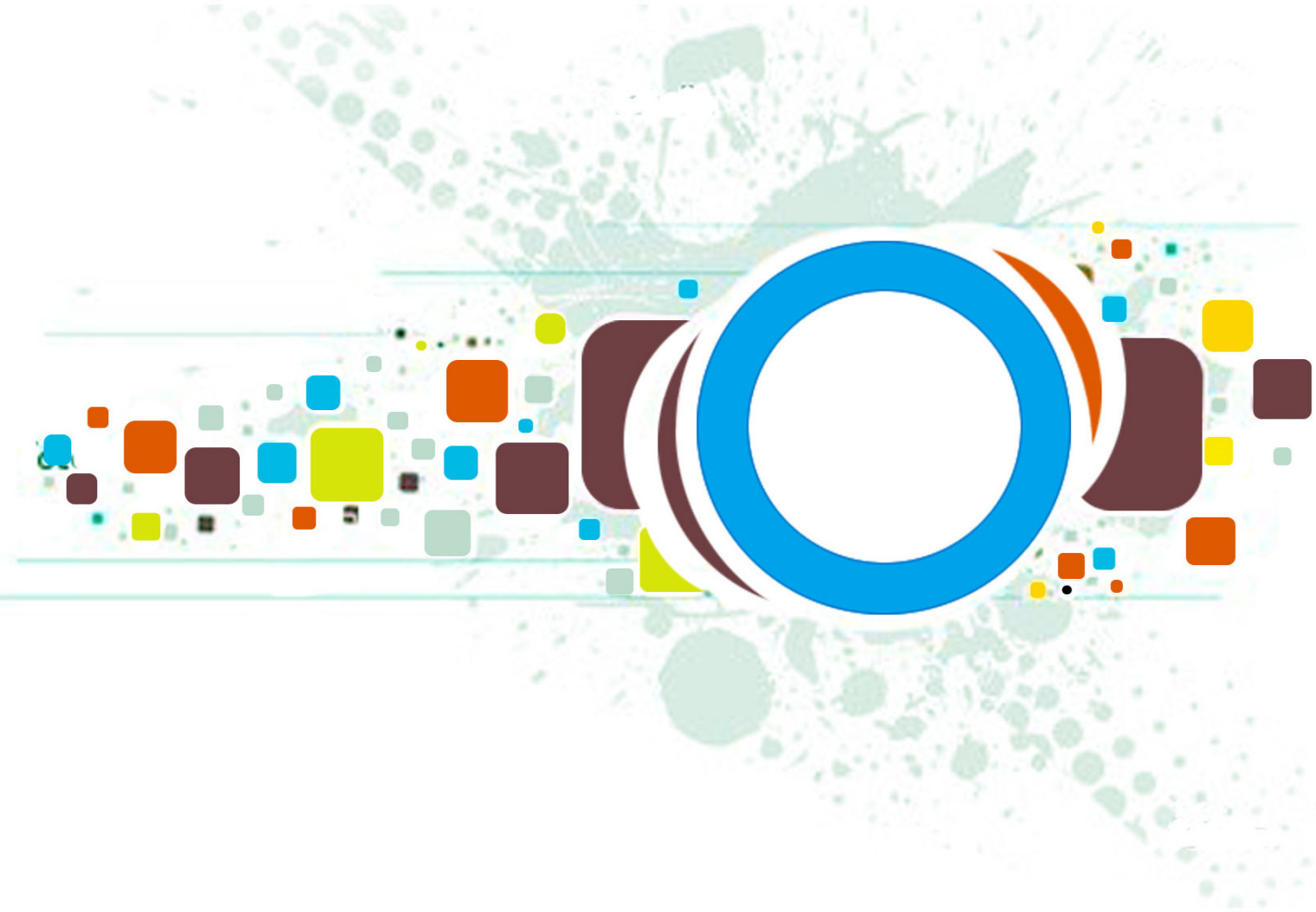
Volume 9 • Issue 1 • January / February 2015

INTERNATIONAL JOURNAL OF

IMAGE PROCESSING (IJIP)

ISSN : 1985-2304

Publication Frequency: 6 Issues Per Year



CSC PUBLISHERS
http://www.cscjournals.org

INTERNATIONAL JOURNAL OF IMAGE PROCESSING (IJIP)

VOLUME 9, ISSUE 1, 2015

EDITED BY
DR. NABEEL TAHIR

ISSN (Online): 1985-2304

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

IJIP Journal is a part of CSC Publishers

Computer Science Journals

http://www.cscjournals.org

INTERNATIONAL JOURNAL OF IMAGE PROCESSING (IJIP)

Book: Volume 9, Issue 1, January / February 2015
Publishing Date: 28-02-2015
ISSN (Online): 1985-2304

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

Published in Malaysia

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

CSC Publishers, 2015

EDITORIAL PREFACE

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

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

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

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

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

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

Editorial Board Members
International Journal of Image Processing (IJIP)

EDITORIAL BOARD

TABLE OF CONTENTS

Volume 9, Issue 1, January / February 2015

Pages

Three-dimensional Face Shape by Local Features Prediction

Rubén García-Zurdo *rubengarcia@cu-cisneros.es*
Psychology Division
Colegio Universitario "Cardenal Cisneros"
Madrid, 28006, Spain

