50 levels for a monochromatic image. Due to this noise false contours appear. However, this noise can be eliminated by simple means.

6.3 Impulsive Noise

When an image is corrupted with individual noisy pixels whose brightness differs significantly from that of the neighborhood, then it represents an impulse noise effect.

6.4 Salt and Pepper Noise

This describes [8] saturated impulsive noise – an image corrupted with white and/or black pixels. The effect is more appropriate for binary images.

6.5 Gaussian Noise

Idealized noise, called white noise or Gaussian noise, has constant power spectrum, meaning that its intensity does not vary with increasing frequency. This noise model is often used.

7. VHDL IMPLEMENTATION OF GA PROCESSOR

The model of GA processor is implemented in VHDL. The screen that captured the VHDL code from the Xilinx ISE software package [9] is given in Figure 5. Each one of the signal implements, the 12 bit random number generator and a sample output captured using the modelsim package, is shown in Figure 6. These bits (GA processor output) are the configuration bits which control the interconnections among the PE's as well as and also the functionality of each PE in the VRC module [10].

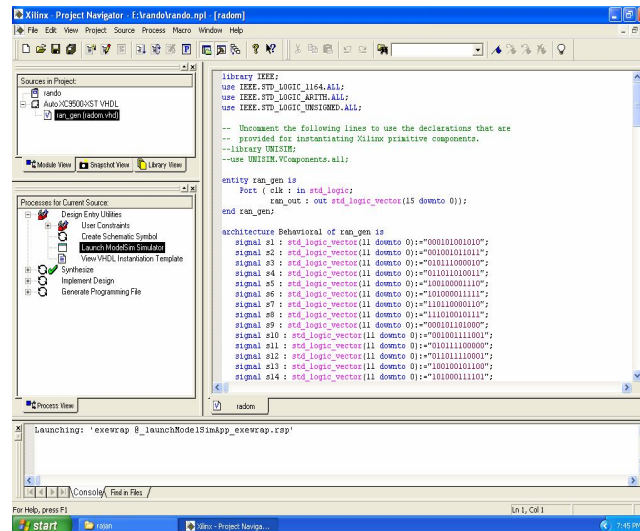


FIGURE 5: Implementation of GA processor using VHDL

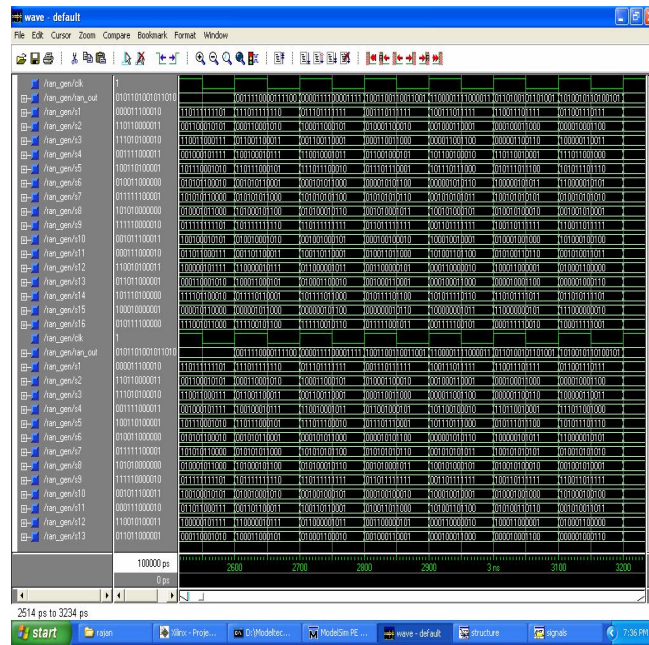


FIGURE 6: Modelsim Captured GA Processor output

8. IMPLEMENTATION RESULT OF EHW FILTER

In this section, the performance of EHW Filter in removing Gaussian noise is presented. The evolved EHW architecture for this case is shown in figure 7.

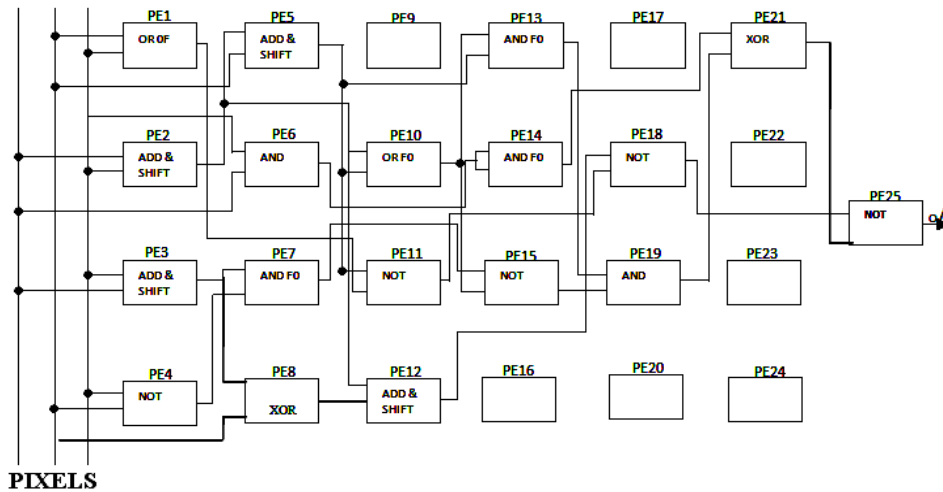


FIGURE 7: Architecture of VRC evolved to filter Gaussian Noise

The standard test image, Lena as shown in figure 8a is used for study and the results. In figure 8b shows Additive Gaussian noise with mean 0 and standard deviation 0.05 is added. The results obtained with Median filter and proposed EHW filter are shown in Figure 8c and 8d respectively. It is observed that Median Filter does not effectively remove Gaussian noise and performs only in edge regions, with blur effects still present in continuous regions. The EHW Filter performed well both in edge and continuous regions and effectively removed Gaussian noise.

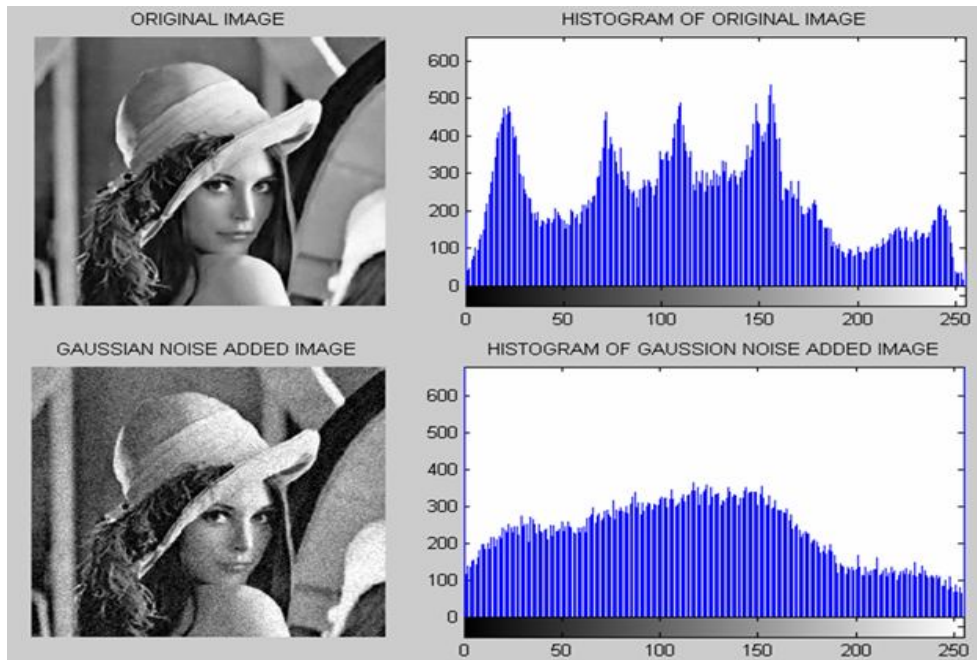


FIGURE 8 a: Original Image,
FIGURE 8 b: Gaussian noise added image

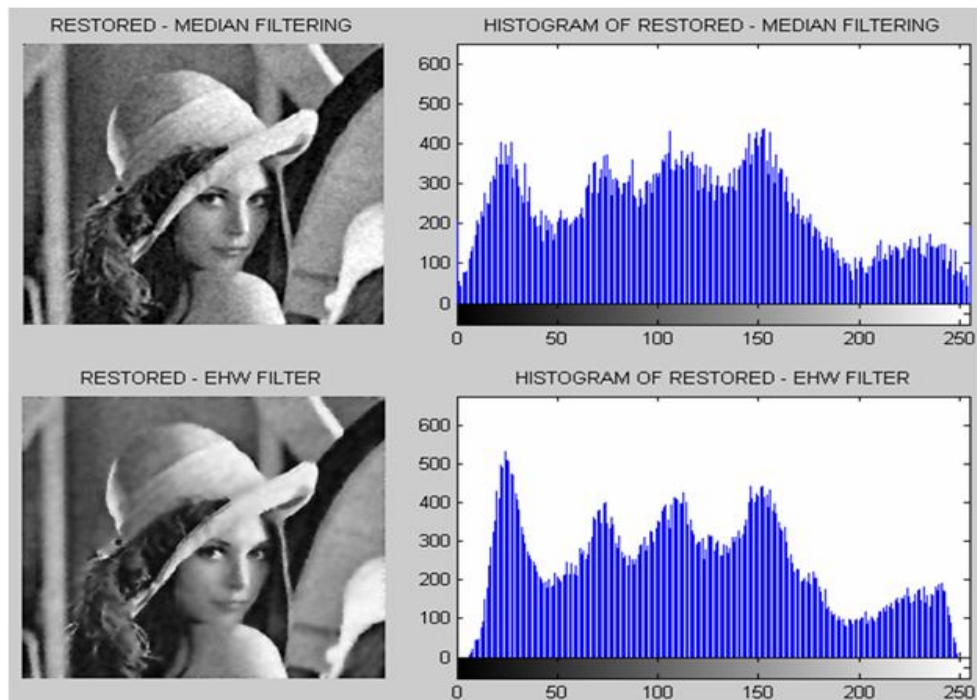


FIGURE 8 c: Restored image using Median
FIGURE 8 d: Restored image using EHW

The standard test image, Lena as shown in figure 9a. In figure 9b shows Salt And Pepper is added for Lena image. The results obtained with sobel filter and proposed EHW filter are shown in Figure 9c and 9d respectively. It is observed that sobel filter does not effectively remove salt

and pepper noise. The EHW Filter performed well both in edge and continuous regions and effectively removed salt and pepper noise.

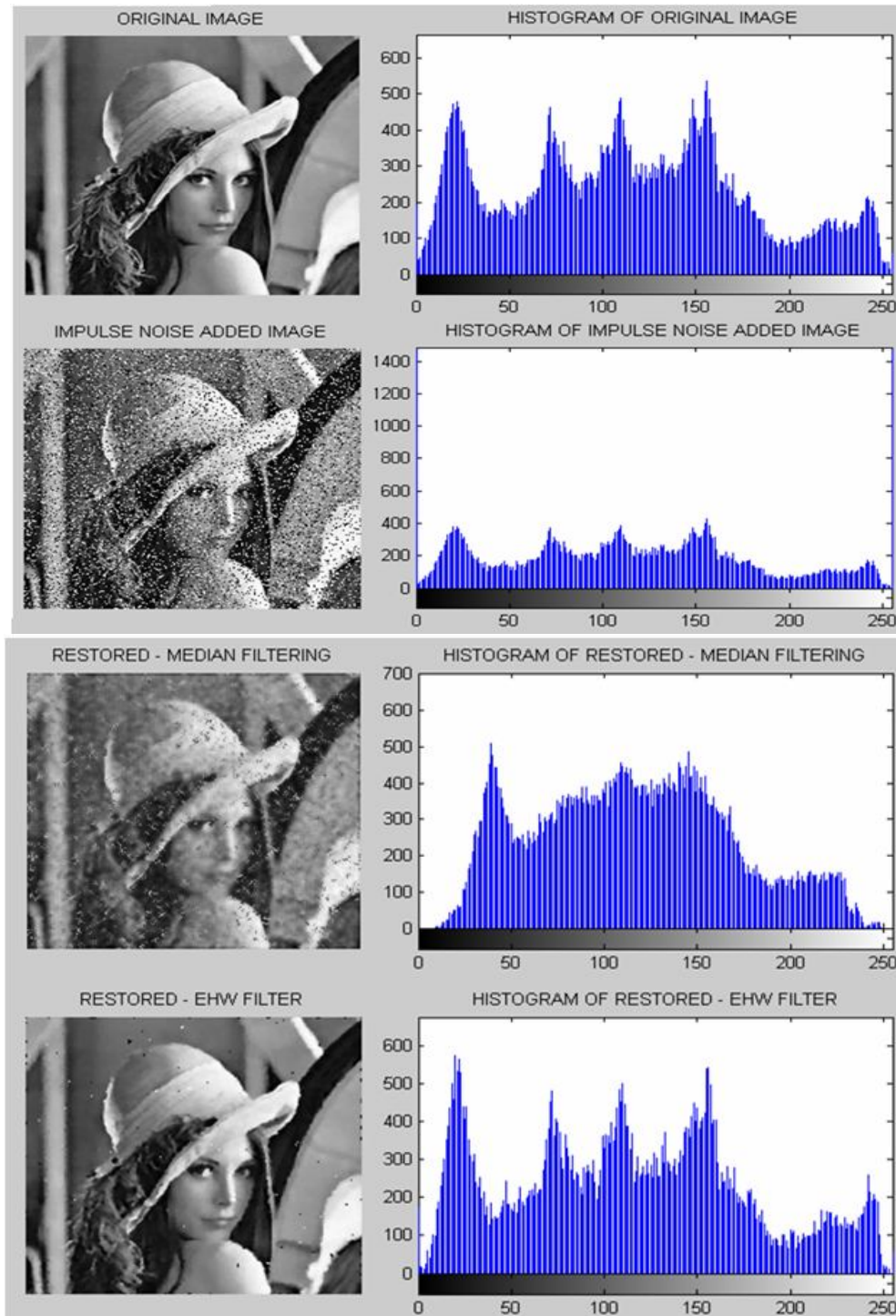


FIGURE 9 a: Original Image,
FIGURE 9 b: Salt and Pepper noise added image
FIGURE 9 c: Restored image using Sober filter
FIGURE 9 d: Restored image using EHW

The comparison results are given in Table 1. It is clear that EHW filter outperforms the other filters for both Gaussian and salt & pepper noise. The nonlinear part of EHW filter preserves the edges and removes the Gaussian noise effectively. The linear part smoothens the salt and pepper noise and removes the spurious part of the image.

Image	PSNR (dB)	MDPP	MSE
Gaussian Noise of SD= 0.05			
Median Filter	24.95	180606	23.08
EHW	27.05	143152	20.14
Salt and Pepper			
Sober Filter	24.17	173449	23.96
EHW	26.69	133399	19.44

TABLE 1: Comparison Study

9. CONSLUSION & FUTURE WORK

In this work, a novel virtual reconfigurable circuit based image filter was presented. It is shown that it is possible to successfully evolve noise removal filters that produce better image quality than a standard median and sober filter. The n-bit configuration information that determines the function of the functional block and the connection to the inputs is chosen optimally. Also, the design guarantees that, in this constrained structure, no random configuration information can ever destroy the chip. The fitness calculation part is the computationally time consuming operation and hence, in this work, this part is performed inside the FPGA along with the VRC. The utilized resources of the FPGA are also found to be practical and any commercially available FPGA can be used for implementing the EHW. If noise corrupting the image varies, a more accurate VRC filter is swapped into the FPGA hardware.

Future Work

To reduce the computational complexity and to enhance the speed, each kernel can be executed on a floating point processor. This can overcome the limitation of architectures which support only binary operations.

