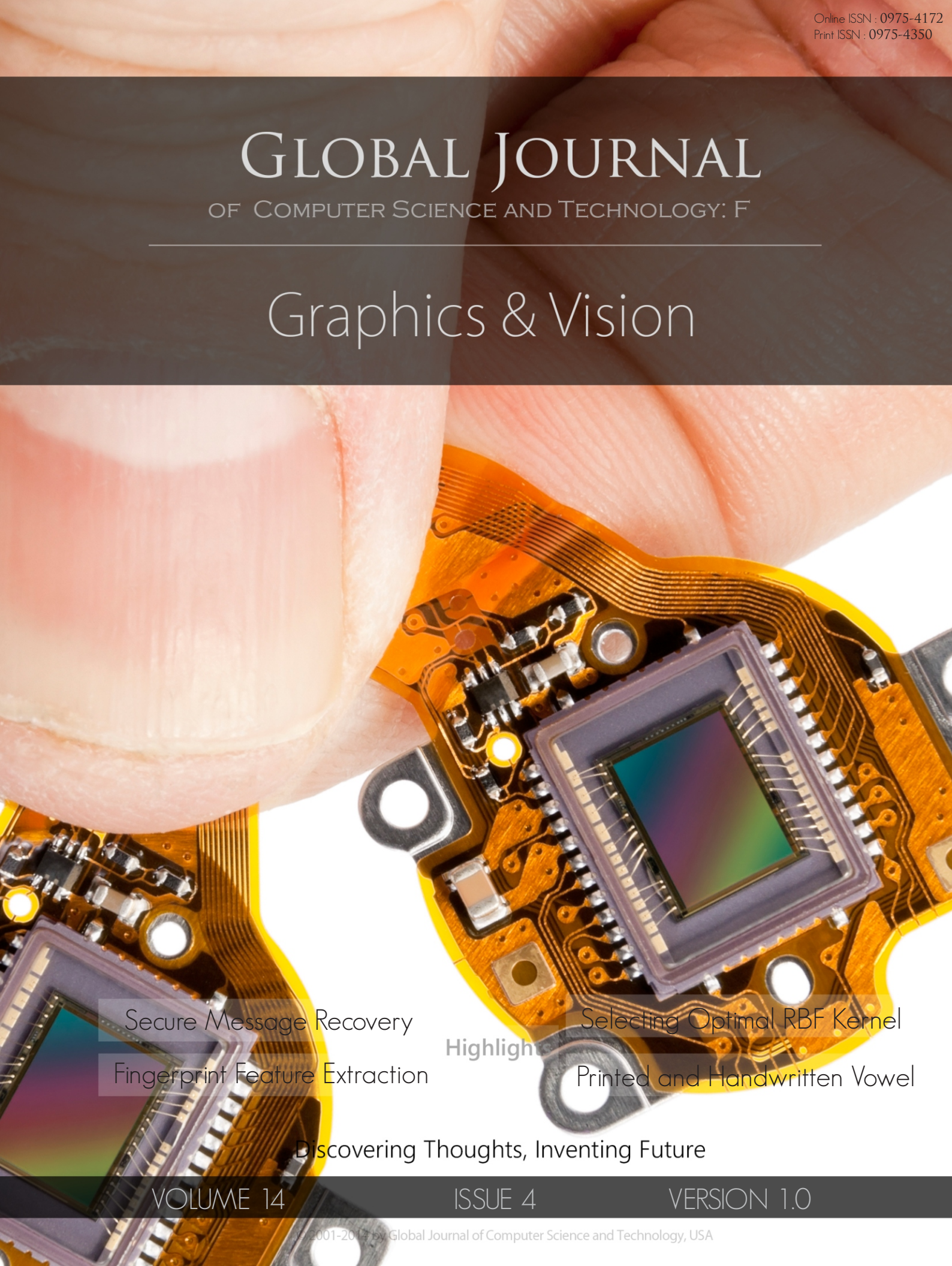
Online ISSN : 0975-4172
Print ISSN : 0975-4350

# Global Journal

## of Computer Science and Technology: F

# Graphics & Vision



Secure Message Recovery

Fingerprint Feature Extraction

Highlights

Selecting Optimal RBF Kernel

Printed and Handwritten Vowel

Discovering Thoughts, Inventing Future

VOLUME 14 ISSUE 4 VERSION 1.0

Global Journal of Computer Science and Technology: F
Graphics & Vision

Global Journal of Computer Science and Technology: F
Graphics & Vision

Volume 14 Issue 4 (Ver. 1.0)

Open Association of Research Society

**Incorporation No.: 0423089**
**License No.: 42125/022010/1186**
**Registration No.: 430374**
**Import-Export Code: 1109007027**
**Employer Identification Number (EIN):**
**USA Tax ID: 98-0673427**

# Global Journals Inc.

*(A Delaware USA Incorporation with "Good Standing"; **Reg. Number: 0423089**)*

*Sponsors:* *Open Association of Research Society*
*Open Scientific Standards*

## *Publisher's Headquarters office*

Global Journals Headquarters
301st Edgewater Place Suite, 100 Edgewater Dr.-Pl,
**Wakefield MASSACHUSETTS,** Pin: 01880,
United States of America
*USA Toll Free: +001-888-839-7392*
*USA Toll Free Fax: +001-888-839-7392*

## *Offset Typesetting*

Global Journals Incorporated
2nd, Lansdowne, Lansdowne Rd., Croydon-Surrey,
Pin: CR9 2ER, United Kingdom

## *Packaging & Continental Dispatching*

Global Journals
E-3130 Sudama Nagar, Near Gopur Square,
Indore, M.P., Pin:452009, India

## *Find a correspondence nodal officer near you*

To find nodal officer of your country, please email us at *local@globaljournals.org*

## *eContacts*

Press Inquiries: *press@globaljournals.org*
Investor Inquiries: *investors@globaljournals.org*
Technical Support: *technology@globaljournals.org*
Media & Releases: *media@globaljournals.org*

## *Pricing (Including by Air Parcel Charges):*

*For Authors:*
22 USD (B/W) & 50 USD (Color)
*Yearly Subscription (Personal & Institutional):*
200 USD (B/W) & 250 USD (Color)

# Integrated Editorial Board
## (Computer Science, Engineering, Medical, Management, Natural Science, Social Science)

**Dr. Bart Lambrecht**
Director of Research in Accounting and FinanceProfessor of Finance
Lancaster University Management School
BA (Antwerp); MPhil, MA, PhD (Cambridge)

**Dr. Carlos García Pont**
Associate Professor of Marketing
IESE Business School, University of Navarra
Doctor of Philosophy (Management), Massachusetts Institute of Technology (MIT)
Master in Business Administration, IESE, University of Navarra
Degree in Industrial Engineering, Universitat Politècnica de Catalunya

**Dr. Fotini Labropulu**
Mathematics - Luther College
University of ReginaPh.D., M.Sc. in Mathematics
B.A. (Honors) in Mathematics
University of Windso

**Dr. Lynn Lim**
Reader in Business and Marketing
Roehampton University, London
BCom, PGDip, MBA (Distinction), PhD, FHEA

**Dr. Mihaly Mezei**
ASSOCIATE PROFESSOR
Department of Structural and Chemical Biology, Mount Sinai School of Medical Center
Ph.D., Etvs Lornd University
Postdoctoral Training,
New York University

**Dr. Söhnke M. Bartram**
Department of Accounting and FinanceLancaster University Management SchoolPh.D. (WHU Koblenz)
MBA/BBA (University of Saarbrücken)

**Dr. Miguel Angel Ariño**
Professor of Decision Sciences
IESE Business School
Barcelona, Spain (Universidad de Navarra)
CEIBS (China Europe International Business School).
Beijing, Shanghai and Shenzhen
Ph.D. in Mathematics
University of Barcelona
BA in Mathematics (Licenciatura)
University of Barcelona

**Philip G. Moscoso**
Technology and Operations Management
IESE Business School, University of Navarra
Ph.D in Industrial Engineering and Management, ETH Zurich
M.Sc. in Chemical Engineering, ETH Zurich

**Dr. Sanjay Dixit, M.D.**
Director, EP Laboratories, Philadelphia VA Medical Center
Cardiovascular Medicine - Cardiac Arrhythmia
Univ of Penn School of Medicine

**Dr. Han-Xiang Deng**
MD., Ph.D
Associate Professor and Research
Department Division of Neuromuscular Medicine
Davee Department of Neurology and Clinical NeuroscienceNorthwestern University
Feinberg School of Medicine

**Dr. Pina C. Sanelli**
Associate Professor of Public Health
Weill Cornell Medical College
Associate Attending Radiologist
NewYork-Presbyterian Hospital
MRI, MRA, CT, and CTA
Neuroradiology and Diagnostic Radiology
M.D., State University of New York at Buffalo,School of Medicine and Biomedical Sciences

**Dr. Roberto Sanchez**
Associate Professor
Department of Structural and Chemical Biology
Mount Sinai School of Medicine
Ph.D., The Rockefeller University

**Dr. Wen-Yih Sun**
Professor of Earth and Atmospheric SciencesPurdue University Director
National Center for Typhoon and Flooding Research, Taiwan
University Chair Professor
Department of Atmospheric Sciences, National Central University, Chung-Li, TaiwanUniversity Chair Professor
Institute of Environmental Engineering, National Chiao Tung University, Hsin-chu, Taiwan.Ph.D., MS The University of Chicago, Geophysical Sciences
BS National Taiwan University, Atmospheric Sciences
Associate Professor of Radiology

**Dr. Michael R. Rudnick**
M.D., FACP
Associate Professor of Medicine
Chief, Renal Electrolyte and Hypertension Division (PMC)
Penn Medicine, University of Pennsylvania
Presbyterian Medical Center, Philadelphia
Nephrology and Internal Medicine
Certified by the American Board of Internal Medicine

**Dr. Bassey Benjamin Esu**
B.Sc. Marketing; MBA Marketing; Ph.D Marketing
Lecturer, Department of Marketing, University of Calabar
Tourism Consultant, Cross River State Tourism Development Department
Co-ordinator , Sustainable Tourism Initiative, Calabar, Nigeria

**Dr. Aziz M. Barbar, Ph.D**.
IEEE Senior Member
Chairperson, Department of Computer Science
AUST - American University of Science & Technology
Alfred Naccash Avenue – Ashrafieh

# Contents of the Issue

Global Journal of Computer Science and Technology: F
Graphics & Vision
Volume 14 Issue 4 Version 1.0 Year 2014
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

# Log-Gabor Orientation with Run-Length Code based Fingerprint Feature Extraction Approach

By Dr. K. Kanagalakshmi & Dr. E. Chandra
*SNS Rajalakshmi College, India*

*Abstract-* This paper aims to design and implement Log-Gabor filtering with Run-length Code based feature Extraction technique. Since minutiae extraction is an essential and core process of fingerprint Identification and Authentication systems, the minutiae features are enhanced in each orientation using Log-Gabor filter and features are extracted using the proposed method. Frequency domain is derived using FFT and they are enhanced by Log-Gabor filter for each orientation. In our method six orientations are considered; binarization, thinning are also followed. Fingerprint features are extracted using proposed method which possesses labeling and Run-length Coding technique. Our method is tested with the benchmark Databases and real time images and the results show the better performance and lower error rate.

*Keywords:* *FFT, log-gabor, minutiae, orientation, frequency.*

*GJCST-F Classification :* *D.4.6*


*Strictly as per the compliance and regulations of:*

# Log-Gabor Orientation with Run-Length Code based Fingerprint Feature Extraction Approach

Dr. K. Kanagalakshmi [α] & Dr. E. Chandra [σ]

*Abstract-* This paper aims to design and implement Log-Gabor filtering with Run-length Code based feature Extraction technique. Since minutiae extraction is an essential and core process of fingerprint Identification and Authentication systems, the minutiae features are enhanced in each orientation using Log-Gabor filter and features are extracted using the proposed method. Frequency domain is derived using FFT and they are enhanced by Log-Gabor filter for each orientation. In our method six orientations are considered; binarization, thinning are also followed. Fingerprint features are extracted using proposed method which possesses labeling and Run-length Coding technique. Our method is tested with the benchmark Databases and real time images and the results show the better performance and lower error rate.

*Keywords: FFT, log-gabor, minutiae, orientation, frequency.*

## I. Introduction

The accuracy of the Fingerprint matching process firmly depends on the feature extraction phase. The true minutiae alone lead the matching process successfully; and reduce the FRR (False Rejection Rate) and FAR (False Acceptance Rate). The Human fingerprint consists of various types of features that are the ridge patterns, traditionally classified according to the decade's old Hendry system: Left loop, Right loop, Arch, Whorl and Tented arch. Fingerprint features are classified into three levels. Level 1 Features are: Arch, Tented arch, Right loop, Left loop, double loop and Whorl. The Level 2 features are Line-unit, Line-fragment, Ending, Bifurcation, Eye and Hook. The Level 3 features are Pores, line shape, incipient ridges, creases, warts and scars [1]. The statistical analysis shows that the level-1 features are not unique but are useful for fingerprint classifications. The level 2 features are having adequate sharp power, used to establish the individuality of fingerprint [2]. Likewise, the level 3 features are permanent, immutable and unique according to the forensic experts. It also can offer discriminatory information for human identification [1]. Rest of the paper comprises of six sections. Section 2 specifies different feature extraction techniques. In section 3, proposed method is described. Section 4 tabularizes the benchmarks used. Experimental tasks and results are provided in section 5. Section 6 concludes the paper.

*Author α: Associate Professor, PG Department of Computer Applications, Vidyasagar College of Arts and Science, Udumalpet, Tiruput(DT), Tamilnadu, India.*
*e-mail: kkanagalakshmi@gmail.com*
*Author σ: Director, Dr. SNS Rajalakshmi College, Coimbatore.*

## II. Background Work

Feature extraction techniques are classified in to different ways; According to image: Direct gray level minutiae extraction, Binary minutiae extraction; According to Levels of features: Level 1 feature extraction, Level 2 feature extraction, Level 3 feature extraction. We desire to extract Binary fingerprint- Level 2 features which posses an adequate and sharp power. Yet other feature extraction techniques are surveyed. They are:

- Orientation or Directional Mapping method
- Gabor-Filter Bank and Orientation method
- Run-Length Coding method
- Chain Code method
- Crossing Number method
- Morphology based method

Run-Length Coding (RLC) method [3, 4] is effective when long sequence of the same symbol occurs. RLC uses the Scan line procedure to extract features. Nalini K. Ratha et al. [5] designed an adaptive flow orientation based feature extraction method to extract binary fingerprint features and also used a waveform projection based ridge segmentation algorithm to locate ridges accurately. Chih-Jen Lee et al. [6] proposed a Gobor-Filter based method for fingerprint recognition. The Gabor-filter based features can also be used for the process of local ridge orientation, core point detection and features extraction. Jain et al. [7] suggested the multichannel approach using Gabor filter for the classification of fingerprints features. Wan S [8] proposed a method based on directional fields of fingerprint image to detect the singular points (cores) and extract features. Neil Yager [9] distinguished the fingerprint features in to different classes. Orientations fields and Gabor-filtering are influential means for classifications of fingerprints features. Classifications and identification of fingerprint features are used for the recognition and features extraction. Sharat Chikkerur et al. [10] proposed an approach of Orientation Map for fingerprint image feature extraction. Feng Zhoo et al. [11] used Crossing Number (CN) method to extract minutiae from the Valley skeleton binary image. The Orientation Maps and Gabor filters are good in fingerprint feature extraction [12]. We

propose a hybrid approach based on Log-Gabor Orientation with RLC method to get accurate minutiae.

## III. Proposed Method

The Proposed method includes three main stages: Image preprocessing, Enhancement and Minutia Extraction and Post processing. The System level design is shown in figure 1.


*Figure 1 :* System Level Design

The flow graph notation of the proposed method is shown in figure 2.

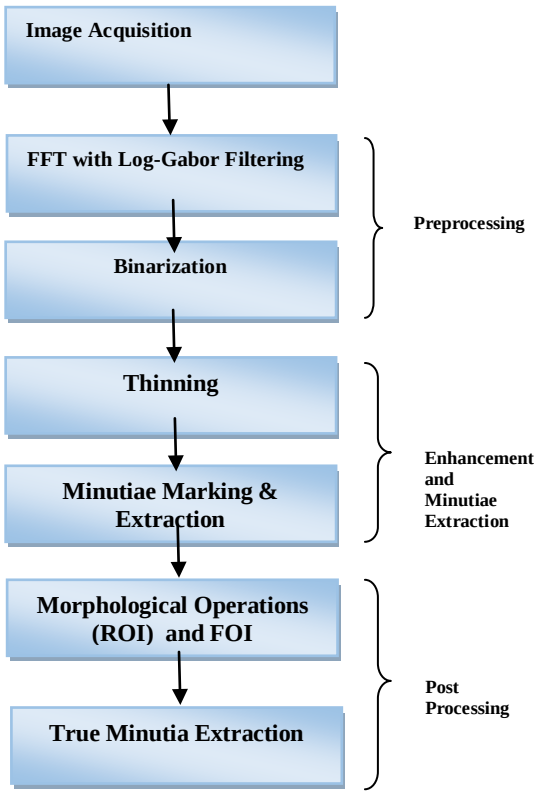



*Figure 2 :* Algorithm Level design

### *a) Preprocessing*

The preprocessing stage includes the Fourier Transformation and the filtering using Log-Gabor Filter followed by Binarization. Steps followed in the first stage are described below.

i. *Image acquisition*

The first step of the algorithm is the image acquisition. The images are acquired from benchmark data sets and also real time fingerprint images.

ii. *FFT and Log-Gabor Filtering*

The image enhancement can be carried out in either spatial or frequency domain. The Log-Gabor can provide a better enhancement on any kind of images with its good smoothening characteristic based on performance and quality measures which were empirically observed [13]. The frequency domain enhancement is carried out for our succeeding work. The frequency values are obtained through the Fast Fourier Transformation. It transforms the image into a frequency image; next the Log-Gabor filter parameters are defined and the Orientations are estimated (six orientations are considered). In this stage, Log-Gabor features and Local ridge orientations are also calculated.

*a. Gabor Features*

The 2-D Gabor filters general format is defined in [14, 15] as follows:

$$G(x,y,\theta_k,f,\theta_k,\sigma_x,\sigma_y) = e^{\left[-\frac{1}{2}\left(\frac{x_\theta 2}{\sigma_x 2}+\frac{y_\theta 2}{\sigma_y 2}\right)\right]} \times e^{i2\pi f x_{\theta_k}} \quad (1)$$

where $x_{\theta_k} = x\cos\theta_k + y\sin\theta_k$ and $y_{\theta_k} = -x\sin\theta_k + y\cos\theta_k$, f is the frequency of the sinusoidal plane wave, $\theta_k$ is the orientation of the Gabor filter, and $\sigma_x, \sigma_y$ are the standard deviations of the Gaussian envelop along the x and y axes, respectively.

The complex form of the eqn. (1) can be expressed as follows:

$G=g_{even} + i\, g_{odd}$ where

$$g_{even}(x,y,\theta_k,f,\theta_k,\sigma_x,\sigma_y) = e^{\left[-\frac{1}{2}\left(\frac{x_\theta 2}{\sigma_x 2}+\frac{y_\theta 2}{\sigma_y 2}\right)2\right]} \times \cos(2\pi f x_{\theta_k}) \quad (2)$$

$$g_{odd}(x,y,\theta_k,f,\theta_k,\sigma_x,\sigma_y) = e^{\left[-\frac{1}{2}\left(\frac{x_\theta 2}{\sigma_x 2}+\frac{y_\theta 2}{\sigma_y 2}\right)2\right]} \times \sin(2\pi f x_{\theta_k}) \quad (3)$$

While most local ridge structures of fingerprint images are with the well defined frequency and orientations, *f* can be set by the reciprocal of the average inter ridge distance and *n* as the number of orientations for calculating $\theta_k = \pi(k-1)/n$ , k=1, 2,….n; and the cosine and sine form and the sinusoidal shape of the Gabor filter is suitable for modeling ridge structures and smoothing noise, respectively. To reduce the complexity of the Gabor equations and make computation faster, we followed the Log-Gabor method with the modified versions of equation 1. The Log-Gabor expression is given below [13].

$$LGF(f) = LG(f) \times FC \quad (4)$$

where LG(f) is the Log-Gabor Radial Component and FC is the angular component. The radial component controls the frequency band and the angular component controls the orientation.

$$\boldsymbol{LG(F) = e^{\left(-\frac{log\left(\frac{r}{rf_0}\right)}{2\log\left(\frac{\sigma}{rf_0}\right)}\right)}} \quad (5)$$

$$\boldsymbol{FC = e^{\left(\frac{-d\theta^2}{2\theta\sigma^2}\right)}} \quad (6)$$

where *r* is the normalized radius from centre, $rf_o$ is the normalized radius from centre of frequency plane corresponding to the wavelength and $d\theta$ is an angular distance of sin and cosine. From the product of the eqn. 5 and 6 the Log-Gabor filter is derived.

*b. Local Ridge Orientation*

The ridge orientation (θ) is computed using the Log-Gabor features as follows:

$$\boldsymbol{\theta = \frac{\Sigma_{k=1}^{n} LG_{\theta_k} \times \theta_k}{\Sigma_{k=1}^{n} LG_{\theta_k}}} \quad (7)$$

Where $\theta_k = \pi(k-1)/n$ , k=1,2….n, and $G_{\theta_k}$ Log-Gabor features.

Log-Gabor Filter is applied on the frequency domain with six orientations ($\theta_1$ – $\theta_6$) in order to eliminate the noise and also enhance the frequency values of an image; and through which even negative frequency values are enhanced [13].

iii. *Binarization*

Binarization is the process of converting the gray-level image [0-255] to binary image [0 or 1]. New value (0 or 1) can be assigned for each pixel according to the intensity mean in a local neighborhood, as follows:

$$I_{new}(p1,p2) = \begin{cases} \boldsymbol{1\ if\ I_{old}(p1,p2 \geq local\ Mean)} \\ \boldsymbol{0\ otherwise} \end{cases} \quad (8)$$

The gray-scale transformations do not depend on the position of the pixel in the image. During the binarization process, the low frequency pixels are omitted [16]. For the binarization process, the Log-Gabor filtered image is used.

*b) Fingerprint Image Enhancement and Minutiae Extraction*

Before extracting minutiae, the fingerprint image is enhanced to get compatible patterns of features. This stage includes three main steps: Thinning, Minutiae Marking (FOI: Feature of Interest) and Extracting minutiae sets.

i. *Thinning*

In order to get skeleton of the fingerprint image, thinning process is followed. A Skeleton is a one-pixel wide ridge [17]. Thinning is a process of translating the thickness of an image into one pixel width representation. From thinning process, thinned and sharp ridges of fingerprint features are derived. It gives a clear structure of the fingerprint image.

The thin operation which we implemented uses the following algorithm [18, 19].

*Step 1:* Divide the image into two distinct subfields in a checkerboard pattern.

*Step 2:* Delete pixel p from the first subfield if and only if the conditions G1, G2, and G3 are all satisfied in the first iteration.

*Step 3:* Delete pixel p from the second subfield if and only if the conditions G1, G2 and G3` are all satisfied during the second sub- iteration.

*Condition G1:*

$$X_H(P) = 1 \tag{9}$$

where $X_H(P) = \sum_{i=1}^{4} bi$

$$b_i = \begin{cases} 1, if\ \ x_{2i-1} = 0\ and\ (x_{2i=1}\ \ or\ x_{2i+1} = 1) \\ 0, \qquad\qquad otherwise. \end{cases}$$

$x_1, x_2, \ldots, x_8$ are the values of the eight neighbors of p, starting with the least neighbor and numbered n counter-clockwise order. Fig. 3 shows the neighbors of p in a checkerboard format.

| | | |
|---|---|---|
| $x_4$ | $x_3$ | $x_2$ |
| $x_5$ | P | $x_1$ |
| $x_6$ | $x_7$ | $x_8$ |

*Figure 3 :* Pixels of N (P)

*Condition G2:*

$$2 \leq \min\{n1(p), n2(p)\} \leq 3 \tag{10}$$

where

$$n1(p) = \sum_{i=1}^{4} x_{2k-1} \vee x_{2k}$$

$$n2(p) = \sum_{i=1}^{4} x_{2k} \vee x_{2k+1}$$

*Condition G3:*

$$\text{G3: } (x_2 \vee x_3 \vee \overline{x_8}) \wedge x_1 = 0 \tag{11}$$

G3 is in the first sub-iteration.

Condition G3`:

$$\textit{G3`: } (x_6 \vee x_7 \vee \overline{x_4}) \wedge x_5 = 0 \tag{12}$$

G3`: 180° rotation in the second.

The given two subscriptions together make an iteration of the thinning algorithm. These iterations are repeated until the specified time. We set it as infinite number of iterations (n='inf'). Therefore, the iterations are repeated until the image stops changing. The conditions are all tested using the pre-computed look up tables.

ii. *Minutia Marking*

In our work, Level 2 features: Terminations and Bifurcations are used to extract. Features are marked using labeling technique and also Run-length Coding algorithm [3]. The algorithm to find the minutiae is given below.

*Step1:* Run-Length Encoding the input image (RLE).

*Step 2:* Scan the runs; assigning preliminary labels for connected components in binary image.

*Step 3:* Determine the equivalence classes(c).

*Step 4:* Concatenate all relevant classes.

*Step 5:* Re-label the runs based on the determined equivalence classes (LB(c)).

Marking or Labeling of connected components is one of the most main operations in pattern recognition. It is essential when an object gets recognized [20]. The proposed algorithm includes scanning, labeling, and determine the equivalence classes of minutiae in order to group and concatenate the relevant classes. Finally, re-labeling of the scanned runs based on the results determined equivalence classes. Our algorithm uses the skeleton image where the ridge flow is 8-connected. The minutiae which are marked (labeled) by scanning the local neighborhood of each ridge pixels in the fingerprint image using (3×3) non-overlapping windows. Based on the label values LB, the ridge pixels are classified into Terminations and Bifurcations. If the pixel is labeled with 0 then it is determined as Isolation. If the pixel is labeled with 1, 2 then it is determined as Termination and Continuing Terminations respectively; and if the pixel is labeled as 3, 4 then it is determined as Bifurcation, Crossing respectively (see Table 1). The templates of the Termination and bifurcations are shown in fig.4.

*Table 1 :* Property of the Label

| Label (LB) | Property |
|---|---|
| 0 | Isolation |
| 1 | Termination |
| 2 | Continuing Termination |
| 3 | Bifurcation |
| 4 | Crossing Point |

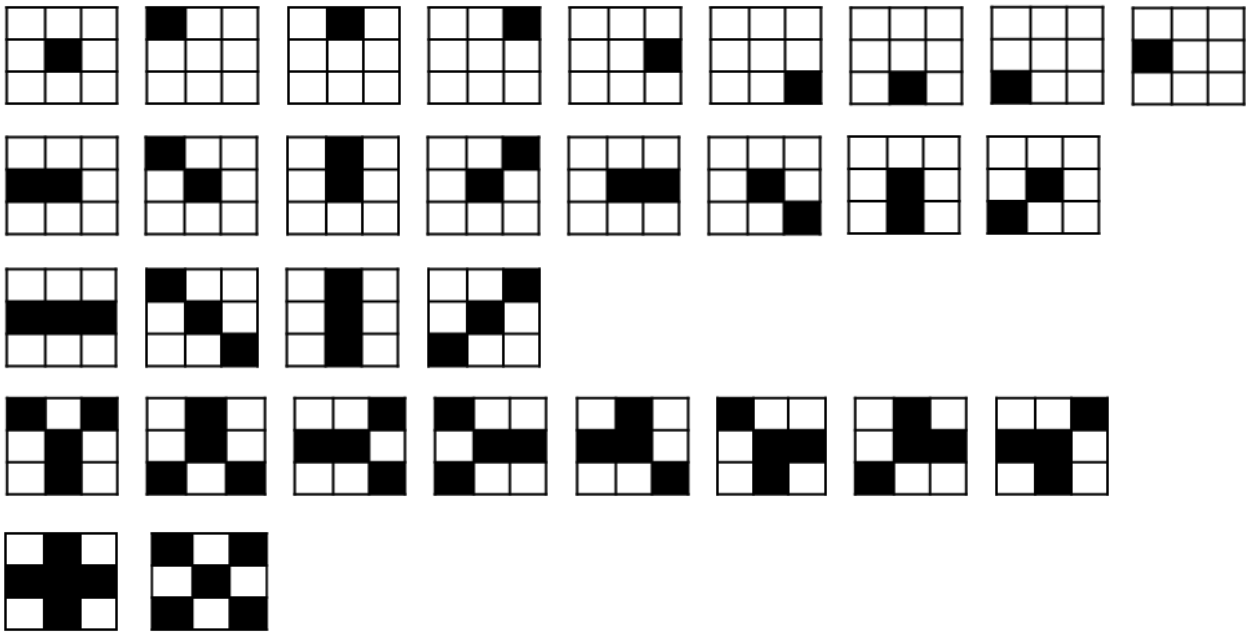


*Figure 4 :* Templates of Isolation (Row 1), Terminations (Row 2), Continuing Terminations (Row 3), Bifurcations (Row 4), and Crossing (Row 5)

iii. *Minutiae Extraction*

Minutiae extraction depends on the labels and properties of the marked minutiae (as defined in table 1). Based on the properties, the ridge terminations and bifurcations are extracted from the fingerprint image (see fig. 5). The red circle refers the ridge endings and the green square refers the ridge bifurcations.

*c) Minutiae Post-processing*

Post processing of minutiae extraction is a vital process to get true minutiae. Since the image comes across different stages of processing, there exist some spurious points in the image. It causes the inaccuracy of minutiae. Hence, the post processing is essential to remove spurs, H-points, break points, closed and border minutia. The main objectives of post process are to remove false minutiae and retain only true minutiae set. In this part, the morphological operations such as setting Region of Interest (ROI), closing and opening are performed. In addition to that the distance between two endpoints is calculated and compared with the threshold value. If they are equal then it considered as true minutia otherwise false. From the post processing, the true minutia set are extracted.

## IV. Benchmarks

In order to check and compare the proposed method, publicly available fingerprint database for FVC in 2000,2002, 2004 and real time data base have been chosen. Each database contains 880(Set A: 100×8, Set B: 10×8)) fingerprints, original from 50 different fingers of untrained volunteers. The same finger is needed to give 5 impressions. Properties of all the selected image databases both public and real time are shown in the table 2.

*Table 2 :* Bench mark Database and Real Time data sets

| Database Name | Sensor Types | Size of the Image | Resolution in dpi |
|---|---|---|---|
| 2000 DB1 | Low-cost Optical Sensor | 300×300 | 500 |
| 2000 DB2 | Low-cost Capacitive Sensor | 256×364 | 500 |
| 2000 DB3 | Optical Sensor | 448×478 | 500 |
| 2002 DB1 | Optical Sensor | 388×374 (142 Kpixels) | 500 |
| 2002 DB2 | Capacitive sensor | 296×560(162 Kpixels) | 569 |
| 2002 DB3 | Capacitive Sensor | 300×300(88 Kpixels) | 500 |
| 2004 DB1 | Optical Sensor | 640×480(45 Kpixels) | 500 |
| 2004 DB2 | Optical Sensor | 328×480(100 Kpixels) | 500 |
| 2004 DB3 | Thermal Sweeping Sensor | 300×480(56 Kpixels) | 500 |
| Real - Time DB | Optical Sensor | 300×300 | 500 |

## V. Experimental Results and Discussions

The proposed algorithm is implemented in MATLAB 7.10 with the standard benchmarks specified in the section 4. The experimental results are shown in fig. 5. The results show the novelty while extracting minutia. From the first step, fingerprint image is captured and then preprocessing stage is carried out; in this stage, the frequency domain enhancement is followed in order to get frequency value. In the second stage, minutia extraction is performed. To eliminate the false minutia, the post processing is also followed thirdly; the extracted minutia set is under the post process. Finally, the true minutiae set are obtained.

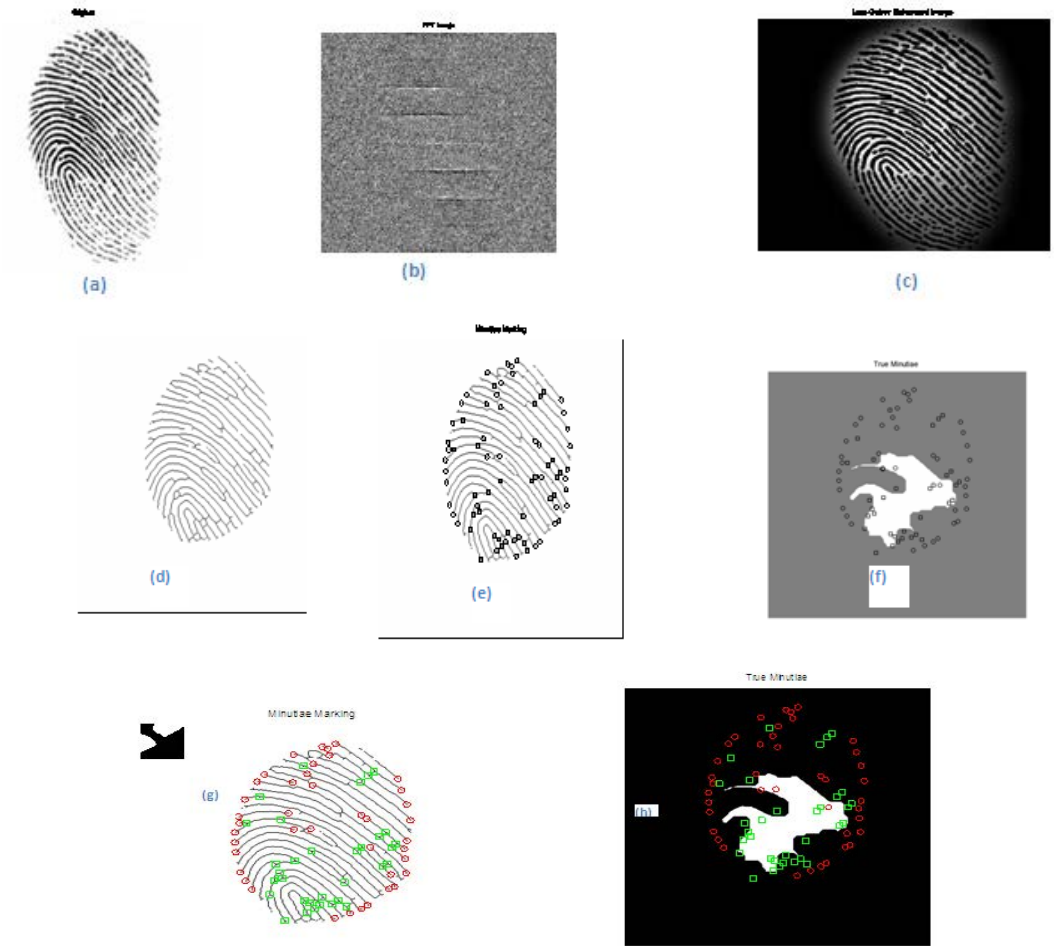



*Figure 5* : Minutiae Extraction stages: (a)Origianl image(101_1.tif) (b) FFT Image (c) Log-Gabor filterd image (d) Enhanced cum thinned image (e) Minutiae Marking (before Post Processing) (f) True Minutiae set (after post processing) (g) and (h) are the color versions of the (e),(f) respectively

Table 3 lists the mean noise for each orientation. The accuracy rates of the proposed algorithm on minutiae before and after pre, post processing are reported in table 4, 5, and 6 respectively. In those tables, the accuracy rate of terminations and bifurcations are computed by the Tt /Te and Bt/Be, respectively. The total accuracy rate is also computed using the following formula [11]:

$$\text{Total Accuracy Rate} = \frac{T_t + B_t}{T_e + B_e} \quad (13)$$

where Tt and Bt are the number of true terminations and true bifurcations and Te , Be are the extracted terminations and extracted bifurcations respectively.