Abstract

A method is presented to estimate the 3D face shape from a frontal image using a multivariate linear regression model between intensity and depth features: block based discrete cosine transform (DCT), block based principal component analysis (PCA) and modular PCA (eigenfeatures) coefficients. After noting that between-illumination coefficients variances are smaller than between-subjects coefficients variances, we try to correct illumination variations by discarding the first coefficient in the DCT method but not in the PCA methods, as the between-illumination coefficients variances are distributed over many of them. The lowest fractional error in depth prediction is obtained by using a low number of coefficients and a high overlap degree between blocks. Modular PCA produces the best results when the test image is frontally illuminated as in the training phase. DCT and local PCA are more robust across point source horizontal angle and ambient illumination variations.

Keywords: Shape from Shading, Three-dimensional Face Shape, Eigenfeatures, Multivariate Regression.

1. INTRODUCTION

Statistical shape from shading aims to detect regularities between global variations in intensity and depth images, usually with faces, the goal being able to build a 3D face shape from its intensity image. Principal component analysis (PCA) is used [1] as a dimensionality reduction technique over a 3D head sample along with an optimization method to predict the coefficients that allow reconstruction of a 3D test image shape. In [2] PCA is applied separately to shape and texture after applying a dense one-to-one correspondence to an internal reference model, allowing reconstruction of a 3D face shape and even generate new face shapes and expressions as a linear transformation of the analysis extracted models. In [3] PCA is applied to a coupled intensity-depth model capturing the joint maximum variation directions between intensity and depth and allowing to implicitly recover shape from intensity.

Some of the 3D face shape reconstruction techniques developed in the face recognition area have not always been applied globally. PCA allows us to discover joint maximum variation direction in a sample of images from the eigenvectors calculated over the covariance matrix, allowing reconstruction of every original image from a reduced number of coefficients as a weighted sum of eigenvectors and to identify it by using the Euclidean distance between the coefficients vector and another reference vector [4]. In modular PCA [5] images are divided into a specific number of sections and individual PCAs are applied for each section, which yield better results in variable illumination and facial expression conditions, as local variations don't affect every section. Therefore, using local representations improves the representation of facial features in variable illumination and expression situations. Eigenfeatures systems are modular PCA systems applied to main facial anatomic features: eyes, nose and mouth. In [6] they are applied with a recognition rate next to global PCA under varying expression and transparent glasses conditions, although not in partial occlusion.

The discrete cosine transform (DCT) has been used as an alternative to PCA facial analysis and recognition [7]. 2D DCT decomposes an image into a weighted sum of cosine basis functions with different orientation and spatial frequency. The first coefficients, corresponding to the lower frequencies and occupying the top positions in a zigzag pattern in the spectral representation, are sufficient to achieve an acceptable reconstruction and perform recognition with good results. Compared to PCA, DCT has the advantage of being based on a fixed set of basis functions that are not to be calculated beforehand. Discarding the first DCT coefficients increases recognition in images with large illumination variations because they contain information about the illumination changes [8]. Scaling all but the the first DCT coefficients according to a global maximum and minimum produces better recognition without the loss of information which is ruled out by discarding some of the initial coefficients [9]. It has been found that integration of global and local DCT features improves recognition [10], and the use of local features from DCT and local PCA in face verification is more robust in pose variations than global features [11].

As a way to improve recognition results by overcoming pose variations, some methods have been developed to construct different virtual pose views by using linear prediction models from local representations. In [12], [13] non frontal views of a face are predicted from a frontal image using multivariate linear regression (MLR) on the DCT coefficients of 8x8 size blocks. In the training phase, a linear transformation matrix is calculated for each pair of blocks occupying the same position in a pair of images of the same subject in different views. A transformation matrix of acceptable size is obtained for a moderate number of image samples by using dimensionality reduction techniques on each block. Blocks with different degrees of overlap are used and results are averaged in order to reduce prediction errors. Best results are obtained with higher overlap degree (87.5%) and using a reduced number of DCT coefficients (25%), as the dimensionality of the representation is kept low relative to the number of training elements. A method is presented in [14] to predict frontal appearance from a non-frontal pose image by calculating a linear operator relating views in different poses, both globally and based on different sized blocks between views whose correspondence is estimated with the help of an average 3D reference model.

The goal of this research is to show that depth prediction from intensity using local features may provide acceptable results. We use MLR to predict the depth corresponding to a frontal intensity image by calculating the linear transformation matrices that relate intensity and depth block pairs for each subject in the training phase. Then we are able to study whether depth predictions from out-of-training images are acceptable. We study the effect produced by the horizontal and vertical overlap degree between blocks, the dimensionality or number of coefficients used and the effect of variable illumination on the predictions of three different local representations: block based DCT, block based local PCA with a common eigenvectors base and modular PCA applied to bigger blocks corresponding to anatomic face features (eigenfeatures).