10. REFERENCES

1. Borgatti, M. Lertora, F. Foret, B. Cali, L. "A reconfigurable system featuring dynamically extensible embedded microprocessor, FPGA, and customizable I/O". IEEE Journal of Solid-State Circuits, Vol.38(3): 521- 529, Mar 2003
2. Christian Haubelt, Thomas Schlichter, and Jurgen Teich, "Improving Automatic Design Space Exploration by Integrating Symbolic Techniques into Multi-Objective Evolutionary Algorithms". International Journal of Computational Intelligence Research, Vol.2(3): 239–254, 2006

3. Guruva Reddy. A, Sri Rama Krishna. K, Giri Prasad. M.N and Chandra Bhushan Rao K, "Autonomously Restructured Fault tolerant image enhancement filter," ICGST International Journal on Graphics, Vision & Image Processing, Vol.08, issue-3, pp.35-40, Oct 2008.
4. Higuchi, T. Iwata, M. Keymeulen, D. Sakanashi, H. Murakawa, M. Kajitani, I. Takahashi, E. Toda, K. Salami, N. Kajihara, N. and Otsu, N, "*Real-world applications of analog and digital evolvable hardware*". IEEE Transactions on Evolutionary Computation, Vol.3(3): 220-235, Sep 1999
5. Higuchi .T, N. Kajihara, "*Evolvable Hardware Chips for Industrial Applications*". Communications of the ACM, Vol.42(4): 60-69, June 2006
6. Clark, G.R, San Diego and La Jolla. "*novel function-level EHW architecture within modern FPGAs*". In Proceedings of Congress on Evolutionary Computation (CEC 99), California Univ, CA, 1999
7. Higuchi.T, Umezono and Tsukuba Ibaraki. "*Evolvable hardware with Genetic learning*". In Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS '96), 1996
8. Jie Li and Shitan Huang. "*Adaptive Salt-&-Pepper Noise Removal: A Function Level Evolution based Approach*". In Proceedings of NASA/ESA Conference on Adaptive Hardware and Systems, June 2008
9. Kyrre Glette and Jim Torresen. "*A Flexible On-Chip Evolution System Implemented on a Xilinx Virtex-II Pro Device*". In Proceedings of International conference on Evolvable systems, 2005
10. Sekanina.L. "*Virtual Reconfigurable Circuits for Real-World Applications of Evolvable Hardware*". In Proceedings of Fifth International Conference on Evolvable Systems (ICES'03), 2003
11. Simon Harding. "*Evolution of Image Filters on Graphics Processor Units Using Cartesian Genetic Programming*". In Proceedings of IEEE Congress on Evolutionary Computation, 2008
12. Goldberg, D. E. "*Genetic Algorithms in Search, Optimization & Machine Learning*", Pearson Education, Inc, 1990.
13. Rafael C. Gonzalez and Richard E. Woods, "*Digital Image Processing*", Second edition, Pearson Education, 2007

Developing 3D Viewing Model from 2D Stereo Pair with its Occlusion Ratio

Himanshu Johari

Jaypee Institute of Information Technology
A-10, Sector-62 Noida-201 307
Uttar Pradesh, India.

himanshujohari@gmail.com

Vishal Kaushik

Jaypee Institute of Information Technology
A-10, Sector-62 Noida-201 307
Uttar Pradesh, India.

contact.vishal007@gmail.com

Pawan Kumar Upadhyay

Jaypee Institute of Information Technology
A-10, Sector-62 Noida-201 307
Uttar Pradesh, India

pawan.upadhyay@jiit.ac.in

Abstract

We intend to make a 3D model using a stereo pair of images by using a novel method of local matching in pixel domain for calculating horizontal disparities. We also find the occlusion ratio using the stereo pair followed by the use of The Edge Detection and Image Segmentation (EDISON) system, on one of the images, which provides a complete toolbox for discontinuity preserving filtering, segmentation and edge detection. Instead of assigning a disparity value to each pixel, a disparity plane is assigned to each segment. We then warp the segment disparities to the original image to get our final 3D viewing Model.

Keywords: 3D model, Stereo Pair, Depth Perception, Parallax Method, Occlusion, Disparity Map.

1. INTRODUCTION

3D models and 3D viewing is catching great pace in the field of computer vision due to its applicability in diverse fields of health, aerospace, textile etc. We in our paper intend to propose a simplistic and a robust method of generating a 3D model given a pair of stereo images. We start by segmenting our image in color space by using the adaptive mean shift segmentation and edge detection. The segmented image hence reproduced has a unique label assigned to every segment. We then calculate the occluded regions for our stereo set, colour them black and let remaining ones be white and go on to calculate the occluded pixel ratio.

Next we try to calculate the pixel disparity by using the method of local matching in pixel domain. The recovery of an accurate disparity map still remains challenging, mainly due to the following reasons:

- (i) Pixels of half occluded regions do not have correspondences in the other image, leading to incorrect matches if not taken into account.
- (ii) Images are disturbed because of sensor noise. This is especially problematic in poorly textured regions due to the low signal-to-noise-ratio (SNR).
- (iii) The constant brightness or color constraint is only satisfied under ideal conditions that can only roughly be met in practice.

We then assign a disparity plane to each segment by associating a segment with the median value of the disparity values of the pixels associated with it. The disparity that we get at this step filters out most of the noise that might hamper the performance of our final output i.e. the 3D model.

The disparity plot on a 3D mesh gives a pretty fair idea of the relative positions of the various objects in the images; But to improve the user understandability we try to regain the lost characteristics of the image by warping the intensity color values of the image on the disparity and plotting it on a 3D view. All the steps will be explained separately in the course of the paper.

The output we present in our paper should however, not be compared to outputs generated by more than two stereo images because of mainly two reasons :

- i. A large portion of the 3D model remains occluded as we cannot estimate the shape or characteristics of the occluded portions.
- ii. A complete 3D model cannot be generated without having covered all faces of any object, which requires a minimum of three cameras.

The idea can however be modified and improvised further to generate a complete 3D model provided we have the complete data set. We in this paper try to analyze the feasibility of our proposed method of generating a 3D model from a stereo pair.

2. LOCAL MATCHING IN PIXEL DOMAIN

2.1 Cost Estimation

Local matching requires to define a matching score and an aggregation window. The most common dissimilarity measures are squared intensity differences (SD) and absolute intensity differences (AD) that are strictly assuming the constant color constraint. Other matching scores such as gradient-based and non-parametric measures are more robust to changes in camera gain. In our approach[1] we are using a self-adapting dissimilarity measure that combines sum of absolute intensity differences (SAD) and a gradient based measure that are defined as follows:

$$C_{SAD}(x, y, d) = \sum_{(i,j) \in N(x,y)} |I_1(i, j) - I_2(i + d, j)| \quad \text{-----(1)}$$

And

$$C_{GRAD}(x, y, d) = \sum_{(i,j) \in N_x(x,y)} |\nabla_x I_1(i, j) - \nabla_x I_2(i + d, j)| + \sum_{(i,j) \in N_y(x,y)} |\nabla_y I_1(i, j) - \nabla_y I_2(i + d, j)| \quad \text{-----(2)}$$

where $N(x, y)$ is a 3×3 surrounding window at position (x, y) , $N_x(x, y)$ a surrounding window without the rightmost column, $N_y(x, y)$ a surrounding window without the lowest row, r_x the forward gradient to the right and r_y the forward gradient to the bottom. Color images are taken into account by summing up the dissimilarity measures for all channels.

An optimal weighting between C_{SAD} and C_{GRAD} is determined by maximizing the number of reliable correspondences that are filtered out by applying a cross-checking test (comparing left-to-right and right-to-left disparity maps) in conjunction with a winner-take-all optimization (choosing the disparity with the lowest matching cost). The resulting dissimilarity measure is given by:

$$C(x, y, d) = (1 - \omega) * C_{SAD}(x, y, d) + \omega * C_{GRAD}(x, y, d) \quad \text{-----(3)}$$

Though this is a simple and robust method there are now better enhanced methods of matching like MACO.[8]. A comparison between various feature extraction and recognition can be studied and appropriately used [9].

2.2 Horizontal Disparity Calculation

Using the method of cost estimation as explained above we calculate the disparity values for every pixel by subtracting the positions of correspondingly matched pixels in the two images. We assume a case of zero horizontal disparity throughout the course of our paper. To improve on the results obtained we repeat the above step twice, firstly keeping the right image at the base and sliding the left image over it and vice versa the second time. The minimum among the two disparity values for every pixel is considered as the final disparity value for that corresponding pixel.

Figure 1, precisely shows our result of horizontal disparity calculation for a given test case of images.

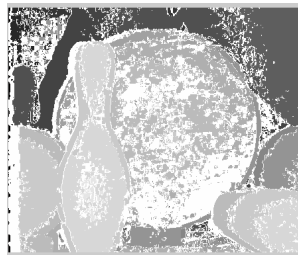


Figure1. Disparity Map

3. OCCLUSION ESTIMATION

As discussed earlier one of the major challenges in generating a 3D model lies in handling occlusion. We in our paper have used a simplistic method to show the half occluded regions i.e. those regions which do not have a corresponding match in the other image.

This method uses the disparity maps generated by the two iterations discussed in the Chapter 2.2. The disparity maps are subjected to scaling (to optimize our results) and appropriate thresh holding is done to uniquely identify the occluded portions.

The extent of half occlusion can be estimated by the absolute occlusion ratio, which is given as :

$$A_{occ} = N_{occ} / (x * y)$$

Here, N_{occ} is the total number occluded pixels and 'x' and 'y' represent the dimension of the image matrix. The multiplication of the x and y dimensions of the matrix give us the total number of pixels in the image. Figure 3, shows the identified half occluded regions. The portions in black are the identified occluded pixels.

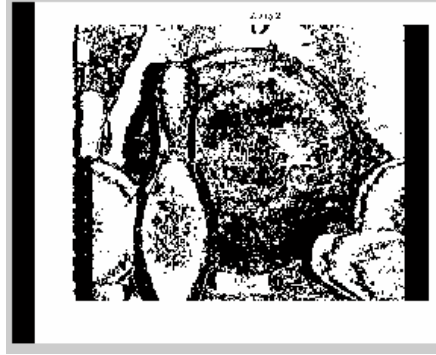


Figure 2. Occluded Regions

4. COLOUR SEGEMENTATION

a. Adaptive Mean Shift Segmentation and Edge detection

There are many ways of segmenting an image in an like Colour Histograms[7], JND histograms[6] etc. In our color image segmentation algorithm a five-dimensional feature space was used [2]. The color space was employed since its metric is a satisfactory approximation to Euclidean, thus allowing the use of spherical windows. The remaining two dimensions were the lattice coordinates. A cluster in this 5D feature space thus contains pixels which are not only similar in color but also contiguous in the image.

The quality of segmentation is controlled by the spatial h_s , and the color h_r , resolution parameters defining the radii of the (3D/2D) windows in the respective domains. The segmentation algorithm has two major steps. First, the image is *filtered* using mean shift in 5D, replacing the value of each pixel with the 3D (color) component of the

5D mode it is associated to. Note that the filtering is discontinuity preserving. In the second step, the basins of attraction of the modes, located within $-9;:_<$ in the color space are recursively *fused* until convergence. The resulting large basins of attraction are the delineated regions, and the value of all the pixels within are set to their average. See [3] and [4] for a complete description and numerous examples of the segmentation algorithm. It is important to emphasize that the segmenter processes gray level and color images in the same way. The only difference is that in the former case the feature space has three dimensions, the gray value and the lattice coordinates.

The mean shift based color image segmentation is already popular in the computer vision community and several

implementations exist. Along with image segmentation the open source EDISON we have used also does gradient based edge detection in the image. However, using the gradient magnitude for decisions causes a well known deficiency, sharp edges with small magnitudes can be detected only at the expense of allowing a large amount of edge clutter. A recently proposed generalization of the gradient based edge detection procedure eliminates this trade-off [5].

The result for the above mentioned process is shown in Figure 3. The code was written in MATLAB and the functions of C++ EDISON code were called using a MEX file.



Figure 3. Segmented Image

5. SNR CALCULATION:

Signal-to-noise ratio (often abbreviated SNR or S/N) is a measure used in to quantify how much a signal has been corrupted by noise. It is defined as the ratio of signal power to the noise power corrupting the signal. A ratio higher than 1:1 indicates more signal than noise. While SNR is commonly quoted for electrical signals, it can be applied to images also.

An alternative definition of SNR is as the reciprocal of the coefficient of variation, i.e., the ratio of mean to standard deviation of a signal or measurement.

$$SNR = \mu/\sigma \tag{4}$$

where μ is the signal mean or expected value and σ is the standard deviation of the noise, or an estimate thereof. Notice that such an alternative definition is only useful for variables that are always positive (such as photon counts and luminance). Thus it is commonly used in image processing, where the SNR of an image is usually calculated as the ratio of the mean pixel value to the standard deviation of the pixel values over a given neighborhood. Sometimes SNR is defined as the square of the alternative definition above.

We have accordingly calculated the SNR for all segments before filtering; some of them are shown in Table1.

Table 1: SNR for various Segments

1.2903	2.6692	3.3551	4.0759	5.3434	7.7233
8.1880	9.3920	9.7224	10.3752	11.2215	12.1938
13.5314	14.5834	15.8698	16.4618	20.2655	21.1895
22.9675	24.9660	25.3624	26.3882	27.8676	28.9518
29.5893	30.6746	31.1829	32.6312	33.1157	34.4666
35.5848	36.0835	37.5370	38.9565	39.7654	40.7948
41.9086	42.3569	43.9115	44.3283	45.5586	46.4970

48.0103	50.2326	55.8950	61.4123	63.9514	65.2723
67.6221	68.3737	70.7674	72.1460	92.3039	93.4421
109.4394	112.9547	114.5421	123.0981	127.1221	130.4643
135.1333	139.1510	144.1111	156.1265	199.2787	254.2282

After this we try to reduce the noise by applying a median filter on each corresponding segment. Hence we assign each segment with its corresponding median disparity value. By the end of filtering we try to achieve a SNR ratio of infinity for each segment. This process is better explained in 5.

6. DISPARITY SEGMENT ASSIGNMENT

The purpose of this step is to increase the accuracy of the disparity plane set by repeating the plane fitting for grouped regions that are dedicated to the same disparity plane. This can be done in two simple steps.

Firstly, we find out the total pixels associated with a segment and their corresponding disparity values. Secondly, we find the median value of all the disparities in that segment and assign that disparity to the segment. This process makes the disparity map neater and also helps in reducing the SNR. The method gets rid of sudden impulses or large unwanted variations in the value of the disparity. Though the method may trade-off with the actual disparity values but it helps the final result of generating a 3D viewing model. The result is shown in Figure 4.

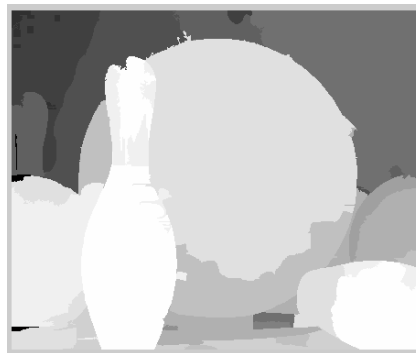


Figure 4. Filtered Disparity

7. DEPTH CALCULATION

In this module we try to calculate the depth of individual segments assuming a parallel camera case. We use the filtered disparity values to calculate the depth using the formula and figure shown below :

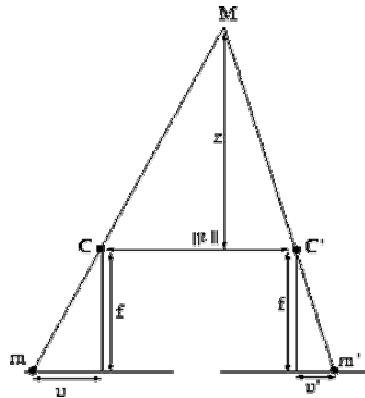


Figure 5

Given a scene point M and its two projection points m of coordinates (u, v) and m' of coordinates (u', v') , the *disparity value* d is defined as

$$d = u' - u \tag{5}$$

Note that $v = v'$ as there is no vertical parallax between the two cameras. The *depth measure* z of M is related to the disparity value d as follows:

$$z = \frac{f \| t \|}{d} \tag{6}$$

8. GENERATING A 3D VIEWING MODEL

a. 3D Plot of filtered disparity

This section deals with generating the final 3D view using the given pair of the images. The disparity calculated in the above step can be used directly to plot on a 3D mesh to get an estimate of the relative distances of various objects in the image. But, there still exists a major problem, i.e. the loss of original 3D intrinsic characteristics of the image in the 3D model

b. Warping image characteristics on disparity

Here we make an attempt to regain the original characteristics of the image. We warp the image intensity values from one of the input images onto the filtered disparity value matrix we got in Chapter 5. This method allows us to simultaneously plot the disparity and the intensity values on a 3D space, hence giving the user a fair idea of the relative depths of various objects identified in the images. The result is shown in Figure 6.