*Table 3 :* Orientations and mean noise

| Image # | Mean Noise in each orientation (1-6) | | | | | |
|---|---|---|---|---|---|---|
| | θ=0 | θ=0.5236 | θ=1.0472 | θ=1.5708 | θ=2.0944 | θ=2.6180 |
| 01 | 0.25 | 0.33 | 0.6187 | 0.67 | 0.41 | 0.2424 |
| 02 | 0.63 | 0.46 | 0.4821 | 0.62 | 0.81 | 0.81 |
| 03 | 0.59 | 0.53 | 0.55 | 0.61 | 0.61 | 0.56 |
| 04 | 0.54 | 0.43 | 0.59 | 0.81 | 0.87 | 0.74 |
| 05 | 0.86 | 0.62 | 0.63 | 0.78 | 1.05 | 1.1629 |

*Table 4 :* Accuracy rate of Ridge Minutiae before Pre-Process

| Image # | Accuracy rate before Pre-processing | | |
|---|---|---|---|
| | Terminations (%) | Bifurcations (%) | Total Rate (%) |
| 01 | 3.86 | 20 | 3.98 |
| 02 | 2.91 | 64.71 | 4 |
| 03 | 26.6 | 3.4 | 10.95 |
| 04 | 4.04 | 17.02 | 4.7 |
| 05 | 3.79 | 23.26 | 4.46 |
| 06 | 5.36 | 2.94 | 4.21 |
| 07 | 3.31 | 21.43 | 3.9 |
| 08 | 3.94 | 4.35 | 3.73 |
| 09 | 1.74 | 60 | 2.44 |
| 10 | 1.07 | 14.29 | 1.4 |
| Average rate | 5.662 | 23.14 | 4.377 |

*Table 5 :* Accuracy rate of Ridge Minutiae after Pre-Process

| Image # | Accuracy rate after Pre-processing | | |
|---|---|---|---|
| | Terminations (%) | Bifurcations (%) | Total Rate (%) |
| 01 | 38.1 | 37.5 | 38 |
| 02 | 91.67 | 57.89 | 76.74 |
| 03 | 83.33 | 6.49 | 28.04 |
| 04 | 45.76 | 24.24 | 38.04 |
| 05 | 39.71 | 30.3 | 32.46 |
| 06 | 93.1 | 4.62 | 20 |
| 07 | 77.27 | 23.08 | 45.1 |
| 08 | 90 | 12.5 | 50 |
| 09 | 50 | 66.67 | 53.13 |
| 10 | 80 | 14.29 | 31.58 |
| Average rate | 68.894 | 27.758 | 41.309 |

*Table 6:* Accuracy rate of Ridge Minutiae after Post-Proces

| Image # | Accuracy rate after Post-processing | | |
|---|---|---|---|
| | Terminations (%) | Bifurcations (%) | Total Rate (%) |
| 01 | 72.73 | 50.00 | 67.86 |
| 02 | 84.62 | 57.89 | 76.74 |
| 03 | 96.15 | 7.46 | 32.26 |
| 04 | 90 | 44.44 | 72.92 |
| 05 | 65.85 | 41.67 | 47.44 |
| 06 | 79.41 | 06.19 | 24.09 |
| 07 | 77.27 | 24.00 | 46 |
| 08 | 90 | 17.39 | 58.49 |
| 09 | 100 | 75.00 | 85 |
| 10 | 80 | 40.00 | 60 |
| Average rate | 83.603 | 36.404 | 57.08 |

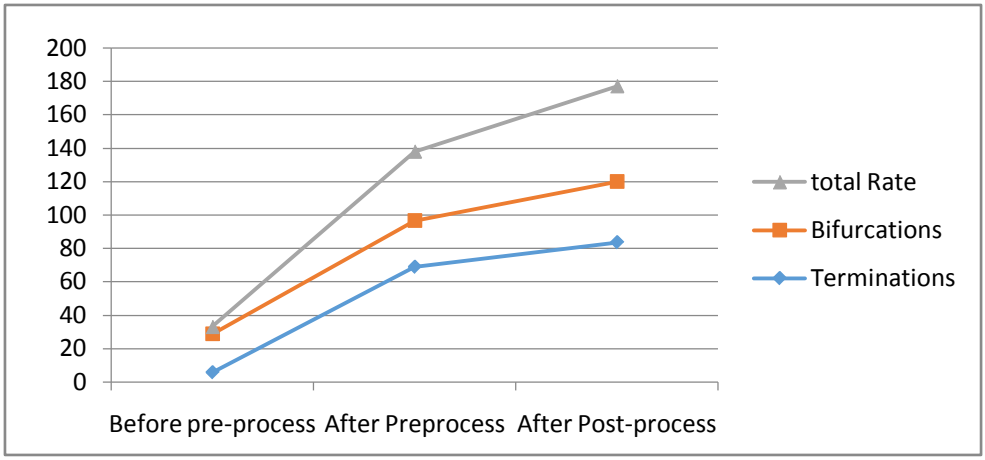



*Figure 6 :* Accuracy rates of Minutiae

Table 6 shows that the accuracy rates of terminations and bifurcations are increased gradually from preprocessing to post process. The accuracy rates are visualized through chart (see Fig. 6). The performance of the proposed algorithm is compared with other methods proposed by Feng Zhao [11], Maio[21], cheng[22], and Kim[23]. Some attributes are computed to measure the performance. They are:

1. True Minutiae (MT): Minutiae which are marked by a human expert.
2. Extracted Minutiae (ME): Minutiae which are extracted (remained) after post-processing.
3. False Minutiae (MF): Minutiae which are extracted by algorithm that do not coincide with True Minutiae (MT).
4. Dropped (Missed) Minutiae (MD): Minutiae marked by human expert that are not extracted by the algorithm.
5. Type-Exchanged Minutiae (MTE): The minutiae extracted by algorithm coincide with True minutiae (MT) but not the type. That is the exchange of terminations and bifurcations.

Table 7 reports the Average Error Rate (AER) after post processing in terms of average error rates of false minutiae (MF / ME), Dropped Minutiae (MD / MT ), and type-exchanged minutiae (MTE / ME). The total error rate can be calculated by summing up them. False Minutiae Rate of the proposed is lesser than the same of Feng Zhao [11] and Kim [23]. Type-exchanged minutiae rate is lesser than results of all the methods except Feng Zhao's result. The results show that our proposed algorithm is better than the other methods in terms of dropped minutiae and total error rates (20.69%). Figure 7 shows the performance of the proposed method according to the Total Error Rate (TER).

*Table 7:* Average Error Rate (AER) after post processing

| | False Minutiae (%) | Dropped Minutiae (%) | Type-Exchanged (%) | Total Error (%) |
|---|---|---|---|---|
| Proposed | 14.61 | 0.4 | 5.68 | **20.69** |
| Feng Zhao[11] | 15.3 | 6.9 | 5.3 | **27.5** |
| Maio[21] | 11.8 | 6.5 | 13.1 | **31.4** |
| Cheng [22] | 9.6 | 15.9 | 10.4 | **35.9** |
| Kim [23] | 25.8 | 13.8 | 6.3 | **45.9** |

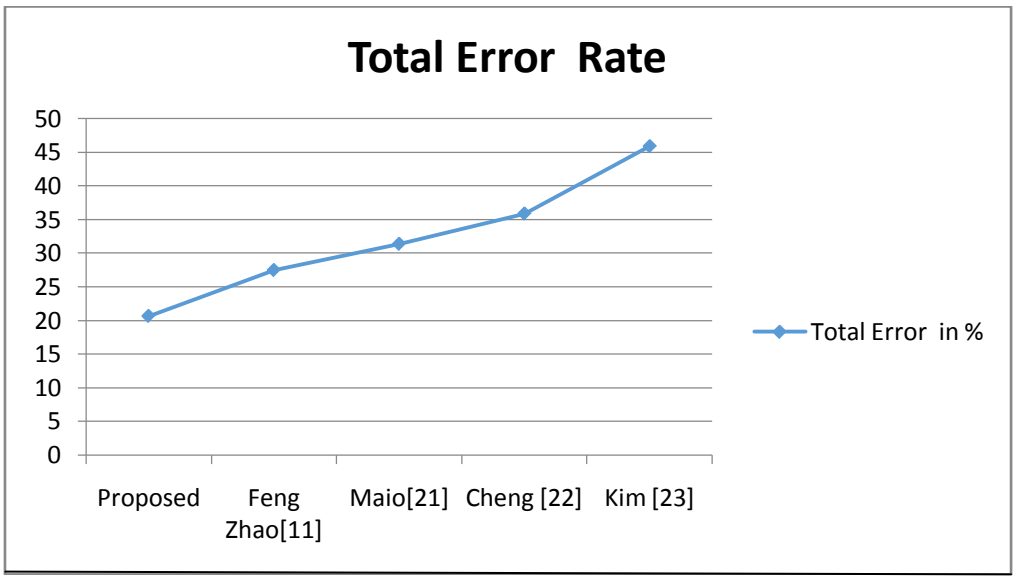



*Figure 7:* Performance chart

## VI. Summary and Conclusion

The proposed algorithm is implemented in order to achieve dual purpose tasks; these are an enhancement cum minutiae feature extraction through the Log-Gabor orientation and RLC methods. Log-Gabor orientation is used to enhance each ridge according to orientation and extract the enhanced minutiae. Enhanced minutiae are extracted for further process. Performance of the proposed method is measured in terms of accuracy rates of minutiae and also average error rates. Those are compared with the existing methods adopted from [11]. Higher the accuracy rate and lower the total error rate advances the performance of our proposed method.

## References Références Referencias


1. Anil.K.Jain, Yi Chen, Meltem Demirkus,(2007), Pores and ridges: High-Resolution Fingerprint Matching Using Level 3 Features, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.29, No.:1.
2. Hans Van Den Nieuwendijk, "Fingerprints", eBook.
3. Haralick, Robert M., and Linda G. Shapiro (1992), Computer and Robot Vision, Volume I, Addison-Wesley, pp. 28-48.
4. V.Rajaraman, Introduction to Information Technology, Third Edition.
5. Nalini K. Ratha, Shaoyun Chen, Anil K. Jain (1995) Adaptive flow orientation based feature extraction in fingerprint images, Pattern Reco-gnition, Vol. 28, issue 11.
6. Chih-Jen Lee and Sheng- De Wang (1999), A Gobor filter-based approach to fingerprint recognition, DOI,0-7803-5650-0/99,IEEE.
7. Jain A, Prabhakar S, Hong L (1999), A Multichannel approach to fingerprint classification, IEEE Transactions on Patt. Anal. Mach. Intel. , 21(4):348-359.
8. Wan S (2002) Fingerprint Classifications by directional fields, Fourth IEEE Int. Conference on Multimodal Interfaces (ICMI '02),IEEE.
9. Neil Yager and Adnan Amin (2004), Fingerprint Classification: a review, Pattern Analysis and Applications, Springer, Vol. 7, number 1, 77-93.
10. Sharat Chikkerur, Chaochang Wu, and Venu Govindaraju (2004), A Systematic Approach for Feature Extraction in Fingerprin Images, Biometric Authentication, LNCS, Vol. 3072, 1-23.
11. Feng Zhoa and Xiaoou Tang (2007), Preprocessing and Post processing for skeleton-based fingerprint minutiae extraction, Pattern Recognition, Vol. 40, 1270-1281, Elsevier.
12. Uday Rajanna, Ali Erol and George Bebis (2009), A Comparative study on Feature Extraction for fingerprint classification and performance

improvement using rank-level fusion, Pattern Analysis and Applications, Springer.
13. Chandra.E and Kanagalakshmi.K (2011), Frequency Domain Enhancement Filters for Fingerprint Images: A Performance Evaluation, CIIT International Journal of Digital Image Processing, Vol.3, No. 16.
14. Yoshihiko Hamamoto, Shunji Uchimura, Masanori, Watanabet, Tetsuya Yasuda, Yoshihiro Mitani and Shingo Tomita (1998), A gobar filter- based method for recognizing handwritten numerals, Pattern Recognitoin, Vol 31, Issue 4, pp: 395-400.
15. Rafael C. Gonzalez and Richard E.Woods, "Digital Image Processing", Third Edition, Pearson.
16. Chandra.E and Kanagalakshmi.K, (2011) Noise Suppression Scheme using Median Filer in Gray and Binary Images, International Journal of Computer Applications (0975 – 8887) Volume 26– No.1, pp. 49-57.
17. Anil K. Jain, and Jianjiang Feng (2011), Laten fingerprint Matching, IEEE Transactions on Pattern analysis and Machine Intelligence, vol. 23, No. 1.
18. MATLAB 7.10, Help desk.
19. Lousia Lam, Seong-Whan Lee, and Ching Y. Suen, (1992), Thinning Methodologies - A Comprehensive Survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 14, No. 9, pp: 869-885.
20. Lifeng He, Yuyan chao, Kenji Suzuki, Kesheng Wu (2009), Fast Connected-component labeling, Pattern Recognition, Vol. 42, Issue 9, pp 1977-1987.
21. D. Maio, D.Maltoni (1997), Direct gray-scale minutiae detection in fingerprints, IEEE Trans. Pattern anal Mech. Intell. 19(1) , 27-40.
22. J. Cheng, J. Tian (2004), Fingerprint enhancement with dyadic scale-space, Patt. Recog. Letters, 25(11), 1273-1284.
23. S. Kim, D. Lee, J. Kim (2001), Algorithm for detection and eliminations of false minutiae in fingerprint images, Proceedings of the third International Conference on Audio and Video-based Biometric Person Authentication (AVBPA'01), Halmstad, Sweden, pp. 235-240.

Global Journal of Computer Science and Technology: F
Graphics & Vision
Volume 14 Issue 4 Version 1.0 Year 2014
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

# Selecting Optimal RBF Kernel with Machine Learning for Feature Extraction and Classification in SAR Images

By P. Deepthi Jordhana & K.Soundararajan
*Intell Engineering College, India*

*Abstract-* Kernel methods are gaining popularity in image processing applications. The accuracy of feature extraction and classification on image data for a given application is greatly influenced by the choice of kernel function and its associated parameters. As on today there existing no formal methods for selecting the kernel parameters. The objective of the paper is to apply machine learning techniques to arrive at suitable kernel parameters and improvise the accuracy of kernel based object classification problem. The graph cut method with Radial Basis function (RBF) is employed for image segmentation, by energy minimization technique. The region parameters are extracted and applied to machine learning algorithm along with RBF's parameters. The region is classified to be man made or natural by the algorithm. Upon each iteration using supervised learning method the kernel parameters are adjusted to improve accuracy of classification. Simulation results based on Matlab are verified for Manmade classification for different sets of Synthetic Aperture RADAR (SAR) Images.

*Keywords:* *machine learning; RBF kernel; image segmentation; graph cut kernel..*

*GJCST-F Classification :* *I.4.0, H.1.2*


*Strictly as per the compliance and regulations of:*

# Selecting Optimal RBF Kernel with Machine Learning for Feature Extraction and Classification in SAR Images

P. Deepthi Jordhana [α] & K.Soundararajan [σ]

*Abstract -* Kernel methods are gaining popularity in image processing applications. The accuracy of feature extraction and classification on image data for a given application is greatly influenced by the choice of kernel function and its associated parameters. As on today there existing no formal methods for selecting the kernel parameters. The objective of the paper is to apply machine learning techniques to arrive at suitable kernel parameters and improvise the accuracy of kernel based object classification problem. The graph cut method with Radial Basis function (RBF) is employed for image segmentation, by energy minimization technique. The region parameters are extracted and applied to machine learning algorithm along with RBF's parameters. The region is classified to be man made or natural by the algorithm. Upon each iteration using supervised learning method the kernel parameters are adjusted to improve accuracy of classification. Simulation results based on Matlab are verified for Manmade classification for different sets of Synthetic Aperture RADAR (SAR) Images.

*Keywords: machine learning; RBF kernel; image segmentation; graph cut kernel.*

## I. Introduction

Automatic identification and reporting of man-made structure in images is useful in several emerging applications including synthetic aperture RADAR (SAR) image analysis, robotic navigation, automatic surveillance, image indexing and retrieval etc. The paper given here focuses on the recognition of man-made structures, which can be categorized to have specific geometric characteristics. Mainly the application of automatic analysis on SAR images is considered here. The automatic man-made object recognition from SAR images is a non-trivial problem due to following reasons.

- The view from which the image is created can be limited in SAR category applications.
- The image is created by RADAR echo signals, which is less informative than normal images.
- The images created under high clutter conditions are noisy and the edge or region extraction is not accurate.

*Author α: Associate Professor, department of electronics and communication engineering, Intell Engineering College, Anantapur, India. e-mail: p.deepthijordhana@gmail.com.*

*Author σ: Retd Professor, department of electronics and communication engineering, JNTU College of Engineering, Anantapur, India. e-mail: soundararajan_jntucea@yahoo.com*

These factors make the computation of the image primitives such as junctions, angles etc., which rely on explicit edge or line detection, prone to errors.

In surveillance and military applications of SAR, Buildings and vehicles are the most important man-made structures, which need to be detected. Some of the previous work on detection of buildings is given at [6][7][8][9] and [10] on normal images. Large number of these techniques uses aerial images for building detection by generating a hypothesis on the presence of surface on building roof top image [6]. The first step is detecting low-level image characteristics such as edges and regions. In next step either geometric feature based hypothesis [7], or a statistical models such as Markov Random Field (MRF) [8] is applied. In [11] a technique was proposed to use graph spectral partitioning for detection. Several techniques used with normal image processing algorithms require complex mathematic operations on images and require noise-free images.

The work at [12] and [13] establishes method to classify the whole image as a landscape or an urban scene. Oliva and Torralba [12] obtain a low dimensional holistic representation of the scene using principal components of the power spectra. The power spectra based features to be noisy for SAR images, which contain a mixture of both the landscape and man-made regions within the same image.

The work at [13] uses the edge coherence histograms over the whole image for the scene classification, using edge pixels at different orientations. Olmos and Trucco [14] proposed a system to detect the presence of man-made objects in underwater images using properties of the contours.

The techniques discussed in [15][16] perform classification in outdoor images using color and texture features, with different classification schemes. These papers report poor performance on the classes containing man-made structures since color and texture features are not very informative for these classes [13]. However for SAR images these techniques cannot be applied. These techniques classify the whole image in a certain class assuming the image to be mainly containing either man-made or natural objects, which is not true for many real-world images. In case of SAR created images, the images are taken over wide area containing mixed real world and man-made objects. The

figure 1, shows typical SAR image consisting man made objects.

*Figure 1 :* An SAR Created Image

In this paper, we propose to detect man-made structures in 2D images, formed by SAR. The proposed method uses Graph cut Image Segmentation method based on kernel mapping functions with machine learning algorithm on the SAR images. The section II illustrates introduction to kernel graph cut Image Segmentation principles and methods. The section III explains the various kernel mapping functions and machine learning algorithm simulated on the SAR images. The section IV has algorithm, simulation results and applications.

## II. Graph Cut Image Segmentation

The purpose of Image Segmentation is to divide an area into regions with a given description. Variational formulations partition an image to minimize an objective functional containing terms with descriptions of its regions and their boundaries. Continuous formulations view images as continuous functions over a continuous domain. The minimization function depends upon gradient descent. As a result, the algorithms converge to a local minimum; can be affected by the initialization. However these algorithms are typically slow and become a major hassle in applications which deal with large Images and thereby large regions.

Discrete formulations take images as discrete functions over a positional array. Graph cut Image Segmentation methods have been proved very efficient using this method. Minimi-zation by graph cuts provides a global optima and are less sensitive to.

The objective is to study kernel mapping to bring graph cut formulation for Multi region Segmentation of Images. The image data is implicitly mapped via a kernel function into data of a higher dimension so that the piecewise constant model, and becomes applicable.

The proposed functional consists of two terms: a kernel-induced term that measures the amount of deviation of the mapped Image data from piecewise constant linear data and regularization term which can be expressed as a function of region indices. The objective minimum functional is found by iterating through 2 steps via a common kernel function.

1. Minimization with respect to the image segmentation by graph cuts and
2. Minimization with respect to the regions parameters via fixed point computation. The pro-posed method has the advantages of being simple modeling and graph cut optimization with accuracy and flexibility.

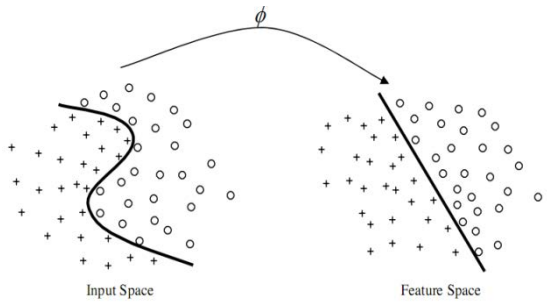



*Figure 2 :* The function $\phi$ maps the input data to Feature space where it is linearly separable

### *a) Graph cut Image Segmentation*

initialization with the advantage of piecewise constant data term.

Let $F: r \in \varphi \subset R^2 \rightarrow F_r = F(r) \in F$ be an image function from a positional array $\varphi$ to a space F consisting of photometric variables such as intensity, disparities, color or texture vectors. F is segmented to $N_{reg}$ regions which finds a suitable partition in the discrete domain to find a region which is compatible with some of the image characteristics. Partitioning of the image domain $\varphi$ equals to assigning each pixel a label l in a finite set of labels L. A region $R_l$ is defined as the set of pixels whose label is l, i.e., $R_l = \{r \in \varphi | r\}$ is labeled . The purpose is to find the labeling which minimizes a given functional.

To calculate Segmentational functional , let λ be an indexing function. λ assigns each point of the image to a region.

$$\lambda: r \in \varphi \rightarrow \lambda(r) \in Ł$$

where Ł is the finite set of region indices whose cardinality is less than $N_{reg}$. The Segmentation function can then be written as given in (1)

$$S(\lambda) = D(\lambda) + a\, R(\lambda) \qquad (1)$$

Where D is the data term and R is prior. α is a positive factor.

### *b) Proposed Functional*

Let $\Phi(.)$ be a nonlinear mapping function from the observation space $F$ to a higher dimensional feature space $J$ .The given labeling assigns each pixel a label and, subsequently,divides the image domain into

multiple regions. Each region is characterized by one label,

$$R_l = \{r \in \Phi / \lambda(r) = l\}, 1 \le l \le N_{reg}$$

Labeling that minimizes the functional in kernel induced space by graph cuts, as represented in (2).

$$S_k(\{v_l\}, \lambda) = \sum_{l \epsilon L} \sum_{r \epsilon R_l} \frac{(\Phi(v_l) - \Phi(F_r))^2 + \alpha \sum_{\{p,q\} \epsilon N} r(\lambda(p), \lambda(q))}{} \quad (2)$$

Generally in machine learning algorithms, the kernel trick is to use a linear classifier map the the original nonlinear data into a higher dimensional space. The Mercer's theorem, states that any continuous, symmetric, positive semidefinite kernel function can be expressed as a dot product in a high-dimensional space, without knowing the mapping explicitly. We can use a kernel function, KF(x,y) for this mapping as given in (3).

$$KF(x, y) = \phi(x)^T . \phi(y), \forall (x, y) \epsilon F^2 \quad (3)$$

Where "." is the dot product in feature space.

Therefore the Segmentation kernel function can be described as in (4). The functon $J_K$ is the non-ecludian distance in the original data space.

$$(\{v_l\}, \lambda) = \sum_{l \epsilon L} \sum_{r \epsilon R_l} J_K (F_r + \alpha) \sum_{\{p,q\} \epsilon N} r(\lambda(p), \lambda(q)) \quad (4)$$

*c) Optimization*

The obtained segmentational functional is minimized with an iterative two-step optimization strategy. The first step consists of fixing the labeling and optimizing $S_k$ with respect to statistical regions parameters using fixed point computation. The second step consists of finding the optimal labeling/partition of the image, with the given region parameters provided by the first step, via graph cut iterations. The algorithm iterates these two steps until convergence. With each iteration $S_k$ is decreased with respect to a parameter.This guarantees the algorithm to converge to a local minimum.

*d) Man-made object classification in SAR images*

The Fig. 1, shows an SAR image created over a region consisting few manmade structures and trees. In several military applications it is very useful if there is an automatic way of identifying these objects.

## III. Kernel Based Methods for Object Classification

This section explains the proposed approach for machine learning (ML) on kernel based object classification on SAR images. The machine learning approaches can be divided into 3 major categories.

- Supervised learning: In supervised learning the input data along with actual output is used to train the machine learning algorithm. The ML algorithm iteratively arrives at optimal hypothesis by every time checking the algorithm output is correct or not.
- Unsupervised learning: Only the input data is given to ML algorithm. The ML algorithm need to cluster the points in feature space and further use statistical means to classify. The ML algorithm training data has no clue of the actual output.
- Reinforcement learning: In this case the actual output is not offered to ML algorithm, instead an indicative of (such as quality factor) correctness or failure is provided.

In this work the supervised learning approach is adopted, where the ML algorithm gets trained with the help of operator visually checking the SAR image. Even though other methods are possible, as a first step towards this classification problem, the supervised ML approach is adopted.

The Fig. 3 has the flow chart representation of implemented ML approach for binary classification of SAR image segments.

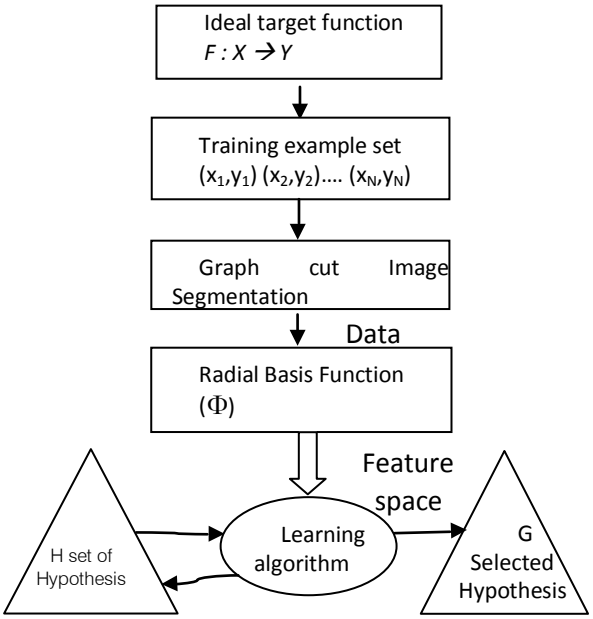



*Figure 3* : Machine learning problem of object classification

The unknown ideal target function is achieved through operator decision on each object of SAR image. The H hypothesis set consists of all pos-sible weight vectors. The selected hypothesis G is the final weight vector achieved after training.

*a) Kernel functions in Machine learning approach*

The basic idea of kernel methods is to (φ) transform the input data points (black dots) in to a high-dimensional feature space, where they can be described by a linear model (straight solid line). The linear model found in feature space corresponds to a non-linear model in the input space (curved solid line). The Fig. 4,

has the basic illustration of mapping involved in kernel functions.

This section describes the low level image parameters and mapped kernel functions which are considered for the ML algorithm. The edges, regions, statistical parameters are considered as input data set X. The output correct decision which is marked by operator is labeled as Y. Hence the training data set can be described as given in (5).

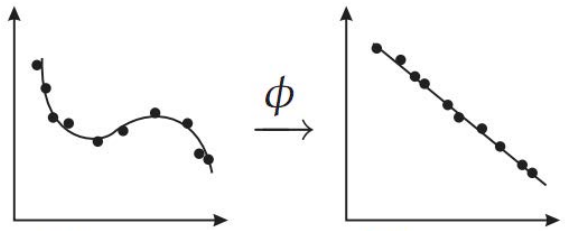



*Figure 4 :* Input space to Linear feature space mapped by kernel function

$$(X_1, y_1), (X_2, y_2), \cdots, (X_N, y_N) \quad (5)$$

The Kernel functions considered here are first order and second order partial derivatives in the x and y directions. If the P being the pixel set for an edge the first order derivative can map the linear edges to constants. Similarly the second order derivative can map the circular manmade objects to constants. Let Φ(.) be the mapping from the input data space to a higher dimensional feature/mapped space as given in (6) and (7).

$$X = (x_0, x_1, \cdots, x_d) \xrightarrow{\varphi} Z = (z_0, z_1, \cdots\cdots\cdots, z_{\hat{d}}) \quad (6)$$

$$X_1, X_2, \cdots, X_n \xrightarrow{\varphi} Z_1, Z_2, \cdots, Z_n \quad (7)$$

The first step in algorithm divides the image domain into multiple regions and edges (X). Each region is characterized by one identifier, Further the region characteristics are mapped to kernel function which becomes input to optimization algorithm.

The perceptron based classification problem for binary decision uses weight vectors as given in (4) defined on feature space.

$$\widehat{W} = (w_0, w_1, \cdots\cdots\cdots, w_{\hat{d}}) \quad (8)$$

$$g(x) = sign\,(\widehat{W}^T \varphi(X)) \quad (9)$$

Based on the sign the correctness is perceptron is used to update the weight vectors iteratively. The final selected hypothesis g becomes the candidate for classifying on the unknown data set.

$$H = \{h\} \quad g \in H$$

The Kernel function that is used here is Radial Basis function (RBF) kernel function which is a very popular kernel function used in Support vector machine classification (SVM)

$$KF(x,y) = \exp(-||x - y||^2/\sigma^2) \quad (10)$$

$||x - y||^2$ is taken to be squared Euclidean distance between the two feature vectors and $\sigma > 0$ is called the width parameter

- The value of kernel decreases with distance and ranges between zero and one, it is referred to as Similarity measure.
- The RBF Kernel specifies an infinite dimension feature space where higher order dimensions decay faster than lower order dimensions.
- It is comparatively faster.
- It provides smoothness over the contours

The numerous type of prevalent functions include

- Polynomial Kernel : $KF(x,y) = (x.y + c)^d$
- Sigmoid kernel : $KF(x,y) = \tanh(c(x.y) + \theta)$

## IV. Algorithm, Results and Applications

This section presents the results and discusses the possible applications of the developed ML Igorithm.

*a) Algorithm*

In the first step, regions are detected using Graph cut Image Segmentation method and pixel groups are made corresponding to each region.

In the second step the pixel coordinates are substituted in kernel mapping functions and feature space values are computed.

- To obtain region based classification, K means clustering algorithm is applied on the SAR Image to obtain clusters of the total image.
- Next on each cluster, Graph cut Image Segmentation algorithm is applied with specific kernel mapping function (RBF) to estimate the local minima convergence points.
- Contours are estimated over these points to finally obtain regions.
- Each region of the image is further divided into sections.
- Depending on the length of each section, the region is classified as Natural or manmade object.

```
If no_of_sections = 1
    Object = Man made
else if no_of_sections <=3
    Object = Natural
else
    Sort all sections in descending order of their length
    Sum the length of first 3 sections as S1
    Sum the length of all sections as S2
If S2>0.8*S1
    Object = Man made
```

else
Object = Natural
end
end

- An unsupervised machine learning algorithm is further applied based on the classification before segregating the sections.

*b) Simulation results*

The ML algorithm with Graph cut Image Segmentation is implemented in MATLAB and simulation results of the same are discussed in this section. The Fig. 5, has the region detection output for a specific SAR image.

(a)

(b)

*Figure 5 : (a)* Input image and *(b)* processed image with regions

The weight vectors for ML algorithm are biased to have reduced probability for target miss ($P_{TM}$) at the cost of increased false alarm ($P_{FA}$). This is because of obvious reason that the algorithm declared man-made objects will be further processed by operators or some other set of algorithms which can ignore if it is found to be not a region of interest. It is considered that, if an actual man-made object is not declared by ML algorithm, then its impact is high on the system operation. The Table I has the four possible scenarios in this binary classification problem.

*Table 1:* Decision analysis for binary classifier

| Algorithm declared result | Actual scene on SAR image | |
|---|---|---|
| | *Manmade object* | *Natural object* |
| Man made object | CORRECT DECISION | FALSE ALARM |
| Natural object | TARGET MISS | CORRECT DECISION |

MATLAB based graphic user interface (GUI) is developed to allow interactive user training for this purpose. The Fig. 6, has this GUI's screen shot.

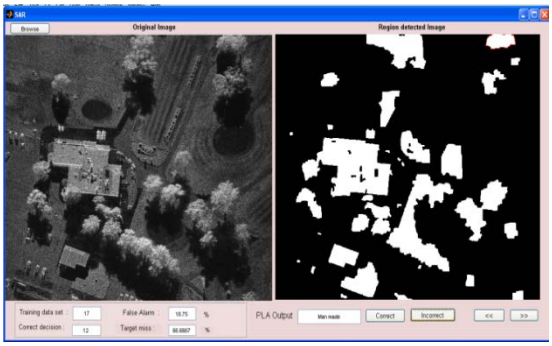


*Figure 6 :* GUI for interactive ML Algorithm

Every iteration highlights the region of the image and the decision made by ML algorithm. Operator declares the decision to be correct or incorrect. The statistics are displayed on the right.

The binary classification ML algorithm is tested with different sizes of data set and its false alarm and target miss rate are tabulated in Table II. The higher training data can make the algorithm more accurate.

*Table 2 :* Improvement in ML based object classification with larger data set

| Sl. No. | Improvement analysis of ML algorithm | | |
|---|---|---|---|
| | *Training data set size* | *False alarm* | *Target miss* |
| 1 | 20 | 18.75% | 66.67% |
| 2 | 50 | 16.78% | 33.33% |
| 3 | 100 | 15.73% | 22.48% |

*c) Applications*

Initially SAR based imaging was mainly used in military applications, but now the applications emerged in civilian applications also. The present work is aimed to be more useful in military applications where an automatic detection of manmade object detection can be used to alert military team. The Reconnaissance and surveillance aircrafts are enabled with SAR imaging technology. Automatic processing on these images can increase the capability to detect small targets within less time. By principle the algorithm does not limit even for ocean applications, however different training data would be required for more accurate results.

The SAR imaging is advantageous in comparison with normal imaging as it can be performed

during day time and also night time. In addition as the RF waves can penetrate through clouds the SAR imaging is preferred.

## V. Conclusions

The Experiment is performed on a set of 15 SAR Images . Each Image is first segmented into regions via Graph cut Image Segmentation using Radial Basis kernel function. Machine learning algorithm is applied to classify each region to either Natural or Manmade Object. It has been observed that with a training set of 100 the achieved classification accuracy is 67%. With higher training data set the accuracy can be further improved. The proposed method establishes the RBF kernel based machine learning approach for arriving at optimal parameters for kernel function and classification problem. The work is aimed to be continued for studying the utilization of different other kernel for the image processing applications and arriving at optimal parameter selection by machine learning approaches.

## VI. Acknowledgment

The authors would like to acknowledge the staff and research review members of JNTU, Anantapur for their suggestions during this research work.

## References Références Referencias

1. Mohammed Ben Salah.Multiregion Image Segmentation by Parametric Kernel Graph cuts.IEEE Trans on Image Processing Vol 20,No 2,2011.
2. J. Shi and J. Malik, "Normalized cuts and image segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 8, pp. 888–905, Aug. 2000.
3. I.B. Ayed, A. Mitiche, and Z. Belhadj, "Multiregion level-set parti-tioning of synthetic aperture radar images," IEEE Trans. Pattern Anal.Mach. Intell., vol. 27, no. 5, pp. 793–800, May 2005.
4. Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," IEEE Trans. Pattern Anal. Mach. Intell., vol.23, no. 11, pp. 1222–1239, Nov. 2001.
5. A. Jain, P. Duin, and J.Mao, "Statistical pattern recognition: A review,"IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 1, pp. 4–37, Jan.2000.
6. J. M. Morel and S. Solimini, Variational Methods in Image Segmentation. Cambridge, MA: Birkhauser, 1995.
7. G. Aubert and P. Kornprobst, Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations. New York: Springer-Verlag, 2006.
8. D. Cremers, M. Rousson, and R. Deriche, "A review of statistical approches to level set segmentation: Integrating color, texture,motion andshape," Int. J. Comput. Vis., vol. 72, no. 2, pp. 195–215, 2007.
9. D. Mumford and J. Shah, "Optimal approximation by piecewisesmooth functions and associated variational problems," Comm. Pure Appl. Math, vol. 42, pp. 577–685, 1989.
10. S. C. Zhu and A. Yuille, "Region competetion: Unifying snakes, region growing, and bayes/MDL for multiband image segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 18, no. 6, pp. 884–900, Jun. 1996.
11. T. F. Chan and L. A. Vese, "Active contours without edges," IEEE Trans. Image Process., vol. 10, no. 2, pp. 266–277, Feb. 2001.
12. L.A.Vese and T. F. Chan, "Amultiphase level set framework for image segmentation using the Mumford ans Shahmodel," Int. J. Comput. Vis.vol. 50, no. 3, pp. 271–293, 2002.
13. V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," in Proc. Int. Conf. Comput. Vis., 1995, pp. 694–699.
14. J. Shi and J. Malik, "Normalized cuts and image segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 8, pp. 888–905, Aug.2000.
15. Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images," in Proc.IEEE Int. Conf. Comput. Vis., 2001, pp. 105–112.
16. P. Felzenszwalb and D. Huttenlocher, "Efficient graph-based image segmentation," Int. J. Comput. Vis., vol. 59, no. 2, pp. 167–181, 2004.
17. Yang, Y. Allen, J. Wright, Y. Ma, and S. S. Sastry, "Unsupervised segmentation of natural images via lossy data compression," Comput. Vis. Image Understand., vol. 110, no. 2, pp. 212–225, 2008..
18. F. Destrempes, M. Mignotte, and J.-F. Angers, "A stochastic method for Bayesian estimation of hidden Markov random field models with application to a color model," IEEE Trans. Image Process., vol. 14,no. 8, pp. 1096–1124, Aug. 2005.
19. C. Vazquez, A. Mitiche, and I. B. Ayed, "Image segmentation as regularized clustering: A fully global curve evolution method," in Proc.IEEE Int. Conf. Image Processing, Oct. 2004, pp. 3467–3470.
20. Y. G. Leclerc, "Constructing simple stable descriptions for image par-titioning," Int. J. Comput. Vis., vol. 3, no. 1, pp. 73–102, 1989.
21. A.Blake, C. Rother, M. Brown, P. Perez, and P. Torr, "Intertive image segmentation using an adaptive GMMRF model," in Proc. Eur. Conf. Comput. Vis., 2004, pp. 428–441.
22. A.W. Doerry and D.F. Dubbert, "A portfolio of fine resolution Ka-band SAR images: part I", available at http://www.sandia.gov/RADAR-/images/ka_band_portfolio.pdf
23. R. Unnikrishnan, C. Pantofaru, and M. Hebert, "A measure for objective evaluation of image segmentation algorithms," in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., Jun. 2005, vol. 3,

pp. 34–41. G. Aubert and P. Kornprobst, Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations. New York: Springer-Verlag, 2006.
24. N. Vu and B. S. Manjunath, "Shape prior segmentation of multiple objects with graph cuts," in Proc. IEEE Int. Conf. Comput. Vis.Pattern Recognit., 2008, pp. 1–8.
25. M. B. Salah, A. Mitiche, and I. B. Ayed, "A continuous labeling for multiphase graph cut image partitioning," in Advances in Visual Computing, Lecture Notes in Computer Science (LNCS), G. B. e. al, Ed.
26. New York: Springer-Verlag, 2008, vol. 5358, pp. 268–277.