2. MULTIVARIATE LINEAR REGRESSION FOR DEPTH PREDICTION FROM LOCAL FEATURES

Let I be an intensity image, Z its corresponding depth image belonging to an N size sample of intensity and depth images, $I_{(x,y)}$ a fixed size block from I starting at (x,y) coordinates and $Z_{(x,y)}$ its corresponding depth block from Z. After performing a dimensionality reduction to D coefficients on both blocks for all N pairs we define a linear regression model:

$$B_{(x,y)} = A_{(x,y)} \quad W_{(x,y)}$$

$$\begin{bmatrix} Z_1^T \\ Z_2^T \\ \vdots \\ Z_N^T \end{bmatrix} = \begin{bmatrix} 1 & I_1^T \\ 1 & I_2^T \\ & \vdots \\ 1 & I_N^T \end{bmatrix} \begin{bmatrix} W_{1,1} & \cdots & W_{1,D} \\ W_{2,1} & \cdots & W_{2,D} \\ \vdots & \vdots & \vdots \\ W_{D+1,1} & \cdots & W_{D+1,D} \end{bmatrix} \quad (1)$$

Where B is the target vectors matrix, A is the extended source vectors matrix and W is a transformation matrix. In order to get W, we turn to the solution of the corresponding linear equations system using the pseudo-inverse matrix:

$$W_{(x,y)} = \left(A^T_{(x,y)} A_{(x,y)}\right)^{-1} A^T_{(x,y)} B_{(x,y)} \quad (2)$$

Every intensity vector may be then transformed into a depth vector by applying equation (1), and an intensity image may be transformed into a depth image by performing the same calculation for every block:

$$\{I_{(x_1,y_1)}, I_{(x_2,y_1)}, \ldots, I_{(x_1,y_2)}, \ldots, I_{(x_n,y_n)}\}$$

If we use overlapping blocks, an averaging operation is performed at the end to estimate depth at each pixel by counting the number of blocks contributing to that pixel.

3. METHOD

We used a sample of 104 pairs of face and depth images provided by the Laboratory for Image and Video Engineering (LIVE) at the University of Texas at Austin, Austin, TX [15]. All images corresponded to a neutral expression frontal view and were aligned by an ordinary procrustes analysis using five landmarks. We selected 65 pairs of images as the training set and the remaining 39 as the test set.

The MLR method of depth prediction from intensity was applied for three different kinds of local features. For DCT we used an 8x8 block size. For local PCA an 8x8 block size and a previously built common eigenvectors base was also used, from all the horizontal and vertical overlapping blocks from 2 images (15 conditions with a total of 3984 blocks). For modular PCA two 32x32 size block grids were used to build a specific eigenvectors base for every block. As shown in Figure 1, both block grids capture important facial features in an overlapping manner. An initial training phase for every condition was performed using the originally illuminated intensity images to build the specific W transform matrixes for every horizontal and vertical overlapping blocks in the intensity and depth images. The first coefficient in the DCT condition was discarded as it captures every block average intensity, and was substituted for every sample average block depth in predictions. We subsequently performed experiments in order to ascertain the effect of different overlap degree, the dimensionality or number of coefficients used in the MLR, and two specific illumination variations, point source horizontal angle and ambient illumination proportion, over the reconstructed depth map fractional error percent:

$$\left(\sum \frac{|Z(x,y) - Z'(x,y)|}{Z(x,y)}\right) \cdot 100$$

Where $Z(x,y)$ represents the true depth at each point, and $Z'(x,y)$ represents the predicted depth at each point.

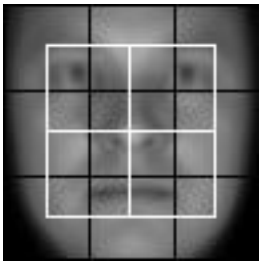


FIGURE 1: The two 32x32 block grids used in the modular PCA condition superimposed to the average intensity image: (a) beginning at (0,0) in black, and (b) beginning at (16,16) in white. Both block grids capture important facial features.

4. NUMBER OF FEATURES AND OVERLAP DEGREE

Figure 2 shows that for all predictors, fractional error increases when increasing the number of coefficients, while it decreases with the degree of overlap between blocks. In the two-way ANOVAs between number of coefficients and overlap degree, both factors proved to be significant for DCT ([$F(2,76) = 96.3$, $p < .001$, $\eta_p^2=.72$] , [$F(2,76 = 99.99$, $p < .001$, $\eta_p^2=.73$]), local PCA ([$F(2,76) = 338.6$, $p < .001$, $\eta_p^2=.9$] , [$F(2,76 = 118.63$, $p < .001$, $\eta_p^2=.76$]) and for number of coefficients in modular PCA[$F(2,76) = 19.83$, $p < .001$, $\eta_p^2=.34$], with assumed sphericity in all cases. When comparing fractional error between different features for a number of 16 coefficients and an overlap of 87.5%, results yield an advantage of local and modular PCAs over DCT, with respective marginal means of 9.64, 9.67 and 10.62%. Figure 3 shows that the difference error under original illumination conditions tends to be lower for modular PCA than for DCT and local PCA.