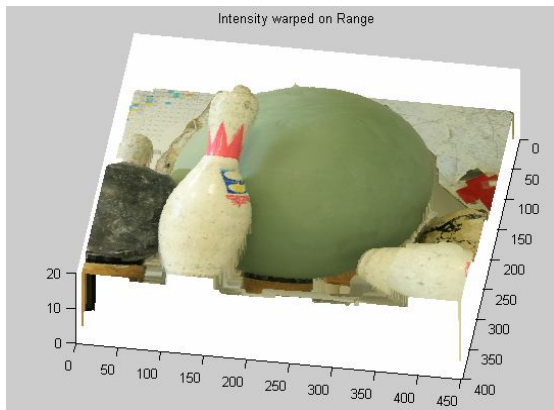


Fig 6(a)

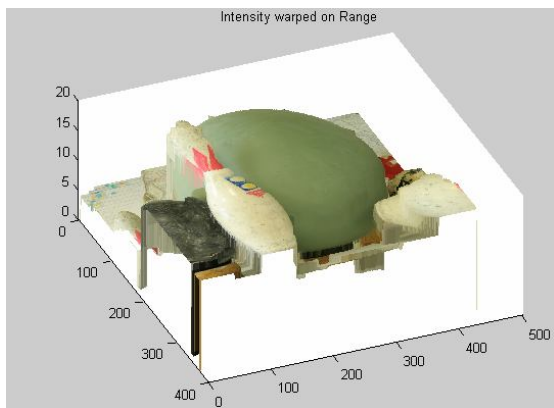


Fig 6(b)

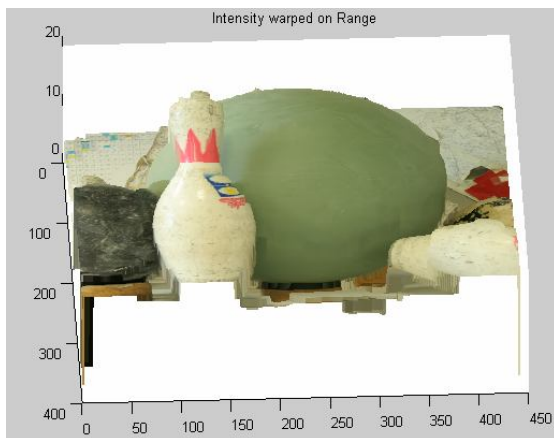


Fig 6(c)

Figure 6. Different Views of 3D Model

9. RESULTS

We have tested our proposed algorithm on a large set of stereo pairs. We have taken stereo pairs from the website of middlebury.edu. The occlusion ratios for different images are shown in Table 2 and some other sample outputs are given in Figures 7(b),8(b),9(b) and 10(b).

Table 2 : Occlusion Ratios of test cases

Sample Pair	Occlusion Ratio
Bowling	0.338
Aloe Plant	0.1801
Baby	0.2223
Pots	0.3939



Fig 7(a) "Baby" stereo pairs (right and left)

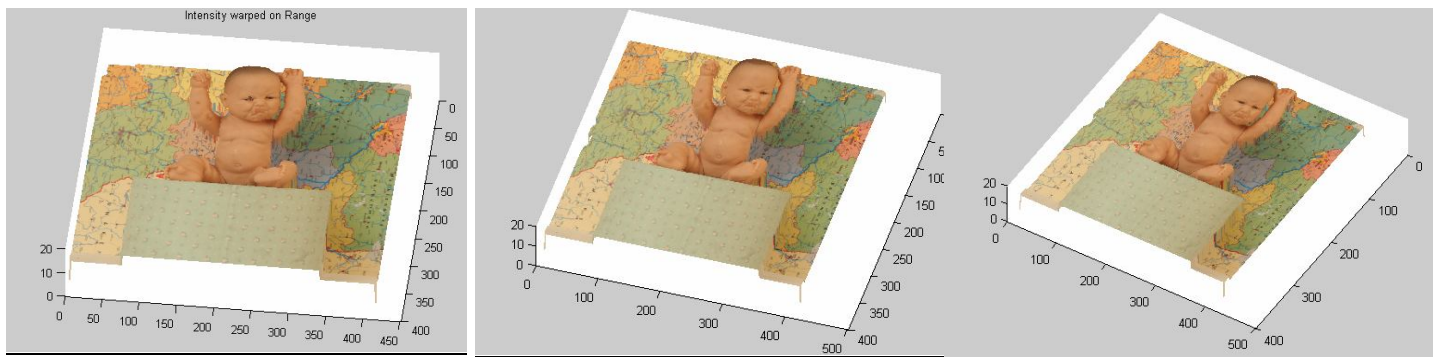


Fig 7(b) Some 3D model views of Stereo Images in 7(a)

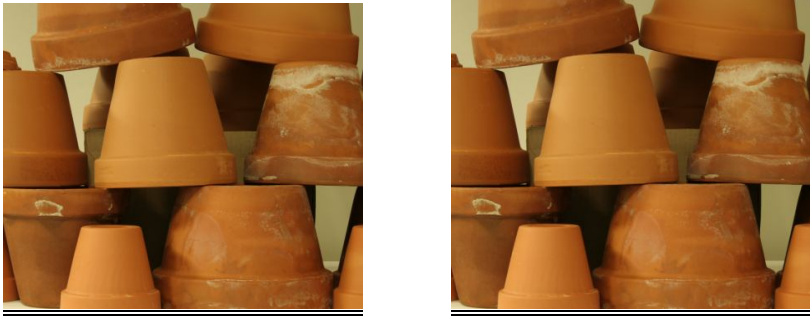


Fig 8(a) "Pots" stereo pairs (right and left)

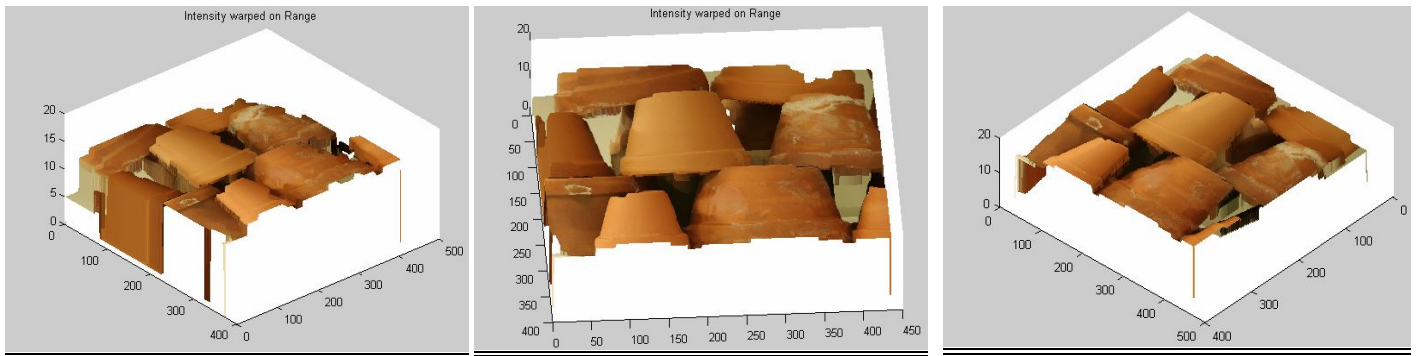


Fig 8(b) Some 3D model views of Stereo Images in 8(a)

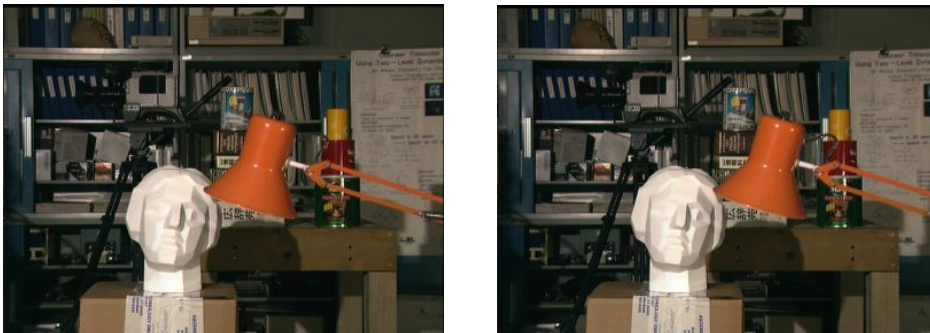


Fig 9(a) "Room" stereo pairs (right and left)

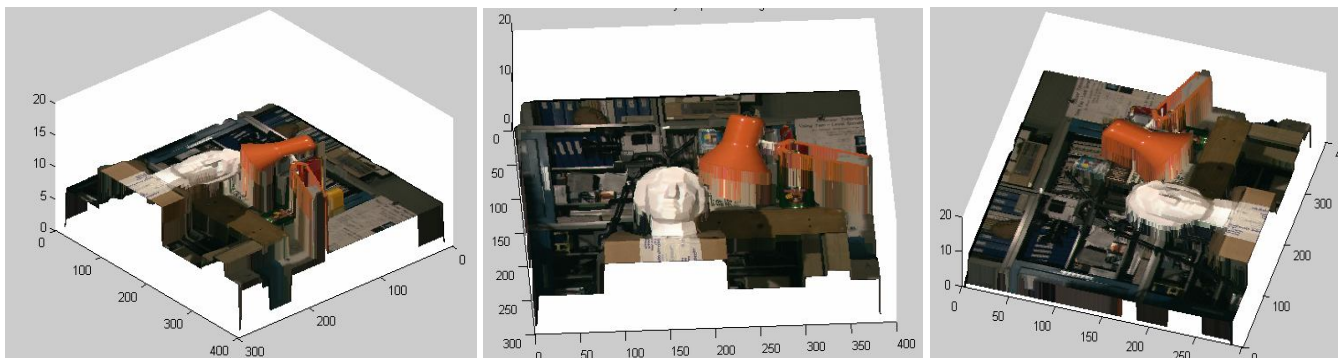


Fig 9(b) Some 3D model views of Stereo Images in 9(a)



Fig 10(a) "Aloe plant" stereo pairs (right and left)

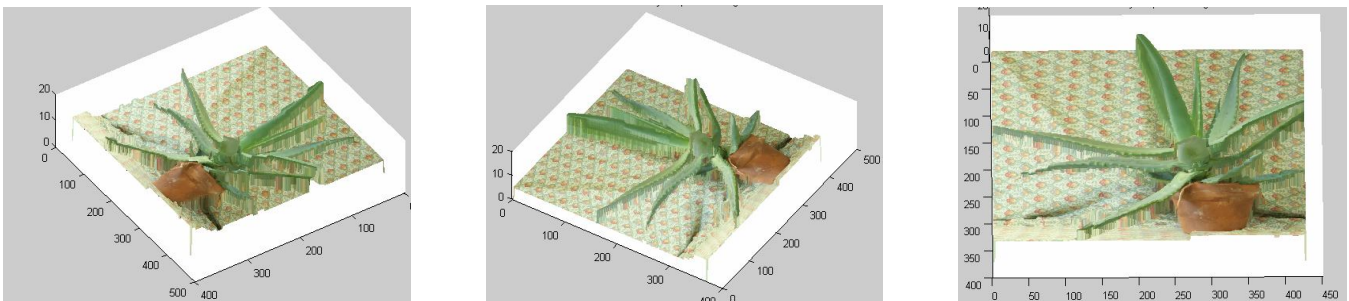


Fig 10(b) Some 3D model views of Stereo Images in 10(a)

10. CONCLUSION AND FUTURE WORK :

As it is clear from our result set that our proposed method works well for all set of stereo pairs. Our output set does not depend on the type of image and works equally well for grayscale and colored images. The number of objects is also not a constraint, just the occlusion ratio increases as the number of objects in the image increases. Our approach can be further used in various applications like :

- 'Virtual catwalk' to allow customers to visualize themselves in clothing prior to purchasing such goods on-line via the Internet.
- Potential to revolutionize autonomous vehicles and the capabilities of robot vision systems

11. REFERENCES

FOR JOURNALS:

- [1] Andreas Klaus, Mario Sormann and Konrad Karner : Segment Based Stereo Matching using Belief Propagation and Self Adapting Dissimilarity Measure, VRVis Research Center, 8010 Graz, Austria, 2006.
- [2] Christopher M. Christoudias, Bogdan Georgescu and Peter Meer : Synergism in Low Level Vision, Rutgers University, 2002.
- [3] D. Comaniciu and P. Meer. Mean shift analysis and applications. In 7th international Conference on computer vision, pages 1197-1203, Kerkyra, Greece, September 1999.
- [4] D. Comaniciu and P. Meer. Mean shift:A robust approach toward feature space analysis. IEEE Trans, Pattern Anal, Machine Intell,24,May 2002.
- [5] T. Pavlidis and Y. T. Liow. Integrating region growing and edge detection. *IEEE Trans. Pattern Anal. Machine Intell.*, 12:225–233, 1990.
- [6] Kishore Bhoyar and Omprakash Kadke, Color Image **Segmentation** Based On JND Color Histogram, International Journal of Image Processing(IJIP), Volume(3), Issue(6), 2010
- [7] M.Swain and D. Ballard,"*Color indexing*", International Journal of Computer Vision, Vol.7, no.1,1991.
- [8] XiaoNian Wang and Ping Jiang, "Multiple Ant Colony Optimizations for Stereo Matching", International Journal of Image Processing(IJIP), Volume(3), Issue(5), 2010
- [9] Luo Juan and Oubong Gwun, "A Comparison of SIFT, PCA-SIFT and SURF" International Journal of Image Processing(IJIP), Volume(3), Issue(5), 2010

FOR CONFERENCES:

- [10] D. Comaniciu and P. Meer. Mean shift analysis and applications. In 7th international Conference on computer vision, Kerkyra, Greece, September 1999.

FOR BOOKS:

Gonzaleez and Woods, Digital Image Processing

Optical Approach Based Omnidirectional Thermal Visualization

Wai-Kit Wong

*Faculty of Engineering and Technology,
Multimedia University,
75450 JLN Ayer Keroh Lama,
Melaka, Malaysia.*

wkwong@mmu.edu.my

Chu-Kiong Loo

*Faculty of Engineering and Technology,
Multimedia University,
75450 JLN Ayer Keroh Lama,
Melaka, Malaysia.*

ckloo@mmu.edu.my

Way-Soong Lim

*Faculty of Engineering and Technology,
Multimedia University,
75450 JLN Ayer Keroh Lama,
Melaka, Malaysia.*

wslim@mmu.edu.my

Abstract

In this paper, a new optical approach based on omnidirectional thermal visualization system is proposed. It will provides observer or image processing tool a 360 degree viewing of surrounding area using a single thermal camera. By applying the proposed omnidirectional thermal visualization system even in poor lighting condition, surrounding area is under proper surveillance and the surrounding heating machineries/items can be monitored indeed. Infrared(IR) reflected hyperbolic mirrors have been designed and custom made for the purpose of reflecting omnidirectional scenes in infrared range for the surrounding area to be captured on a thermal camera, thus producing omnidirectional thermal visualization images. Five cost effective and market-common IR reflected materials used to fabricate the designed hyperbolic mirror are studied, i.e. stainless steel, mild steel, aluminum, brass, and chromium. Among these materials, chromium gives the best IR reflectivity, with $\epsilon_r = 0.985$. Specifically, we introduce log-polar mapping for unwarping the captured omnidirectional thermal image into a panoramic view, hence providing observers or image processing tools a complete wide angle of view. Three mapping techniques are proposed in this paper namely the point sample, mean sample and interpolation mapping techniques. Point sample mapping technique provides the greatest interest due to its lower complexity and moderate output image quality.

Keywords: Image processing, Image representation, Omnidirectional vision, Thermal imaging, Architecture for imaging and vision systems.

1. INTRODUCTION

Conventional visualization method has limited field that make them restrictive in a variety of vision applications. Omnidirectional visualization design aims to expand the camera field of view in the horizontal plane, or with a visual field that covers (approximately) the entire sphere. At anytime, it

gives the views in all direction (360 degree coverage). Such omnidirectional visualization system would be applicable in various areas, such as remote surveillance, autonomous navigation, video conferencing, panoramic photography, robotics and the other areas where large field coverage is needed.