This page is intentionally left blank

Global Journal of Computer Science and Technology: F
Graphics & Vision
Volume 14 Issue 4 Version 1.0 Year 2014
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

# Cyclosparsity: A New Concept for Sparse Deconvolution

By Khalid Sabri
*University of Chouaib Doukkali, Morroco*

*Abstract-* Periodic random impulse signals are appropriate tools for several situations of interest and are a natural way for modeling highly localized events occuring randomly at given times. Nevertheless, the impulses are generally hidden and swallowed up in noise because of unwanted convolution. Thus, the resulting signal is not legible and may lead to erroneaous analysis, and hence, the need of deconvolution to restore the random periodic impulses. The main purpose of this study is to introduce the concept of cyclic sparsity or cyclosparsity in deconvolution framework for signals that are jointly sparse and cyclostationary like periodic random impulses. Indeed, all related works in this area exploit only one property, either sparsity or cyclostationarity and never both properties together. Although, the key feature of the cyclosparsity concept is that it gathers both properties to better characterize this kind of signals. We show that deconvolution based on cyclic sparsity hypothesis increases the performances and reduces significantly the computation cost as well. Finally, we use computer simulations to investigate the behavior in deconvolution framework of the algorithms Matching Pursuit (MP) [13], Orthogonal Matching Pursuit (OMP) [14], Orthogonal Least Square (OLS) [15], Single Best Replacement (SBR), [19, 20, 21] and the proposed extensions to cyclic sparsity context: Cyclo-MP, Cyclo-OMP, Cyclo-OLS and Cyclo-SBR.

*Keywords: cyclosparsity, sparsity, cyclostationary, periodic random impulses, deconvolution, greedy.*

*GJCST-F Classification : G.1.3*

*Strictly as per the compliance and regulations of:*

# Cyclosparsity: A New Concept for Sparse Deconvolution

Khalid Sabri

*Abstract-* Periodic random impulse signals are appropriate tools for several situations of interest and are a natural way for modeling highly localized events occurring randomly at given times. Nevertheless, the impulses are generally hidden and swallowed up in noise because of unwanted convolution. Thus, the resulting signal is not legible and may lead to erroneous analysis, and hence, the need of deconvolution to restore the random periodic impulses. The main purpose of this study is to introduce the concept of cyclic sparsity or cyclosparsity in deconvolution framework for signals that are jointly sparse and cyclostationary like periodic random impulses. Indeed, all related works in this area exploit only one property, either sparsity or cyclostationarity and never both properties together. Although, the key feature of the cyclosparsity concept is that it gathers both properties to better characterize this kind of signals. We show that deconvolution based on cyclic sparsity hypothesis increases the performances and reduces significantly the computation cost as well. Finally, we use computer simulations to investigate the behavior in deconvolution framework of the algorithms Matching Pursuit (MP) [13], Orthogonal Matching Pursuit (OMP) [14], Orthogonal Least Square (OLS) [15], Single Best Replacement (SBR), [19, 20, 21] and the proposed extensions to cyclic sparsity context: Cyclo-MP, Cyclo-OMP, Cyclo-OLS and Cyclo-SBR.

*Keywords: cyclosparsity, sparsity, cyclostationary, periodic random impulses, deconvolution, greedy.*

## I. Introduction

Cyclostationarity is very useful tool for studying periodically correlated signals by means of cyclic statistics [1].

As cyclostationary signals often encountered in practice, cyclostationary modeling being used in various domains such as mechanics [2], telecommunications [3], biomechanics [5] and allows good performances.

In this study we focus on particular case of cyclostationary signals which are wide sense second order cyclostationary i.e. first order and second order cyclostationary. Also, the signals are assumed to be made of periodic impulses with random amplitudes namely few nonzero impulses per period. Given the Impulse Response (IR), the aim is to retrieve the original object which has been distorted by passage through a known linear and time-invariant system in presence of noise. Indeed, enhancing the resolution of the signal and the Signal to Noise Ratio (SNR) from the knowledge of the IR corresponds to a deconvolution problem.

*Author : STIC Laboratory, Faculty of sciences, University Chouaib Doukkali, El Jadida, Morocco; (email: sabri.k@ucd.ac.ma)*

The deconvolution of cyclostationary signals has been addressed by several authors with different approaches. In [6], a bayesian deconvolution algorithm based on Markov chain Monte Carlo is presented. Cyclic statistics are often used for deconvolution, in [7, 8] the deconvolution is based on cyclic cepstrum, whereas, in [9, 10] the deconvolution is based on cyclic correlation. The main drawback of these methods is their inability to detect and restore impulses drowned in noise.

Actually, signal deconvolution belongs to inverse problems and is particularly well-known to be an ill-posed problem since the IR acts as a low-pass filter and the convolved signal is always noisy. Fortunately, regularization methods lead to acceptable solutions accounting for a priori information on the original object [38].

Analyzing periodic random impulse signals in details uncover another a priori information which is sparsity, this is because only few impulses are nonzero. Thus, data are sparse on the direct domain.

In this study, we will not exploit cyclic statistics but only the periodic character jointly with sparsity of periodic random impulses. This is possible thanks to the new concept of cyclosparsity that gathers both properties to better characterize these signals. Thus, cyclosparse deconvolution can be performed taking benefit of the correlation between the signal at different cycles (or periods).

Lately, in a different framework, sparse approximation of signals has drawn significant interest in many areas. The key idea is that a signal can be very well approximated with only few elementary signals (hereinafter referred to as atoms) taken from a redundant family (often referred to as dictionary), while its projection onto a basis of elementary signals may lead to a larger number of nonzero coefficients. Such a basic idea is the origin of recent theoretical development and many practical applications in denoising, compression, blind source separation and inverse problems [35, 11, 12].

Contrarily to orthogonal transforms, a redundant dictionary leads to non-unique representations of a given signal and several methods and algorithms have been developed to find the sparse approximations, i.e. the approximation with the smaller number of nonzero coefficients. In other words, minimizing the number of nonzero coefficients in a linear combination

approximating the data leads to an exhaustive search which is a NP-hard problem. Various methods and algorithms have been proposed to attempt to solve this problem and some sufficient conditions for these algorithms to reach the sparse solution have been established. Algorithms can be roughly classified in two approaches: greedy pursuit algorithms and convex relaxation. Greedy pursuit algorithms iteratively improve the approximation selecting at each iteration an additional elementary signal and many algorithms have been proposed based on such a scheme [13, 14, 15].

The principle of convex relaxation methods is to replace the minimization of the number of elements by the minimization of another functional which can be minimized more easily and still guarantee the solution to have a large number of zero coefficients. A $\ell_1$-norm is mainly used to this end [16, 17, 18].

Our work consists on extending sparse approximation to cyclostationary signals with periodic random impulses, where the aim is to find jointly a sparse approximation of each cyclic period (or cycle), accounting for the same elementary signals in each approximation, but shifted with a multiple of the cyclic period and with different coefficients. However, we insist on the application of cyclosparse approximation to deconvolution i.e. the dictionary is given by the Toeplitz matrix formed by the IR.

The paper is organized as follows; section 2 defines the problem statement and motivations of the study. The main contribution of this paper is described in sections 3, 4 and 5. The concept of cyclosparsity is introduced in section 3. In section 4, we summarize the statement of sparse and cyclosparse approximation problems. We also point-out the link with cyclosparse deconvolution and we insist on the differences with joint sparsity. At the end of this section, we came to the conclusion that none of the usual algorithms ensure to reach the actual solution or even to reach the same solution. Thus some numerical experiments have to be performed to compare these algorithms. Such a cyclosparse model is taken into account in the sparse deconvolution by greedy algorithms in section 5, which fortunately reduces significantly the computation cost. In this paper, we focus on greedy algorithms namely

Matching Pursuit (MP) [13], Orthogonal Matching Pursuit (OMP) [14], Orthogonal Least Square (OLS) [15] and Single Best Replacement (SBR), [19, 20, 21] (which, despite a different aim, has a structure very similar to greedy algorithms). And we propose to generalize these algorithms to the cyclosparse context: Cyclo-MP, Cyclo-OMP, Cyclo-OLS and Cyclo-SBR. We propose furthermore to test all the algorithms on the same statistical basis, i.e. with the same stopping rule deduced from statistical properties of the noise. Also it seems to be necessary to obtain satisfactory deconvolution results, as shown in the simulation results section 6.

## II. Background

### a) Problem Formulation

Consider the situation where a known system $\mathcal{H}(t)$ is excited by a cyclostationary signal $x(t)$ consisting of periodic random impulses. By periodic, we mean that the signal can be divided into portions of length $T$ (which is known as the cyclic period of the signal) with $d$ impulses in each portion. Moreover, the delay factor $\tau_i$ of the *ith* impulse $x_i$ is constant for all portions. Note that in general $\tau_i$ will be different for different $i$ although in most of the cases they may be integral multiples of a constant $\tau$. An example of $x(t)$ is shown in Fig. 1 with $d = 5$.

Reconsider the system as described above. Since the impulses are periodic, we can consider a period of time $T$ and write that portion of the output as $\sum_{k=1}^{d} x_k \mathcal{H}(t-\tau_k) + n(t),\ 0 \le t < T$. This relationship can be generalized to cover the whole signal as,

$$y(t) = \sum_{i=0}^{K-1} \sum_{k=1}^{d} x_{i,k}\, \mathcal{H}(t-(\tau_k + iT)) + n(t), \qquad (1)$$

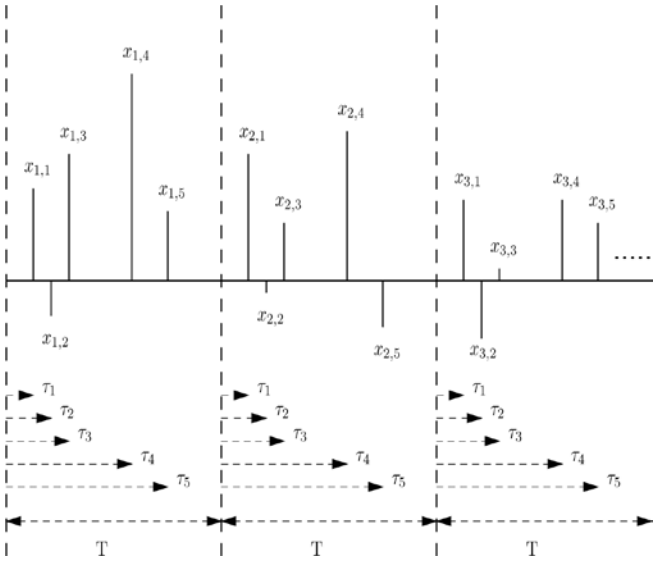



*Figure 1:* Example of cyclosparse signal $x(t)$

Where $d$ is the number of effective impulses in the period $T$ with $x_{i,k}$ and $\tau_k$ being their amplitude and delay factors respectively. $K$ denotes the number of period per signal and the sub-index $i$ stands for the period index, so $x_{i,k}$ represents the impulse with $\tau_k$ as delay factor in the *ith*-period. $n(t)$ represents the random noise of the system.

*b) Motivation of the study*

The choice of the signal modeling 1 is not arbitrary but can be justified in practice. Periodic random impulse processes are suitable tools that allow physicists to model many situations of interest, indeed they are a natural way for modeling highly localized events occurring randomly at given times or points of the state space [7, 8, 10, 6].

Another additional property of cyclostationary signals with periodic random impulses is sparsity, as only few impulses are nonzero i.e. $K \times d$ nonzero impulses. Hence, these signals are sparse on the direct domain. Unfortunately, all related works in this area exploit only one property, either sparsity or cyclostationarity and never both properties jointly. Convinced that combining simultaneously sparsity and cyclostationarity may lead to an enhancement of performances. So we wondered whether it is possible to build up an approach based on this idea. These were our motivations for introducing the new concept of cyclosparsity that gathers both properties to better characterize this kind of signals. The contributions of this article with respect to the short one [4], lie in:

- studying and generalizing the cyclosparsity concept to the algorithms OLS and SBR, which gives rise to Cyclo-OLS and Cyclo-SBR.
- providing insights into how the cyclosparsity works, and reviewing its implications in improving the selection step and reducing the computational cost of the four generalized greedy algorithms.
- jointly comparing the performances of all algorithms by varying the SNR for different number of cycles (K=2, 4, 8 and 16).

## III. Concept of Cyclosparsity

The signal object of this study is assumed to be cyclostationary with random impulses i.e. consists of *d* periodic random impulses with *d* represents the number of impulses by cycle. On the one hand, only few elements are nonzero by cycle, so the signal is considered to be sparse. Furthermore, the positions of these nonzero elements (impulses) are cyclic/periodic i.e. they keep the same positions whatever the cycle. Hence, the $\ell_0$-norms of the signal for each cycle are equal. On the other hand, the amplitudes of these nonzero elements are different and assumed to be random (see Fig. 1). Thus, the way to better characterize cyclostationary random impulses is to combine both properties i.e. cyclostationarity and sparsity. This hypothesis is the key idea of the cyclosparsity concept and is often satisfied in practice.

Before announcing the theorem defining cyclosparse process, we first introduce the following notations:

- $\boldsymbol{x} = [x_1, \ldots, x_{L_x}]^{\mathrm{T}}$ being the column vector of length $L_x$ constructed from the signal $x(t)$.
- $\boldsymbol{x}_i = \boldsymbol{x}(i, i+T, i+2T, \ldots, i+(K-1)T)$ being a column vector with $1 \le i \le T$ sweeps all *ith* elements of each period (a total of $K$ elements).
- $\boldsymbol{x}_{i,:} = \boldsymbol{x}((i-1)T+1, :iT)$ being a column vector with $1 \le i \le K$ sweeps all elements of the *ith* period (a total of *T* elements).
- $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T] = [\boldsymbol{x}_{1,:}, \ldots, \boldsymbol{x}_{K,:}]^{\mathrm{T}}$ being a $K \times T$ matrix where the elements are $x_{i,j}$ with *i* and *j* stand respectively for the period/cycle index and the position on each period.

The superscript T stands for the transpose of a vector or a matrix.

*Theorem.* A periodic random impulse signal $x(t)$ of period *T* is cyclosparse if the cardinal of nonzero elements of the whole signal is equal to the cardinal of nonzero elements of the signal over any period multiplied by the number of cycles *K*. This can be formulated into the following mathematical relationship:

$$\|\boldsymbol{x}\|_0 \le K \|\boldsymbol{X}\|_{\infty,0} \tag{2}$$

the $\ell_{\infty,0}$-norm is applied to the column vector $\boldsymbol{x}$ after being reshaped as $K \times T$ matrix.

*Proof.* The $\ell_0$-norm of $\boldsymbol{x}$ is defined as follows

$$\|\boldsymbol{x}\|_0 = \sum_{k=1}^{KT} |x_k|^0 \tag{3}$$

As is standard, $|x_k|^0 = 1$ if $x_k \neq 0$ and $|x_k|^0 = 0$ if $x_k = 0$. Thus, $\|\boldsymbol{x}\|_0 = Card\{k, x_k \neq 0\}$ which indicates

the number of nonzero components of the column vector $\boldsymbol{x}$. Using these notations, Eq. 3 can be written as,

$$\|\boldsymbol{x}\|_0 = \|\boldsymbol{x}_{1,:}\|_0 + \|\boldsymbol{x}_{2,:}\|_0 + \cdots + \|\boldsymbol{x}_{K,:}\|_0$$
$$= \sum_{k=1}^{T} |x_{1,k}|^0 + \sum_{k=1}^{T} |x_{2,k}|^0 + \cdots + \sum_{k=1}^{T} |x_{K,k}|^0$$

Reorganizing the previous sum leads to,

$$\|\boldsymbol{x}\|_0 = |x_{1,1}|^0 + |x_{1,2}|^0 + \cdots + |x_{1,T}|^0$$
$$+ |x_{2,1}|^0 + |x_{2,2}|^0 + \cdots + |x_{2,T}|^0$$
$$+ \vdots + \vdots + \cdots + \vdots$$
$$+ |x_{K,1}|^0 + |x_{K,2}|^0 + \cdots + |x_{K,T}|^0$$

It is easy to see that each vertical sum corresponds to a $\ell_\infty$-norm along the dimension of its corresponding column multiplied by the scalar $K$.

$$\|\boldsymbol{x}\|_0 \leq K\left(\left|\max_i |\boldsymbol{x}_1|\right|^0 + \left|\max_i |\boldsymbol{x}_2|\right|^0 + \cdots + \left|\max_i |\boldsymbol{x}_T|\right|^0\right)$$
$$\leq K \left\|\max_i |\boldsymbol{x}_i|\right\|_0$$
$$\leq \|\boldsymbol{X}\|_{\infty,0}$$

Where $\max_i |\boldsymbol{x}_i|$ stands for $\ell_\infty$-norm of $\boldsymbol{x}_i$ and the composite norm $\|\max_i |\boldsymbol{x}_i|\|_0$ represents the mixing norm $\ell_{\infty,0}$ of $\boldsymbol{X}$. The minimization of this norm encourages first diversity along $i$ and then sparsity of the resulting vector.

The cyclosparsity in the case of deconvolution involves that two kinds of atoms participate to the linear combination in order to build data $\boldsymbol{y}$. Thus, we distinguish,

- Fundamental atoms, which correspond to nonzero impulses of the first cycle.
- Harmonic atoms, which are deduced from fundamental atoms by shifting with multiple of the cycle i.e. atoms corresponding to nonzero impulses of the remaining cycles.

Before studying in detail the cyclosparse approximation we first recall the principle of sparse approximation.

## IV. From Sparse Approximation to Cyclosparse Approximation

### a) *Sparse approximation*

The problem of sparse signal approximation consists in approximating a signal as a linear combination of a restricted number of elementary signals selected in a redundant collection (dictionary). It can be written as:

Find sparse $\boldsymbol{x}$ such that $\boldsymbol{\Phi x} \approx \boldsymbol{y}$,

where $\boldsymbol{y}$ corresponds to measured data and $\boldsymbol{\Phi}$ is a known matrix with atoms $\{\boldsymbol{\phi}_k\}_{k=1\ldots Q}$ as columns. Sparse approximations have to deal with a compromise between a good approximation and the number of involved elementary signals. Mathematically such compromise arises from minimizing the following criterion:

$$\mathcal{J}(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{\Phi x}\|_2^2 + \beta\|\boldsymbol{x}\|_0 \qquad (4)$$

The parameter β controls the trade-off between the sparsity of the solution and the quality of the approximation. The lower is β, less sparse is the solution and better is the approximation. Hence, β is the key parameter to reach the compromise [38].

Of course, minimizing such a criterion corresponds to a combinatory optimization problem which is widely known to be NP hard. However, two approaches are usually used to avoid sweeping every combination: 1) Greedy algorithms, which iteratively ripen the approximation by successively identifying additional elementary signals that improves the approximation quality [13, 22]; 2) Convex relaxation algorithms i.e. based on the relaxation of the criterion (4), which replace the combinatorial problem with an easier optimization problem often chosen convex [16]. In the latter, the $\ell_0$-norm is often relaxed with a $\ell_p$-norm, where $\|\boldsymbol{x}\|_p = (\sum_k |x_k|^p)^{\frac{1}{p}}$ (note that $p = \infty$ is a limiting case for which $\|\boldsymbol{x}\|_\infty = \max_k |x_k|$). For $p = 1$ this problem corresponds to the Least Absolute Shrinkage And Selection Operator Regression (LASSO) [18] or Basis Pursuit DeNoising (BPDN) in signal processing [16].

Serious efforts have been made over many years to relate the solution of the original problem and the results of these two approaches and related algorithms. Some theoretical results give sufficient conditions of equivalence, depending on the dictionary and eventually of the solution [23, 22]. One such sufficient condition, not depending on the elementary signals of solution, is related to the coherence parameter $\mu$ of a dictionary which corresponds to the maximum absolute inner product between two distinct atoms in the dictionary:

$$\mu \stackrel{\text{def}}{=} \max_{j \neq k} |\langle \boldsymbol{\phi}_j , \boldsymbol{\phi}_k \rangle| = \max_{j \neq k} |(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})_{jk}|.$$

In particular, it has been shown [24, 25, 22] that recovery condition for which the solution is unique and can be computed with various algorithms (BP, OMP, . . .) is

$$\|\boldsymbol{x}\|_0 < \frac{1}{2}\left(1+\frac{1}{\mu}\right) \text{ with } \mu < 1. \qquad (5)$$

This is a sufficient condition under which both $\ell_1$ criterion and greedy approaches can recover an exactly sparse signal.

*b) Cyclosparse approximation*

The extension of sparse approximation for cyclostationary and sparse signals is the main aim of this study. A cyclosparse solution is given by minimizingthe following criterion:

$$\mathcal{J}(\boldsymbol{x}) \sum_{i=1}^{T} \left\|\boldsymbol{y} - \boldsymbol{\Phi}_{\{i+mT;\ m=0,\dots,K-1\}}\boldsymbol{x}\right\|^2 \beta K \|\boldsymbol{X}\|_{\infty,0}\ ; \qquad (6)$$

with *K* and *T* denote respectively the number of cycles/periods and the cyclic period. Actually, the inner $\ell_\infty$-norm encourages diversity along the cycles so the $\ell_{\infty,0}$ mixed-norm measures the cyclosparsity along the whole signal. Both approaches used to avoid exploring every combination of the sparse approximation problem can be extended to cyclosparse context: greedy algorithms and convex relaxation. In this paper, we focus on greedy algorithms.

*c) Difference between joint sparse approximation and cyclosparse deconvolution*

In spite of the fact that the joint sparse approximation [26, 27, 28, 29] and cyclosparse deconvolution seem to have similar formulations, they are completely different, and the most basic differences are listed as follows:

1. In the case of deconvolution, the dictionary is imposed by the IR and cannot be chosen unlike sparse approximation where the dictionary is generally chosen as union of bases or wavelet dictionary. Also, the dictionary size is of the same order as data size, so the dictionary does not correspond to a redundant set of elementary signals as in sparse approximation.
2. The deconvolution aims to retrieve the object $\boldsymbol{x}$, i.e., to find the coefficients associated to the actual elementary signals forming the data. Whereas, the problem of sparse approximation searches for a good approximation of the object involving few atoms.
3. The joint sparsity is used for multi-dimensional signals or images that share almost the same dictionary. In the context of our study, we deal with mono-dimensional signals which cannot be split to apply joint sparsity.

The first point is a matter of prime importance in terms of computation of the solution. Indeed, as for deconvolution, the dictionary atoms correspond to shifted versions of the IR, hence they are highly correlated. So, the theoretical properties which guarantee the solution of greedy algorithms or convex relaxation to correspond to the sparse approximation are often not satisfied. For example, the coherence is generally very high $(\mu > \frac{1}{2})$ so theorems based on such quantities guarantee to reach the solution only if it is composed of a single elementary signal. This can be demonstrated by considering a cyclosparse object $\boldsymbol{x}$ with $\|\boldsymbol{x}\|_0 = K\|\boldsymbol{X}\|_{\infty,0} = Kd$, relationship 5 becomes $\mu < \frac{1}{2Kd-1}$. To be in conformity with $\mu > \frac{1}{2}$, $Kd$ must be set to 1, which means that the signal is composed of single period with one impulse i.e. a single elementary signal. More precisely, it can easily be seen in this case that the scalar product $\langle \mathbf{h}_j , \mathbf{h}_k \rangle$ roughly corresponds (up to the boundary conditions taken into account for the convolution) to a value of the auto-correlation of the IR:$\langle \mathbf{h}_j , \mathbf{h}_k \rangle = \langle \mathcal{H} * \delta_j , \mathcal{H} * \delta_k \rangle = \langle \delta_j , \mathcal{H}^- * \mathcal{H} * \delta_k \rangle = \mathcal{H}^- * \mathcal{H}(k-j)$,where $\mathcal{H}^-(j) = \mathcal{H}(-j)$ and $\delta$ represents the Kronecker symbol. For example, if the IR has a Gaussian-like shape, the autocorrelation of the IR also has a Gaussian-like shape with a twice as large standard deviation, so a correct sampling of the Gaussian leads to $\mu > \frac{1}{2}$. On the other hand, it is obvious that two spikes are easy to detect if they are sufficiently far away from each other. A sufficient recovery condition has been established on the basis of such distance for sparse deconvolution [30] and also for super-resolution [37], but this condition may not be carried out for real data. Therefore, no theoretical results guarantee that the algorithms converge towards the sparse solution. So the various algorithms have to be tested on realistic data. Moreover, as each algorithm could give different solutions, one should compare the efficiency of each algorithm on a coherent basis, in particular with respect to the tuning parameters.

The second point points-out the issues of sparse deconvolution and sparse approximation. Indeed, the objective of sparse deconvolution is to recover the spike-like objects, i.e. to detect the unresolved objects and to estimate their amplitude, and not only to have a sparse solution which gives a good approximation of the data. So if the algorithms fail to reach the real solution, this leads to false alarms and missing detection of objects. Moreover, when dealing with real data, it is not guaranteed that the sparse solution corresponds to the true objects. To avoid such a problem in this study, we will only compare hereafter the algorithms on simulated data for which the true solution is known.

Finally, the third point establishes that joint sparsity modeling is not suitable for periodic random

impulses. Actually, the major difference between both concepts arises in the fact that joint sparsity assumes that multi-dimensional signals are disjointed on the border. This means that each signal does not contribute on its neighbors because the convolution is independently made between each mono-dimensional signal and the IR. However, for cyclosparsity, the cycles are related and connected because of the convolution with the IR. Hence, the cycles are not independent anymore and cannot be split as for joint sparsity.

Consequently, we came to the conclusion that the joint sparsity cannot be used to characterize and model cyclostationary signals with periodic random impulses. These were the reasons behind developing the concept of cyclosparsity.

In addition, as the algorithms may give different solutions, one should test all the existing algorithms. In this study we focus on the greedy algorithms. The first reason is that greedy algorithms are classically used in many domains for a long time, and even before the first publications on sparse representations. So the use of greedy algorithms is quite natural. Moreover, one important aspect of this study is to be able to compare the algorithms on the same basis in terms of parameters tuning. The noise level of real data can often be modeled or at least estimated from the data so the noise variance can be considered as a known parameter. In this study, we will use the noise variance to build the condition to stop the iterations of the various greedy algorithms on the same statistical basis. Such a condition is not as easy to build on the same basis using convex relaxation.

## V. Cyclosparse Deconvolution

### a) *Structure of the dictionary H*

Let us first specify the boundary condition accounted for in the convolution operator. We assume that the convolution (Eq. 1) is computed with the zero-padded edges. Using this option the resulting signal has length $L_y = L_x + L_h - 1$ where $L_x$ and $L_h$ stand respectively for the length of the signal to reconstruct and the Point Spread Function (PSF). Of course, such boundary hypothesis influences size and structure of the dictionary $\mathbf{H}$ formed from the IR. In particular, for physical reasons, the IR is normalized such that $\sum_{j=1}^{L_h} \mathcal{H}_j^2 = 1$ so the columns of the matrix which correspond to shifted versions of the IR should have a constant norm $\mathbf{h}_i^T \mathbf{h}_i = \sum_{j=1}^{L_h} \mathcal{H}_j^2$.

Note that $\mathbf{H}$ is a sparse matrix of dimension $L_y \times L_x$ with $L_h \times L_x$ nonzero elements (the length of the PSF is generally largely smaller than the length of the signal). But as the $L_x$ atoms of the dictionary correspond to shifted versions of the PSF, matrix $\mathbf{H}$ is composed only with the $L_h$ elements of the PSF.

Moreover, as matrix $\mathbf{H}$ models a convolution operator, it has a Toeplitz structure (diagonal-constant matrix) and each operation involving $\mathbf{H}$ may be computed as a result of a convolution.

### b) *Problem statement*

Before discussing the general case modeling of cyclosparse deconvolution, let us consider an example of cyclosparse object $\boldsymbol{x}$, with two periodic random impulses of positions $d_1 + mT$ and $d_2 + mT$, in order to make clear the formulation. Thus, the atoms that participate to build-up $\boldsymbol{y}$ are $\mathbf{h}_{d_1+mT}$ and $\mathbf{h}_{d_2+mT}$ with $m = 0, \ldots, K-1$. In addition, the atoms $\mathbf{h}_{d_1+mT}$ are not correlated (likewise $\mathbf{h}_{d_2+mT}$ ) (i.e. the scalar product of these vectors for different values of $m$ is null) as the IR length $L_h$ is generally smaller than the cyclic period $T$. If it is not the case, the IR $\mathcal{H}$ can be truncated such as $L_h < T$.

Consequently, for any cyclosparse object $\boldsymbol{x}$, the model 1 can be written with matrix notations as follows,

$$\boldsymbol{y} = \mathbf{H}\,\mathbf{D}\,\boldsymbol{x} + \boldsymbol{n}$$
$$= \tilde{\mathbf{H}}\,\boldsymbol{x} + \boldsymbol{n}$$

$\tilde{\mathbf{H}}^{\mathrm{T}}\tilde{\mathbf{H}} = \mathbf{D}\left(\mathbf{H}^{\mathrm{T}}\mathbf{H}\right)\mathbf{D}$ with $\mathbf{D}$ is the cyclosparsity operator which is a $L_x \times L_x$ diagonal matrix

$\mathbf{D}^{\mathrm{T}}=\mathbf{D} = \sum_{m=0}^{K-1} \sum_{n=1}^{\|\boldsymbol{x}\|_{\infty,0}} \boldsymbol{e}_{d_{n+mT}} \boldsymbol{e}_{d_{n+mT}}^{\mathrm{T}} =$ $\mathrm{diag}\left(\sum_{m=0}^{K-1} \sum_{n=1}^{\|\boldsymbol{x}\|_{\infty,0}} \boldsymbol{e}_{d_{n+mT}}\right)$ where $\boldsymbol{e}_{d_{n+mT}}$ are the canonical basis vectors and $d_n$ is the index of the *nth* impulse with $\tau_n$ as factor delay. Thus, $\tilde{\mathbf{H}}$ is a $L_y \times L_x$ matrix with particular structure i.e. the nonzero columns for each cycle are deduced by shifting the columns of the first cycle with a multiple of the cyclic period $T$. Hence, $\tilde{\mathbf{H}}$ points out the cyclosparsity property in the convolution case.

### c) *Cyclosparse greedy algorithms*

Let the sub-matrix $\mathbf{H}_\Lambda$ built-up from the columns of $\mathbf{H}$ where the indices are in $\Lambda$, $\mathbf{h}_i = \mathbf{H}_{\{i\}}$, and $\Lambda^{(k)}$ is the set of the selected indices at iteration $k$. The vectors are defined as follows, $\boldsymbol{x} = [x_1, \ldots, x_{L_x}]^{\mathrm{T}}$, $\boldsymbol{y} = [y_1, \ldots, y_{L_y}]^{\mathrm{T}}$, $\boldsymbol{n} = [n_1, \ldots, n_{L_y}]^{\mathrm{T}}$ and $\boldsymbol{r} = [r_1, \ldots, r_{L_y}]^{\mathrm{T}}$ which denotes the residual. $L_x$, $L_y$ and $L_h$ stand respectively for the length of $\boldsymbol{x}, \boldsymbol{y}$ and $\mathcal{H}$. Finally, let the vector $m = [0, 1, \ldots., (K-1)]$ be the vector of period indices.

In this study the interest is focused on the extension of greedy sparse approximation algorithms to cyclosparsity context for deconvolution. The dictionary is given by the Toeplitz matrix $\mathbf{H}$ deduced from the IR $\mathcal{H}$. Greedy algorithms are iterative algorithms composed of two major steps at each iteration: 1) the selection of an additional elementary signal in the dictionary; 2) the update of the solution and the corresponding

approximation. A stopping rule helps decide whether to stop or continue the iteration.

Let us consider $\boldsymbol{x}^{(k)}$ be the solution of the *kth* iteration, $\boldsymbol{x}_{\Lambda}^{(k)}$ being its coefficients at indices $\Lambda$ and $\boldsymbol{r}^{(k)} = \boldsymbol{y} - \mathbf{H}\,\boldsymbol{x}^{(k)}$ the residual corresponding to this solution (approximation error). The typical structure of a greedy algorithm is:

Initialize $k = 0$, $\Lambda^{(k)} = \emptyset$ and $\boldsymbol{r}^{(k)} = \boldsymbol{y}$.

Iterate on $k = k + 1$ until the stopping rule is satisfied:

- Select the index $i^{(k)}$ corresponding to an atom $\mathbf{h}_i$ improving the approximation.
- Update the solution $\boldsymbol{x}^{(k)}$, with nonzero elements at indices $\Lambda^{(k)} = \Lambda^{(k-1)} \cup \{i^{(k)}\}$, and the corresponding residual $\boldsymbol{r}^{(k)}$.

The various algorithms differ on the selection or the updating steps. The most popular greedy algorithm must be the Matching Pursuit [13] and its orthogonal version OMP [14]. Both algorithms and two other algorithms will be extended to cyclosparsity context.

*Cyclo-Matching Pursuit (Cyclo-MP)*

The Cyclo-MP (or Cyclic-MP) is the extension of the Matching Pursuit (MP) [13]. The *K* additional atoms jointly maximize the scalar product with the residual. The update corresponds to an orthogonal projection of the residual on the selected atoms, so only the solution at the selected indices is updated. Note that with such a scheme it is possible to select already selected atoms.

In order to avoid overloading equations, the index $i^{(k)} + mT$ will be replaced by $i_{mT}^{(k)}$ in the following.

- Selection of *K* atoms: $\Lambda^{(k)} = \Lambda^{(k-1)} \cup \{i_{mT}^{(k)};\ m = 1, \ldots, K-1\}$,

$$i^{(k)} = \underset{i}{\operatorname{argmax}} \sum_{m=0}^{K-1} \left|\mathbf{h}_{i+mT}^{\mathrm{T}}\, \boldsymbol{r}^{(k-1)}\right| \qquad (7)$$

- Update:
  solution:

$$\boldsymbol{x}_{i_{mT}^{(k)}}^{(k)} = \boldsymbol{x}_{i_{mT}^{(k)}}^{(k-1)} + \left(\mathbf{H}_{i_{mT}^{(k)}}^{\mathrm{T}} \mathbf{H}_{i_{mT}^{(k)}}\right)^{-1} \mathbf{H}_{i_{mT}^{(k)}}^{\mathrm{T}}\, \boldsymbol{r}^{(k-1)} \qquad (8)$$

  residual:

$$\boldsymbol{r}^{(k)} = \boldsymbol{r}^{(k-1)} - \mathbf{H}_{i_{mT}^{(k)}} \left(\mathbf{H}_{i_{mT}^{(k)}}^{\mathrm{T}} \boldsymbol{H}_{i_{mT}^{(k)}}\right)^{-1} \mathbf{H}_{i_{mT}^{(k)}}^{\mathrm{T}}\, \boldsymbol{r}^{(k-1)} \qquad (9)$$

- Stoping criterion

where the matrix $\mathbf{H}_{i_{mT}^{(k)}} = [\,\mathbf{h}_i,\ \mathbf{h}_{i+T}, \ldots, \mathbf{h}_{i+(K-1)T}]$ gathers the *K* selected atoms. Appendix A explains why the selection step of the Cyclo-MP (Eq. 7) is more efficient?

*Cyclo-Orthogonal Matching Pursuit (Cyclo-OMP)*

The Cyclo-OMP (or Cyclic-OMP) is the extension of the Orthogonal Matching Pursuit (OMP) [14]. The Cyclo-OMP differs from the Cyclo-MP on the updating step as an orthogonal projection of the data on the whole selected atoms is performed. This avoids the selection of already selected atoms but increases the computation cost as the amplitudes associated to all the selected atoms are updated.