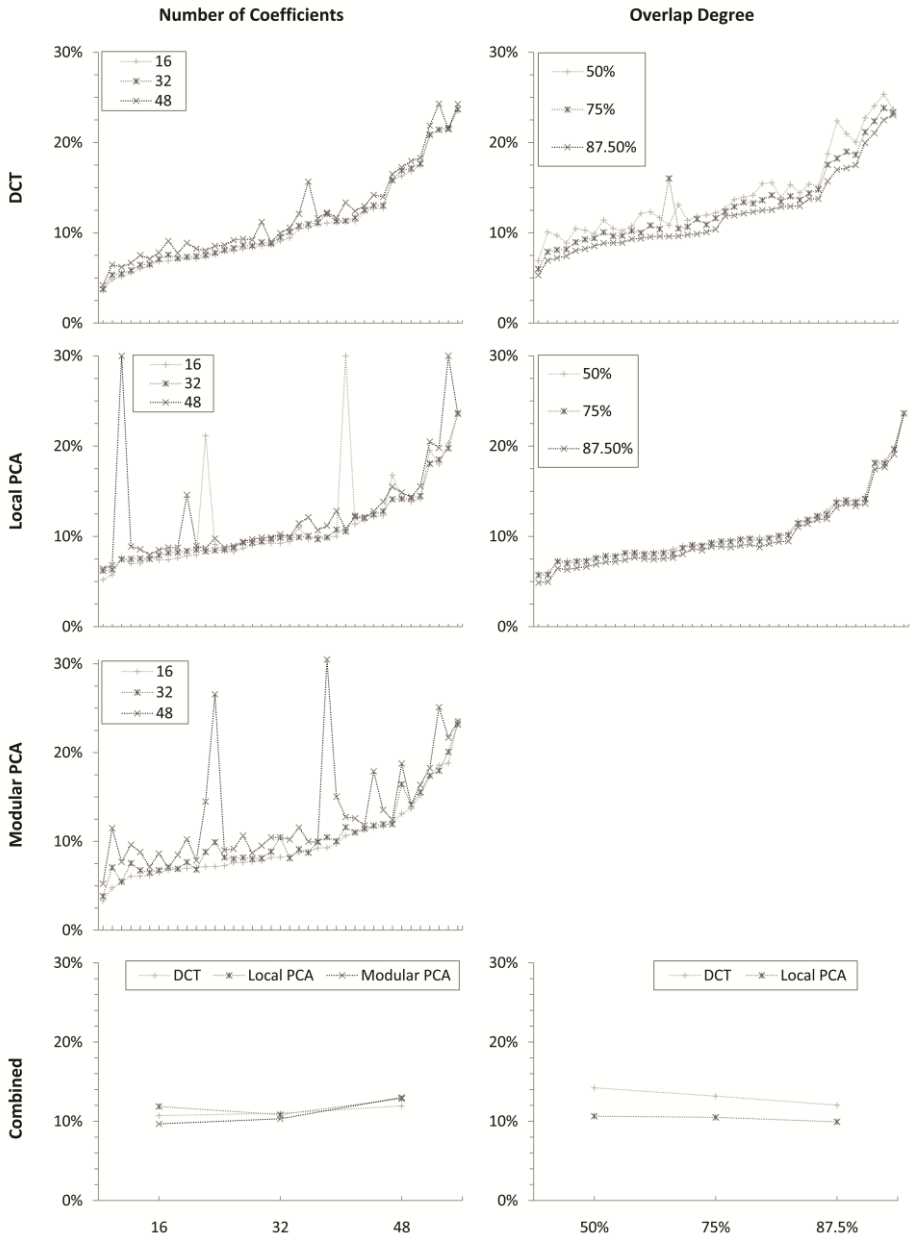


FIGURE 2: Fractional error by number of coefficients and overlap degree for DCT, local PCA and modular PCA conditions. The first three rows show the individual fractional errors. The last row shows combined marginal means for DCT, local PCA and modular PCA conditions.