The classical acquisition of omnidirectional images is based on the use of either mechanical or optical devices, which are discussed in details in Section 2. The mechanical solutions are based on motorized linear or array-based cameras, usually with a 360 degree rotation, scanning the visual world. The main advantage of the mechanical solution is the possibility of acquiring very high-resolution images and its major drawback is the long time requirement to mechanically scan the scene to obtain a single omnidirectional image. Hence, it cannot provide video rate real-time omnidirectional images. On the other hand, optical solutions provide lower resolution images but they are the most appropriate solutions for real-time applications. Two optical alternatives have been proposed in literature, namely the use of special purpose lens (such as fish eye lens [1]) and the use of hyperbolic optical mirrors [2]. As real-time processing and applications are intend to be investigated, only the optical approach will be pursued.

Hyperbolic optical mirror is used in this paper because it is less expensive and not complex to design as compare to fish eye lens with almost the same reflective quality. The structure for the hyperbolic optical mirror is studied and designed, as can be viewed in Section 3 in details. One goal of this research paper is to use a thermal camera to replace digital camera in order to obtain thermal/infrared (IR) visualization. Therefore, the material used to design the hyperbolic optical mirror needs to be with highest IR reflectivity. In this research paper, five different cost effective IR reflected materials, used to fabricate the designed hyperbolic mirror are studied, i.e. stainless steel, mild steel, aluminum, brass, and chromium. Among these materials, chromium gives the best outcome in IR reflectivity, with IR relative ratio $\epsilon_r = 0.985$.

The image captured from the omnidirectional thermal imaging system is not immediately understandable because of the geometric distortion introduced by the optical mirror. Hence, an unwarping process needs to be introduced in order to map the original omnidirectional thermal image into a plane that provides a complete panoramic thermal image. The advantage of such an unwarping process lays in providing the observers with a complete view of its surrounding in one image which can be refreshed at video rate. Two unwarping methods are discussed in this paper namely the pano-mapping table's method [3] and log-polar mapping method [4]. Log-polar mapping is in favor to be selected for the proposed omnidirectional thermal visualization system because log-polar mapping does not required complicated landmark learning steps and it has lower computational complexity in calculating radial distance, comparing to pano-mapping table method. Log-polar mapping method also has higher compression capability too. Three mapping techniques for log-polar mapping are proposed in this paper namely point sample, mean and interpolation mapping techniques. Point sample mapping technique is applied in the proposed omnidirectional thermal visualization system due to its lower complexity and moderate output image quality.

The rest of the paper is organized as follows: Section 2 shows the mechanical approach versus optical approach in omnidirectional visualization and the proposed system model for our work. Section 3 presents the hyperbolic optical mirror design and discusses the materials used to construct the IR reflected hyperbolic optical mirror (or in short, omnidirectional hot mirror). Section 4 discusses two unwarping methods, namely the pano-mapping table and log-polar mapping. Three mapping techniques for log-polar mapping namely point sample, mean and interpolation mapping techniques are also discussed here. Finally, section 5 gives concluding remarks and envies some future work.

2. MECHANICAL APPROACH V.S. OPTICAL APPROACH

Omnidirectional visualization has significant potentials in various applications such as surveillance quality control, mobile robot navigation and others. The methods used to acquire omnidirectional views can be classified into two approaches [5]:

- 1.) Mechanical approach: the methods to gather images to generate an omnidirectional image.
- 2.) Optical approach: the methods to capture an omnidirectional image at once.

In addition, they are classified into two categories by the viewpoint of the image [5]: single viewpoint and multiple viewpoints.

For the mechanical approach, the images captured on a single viewpoint are continuous. One example is the rotating camera system [6]-[9]. In such a system, the camera rotates around the center of the lens. It generates an omnidirectional image from a single viewpoint. However, since it is necessary to rotate a video camera for a full cycle in order to obtain a single omnidirectional image, it is impossible to generate omnidirectional image at video rate. Other disadvantages of rotating camera system are that it requires the use of moving parts and precise positioning. The image captured at multiple viewpoints is relatively easy to be constructed. A single camera or more cameras are used to gather multiple images at multiple viewpoints and the images are combined into an omnidirectional image. Quick Time VR system [10] adopted such technologies and it has many market applications. However, the images generated by the system are not always continuous and consistent, and it also cannot capture the dynamic scene at video rate.

Since mechanical approach leads to many problems on discontinuity and inconsistent, optical approach is use. This approach is the most appropriate for real-time applications and it has single viewpoint. Two alternatives have been proposed, namely the use of special purpose lens (such as the fish eye lens [1]) and the use of hyperbolic optical mirrors [2]. Fish eye lens are used in place of a conventional camera lens due to the reason that they have very short focal length that enables the camera to view objects as much as in a hemisphere scene. Fish eye lens have been widely use for wide angle imaging areas as noted in [11, 12]. However, Nalwa's works in [13] found out that it is difficult to design the fish eye lens that ensure all the incoming principal rays intersect at a single point to yield a fixed viewpoint. The acquired image, using fish eye lens, normally does not permit the distortion free perspective images constructed from the viewed scene. Hence, to capture an omnidirectional view, the design of the optimal fish eye lens must be quite complex and large, and hence expensive. Since fish eye lens are cost expensive, complex in design and providing almost the same reflective quality as hyperbolic optical mirror, hyperbolic optical approach are planned to adopt in this research.

The optical approach based omnidirectional thermal visualization system model proposed in this paper is shown in Fig. 1. The system required a custom made omnidirectional hot mirror, a camera mirror holder, a fine resolution thermal camera and a laptop/PC with Matlab programming.

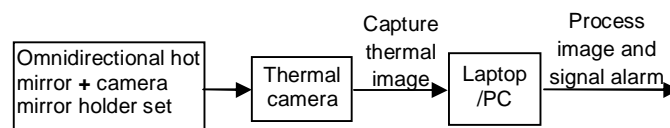


Fig. 1: Optical Approach Based Omnidirectional Thermal Visualization System Model



Fig. 2: Overall Fabricated Omnidirectional Thermal Visualization System Model

Custom made omnidirectional hot mirror will be discussed in details in Section 3. The camera mirror holder is custom designed in order to hold the camera and the omnidirectional hot mirror in a proper adjustable way. It is made with aluminum material due to its properties (light weight and acceptable tensile strength), as shown in Fig. 2.

The thermal camera used in this research is an affordable and accurate temperature model: ThermoVision A-20M manufactured by FLIR SYSTEM [14]. The thermal camera has a temperature sensitivity of 0.10 in a range from -20°C to 900°C and it can capture thermal image with fine resolution up to 320×240 pixels, offering more than 76,000 individual measurement points per image at a refresh rate of 50/60 Hz. The A-20M features a choice of connectivity options. For fast image and data transfer of real-time fully radiometric 16-bit images, an IEEE-1394 FireWire digital output can be chosen. For network and/or multiple camera installations, Ethernet connectivity is also available. Each A-20M can be equipped with its own unique URL allowing it to be addressed independently via its Ethernet connection and it can be linked together with router to form a network.

A laptop or PC can be used for image processor placing either on site or in a monitoring room. Matlab version 2007 or higher version is suggested because its user friendly environment (in m-file) is developed for performing log-polar mapping technique to unwrap the captured omnidirectional thermal image into panoramic form. For machine condition monitoring purposes, the software can be used to partition the panoramic thermal images easily according to each single machine to be monitored, processing them smoothly with the user-programmed monitoring algorithm and the signaling alarm whenever security threat alerts. It also can be used for trespasser detection. The overall fabricated system model is shown in Fig. 2.

3. Omnidirectional Hot Mirror

In this section, the structural design for omnidirectional hot mirror and the materials used for fabrication will be described.

3.1 Structural Design for Omnidirectional Hot Mirror

The best shape of practical use omnidirectional mirror is hyperbolic. As derived by Chahl and Srinivasan in [15], all the polynomial mirror shapes (conical, spherical, parabolic, etc) do not provides a central perspective projection, except for the hyperbolic one. They also shown that the hyperbolic mirror guarantee a linear mapping between the angle of elevation θ and the radial distance from the center of the image plane ρ . Another advantage of hyperbolic mirror is the resolution in the omnidirectional image captured is increasing with growing eccentricity when

using it with a camera/imager of homogenous pixel density. Hence, it will guarantee a uniform resolution for the panoramic image after unwarping.

The shape of the hyperbolic omnidirectional hot mirror used in our project is design according to the derivation work done in [16,17], such that at a given distance d from the origin, the vertical dimension h is linearly mapped to the radial distance from the center of the image plane ρ as shown in Fig. 3. The relation between a world point with coordinates $[d, h]^T$, the cross section function $F(t)$ with $t(\rho)$, and an image point with coordinates $[\rho, 0]^T$ is defined by the law of reflection [18], as refer to Fig. 3 and Fig. 4. According to [16], instead of deriving directly $F(t)$, an expression for $h(t)$ is fist sought. Referring to Fig. 3, the world point at a distance d is given by the following relation [16]:

$$h(t) = F(t) - \cot(\varphi)(d - t) \tag{1}$$

where it carry the meaning that a ray emanating from an image point is reflected by the mirror and reaches a world point. The incident and coincident rays on the mirror [16] as shown in Fig.5 are described by their directional vectors \vec{i} respectively \vec{c} , both of norm equal to one. The law of reflection [18] imposes that the angle between incident ray and the normal to the surface is equal to the angle between the normal and the coincident ray. The normal vector \vec{n} corresponds to a

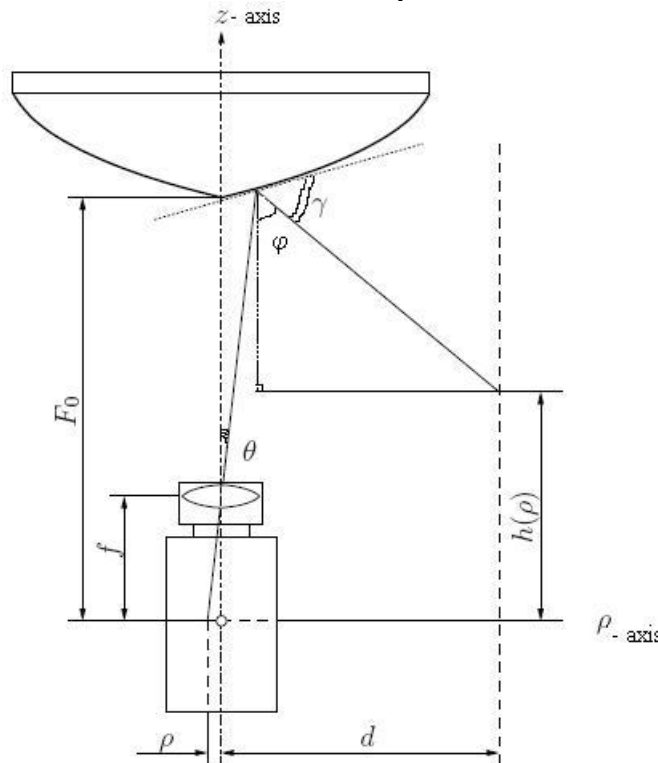


Fig. 3: Mirror design structural and its related parameters.

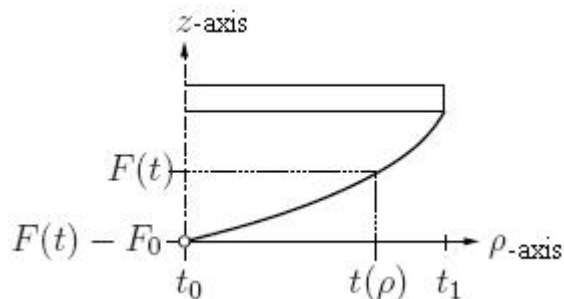


Fig. 4: Mirror cross section view.

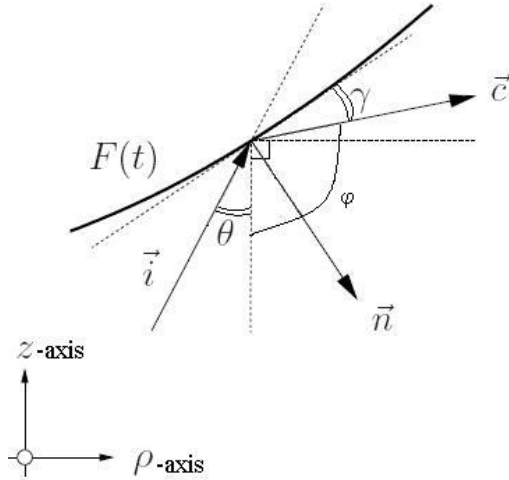


Fig. 5: incident and coincident rays on mirror surface

function of derivatives of the cross section function $F(t)$ at point t . Expressed in scalar product, the following condition must hold [16]:

$$-\vec{i} \cdot \vec{n} = \vec{n} \cdot \vec{c} \tag{2}$$

where the components of the vectors are as follows

$$\vec{i} = \frac{\vec{r}}{|\vec{r}|} = \begin{bmatrix} i_p \\ i_z \end{bmatrix}, \vec{r} = \begin{bmatrix} F(t) - f \\ t \end{bmatrix}, \vec{n} = \begin{bmatrix} \delta F(t) \\ -\delta t \end{bmatrix}, \vec{c} = \begin{bmatrix} c_p \\ c_z \end{bmatrix}.$$

The law of reflection impose $2\gamma + \theta + \varphi = \pi$ and therefore the slopes of the coincident ray is given by:

$$\frac{c_z}{c_p} = \cot(2\gamma + \theta) = -\cot(\varphi) \tag{3}$$

From (1), the term $-\cot(\varphi)$ can be expressed as a function of (3):

$$h(t) = F(t) + \frac{c_z}{c_p}(d - t) \tag{4}$$

Solving (2) by substituting each respectively scalar components:

$$\begin{aligned} -i_p \delta F(t) - i_z(-\delta t) &= \delta F(t) c_p - c_z \delta t \\ c_z &= i_p \frac{\delta F(t)}{\delta t} + c_p \frac{\delta F(t)}{\delta t} - i_z ; i_z = \frac{F(t)-f}{t} i_p \\ \frac{c_z}{c_p} &= F'(t) + \frac{i_p}{c_p} \left(F'(t) - \frac{F(t)-f}{t} \right) \end{aligned} \tag{5}$$

The condition for two unity vectors, $\|\vec{i}\| = \|\vec{c}\| = 1$.

$$\sqrt{i_p^2 + i_z^2} = \sqrt{c_p^2 + c_z^2} = 1 \tag{6}$$

From (6),

$$\begin{aligned} i_p^2 &= 1 - i_z^2 ; i_z = \frac{c_z}{c_p} i_p \\ i_p^2 &= 1 - \left(\frac{c_z^2}{c_p^2} i_p^2 \right) \end{aligned}$$

$$i_p^2 = \frac{1}{1 + \frac{c_z^2}{c_p^2}} \tag{7}$$

From (6),

$$\begin{aligned} c_p^2 &= 1 - c_z^2 \quad ; \quad c_z = \frac{c_p}{i_p} i_z \\ c_p^2 &= 1 - \left(\frac{c_p^2}{i_p^2} i_z^2 \right) \quad ; \quad i_z^2 = \left(\frac{F(t) - f}{t} \right) i_p^2 \\ c_p^2 &= \frac{1}{1 + \left(\frac{F(t) - f}{t} \right)^2} \end{aligned} \tag{8}$$

Taking (7) divide by (8), gives

$$\left(\frac{i_p}{i_c} \right)^2 = \frac{1 + \left(\frac{F(t) - f}{t} \right)^2}{1 + \left(\frac{c_z}{c_p} \right)^2} \tag{9}$$

Combining (5) and (9) results in a cubic expression [16]:

$$\begin{aligned} &\left(\left(F'(t) - \frac{F(t) - f}{t} \right)^2 - \left(1 + \left(\frac{F(t) - f}{t} \right)^2 \right) \right) \left(\frac{c_z}{c_p} \right)^2 + 2F'(t) \left(1 + \left(\frac{F(t) - f}{t} \right)^2 \right) \left(\frac{c_z}{c_p} \right) + \\ &\left(\left(F'(t) - \frac{F(t) - f}{t} \right)^2 - F'(t)^2 \left(1 + \left(\frac{F(t) - f}{t} \right)^2 \right) \right) = 0 \end{aligned} \tag{10}$$

Solving (10) for $\frac{c_z}{c_p}$, get:

$$\frac{c_z}{c_p} = \frac{F(t) - f}{t} \tag{11}$$

$$\frac{c_z}{c_p} = \frac{2tF'(t) - (F(t) - f)(1 - F'(t)^2)}{2(F(t) - f)F'(t) + t(1 - F'(t)^2)} \tag{12}$$