- Selection: same as for the Cyclo-MP (7)
- Update:

solution:

$$\boldsymbol{x}_{\Lambda^{(k)}}^{(k)} = \left(\mathbf{H}_{\Lambda^{(k)}}^{\mathrm{T}} \mathbf{H}_{\Lambda^{(k)}}\right)^{-1} \mathbf{H}_{\Lambda^{(k)}}^{\mathrm{T}} \boldsymbol{y} \qquad (10)$$

residual:

$$\boldsymbol{r}^{(k)} = \boldsymbol{y} - \mathbf{H}_{\Lambda^{(k)}} \boldsymbol{x}_{\Lambda^{(k)}}^{(k)}$$

- Stopping criterion

where the matrix $\mathbf{H}_{i_{mT}^{(k)}} = [\,\mathbf{h}_i,\ \mathbf{h}_{i+T}, \ldots, \mathbf{h}_{i+(K-1)T}]$ gathers the *K* selected atoms. Appendix A explains why the selection step of the Cyclo-MP (Eq. 7) is more efficient?

*Cyclo-Orthogonal Matching Pursuit (Cyclo-OMP)*

The Cyclo-OMP (or Cyclic-OMP) is the extension of the Orthogonal Matching Pursuit (OMP) [14]. The Cyclo-OMP differs from the Cyclo-MP on the updating step as an orthogonal projection of the data on the whole selected atoms is performed. This avoids the selection of already selected atoms but increases the computation cost as the amplitudes associated to all the selected atoms are updated.

- Selection: same as for the Cyclo-MP (7)
- Update:

solution:

$$\boldsymbol{x}_{\Lambda^{(k)}}^{(k)} = \left(\mathrm{H}_{\Lambda^{(k)}}^{\mathrm{T}} \mathrm{H}_{\Lambda^{(k)}}\right)^{-1} \mathrm{H}_{\Lambda^{(k)}}^{\mathrm{T}} \boldsymbol{y} \qquad (10)$$

residual:

$$\boldsymbol{r}^{(k)} = \boldsymbol{y} - \mathbf{H}_{\Lambda^{(k)}} \boldsymbol{x}_{\Lambda^{(k)}}^{(k)}$$

- Stopping criterion

*Cyclo-Orthogonal Least Squares (Cyclo-OLS)*

The Cyclo-OLS (or Cyclic-OLS) is the extension of the Orthogonal Least Squares (OLS) [15]. The Cyclo-OLS differs from the Cyclo-OMP on the selecting step as the selected atoms minimize the approximation error. The Cyclo-OLS is the more coherent greedy algorithm as both in the selection and the updating steps it aims to minimize the approximation error. However, the computation cost of the Cyclo-OLS highly increases compared to the Cyclo-OMP.

- Selection: $\Lambda^{(k)} = \Lambda^{(k-1)} \cup \{i_{mT}^{(k)}; m = 1, \ldots, K-1\}$,

$$i^{(k)} = \underset{i}{\operatorname{argmin}} \|\boldsymbol{y} - \mathbf{H}_{\Lambda}\left(\mathbf{H}_{\Lambda}^{\mathrm{T}} \boldsymbol{H}_{\Lambda}\right)^{-1} \mathbf{H}_{\Lambda}^{\mathrm{T}} \boldsymbol{y}\|^2 \qquad (11)$$

with $\Lambda = \Lambda^{(k-1)} \cup \{i_{mT}^{(k)}; m = 1, \ldots, K-1\}$,

- Update: same as for the Cyclo-OMP solution (10)
- Stopping criterion

*Appendix B shows why the selection step of the Cyclo-OLS (Eq. 11) is more efficient?*

*Cyclo-Single Best Replacement (Cyclo-SBR):*

The Cyclo-SBR (or Cyclic-SBR) is the extension of the Single Best Replacement (SBR) algorithm [19, 20, 21]. The SBR algorithm is not strictly speaking a greedy algorithm. It is an iterative algorithm which aims to minimize criterion (4). The SBR has been inspired by the Single Most Likely Replacement (SMLR) algorithm [31] proposed for Bernoulli-Gaussian deconvolution. Of course, the SBR is not guaranteed to converge towards the global minimum of (4), i.e. to the sparse approximation, but it is an interesting alternative to greedy algorithms as it has a very similar iterative scheme and has been shown to give better results for deconvolution [19, 20, 21].

For a given parameter □, the selection and update steps of the Cyclo-SBR can be written as:

- Selection:

$$\Lambda^{(k)} = \Lambda^{(k-1)} \underbrace{\cup \{i_{mT}^{(k)}; m = 1, \dots, K-1\}}_{\text{Addition}}$$

or

$$\Lambda^{(k)} = \Lambda^{(k-1)} \underbrace{\setminus \{i_{mT}^{(k)}; m = 1, \dots, K-1\}}_{\text{Removal}}$$

$$i^{(k)} = \underset{i}{\operatorname{argmin}}\ \mathcal{J}_\Lambda = \underset{i}{\operatorname{argmin}} \|\boldsymbol{y} - \mathbf{H}_\Lambda \left(\mathbf{H}_\Lambda^T \mathbf{H}_\Lambda\right)^{-1} \mathbf{H}_\Lambda \boldsymbol{y}\|^2 + \beta \,\#\{\Lambda\} \quad (12)$$

$$\text{with} \begin{cases} \Lambda = \Lambda^{(k-1)} \cup \left\{i_{mT}^{(k)}; m = 1, \dots, K-1\right\} \\ \text{or} \\ \Lambda = \Lambda^{(k-1)} \setminus \{i_{mT}^{(k)}; m = 1, \dots, K-1\} \end{cases} \quad (13)$$

- Update: same as for the Cyclo-OMP solution (10)
- Stopping criterion: no replacement is accepted

This means that the selection step compares both selecting $K$ new atoms and deleting $K$ previously selected atoms, the replacement which minimizes criterion (4) being selected. In case of suppression, the set of selected atoms is updated as $\Lambda^{(k)} = \Lambda^{(k-1)} \setminus \left\{i_{mT}^{(k)}; m = 1, \dots, K-1\right\}$. Note that the selection step for adding $K$ new atoms is strictly identical to the Cyclo-OLS one, the updating step being identical to that of the Cyclo-OMP and the Cyclo-OLS. The iterations of the Cyclo-SBR stops when no replacement allows decreasing the criterion. Appendix C illustrates the reason why the selection step of the Cyclo-SBR (Eq. 12) is more efficient?

*d) Stopping rule*

The only parameter to set for using these greedy algorithms is the stopping rule. In terms of sparse approximation, comparing the norm of the residual to a threshold is a natural stopping rule, as it corresponds to an expected quality of approximation. On the other hand, for deconvolution, the residual for the true object corresponds to the noise. So a statistical test on the residual may be used as stopping rule, which decides whether the residual can be distinguished from noise. For Gaussian, centered, independent and identically distributed (*i.i.d.*) noise, of known variance $\sigma^2$, the norm $\|\boldsymbol{n}\|_2^2/\sigma^2 = \sum_{k=1}^{N} n_k^2/\sigma^2$ follows a Chi-square distribution with $N$ degrees of freedom. So, the Chi-square distribution may be used to determine the value of $\epsilon$ for which $Pr\left(\left\|\boldsymbol{r}^{(k)}\right\| \leq \epsilon\right) = \eta$ for a given probability, e.g. $\eta = 95\%$.

*e) Discussion*

- Note that for the four algorithms, the selection step is made jointly over the whole set of cycle indices $m$ thanks to cyclosparsity. Also note that the original version of the algorithms can be retrieved taking into account a single period for $m = 0$.
- An objective and meaningful comparison of the greedy algorithms and the Cyclo-SBR requires a common tuning of the parameters. This means that the parameter □ of the Cyclo-SBR has to be set in agreement with the stopping rule of the greedy algorithms. This can be done using a continuation path technique for the SBR [19, 20, 21] named CSBR (Continuation SBR). Roughly, the SBR algorithm is executed for decreasing values of $\beta = \beta_q$ and stopped at the higher value of □ for which the approximation is acceptable, i.e. with the same stopping rule as the greedy algorithms presented § 5.4. Note that the critical values $\beta_q$ for which the selected atoms may change have been shown to be a by-product of the SBR algorithm [19, 20, 21] and do not require additional computation.
- Another advantage of cyclic greedy algorithms is the significant reduction of the computation cost. Cyclic greedy algorithms select $K$ atoms at time utilizing the residual at iteration $k$ i.e. one scalar product with atoms. Unlike greedy algorithms that select one atom at time utilizing the residual at iteration $k$, to select an additional atom, the residual at iteration $(k+1)$ must be calculated and then the scalar product with atoms. In consequence, to select $K$ atoms, $K$ scalar product must be performed. Thus, the computation cost for the

selection step is roughly divided by $K$. For that matter Appendix E summarizes the computational cost of the studied greedy algorithms.

- Proceeding [32, 14, 19, 20, 21] one can use the matrix inversion lemma to compute iteratively $(\mathbf{H}^{\mathrm{T}}_{\Lambda^{(k-1)} \cup \{i_{mT}^{(k)}; m=1,\ldots,K-1\}} \mathbf{H}_{\Lambda^{(k-1)} \cup \{i_{mT}^{(k)}; m=1,\ldots,K-1\}})^{-1}$ at a low cost, from the knowledge of $(\mathbf{H}^{\mathrm{T}}_{\Lambda^{(k-1)}} \mathbf{H}_{\Lambda^{(k-1)}})^{-1}$ (see Appendix D for more details).
- The cyclic greedy algorithms have been presented with matrix notations, which are very useful to understand the algorithms but may not be used directly for their implementation in the case of large size data. For the deconvolution case, according to the dictionary and matrix structures some efficient implementation has to be accounted for to reduce the computation cost and the memory storage. As matrix H models a convolution operator, it has a Toeplitz structure and each operation involving H (Matrices and vectors operation) may be computed as a result of a convolution. An efficient implementation is proposed in [33], based on the convolution operator and not on vector and matrix products as is usually done for sparse approximations. We used therefore the same practical implementation for cyclic greedy algorithms for the deconvolution.

## VI. Simulation

### a) Description

The proposed methods are tested using synthetic signals in order to evaluate their effectiveness. To do so, we consider simulation example with the following parameters. A cyclostationary signal based on periodic random impulses $L_x = 256$ and $T = 32$, so the number of periods is $K = 8$). This input signal consists of $d = 5$ periodic random impulses of the same positions (7, 9, 11, 13 and 15) in each cycle. The signal is then filtered by an ARMA system where the transfer function is given as: $H(z) = \frac{1+b_1 z^{-1}}{1+a_1 z^{-1}+a_2 z^{-2}}$ where $b_1 = -0.6$, $a_1 = -0.9$ and $a_2 = 0.6$; the time representation of the IR $(L_h = 15)$ is reported in Fig. 2-b.

Then *i.i.d.* Gaussian noise $(L_y = 270)$ is added to the convolved signal, as illustrated by (1), such that the SNR is 14dB. The resulting signal is reported in Fig. 2-a.

For the first evaluation we consider the time representations of the reconstructed signals given by each algorithm versus the true signal. Fig. 3 reports the true signal (blue line) described by the relationship (1) and the estimated signal (colored line) for each method. We note that we constrained the plot to the three first periods in order to avoid overloading Fig. 3. Regarding each algorithm and its corresponding extension, we note from Fig. 3 that deconvolution across cyclosparsity hypothesis allows to:

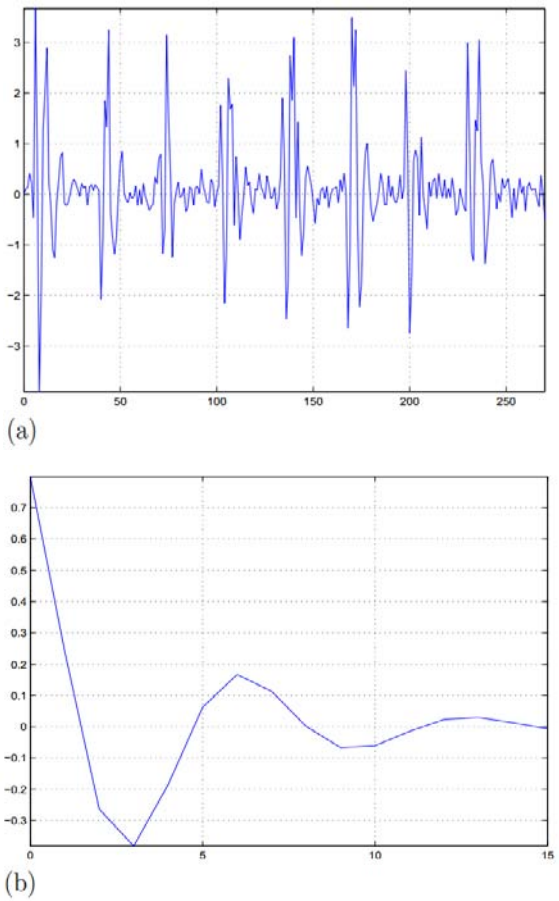



*Figure 2 (a) :* The observed signal (SNR= 14dB).
*(b):* The IR used for all simulations

- Detect and restore impulses even drowned in noise
- Reduce false and missing detections
- Estimate well the amplitude of impulses

Comparing all algorithms, we deduce that the Cyclo-OMP, the Cyclo-OLS and the Cyclo-SBR provided the best estimations. This point will be examined in detail using other evaluations (mean squared error and histogram).

### b) Mean Squared Error

We provide here a comparison between the proposed approaches against the original ones. The aim is to show the performance of cyclosparse greedy deconvolution in various *i.i.d.* noisy environments. The simulation is made with the same parameters as the first example except SNR. Actually, the SNR will vary from 1dB to 30dB. And for each value of the SNR, 500 Monte Carlo (MC) runs will be implemented. Thus, for each MC run,

- The periodic random impulses keep the same positions but with random amplitudes

- The input signal is filtered by the IR of Fig. 2-b
- *i.i.d.* Gaussian noise is added to the convolved signal, as illustrated by (1), such that the SNR is set to the desired value

Another measure adopted to evaluate our system involved changing the SNR while every other variable used in MC simulations remained constant. We aim to examine the effects of increasing noise on the performances of these methods. The evaluation quantities for our simulation study, comparing the performances of these methods, were average Mean Squared Error (MSE) and average histogram.

The MSE provides a measure of the quality of the reconstructed signal. The MSE of the estimate $\hat{\boldsymbol{x}}$ with respect to $\boldsymbol{x}$ is defined as, $\mathrm{MSE}(\hat{\boldsymbol{x}}) = \mathrm{E}[(\hat{\boldsymbol{x}} - \boldsymbol{x})]$. These MSE will be averaged over the number of MC runs. Fig. 4 shows the variation of each output'MSE, for the proposed methods and the original ones as well, with the SNR. We note from the trend of the graph that the MSE decreases with increasing SNR. This is because higher SNR implies lower noise effect on observed data $\boldsymbol{y}$. This effect allows fewer amplitude estimation errors after detection takes place; hence, good performances of the algorithms. However, we note from Fig. 4 (for $K = 8$) that higher MSE occurs for lower SNR. This is also as a result of higher noise effect for lower SNR. Also, the MSE is nearly the same for all cyclo-algorithms except Cyclo-MP, with highest MSE occurring for lower SNR.

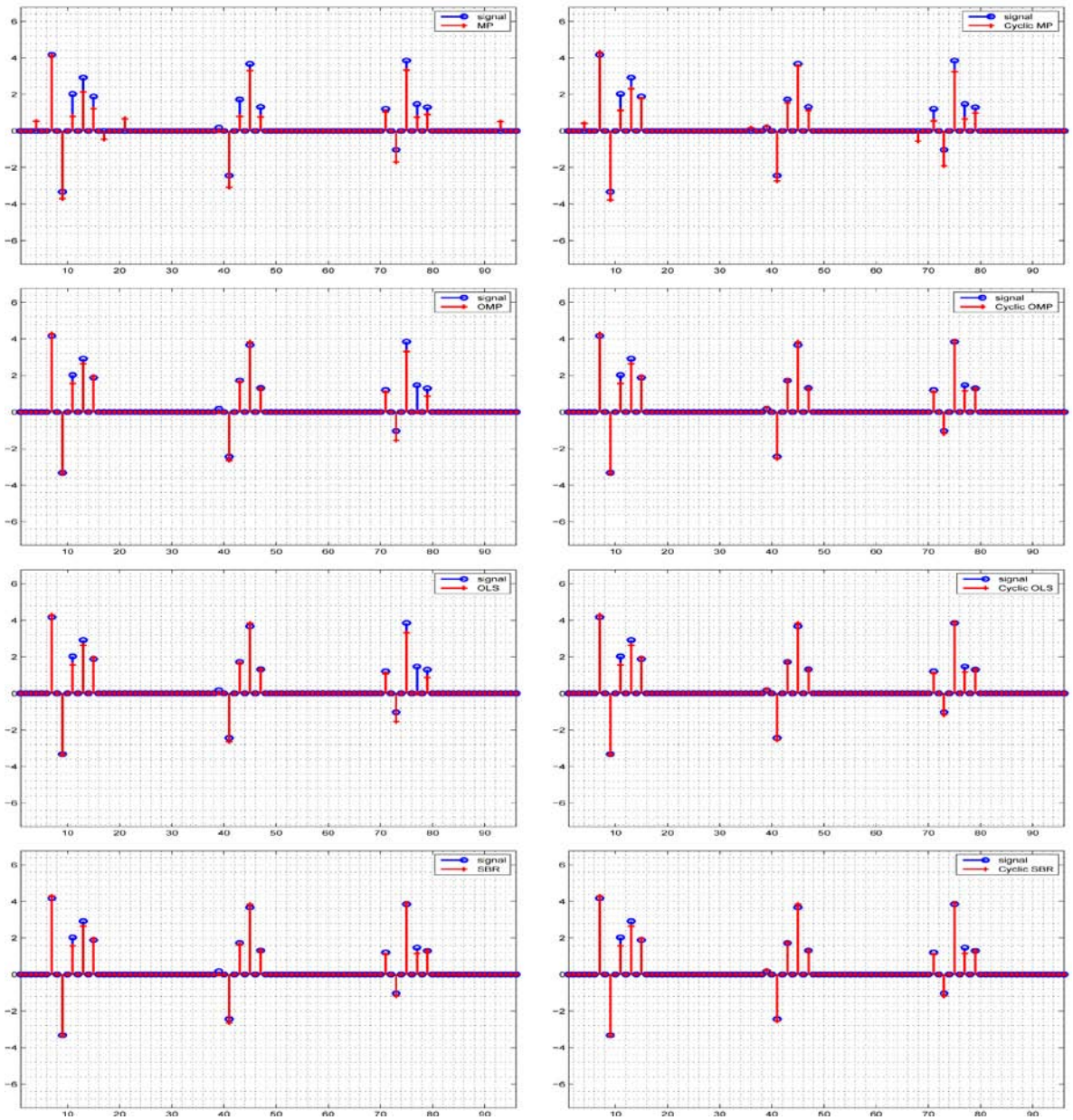



*Figure 3 :* The reconstructed signal versus the original one (the SNR is set to 14dB)

Furthermore, as can be seen from the same figure (Fig. 4, for $K = 8$), the algorithms' behavior with respect to the MSE against noise can be decomposed into two parts: SNR less or greater than 6dB. For SNR greater than 6dB, the MSE of the algorithms can be sorted in descending order as follow, MP; Cyclo-MP; (OMP, OLS and SBR); (Cyclo-OMP, Cyclo-OLS and Cyclo-SBR). However for SNR less than 6dB, the MSE of the algorithms can be sorted as, (MP, OMP, OLS and SBR); Cyclo-MP; (Cyclo-OMP and Cyclo-OLS); Cyclo-SBR. We conclude therefore, that cyclo-algorithms perform well even for lower SNR

*c) Histogram*

The histogram shows the distribution of data values. Thus, performing the histogram to the reconstructed signal will show the number of the true impulses and false/missing detections that happen within the true impulses as well. Therefore, this will help us know how much detection is in error for each algorithm. We made MC simulations in order to help us determine an average histogram over the number of 500 MC runs. Fig. 5 shows the variation of each output'average histogram obtained by varying SNR from 1dB to 30dB for the proposed methods and the original ones as well. As the histogram is almost periodic, we constrained the plot to the first period in order to avoid overloading Figures 5, 6, 7 and 8.

We note from Fig. 5 that false detections increase with decreasing SNR. This is because higher SNR implies lower noise effect on observed data $\boldsymbol{y}$. The

histogram is nearly the same for all cyclo-algorithms except Cyclo-MP, with highest missing detections occurring even for higher SNR. However, we note from Fig. 5 that higher missing/false detections occur mainly for lower SNR. This is also as a result of higher noise effect for lower SNR.

The histogram indicates good detection, false detection/alarms and missing detection. Based on these three criteria, we can classify the histogram similarly to the MSE. As can be seen from this Fig. 5, the algorithms' behavior with respect to the histogram against noise can be decomposed into two parts: SNR less or greater than 6dB. Consequently, the histogram confirms the MSE behavior of the algorithms and leads to the same sorting of the algorithms.

*d) Influence of the number of cycles*

Another parameter which can influence the performances of the cyclo-algorithms is the number of cycles/periods $K$. To examine this, we performed three simulations in which $K$ was gradually increased. The simulations are made with the same parameters as the second example except data size. Actually, changing the number of cycles means changing data size as well. So, for $K$ equal to 2, 4, 8 and 16, correspond respectively to data size 64, 128, 256 and 512. Simulation results for each data size set were reported in Fig. 4 for the average MSE and Figures 6, 7, 5 and 8 for the average histogram. As expected from theory, increasing the number of cycles leads to good performances with less false/missing detections and errors in the estimation of the impulses amplitudes, for cyclo-algorithms in comparison with their corresponding algorithms. We note also that the Cyclo-OLS and Cyclo-OMP converge gradually (especially for lower SNR) to the Cyclo-SBR as $K$ increases. This is because less error occurred in the selection step for adding new atoms, so no need for the Cyclo-SBR to correct any error on the selection step by removing already added atoms.

## VII. Discussion

The objective of the previous simulations is to evaluate the contribution of cyclosparsity for greedy deconvolution. It is apparent therefore, that deconvolution across cyclosparsity hypothesis allows detecting and restoring impulses even drowned in noise provided that these impulses being significant for the other cycles of the signal. Also, increasing the number of cycles i.e. more atoms involved in the average leads to a considerable enhancement of the performances of cyclo-algorithms.

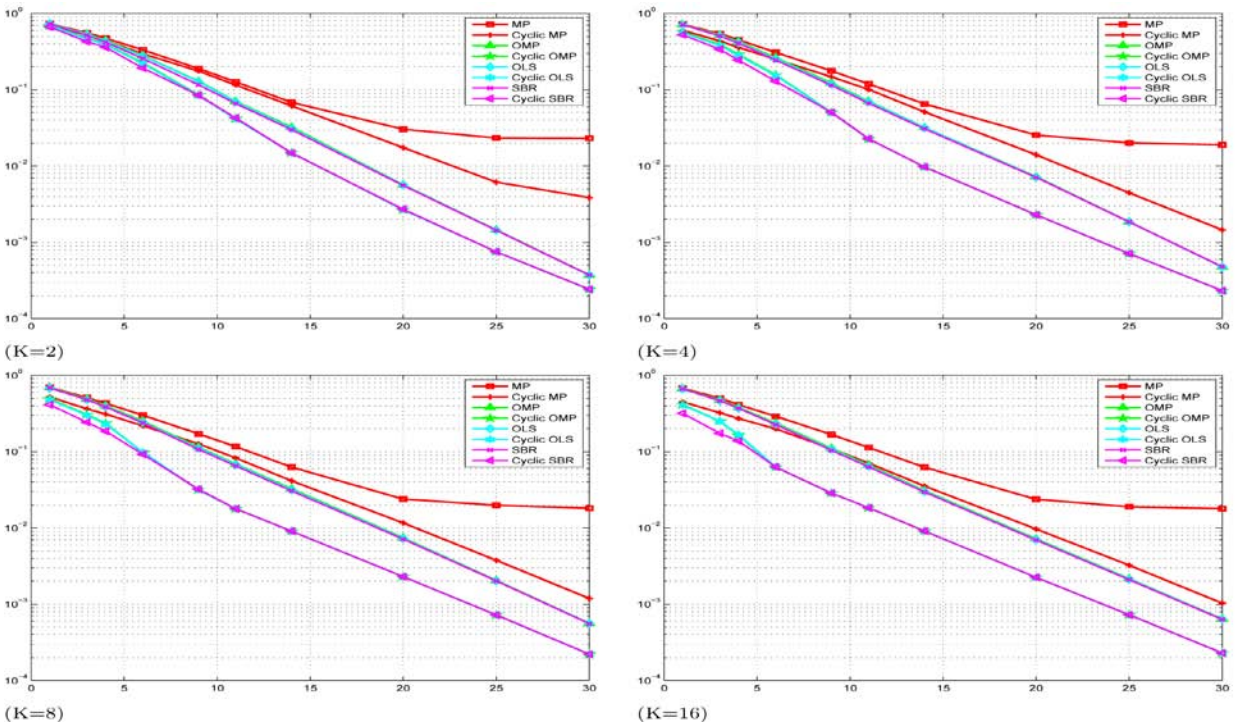



*Figure 4 :* The effect of varying SNR from 1dB to 30dB over MC runs on the MSE

The Cyclic algorithms perform better than their corresponding classical ones even for lower SNR. The Cyclo-OLS and Cyclo-OMP reach the performances of

the Cyclo-SBR, especially for higher SNR, because averaging over *m* reduces false/missing detections, so the Cyclo-SBR seldom if ever removes already added atoms. The Cyclo-MP has the bad performances. What happened to the behavior of the Cyclo-MP can be explained by the distance between adjacent impulses. When nonzero elements are so close and strongly correlated, false detections occur often because the orthogonal projections are made over only the *K* selected atoms unlike the other algorithms where the orthogonal projections are made over the whole selected atoms. However, by increasing the parameter *K*, the Cyclo-MP behavior converges to the behavior of OMP, OLS and SBR for higher SNR, whereas for lower SNR the Cyclo-MP behavior converges to the behavior of Cyclo-OMP and Cyclo-OLS. This means that cyclosparsity allows to the Cyclo-MP to overcome the drawback of the MP.

## VIII. Conclusion

- The objective of the paper is the introduction of the concept of cyclosparsity for cyclostationary signals based on periodic random impulses. Then, integrate this concept for greedy sparse algorithms in order to increase the performances of the deconvolution and reduce significantly the computation cost as well.
- The performance of the new algorithms using computer simulated cyclostationary signals was demonstrated. It is apparent therefore that the proposed methods compare favorably with the original ones. Reasons for the improved performance of the proposed methods over the original ones include the following: the cyclosparsity model makes possible the exploitation of the information given by the periodicity which allows less false alarms and missing detections as well.
- The unique additional information required by cyclic greedy algorithms is the cyclic period *T*. In general, the cyclic period is related to the studied system. For rotating machines, the cyclic period is a multiple of the shaft rotation. Furthermore, the problem of estimating the cyclic period or the cyclic frequency has been addressed in several articles as [34].
- We investigate to apply the proposed algorithms to vibratory signals namely bearing signals for diagnostic. Actually, bearing with inner rice or outer rice default signals are known to be random periodic impulse signals. These signals are convolved by the IR of the mechanical structure of the rotating machine and then noise is added to the convolved signal. Thus, the resulting signal is not legible, and hence, the need of deconvolution to restore the random periodic impulses in order to estimate the degree of the default.

*Appendix A. Comparison between the selection step of the Cyclo-MP (Eq. 7) and the one of the MP*

The selection step of the Cyclo-MP and Cyclo-OMP is given by,

$$i^{(k)} = \arg\max_{i} \sum_{m=0}^{K-1} |\mathbf{h}_{i+mT}^{\mathrm{T}} \boldsymbol{r}^{(k-1)}| \qquad (A.1)$$

Let develop the term $\sum_{m=0}^{K-1} |\mathbf{h}_{i+mT}^{T} \boldsymbol{r}^{(k-1)}|$,

$$\sum_{m=0}^{K-1} |\mathbf{h}_{i+mT}^{\mathrm{T}} \boldsymbol{r}^{(k-1)}| = \sum_{m=0}^{K-1} |C_{r^{(k-1)}h_i}(i+mT)| = \sum_{m=0}^{K-1} |\mathcal{H}^{-} * \boldsymbol{r}^{(k-1)}(i+mT)|$$

$$= \left|C_{r^{(k-1)}h_i}(i)\right| + \left|C_{r^{(k-1)}h_i}(i+T)\right| + \cdots + |C_{r^{(k-1)}h_i}(i+(K-1)T)|$$

$$= |\mathcal{H}^{-} * \boldsymbol{r}^{(k-1)}(i)| + |\mathcal{H}^{-} * \boldsymbol{r}^{(k-1)}(i+T)| + \cdots + |\mathcal{H}^{-} * \boldsymbol{r}^{(k-1)}(i+(K-1)T)|$$

where $C_{r^{(k-1)}h_i}^{T}$ represents the correlation between $\boldsymbol{r}^{(k-1)}$ and $\mathbf{H}$ and $\mathcal{H}^{-}(j) = \mathcal{H}(-j)$.

The selection step of the MP and OMP is exclusively based on the term $|C_{r^{(k-1)}h_i}^{\mathrm{T}}(i)|$ (equivalent to $|\mathcal{H}^{-} * \boldsymbol{r}^{(k-1)}(i)|$). So the selected atom is the one who maximizes it. Whereas, the Cyclo-MP and Cyclo-OMP are based on the correlation at *i* for the multiple of the cyclic period *T*, i.e.

$$\left|C_{r^{(k-1)}h_i}(i)\right| + \left|C_{r^{(k-1)}h_i}(i+T)\right| + \cdots + |C_{r^{(k-1)}h_i}(i+(K-1)T)|$$

equivalent to

$$|\mathcal{H}^{-} * \boldsymbol{r}^{(k-1)}(i)| + |\mathcal{H}^{-} * \boldsymbol{r}^{(k-1)}(i+T)| + \cdots + |\mathcal{H}^{-} * \boldsymbol{r}^{(k-1)}(i+(K-1)T)|$$

for the step selection. Thus, the *K* selected atoms should maximize jointly the sum.

Obviously, this has the advantage to avoid penalizing atoms associated to impulses with small amplitude thanks to the sum over periods that allows a joint selection of *K* atoms at once. And hence, the need to involve more atoms in the sum. So, when *K* increases the sum covers more atoms, therefore the chance to have atoms that bear on the sum increase significantly. Then less errors occur (especially when impulses are close) in the selection step.

*Appendix B. Comparison between the selection step of the Cyclo-OLS (Eq. 11) and the one of the OLS*

The selection step of the Cyclo-OLS is given by,

$$i^{(k)} = \underset{i}{\operatorname{argmin}} \|\boldsymbol{y} - \mathbf{H}_{\Lambda^{(k)}}\left(\mathbf{H}_{\Lambda^{(k)}}^{\mathrm{T}}\mathbf{H}_{\Lambda^{(k)}}\right)^{-1}\mathbf{H}_{\Lambda^{(k)}}^{\mathrm{T}}\boldsymbol{y}\|^2 \quad (B.1)$$

where $\mathbf{H}_{\Lambda^{(1)}} = [\mathbf{h}_i, \mathbf{h}_{i+T}, \ldots, \mathbf{h}_{i+(K-1)T}]$.

For this case, $\mathbf{H}_{\Lambda^{(1)}}^{\mathrm{T}}\mathbf{H}_{\Lambda^{(1)}} = \mathbf{I}_K$ with $\mathbf{I}_K$ stands for the $K\times K$ identity matrix. This is because, the IR is normalized such that $\sum_{j=1}^{L_h}\mathcal{H}_j^2 = 1$ so the columns of the matrix which correspond to shifted versions of the IR should have a constant norm $\mathbf{h}_{i+mT}^{\mathrm{T}}\mathbf{h}_{\mathrm{i+mT}} = \sum_{\mathrm{j}=1}^{L_h}\mathcal{H}_j^2$. Also, for a given $i$ the atoms $\mathbf{h}_{i+mT}$ (for all $m$) are not correlated (this is because the scalar product of these vectors for different values of $m$ is null) as the IR length $L_h$ is assumed to be smaller than the cyclic period $T$. Thus, Eq. B.2 becomes,

$$i^{(1)} = \underset{i}{\operatorname{argmin}} \|\boldsymbol{y} - \mathbf{H}_{\Lambda^{(1)}}\mathbf{H}_{\Lambda^{(1)}}^{\mathrm{T}}\boldsymbol{y}\|^2$$

*Link between the selections criteria of the Cyclo-MP/Cyclo-OMP and Cyclo-OLS*

$$i^{(1)} = \underset{i}{\operatorname{argmax}} \sum_{m=0}^{(K-1)} |\mathbf{h}_{i+mT}^{\mathrm{T}}\boldsymbol{r}^{(0)}|$$

$$= \underset{i}{\operatorname{argmax}} \|\mathbf{H}_{\Lambda^{(1)}}^{\mathrm{T}}\mathbf{r}^{(0)}\|^2$$

Which is equivalent to

$$\mathrm{i}^{(1)} = \underset{\mathrm{i}}{\operatorname{argmin}} \min_{\alpha} \|\mathbf{r}^{(0)} - \mathbf{H}_{\Lambda^{(1)}}\alpha\|^2$$

*Proof*

$$\min_{\alpha} \|\boldsymbol{r}^{(0)} - \mathbf{H}_{\Lambda^{(1)}}\alpha\|^2 = \min_{\alpha} \boldsymbol{r}^{(0)^{\mathrm{T}}}\boldsymbol{r}^{(0)} - 2\boldsymbol{r}^{(0)^{\mathrm{T}}}\mathbf{H}_{\Lambda^{(1)}}\alpha + \alpha^{\mathrm{T}}\mathbf{H}_{\Lambda^{(1)}}^{\mathrm{T}}\mathbf{H}_{\Lambda^{(1)}}\alpha$$
$$= \min_{\alpha} -2\,\boldsymbol{r}^{(0)^{\mathrm{T}}}\mathbf{H}_{\Lambda^{(1)}}\alpha + \alpha^{\mathrm{T}}\alpha \;(\text{as } \mathbf{H}_{\Lambda^{(1)}}^{\mathrm{T}}\mathbf{H}_{\Lambda^{(1)}} = \mathbf{I}_K)$$

The solution is $\alpha = \mathbf{H}_{\Lambda^{(1)}}^{\mathrm{T}}\boldsymbol{r}^{(0)}$. Therefore,

$$\min_{i} \|\boldsymbol{r}^{(0)} - \mathbf{H}_{\Lambda^{(1)}}\mathbf{H}_{\Lambda^{(1)}}^{\mathrm{T}}\boldsymbol{r}^{(0)}\|^2 = \min_{\mathrm{i}} \boldsymbol{r}^{(0)^{\mathrm{T}}}\boldsymbol{r}^{(0)} - 2\boldsymbol{r}^{(0)^{\mathrm{T}}}\mathbf{H}_{\Lambda^{(1)}}\mathbf{H}_{\Lambda^{(1)}}^{T}\boldsymbol{r}^{(0)} + \boldsymbol{r}^{(0)^{\mathrm{T}}}\mathbf{H}_{\Lambda^{(1)}}\mathbf{H}_{\Lambda^{(1)}}^{\mathrm{T}}\boldsymbol{r}^{(0)} = \min_{\mathrm{i}} -\boldsymbol{r}^{(0)^{\mathrm{T}}}\mathbf{H}_{\Lambda^{(1)}}\mathbf{H}_{\Lambda^{(1)}}^{\mathrm{T}}\boldsymbol{r}^{(0)}$$
$$= \max_{\mathrm{i}} \|\mathbf{H}_{\Lambda^{(1)}}^{\mathrm{T}}\boldsymbol{r}^{(0)}\|^2$$

As for the first iteration $\boldsymbol{y} = \boldsymbol{r}^{(0)}$ the selection step of the Cyclo-OLS for the first $K$ atoms is identical to the Cyclo-MP/Cyclo-OMP and is based on the sum of the correlation between $\boldsymbol{y}$ and $\mathcal{H}$ at $i + mT$.

For the other iterations, the selection step of the Cyclo-OLS is given by the minimization of $\|\mathbf{y} - \mathbf{H}_{\Lambda^{(k)}}\hat{\boldsymbol{x}}\|^2$ with $\hat{\boldsymbol{x}}^{(k)} = (\mathbf{H}_{\Lambda^{(k)}}^{\mathrm{T}}\mathbf{H}_{\Lambda^{(k)}})^{-1}\mathbf{H}_{\Lambda^{(k)}}^{\mathrm{T}}\boldsymbol{y}$ being the orthogonal projection of $\boldsymbol{y}$ on the atoms of index in $\Lambda^{(\mathrm{k})} = \Lambda^{(\mathrm{k}-1)} \cup \left\{i_{mT}^{(k)}; m = 1, \ldots, K-1\right\}$, and $\mathbf{H}_{\Lambda^{(k)}}\hat{\boldsymbol{x}}^{(k)}$ represents the contribution of the estimated impulses (till iteration $k$) $\hat{\boldsymbol{x}}^{(k)}$ on $\boldsymbol{y}$.