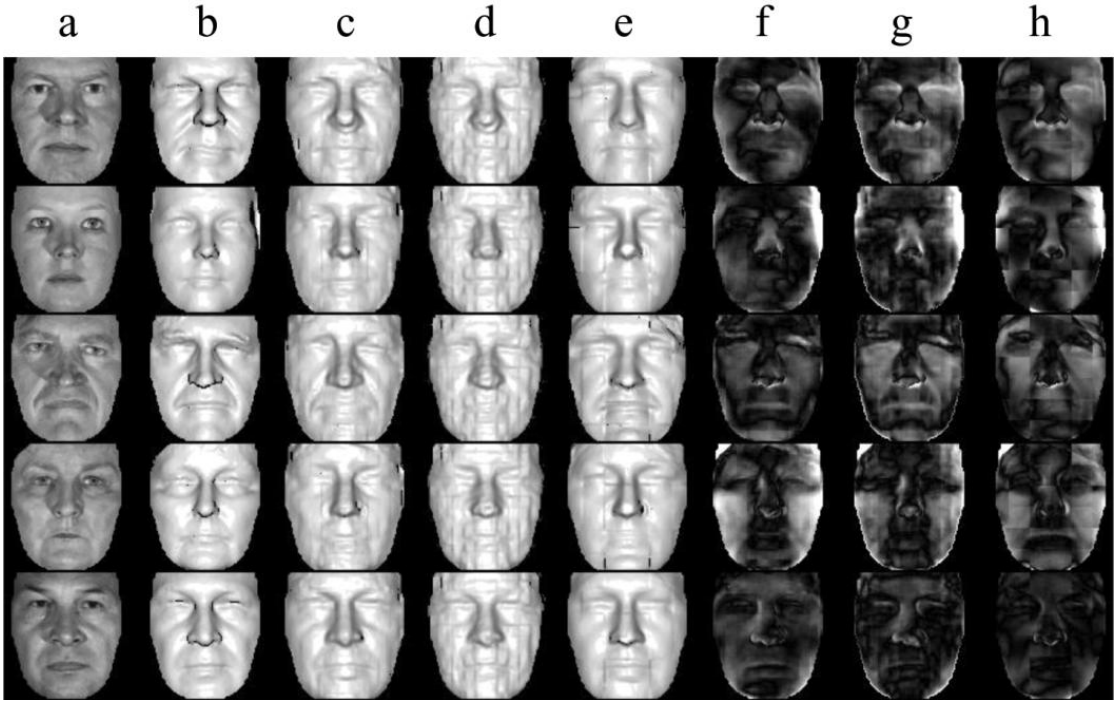


FIGURE 3: Predicted depths as frontal re-illumination for several out-of-training images using different local predictors. For every row: a) the input image, b) true depth, c) predicted depth for DCT, d) local PCA, e) modular PCA, f) error or difference between prediction and true depth for DCT, g) local PCA and h) modular PCA. Lighter values indicate greater error.

5. ILLUMINATION VARIATIONS

We conceptualized image illumination as a four spherical harmonic approximation [16] and rendered new intensity images by varying the light source angle over the horizontal plane from 0º (frontal position, lighting vector [0,0,1]) to 45º (to the left) at five levels at regular intervals and the proportion of ambient illumination over the total illumination from 0.2 to 0.9, at five levels at regular intervals, and a level 0 condition. The effect of illumination variations over the logarithm of coefficients variances and prediction fractional error was studied.

Figure 4 shows that most of between-illumination variance is concentrated in the first coefficient for DCT [8]. In the rest of the conditions the between-illumination variance is shared between the different coefficients. It's noted that between-illumination variance is almost negligible in the case of local PCA, while in the first block grid from modular PCA (a), which includes the blocks of the edge of the face, coefficients variance is greater than in the second grid (b). Between-subjects variances were greater than between-illumination for all conditions when performing a t-test. For DCT [$t(64) = 4.67$, $p < .001$, $d' = 0.82$, $r = .38$], local PCA [$t(64) = 4.68$, $p < .001$, $d' = 0.83$, $r = .38$], modular PCA grid a [$t(64) = 41.91$, $p < .001$, $d' = 6.26$, $r = .95$] and modular PCA grid b [$t(64) = 50.71$, $p < .001$, $d' = 7.98$, $r= .97$].

Figure 5 shows that the effect of horizontal light source angle and proportion of ambient illumination is small but consistent for the DCT condition ([$F(4,152) = 25.1$, $p < .001$, $\eta_p^2=.49$], [$F(5,190) = 12.18$, $p < .001$, $\eta_p^2= .24$]) and local PCA condition([$F(4,152) = 92.6$, $p < .001$, $\eta_p^2= .71$], [$F(5,190) = 18.15$, $p < .001$, $\eta_p^2= .32$]). But it is more noticeable in the case of modular PCA ([$F(4,152) = 15.97$, $p < .001$, $\eta_p^2= .28$], [$F(5,190) = 10.33$, $p < .001$, $\eta_p^2= .21$]). Specifically, increasing the horizontal light source angle is accompanied by an increase in the fractional error in modular PCA, while increasing the proportion of ambient illumination results in a decrease of fractional error as shown in the combined averages row of Figure 5.