(11) corresponding to the slope of the transmitted ray, while (12) corresponds to the slope of the reflected ray, which is the sought one [16]. The expression for $h(t)$ in (4) can now be written as follows:

$$h(t) = F(t) + \frac{2tF'(t) - (F(t) - f)(1 - F'(t)^2)}{2(F(t) - f)F'(t) + t(1 - F'(t)^2)} (d - t) \tag{13}$$

Solving (13) for the derivative of the cross section function $F'(t)$ results in a cubic differential equation [16]:

$$F'(t)^2 + 2 \frac{t(d-t) + (F(t) - f)(F(t) - h(t))}{(F(t) - f)(d-t) - t(F(t) - h(t))} F'(t) - 1 = 0 \tag{14}$$

The differential equation, which defines the convex mirror shape is consequently given by the following expression [16]:

$$F'(t) + \frac{t(d-t) + (F(t) - f)(F(t) - h(t))}{(F(t) - f)(d-t) - t(F(t) - h(t))} - \sqrt{\left(\frac{t(d-t) + (F(t) - f)(F(t) - h(t))}{(F(t) - f)(d-t) - t(F(t) - h(t))} \right)^2 + 1} = 0 \tag{15}$$

From(15), there are three parameters influencing the mirror shape, which are the focal length f of the camera, the distance d that corresponds to the perimeter of the projected cylinder, and

the function $h(t)$, that corresponds to the vertical dimension of a world point. The parameter $h(t)$ also defines the characteristic of the thermal camera. The following condition must hold [16]:

$$h(\rho) = a\rho + b \tag{16}$$

where a is the gain and b is the offset of the function $h(\rho)$ in order to have a linear relationship between the coordinate of a world point and the coordinate of an image point. When substituting $h(\rho)$ in (16) into (15), the image coordinate ρ must be replaced by t . The relation between ρ and t results from the projection by a thermal camera as:

$$\rho = \frac{ft}{F(t)-f} \tag{17}$$

The research group of OMNIVIEWS project from Czech Technical University further developed Matlab software for designing omnidirectional optical mirror [16, 17]. From the Matlab software, practitioner can design their own hyperbolic omnidirectional hot mirror by inputting some parameters specifying the mirror dimension. The first parameter is the focal length of the camera f , in which the thermal camera in use is 12.5 mm and the distance d (ρz -plane) from the origin is set to 2m. The image plane height h is set to 20 cm. The radius of the mirror rim is chosen $t_1=3.6\text{cm}$ as modified from Svoboda work in [17], with radius for fovea region 0.6cm and retina region 3.0cm. Fovea angle is set in between 0° to 45° , whereas retina angle is from 45° to 135° . The coordinates as well as the plot of the mirror shape is generated using Matlab and it is shown in Fig. 6. The coordinates as well as the mechanical drawing using Autocad was provided to precision engineering company to fabricate/custom made the hyperbolic mirror.

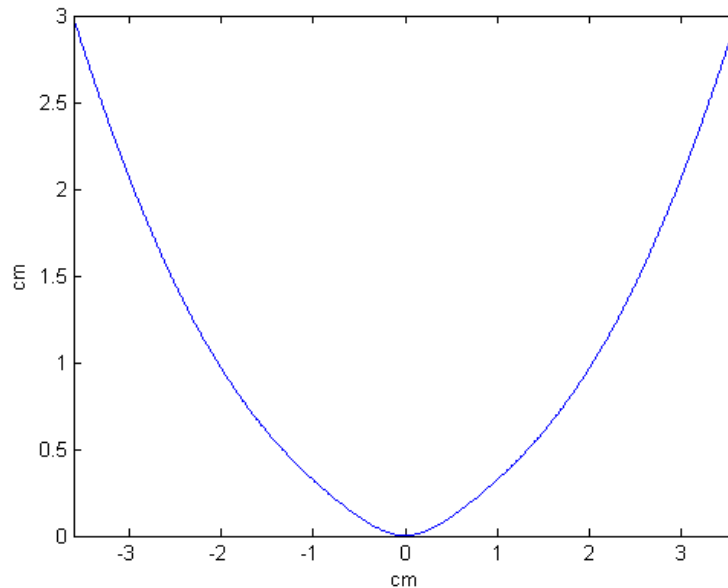


Fig. 6: Mirror coordinates plot in Matlab

3.2 Material for Fabricating Omnidirectional Hot Mirror

In particular, all kinds of flat and polished metal surfaces may easily reflecting infrared radiation, even if some of them are not reflecting visible light radiation. As a result, analysis of infrared thermal images must consider the thermal reflections capability of the use of a reflector or a mirror. In classical antiquity, mirrors were made directly of solid metal such as gold, silver, copper, tin, aluminum and etc [19]. In modern world, mirrors are manufactured by applying a reflective coating to a suitable substrate. For scientific optical work, dielectric materials are deposited, to form an optical coating. By carefully selecting the type and thickness of the dielectric layers, the range of the wavelengths and amount of light reflected from the mirror can be specified.

Visible light mirror's substrate is normally glass, due to its transparency, ease of fabrication, rigidity and ability to take a smooth finish. The reflective coating is either silver (much expensive) or tin-mercury (cheaper) [20], which is applied to the back surface of the glass, so that it is protected from corrosion and accidental damage. However, this type of visible light mirror is not suitable to be applied in IR reflectivity. The reflectivity of mirror is a function of wavelength depends on both the thickness of coating and how it is applied.

Although most mirrors are designed for visible light, there are also mirrors that designed for other types of waves or wavelengths spectrum in used. The mirror that used in IR spectrum is so called "hot mirror" [21]. Hot mirror is a specialized dielectric mirror that can reflect infrared light. Normally, the materials used to build hot mirror are metals [22]. Among the metals used to fabricate hot mirror such as aluminum, silver, gold and copper based on IR reflectivity were studied by Jitra in [22]. According to the studies, the most convenient materials are silver, gold and copper because these materials have very high reflectance in IR spectral range, which is upto 90-95% of IR wavelength region [23] when they are new. If they are deposited in very thin layer, they may have high light transmittance in visible spectral range too.

However, gold and copper are very expensive materials in which they were too expensive for widespread use, as well as being prone to corrosion. They are not cost effective to be chosen for hot mirrors. And, if such hot mirrors are used on remote site, they are easily to get lost or steal by theft. Silver tends to be very soft mechanically and easily abraded, as well as susceptible to tarnishing and corrosion over time from ordinary atmospheric contaminants/conditions, such as by reacting with oxygen, chlorine, sulfur, ozone and others. Because of these reasons, silver coatings are often not used in optical mirror systems unless they are protected from the elements stated above [24]. In Jitra's studies [22], aluminium is also not a suitable material because of its high visible reflectance.

In our research project, five types of cost effective and market-common IR reflected materials are chosen to fabricate the designed hot mirrors. They are: brass, mild steel, stainless steel, aluminum and chromium. Since fabricating mirror with whole chromium material is too expensive, it was substituted with aluminum coated with chromium. These fabricated hot mirrors are tested for reflectivity to find out the optimum one. The experimental results are shown and discussed in Section 5.1.

4. Unwarping Process

One of the important processes in optical based omnidirectional thermal visualization is the unwarping process. Unwarping process is required to map the captured omnidirectional image into a panoramic image, providing the observers and image processing tools with a complete wide angle of view for the surrounding of the area under surveillance. Many conventional unwarping methods require some information of certain omnidirectional imaging tools' intrinsic parameters, such as lens' focal length, coefficients of the mirror surface shape equation and the others. to perform calibration before unwarping take place [25-29]. This is not applicable in some situations where the information of such omnidirectional imaging tools' parameters are unknown. Therefore, it is desirable to have some unwarping methods which are universal to all kinds of omnidirectional imaging tools. In this section, two such unwarping methods will be discussed, one method is the recent works of S. W. Jeng et. al. in [3] by using pano-mapping table and another one is our proposed work by the use of log-polar mapping [4].

4.1 Pano-mapping Table

The proposed work of S. W. Jeng et. al. in [3], by using pano-mapping table for unwarping omnidirectional images into panoramic view images, consists of three major stages:

- (i) Landmark Learning Stage [3]: This stage is to setup some pairs of selected world space points (Authors in [3] name it as landmark points) with known positions and known pixels. The steps start by selecting a number (normally > 5) of landmark point pairs, each with the world space points which is easy to be identified (for example, a corner in a room). The coordinates of the selected world space points are measured. Take the coordinates data of the point pair (O_c, O_m)

as the first set of the learned data, where O_c is the image center with known world coordinates (u_o, v_o) and O_m is the mirror center with known world coordinates (x_o, y_o, z_o) . After learning, there will be n sets of landmark point pairs data, each set includes the coordinates (u_k, v_k) of the image point and (x_k, y_k, z_k) of the world space point respectively, where $k=0,1,\dots, n-1$.

(ii)Table Creation Stage [3]: This stage is building a two-dimensional mapping table using the coordinates data of the landmark point pairs. The pano-mapping table is a two-dimensional array, acting as a medium for unwarping omnidirectional images, after it has been conducted. The table is shown in Table 1 below.

	θ_1	θ_2	θ_3	θ_4	...	θ_M
	(u_{11}, v_{11})	(u_{21}, v_{21})	(u_{31}, v_{31})	(u_{41}, v_{41})	...	(u_{M1}, v_{M1})
ρ_1						
	(u_{12}, v_{12})	(u_{22}, v_{22})	(u_{32}, v_{32})	(u_{42}, v_{42})	...	(u_{M2}, v_{M2})
ρ_2						
	(u_{13}, v_{13})	(u_{23}, v_{23})	(u_{33}, v_{33})	(u_{43}, v_{43})	...	(u_{M3}, v_{M3})
ρ_3						
	(u_{14}, v_{14})	(u_{24}, v_{24})	(u_{34}, v_{34})	(u_{44}, v_{44})	...	(u_{M4}, v_{M4})
ρ_4						
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
	(u_{1N}, v_{1N})	(u_{2N}, v_{2N})	(u_{3N}, v_{3N})	(u_{4N}, v_{4N})	...	(u_{MN}, v_{MN})
ρ_N						

Table 1: Example of pano-mapping table of size $M \times N$ [3].

The horizontal and vertical axes on the pano-mapping table specifies the range of the azimuth angle θ and the elevation angle ρ respectively of all possible incident light rays going through the mirror center. The range 2π radians (360°) of the azimuth angles are divided equally into M -units and the range of the elevation angles are divided equally to N -units to create a table T_{pm} of $M \times N$ entries as shown in Table 1. Each Entry E with corresponding angle pair (θ, ρ) in T_{pm} maps to a pixel p with coordinates (u, v) in the input omnidirectional image I .

Due to nonlinear property of the mirror surface shape, the radial-directional mapping should be specified by a nonlinear function f_r , such that the radial distance r from each image pixel p with coordinates (u, v) in I to the image center O_c at (u_0, v_0) may be computed by $r = f_r(\rho)$. The pairs $(r, \varphi) = (f_r(\rho), \theta)$ of all the image pixels form a polar coordinate system with the image coordinates (u, v) specified by:

$$\begin{aligned} u &= r \times \cos\varphi = f_r(\rho) \times \cos\theta, \\ v &= r \times \sin\varphi = f_r(\rho) \times \sin\theta \end{aligned} \tag{18}$$

In [3], the authors also name f_r as the ‘radial stretching function’, and they attempt to describe it with the following 4th degree polynomial function:

$$r = f_r(\rho) = a_0 + a_1 \times \rho^1 + a_2 \times \rho^2 + a_3 \times \rho^3 + a_4 \times \rho^4 \tag{19}$$

where $a_0, a_1, a_2, a_3,$ and a_4 are five coefficients to be estimated using the data of the landmark point pairs, as described in the following algorithm. A similar idea of approximation can be found in [30].

Algorithm 1. Estimation of coefficients of radial stretching [3]:

Step 1: *Coordinates transformation in world space*: -Transform the world coordinates (x_k, y_k, z_k) of each selected landmark point $P_k, k=1, 2, \dots, n-1$, with respect to origin of world coordinate O_w into coordinates with respect to O_m by subtracting from (x_k, y_k, z_k) the coordinates values $(x_0, y_0, z_0)=(-D, 0, H)$ of O_m . Hereafter, (x_k, y_k, z_k) will be denote this coordinates transformation result.

Step 2: *Elevation angle and radial distance calculation*: -The coordinate’s data of each landmark point pair (P_k, ρ_k) , including the world coordinates (x_k, y_k, z_k) and the image coordinates

(u_k, v_k) are use to calculate the elevation angle ρ_k of P_k in the world space and the radial

distance r_k of ρ_k in the image plane by the following equations:

$$\begin{aligned} \rho_k &= \tan^{-1} \left[\frac{z_k}{D_k} \right], \\ r_k^2 &= u_k^2 + v_k^2 \end{aligned} \tag{20}$$

where D_k is the distance from the landmark point P_k to the mirror center O_m in the X-Y plane of the world coordinates system, computed by $D_k = \sqrt{x_k^2 + y_k^2}$.

Step 3: *Calculation of coefficients of the radial stretching function*: -Substitute all the data $\rho_0,$

$\rho_1, \dots, \rho_{n-1}$ and r_0, r_1, \dots, r_{n-1} computed in step 2 into (19) to get n simultaneous

equations:

$$\begin{aligned} r_0 &= f_r(\rho_0) = a_0 + a_1 \times \rho_0^1 + a_2 \times \rho_0^2 + a_3 \times \rho_0^3 + a_4 \times \rho_0^4, \\ r_1 &= f_r(\rho_1) = a_0 + a_1 \times \rho_1^1 + a_2 \times \rho_1^2 + a_3 \times \rho_1^3 + a_4 \times \rho_1^4, \\ &\vdots \\ r_{n-1} &= f_r(\rho_{n-1}) = a_0 + a_1 \times \rho_{n-1}^1 + a_2 \times \rho_{n-1}^2 + a_3 \times \rho_{n-1}^3 + a_4 \times \rho_{n-1}^4 \end{aligned} \tag{21}$$

and solve them to get the desired coefficients (a_0, a_1, a_2, a_3 , and a_4) of the radial stretching function f_r by a numerical analysis method [31].

Now the entries of the pano-mapping table T_{pm} can be filled with the corresponding image coordinates using (18) and (19) with the following algorithm. Note that there are $M \times N$ entries in the table.

Algorithm 2. Filling entries of pano-mapping table [3]:

Step 1: Divide the range 2π of the azimuth angles into M intervals, and compute the i -th

azimuth angle θ_i by:

$$\theta_i = i \times (2\pi/M) \tag{22}$$

for $i=0,1, \dots, M-1$.

Step 2: Divide the range $[\rho_e - \rho_s]$ of the elevation angles into N intervals, as shown in Fig. 7, and compute the j -th elevation angle ρ_j by:

$$\rho_j = j \times [(\rho_e - \rho_s)/N] + \rho_s \tag{23}$$

for $j=0,1, \dots, N-1$.

Step 3: Fill the entry E_{ij} with the corresponding image coordinates (u_{ij}, v_{ij}) computed according to (18) and (19) as follows:

$$\begin{aligned} u_{ij} &= r_j \times \cos\theta_i, \\ v_{ij} &= r_j \times \sin\theta_i \end{aligned} \tag{24}$$

where r_j is computed by:

$$r_j = f_r(\rho_j) = a_0 + a_1 \times \rho_j^1 + a_2 \times \rho_j^2 + a_3 \times \rho_j^3 + a_4 \times \rho_j^4 \tag{25}$$

with $(a_0, a_1, a_2, a_3, a_4)$ being those computed by Algorithm 1.

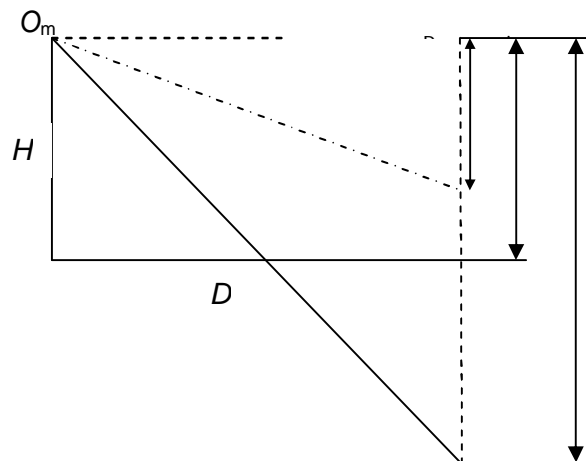


Fig. 7: Lateral-view configuration for generating a panoramic image [3].