Hence, the $K$ selected atoms $\{i_{mT}^{(k)}\}$ should minimize jointly the MSE between $\boldsymbol{y}$ and $\mathbf{H}_{\Lambda^{(k)}}\hat{\boldsymbol{x}}^{(k)}$. Since the minimization is made simultaneously, the impulses of small amplitudes are not penalized if the remaining atoms (for other value of $m$) bear on the minimization of the MSE.

Consequently, the selection step of the Cyclo-OLS is more efficient than the one of the OLS which minimizes independently the MSE for each atom.

*Appendix C. Comparison between the selection step of the Cyclo-SBR (Eq. 12) and the one of the SBR*

As the Cyclo-SBR has a similar selection criterion (Eq. 12) as the Cyclo-OLS except the second term, $\beta \,\#\, \{\Lambda^{(\mathrm{k})}\}$, which does not really depend on the selected atoms as it indicates how many atoms are added. Note that for $\beta = 0$ the Cyclo-SBR is reduced to the Cyclo-OLS. Also, for the first iteration ($k$=1), there is no added atoms as $\Lambda^{(0)} = \emptyset$ and hence, removal test is inconceivable. Only addition test is possible and is identical to the Cyclo-OLS with an additional value $\beta \,\#\, \{\Lambda^{(1)}\} = \beta K$. If we leave out the second term of the criterion (Eq. 12), we conclude that Cyclo-SBR behaves in the same way as the Cyclo-OLS for the selection step (mainly for adding $K$ new atoms). Unfortunately, this is not the case for the SBR as it behaves as the OLS (for addition test). In consequence, the selection step of the Cyclo-SBR (Eq. 12) is more efficient than the one of the SBR.

*Appendix D. Use of the Matrix inversion lemma*

In addition to the matrix-vector products, the Cyclo-OMP, Cyclo-OLS and Cyclo-SBR algorithms require the inversion of the matrix $\mathbf{F}_\Lambda = (\mathbf{H}_\Lambda^{\mathrm{T}}\mathbf{H}_\Lambda)^{-1}$, (for the sake of simplicity the superscript ($k$) is omitted hereafter) with a growing set of indices □, in particular in the updating step (10) of the Cyclo-OMP and even in the selecting steps (11) and (12) of the Cyclo-OLS and Cyclo-SBR respectively. Following [32, 14, 19, 20, 21] one can take advantage of the matrix inversion lemma to compute iteratively $\mathbf{F}_{\Lambda\cup\{i,\ldots,i+(K-1)T)\}}$ at a low computation cost, from the knowledge of $\mathbf{F}_\Lambda = (\mathbf{H}_\Lambda^{\mathrm{T}}\mathbf{H}_\Lambda)^{-1}$. Indeed, using a block matrices notation, it can be shown that as $\mathbf{H}_{\hat{\Lambda}\cup\{i+mT\}} = [\mathbf{H}_{\hat{\Lambda}}|\mathbf{h}_{(i+mT)}]$ and $\hat{\Lambda} = \Lambda$, $\mathbf{F}_{\hat{\Lambda}\cup\{i+mT\}} = \left[\frac{\mathbf{F}_{11}|\mathbf{f}_{12}}{\mathbf{f}_{12}^{\mathrm{T}}|\mathbf{f}_{22}}\right]$ :

$$\text{with} \begin{cases} \mathrm{f}_{22} = (\mathbf{h}_{i+mT}^{\mathrm{T}}\mathbf{h}_{i+mT} - \mathbf{h}_{i+mT}^{\mathrm{T}}\mathbf{H}_{\hat{\Lambda}}\mathbf{F}_{\hat{\Lambda}}\mathbf{H}_{\hat{\Lambda}}^{\mathrm{T}}\mathbf{h}_{i+mT})^{-1}, \\ \mathbf{f}_{12} = -\mathrm{f}_{22}\mathbf{F}_{\hat{\Lambda}}\mathbf{H}_{\hat{\Lambda}}^{\mathrm{T}}\mathbf{h}_{i+mT}. \\ \mathbf{F}_{11} = \mathbf{F}_{\hat{\Lambda}} + \mathrm{f}_{22}\mathbf{F}_{\hat{\Lambda}}\mathbf{H}_{\hat{\Lambda}}^{\mathrm{T}}\mathbf{h}_{i+mT}\mathbf{h}_{i+mT}^{\mathrm{T}}\mathbf{H}_{\hat{\Lambda}}\mathbf{F}_{\hat{\Lambda}}, \end{cases} \quad (D.1)$$

Such notation can also be used to compute $\mathbf{F}_{\widehat{\Lambda}}$ from $\mathbf{F}_{\widehat{\Lambda}\cup\{i+mT\}}$, which is required in the selection step (12) of the Cyclo-SBR, as

$$\mathbf{F}_{\widehat{\Lambda}} = \mathbf{F_{11}} - \mathrm{f}_{22}^{-1}\mathbf{f}_{12}\,\mathbf{f}_{12}^{\mathrm{T}} \qquad \text{(D.2)}$$

It should be noted that the relations D.1 and D.2 will be repeated iteratively for each value of $m = [0, 1, \ldots, (K-1]$ with of course an updated set $\widehat{\Lambda} = \widehat{\Lambda} \cup \{i + mT\}$ when addition or $\widehat{\Lambda} = \widehat{\Lambda}\backslash\{i + mT\}$ when removal.

Using these relations, the computation of the matrix $\mathbf{F}_{\widehat{\Lambda}\cup\{i+mT\}}$, with an increasing (or eventually decreasing for the Cyclo-SBR) set of indices $\widehat{\Lambda}$ can be performed at a relative low cost.

Furthermore, the selection steps of the Cyclo-OLS and Cyclo-SBR do not require the computation of the solution but only the computation of the criterion $\mathcal{J}_{\Lambda}$ of (12) (note that eq. (11) is identical to eq. (12) for parameter $\beta = 0$) which can be updated using the previous block matrix notation, in case of addition or removal of the $K$ atoms:

$$\mathcal{J}_{\widehat{\Lambda}\cup\{i,\ldots,i+(K-1)T\}} - \mathcal{J}_{\Lambda} = \sum_{m=0}^{K-1}\{-\mathrm{f}_{22}(\boldsymbol{y}^{\mathrm{T}}\mathbf{H}_{\widehat{\Lambda}}\mathbf{F}_{\widehat{\Lambda}}\mathbf{H}_{\widehat{\Lambda}}^{\mathrm{T}}\mathbf{h}_{i+mT} - \boldsymbol{y}^{\mathrm{T}}\mathbf{h}_{i+mT})^2 + \beta\} \quad \text{with } \widehat{\Lambda} = \widehat{\Lambda} \cup \{i + mT\} \quad \text{(D.3)}$$

$$\mathcal{J}_{\widehat{\Lambda}\backslash\{i,\ldots,i+(K-1)T\}} - \mathcal{J}_{\Lambda} = \sum_{m=0}^{K-1}\{-\mathrm{f}_{22}^{-1}([\mathbf{f}_{12}^{\mathrm{T}}|\mathrm{f}_{22}]\mathbf{H}_{\widehat{\Lambda}}^{\mathrm{T}}\boldsymbol{y})^2 - \beta\} \quad \text{with } \widehat{\Lambda} = \widehat{\Lambda}\backslash\{i\ +\ mT\} \qquad \text{(D.4)}$$

where $\mathrm{f}_{22}, \mathbf{f}_{12}, \mathbf{F}_{\widehat{\Lambda}}$, and $\mathbf{H}_{\widehat{\Lambda}}$ are updated with the relations D.1 and D.2 for each value of $m$.

*Appendix E. Computational cost*

Since the computational cost of the studied cyclo-algorithms is roughly the one of their corresponding greedy algorithm divided by $K$, therefore, the computational cost is given for MP, OMP, OLS and SBR, which can be respectively retrieved from Cyclo-MP, Cyclo-OMP, Cyclo-OLS and Cyclo-SBR taking into account a single period for $m$=0. As the cost of an addition operation is generally negligible compared to a multiplication operation, only multiplication operation is considered in the computation cost. The multiplications required for each algorithm at a given iteration $k$ are summarized in table E.1. For the SBR, we suppose only addition of atoms (atom removal does not happen), this corresponds to the worst case. It should be noted that this computational cost is founded on the efficient implementation (proposed in [33]) which is based on the convolution operator and not on vector and matrix products as is usually done for sparse approximations. In other respects, many applications in signal and image processing where the computations are expensive from the execution time and from memory storage point of view use parallel approach as [36].

*Table E.1 :* multiplication required for each algorithm at a given iteration k

| | | Algorithms | | | |
|---|---|---|---|---|---|
| | Multi. | MP | OMP | OLS | SBR |
| Full convolution | $L_x L_h$ | 1 (7) | 1 (7) | 0 | 0 |
| Sparse convolution | $kL_h$ | 1 (9) | k (11) | 0 | 0 |
| Vector-Matrix product | $k^2$ | 0 | 1 (10) | 0 | 0 |
| Update **F** (addition) | $3k^2 + k$ | 0 | 1 (10) | 1 (11) | 1/0 (12) |
| Addition test | $k^2 + k + L_h$ | 0 | 0 | $L_x - k + 1$ (D.3) | $L_x - k + 1$ (D.3) |
| Removal test | $k$ | 0 | 0 | 0 | $k - 1$ (D.4) |
| Update **F** (removal) | $k^2 + k$ | 0 | 0 | 0 | 0/1 (D.2) |

A rough estimate of the maximum number of multiplications of the algorithms for a number $M$ of iterations is given in table E.2

*Table E.2 :* Maximum number of multiplications of the various algorithms for $M$ of iterations

| | |
|---|---|
| MP | $L_x L_h M + \frac{1}{2}(M+1)L_h M$ |
| OMP | $L_x L_h M + (\frac{1}{3}M^2 + \frac{1}{2}M + \frac{1}{6})L_h M + (\frac{4}{3}M^2 + \frac{5}{2}M + \frac{7}{6})M$ |
| OLS | $L_x L_h M + (\frac{1}{3}M^2 + M + \frac{2}{3})L_x M + \frac{1}{2}(-M+1)L_h M + \frac{1}{2}(-\frac{1}{2}M^3 + M^2 + 4M + 3)M$ |
| SBR | $L_x L_h M + (\frac{1}{3}M^2 + M + \frac{2}{3})L_x M + \frac{1}{2}(-M+1)L_h M + \frac{1}{2}(-\frac{1}{2}M^3 + \frac{5}{3}M^2 + 4M + \frac{7}{3})M$ |

The MP has a low computation cost, but may select several times the same atom as the amplitudes are not computed from a joint orthogonal projection. Compared to the MP, the OMP just adds, for each iteration, an orthogonal projection step to compute the amplitude of the selected atoms and *k*-sparse convolution for the residual update, so the additional computation cost is relatively low. The selection step of the OLS is based on the orthogonal projection used in the update step of the OMP, but the computation cost is dramatically reduced thanks to the use of the block matrix inversion and eq. (D.3). Finally, the SBR has a computation cost very similar to the OLS as the removing tests can be computed at a low cost using eq. (D.4).

## References Références Referencias


1. W. A. Gardner, *Introduction to random processes: with application to signals and systems,* 1986.
2. J. Antoni and R. B. Randall, "The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines," *Mechanical Systems and Signal Processing,* vol. 20, no. 2, pp. 308–331, Feb. 2006.
3. W. A. Gardner, *Cyclostationarity in communications and signal processing,* IEEE PRESS, 1993.
4. Khalid Sabri, "Cyclic Sparse Greedy Deconvolution," IJIGSP, vol. 4, no. 12, pp. 1–8, 2012.
5. K. Sabri, M. El Badaoui, F. Guillet, A. Belli, G. Millet and J.-B. Morin, "Cyclostationary modeling of ground reaction force signals," Signal Processing, vol. 90, no. 4, pp. 1146–1152, April 2010.
6. C. Andrieu, P. Duvaut, and A. Doucet, "Bayesian deconvolution of cyclostationary processes based on point processes,"in EUSIPCO, Trieste, Italy, 10-13 Sep 1996.
7. D. Hatzinakos J. Ilow, "Recursive least squares algorithm for blind deconvolution of channels with cyclostationary inputs,"in IEEE Military Communications Conference, Boston, Massachusetts, 1993, pp. 123–126.
8. D. Hatzinakos, "Nonminimum phase channel deconvolution using the complex cepstrum of the cyclic autocorrelation,"IEEE Trans. on SP, vol. 42, no. 11, pp. 3026–3042, Nov. 1994.
9. L. Cerrato and B. Eisenstein, "Deconvolution of cyclostationary signals," IEEE Trans. on Acoustics, Speech, and SP, vol. 25, no. 6, pp. 466–476, Dec. 1977.
10. K.M. Cheung and S.F. Yau, "Blind deconvolution of system with unknown response excited by cyclostationary impulses,"in IEEE, ICASSP, Detroit, MI,, 9-12 May 1995, pp. 1984–1987.
11. R. Gribonval and M. Zibulevsky, Sparse Component Analysis, chapter Handbook of Blind Source Separation: Independent Component Analysis and Applications, pp. 367–420, Pierre Comon and Christian Jutten (Editors), ELSEVIER, 2010.
12. J. Bobin, J-L. Starck, Y. Moudden, and M.J. Fadili, Advances in Imaging and Electron Physics, chapter Blind Source Separation: The Sparsity Revolution, pp. 221–298, Peter Hawkes Ed., Academic Press, ELSEVIER, 2008.
13. S. Mallat and Z. Zhang, "Matching pursuits with time frequency dictionaries," IEEE Trans. on SP., vol. 41, pp. 3397–3415, 1993.
14. Y.C. Pati, Ramin Rezaiifar, and P.S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in the 27th Annual Asilomar, Pacific Grove, CA, USA, 1993.
15. S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," International Journal of Control, vol. 50, no. 5, pp. 1873–1896, 1989.
16. S. Shaobing Chen, David L. Donoho, and Michael A. Saunders, "Atomic decomposition by basis pursuit," SIAM Journal on Scientific Computing, vol. 20, no. 1, pp. 33–61, 1998.
17. Bashkar Rao and Kenneth Kreutz-Delgado, "An affine scaling methodology for best basis selection," IEEE Transaction On Signal Processing, vol. 47, no. 1, pp. 187200, Jan. 1999.
18. Robert Tibshirani, "Regression shrinkage and selection via the lasso," Journal of the Royal Statistical Society Serie B, vol. 58, no. 1, pp. 267–288, 1996.
19. J. Duan, C. Soussen, D. Brie, and J. Idier, "On the properties of the solution path of the constrained and penalized l2-l0 problems," Tech. Rep., Centre de Recherche en Automatique de Nancy, Feb. 2009.
20. Soussen C., J. Idier, D. Brie, and J. Duan, "From bernoulli-gaussian deconvolution to sparse signal restoration," IEEE Trans. on SP, vol. 59, no. 10, pp. 4572–4584, oct 2011.
21. Duan J., Soussen C., Brie D., and Idier J., "Détection conjointe de discontinuités d'ordres différents dans un signal par minimisation de crit`ere L2-L0," in Actes 22ecoll. GRETSI, Dijon, Sep. 2009.
22. Joel A. Tropp, "Greed is good: Algorithmic results for sparse approximation," IEEE Trans. on Information Theory, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
23. D.L. Donoho, M. Elad, and V.N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," vol. 52, no. 1, pp. 6–186, Jan. 2006.
24. J.J. Fuchs, "On sparse representations in arbitrary redundant basis," vol. 50, pp. 1341–1344, 2004.

25. D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," vol. 47, pp. 28452862, 2001.
26. Massimo Fornasier and Holger Rauhut, "Recovery algorithm for vector-valued data with joint sparsity constraints," SIAM Journal on Numerical Analysis, vol. 46, no. 2, pp. 577-613, 2008.
27. J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation part I: Greedy pursuit,"*Signal Processing, special issue* "Sparse approximations in signal and image processing,", vol. 86, pp. 572–588, Apr. 2006.
28. A. J. Tropp, "Algorithms for simultaneous sparse approximation. part II: Convex relaxation," vol. 86, pp. 589–602, Apr. 2006.
29. M. Kowalski and B. Torrésani, "Sparsity and persistence: mixed norms provide simple signal models with dependent coefficients," Signal Image and Video Processing, vol. 3, no. 3, pp. 251–264, Sep. 2009.
30. C. Dossal and S. Mallat, "Sparse spike deconvolution with minimum scale," in SPARS05, Rennes, France, Nov. 2005.
31. J. J. Kormylo and J. M. Mendel, "Maximum-likelihood detection and estimation of bernoulli-gaussian processes," vol. 28, no. 3, pp. 482–488, May 1982.
32. Y. Goussard, G. Demoment, and G. Idier, "A new algorithm for iterative deconvolution of sparse spike trains," in IEEE ICASSP, Albuquerque, NM, April 1990, pp. 1547–1550.
33. Khalid Sabri and Hervé Carfantan, "High resolution sparse spike hyperspectral images deconvolution," Submitted to IEEE Journal of Selected Topics in Signal Processing, 2014.
34. A.V. Dandawate and G.B. Giannakis, "Statistical tests for presence of cyclostationarity," IEEE Trans. on Signal Processing, vol. 42, no. 9, pp. 2355 – 2369, Sep. 1994.
35. M. Ehler, "Shrinkage rules for variational minimization problems and applications to analytical ultracentrifugation," Journal of Inverse and Ill-Posed Problems, vol.19, no. 4-5, pp. 593–614, 2011.
36. D. M. Onchis, "Signal Reconstruction in Multi-Windows Spline-Spaces Using the Dual System," IEEE Signal Processing Letters, Vol. 19, no. 11, 729–732, Nov. 2012.
37. E. J. Candes and C. Fernandez-Granda, "Towards a Mathematical Theory of Super-Resolution," journal CoRR, vol.abs/1203.5871, 2012.
38. Bayesian Approach to Inverse Problems, jISTE Ltd and John Wiley & Sons Inc, April 2008.

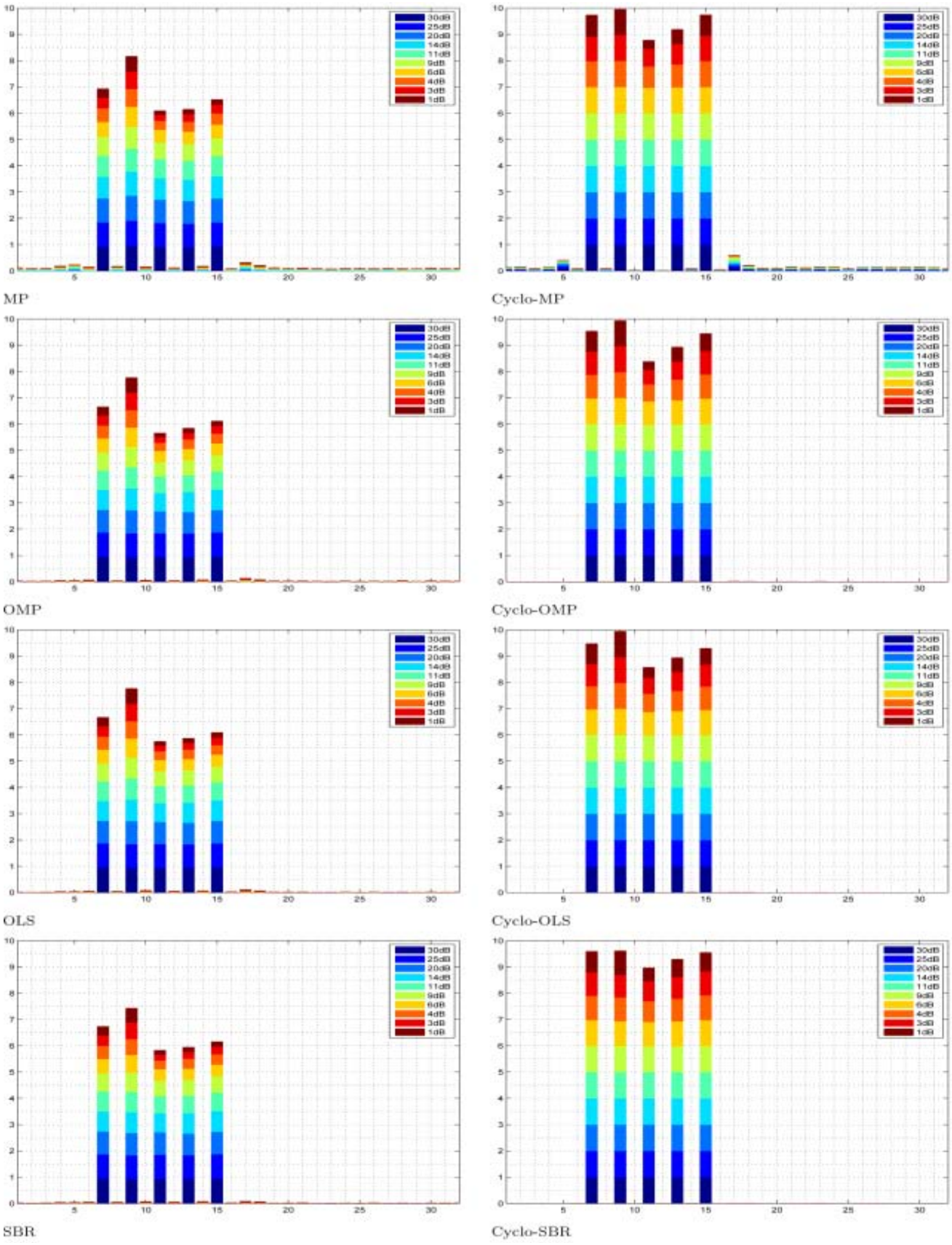



*Figure 5 :* The effect of varying SNR from 1dB to 30dB over MC runs on the histogram (K=8)

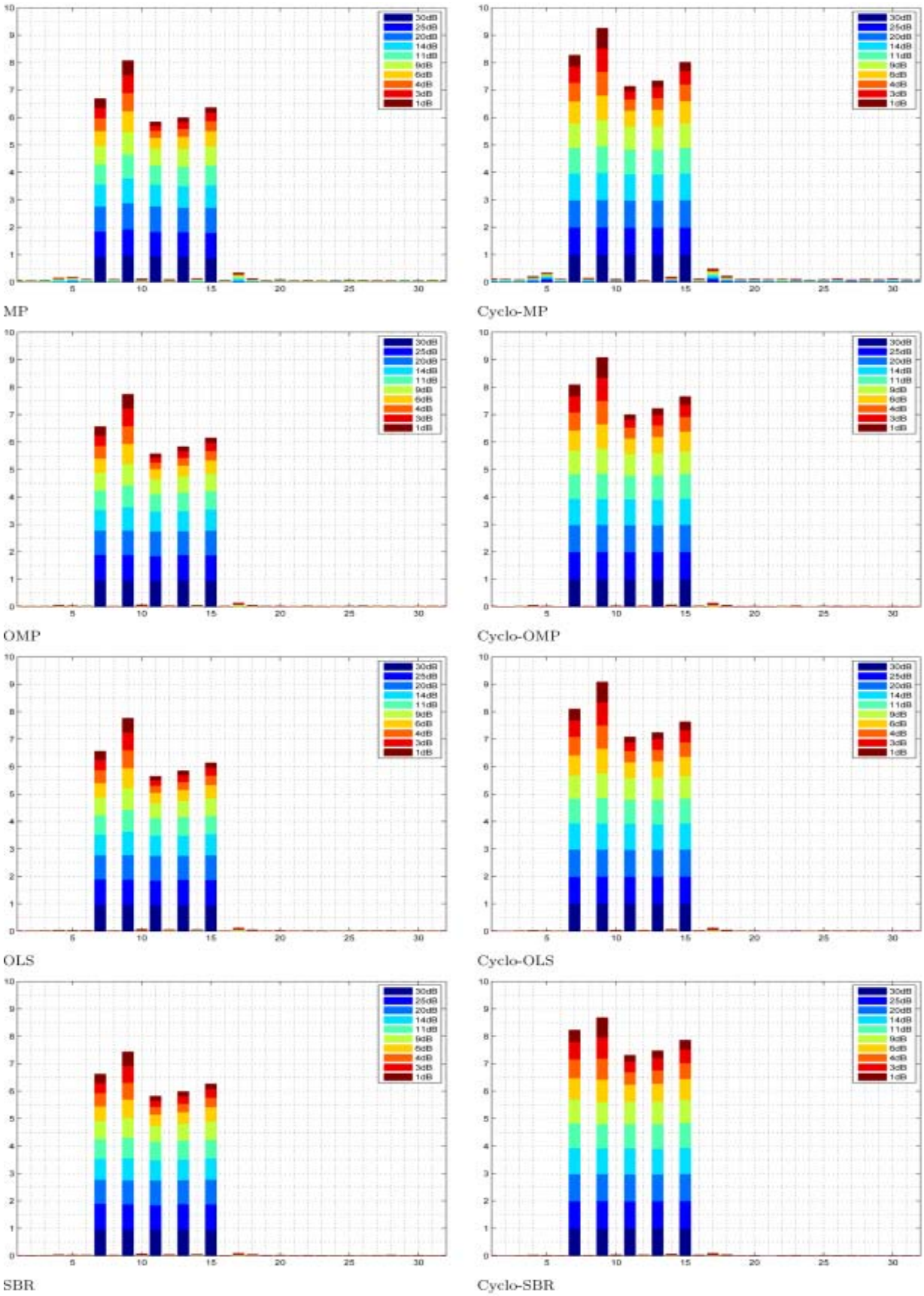



*Figure 6 :* The effect of varying SNR from 1dB to 30dB over MC runs on the histogram (K=2)

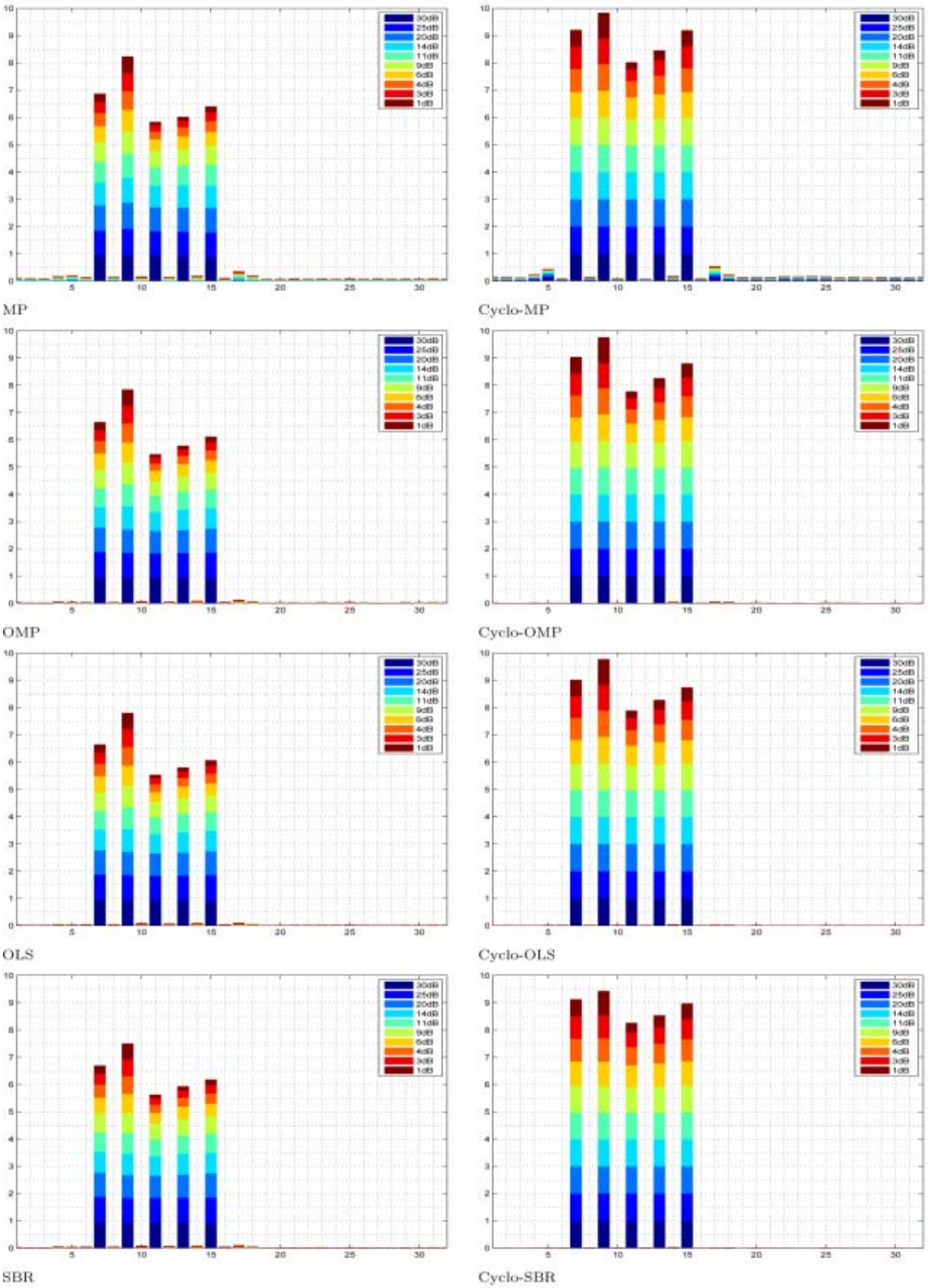



*Figure 7 :* The effect of varying SNR from 1dB to 30dB over MC runs on the histogram (K=4)

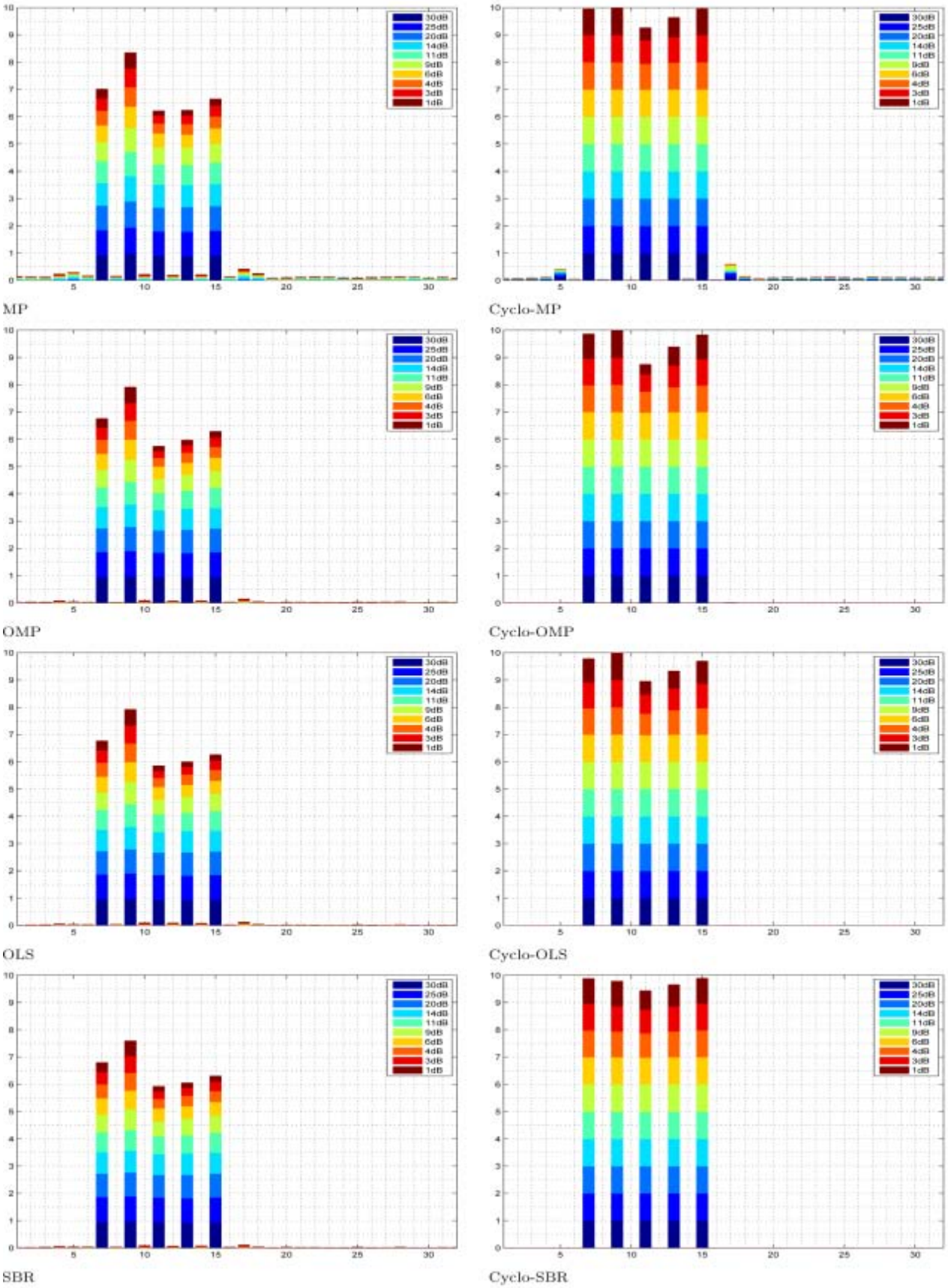



*Figure 8 :* The effect of varying SNR from 1dB to 30dB over MC runs on the histogram (K=16)

This page is intentionally left blank

Global Journal of Computer Science and Technology: F
Graphics & Vision
Volume 14 Issue 4 Version 1.0 Year 2014
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

# Zone-Features' based Nearest Neighbor Classification of Images of Kannada Printed and Handwritten Vowel and Consonant Primitives

By Basavaraj. S. Anami, Deepa S Garag & P. Shivakumara
*K.L.E. Institute of Technology, India*

*Abstract-* The characters of any languages having scripts are formed by basic units called primitives. It is necessary to practice writing the primitives and their appropriate combinations while writing different characters. In order to automate character generation, primitives" recognition becomes important. In this paper, we propose a zone-features based nearest neighbor classification of Kannada printed and handwritten vowel and consonant primitives. The normalized character image is divided into 49 zones, each of size 4x4 pixels. The classifier based on nearest neighbor using Euclidean distances is deployed. Experiments are performed on images of printed and handwritten primitives of Kannada vowels and consonants. We have considered 9120 images of printed and 3800 images of handwritten 38 primitives. A K-fold cross validation method is used for computation of results. We have observed average recognition accuracies are in the range [90%, 93%] and [93% to 94%] for printed and handwritten primitives respectively. The work is useful in multimedia teaching, animation; Robot based assistance in handwriting, etc.

*Keywords: classification, feature extraction, K_Fold cross validation, majority voting, nearest neighbor, printed primitives, handwritten primitives.*

*GJCST-F Classification : I.4.0*


*Strictly as per the compliance and regulations of:*

# Zone-Features' based Nearest Neighbor Classification of Images of Kannada Printed and Handwritten Vowel and Consonant Primitives

Basavaraj. S. Anami [α], Deepa S Garag [σ] & P. Shivakumara [ρ]

*Abstract -* The characters of any languages having scripts are formed by basic units called primitives. It is necessary to practice writing the primitives and their appropriate combinations while writing different characters. In order to automate character generation, primitives" recognition becomes important. In this paper, we propose a zone-features based nearest neighbor classification of Kannada printed and handwritten vowel and consonant primitives. The normalized character image is divided into 49 zones, each of size 4x4 pixels. The classifier based on nearest neighbor using Euclidean distances is deployed. Experiments are performed on images of printed and handwritten primitives of Kannada vowels and consonants. We have considered 9120 images of printed and 3800 images of handwritten 38 primitives. A K-fold cross validation method is used for computation of results. We have observed average recognition accuracies are in the range [90%, 93%] and [93% to 94%] for printed and handwritten primitives respectively. The work is useful in multimedia teaching, animation; Robot based assistance in handwriting, etc.

*Keywords: classification, feature extraction,* K_Fold *cross validation, majority voting, nearest neighbor, printed primitives, handwritten primitives.*

## I. Introduction

Natural language processing (NLP) is a field of computer science that deals with understanding and generation of natural languages. Natural language understanding enables computers to understand the natural language and extract meaning from it. Natural language generation involves both spoken and written information. Some applications of both text to speech and speech to text conversion are man-machine interfaces to computers, systems that read and understand printed and hand written text, speed understanding system, text analysis and understanding systems,computer aided instruction systems etc. Ediphones and Dictaphones are examples of speech to text conversion systems.