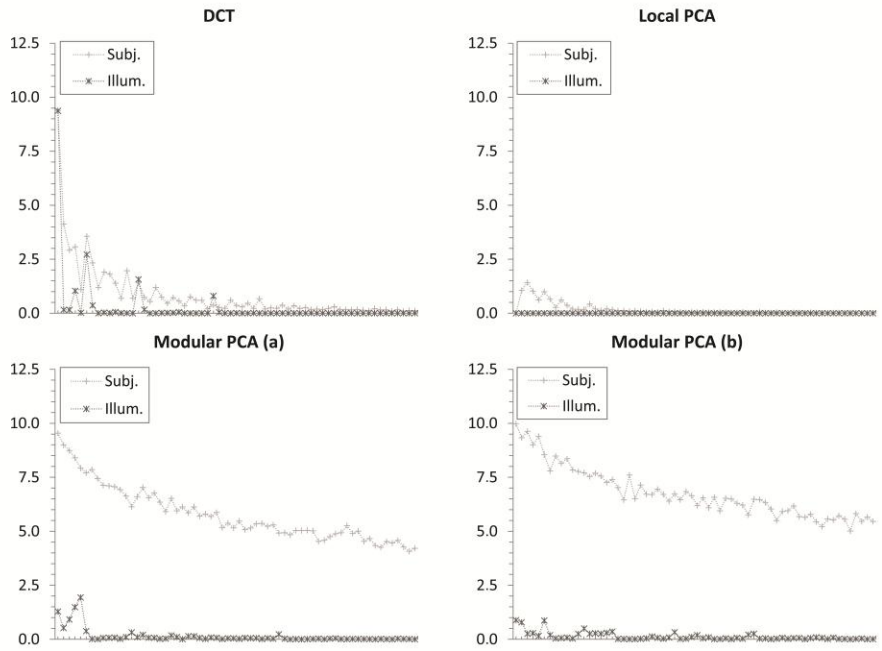


FIGURE 4: Logarithm of between-subjects and between-illumination variances for DCT, local PCA and modular PCA conditions. For modular PCA results for grid starting at (0,0) (a) and starting at (16,16) (b) are shown.

6. EXTERNAL IMAGE TEST

Finally, we performed an experiment with 10 real world images from the Yale B database. In Figure 6, we show the input image of each subject, with a frontal and side view of the modular PCA predicted depth map as frontal re-illumination. The results using modular PCA are visually acceptable and resemble some of the subjects' distinctive features in the frontal and also in the side view.

7. CONCLUSIONS AND FUTURE WORK

We applied a new method that allows reconstruction of 3D face shape from local features prediction using a MLR model with acceptable results for out-of-training and external images. The best results are obtained by using the eigenfeatures coefficients from modular PCA in images frontally illuminated in the same way as in the training phase, but it deteriorates as the illumination angle departs from the frontal position. The presence of ambient illumination compensates for this effect. Local PCA and DCT with the first coefficient removed produces poorer depth predictions that are robust against illumination variations. The proposed method is in contrast with state-of-the-art methods to the extent that it is local. It applies a local regression method to face features instead of doing it to the whole face globally. Global methods try to extract regularities between the whole face intensity and depth. We feel that a local approach may improve their results by taking into account variation modes between intensity and depth that are specific to certain facial features but not to the whole face as a unit.

A drawback of every method that uses aligned face images is that no alignment method produces a perfect match between facial landmarks. In future research, we'll try to improve the results by improving the images alignment by morphing every image to the average shape of a set of manually located features points [17] and by using more elaborate regression models, as partial least squares [18] and intensity-depth coupled models [3]. This will allow us to yield a quantitative comparison to state-of-the-art global methods.