(iii) Image Unwarping Stage [3]: This stage is to construct a panoramic image Q using the pano-mapping table T_{pm} as a medium to map each pixel q in Q to an entry E_{ij} with coordinates values (u_{ij}, v_{ij}) in the pano-mapping table T_{pm} and to assign the color value of the pixel at coordinates (u_{ij}, v_{ij}) of image I to pixel q in Q . For the details of the generation of panoramic image from a given input omnidirectional image I and a pano-mapping table T_{pm} , a corresponding generic panoramic image Q may be generated from I , which is exactly of the same size as T_{pm} . The steps are as follows: First, for each entry E_{ij} of T_{pm} with azimuth angle of θ_i and elevation angle ρ_j , take out the coordinates (u_{ij}, v_{ij}) filled in E_{ij} , then assign the color values of the pixel p_{ij} of I at coordinates (u_{ij}, v_{ij}) to the pixel q_{ij} of Q at coordinates (i, j) . After all entries of the table are processed, the final Q becomes a generic panoramic image corresponding to I . In the process, Q may regard as the output of the pano-mapping f_{pm} described by the pano-mapping tables T_{pm} with I as the input, i.e. $f_{pm}(I) = Q$.

Pano-mapping method shows best results in unwarping omnidirectional images taken by any omnidirectional imaging tools without requirement of omnidirectional tools' parameters. It is a universal unwarping method. However, it requires complicated landmark learning steps to define/search for different landmark points for table creation. It also has high computational complexity in which it requires up to 4th order calculation in each radial distance r as shown in Step 3 of Algorithm 1. Another demerit of such unwarping method is that it has no data compression capability. As observe from the experimental results in [3], the output panoramic image has resolution three to four folds greater than the input omnidirectional images, in which image expansion at output.

In next subsection, another universal unwarping method will be proposed, so-called the log-polar mapping method. In contrast with pano-mapping table method, log-polar mapping method does not require complicated landmark learning steps. It only needs to define one single point, which is the center of image, so-called the fovea. From the center of fovea, user can develop the fovea region and periphery region covering the whole omnidirectional image, bio-inspired the

mammalian viewing image. The computational complexity to calculate each radial distance ρ is

low, which requires cosine and sine calculation function only. Log-polar mapping method also has higher compression capability up to 4 folds.

4.2 Log-Polar Mapping

Log-polar mapping geometry is an example of foveated or space-variant image representation used in the active vision systems motivated by human visual system [32]. It is a spatially-variant image representation in which pixel separation increases linearly with distance from a central point [33]. It provides a way of concentrating computational resources on regions of interest, whilst retaining low-resolution information from a wider field of view. One advantage of this kind of sampling is data reduction. Foveal image representations like this are most useful in the context of active vision system where the densely sampled central region can be directed to pick up the most salient information. Human eyes are very roughly organized in this way.

In robotics, there has been a trend to design and use true retina-like sensors (e.g. SVAVISCA log-polar camera [34] and CMOS foveated image sensor [35]) or simulate the log-polar images by software conversion (e.g. application to face detection and tracking [36], application to monocular active tracking [37], etc). In the software conversion of log-polar images, practitioners in pattern recognition usually named it as log-polar mapping. An example illustrating the log-polar mapping of a hypothetical N rings retina, containing a 3 rings fovea, is shown in Fig. 8.

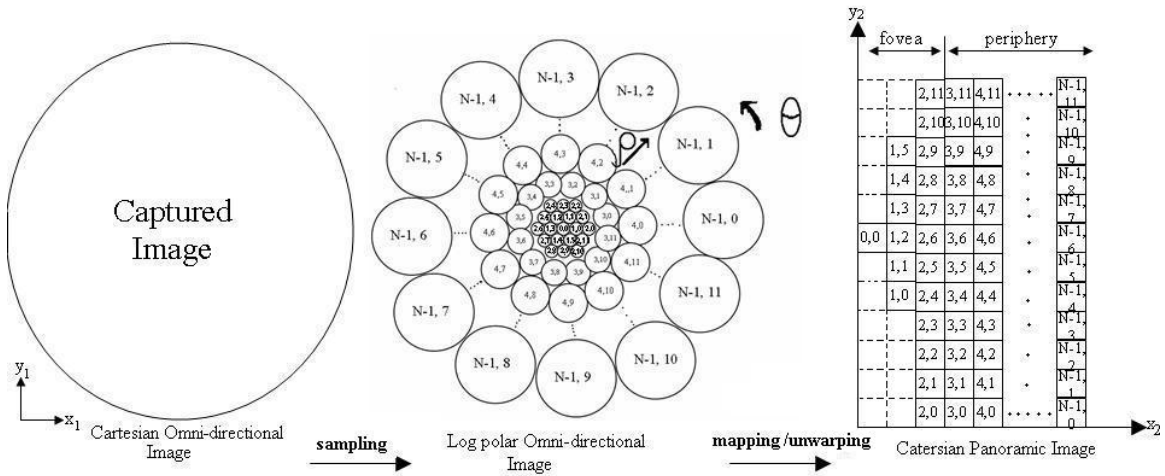


Fig. 8: An example of log-polar mapping of a N rings retina containing 3 fovea. To simplify the figure, no overlapping was used. Note that periphery portion is log-polar.

The log-polar mapping can be summarized as following: Initially, the geometry of each section of the partitioned thermal image is in Cartesian form (x_1, y_1) . Log-polar sampling is uses to sample the Cartesian input image into log-polar sampling image. After that, the log-polar image is mapped to another Cartesian form (x_2, y_2) whereby in this process, the log-polar sampling image is unwrapped into a log-polar mapping image. Since the output log-polar mapping image provides data compression and it is also in Cartesian form, subsequent image processing task will become much easier.

The centre of pixel for log-polar sampling expression is described in [4]:

$$\rho(x_1, y_1) = \log_{\lambda} \frac{R}{r_0} \tag{26}$$

$$\theta(x_1, y_1) = \frac{N_{\theta}}{2\pi} \arctan \frac{y_1}{x_1} \tag{27}$$

The centre of pixel for log-polar mapping expression is described in [4]:

$$x_2(\rho, \theta) = \lambda^{\rho} r_0 \cos \left(\frac{2\pi\theta}{N_{\theta}} \right) \tag{28}$$

$$y_2(\rho, \theta) = \lambda^{\rho} r_0 \sin \left(\frac{2\pi\theta}{N_{\theta}} \right) \tag{29}$$

where $R = \sqrt{x_1^2 + y_1^2}$ is the distance between the given point and the center of mapping

r_0 is a scaling factor which will define the size of the circle at $\rho(x_1, y_1) = 0$.

λ is the base of the algorithm,

$$\lambda = \frac{1 + \sin \left(\frac{\pi}{N_{\theta}} \right)}{1 - \sin \left(\frac{\pi}{N_{\theta}} \right)} \tag{30}$$

N_{θ} is the total number of pixels per ring in log-polar geometry.

The number of rings in the fovea region is given by [4]:

$$N_{fov} = \frac{\lambda}{\lambda - 1} \tag{31}$$

To sample the Cartesian pixels (x_1, y_1) into log polar pixels (ρ, θ) , at each center point

calculated using (1) and (2), the corresponding log-polar pixel (ρ_n, θ_n) is cover a region of

Cartesian pixels with radius:

$$r_n = \lambda r_{n-1} \tag{32}$$

where $n=0, 1, \dots, N-1$. Fig. 9 shows the conventional circle sampling method of log-polar mapping [32, 35].

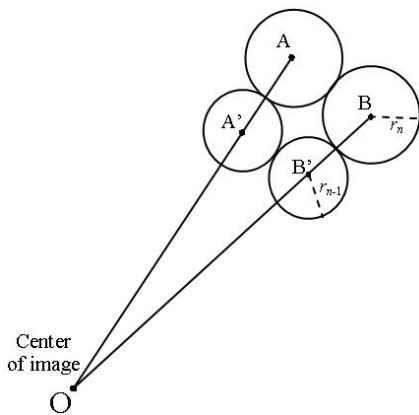


Fig. 9: Conventional circle sampling method for log polar image.

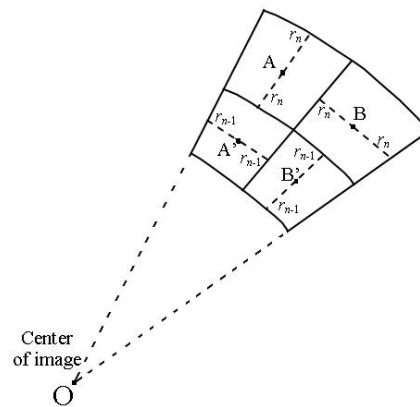


Fig. 10: Sector sampling method for log-polar image.

One of the disadvantages of using circle sampling is that certain region of Cartesian pixels outside sampling circle did not cover by any log-polar pixels. Therefore, some researchers [36, 38, 39] had adopted sector sampling method as shown in Fig. 10, which could maximize the coverage of Cartesian pixels for each log-polar pixel. Log-polar pixel image will later perform mapping (unwarping) process to convert into panoramic image. There are three types of mapping techniques proposed in this paper to help generating a good quality panoramic image. They are point sample, mean and interpolation mapping technique.

4.2.1 Point sample

Point sample is the simplest mapping techniques. During mapping process, the (ρ, θ) pixels will

map to each corresponding (x_2, y_2) pixels as shown in Fig. 11. The region of panoramic Cartesian pixels cover by an individual log-polar pixel will has the same color intensity follow the original Cartesian center sampling point's pixel from the respective log-polar pixel.

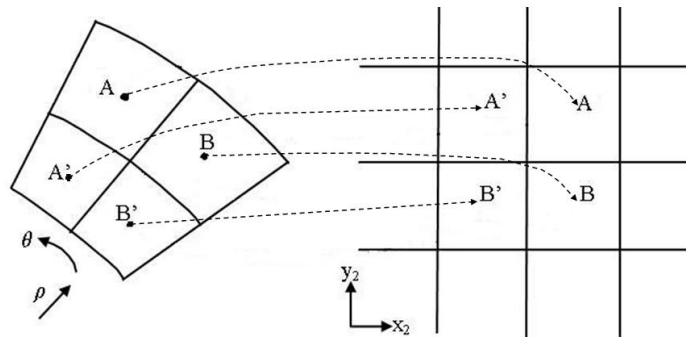


Fig. 11: Point sample mapping process

4.2.2 Mean

In mean mapping technique, the intensity value in each individual log-polar pixel equals the mean intensity value of all pixels inside the sampling sector on the original Cartesian image (x_1, y_1) :

$$\text{mean} = \frac{\text{Total } (x_1, y_1) \text{ pixels' value cover by a particular log-polar pixel}}{\text{Total number of } (x_1, y_1) \text{ pixels cover in that particular log-polar pixel}} \quad (33)$$

Fig. 12 shows the mean mapping process.

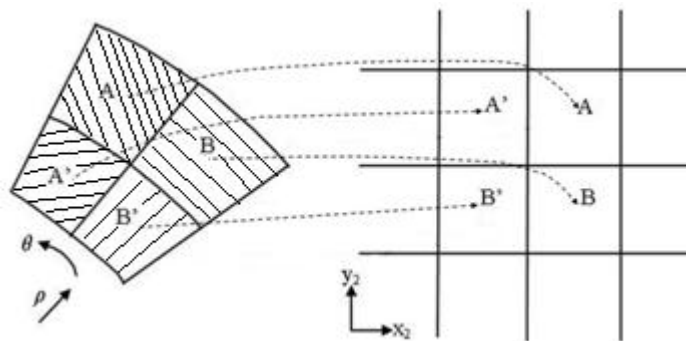


Fig.12: Mean mapping process

4.2.3 Interpolation

This method performs interpolation on point sample or mean mapping image. Interpolation is the process of determining the values of a function at the positions lying between its samples [40]. Interpolation reconstructs the information lost in the sampling process by smoothing the data samples with an interpolation function. Several interpolation methods have been developed and can be found in the literature [40- 42]. The most commonly used interpolation methods in image processing are the nearest neighbor, bi-linear and bi-cubic interpolation techniques.

(i) Nearest Neighbour Interpolation

Nearest neighbor is considered the simplest interpolation method from a computational standpoint [40], where each interpolated output pixel is assigned the value of the nearest sample point in the input image. The interpolation kernel for the nearest neighbor algorithm is defined as [40]:

$$h(x) = \begin{cases} 1 & 0 \leq |x| < 0.5 \\ 0 & 0.5 \leq |x| \end{cases} \quad (34)$$

The frequency response of the nearest neighbor kernel is [40]:

$$H(\omega) = \text{sinc}\left(\frac{\omega}{2}\right) \quad (35)$$

The kernel and its Fourier transform are shown in Fig. 13a. Convolution in the spatial domain with the rectangle function h is equivalent in the frequency domain to multiplication

with a sinc function [40]. Due to prominent side lobes and infinite extent, sinc function makes a poor low-pass filter. This technique achieves magnification by pixel replication, and magnification by sparse point sampling. Fig. 13b shows how the new brightness is assigned in nearest neighbor interpolation method. Dashed (- -) lines show how the inverse planar transformation map the grids of the output image into the input image (point sample or mean mapping image). Solid lines show the grid of the input image. The output pixel is assigned with the value of the pixel that the point falls within. No other pixels are considered.

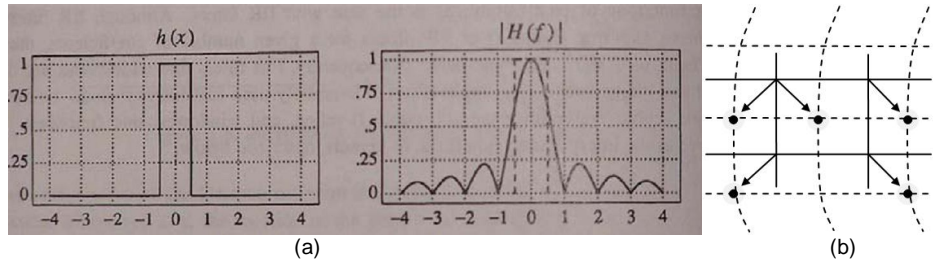


Fig. 13: (a) The nearest neighbor kernel and its F.T. [40] (b) Nearest neighbor interpolation

(ii) Bi-linear Interpolation

The second interpolation method discuss in this paper is bi-linear interpolation. Bi-linear interpolation explores four points neighboring the initial point, and assumes that the brightness function is bi-linear in this neighborhood. In the spatial domain, bilinear interpolation is equivalent to convolving the sampled input with the following kernel [40]:

$$h(x) = \begin{cases} 1-x & \text{if } x \in [0, 1] \\ x+1 & \text{if } x \in [-1, 0] \\ 0 & \text{otherwise} \end{cases} \quad (36)$$

The frequency response of the bilinear interpolation kernel is [40]:

$$H(\omega) = \text{sinc}^2\left(\frac{\omega}{2}\right) \quad (37)$$

The kernel and its Fourier transform are shown in Fig. 14a. Note that the frequency response of the bi-linear interpolation kernel is superior to the nearest neighbor interpolation since the side lobes are less prominent and the performance is improved in the stopband. A passband is moderately attenuated and this results in image smoothing. Fig 14b shows how the new brightness is assigned in bi-linear interpolation method. Dashed (- -) lines show how the inverse planar transformation map the grids of the output image into the input image (point sample or mean mapping image). Solid lines show the grid of the input image. The output pixel value is a weighted average of pixels in the nearest 2-by-2 neighborhood. Bi-linear interpolation can cause a small decrease in resolution and blurring due to its averaging nature. Bi-linear interpolation produces reasonably good results at moderate cost. But for better performance, more complicated algorithms are needed.