India is a multi lingual and multi script country with 22 scheduled languages. Every state has its language andcertain dialects. To name a few, Karnataka state has Kannada, Maharashtra has Marathi, Uttarpradesh has Hindi and Gujarath has Gujarathi. These languages are written using only twelve scripts. Devnagiri script used to write Hindi, Marathi, and Sanskrit. All other languages have their own script. Most of the Indian languages are derived from Ancient Brahmi script, Kannada is one amongst them and are called phonetic languages. Therefore, writing these scripts map the sounds of alphabet to specific shapes. The text in all these languages, except Urdu, is written from left to right. Kannada is a scheduled language in India and is spoken by about 60 million people of the south Indian state of Karnataka. The Kannada alphabets are developed from the Kadamba and Chalaukya scripts, descendents of Brahmi script, which are used between the 5th and 7th centuries A.D. The basic structure of Kannada script is distinctly different from the Roman script. Kannada script evolved primarily from stone carving; hence most of the characters are round, curvy and symmetric with straight strokes/wedges. All these characters in a word are isolated.

Kannada script has compound characters and hence different from English. In addition to this, upper and lower case characters are absent in Kannada script. The typical Kannada text looks like as given in Box1. The content of thebox 1 is read as "udaharanegalu" and its English equivalent is "Examples".


*Box 1 :* Sample Kannada word and its primitives

*Author α, σ : K.L.E. Institute of Technology, Hubli-580030, Karnataka .*
*Author ρ: University of Malaya (UM), Kuala Lumpur-50603, Malaysia .*
*e-mails: anami_basu@hotmail.com, deepagarag@rediffmail.com, shiva@um.edu.my.*

The alphabet of a language is divided into basic characters called vowels and consonants. Two or more basic characters are combined to form compound characters. We recall here the way we have learnt the alphabet and started writing the text in a language, in kindergarten schools, children are made to practice writing the characters and thereby memorize them. The character writing involves combining the primitives. Every character written from an alphabet follows a definite way and depends upon the type of the writer, whether left-hand-writer or right-handwriter. The character generation at once is different from how it is written. Each character in the alphabet has definite way of writing it and is combination of sub parts called primitives. This combination of primitives is a systematic approach in generating or building the characters. The character construction is basic to any medium of learning. While reading, we read the whole character and while writing or constructing a character, we write primitives in an order.

The automation of character construction requires the recognition of primitives from the database of primitives in a language. Many researchers have worked on character recognition, wherein the whole character is considered as one single unit. The focus of the present work is to recognize the images of primitives of Kannada language useful in the construction of characters using syntactic approach. The work is useful for novice learners, multimedia applications, translate-ration and translation etc. The automated script writing and learning by taking technological leverage is considered a new area of research.

In this paper, we have considered the different font types and font sizes of vowels and consonants characters supported by Kannada language software, namely, Nudi and Baraha. We have identified with the help of language experts the primitives of vowels and consonants and manually separated and their images are stored. These images of primitives are preprocessed through binarization, thinning and resizing. The simple zone based features are obtained for these primitives. Nearest neighbor classification is adopted with Euclidean distance measure for recognition of primitives. We have tested for all the combinations of printed primitives with different fonts" types and sizes.

The remaining part of the paper is organized into four sections. Section 2 deals with detailed survey on automatic primitive recognition. Section 3 deals with the proposed methodology, wherein different stages of the methodology are discussed. The experimental results and discussion are given in section 4. Conclusion and Future work are given in Section 5.

## II. Literature Survey

To know the state-of-the-art in automatic primitive recognition, we carried out the literature survey and following isthe gist of cited papers.

[Leena R Ragha, et. al, 2010] have investigated the moments features on Kannada handwritten basic character set of 49 letters. Four directional images using Gabor wavelets from the dynamically preprocessed original images are found. Then moments features are extracted from them. The comparison of moments features of 4 directional images with original images when tested on Multi Layer Perceptron with Back Propagation Neural Network shows an average improvement of 13% from 72% to 85%. The mean performance of the system with these two features together obtained is 92%.

[Karthik Sheshadri et.al, 2010] have proposed Kannada Character Recognition method based on k-means clustering. A segmentation technique to deco-mpose each character into components from three base classes is used to reduce the magnitude of the problem. The k-Means clustering technique provides a natural degree of font independence and this is used to reduce the size of the training data set to about one-tenth of those used in related works. Accuracy comparisons with related work, shows that the proposed method yields a better peak accuracy. The relative merits of probabilistic and geometric seeding in k-means are also discussed.

[Leena R Ragha, et. al, 2011] have presented the use of moments features on Kannada Kagunita. Four directional images are found using Gabor wavelets from the dynamically preprocessed original image. The Kagunita set is analysed and the regions with vowel and consonant information are identified and cut from the preprocessed original image to form a set of cut images. Moments and statistical features are extracted from original images, directional images and cut images. These features are used for both vowel and consonant recognition using multi-layer perceptron with backpropagation learning.

[Sangame S.K, et. al, 2011] have presented an unconstrained handwritten Kannada basic character recognition using invariant moments and chain code features. Invariant moments feature are extracted from zoned images and chain code. A Euclidean distance based K-NN classifier is used to classify the handwritten Kannada vowels. The method is invariant to size, slant, orientation and translation.

[B.V.Dhandra et. al, 2011] have proposed zone based features for recognition of the mixer of handwritten and printed Kannada digits. The kNN and SVM are used to classify the mixed handwritten and printed Kannadadigits. The reported recognition rates are 97.32% and 98.30% for mixed handwritten and printed Kannada digits using KNN and SVM classifiers respectively.

[G. G. Rajput et.al, 2011] have discussed the implementation of shape based features, namely, Invariant Fourier descriptors and normalized chain codes for OCR of binary images of Kannada hand-written characters. SVM classifier is used for recognition.

The result is computed using five-fold cross validation. The mean performance of the recognition reported for the two shape based features together is 98.45% and 93.92%, for numeral characters and vowels, respectively. The mean recognition rate of 95% is obtained for both vowels and characters taken together.

[K S Prasanna Kumar et. al, 2012] have presented an algorithm to optical character recognition (OCR) for Kannada numerals. The segmentation of a numeral into four equal parts and one of these parts i.e., left bottom segment, is used to extract recognition features. A conflict resolution algorithm is proposed to resolve the conflicting features. A minimum number of features are extracted so as to improve the response time.

[Umapada Pal et. al, 2012] have given a state-of-the-art survey about the techniques available in the area of offline handwriting recognition (OHR) in Indian regional scripts. Various feature extraction and classification techniques associated with the offline handwriting recognition of the regional scripts are discussed in this survey. A separate section is dedicated to the observations made, future scope, and existing difficulties related to handwriting recognition in Indian regional scripts.

[Kauleshwar Prasad, et.al, 2013] have focused on recognition of English alphabet in a given scanned text document with the help of neural networks. The first step involves image acquisition that involves the sub steps,namely, noise filtering, smoothing and normalization of scanned image. The second step, decomposition, involves segmentation of acquired image into sub images. The features improve recognition rate and misclassification. Character extraction and edge detection algorithms are devised. Neural network is used to classify and recognize the handwritten characters.

[Mamatha Hosalli Ramappa, et.al, 2013] have examined a variety of feature extraction approaches and classification methods for different optical character recognition applications. These approaches are designed to recognize handwritten numerals of Kannada script. Eight different features are computed from zonal extraction, image fusion, radon transform, etc and ten different classifiers like Euclidean distance, Chebyshev distance, etc are deployed. Artificial Immune system and classifier fusion are considered.

[Om Prakash Sharma et. al, 2013] have presented recent trends and tools used for feature extraction that helps in efficient classification of the handwritten alphabets. The use of Euler Number in addition to zoning increases the speed and the accuracy of the classifier as it reduces the search space by dividing the character set into three groups.

[G. G. Rajput et.al, 2013] have proposed a zone based method for recognition of handwritten characters in Kannada language. The normalized character image is divided into 64 zones and each is of size 8x8 pixels. For each zone,from left to right and from top to bottom, the crack code, representing the line between the object pixel and the background (the crack), is generated by traversing it in anticlockwise direction. A feature vector of size 512 is obtained for each character. A multi-class SVM is used for the classification purpose. The data set has 24500 images with 500 samples of each character. Five-fold cross validation is used and yielded 87.24% recognition accuracy.

[Swapnil A. Vaidya et. al, 2013] have given an overview of the ongoing research in OCR systems for Kannada scripts. They have provided a starting point for the researchers in the field of OCR. The state-of-the-art OCR techniques used in recognition of Kannada scripts, recognition accuracies and the resources available are discussedin fair detail.

[H. Imran Khan et. al, 2013] have proposed a chain code based feature extraction method for developing HCR system. A eight -neighborhood method is implemented, which allows generation of eight different codes for each character. These codes are used as features of the characters‟ images. These features are used for training and testing the k-Nearest Neighbor (KNN) classifier.

[Nithya E. et.al, 2013] have proposed an OCR system for complex printed Kannada characters. The input to the system is a scanned image of a page of text containing complex Kannada characters and the output is in a machine editable form. The pre-processing step converts the input document into binary form. The lines from the document image are extracted and further segmented into the lines, characters and sub characters. The histogram and connected component methods are used for segmentation and correlation is used for recognition of characters.

[Mamatha.H.R. et, al, 2013] have attempted to measure the performance of the classifier by testing with twodifferent datasets of different sizes. A framework based on the combined concepts of decision fusion and feature fusion for the isolated handwritten Kannada numerals classification is proposed. The combined approach has increased the recognition accuracy by 13.95%.

[Manjunath A. E., et.al, 2013] have proposed a Kannada OCR (Optical Character Recognition) in hand held devices. The work involves Kohonen‟s algorithm. The Kohonen network is trained with initial images. The images are thinned Hilditch algorithm. The distortions present in the images are eliminated and images are converted to grey scale images. The grey scale images are segmented and the result is displayed along with a vocal output.

[Sandhya.N et.al, 2014] have proposed a new classification method for Kannada characters, which is

used as a preliminary step for recognition. An analysis of Kannada characters is carried out. The syntactic features are identified. At first, the basic features and their exact positions in the characters are identified and recorded. Further, by using a decision tree, the characters are classified. The experimental results show that the syntactic based method using basic features gives good contribution and reliability for Kannada character classification.

From the literature survey, it is observed that researchers have worked on Kannada character recognition. The feature extraction techniques such as template matching, Zernike moments, geometric moment invariants, directional, positional, Fourier transform, etc are used. The classification techniques such as, neural network, support vector machines, nearest neighbor, etc are used. No specific work is observed on recognition of Kannada language primitives in the light of character construction is cited in the literature [1-18]. This is the motivation for the present work on printed and handwritten Kannada language vowels and consonants primitives" recognition.

## III. Proposed Methodology

The proposed methodology consists of four major steps, namely, identifying the primitives in printed and handwritten Kannada vowels and consonants and obtaining their images, preprocessing of primitives" images, primitives" recognition and classification of primitives. The steps are shown in the Figure 1.

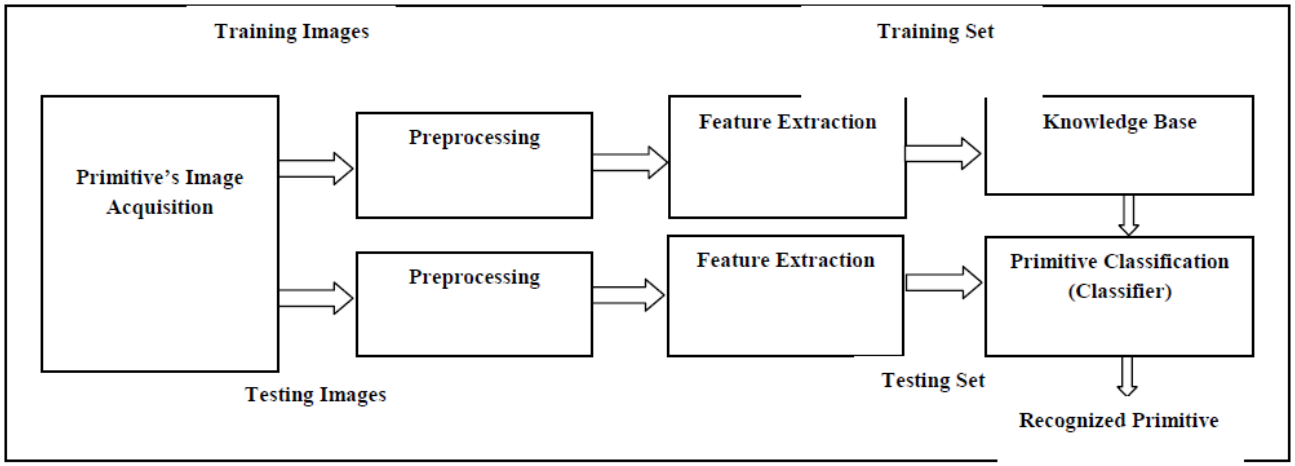



*Figure 1 :* Phases in Primitive Recognition

### *a) Identified Primitives and Image Acquisition*

This step consists of two tasks, namely, identifying primitives of printed as well as handwritten Kannada vowels and consonants and obtaining their images.

i. *Identified Primitives*

The Kannada language characters are classified into Swaras (vowels), Vyanjanas (consonants), Yogavaahakas (partly vowels and partly consonants), Kagunitha (combination of consonants and vowels) and Wothakshara (conjunct consonants) as given in Table 1. Kannada language script consists of more than 250 basic, modified and compound character shapes giving rise to 18511 distinct characters. We have used the word Kannada and Kannada language interchangeably in this paper. Kannada characters are curve shaped with some regions highly denser than others. Some shapes are wider and some are longer than others, as visible in Table1.

*Table 1 :* Kannada Vowels, Consonants and Sample Kagunitha

| Vowels | Yogavaahaka's | | Consonants | | Kagunitha |
|---|---|---|---|---|---|
| Swaras | Anusvara | Visarga | Structured | Unstructured | |
| ಅ ಆ ಇ ಈ ಉ ಊ ಋ ಎ ಏ ಐ ಒ ಓ ಔ | ಅಂ | ಅಃ | ಕ ಖ ಗ ಘ ಙ ಚ ಛ ಜ ಝ ಞ ಟ ಠ ಡ ಢ ಣ ತ ಥ ದ ಧ ನ ಪ ಫ ಬ ಭ ಮ | ಯ ರ ಲ ವ ಶ ಷ ಸ ಹ ಳ | ಕ ಕಾ ಕಿ ಕೀ ಕು ಕೂ ಕೃ ಕೆ ಕೇ ಕೈ ಕೊ ಕೋ ಕೌ ಕಂ ಕಃ |

We have consulted the Kannada language experts and identified the primitives. These 38 primitives are categorized into basic primitives (BP) and character cum primitives (CcP) as shown in the Figure 2. A single primitive, which is also a complete vowel or complete consonant, is defined as Character cum Primitive. One or more basic primitives are joined at appropriate positions to form the given vowels and consonants. It is also observed that symmetry exists in most of the Kannada characters.

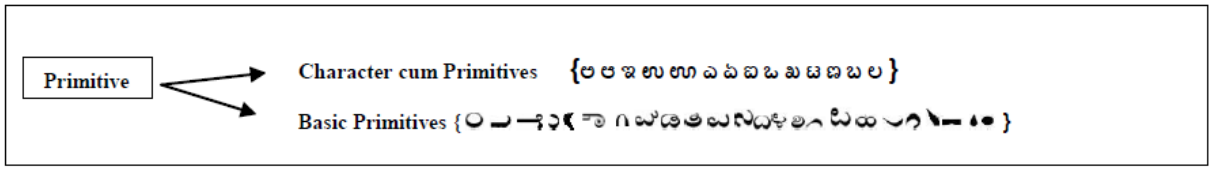



*Figure 2 :* BP and CcP in Kannada Vowels and Consonants

ii. *Image Acquisition of Primitives*

Since there is no standard database available for Kannada primitives, we have created a database of primitives for printed and handwritten Kannada vowels and consonants in consultation with language experts. We have identified 38 primitives to construct all the Kannada vowels and consonants and are given in Table 2. For example, the basic primitives required for constructing the two vowels ಈ (pronounced as „e") and ಔ (pronounced as „aou") are given in Box2. The „+" symbol represents the connection of primitives at appropriate positions for constructing a character. A database of 9120 primitives" images is prepared consisting of different font sizes and font types of printed Kannada vowels and consonants. Handwritten 3800primitives" images are collected from 100 writers who have written all the 38 primitives on plain A4 sheet. These sheets are scanned through HP Laser Jet M 1213nf MFP scanner having resolution of 300 dpi. The primitives are extracted manually from the printed as well as handwritten Kannada vowels and consonants and the images are stored as binary images in .jpg format.

*Table 2 :* List of Kannada Vowel and Consonant Primitives

| Primitive Number | Printed Primitive | Handwritten Primitive | Primitive Number | Printed Primitive | Handwritten Primitive | Primitive Number | Printed Primitive | Handwritten Primitive | Primitive Number | Printed Primitive | Handwritten Primitive |
|---|---|---|---|---|---|---|---|---|---|---|---|
| P1 | | | P11 | | | P21 | | | P31 | | |
| P2 | | | P12 | | | P22 | | | P32 | | |
| P3 | | | P13 | | | P23 | | | P33 | | |
| P4 | | | P14 | | | P24 | | | P34 | | |
| P5 | | | P15 | | | P25 | | | P35 | | |
| P6 | | | P16 | | | P26 | | | P36 | | |
| P7 | | | P17 | | | P27 | | | P37 | | |
| P8 | | | P18 | | | P28 | | | P38 | | |
| P9 | | | P19 | | | P29 | | | | | |
| P10 | | | P20 | | | P30 | | | | | |

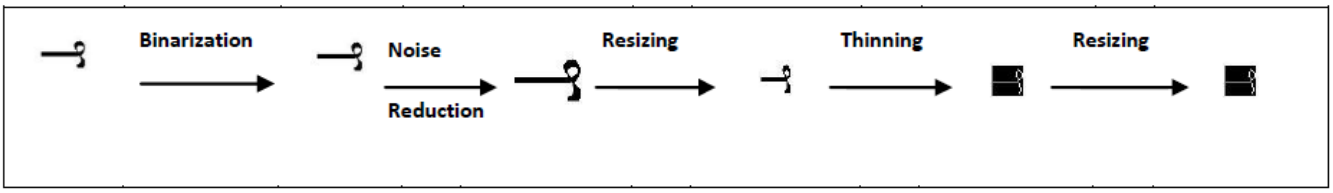
*Box 2 :* Sample Kannada Vowels and required Primitives

| Character | Primitives |
|---|---|
| ಈ | 0 + ು + ೀ |
| ಔ | ಒ + ೗ |

*b) Preprocessing*

We have preprocessed the images of the primitives to make them suitable for feature extraction. Preprocessing involves binarization, noise reduction, size normalization and thinning. The binarization is categorized into two main classes, namely, global and local. We have adopted global approach for converting gray scale image to binary image. Image binarization is performed using Otsu's method. The salt and pepper noise present in the image is removed by applying median filter. The process of thinning involves reducing thickness of each line of the pattern to just a single pixel wide is carried out. Size normalization is required as the size of the primitives vary from one vowel to another. In

order to bring uniformity among the images of primitives, each image is normalized to the size of 28*28 after finding the bounding box of each image without disturbing the aspect ratio using bilinear standard transformation. The images of primitive obtained after applying all the preprocessing steps to a given sample primitive is shown in Box 3.




*Box 3 :* Preprocessing steps

*c) Feature Extraction and Knowledge Base*

We have used zone based feature extraction technique. The number of zones and their sizes are decided based on the classifier accuracy. Figure 3 shows the behavior of the classifier for image sizes ranging from 20*20 to 40*40. Since the primitives, with the image sizes 28*28, have given maximum classification accuracy, the image size of 28*28 is chosen for all the primitives. Figure 4 shows the behavior of the classifier for different zone sizes for the given image size of 28*28. Since the zone size of 4*4 gives maximum accuracy, this size is chosen and the image is divided into 49 zones (7*7) and each is of size 4*4. The information present in each zone is used in defining the features. Figure 5a shows the 7*7 zones of the primitive P7 ( ) and Figure 5b shows the zone feature values obtained for the primitive P7 ( ).

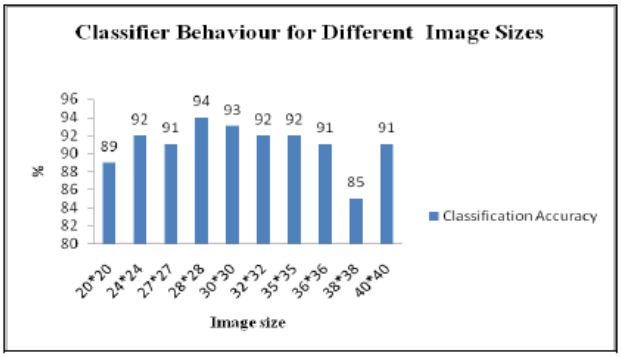



*Figure 3 :* Behavior of the Classifier for different Image Sizes

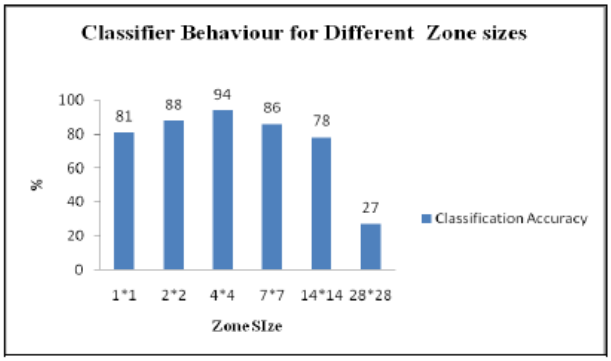



*Figure 4 :* Behavior of the Classifier for different Zone Sizes

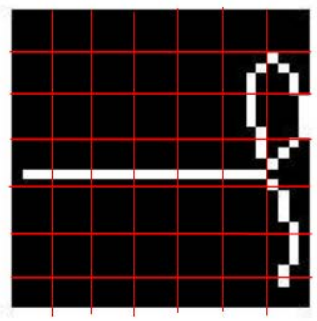


*Figure (a) :* 7×7 Zones of primitive P9( )

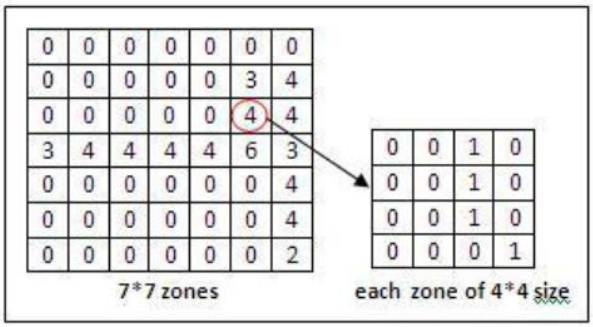



*Figure (b) :* Sample zone feature values for the primitive P9( )

*Figure. 5 :* Intermediate steps for feature extraction

A total of 49 features values are extracted from each primitive and this will serve as the feature vector. The feature vector for image (i ) denoted by Fi = {z1, z2, z3 … z49}, where zi denotes ith zone value. There could be some zones, which do not contain any part of the primitive at all; therefore the corresponding zone value in the feature vector is zero. The set of feature vectors obtained from the training samples is used as the Knowledge Base (KB). This knowledge base is used to recognize the test samples. We have used nearest neighbor classifier for recognition.

*d) Classification*

Firstly, a training set consisting of images whose class labels are known is formed. The training set is used to build a classification model which is subsequently applied to the test set, which consists of images with unknown class labels. Evaluation of the performance of a classifier is done based on the counts of test images being correctly classified. The classifier used in the work is the nearest neighbor classifier. It considers each given input pattern and classifies itto a certain class by calculating the distance between the input pattern and the training patterns. The decision is generally based on the class values of the nearest neighbors. In this work, we have computed the distance between features of the test sample and the features of every training sample using the L2 Euclidean distance measures given inBox 4.

$$d(p,q) = \left(\sum_{k=1}^{n} | p_k - q_k |^r\right)^{\frac{1}{r}} \quad \ldots\ldots \quad (1)$$

Where r is a parameter varied from 1 to ∞, n is the number of attributes and $p_k$ and $q_k$ are, respectively, the $k^{th}$ attributes (also called components) of the images p and q. $r = 2$ $L_2$ - Euclidean distance

*Box 4 :* Distance Measures used

The minimum distance between the test image and training image data in the knowledge base is used to decide the type of the primitive. A nearest neighbor classifier using Euclidean distance measure is used for the classification purpose. The procedure adopted for primitive recognition is given in the form of Algorithm1.

i. *Algorithm 1: Recognition of primitives*

Input : Kannada vowel primitive images.
Output: Recognition of the primitive.
Description: Zone based features and the nearest neighbor classifier are used in the work. Each zone is of 4x4 pixels. Euclidean distance measure is being used.

Start

*Step 1.* Accept and preprocess the input image to eliminate the noise using median filter and perform thinning.

*Step 2.* Fit the input image in a bounding box and crop the image to resize to 28*28 pixels.

*Step 3.* Extract 49 zone values, define feature vector and store.

*Step 4 .* Repeat steps 1 to 3 until the training images are exhausted.

*Step 5.* Accept the test sample.

*Step 6.* Compute Euclidean distance of the test sample with all the trained images.

*Step 7.* Declare the class of the primitive as the class with minimum distance using Nearest neighbor
classification.

*Step 8.* Repeat steps 5 and 7 until test images are exhausted.

*Step 9.* Obtain the accuracy of classification.

Stop.

The accuracy of the classifier is evaluated through k-fold cross-validation method. In this method, each time one of the k subsets of images is used as the test set of images and the other (k-1) subsets are put together to form a training set of images. The advantage of this method is that it matters less how the images are divided. The variance of the resulting estimate is reduced as k is increased. We have computed the average error across all k trials.

## IV. Results and Discussion

In experimentation of the methodology, we have considered 60 font styles and 100 font sizes. The different combinations are tried and are as given in Table 3. We have totally five combinations of font styles and font sizes. The bit 0 indicates varying and 1 indicates constant. The experimentation is done on 9120 (38*240) images of printed Kannada vowel primitives. We have considered 240 images with varying font size and font styles for each primitive. The font size and font styles used are given in Table 4. The entire image set is partitioned into training set and test set and classified using K-fold cross validation method.

*Table 3 :* Combinations of Font Sizes(FS) and Font Styles(FSt)

| Combinations | Font Size(S) | Font Styles(F) | Remarks |
|---|---|---|---|
| 1. | 0 | 0 | Both Font Size and Font Style are Fixed |
| 2. | 0 | 1 | Font Size is Fixed and Font Style is Varied |
| 3. | 1 | 0 | Font Size is Varied and Font Style is Fixed |
| 4. | 1 | 1 | Both Font Size and Font Style are Varied (Nonuniform mix) |
| 5. | 1 | 1 | Both Font Size and Font Style are Varied (Uniform mix) |

*Table 4 :* Font Sizes(FS) and Font Styles(FSt) used

| Sl No | Font Size | Font Style |
|---|---|---|
| 1. | 12,14,16,……,110 (100 sizes) | Baraha 01 |
| 2. | 12,14,16,……,110 (100 sizes) | Baraha 02 |
| ..... | 12,14,16,……,110 (100 sizes) | ….. |
| ..... | 12,14,16,……,110 (100 sizes) | Baraha 30 |
| ..... | 12,14,16,……,110 (100 sizes) | Nudi 01 |
| ..... | 12,14,16,……,110 (100 sizes) | ….. |
| 60. | 12,14,16,……,110 (100 sizes) | Nudi 30 |

*a) Both Font Size and Font Style Fixed*

An experiment is carried out for a total of 39 primitives out of which 14 are character cum primitives and 24 are basic primitives. The image data set has 1520 (38*40) images. For example, we have considered 40 images of font size 60 and font style - Nudi 0.1. The zones based features and nearest neighbor classifier have given 100% recognition accuracy for both the types of primitives. Table 5 gives the results obtained for Character cum Primitives and Table 6 gives the results obtained for Basic primitives.

*Table 5 :* CcP's recognition results for SFS and SFSt

| Same Size and Same Font | |
|---|---|
| Image size=28*28 Zone=4*4 No of Samples=38*40=1520 | |
| Primitive | Classification Accuracy (%) |
| P1 | 100 |
| P2 | 100 |
| P3 | 100 |
| P4 | 100 |
| P5 | 100 |
| P6 | 100 |
| P7 | 100 |
| P8 | 100 |
| P9 | 100 |
| P10 | 100 |
| P11 | 100 |
| P12 | 100 |
| P13 | 100 |
| P14 | 100 |
| Average | 100 |

*Table 6 :* BP's recognition results for Same FS(SFS) and Same FSt(SFSt)

| Same Size and Same Font | | | | |
|---|---|---|---|---|
| Image size=28*28 Zone=4*4 No of Samples=38*40=1520 | | | | |
| Primitive | Classification Accuracy(%) | | Primitive | Classification Accuracy(%) |
| P15 | 100 | | P27 | 100 |
| P16 | 100 | | P28 | 100 |
| P17 | 100 | | P29 | 100 |
| P18 | 100 | | P30 | 100 |
| P19 | 100 | | P31 | 100 |
| P20 | 100 | | P32 | 100 |
| P21 | 100 | | P33 | 100 |
| P22 | 100 | | P34 | 100 |
| P23 | 100 | | P35 | 100 |
| P24 | 100 | | P36 | 100 |
| P25 | 100 | | P37 | 100 |
| P26 | 100 | | P38 | 100 |
| Average | 100 | | | |

*b) Font Size Fixed and Font Style Varied*

An experiment is carried out for 2280 (38*60) images. For each primitive, 60 images of fixed font size - 60 and 60 different font styles, as given in Table 4, are considered. The classifier is subjected to k-fold cross-validation. We have considered 1520 primitive images in each validation step for training and 760 primitive images for testing. Table 7 and Table 8 give recognition accuracy for this combination of CcP and BP primitives using 3- fold cross validation method. The range of recognition accuracies obtained for both, character cum primitives and basic primitives are given in Table 15.

*Table 7 :* 3-Fold cross validation of CcP's

| 3_Fold Cross Validation | | | | |
|---|---|---|---|---|
| Same Size and Different Font using Euclidean Distance | | | | |
| Image size=28*28 Zone size=4*4 No of Samples=38*60=2280 | | | | |
| Primitive | 1_Fold | 2_Fold | 3_Fold | Average |
| P1 | 70 | 80 | 80 | 76.666667 |
| P2 | 65 | 75 | 50 | 63.333333 |
| P3 | 70 | 100 | 90 | 86.666667 |
| P4 | 100 | 95 | 100 | 98.333333 |
| P5 | 95 | 100 | 100 | 98.333333 |
| P6 | 75 | 85 | 90 | 83.333333 |
| P7 | 60 | 85 | 100 | 81.666667 |
| P8 | 70 | 70 | 85 | 75 |
| P9 | 70 | 75 | 85 | 76.666667 |
| P10 | 75 | 75 | 95 | 81.666667 |
| P11 | 95 | 85 | 65 | 81.666667 |
| P12 | 95 | 85 | 90 | 90 |
| P13 | 70 | 75 | 60 | 68.333333 |
| P14 | 80 | 85 | 85 | 83.333333 |
| Average | 77.857143 | 83.571429 | 83.928571 | 81.785714 |

*Table 8 :* 3-Fold cross validation of BP's

| 3_Fold Cross Validation | | | | |
|---|---|---|---|---|
| Same Size and Different Font using Euclidean Distance | | | | |
| Image size=28*28 Zone size=4*4 No of Samples=38*60=2280 | | | | |
| Primitive | 1_Fold | 2_Fold | 3_Fold | Average |
| P15 | 75 | 80 | 85 | 80 |
| P16 | 85 | 95 | 80 | 86.666667 |
| P17 | 100 | 100 | 100 | 100 |
| P18 | 85 | 80 | 80 | 81.666667 |
| P19 | 80 | 60 | 60 | 66.666667 |

| | | | | |
|---|---|---|---|---|
| P20 | 75 | 85 | 95 | 85 |
| P21 | 25 | 80 | 90 | 65 |
| P22 | 20 | 25 | 25 | 23.333333 |
| P23 | 30 | 25 | 20 | 25 |
| P24 | 75 | 90 | 85 | 83.333333 |
| P25 | 70 | 95 | 100 | 88.333333 |
| P26 | 55 | 55 | 60 | 56.666667 |
| P27 | 70 | 95 | 65 | 76.666667 |
| P28 | 70 | 55 | 60 | 61.666667 |
| P29 | 75 | 80 | 75 | 76.666667 |
| P30 | 80 | 75 | 90 | 81.666667 |
| P31 | 95 | 65 | 80 | 80 |
| P32 | 85 | 95 | 95 | 91.666667 |
| P33 | 65 | 80 | 85 | 76.666667 |
| P34 | 85 | 70 | 80 | 78.333333 |
| P35 | 85 | 90 | 95 | 90 |
| P36 | 95 | 85 | 85 | 88.333333 |
| P37 | 85 | 95 | 90 | 90 |
| P38 | 90 | 95 | 95 | 93.333333 |
| Average | 73.125 | 77.083333 | 78.125 | 76.111111 |

*c) Font Size Varied and Font Style Fixed*

An experiment is carried out on 1520(38*40) images. We have considered 40 images of different font sizes, as given in Table 4, font style of Nudi 01 for each primitive. Table 9 and Table 10 give results of classification for varying font sizes and fixed font style using 4- fold cross validation using Euclidean distance. The range of recognition accuracies obtained for this combination in case of both character cum primitives and basic primitives are given in Table 15.

*Table 9 :* 4-Fold cross validation of CcP's

| 2_Fold Cross Validation | | | |
|---|---|---|---|
| Different Size and Same Font using Euclidean Distance | | | |
| Image size=28*28 Zone size=4*4 No of Samples=38*40=1520 | | | |
| Primitive | 1_Fold | 2_Fold | Average |
| P1 | 100 | 100 | 100 |
| P2 | 95 | 100 | 97.5 |
| P3 | 100 | 100 | 100 |
| P4 | 100 | 100 | 100 |
| P5 | 100 | 100 | 100 |
| P6 | 100 | 100 | 100 |
| P7 | 100 | 100 | 100 |
| P8 | 100 | 100 | 100 |
| P9 | 100 | 100 | 100 |
| P10 | 100 | 100 | 100 |
| P11 | 100 | 100 | 100 |
| P12 | 100 | 100 | 100 |
| P13 | 100 | 100 | 100 |
| P14 | 100 | 100 | 100 |
| Average | 99.64286 | 100 | 99.82143 |

*Table 10 :* 4-Fold cross validation of BP'

| 2_Fold Cross Validation | | | |
|---|---|---|---|
| Different Size and Same Font using Euclidean Distance | | | |
| Image size=28*28 Zone size=4*4 No of Samples=38*40=1520 | | | |
| Primitive | 1_Fold | 2_Fold | Average |
| P15 | 100 | 100 | 100 |
| P16 | 100 | 100 | 100 |
| P17 | 100 | 100 | 100 |
| P18 | 100 | 100 | 100 |
| P19 | 100 | 100 | 100 |

| | | | |
|---|---|---|---|
| P20 | 100 | 100 | 100 |
| P21 | 85 | 85 | 85 |
| P22 | 65 | 70 | 67.5 |
| P23 | 70 | 80 | 75 |
| P24 | 100 | 100 | 100 |
| P25 | 100 | 95 | 97.5 |
| P26 | 100 | 100 | 100 |
| P27 | 95 | 100 | 97.5 |
| P28 | 100 | 95 | 97.5 |
| P29 | 100 | 100 | 100 |
| P30 | 100 | 100 | 100 |
| P31 | 100 | 100 | 100 |
| P32 | 100 | 100 | 100 |
| P33 | 100 | 100 | 100 |
| P34 | 100 | 100 | 100 |
| P35 | 100 | 100 | 100 |
| P36 | 100 | 100 | 100 |
| P37 | 100 | 100 | 100 |
| P38 | 100 | 100 | 100 |
| Average | 96.458333 | 96.875 | 96.666667 |

*Table 11 :* 5-Fold cross validation of CcP's (Non -uniform mix)

| 5_Fold Cross Validation | | | | | | |
|---|---|---|---|---|---|---|
| Different Size and Different Font using Euclidean Distance | | | | | | |
| Image size=28*28 Zone size=4*4 No of Samples=38*100=3800 | | | | | | |
| Primitive | 1_Fold | 2_Fold | 3_Fold | 4_Fold | 5_Fold | Average |
| P1 | 95 | 95 | 90 | 100 | 100 | 95 |
| P2 | 95 | 85 | 95 | 100 | 100 | 95 |
| P3 | 100 | 100 | 100 | 100 | 90 | 98 |
| P4 | 100 | 100 | 100 | 100 | 100 | 100 |
| P5 | 95 | 100 | 100 | 100 | 100 | 99 |
| P6 | 100 | 100 | 95 | 100 | 100 | 99 |
| P7 | 100 | 95 | 100 | 100 | 100 | 99 |
| P8 | 95 | 95 | 100 | 100 | 90 | 96 |
| P9 | 95 | 100 | 90 | 100 | 100 | 97 |
| P10 | 95 | 95 | 100 | 100 | 100 | 98 |
| P11 | 90 | 100 | 100 | 100 | 100 | 98 |
| P12 | 95 | 90 | 100 | 100 | 100 | 97 |
| P13 | 85 | 95 | 95 | 90 | 100 | 93 |
| P14 | 100 | 100 | 100 | 100 | 100 | 100 |
| Average | 95.714286 | 96.071429 | 97.5 | 99.285714 | 98.571429 | 97.428571 |

*d) Both Font Size and Font Style Varied (Non-uniform mix)*

An experiment is carried out on 3800 images. In non-uniform mix, we have considered 60 images of different font sizes, as given in the Table 4, of Nudi 01 font style and 40 images of 40 different font sizes for each primitive. Table 11 and Table 12 show the results for both varying font sizes and font styles using 5- fold cross validation using Euclidean distance. The recognition accuracy obtained for basic primitives and for character cum primitives is given in Table 15.