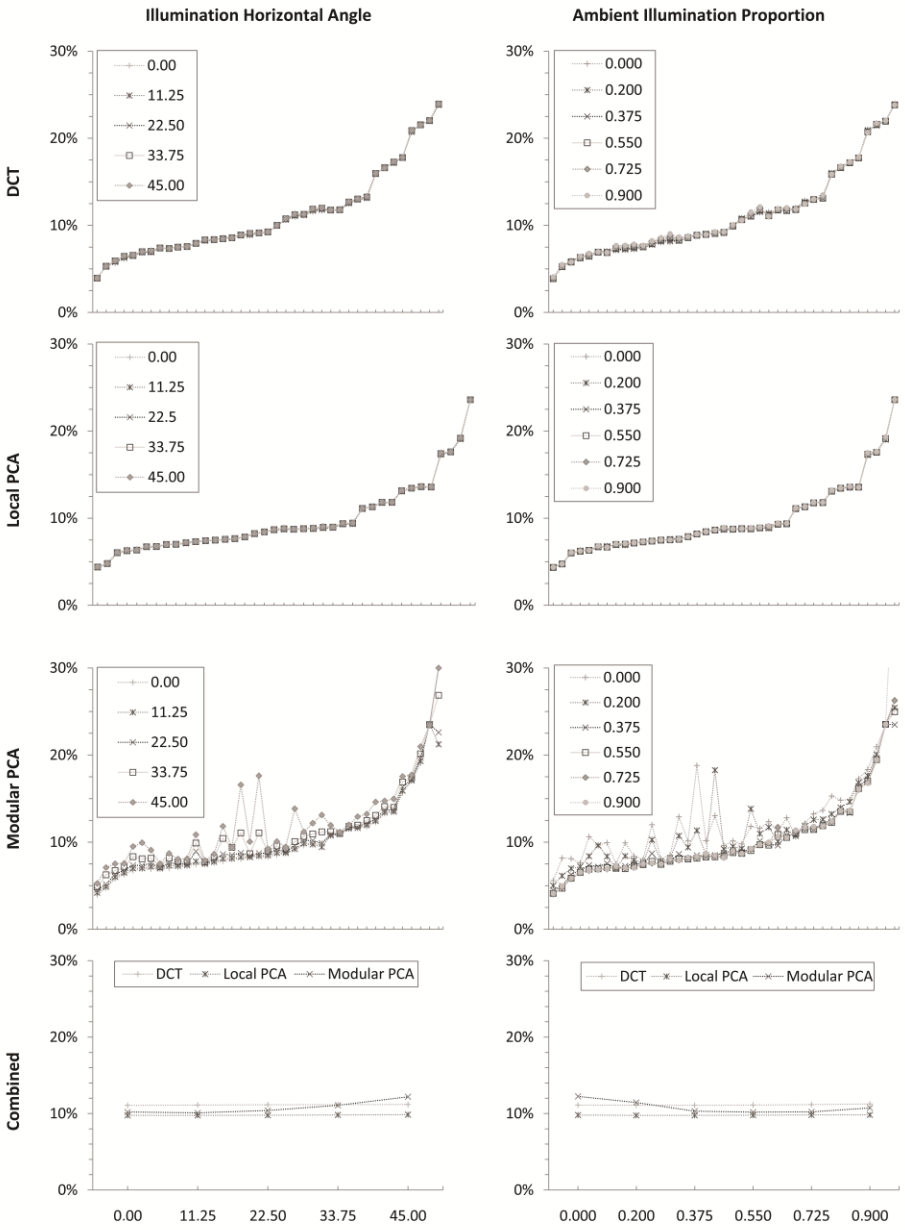


FIGURE 5: Fractional error by illumination horizontal angle and ambient illumination proportion for DCT, local PCA and modular PCA conditions. The first three rows show the individual fractional errors. The last row shows combined marginal means for DCT, local PCA and modular PCA conditions.

FIGURE 6: Predicted depths as frontal re-illumination for external images from the Yale B database using the modular PCA (eigenfeatures) predictor with 16 coefficients and a 87.5% overlap degree. Both grids starting at (0,0) and starting at (16,16) were used and their results were averaged. For every row in each panel the input image and two reconstructed views are shown.

8. REFERENCES

[1] J. J. Atick, P. A. Griffin and A. N Redlich. "Statistical approach to shape from shading: Reconstruction of three-dimensional face surfaces from single two-dimensional images." *Neural Computation*, vol. 8, no 6, pp. 1321-1340, 1996.

[2] V. Blanz and T. Vetter. "A morphable model for the synthesis of 3D faces." *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, 1999, pp. 187-194.

[3] M. Castelán, W. A. Smith and E. R. Hancock. "A coupled statistical model for face shape recovery from brightness images." *IEEE Transactions on Image Processing*, 2007, vol. 16, no 4, pp. 1139-1151.

[4] M. Turk and A. Pentland. "Eigenfaces for recognition." *Journal of cognitive neuroscience*, vol. 3, no 1, pp. 71-86, 1991.

[5] R. Gottumukkal, and V. K. Asari. "An improved face recognition technique based on modular PCA approach." *Pattern Recognition Letters*, vol. 25, no 4, pp. 429-436, 2004.

[6] P. Quintiliano, A. N. Santa-Rosa and R. Guadagnin. "Face recognition based on eigenfeatures." *Multispectral Image Processing and Pattern Recognition*, 2001, pp. 140-145.

[7] Z. M. Hafed and M.D. Levine. "Face recognition using the discrete cosine transform." *International Journal of Computer Vision*, vol. 43, no 3, pp. 167-188, 2001.

[8] W. Chen, M. J. Er, and S. Wu. "Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain." *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2006, vol. 36, no2, pp. 458-466.

[9] V. P. Vishwakarma, S. Pandey and M. N. Gupta. "Adaptive histogram equalization and logarithm transform with rescaled low frequency DCT coefficients for illumination normalization." *International Journal of Recent Trends in Engineering*, vol. 1, no 1, pp. 318-322, 2009.

[10] A. R. Chadha, P. P. Vaidya, and M. M. Roja. "Face recognition using discrete cosine transform for global and local features." *International Conference on Recent Advancements in Electrical, Electronics and Control Engineering*, 2011, pp. 502-505.

[11] C. Sanderson, M. Saban and Y. Gao. "On local features for GMM based face verification." *International Conference on Information Technology and Applications*, 2005, vol. 1, pp. 650-655.

[12] C. Sanderson, S. Bengio and Y. Gao. "On transforming statistical models for non-frontal face verification." *Pattern Recognition*, vol. 39, no 2, pp. 288-302, 2006.

[13] Y. Wong, C. Sanderson and B.C. Lovell. "Regression based non-frontal face synthesis for improved identity verification." *Computer Analysis of Images and Patterns*, Springer Berlin Heidelberg, 2009, pp. 116-124.

[14] Chai, X., Shan, S., Chen, X., & Gao, W. "Local linear regression for pose invariant face recognition." *International Conference on Automatic Face and Gesture Recognition*, 2006, pp. 631-636.

[15] S. Gupta, K. R. Castleman, M. K. Markey, and A. C. Bovik. "Texas 3D face recognition database." *IEEE Southwest Symposium on Image Analysis & Interpretation*, 2010, pp. 97-100.

[16] R. Basri, and D. W. Jacobs. "Lambertian reflectance and linear subspaces." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, vol. 25, no 2, pp. 218-233.

[17] I. Craw and P. Cameron. "Parameterising images for recognition and reconstruction." *British Machine Vision Conference*, Springer London, 1991, pp. 367-370.

[18] M. Castelán and J. Van Horebeek. "3D face shape approximation from intensities using Partial Least Squares." *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2008, pp. 1-8.