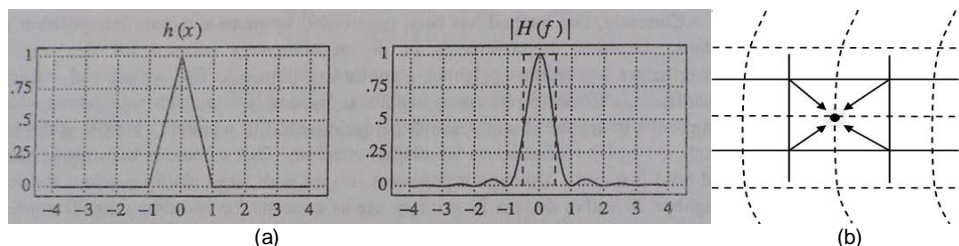


Fig. 14: (a) The bi-linear kernel and its F.T. [40] (b) Bi-linear interpolation

(iii) Bi-cubic Interpolation

Bi-cubic interpolation is a third degree interpolation algorithm that fairly well approximates the theoretically optimum sinc interpolation function. The interpolation kernel (named ‘Mexican hat’) is defined via [40]:

$$h(x) = \begin{cases} (a+x)|x|^3 - (a+3)|x|^2 + 1 & 0 \leq |x| < 1 \\ a|x|^3 - 5a|x|^2 + 8a|x| - 4a & 1 \leq |x| < 2 \\ 0 & 2 \leq |x| \end{cases} \quad (38)$$

Choice for a is -1 as suggested by [43] for bicubic interpolation, (38) becomes:

$$h(x) = \begin{cases} 1 - 2|x|^2 + |x|^3 & 0 \leq |x| < 1 \\ 4 - 8|x| + 5|x|^2 - |x|^3 & 1 \leq |x| < 2 \\ 0 & 2 \leq |x| \end{cases} \quad (39)$$

The frequency response of the kernel is [40]:

$$H(\omega) = \frac{12}{\omega^2} \left(\text{sinc}^2\left(\frac{\omega}{2}\right) - \text{sinc}(\omega) \right) + \frac{8}{\omega^2} \left(3\text{sinc}^2(\omega) - 2\text{sinc}(\omega) - \text{sinc}(2\omega) \right) \quad (40)$$

The kernel and its Fourier transform are shown in Fig. 15a. Note that the frequency response of the bi-cubic interpolation preserves fine details in the image very well. Fig 15b shows how the new brightness is assigned in bi-cubic interpolation method. Dashed (- -) lines show how the inverse planar transformation map the grids of the output image into the input image (point sample or mean mapping image). Solid lines show the grid of the input image. The output pixel value is a weighted average of pixels in the nearest 4-by-4 neighborhood. Bi-cubic interpolation does not suffer from the step like boundary problem of nearest neighborhood interpolation, and copes with bilinear interpolation blurring as well.

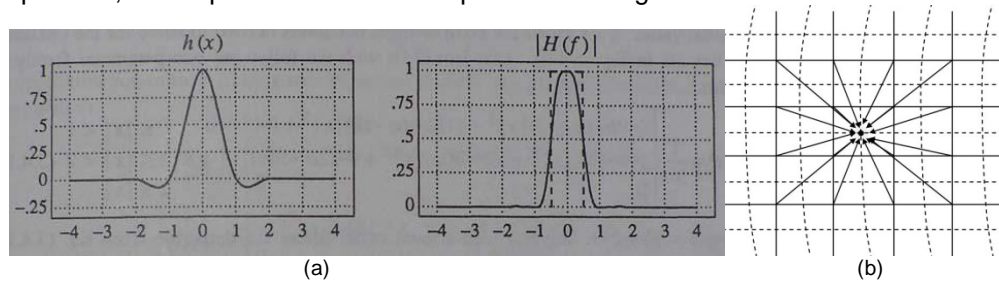


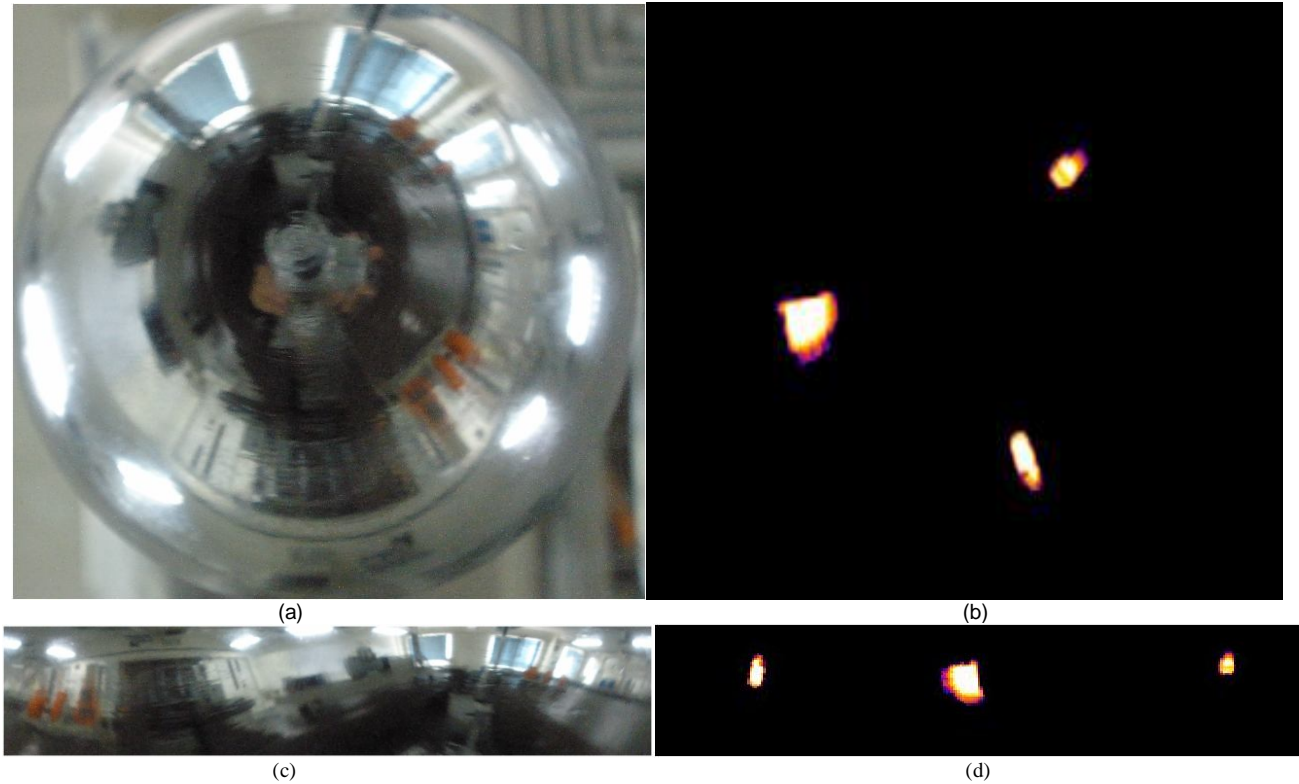
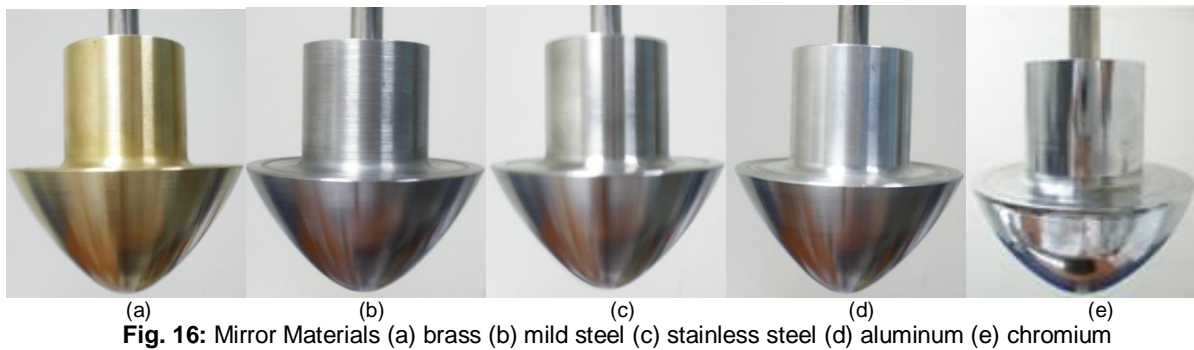
Fig. 15: (a) The bi-cubic kernel and its F.T. [40] (b) Bi-cubic interpolation

5. Experimental Results

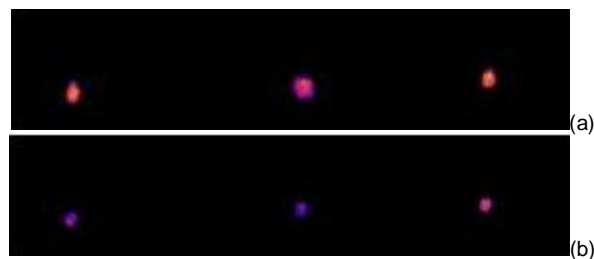
This section will briefly illustrate the materials selected to fabricate the IR omnidirectional hot mirror, calculation on the relative permittivity of each selected materials to test for IR reflection, comparisons on the performance of the unwarping method (pano-mapping vs. log-polar mapping) and the mapping methods use in log-polar mapping (point sample, mean and interpolation).

5.1 Mirror Materials

Five types of cost effective and common materials are chosen to fabricate the designed hot mirror as stated in section 3.1. They are brass, mild steel, stainless steel, aluminum, and chromium as shown in Fig. 16 (a, b, c, d and e respectively). Since fabricating mirror with whole chromium material is too expensive, it is substituted with aluminum coated with chromium. These fabricated hot mirrors are to be tested for reflectivity to find out the optimum one. A three machines' case is studied here and shown in Fig. 17. In Fig.17c, machine A (leftmost) and machine C (rightmost) are vibro test machines with same model and same specifications, whereas machine B (center) is a fatigue test machine. Three machines are considered function in normal when their motor's temperature is around 80°C after two hours or longer functioning time, and considered to be overheated when temperature reach 90°C.



The panoramic form of thermal images captured on different hot mirrors are shown in Fig. 18 (a, b, c, d and e respectively). The weights and temperature readings of different hot mirrors for a selected section (machine A select as random, results will be expected same if machine B or C is choose) are recorded as shown in Table 1. The contact temperature on machine A is also measured using TES 1310 digital thermometer manufactured by TES Electrical Electronic Corp.



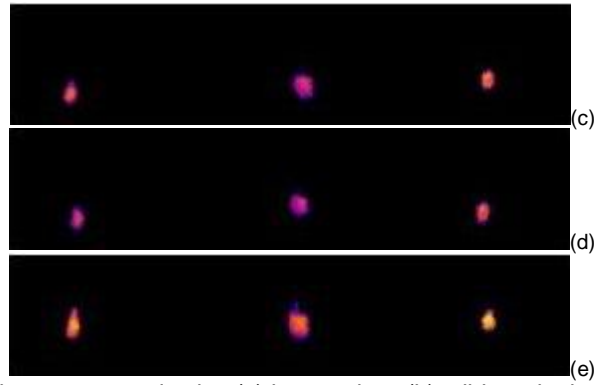


Fig. 18: Unwarp thermal image captured using (a) brass mirror (b) mild steel mirror (c) stainless steel mirror (d) aluminum mirror (e) chromium mirror on the same scene.

	Weight	Temperature reading on Mac. A	Relative reflectivity ϵ_r
Brass	1015g	74.2°C	0.9275
Mild steel	931g	70.8°C	0.8850
Stainless steel	940g	73.7°C	0.9213
Aluminum	325g	72.9°C	0.9113
Chromium	356g	78.8°C	0.9850
TES Thermometer	NIL	80°C	NIL

Table 1: weights, temperature readings and relative permittivity of different hot mirrors on a selected machine

The relative reflectivity for each hot mirror's is calculated using the relation ratio:

$$\epsilon_r = \frac{\text{Actual temperature}}{\text{IR reflected temperature}} \quad (41)$$

where Actual temperature is the temperature measured using TES thermometer and IR reflected temperature is the temperature reflected on the hot mirror captured using thermal camera. The closest the ϵ_r to 1, the better the IR reflectivity of a mirror. From the experimental results shown, chromium gives the highest ϵ_r , with the IR reflectivity of 98.5%. This follow by brass (92.75%), stainless steel (92.13%), aluminum (91.13%) and mild steel (88.5%) respectively. Chromium is adopted in the proposed omnidirectional thermal visualization system since it provides the best IR reflectivity and its moderate weight (356g, second lightest material among the tested).

5.2 Pano-mapping vs. Log-polar Mapping

An example of omnidirectional image is taken from [3] and its pano-mapping form is shown in Fig. 19a and Fig. 19c respectively. Fig. 19b shown an unwarp panoramic form of Fig. 19a using log-polar mapping. Comparing Fig. 19b and Fig. 19c, pano-mapping method expands the panoramic image into 4 times of that original omnidirectional image, whereas log-polar mapping method has higher compression capability upto 4 folds of the original omnidirectional image. Pano-mapping method required complicated landmark learning steps whereby it is not required in log-polar mapping method. Pano-mapping method also has high computational complexity which required upto 4th order calculation in each radial distance, whereas for log-polar mapping method, the

computational to calculate each radial distance ρ is relatively low, which required 1st order cosine

and sine function only. Since log-polar mapping has that much of merits in compare to pano-mapping, log-polar mapping is adopted as unwarping method in the proposed omnidirectional thermal visualization system.

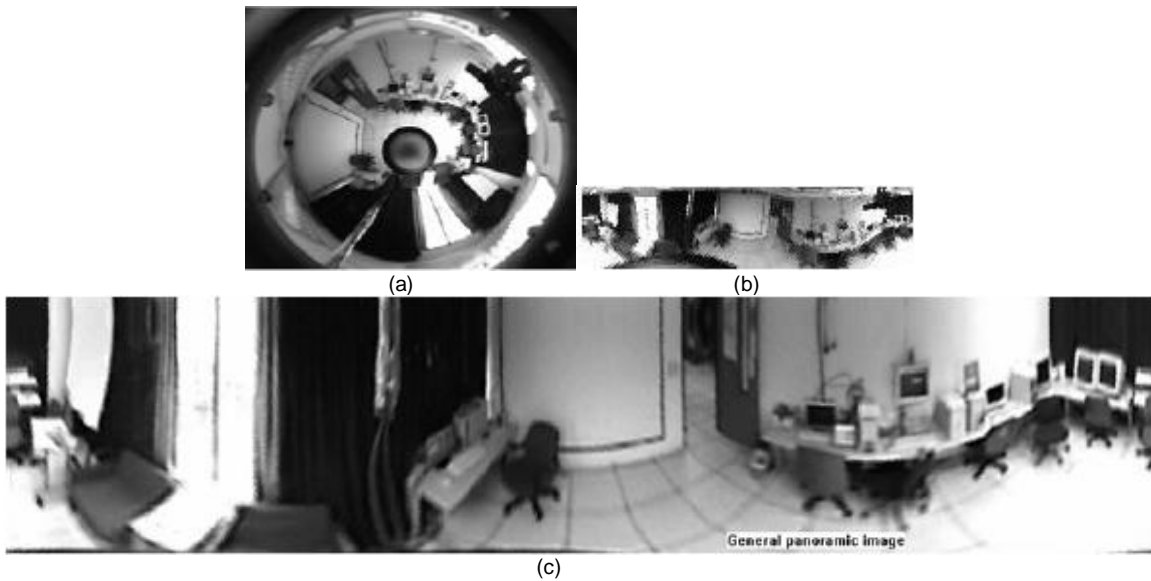


Fig. 19: (a) Omnidirectional image taken from [3] (b) Unwarp panoramic form of a using log-polar mapping (c) Unwarp panoramic form of a using pano-mapping

5.3 Different Mapping Method in Log-polar Mapping

This subsection investigates different mapping methods proposed in log-polar mapping. They are point sample, mean and interpolation mapping techniques as discussed in section 4.2. Fig. 17a shows a digital image captured using the proposed chromium hot mirror as shown in Fig. 16e. As can be observed in Fig. 17a, the mirror that good in IR range does not necessary good in human visual range as well. The mapping output of Fig. 17a is shown in Fig. 20 (a to h) using different mapping techniques (point sample, mean, nearest neighbor interpolation of point sample, nearest neighbor interpolation of mean, bi-linear interpolation of point sample, bi-linear interpolation of mean, bi-cubic interpolation of point sample and bi-cubic interpolation of mean respectively). The mapping output of Fig. 17b, which is the thermal image of Fig. 17a are shown in Fig. 21 (a to h), using different mapping techniques. For interpolation techniques, the number of pixels considered affects the complexity of the computation. Therefore the bi-cubic method (weighted average of 4-by-4 neighborhood) takes longer computation times than bi-linear method (weighted average of 2-by-2 neighborhood). And, the bi-linear method takes longer times than nearest neighbor (assigned value of pixel that point nearby). However, the greater the number of pixels is considered, the more accurate the effect is; and so there is a tradeoff between processing time and quality.