*Table 12 :* 5-Fold cross validation of BP's (Non -uniform mix)

| 5_Fold Cross Validation | | | | | | |
|---|---|---|---|---|---|---|
| Different Size and Different Font using Euclidean Distance | | | | | | |
| Image size=28*28 Zone size=4*4 No of Samples=38*100=3800 | | | | | | |
| Primitive | 1_Fold | 2_Fold | 3_Fold | 4_Fold | 5_Fold | Average |
| P15 | 100 | 100 | 100 | 100 | 90 | 98 |
| P16 | 90 | 95 | 100 | 100 | 100 | 97 |
| P17 | 100 | 100 | 100 | 100 | 100 | 100 |
| P18 | 95 | 90 | 100 | 100 | 100 | 97 |
| P19 | 90 | 90 | 95 | 100 | 80 | 91 |
| P20 | 90 | 90 | 100 | 100 | 90 | 94 |

| | | | | | | |
|---|---|---|---|---|---|---|
| P21 | 50 | 60 | 50 | 35 | 50 | 49 |
| P22 | 35 | 20 | 35 | 60 | 40 | 38 |
| P23 | 55 | 75 | 65 | 70 | 85 | 70 |
| P24 | 100 | 95 | 100 | 100 | 100 | 99 |
| P25 | 95 | 100 | 100 | 90 | 90 | 93 |
| P26 | 90 | 90 | 90 | 100 | 100 | 94 |
| P27 | 90 | 90 | 90 | 100 | 100 | 96 |
| P28 | 85 | 95 | 100 | 90 | 100 | 93 |
| P29 | 85 | 95 | 100 | 90 | 90 | 91 |
| P30 | 100 | 100 | 95 | 100 | 100 | 98 |
| P31 | 95 | 95 | 90 | 100 | 100 | 96 |
| P32 | 100 | 100 | 100 | 100 | 90 | 98 |
| P33 | 100 | 95 | 100 | 100 | 100 | 99 |
| P34 | 100 | 100 | 100 | 90 | 100 | 98 |
| P35 | 100 | 95 | 100 | 100 | 100 | 99 |
| P36 | 100 | 100 | 100 | 100 | 90 | 98 |
| P37 | 100 | 95 | 100 | 90 | 100 | 97 |
| P38 | 95 | 100 | 100 | 100 | 100 | 99 |
| Average | 89.166667 | 89.583333 | 92.083333 | 92.291667 | 91.458333 | 90.916667 |

*e) Both Font Size and Font Style Varied (Uniform mix)*

An experiment is carried out on 3900 (39*100) images. We have considered100 images of 10 font styles and 10 varying sizes, for each font style, as given in Table 4, for each primitive. Table 13 and Table 14 show the results for both varying font sizes and font styles using 5- fold cross validation using Euclidean distance. The recognition accuracy obtained for basic primitives and for character cum primitives is given in Table 15.

*Table 13 :* 5-Fold cross validation of CcP's (Uniform mix)

| 5_Fold Cross Validation | | | | | | |
|---|---|---|---|---|---|---|
| Different Size and Different Font using Euclidean Distance | | | | | | |
| Image size=28*28 Zone size=4*4 No of Samples=38*100=3800 | | | | | | |
| Primitive | 1_Fold | 2_Fold | 3_Fold | 4_Fold | 5_Fold | Average |
| 1 | 100 | 100 | 100 | 90 | 100 | 98 |
| 2 | 95 | 100 | 100 | 90 | 100 | 97 |
| 3 | 100 | 95 | 100 | 90 | 90 | 95 |
| 4 | 100 | 95 | 100 | 100 | 100 | 99 |
| 5 | 100 | 100 | 100 | 100 | 100 | 100 |
| 6 | 100 | 100 | 95 | 100 | 90 | 97 |
| 7 | 100 | 100 | 100 | 100 | 95 | 99 |
| 8 | 100 | 90 | 100 | 90 | 85 | 93 |
| 9 | 90 | 100 | 100 | 90 | 100 | 96 |
| 10 | 100 | 90 | 95 | 100 | 95 | 96 |
| 11 | 100 | 100 | 90 | 100 | 85 | 95 |
| 12 | 100 | 100 | 90 | 85 | 90 | 93 |
| 13 | 100 | 100 | 90 | 95 | 95 | 96 |
| 14 | 95 | 100 | 100 | 95 | 95 | 97 |
| Average | 98.571429 | 97.857143 | 97.142857 | 94.642857 | 94.285714 | 96.5 |

*Table 14 :* 5-Fold cross validation of BP's(Uniform mix)

| 5_Fold Cross Validation | | | | | | |
|---|---|---|---|---|---|---|
| Different Size and Different Font using Euclidean Distance | | | | | | |
| Image size=28*28 Zone size=4*4 No of Samples=38*100=3800 | | | | | | |
| Primitive | 1_Fold | 2_Fold | 3_Fold | 4_Fold | 5_Fold | Average |
| P15 | 95 | 85 | 70 | 75 | 60 | 77 |
| P16 | 90 | 85 | 95 | 100 | 90 | 92 |
| P17 | 100 | 90 | 95 | 100 | 95 | 96 |
| P18 | 85 | 95 | 95 | 100 | 95 | 94 |
| P19 | 80 | 95 | 100 | 100 | 100 | 95 |
| P20 | 100 | 90 | 90 | 100 | 85 | 93 |

| | | | | | | |
|---|---|---|---|---|---|---|
| P21 | 100 | 100 | 100 | 100 | 100 | 100 |
| P22 | 40 | 30 | 50 | 50 | 50 | 44 |
| P23 | 100 | 90 | 100 | 100 | 100 | 98 |
| P24 | 95 | 95 | 100 | 95 | 65 | 90 |
| P25 | 100 | 95 | 95 | 80 | 90 | 92 |
| P26 | 70 | 95 | 55 | 55 | 100 | 75 |
| P27 | 100 | 85 | 85 | 75 | 55 | 80 |
| P28 | 95 | 75 | 85 | 45 | 95 | 79 |
| P29 | 80 | 80 | 85 | 55 | 90 | 78 |
| P30 | 100 | 100 | 100 | 95 | 90 | 97 |
| P31 | 100 | 90 | 95 | 90 | 95 | 94 |
| P32 | 100 | 85 | 95 | 90 | 70 | 88 |
| P33 | 95 | 95 | 90 | 100 | 80 | 92 |
| P34 | 100 | 100 | 95 | 80 | 70 | 89 |
| P35 | 100 | 95 | 95 | 90 | 85 | 93 |
| P36 | 100 | 100 | 90 | 85 | 50 | 85 |
| P37 | 100 | 95 | 100 | 70 | 75 | 88 |
| P38 | 95 | 100 | 100 | 95 | 85 | 95 |
| | 92.5 | 89.375 | 90 | 84.375 | 82.083333 | 87.666667 |

*Table 15 :* Combinations of font sizes and font styles and their recognition accuracy for CcP's and BP's

| | Printed Basic Primitives | | Printed Character cum Primitives | |
|---|---|---|---|---|
| Combinations | Euclidean Distance | | Euclidean Distance | |
| | Range | Avg | Range | Avg |
| Both Font Size and Font Style are Fixed | 100% | 100% | 100% | 100% |
| Font Size is Fixed and Font Style is Varied | 25%--93% | 76% | 63%--98% | 82% |
| Font Size is Varied and Font Style is Fixed | 68%-100% | 97% | 98%-100% | 99% |
| Both Font Size and Font Style are Varied(Non-uniform | 38%-100% | 90% | 93%-100% | 97% |
| Both Font Size and Font Style are Varied(Uniform mix) | 44%-100% | 88% | 93%-100% | 87% |
| Average | | 90% | | 93% |
| | Handwritten Basic Primitives | | Handwritten Character cum Primitives | |
| | Euclidean Distance | | Euclidean Distance | |
| Average | 93% | | 94% | |

## V. Conclusion

We have proposed a zone features based methodology for recognition of both printed and handwritten 38 primitives of 49 Kannada vowels and consonants together. Zones of 4*4 and images of 28*28 are used. The nearest neighbor classifier is used with Euclidean distance measure. The accuracy of the classifier is verified with k-fold cross validation, for k = 2,3 and 5. We have experimented with four combinations of font sizes and font styles for printed primitives and obtained average recognition accuracy in the range [89%, 94%]. Further, we have obtained accuracy in the range [90%, 94% ]for handwritten primitives. This work is basic to automation of writing of Kannada kagunitha and wothakshara"s.

## References Références Referencias


1. Leena R Ragha and M Sasikumar , (2011) "Feature Analysis for Handwritten Kannada Kagunita Recognition,International Journal of Computer Theory and Engineering, Vol.3, No.11793-8201
2. S. K. Sangame, R.J. Ramteke and Yogesh V. G. (2011) "Recognition of isolated handwritten Kannada characters using invariant moments and chain code" world journal of science and technology, 1(8): 115-120 ISSN: 2231 – 2587
3. K S Prasanna Kumar, (2012) "Optical Character Recognition (OCR) for Kannada numerals using Left Bottom 1/4th segment minimum features extraction" International Journal of Computer Techology & Applications, Vol 3 (1) , 221-225 ISSN:2229-6093
4. Kauleshwar Prasad, Devvrat C. Nigam, Ashmika Lakhotiya and Dheeren Umre,(2013) "Character Recognition Using Matlab"s Neural Network Toolbox", International Journal of u- and e- Service, Science and Technology Vol. 6, No. 1
5. Mamatha Hosalli R amappa and Srikantamurthy Krishnamurthy, (2013) "A Comparative Study of Different Feature Extraction and Classification Methods for Recognition of Handwritten Kannada Numerals" International Journal of Database Theory and Application Vol. 6, No. 4
6. Manjunath A E Sharath B(2013) "Implementing Kan-nada Optical Character Recognition on the Android

Operating System for Kannada Sign Boards"International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 1

7. Umapada Pal, Ramachandran Jayadevan and Nabin Sharma (2012), "Handwriting Recognition in Indian Regional Scripts: A Survey of Offline Techniques", ACM Transactions on Asian Language Information Processing, Vol. 11, No. 1, Article 1, Publication date: March 2012.
8. Leena R Ragha and M Sasikumar , (2010) "Using Moments Features from Gabor Directional Images for Kannada Handwriting Character Recognition", International Conference and Workshop on Emerging Trends in Technology (ICWET 2010) – TCET, Mumbai, India
9. Om Prakash Sharma, M. K. Ghose, Krishna Bikram Shah, Benoy Kumar Thakur (2013) "Recent Trends and Tools for Feature Extraction in OCR Technology", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-6, January 2013
10. Karthik Sheshadri, Pavan Kumar T Ambekar, Deeksha Padma Prasad and Dr. Ramakanth P Kumar(2010) "An OCR system for Printed Kannada using k-means clustering" 978-1-4244-5697-0/10/$25.00 ©2010 IEEE.
11. B.V.Dhandra, Gururaj Mukarambi, Mallikarjun Hangarge (2011), "Zone Based Features for Handwritten and Printed Mixed Kannada Digits Recognition", International Conference on VLSI, Communication & Instrumentation (ICVCI) 2011 Proceedings published by International Journal of Computer Applications®.
12. H. Imran Khan, Smitha U. V, Suresh Kumar D. S (2013), "Isolated Kannada Character Recognition using Chain Code Features", International Journal of Science and Research (IJSR), India Online ISSN: 2319-7064, Volume2 Issue 8, August 2013
13. Mamatha.H.R , Sucharitha Srirangaprasad, Srikantamurthy K (2013) "Data fusion based framework for the recognition of Isolated Handwritten Kannada Numerals", International Journal of Advanced Computer Science and Applications, Vol. 4, No. 6, 2013
14. Swapnil A. Vaidya, Balaji R. Bombade(2013) "A Novel Approach of Handwritten Character Recognition using Positional Feature Extraction", International Journal of Computer Science and Mobile Computing Vol.2 Issue.6, June- 2013, pg. 179-186
15. Nithya E, Dr. Ramesh Babu D R (2013) "OCR System for Complex Printed Kannada Characters", International Journal of Advanced Research in Computer Science and Software Engineering 3(6), June - 2013, pp. 102-105
16. G. G. Rajput, Rajeswari Horakeri(2011) "Shape Descriptors based Handwritten Character Recognition Engine with Application to Kannada Characters", International Conference on Computer & Communication Technology (ICCCT)-2011.
17. G. G. Rajput, Rajeswari Horakeri(2013) "Zone based Handwritten Kannada Character Recognition Using Crack code and SVM", 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)-2013.
18. Sandhya.N, Krishnan.R, D.R.Ramesh Babu (2014) "Feature Based Kannada Character Classification Method of Kannada Character Recognition" International Journal of Scientific & Engineering Research, Volume 5, Issue 2, February-2014 729 ISSN 2229-5518.

Global Journal of Computer Science and Technology: F
Graphics & Vision
Volume 14 Issue 4 Version 1.0 Year 2014
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

# Secure Message Recovery and Batch Verification using Digital Signature

By Makkar,Sarika & Thakral,Silvi
*Manav Institute of Technology and Management, India*

*Abstract-* This paper about the study of Secure message Recovery and batch verification using Digital Signature. Security is increased in batch verification through triple DES algorithm. Encryption is used for the Security in which the plaintext is transforming into the cipher text. A digital signature scheme involves two phases, the signature generation phase which is performed at the sender side and the signature verification phase that is performed by the receiver of that message. In computer to computer communication, the computer at sender's end usually transforms a plaintext into cipher text using encryption. When the message is recovered at the Receiver Side than the original text is converted in to the encrypted text. That encrypted text is secure for the authenticated person. After recover the message if authentic person wants to get the original text then he/she enter the key and take the plaintext.

*Keywords:* *digital signature, forgeries, encryption, triple des algorithm.*

*GJCST-F Classification :* *D.4.6, D.4.4, H.3.7*


*Strictly as per the compliance and regulations of:*

# Secure Message Recovery and Batch Verification using Digital Signature

Makkar,Sarika [α] & Thakral,Silvi [σ]

*Abstract -* This paper about the study of Secure message Recovery and batch verification using Digital Signature. Security is increased in batch verification through triple DES algorithm. Encryption is used for the Security in which the plaintext is transforming into the cipher text. A digital signature scheme involves two phases, the signature generation phase which is performed at the sender side and the signature verification phase that is performed by the receiver of that message. In computer to computer communication, the computer at sender's end usually transforms a plaintext into cipher text using encryption. When the message is recovered at the Receiver Side than the original text is converted in to the encrypted text. That encrypted text is secure for the authenticated person. After recover the message if authentic person wants to get the original text then he/she enter the key and take the plaintext.

*Keywords: digital signature, forgeries, encryption, triple des algorithm.*

## I. Introduction

Digital signature is an authentication process that is used to prove the identity of source and integrity of message. A digital signature scheme involves two phases, the signature generation phase which is performed at the sender side and the signature verification phase that is performed by the receiver of that message. in this pair of key is used private key and public key. Private key is Secret and public key known all the users.

Digital signature provides the following security services:

*a) Message integrity*

It guards against the In appropriate information modification or damage. Message integrity ensures the information nonrepudiation and authenticity By using this, users are able to ensure that the message has not been altered during transmission. A loss of message integrity means that there is insertion, deletion or modification in message or replay of the message.

*b) Authentication*

This property defines being real and being able to be trusted and verifiable. The functionality of the authentication service is to guarantee the recipient that message is from the source that it state to be. two aspects are involved: first at the connection initiation time, the entities are authentic that is each entity is the entity which it state to be. Second the process of authentication must assure that the connection is not interfere by the third party in such a way that a third party can impersonate one of the two legal parties for unauthorized transmission or reception of messages.

*c) Nonrepudiation*

It prevents from denying transmission of a message by either sender or receiver. Thus if the message is sent then the receiver can validate that the claimed sender has sent the message. This is called origin nonrepudiation. Similarly, when a message is received the source can validate that the claimed receiver has in fact receive the message.

Thus digital signature must have to posses the following properties:

1. The digital signature must validate the sender and date and time of the digital signature.
2. Digital signature must authenticate the content of message at the time of digital signature.
3. In case of any dispute, digital signature must be verifiable by third party to resolve it.

## II. Digital Signature Requirements

The points described below states the requirement of the digital signature:

1. Digital signature (a bit pattern) must depend upon the message that is to be signed by the sender.
2. It must make use of some information related to sender that is unique to it to prevent against denial and forgeries.
3. Digital signature must be comparatively easy to compute on message.

It must be comparatively easy to recognize and validate digital signature

## III. Encryption And Decryption

Encryption is used for the Security in which the plaintext is transforming into the cipher text. In computer to computer communication, the computer at sender's end usually transforms a plaintext into cipher text using encryption the encrypted cipher text message is sent to the receiver over a network then the receiver takes encrypted message and performs the reverse of encryption. I.e. performs the decryption process obtain the plaintext.

*Author α σ : Student-M.tech., Assitant Professor, Manav Institute of Technology & Management, jevra(Hissar).*
*e-mails: sarikamakkar0@gmail.com , silvithakral1@gmail.com.*

*a) Plaintext and cipher text*

Any communication in the language that we speak that is the human language, takes the form of plain text or clear text. That is, a message in plaintext can be understood by anybody knowing the language as long as the message is not codified in any manner. Plain text signifies a message that can be understood by the sender, the recipient and also by anyone else who gets an access to the message. when a plaintext message is codified using is codified using any suitable

## IV. Digital Signature Modes

There Are Two Modes of Operation, Appendix Mode And Recovery Mode.

*a) Appendix mode*

In appendix mode the creator of the message attach a code with the message that act as a signature. Typically the signature is produced by taking the hash of the message and encrypts it with the private key of sender. This signature guarantees the integrity of message and claimed identity of source.

In the figure 3 first a hash code generation algorithm has been applied on the message and then it is encrypted with the private key of the sender. The generated code then appends to the message and transmitted to receiver via network. Receiver verifies the signature using three items, the public key of sender, the packet and the signature. The receiver first cut off the message from digital signature. It first computes the hash of the message and decrypts the received signature with the public key of sender. If both values are equal then the message will be considered as authentic otherwise it has been modified during transmission.

*b) Recovery mode*

In message recovery mode the signed message is implanted in the digital signature and it can be recovered from it. The well-known digital signature scheme with message recovery is the RSA digital signature scheme which security is based upon solving the factor of large prime numbers. Later Nyberg and Rueppel also proposed the digital signature scheme with message recovery based upon the discrete logarithm. Some of these schemes have the capability of privacy of signed message and thus only the legal receiver can recover the message and verify its authenticity. However the scheme only allows a signer to sign each message independently.

As shown in the figure 4, the receiver requires only two parameters to verify the digital signature of the message, the public key of sender and the digital signature. The receiver first recovers the message from the received signature and then performs computation for digital signature verification.

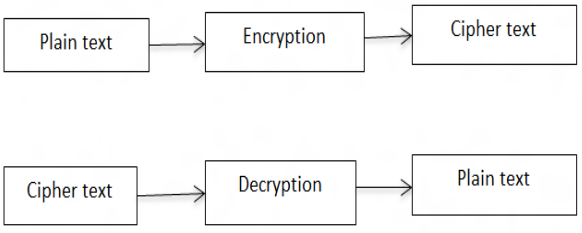



*Figure1:* Encryption and Decryption Process

## V. Objective

1. An unauthorized person cannot get the original text.
2. If any person tries to get the plaintext than he/she get the encrypted form text.
3. If the authenticated person knows the key than get the plaintext.
4. It must detect integrity violence. An attacker must not be able to replace false packets for legitimate ones i.e. multiple packets should not be modified
5. It must detect integrity violence. An attacker must not be able to replace false packets for legitimate ones i.e. multiple packets should not be modified.

## VI. Proposed Methodology

*a) Triple DES Algorithm*

Triple DES Algorithm is same as the DES with two 56 bit key is applied. Given a plaintext message first key is used to DES encrypt the message. The second key is used to decrypt the encrypted message. The twice scrambled message is encrypted again with the first key to yield the final cipher text.it uses three 56 bits DES keys giving a total key length of 168 bits. The block size is is 64 bits and the key sizes are 168, 112, or 56 bits with respect to keying option 1, 2, or 3. The input key sizes are 3 64 bit keys, which are shortened to 56 bits because of the internal key scheduler.

The block of data is encrypted 3 times with each of the keys according to the keying options:

*Keying Option 1 :* All of the keys are independent
*Keying Option 2 :* K1 and K2 are independent and K3 = K1

*Keying Option3 :* All keys are identical K1=K2=K3

Triple DES Algorithm has following steps:-

*Step1 :* Encrypt the data using DES with the first 56 bit key.
*Step2 :* Decrypt the data using Second 56 bit key.
*Step3 :* Encrypt the data using DES third key 56 bit

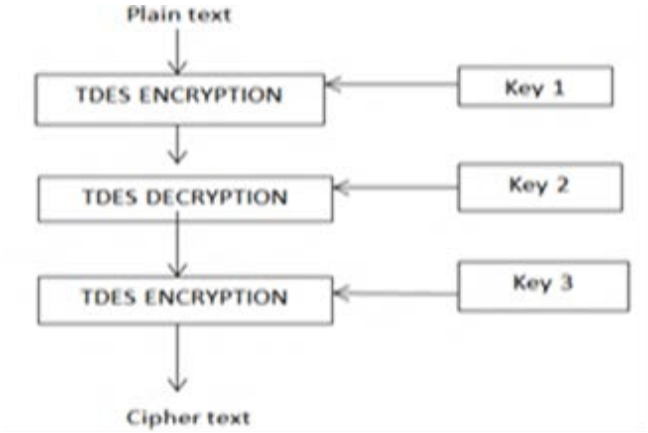



*Figure 2 :* Tdes Encryption Decryption Process

i. *Advantages*

1. TDES Algorithm not Easy to break.
2. It is more Secure rather than DES.

ii. *Disadvantage*

1. This algorithm take 3 times more than DES.

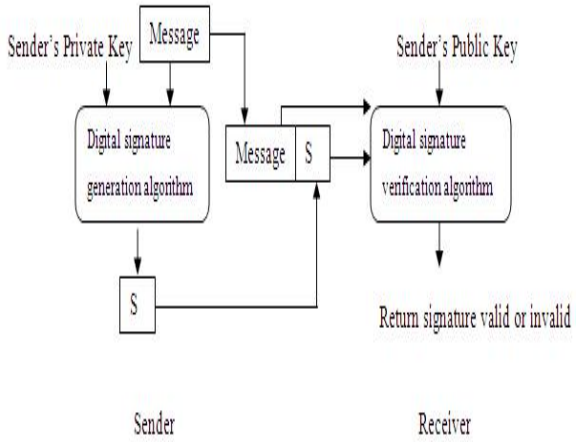



*Figure 3 :* Appendix Mode Digital Signature

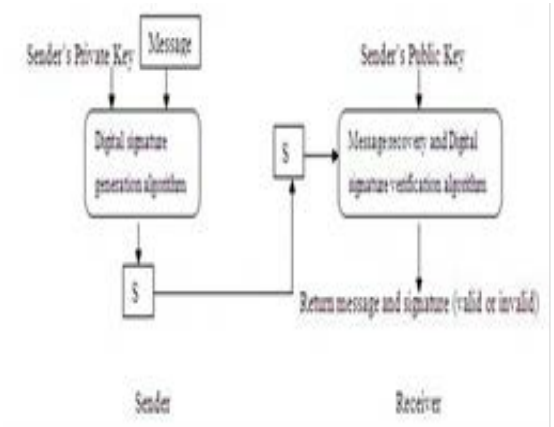



*Figure 4 :* Recovery Mode Digital Signature

## VII. Conclusion

In This Present Work We Increase The Security Of Batch Verification. After Rec-Overy The Message Is Not In The Original Form. No One Can Get The Plaintext Whether He/She Not Enter The Key. If Any Unauthentic Person Tries To See The Plaintext He/She Only Gets The Encrypted Text.

## References Références Referencias


1. C. Boyd and C. Pavlovski(2000), "Attacking and Repairing Batch Verification Schemes", Proc. Sixth Int'l Conf. Theory and application of Cryptology and Information Security Advances in Cryptology (ASIANCRYPT '00).
2. Chin-Chen Chang and Ya-Fen Chang (2004) "Signing a digital signature without using one-way hash functions and message redundancy schemes" IEEE Communication Letters.
3. F. G. Zhang (2005) "Cryptanalysis of Chang et al.'s Signature Scheme with Message Recovery", IEEE Communication Letters.
4. Hung, yu chien (2006), "Forgery Attacks on Digital Signature Schemes without sing One-way Hash and Message Redundancy", IEEE communications letters.
5. J.M. Park, E.K.P. Chong, and H.J. Siegel (2002) "Efficient Multicast Packet Authentication Using Signature Amortization" Proc. IEEE Symp. Security and Privacy.
6. Jie Liu, Jianhua Li (2006), "Improvement on a Digital Signature Scheme without using One-way Hash and Message Redundancy", Department of Electronic Engineering, Shanghai Jiao Tong University.
7. L. Kang and X. H. Tang (2006)"*Digital signature scheme without hash functions and message redundancy",* Journal on Communications.
8. M. S. Hwang, C. C. Chang, and K. F. Hwang(2002) *"An ElGamal-like cryptosystem for enciphering large messages"* IEEE Transactions on Knowledge and Data Engineering.
9. S. P. Shieh, C. T. Lin, W. B. Yang, and H. M. Sun (2000) "Digital multisignature schemes for authenticating delegates in mobile code systems".
10. S. W. Changchien and M. S. Hwang (2002) "*A batch verifying and detecting multiple RSA digital signatures"* International Journal of Computational and Numerical Analysis and Applications.
11. S. J. Hwang and E.-T. Li (2003) *"Cryptanalysis of Shieh-Lin-Yang-Sun signature scheme"* IEEE Commun. Lett.

# Global Journals Inc. (US) Guidelines Handbook 2014

www.GlobalJournals.org

## FELLOW OF ASSOCIATION OF RESEARCH SOCIETY IN COMPUTING (FARSC)

Global Journals Incorporate (USA) is accredited by Open Association of Research Society (OARS), U.S.A and in turn, awards "FARSC" title to individuals. The 'FARSC' title is accorded to a selected professional after the approval of the Editor-in-Chief/Editorial Board Members/Dean.

The "FARSC" is a dignified title which is accorded to a person's name viz. Dr. John E. Hall, Ph.D., FARSC or William Walldroff, M.S., FARSC.

FARSC accrediting is an honor. It authenticates your research activities. After recognition as FARSC, you can add 'FARSC' title with your name as you use this recognition as additional suffix to your status. This will definitely enhance and add more value and repute to your name. You may use it on your professional Counseling Materials such as CV, Resume, and Visiting Card etc.

*The following benefits can be availed by you only for next three years from the date of certification:*

FARSC designated members are entitled to avail a 40% discount while publishing their research papers (of a single author) with Global Journals Incorporation (USA), if the same is accepted by Editorial Board/Peer Reviewers. If you are a main author or co-author in case of multiple authors, you will be entitled to avail discount of 10%.

Once FARSC title is accorded, the Fellow is authorized to organize a symposium/seminar/conference on behalf of Global Journal Incorporation (USA).The Fellow can also participate in conference/seminar/symposium organized by another institution as representative of Global Journal. In both the cases, it is mandatory for him to discuss with us and obtain our consent.

You may join as member of the Editorial Board of Global Journals Incorporation (USA) after successful completion of three years as Fellow and as Peer Reviewer. In addition, it is also desirable that you should organize seminar/symposium/conference at least once.

We shall provide you intimation regarding launching of e-version of journal of your stream time to time. This may be utilized in your library for the enrichment of knowledge of your students as well as it can also be helpful for the concerned faculty members.

The FARSC can go through standards of OARS. You can also play vital role if you have any suggestions so that proper amendment can take place to improve the same for the benefit of entire research community.

As FARSC, you will be given a renowned, secure and free professional email address with 100 GB of space e.g. johnhall@globaljournals.org. This will include Webmail, Spam Assassin, Email Forwarders, Auto-Responders, Email Delivery Route tracing, etc.


The FARSC will be eligible for a free application of standardization of their researches. Standardization of research will be subject to acceptability within stipulated norms as the next step after publishing in a journal. We shall depute a team of specialized research professionals who will render their services for elevating your researches to next higher level, which is worldwide open standardization.


The FARSC member can apply for grading and certification of standards of their educational and Institutional Degrees to Open Association of Research, Society U.S.A. Once you are designated as FARSC, you may send us a scanned copy of all of your credentials. OARS will verify, grade and certify them. This will be based on your academic records, quality of research papers published by you, and some more criteria. After certification of all your credentials by OARS, they will be published on your Fellow Profile link on website https://associationofresearch.org which will be helpful to upgrade the dignity.

The FARSC members can avail the benefits of free research podcasting in Global Research Radio with their research documents. After publishing the work, (including published elsewhere worldwide with proper authorization) you can upload your research paper with your recorded voice or you can utilize chargeable services of our professional RJs to record your paper in their voice on request.


The FARSC member also entitled to get the benefits of free research podcasting of their research documents through video clips. We can also streamline your conference videos and display your slides/ online slides and online research video clips at reasonable charges, on request.

The FARSC is eligible to earn from sales proceeds of his/her researches/reference/review Books or literature, while publishing with Global Journals. The FARSC can decide whether he/she would like to publish his/her research in a closed manner. In this case, whenever readers purchase that individual research paper for reading, maximum 60% of its profit earned as royalty by Global Journals, will be credited to his/her bank account. The entire entitled amount will be credited to his/her bank account exceeding limit of minimum fixed balance. There is no minimum time limit for collection. The FARSC member can decide its price and we can help in making the right decision.

The FARSC member is eligible to join as a paid peer reviewer at Global Journals Incorporation (USA) and can get remuneration of 15% of author fees, taken from the author of a respective paper. After reviewing 5 or more papers you can request to transfer the amount to your bank account.

## MEMBER OF ASSOCIATION OF RESEARCH SOCIETY IN COMPUTING (MARSC)

The ' MARSC ' title is accorded to a selected professional after the approval of the Editor-in-Chief / Editorial Board Members/Dean.
The “MARSC” is a dignified ornament which is accorded to a person’s name viz. Dr. John E. Hall, Ph.D., MARSC or William Walldroff, M.S., MARSC.

MARSC accrediting is an honor. It authenticates your research activities. After becoming MARSC, you can add 'MARSC' title with your name as you use this recognition as additional suffix to your status. This will definitely enhance and add more value and repute to your name. You may use it on your professional Counseling Materials such as CV, Resume, Visiting Card and Name Plate etc.

*The following benefitscan be availed by you only for next three years from the date of certification.*


MARSC designated members are entitled to avail a 25% discount while publishing their research papers (of a single author) in Global Journals Inc., if the same is accepted by our Editorial Board and Peer Reviewers. If you are a main author or co-author of a group of authors, you will get discount of 10%.

As MARSC, you will be given a renowned, secure and free professional email address with 30 GB of space e.g. johnhall@globaljournals.org. This will include Webmail, Spam Assassin, Email Forwarders, Auto-Responders, Email Delivery Route tracing, etc.


We shall provide you intimation regarding launching of e-version of journal of your stream time to time. This may be utilized in your library for the enrichment of knowledge of your students as well as it can also be helpful for the concerned faculty members.

The MARSC member can apply for approval, grading and certification of standards of their educational and Institutional Degrees to Open Association of Research, Society U.S.A.


Once you are designated as MARSC, you may send us a scanned copy of all of your credentials. OARS will verify, grade and certify them. This will be based on your academic records, quality of research papers published by you, and some more criteria.

It is mandatory to read all terms and conditions carefully.

# AUXILIARY MEMBERSHIPS

## Institutional Fellow of Open Association of Research Society (USA)-OARS (USA)

Global Journals Incorporation (USA) is accredited by Open Association of Research Society, U.S.A (OARS) and in turn, affiliates research institutions as "Institutional Fellow of Open Association of Research Society" (IFOARS).

The "FARSC" is a dignified title which is accorded to a person's name viz. Dr. John E. Hall, Ph.D., FARSC or William Walldroff, M.S., FARSC.

The IFOARS institution is entitled to form a Board comprised of one Chairperson and three to five board members preferably from different streams. The Board will be recognized as "Institutional Board of Open Association of Research Society"-(IBOARS).

*The Institute will be entitled to following benefits:*

The IBOARS can initially review research papers of their institute and recommend them to publish with respective journal of Global Journals. It can also review the papers of other institutions after obtaining our consent. The second review will be done by peer reviewer of Global Journals Incorporation (USA) The Board is at liberty to appoint a peer reviewer with the approval of chairperson after consulting us.

The author fees of such paper may be waived off up to 40%.

The Global Journals Incorporation (USA) at its discretion can also refer double blind peer reviewed paper at their end to the board for the verification and to get recommendation for final stage of acceptance of publication.

The IBOARS can organize symposium/seminar/conference in their country on behalf of Global Journals Incorporation (USA)-OARS (USA). The terms and conditions can be discussed separately.

The Board can also play vital role by exploring and giving valuable suggestions regarding the Standards of "Open Association of Research Society, U.S.A (OARS)" so that proper amendment can take place for the benefit of entire research community. We shall provide details of particular standard only on receipt of request from the Board.


The board members can also join us as Individual Fellow with 40% discount on total fees applicable to Individual Fellow. They will be entitled to avail all the benefits as declared. Please visit Individual Fellow-sub menu of GlobalJournals.org to have more relevant details.

We shall provide you intimation regarding launching of e-version of journal of your stream time to time. This may be utilized in your library for the enrichment of knowledge of your students as well as it can also be helpful for the concerned faculty members.


After nomination of your institution as "Institutional Fellow" and constantly functioning successfully for one year, we can consider giving recognition to your institute to function as Regional/Zonal office on our behalf.
The board can also take up the additional allied activities for betterment after our consultation.

**The following entitlements are applicable to individual Fellows:**


Open Association of Research Society, U.S.A (OARS) By-laws states that an individual Fellow may use the designations as applicable, or the corresponding initials. The Credentials of individual Fellow and Associate designations signify that the individual has gained knowledge of the fundamental concepts. One is magnanimous and proficient in an expertise course covering the professional code of conduct, and follows recognized standards of practice.

Open Association of Research Society (US)/ Global Journals Incorporation (USA), as described in Corporate Statements, are educational, research publishing and professional membership organizations. Achieving our individual Fellow or Associate status is based mainly on meeting stated educational research requirements.

Disbursement of 40% Royalty earned through Global Journals : Researcher = 50%, Peer Reviewer = 37.50%, Institution = 12.50% E.g. Out of 40%, the 20% benefit should be passed on to researcher, 15 % benefit towards remuneration should be given to a reviewer and remaining 5% is to be retained by the institution.

We shall provide print version of 12 issues of any three journals [as per your requirement] out of our 38 journals worth $ 2376 USD.

**Other:**

**The individual Fellow and Associate designations accredited by Open Association of Research Society (US) credentials signify guarantees following achievements:**

- The professional accredited with Fellow honor, is entitled to various benefits viz. name, fame, honor, regular flow of income, secured bright future, social status etc.

- In addition to above, if one is single author, then entitled to 40% discount on publishing research paper and can get 10%discount if one is co-author or main author among group of authors.
- The Fellow can organize symposium/seminar/conference on behalf of Global Journals Incorporation (USA) and he/she can also attend the same organized by other institutes on behalf of Global Journals.
- The Fellow can become member of Editorial Board Member after completing 3yrs.
- The Fellow can earn 60% of sales proceeds from the sale of reference/review books/literature/publishing of research paper.
- Fellow can also join as paid peer reviewer and earn 15% remuneration of author charges and can also get an opportunity to join as member of the Editorial Board of Global Journals Incorporation (USA)
- • This individual has learned the basic methods of applying those concepts and techniques to common challenging situations. This individual has further demonstrated an in–depth understanding of the application of suitable techniques to a particular area of research practice.

**Note :**

"

- In future, if the board feels the necessity to change any board member, the same can be done with the consent of the chairperson along with anyone board member without our approval.
- In case, the chairperson needs to be replaced then consent of 2/3rd board members are required and they are also required to jointly pass the resolution copy of which should be sent to us. In such case, it will be compulsory to obtain our approval before replacement.
- In case of "Difference of Opinion [if any]" among the Board members, our decision will be final and binding to everyone.

"

# Process of submission of Research Paper

The Area or field of specialization may or may not be of any category as mentioned in 'Scope of Journal' menu of the GlobalJournals.org website. There are 37 Research Journal categorized with Six parental Journals GJCST, GJMR, GJRE, GJMBR, GJSFR, GJHSS. For Authors should prefer the mentioned categories. There are three widely used systems UDC, DDC and LCC. The details are available as 'Knowledge Abstract' at Home page. The major advantage of this coding is that, the research work will be exposed to and shared with all over the world as we are being abstracted and indexed worldwide.

The paper should be in proper format. The format can be downloaded from first page of 'Author Guideline' Menu. The Author is expected to follow the general rules as mentioned in this menu. The paper should be written in MS-Word Format (*.DOC,*.DOCX).

The Author can submit the paper either online or offline. The authors should prefer online submission.Online Submission: There are three ways to submit your paper:

**(A) (I) First, register yourself using top right corner of Home page then Login. If you are already registered, then login using your username and password.**

**(II) Choose corresponding Journal.**

**(III) Click 'Submit Manuscript'. Fill required information and Upload the paper.**

**(B) If you are using Internet Explorer, then Direct Submission through Homepage is also available.**

**(C) If these two are not convenient, and then email the paper directly to dean@globaljournals.org.**

Offline Submission: Author can send the typed form of paper by Post. However, online submission should be preferred.

# Preferred Author Guidelines

**MANUSCRIPT STYLE INSTRUCTION (Must be strictly followed)**

Page Size: 8.27" X 11'"

- Left Margin: 0.65
- Right Margin: 0.65
- Top Margin: 0.75
- Bottom Margin: 0.75
- Font type of all text should be Swis 721 Lt BT.
- Paper Title should be of Font Size 24 with one Column section.
- Author Name in Font Size of 11 with one column as of Title.
- Abstract Font size of 9 Bold, "Abstract" word in Italic Bold.
- Main Text: Font size 10 with justified two columns section
- Two Column with Equal Column with of 3.38 and Gaping of .2
- First Character must be three lines Drop capped.
- Paragraph before Spacing of 1 pt and After of 0 pt.
- Line Spacing of 1 pt
- Large Images must be in One Column
- Numbering of First Main Headings (Heading 1) must be in Roman Letters, Capital Letter, and Font Size of 10.
- Numbering of Second Main Headings (Heading 2) must be in Alphabets, Italic, and Font Size of 10.

**You can use your own standard format also.**
**Author Guidelines:**

1. General,

2. Ethical Guidelines,

3. Submission of Manuscripts,

4. Manuscript's Category,

5. Structure and Format of Manuscript,

6. After Acceptance.

**1. GENERAL**

Before submitting your research paper, one is advised to go through the details as mentioned in following heads. It will be beneficial, while peer reviewer justify your paper for publication.

**Scope**

The Global Journals Inc. (US) welcome the submission of original paper, review paper, survey article relevant to the all the streams of Philosophy and knowledge. The Global Journals Inc. (US) is parental platform for Global Journal of Computer Science and Technology, Researches in Engineering, Medical Research, Science Frontier Research, Human Social Science, Management, and Business organization. The choice of specific field can be done otherwise as following in Abstracting and Indexing Page on this Website. As the all Global

Journals Inc. (US) are being abstracted and indexed (in process) by most of the reputed organizations. Topics of only narrow interest will not be accepted unless they have wider potential or consequences.

**2. ETHICAL GUIDELINES**

Authors should follow the ethical guidelines as mentioned below for publication of research paper and research activities.

Papers are accepted on strict understanding that the material in whole or in part has not been, nor is being, considered for publication elsewhere. If the paper once accepted by Global Journals Inc. (US) and Editorial Board, will become the copyright of the Global Journals Inc. (US).

**Authorship: The authors and coauthors should have active contribution to conception design, analysis and interpretation of findings. They should critically review the contents and drafting of the paper. All should approve the final version of the paper before submission**

The Global Journals Inc. (US) follows the definition of authorship set up by the Global Academy of Research and Development. According to the Global Academy of R&D authorship, criteria must be based on:

1) Substantial contributions to conception and acquisition of data, analysis and interpretation of the findings.

2) Drafting the paper and revising it critically regarding important academic content.

3) Final approval of the version of the paper to be published.

All authors should have been credited according to their appropriate contribution in research activity and preparing paper. Contributors who do not match the criteria as authors may be mentioned under Acknowledgement.

Acknowledgements: Contributors to the research other than authors credited should be mentioned under acknowledgement. The specifications of the source of funding for the research if appropriate can be included. Suppliers of resources may be mentioned along with address.

**Appeal of Decision: The Editorial Board's decision on publication of the paper is final and cannot be appealed elsewhere.**

**Permissions: It is the author's responsibility to have prior permission if all or parts of earlier published illustrations are used in this paper.**

Please mention proper reference and appropriate acknowledgements wherever expected.

If all or parts of previously published illustrations are used, permission must be taken from the copyright holder concerned. It is the author's responsibility to take these in writing.

Approval for reproduction/modification of any information (including figures and tables) published elsewhere must be obtained by the authors/copyright holders before submission of the manuscript. Contributors (Authors) are responsible for any copyright fee involved.

**3. SUBMISSION OF MANUSCRIPTS**

Manuscripts should be uploaded via this online submission page. The online submission is most efficient method for submission of papers, as it enables rapid distribution of manuscripts and consequently speeds up the review procedure. It also enables authors to know the status of their own manuscripts by emailing us. Complete instructions for submitting a paper is available below.

Manuscript submission is a systematic procedure and little preparation is required beyond having all parts of your manuscript in a given format and a computer with an Internet connection and a Web browser. Full help and instructions are provided on-screen. As an author, you will be prompted for login and manuscript details as Field of Paper and then to upload your manuscript file(s) according to the instructions.

To avoid postal delays, all transaction is preferred by e-mail. A finished manuscript submission is confirmed by e-mail immediately and your paper enters the editorial process with no postal delays. When a conclusion is made about the publication of your paper by our Editorial Board, revisions can be submitted online with the same procedure, with an occasion to view and respond to all comments.

Complete support for both authors and co-author is provided.

**4. MANUSCRIPT'S CATEGORY**

Based on potential and nature, the manuscript can be categorized under the following heads:

Original research paper: Such papers are reports of high-level significant original research work.

Review papers: These are concise, significant but helpful and decisive topics for young researchers.

Research articles: These are handled with small investigation and applications.

Research letters: The letters are small and concise comments on previously published matters.

**5. STRUCTURE AND FORMAT OF MANUSCRIPT**

The recommended size of original research paper is less than seven thousand words, review papers fewer than seven thousands words also.Preparation of research paper or how to write research paper, are major hurdle, while writing manuscript. The research articles and research letters should be fewer than three thousand words, the structure original research paper; sometime review paper should be as follows:

**Papers**: These are reports of significant research (typically less than 7000 words equivalent, including tables, figures, references), and comprise:

(a)Title should be relevant and commensurate with the theme of the paper.

(b) A brief Summary, "Abstract" (less than 150 words) containing the major results and conclusions.

(c) Up to ten keywords, that precisely identifies the paper's subject, purpose, and focus.

(d) An Introduction, giving necessary background excluding subheadings; objectives must be clearly declared.

(e) Resources and techniques with sufficient complete experimental details (wherever possible by reference) to permit repetition; sources of information must be given and numerical methods must be specified by reference, unless non-standard.

(f) Results should be presented concisely, by well-designed tables and/or figures; the same data may not be used in both; suitable statistical data should be given. All data must be obtained with attention to numerical detail in the planning stage. As reproduced design has been recognized to be important to experiments for a considerable time, the Editor has decided that any paper that appears not to have adequate numerical treatments of the data will be returned un-refereed;

(g) Discussion should cover the implications and consequences, not just recapitulating the results; conclusions should be summarizing.

(h) Brief Acknowledgements.

(i) References in the proper form.

Authors should very cautiously consider the preparation of papers to ensure that they communicate efficiently. Papers are much more likely to be accepted, if they are cautiously designed and laid out, contain few or no errors, are summarizing, and be conventional to the approach and instructions. They will in addition, be published with much less delays than those that require much technical and editorial correction.

The Editorial Board reserves the right to make literary corrections and to make suggestions to improve briefness.

It is vital, that authors take care in submitting a manuscript that is written in simple language and adheres to published guidelines.

**Format**

*Language: The language of publication is UK English. Authors, for whom English is a second language, must have their manuscript efficiently edited by an English-speaking person before submission to make sure that, the English is of high excellence. It is preferable, that manuscripts should be professionally edited.*

Standard Usage, Abbreviations, and Units: Spelling and hyphenation should be conventional to The Concise Oxford English Dictionary. Statistics and measurements should at all times be given in figures, e.g. 16 min, except for when the number begins a sentence. When the number does not refer to a unit of measurement it should be spelt in full unless, it is 160 or greater.

Abbreviations supposed to be used carefully. The abbreviated name or expression is supposed to be cited in full at first usage, followed by the conventional abbreviation in parentheses.

Metric SI units are supposed to generally be used excluding where they conflict with current practice or are confusing. For illustration, 1.4 l rather than 1.4 × 10-3 m3, or 4 mm somewhat than 4 × 10-3 m. Chemical formula and solutions must identify the form used, e.g. anhydrous or hydrated, and the concentration must be in clearly defined units. Common species names should be followed by underlines at the first mention. For following use the generic name should be constricted to a single letter, if it is clear.

**Structure**

All manuscripts submitted to Global Journals Inc. (US), ought to include:

Title: The title page must carry an instructive title that reflects the content, a running title (less than 45 characters together with spaces), names of the authors and co-authors, and the place(s) wherever the work was carried out. The full postal address in addition with the e-mail address of related author must be given. Up to eleven keywords or very brief phrases have to be given to help data retrieval, mining and indexing.

*Abstract, used in Original Papers and Reviews:*

Optimizing Abstract for Search Engines

Many researchers searching for information online will use search engines such as Google, Yahoo or similar. By optimizing your paper for search engines, you will amplify the chance of someone finding it. This in turn will make it more likely to be viewed and/or cited in a further work. Global Journals Inc. (US) have compiled these guidelines to facilitate you to maximize the web-friendliness of the most public part of your paper.

Key Words

A major linchpin in research work for the writing research paper is the keyword search, which one will employ to find both library and Internet resources.

One must be persistent and creative in using keywords. An effective keyword search requires a strategy and planning a list of possible keywords and phrases to try.

Search engines for most searches, use Boolean searching, which is somewhat different from Internet searches. The Boolean search uses "operators," words (and, or, not, and near) that enable you to expand or narrow your affords. Tips for research paper while preparing research paper are very helpful guideline of research paper.

Choice of key words is first tool of tips to write research paper. Research paper writing is an art.A few tips for deciding as strategically as possible about keyword search:

- One should start brainstorming lists of possible keywords before even begin searching. Think about the most important concepts related to research work. Ask, "What words would a source have to include to be truly valuable in research paper?" Then consider synonyms for the important words.
- It may take the discovery of only one relevant paper to let steer in the right keyword direction because in most databases, the keywords under which a research paper is abstracted are listed with the paper.
- One should avoid outdated words.

Keywords are the key that opens a door to research work sources. Keyword searching is an art in which researcher's skills are bound to improve with experience and time.

Numerical Methods: Numerical methods used should be clear and, where appropriate, supported by references.

*Acknowledgements: Please make these as concise as possible.*

References

References follow the Harvard scheme of referencing. References in the text should cite the authors' names followed by the time of their publication, unless there are three or more authors when simply the first author's name is quoted followed by et al. unpublished work has to only be cited where necessary, and only in the text. Copies of references in press in other journals have to be supplied with submitted typescripts. It is necessary that all citations and references be carefully checked before submission, as mistakes or omissions will cause delays.

References to information on the World Wide Web can be given, but only if the information is available without charge to readers on an official site. Wikipedia and Similar websites are not allowed where anyone can change the information. Authors will be asked to make available electronic copies of the cited information for inclusion on the Global Journals Inc. (US) homepage at the judgment of the Editorial Board.

The Editorial Board and Global Journals Inc. (US) recommend that, citation of online-published papers and other material should be done via a DOI (digital object identifier). If an author cites anything, which does not have a DOI, they run the risk of the cited material not being noticeable.

The Editorial Board and Global Journals Inc. (US) recommend the use of a tool such as Reference Manager for reference management and formatting.

Tables, Figures and Figure Legends

*Tables: Tables should be few in number, cautiously designed, uncrowned, and include only essential data. Each must have an Arabic number, e.g. Table 4, a self-explanatory caption and be on a separate sheet. Vertical lines should not be used.*

*Figures: Figures are supposed to be submitted as separate files. Always take in a citation in the text for each figure using Arabic numbers, e.g. Fig. 4. Artwork must be submitted online in electronic form by e-mailing them.*

Preparation of Electronic Figures for Publication

Even though low quality images are sufficient for review purposes, print publication requires high quality images to prevent the final product being blurred or fuzzy. Submit (or e-mail) EPS (line art) or TIFF (halftone/photographs) files only. MS PowerPoint and Word Graphics are unsuitable for printed pictures. Do not use pixel-oriented software. Scans (TIFF only) should have a resolution of at least 350 dpi (halftone) or 700 to 1100 dpi (line drawings) in relation to the imitation size. Please give the data for figures in black and white or submit a Color Work Agreement Form. EPS files must be saved with fonts embedded (and with a TIFF preview, if possible).

For scanned images, the scanning resolution (at final image size) ought to be as follows to ensure good reproduction: line art: >650 dpi; halftones (including gel photographs) : >350 dpi; figures containing both halftone and line images: >650 dpi.

Color Charges: It is the rule of the Global Journals Inc. (US) for authors to pay the full cost for the reproduction of their color artwork. Hence, please note that, if there is color artwork in your manuscript when it is accepted for publication, we would require you to complete and return a color work agreement form before your paper can be published.



*Figure Legends: Self-explanatory legends of all figures should be incorporated separately under the heading 'Legends to Figures'. In the full-text online edition of the journal, figure legends may possibly be truncated in abbreviated links to the full screen version. Therefore, the first 100 characters of any legend should notify the reader, about the key aspects of the figure.*

**6. AFTER ACCEPTANCE**

Upon approval of a paper for publication, the manuscript will be forwarded to the dean, who is responsible for the publication of the Global Journals Inc. (US).

**6.1 Proof Corrections**

The corresponding author will receive an e-mail alert containing a link to a website or will be attached. A working e-mail address must therefore be provided for the related author.

Acrobat Reader will be required in order to read this file. This software can be downloaded

(Free of charge) from the following website:

www.adobe.com/products/acrobat/readstep2.html. This will facilitate the file to be opened, read on screen, and printed out in order for any corrections to be added. Further instructions will be sent with the proof.

Proofs must be returned to the dean at dean@globaljournals.org within three days of receipt.

As changes to proofs are costly, we inquire that you only correct typesetting errors. All illustrations are retained by the publisher. Please note that the authors are responsible for all statements made in their work, including changes made by the copy editor.

**6.2 Early View of Global Journals Inc. (US) (Publication Prior to Print)**

The Global Journals Inc. (US) are enclosed by our publishing's Early View service. Early View articles are complete full-text articles sent in advance of their publication. Early View articles are absolute and final. They have been completely reviewed, revised and edited for publication, and the authors' final corrections have been incorporated. Because they are in final form, no changes can be made after sending them. The nature of Early View articles means that they do not yet have volume, issue or page numbers, so Early View articles cannot be cited in the conventional way.

**6.3 Author Services**

Online production tracking is available for your article through Author Services. Author Services enables authors to track their article - once it has been accepted - through the production process to publication online and in print. Authors can check the status of their articles online and choose to receive automated e-mails at key stages of production. The authors will receive an e-mail with a unique link that enables them to register and have their article automatically added to the system. Please ensure that a complete e-mail address is provided when submitting the manuscript.

**6.4 Author Material Archive Policy**

Please note that if not specifically requested, publisher will dispose off hardcopy & electronic information submitted, after the two months of publication. If you require the return of any information submitted, please inform the Editorial Board or dean as soon as possible.

**6.5 Offprint and Extra Copies**

A PDF offprint of the online-published article will be provided free of charge to the related author, and may be distributed according to the Publisher's terms and conditions. Additional paper offprint may be ordered by emailing us at: editor@globaljournals.org .

You must strictly follow above Author Guidelines before submitting your paper or else we will not at all be responsible for any corrections in future in any of the way.

Before start writing a good quality Computer Science Research Paper, let us first understand what is Computer Science Research Paper? So, Computer Science Research Paper is the paper which is written by professionals or scientists who are associated to Computer Science and Information Technology, or doing research study in these areas. If you are novel to this field then you can consult about this field from your supervisor or guide.

## TECHNIQUES FOR WRITING A GOOD QUALITY RESEARCH PAPER:

**1. Choosing the topic:** In most cases, the topic is searched by the interest of author but it can be also suggested by the guides. You can have several topics and then you can judge that in which topic or subject you are finding yourself most comfortable. This can be done by asking several questions to yourself, like Will I be able to carry our search in this area? Will I find all necessary recourses to accomplish the search? Will I be able to find all information in this field area? If the answer of these types of questions will be "Yes" then you can choose that topic. In most of the cases, you may have to conduct the surveys and have to visit several places because this field is related to Computer Science and Information Technology. Also, you may have to do a lot of work to find all rise and falls regarding the various data of that subject. Sometimes, detailed information plays a vital role, instead of short information.

**2. Evaluators are human:** First thing to remember that evaluators are also human being. They are not only meant for rejecting a paper. They are here to evaluate your paper. So, present your Best.

**3. Think Like Evaluators:** If you are in a confusion or getting demotivated that your paper will be accepted by evaluators or not, then think and try to evaluate your paper like an Evaluator. Try to understand that what an evaluator wants in your research paper and automatically you will have your answer.

**4. Make blueprints of paper:** The outline is the plan or framework that will help you to arrange your thoughts. It will make your paper logical. But remember that all points of your outline must be related to the topic you have chosen.

**5. Ask your Guides:** If you are having any difficulty in your research, then do not hesitate to share your difficulty to your guide (if you have any). They will surely help you out and resolve your doubts. If you can't clarify what exactly you require for your work then ask the supervisor to help you with the alternative. He might also provide you the list of essential readings.

**6. Use of computer is recommended:** As you are doing research in the field of Computer Science, then this point is quite obvious.

**7. Use right software:** Always use good quality software packages. If you are not capable to judge good software then you can lose quality of your paper unknowingly. There are various software programs available to help you, which you can get through Internet.

**8. Use the Internet for help:** An excellent start for your paper can be by using the Google. It is an excellent search engine, where you can have your doubts resolved. You may also read some answers for the frequent question how to write my research paper or find model research paper. From the internet library you can download books. If you have all required books make important reading selecting and analyzing the specified information. Then put together research paper sketch out.

**9. Use and get big pictures:** Always use encyclopedias, Wikipedia to get pictures so that you can go into the depth.

**10. Bookmarks are useful:** When you read any book or magazine, you generally use bookmarks, right! It is a good habit, which helps to not to lose your continuity. You should always use bookmarks while searching on Internet also, which will make your search easier.

**11. Revise what you wrote:** When you write anything, always read it, summarize it and then finalize it.



**12. Make all efforts:** Make all efforts to mention what you are going to write in your paper. That means always have a good start. Try to mention everything in introduction, that what is the need of a particular research paper. Polish your work by good skill of writing and always give an evaluator, what he wants.

**13. Have backups:** When you are going to do any important thing like making research paper, you should always have backup copies of it either in your computer or in paper. This will help you to not to lose any of your important.

**14. Produce good diagrams of your own:** Always try to include good charts or diagrams in your paper to improve quality. Using several and unnecessary diagrams will degrade the quality of your paper by creating "hotchpotch." So always, try to make and include those diagrams, which are made by your own to improve readability and understandability of your paper.

**15. Use of direct quotes:** When you do research relevant to literature, history or current affairs then use of quotes become essential but if study is relevant to science then use of quotes is not preferable.

**16. Use proper verb tense:** Use proper verb tenses in your paper. Use past tense, to present those events that happened. Use present tense to indicate events that are going on. Use future tense to indicate future happening events. Use of improper and wrong tenses will confuse the evaluator. Avoid the sentences that are incomplete.

**17. Never use online paper:** If you are getting any paper on Internet, then never use it as your research paper because it might be possible that evaluator has already seen it or maybe it is outdated version.

**18. Pick a good study spot:** To do your research studies always try to pick a spot, which is quiet. Every spot is not for studies. Spot that suits you choose it and proceed further.

**19. Know what you know:** Always try to know, what you know by making objectives. Else, you will be confused and cannot achieve your target.

**20. Use good quality grammar:** Always use a good quality grammar and use words that will throw positive impact on evaluator. Use of good quality grammar does not mean to use tough words, that for each word the evaluator has to go through dictionary. Do not start sentence with a conjunction. Do not fragment sentences. Eliminate one-word sentences. Ignore passive voice. Do not ever use a big word when a diminutive one would suffice. Verbs have to be in agreement with their subjects. Prepositions are not expressions to finish sentences with. It is incorrect to ever divide an infinitive. Avoid clichés like the disease. Also, always shun irritating alliteration. Use language that is simple and straight forward. put together a neat summary.

**21. Arrangement of information:** Each section of the main body should start with an opening sentence and there should be a changeover at the end of the section. Give only valid and powerful arguments to your topic. You may also maintain your arguments with records.

**22. Never start in last minute:** Always start at right time and give enough time to research work. Leaving everything to the last minute will degrade your paper and spoil your work.

**23. Multitasking in research is not good:** Doing several things at the same time proves bad habit in case of research activity. Research is an area, where everything has a particular time slot. Divide your research work in parts and do particular part in particular time slot.

**24. Never copy others' work:** Never copy others' work and give it your name because if evaluator has seen it anywhere you will be in trouble.

**25. Take proper rest and food:** No matter how many hours you spend for your research activity, if you are not taking care of your health then all your efforts will be in vain. For a quality research, study is must, and this can be done by taking proper rest and food.

**26. Go for seminars:** Attend seminars if the topic is relevant to your research area. Utilize all your resources.

**27. Refresh your mind after intervals:** Try to give rest to your mind by listening to soft music or by sleeping in intervals. This will also improve your memory.

**28. Make colleagues:** Always try to make colleagues. No matter how sharper or intelligent you are, if you make colleagues you can have several ideas, which will be helpful for your research.

**29. Think technically:** Always think technically. If anything happens, then search its reasons, its benefits, and demerits.

**30. Think and then print:** When you will go to print your paper, notice that tables are not be split, headings are not detached from their descriptions, and page sequence is maintained.

**31. Adding unnecessary information:** Do not add unnecessary information, like, I have used MS Excel to draw graph. Do not add irrelevant and inappropriate material. These all will create superfluous. Foreign terminology and phrases are not apropos. One should NEVER take a broad view. Analogy in script is like feathers on a snake. Not at all use a large word when a very small one would be sufficient. Use words properly, regardless of how others use them. Remove quotations. Puns are for kids, not grunt readers. Amplification is a billion times of inferior quality than sarcasm.

**32. Never oversimplify everything:** To add material in your research paper, never go for oversimplification. This will definitely irritate the evaluator. Be more or less specific. Also too, by no means, ever use rhythmic redundancies. Contractions aren't essential and shouldn't be there used. Comparisons are as terrible as clichés. Give up ampersands and abbreviations, and so on. Remove commas, that are, not necessary. Parenthetical words however should be together with this in commas. Understatement is all the time the complete best way to put onward earth-shaking thoughts. Give a detailed literary review.

**33. Report concluded results:** Use concluded results. From raw data, filter the results and then conclude your studies based on measurements and observations taken. Significant figures and appropriate number of decimal places should be used. Parenthetical remarks are prohibitive. Proofread carefully at final stage. In the end give outline to your arguments. Spot out perspectives of further study of this subject. Justify your conclusion by at the bottom of them with sufficient justifications and examples.

**34. After conclusion:** Once you have concluded your research, the next most important step is to present your findings. Presentation is extremely important as it is the definite medium though which your research is going to be in print to the rest of the crowd. Care should be taken to categorize your thoughts well and present them in a logical and neat manner. A good quality research paper format is essential because it serves to highlight your research paper and bring to light all necessary aspects in your research.

## Informal Guidelines of Research Paper Writing

**Key points to remember:**

- Submit all work in its final form.
- Write your paper in the form, which is presented in the guidelines using the template.
- Please note the criterion for grading the final paper by peer-reviewers.

**Final Points:**

A purpose of organizing a research paper is to let people to interpret your effort selectively. The journal requires the following sections, submitted in the order listed, each section to start on a new page.

The introduction will be compiled from reference matter and will reflect the design processes or outline of basis that direct you to make study. As you will carry out the process of study, the method and process section will be constructed as like that. The result segment will show related statistics in nearly sequential order and will direct the reviewers next to the similar intellectual paths throughout the data that you took to carry out your study. The discussion section will provide understanding of the data and projections as to the implication of the results. The use of good quality references all through the paper will give the effort trustworthiness by representing an alertness of prior workings.



Writing a research paper is not an easy job no matter how trouble-free the actual research or concept. Practice, excellent preparation, and controlled record keeping are the only means to make straightforward the progression.

**General style:**

Specific editorial column necessities for compliance of a manuscript will always take over from directions in these general guidelines.

To make a paper clear

· Adhere to recommended page limits

Mistakes to evade

- Insertion a title at the foot of a page with the subsequent text on the next page
- Separating a table/chart or figure - impound each figure/table to a single page
- Submitting a manuscript with pages out of sequence

In every sections of your document

· Use standard writing style including articles ("a", "the," etc.)

· Keep on paying attention on the research topic of the paper

· Use paragraphs to split each significant point (excluding for the abstract)

· Align the primary line of each section

· Present your points in sound order

· Use present tense to report well accepted

· Use past tense to describe specific results

· Shun familiar wording, don't address the reviewer directly, and don't use slang, slang language, or superlatives

· Shun use of extra pictures - include only those figures essential to presenting results

**Title Page:**

Choose a revealing title. It should be short. It should not have non-standard acronyms or abbreviations. It should not exceed two printed lines. It should include the name(s) and address (es) of all authors.

**Abstract:**

The summary should be two hundred words or less. It should briefly and clearly explain the key findings reported in the manuscript--must have precise statistics. It should not have abnormal acronyms or abbreviations. It should be logical in itself. Shun citing references at this point.

An abstract is a brief distinct paragraph summary of finished work or work in development. In a minute or less a reviewer can be taught the foundation behind the study, common approach to the problem, relevant results, and significant conclusions or new questions.

Write your summary when your paper is completed because how can you write the summary of anything which is not yet written? Wealth of terminology is very essential in abstract. Yet, use comprehensive sentences and do not let go readability for briefness. You can maintain it succinct by phrasing sentences so that they provide more than lone rationale. The author can at this moment go straight to shortening the outcome. Sum up the study, with the subsequent elements in any summary. Try to maintain the initial two items to no more than one ruling each.

- Reason of the study - theory, overall issue, purpose
- Fundamental goal
- To the point depiction of the research
- Consequences, including definite statistics - if the consequences are quantitative in nature, account quantitative data; results of any numerical analysis should be reported
- Significant conclusions or questions that track from the research(es)

Approach:

- Single section, and succinct
- As a outline of job done, it is always written in past tense
- A conceptual should situate on its own, and not submit to any other part of the paper such as a form or table
- Center on shortening results - bound background information to a verdict or two, if completely necessary
- What you account in an conceptual must be regular with what you reported in the manuscript
- Exact spelling, clearness of sentences and phrases, and appropriate reporting of quantities (proper units, important statistics) are just as significant in an abstract as they are anywhere else

**Introduction:**

The **Introduction** should "introduce" the manuscript. The reviewer should be presented with sufficient background information to be capable to comprehend and calculate the purpose of your study without having to submit to other works. The basis for the study should be offered. Give most important references but shun difficult to make a comprehensive appraisal of the topic. In the introduction, describe the problem visibly. If the problem is not acknowledged in a logical, reasonable way, the reviewer will have no attention in your result. Speak in common terms about techniques used to explain the problem, if needed, but do not present any particulars about the protocols here. Following approach can create a valuable beginning:

- Explain the value (significance) of the study
- Shield the model - why did you employ this particular system or method? What is its compensation? You strength remark on its appropriateness from a abstract point of vision as well as point out sensible reasons for using it.
- Present a justification. Status your particular theory (es) or aim(s), and describe the logic that led you to choose them.
- Very for a short time explain the tentative propose and how it skilled the declared objectives.

Approach:

- Use past tense except for when referring to recognized facts. After all, the manuscript will be submitted after the entire job is done.
- Sort out your thoughts; manufacture one key point with every section. If you make the four points listed above, you will need a least of four paragraphs.



- Present surroundings information only as desirable in order hold up a situation. The reviewer does not desire to read the whole thing you know about a topic.
- Shape the theory/purpose specifically - do not take a broad view.
- As always, give awareness to spelling, simplicity and correctness of sentences and phrases.

**Procedures (Methods and Materials):**

This part is supposed to be the easiest to carve if you have good skills. A sound written Procedures segment allows a capable scientist to replacement your results. Present precise information about your supplies. The suppliers and clarity of reagents can be helpful bits of information. Present methods in sequential order but linked methodologies can be grouped as a segment. Be concise when relating the protocols. Attempt for the least amount of information that would permit another capable scientist to spare your outcome but be cautious that vital information is integrated. The use of subheadings is suggested and ought to be synchronized with the results section. When a technique is used that has been well described in another object, mention the specific item describing a way but draw the basic principle while stating the situation. The purpose is to text all particular resources and broad procedures, so that another person may use some or all of the methods in one more study or referee the scientific value of your work. It is not to be a step by step report of the whole thing you did, nor is a methods section a set of orders.

Materials:

- Explain materials individually only if the study is so complex that it saves liberty this way.
- Embrace particular materials, and any tools or provisions that are not frequently found in laboratories.
- Do not take in frequently found.
- If use of a definite type of tools.
- Materials may be reported in a part section or else they may be recognized along with your measures.

Methods:

- Report the method (not particulars of each process that engaged the same methodology)
- Describe the method entirely
- To be succinct, present methods under headings dedicated to specific dealings or groups of measures
- Simplify - details how procedures were completed not how they were exclusively performed on a particular day.
- If well known procedures were used, account the procedure by name, possibly with reference, and that's all.

Approach:

- It is embarrassed or not possible to use vigorous voice when documenting methods with no using first person, which would focus the reviewer's interest on the researcher rather than the job. As a result when script up the methods most authors use third person passive voice.
- Use standard style in this and in every other part of the paper - avoid familiar lists, and use full sentences.

What to keep away from

- Resources and methods are not a set of information.
- Skip all descriptive information and surroundings - save it for the argument.
- Leave out information that is immaterial to a third party.

**Results:**

The principle of a results segment is to present and demonstrate your conclusion. Create this part a entirely objective details of the outcome, and save all understanding for the discussion.

The page length of this segment is set by the sum and types of data to be reported. Carry on to be to the point, by means of statistics and tables, if suitable, to present consequences most efficiently.You must obviously differentiate material that would usually be incorporated in a study editorial from any unprocessed data or additional appendix matter that would not be available. In fact, such matter should not be submitted at all except requested by the instructor.

Content

- Sum up your conclusion in text and demonstrate them, if suitable, with figures and tables.
- In manuscript, explain each of your consequences, point the reader to remarks that are most appropriate.
- Present a background, such as by describing the question that was addressed by creation an exacting study.
- Explain results of control experiments and comprise remarks that are not accessible in a prescribed figure or table, if appropriate.
- Examine your data, then prepare the analyzed (transformed) data in the form of a figure (graph), table, or in manuscript form.

What to stay away from

- Do not discuss or infer your outcome, report surroundings information, or try to explain anything.
- Not at all, take in raw data or intermediate calculations in a research manuscript.
- Do not present the similar data more than once.
- Manuscript should complement any figures or tables, not duplicate the identical information.
- Never confuse figures with tables - there is a difference.

Approach

- As forever, use past tense when you submit to your results, and put the whole thing in a reasonable order.
- Put figures and tables, appropriately numbered, in order at the end of the report
- If you desire, you may place your figures and tables properly within the text of your results part.

Figures and tables

- If you put figures and tables at the end of the details, make certain that they are visibly distinguished from any attach appendix materials, such as raw facts
- Despite of position, each figure must be numbered one after the other and complete with subtitle
- In spite of position, each table must be titled, numbered one after the other and complete with heading
- All figure and table must be adequately complete that it could situate on its own, divide from text

**Discussion:**

The Discussion is expected the trickiest segment to write and describe. A lot of papers submitted for journal are discarded based on problems with the Discussion. There is no head of state for how long a argument should be. Position your understanding of the outcome visibly to lead the reviewer through your conclusions, and then finish the paper with a summing up of the implication of the study. The purpose here is to offer an understanding of your results and hold up for all of your conclusions, using facts from your research and generally accepted information, if suitable. The implication of result should be visibly described. Infer your data in the conversation in suitable depth. This means that when you clarify an observable fact you must explain mechanisms that may account for the observation. If your results vary from your prospect, make clear why that may have happened. If your results agree, then explain the theory that the proof supported. It is never suitable to just state that the data approved with prospect, and let it drop at that.

- Make a decision if each premise is supported, discarded, or if you cannot make a conclusion with assurance. Do not just dismiss a study or part of a study as "uncertain."
- Research papers are not acknowledged if the work is imperfect. Draw what conclusions you can based upon the results that you have, and take care of the study as a finished work
- You may propose future guidelines, such as how the experiment might be personalized to accomplish a new idea.
- Give details all of your remarks as much as possible, focus on mechanisms.
- Make a decision if the tentative design sufficiently addressed the theory, and whether or not it was correctly restricted.
- Try to present substitute explanations if sensible alternatives be present.
- One research will not counter an overall question, so maintain the large picture in mind, where do you go next? The best studies unlock new avenues of study. What questions remain?
- Recommendations for detailed papers will offer supplementary suggestions.

Approach:

- When you refer to information, differentiate data generated by your own studies from available information
- Submit to work done by specific persons (including you) in past tense.
- Submit to generally acknowledged facts and main beliefs in present tense.

Please carefully note down following rules and regulation before submitting your Research Paper to Global Journals Inc. (US):

**Segment Draft and Final Research Paper:** You have to strictly follow the template of research paper. If it is not done your paper may get rejected.

- The **major constraint** is that you must independently make all content, tables, graphs, and facts that are offered in the paper. You must write each part of the paper wholly on your own. The Peer-reviewers need to identify your own perceptive of the concepts in your own terms. NEVER extract straight from any foundation, and never rephrase someone else's analysis.

- Do not give permission to anyone else to "PROOFREAD" your manuscript.

- Methods to avoid Plagiarism is applied by us on every paper, if found guilty, you will be blacklisted by all of our collaborated research groups, your institution will be informed for this and strict legal actions will be taken immediately.)
- To guard yourself and others from possible illegal use please do not permit anyone right to use to your paper and files.

## Criterion for Grading a Research Paper (Compilation) by Global Journals Inc. (US)

**Please note that following table is only a Grading of "Paper Compilation" and not on "Performed/Stated Research" whose grading solely depends on Individual Assigned Peer Reviewer and Editorial Board Member. These can be available only on request and after decision of Paper. This report will be the property of Global Journals Inc. (US).**

| Topics | Grades | | |
|---|---|---|---|
| | **A-B** | **C-D** | **E-F** |
| ***Abstract*** | Clear and concise with appropriate content, Correct format. 200 words or below | Unclear summary and no specific data, Incorrect form<br><br>Above 200 words | No specific data with ambiguous information<br><br>Above 250 words |
| ***Introduction*** | Containing all background details with clear goal and appropriate details, flow specification, no grammar and spelling mistake, well organized sentence and paragraph, reference cited | Unclear and confusing data, appropriate format, grammar and spelling errors with unorganized matter | Out of place depth and content, hazy format |
| ***Methods and Procedures*** | Clear and to the point with well arranged paragraph, precision and accuracy of facts and figures, well organized subheads | Difficult to comprehend with embarrassed text, too much explanation but completed | Incorrect and unorganized structure with hazy meaning |
| ***Result*** | Well organized, Clear and specific, Correct units with precision, correct data, well structuring of paragraph, no grammar and spelling mistake | Complete and embarrassed text, difficult to comprehend | Irregular format with wrong facts and figures |
| ***Discussion*** | Well organized, meaningful specification, sound conclusion, logical and concise explanation, highly structured paragraph reference cited | Wordy, unclear conclusion, spurious | Conclusion is not cited, unorganized, difficult to comprehend |
| ***References*** | Complete and correct format, well organized | Beside the point, Incomplete | Wrong format and structuring |



# INDEX

save our planet

# Global Journal of Computer Science and Technology

Visit us on the Web at www.GlobalJournals.org | www.ComputerResearch.org
or email us at helpdesk@globaljournals.org

ISSN 9754350