Fig. 20: Output of digital image using different mapping technique (a) point sample (b) mean (c) nearest neighbor interpolation of point sample (d) nearest neighbor interpolation of mean (e) bi-linear interpolation of point sample (f) bi-linear interpolation of mean (g) bi-cubic interpolation of point sample (h) bi-cubic interpolation of mean

The differences of mapping are obvious in quality for digital images but not significant in thermal images. Since interpolation and mean mapping techniques require high complexity as compare to point sample mapping technique, and all these mapping techniques provide almost the same image quality in unwarping thermal image, the less complexity technique is adopted, which is the point sample mapping technique(without interpolation).

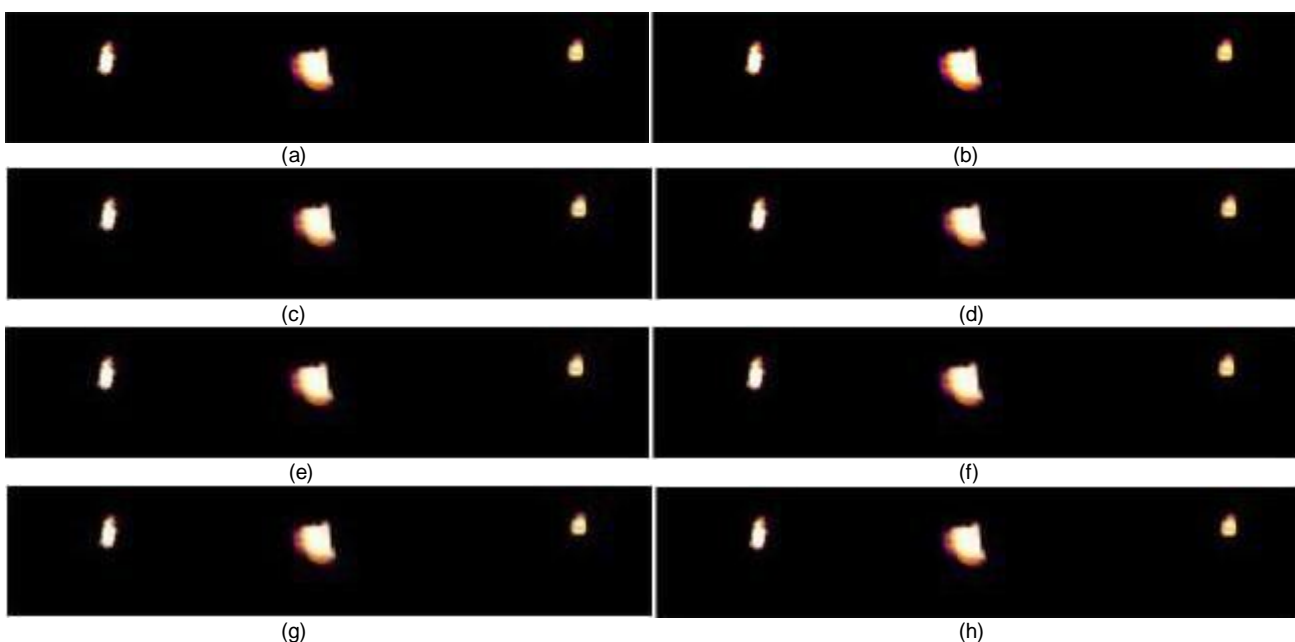


Fig. 21: Output of thermal image using different mapping technique (a) point sample (b) mean (c) nearest neighbor interpolation of point sample (d) nearest neighbor interpolation of mean (e) bi-linear interpolation of point sample (f) bi-linear interpolation of mean (g) bi-cubic interpolation of point sample (h) bi-cubic interpolation of mean

6. Conclusion

In this paper, a new optical approach based omnidirectional thermal visualization system using minimum hardware is proposed for wide angle visualization surveillance , condition monitoring and others even in poor lighting condition. IR reflected hot mirrors have been successfully designed and custom made for the purpose of reflecting 360 degree scene in IR range for the area to be captured by a single thermal camera, thus producing omnidirectional thermal visualization images. Five cost effective and market-common IR reflected materials are studied, i.e. brass, mild steel, stainless steel, aluminum and chromium. Among these materials, chromium gives the best reflectivity. The approach also adopt log-polar mapping and point sample mapping technique to unwarp the captured omnidirectional thermal image into panoramic form, providing observer or image processing tools a complete wide angle of view. In future, higher level of image processing capabilities will be embedded, such as algorithm development for trespasser detection, machine condition monitoring, surveillance and mobile navigation. More sets of omnidirectional thermal visualization system can also be networked for monitoring a factory/site

providing a new platform for remote monitoring using minimum hardware. All these will be addressed in the future work.

References:

- [1] J. Kumar, and M. Bauer, "Fisheye lens design and their relative performance", Proc. SPIE, Vol. 4093, p.p. 360-369, 2000.
- [2] T. Padjla and H. Roth, "Panoramic Imaging with SVAVISA Camera- Simulations and Reality", Research Reports of CMP, Czech Technical University in Prague, No. 16, 2000.
- [3] S. W. Jeng, and W. H. Tsai, "Using pano-mapping tables for unwarping of omni-images into panoramic and perspective-view images", IET Image Process., 1(2), p.p. 149-155, 2007.
- [4] F. Berton, A brief introduction to log-polar mapping, *Technical report*, LIRA-Lab, University of Genova, 10 Feb 2006.
- [5] T. Kawanishi, K. Yamazawa, H. Iwasa, H. Takemura and N. Yokoya, "Generation of High-resolution Stereo Panoramic Images by Omnidirectional Imaging Sensor Using Hexagonal Pyramidal Mirrors", Proc. 14th Int. Conf. in Pattern Recognition, Vol. 1, p.p. 485-489, Aug, 1998.
- [6] H. Ishiguro, M. Yamamoto, and S. Tsuji, "Omni-Directional Stereo", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 14, No. 2, p.p. 257-262, Feb 1992.
- [7] H-C. Huang and Y. P. Hung, "Panoramic Stereo Imaging System with Automatic Disparity Warping and Seaming", Graphical Models and Image Processing, Vol. 60, No. 3, p.p. 196-208, May 1998.
- [8] S. Peleg, and M. Ben-Ezra, "Stereo Panorama with a Single Camera", Proc. IEEE Conf. Computer Vision and Pattern Recognition, p.p. 395-401, June 1999.
- [9] H. Shum, and R. Szeliski, "Stereo Reconstruction from Multi-perspective Panoramas", Proc. Seventh Int. Conf. Computer Vision, p.p. 14-21, Sep 1999.
- [10] S. E. Chen, "Quick Time VR: An Image-Based Approach to virtual Environment Navigation", Proc. of the 22nd Annual ACM Conf. on Computer Graphics, p.p. 29-38, 1995.
- [11] S. J. Oh, and E. L. Hall, "Guidance of a Mobile Robot Using an Omnidirectional Vision Navigation System", Proc. of the Society of Photo-Optical Instrumentation Engineers, SPIE, 852, p.p. 288-300, Nov, 1987.
- [12] D. P. Kuban, H. L. Martin, S. D. Zimmermann and N. Busico, "Omniview Motionless Camera Surveillance System", United States Patent No. 5, 359, 363, Oct, 1994.
- [13] V. Nalwa, "A True Omnidirectional Viewer", Technical Report, Bell Laboratories, Holmdel, NJ07733, USA, Feb 1996.
- [14] <http://www.flirthermography.com>
- [15] J. Chahl, and M. Srinivasan, "Reflective surfaces for panoramic imaging", Applied Optics, 36(31), p.p. 8275-85, Nov, 1997.
- [16] S. Gachter, "Mirror Design for an Omnidirectional Camera with a Uniform Cylindrical Projection When Using the SVAVISA Sensor", Research Reports of CMP, OMNIVIEWS Project, Czech Technical University in Prague, No. 3, 2001. Redirected from: <http://cmp.felk.cvut.cz/projects/omniviews/>
- [17] T. Svoboda, "Central Panoramic Cameras Design, Geometry, Egomotion", PhD Thesis, Center of Machine Perception, Czech Technical University in Prague, 1999.
- [18] W. C. Robertson, "Stop Faking It! Light", National Science Teacher Association Press, 2003.
- [19] J. M. Enoch, "History of Mirrors Dating Back 8000 Years", Optometry and Vision Science, Vol. 83, Issue 10, p.p. 775-781, 2006.
- [20] P. Hadsund, "The Tin-Mercury Mirror: Its Manufacturing Technique and Deterioration Processes", Studies in Conservation, Vol. 38, No. 1, p.p. 3-16, 1993.
- [21] Answer Corporation (2009) redirected from <http://www.answers.com/topic/mirror>
- [22] J. Mohelnikova, "Materials for reflective coatings of window glass applications", Construction and Building Materials, Elsevier, Vol. 23, Issue 5, p.p. 1993-1998, 2009.
- [23] Evaporated Coating Inc., "Mirror Coatings for Plastic Optical Components" redirected from <http://www.evaporatedcoatings.com/pdf/800p.pdf>
- [24] J. D. Wolfe, "Durable silver mirror with ultra-violet thru far infra-red reflection" Agent: James S. Tak Attorney For Applicant - Livermore, CA, US USPTO Application #: 20060141272
- [25] X. H. Ying and Z. Y. Hu, "Catadioptric camera calibration using geometric invariants", IEEE Trans. Pattern Anal. Mach. Intell., 26(10), p.p. 1260-1271, 2004.
- [26] C. Geyer and K. Daniilidis, "Catadioptric camera calibration", Proc. IEEE 7th Int. Conf. Comput. Vis., p.p. 398-404, Sep, 1999.
- [27] T. Mashita, Y. Iwai and M. Yachida, "Calibration method for misaligned catadioptric camera", IEICE Trans. Inf. Syst., E89D, (7), p.p. 1984-1993, 2006.
- [28] D. Strelow, J. Mishler, D. Koes and S. Singh, "Precise omnidirectional camera calibration", Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR 2001), Kauai, HI, p.p. 689-694, Dec, 2001.
- [29] D. Scaramuzza, A. Martinelli and R. Siegwart, "A flexible technique for accurate omnidirectional camera calibration and structure from motion", Proc. Fourth IEEE Int. Conf. Comput. Vis. Syst., 2006.
- [30] G. Scotti, L. Marcenaro, C. Coelho, F. Selvaggi, and C. S. Regazzoni, "Dual camera intelligent sensor for high definition 360 degrees surveillance", IEE Proc. Vis. Image Signal Process., 152(2), p.p. 250-257, 2005.
- [31] L. R. Burden and J. D. Faires, "Numerical analysis", Brooks cole, Belmont, C. A., 2000, 7th edition, ISBN: 0534382169.
- [32] H. Araujo, J. M. Dias, "An Introduction To The Log-polar Mapping", *Proceedings of 2nd Workshop on Cybernetic Vision*, p.p. 139-144, 1996.

- [33] C. F. R. Weiman and G. Chaikin, "Logarithmic Spiral Grids For Image Processing And Display", *Computer Graphics and Image Processing*, Vol. 11, p.p. 197-226, 1979.
- [34] LIRA Lab, Document on specification, *Tech. report*, Espirit Project n. 31951 – SVAVISCA- available at <http://www.lira.dist.unige.it>.
- [35] R. Wodnicki, G. W. Roberts, and M. D. Levine, "A foveated image sensor in standard CMOS technology", *Custom Integrated Circuits Conf. Santa Clara*, p.p. 357-360, May 1995.
- [36] F. Jurie, "A new log-polar mapping for space variant imaging: Application to face detection and tracking", *Pattern Recognition, Elsevier Science*, 32:55, p.p. 865-875, 1999.
- [37] V. J. Traver, "Motion estimation algorithms in log-polar images and application to monocular active tracking", PhD thesis, Dep. Llenguatges I Sistemes Informatics, University Jaume I, Castellon (Spain), Sep, 2002.
- [38] M. Bolduc and M.D. Levine, "A review of biologically-motivated space variant data reduction models for robotic vision", *Computer Vision and Image Understanding*, Vol. 69, No. 2, p.p. 170-184, February 1998.
- [39] C.G. Ho, R.C.D. Young and C.R. Chatwin, "Sensor geometry and sampling methods for space variant image processing", *Pattern Analysis and Application*, Springer Verlag, p.p. 369-384, 2002.
- [40] P. Miklos, "Image Interpolation Techniques", 2nd Serbian-Hungarian Joint Symposium on Intelligent System (SISY 2004), p.p.1-6, 2004.
Directed from: <http://bmf.hu/conference/sisy2004/Poth.pdf>
- [41] Xin Li and Michael T. Orchard, "New Edge Directed Interpolation", *IEEE Int. Conf. on Image Processing*, Vancouver, Vol.2, p.p. 311-314, Sept. 2000.
- [42] Ron Kimmel, "Demosaicing: Image Reconstruction from Color CCD Samples," *IEEE Trans. on Image Processing*, Vol. 8, No. 9, pp. 1221-1228, September 1999.
- [43] <http://en.wikipedia.org/wiki/Interpolation>

CALL FOR PAPERS

Journal: International Journal of Image Processing (IJIP)

Volume: 4 **Issue:** 3

ISSN: 1985-2304

URL: <http://www.cscjournals.org/csc/description.php?JCode=IJIP>

About IJIP

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.

IJIP 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
- 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

- Image representation, sensing
- Implementation and architectures
- Materials for electro-photography
- New visual services over ATM/packet network
- Object modeling and knowledge acquisition
- Photographic emulsions
- Prepress and printing technologies
- Remote image sensing
- Storage and transmission
- analysis and recognition
- Imaging systems and image scanning
- Latent image
- Network architecture for real-time video transport
- Non-impact printing technologies
- Photoconductors
- Photopolymers
- Protocols for packet video
- Retrieval and multimedia
- Video coding algorithms and technologies for ATM/p

CFP SCHEDULE

Volume: 4

Issue: 4

Paper Submission: July 31 2010

Author Notification: September 1 2010

Issue Publication: September 2010

CALL FOR EDITORS/REVIEWERS

CSC Journals is in process of appointing Editorial Board Members for ***International Journal of Image Processing (IJIP)***. CSC Journals would like to invite interested candidates to join **IJIP** network of professionals/researchers for the positions of Editor-in-Chief, Associate Editor-in-Chief, Editorial Board Members and Reviewers.

The invitation encourages interested professionals to contribute into CSC research network by joining as a part of editorial board members and reviewers for scientific peer-reviewed journals. All journals use an online, electronic submission process. The Editor is responsible for the timely and substantive output of the journal, including the solicitation of manuscripts, supervision of the peer review process and the final selection of articles for publication. Responsibilities also include implementing the journal's editorial policies, maintaining high professional standards for published content, ensuring the integrity of the journal, guiding manuscripts through the review process, overseeing revisions, and planning special issues along with the editorial team.

A complete list of journals can be found at <http://www.cscjournals.org/csc/byjournal.php>. Interested candidates may apply for the following positions through <http://www.cscjournals.org/csc/login.php>.

Please remember that it is through the effort of volunteers such as yourself that CSC Journals continues to grow and flourish. Your help with reviewing the issues written by prospective authors would be very much appreciated.

Feel free to contact us at coordinator@cscjournals.org if you have any queries.

Contact Information

Computer Science Journals Sdn Bhd

M-3-19, Plaza Damas Sri Hartamas
50480, Kuala Lumpur MALAYSIA

Phone: +603 6207 1607
 +603 2782 6991
Fax: +603 6207 1697

BRANCH OFFICE 1

Suite 5.04 Level 5, 365 Little Collins Street,
MELBOURNE 3000, Victoria, AUSTRALIA


Fax: +613 8677 1132

BRANCH OFFICE 2

Office no. 8, Saad Arcad, DHA Main Bulevard
Lahore, PAKISTAN

EMAIL SUPPORT

Head CSC Press: coordinator@cscjournals.org
CSC Press: cscpress@cscjournals.org
Info: info@cscjournals.org



COMPUTER SCIENCE JOURNALS SDN BHD
M-3-19, PLAZA DAMAS
SRI HARTAMAS
50480, KUALA LUMPUR
MALAYSIA