A Comparison of Accuracy Measures for Remote Sensing Image Classification: Case Study In An Amazonian Region Using Support Vector Machine

G. B. Scofield *graziela.scofield@cemaden.gov.br*
Center for Natural Disaster Monitoring and Alert - CEMADEN
São José dos Campos, São Paulo, 12630-000, Brazil

E. Pantaleão *epantaleao@ufu.br*
Faculty of Computer, Federal University of Uberlândia - UFU
Patos de Minas, Minas Gerais, 38700-126, Brazil

R. G. Negri *rogerio.negri@ict.unesp.br*
Environmental Engineering Departament,
São Paulo State University - UNESP
São José dos Campos, São Paulo, 12247-004, Brazil

Abstract

This work investigated the consistency of both the category-level and the map-level accuracy measures for different scenarios and features using Support Vector Machine. It was verified that the classification scenario and the features adopted have not influenced the accuracy measure consistency and all accuracy measures are highly positively correlated.

Keywords: Image Classification, Accuracy Measures, Category-level, Map-level, Comparison.

1. INTRODUCTION

In geoscience, Pattern Recognition methods have been shown useful for detecting targets in images obtained by satellites for many different purposes. Monitoring natural resources such as forests, rivers and glaciers [1; 2; 3], study areas affected by natural disasters [4] and urban planning and road monitoring for governmental purposes [5] are some examples. The Pattern Recognition techniques used to identify targets on images are called Image Classification.

Several accuracy measures are developed to quantify category-level and map-level accuracy measures based on the *confusion matrix*, which is computed from the Image Classification results. Each accuracy measure may be more relevant than others for a particular objective, since different measures incorporate different information about the confusion matrix.

In [6], the consistency of both category-level and map-level measures was investigated through the statistical correlation analysis of the confusion matrices using the probability of concordance. These confusion matrices were collected from previous studies presented in literature.

The objective of this work is to investigate the consistency of category-level and map-level accuracy measures, for image classification results with different numbers of classes. The image classification process was done using the Support Vector Machine (SVM) method and LISS-3 multi-spectral optical images, onboard of the Indian Remote Sensing Satellite (IRS).

2. ACCURACY MEASURES

Confusion matrix is one of the most effective ways to represent the Image Classification accuracy, being able to describe inclusion and exclusion errors. A generic error matrix is represented in Table 1, where for a given class ω_i, p_{ij} is the number of pixels from ω_i sample that

was classified as ω_j, p_{i+} and p_{+i} denotes the marginal producer and user, respectively, to a given class ω_i, and m is the total of validation pixels.

		User				
		ω_1	ω_2	$\cdots$	ω_k	
Producer	ω_1	p_{11}	p_{12}	$\cdots$	p_{1k}	p_{1+}
	ω_2	p_{21}	p_2	$\cdots$	p_{2k}	p_{2+}
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	ω_k	p_{k1}	p_{k2}	$\cdots$	p_{kk}	p_{k+}
		p_{+1}	p_{+2}	$\cdots$	p_{+k}	m

TABLE 1: A Generic Error Matrix and Its Elements.

Several measures of thematic map accuracy were developed using the confusion matrix [7; 8; 9; 10]. These measures are organized into two groups: category-level and map-level accuracy measures. The category and map-level measures shown in Table 2 were analyzed.

	Measure name	Formula	Range
Category-level	User accuracy [11]	$ua_i = p_{ii}/p_{i+}$	[0,1]
	Producer accuracy [11]	$pa_i = p_{ii}/p_{+i}$	[0,1]
	User conditional kappa [9]	$uck_i = (ua_i - p_{+i})/(1 - p_{+i})$	$]-\infty, 0]$
	Producer conditional kappa [9]	$pck_i = (pa_i - p_{i+})/(1 - p_{i+})$	$]-\infty, 0]$
	Modified user conditional kappa [10]	$muck_i = (ua_i - \frac{1}{m})/(1 - \frac{1}{m})$	[−1,1]
	Modified producer conditional kappa [10]	$mpck_i = (pa_i - \frac{1}{m})/(1 - \frac{1}{m})$	[−1,1]
Map-level	Overall accuracy [8]	$oa = \sum_{i=1}^{m} p_{ii}$	[0,1]
	Average accuracy user perspective [11]	$aau = \frac{1}{m}\sum_{i=1}^{m} ua_i$	[0,1]
	Average accuracy producer perspective [11]	$aap = \frac{1}{m}\sum_{i=1}^{m} pa_i$	[0,1]
	Double average user and producer accuracy [12]	$daup = \frac{aau - aap}{2}$	[0,1]
	Average of Hellden mean accuracy index [13]	$amah = \frac{1}{m}\sum_{i=1}^{m} \frac{2p_{ii}}{p_{i+} + p_{+i}}$	[0,1]
	Average of Short mapping accuracy index [13]	$ams = \frac{1}{m}\sum_{i=1}^{m} \frac{p_{ii}}{p_{i+} + p_{+i} - p_{ii}}$	[0,1]
	Combining accuracy from both user and producer perspectives [11]	$caup = \frac{aau - amah}{2}$	[0,1]
	Kappa [7]	$kap = \frac{oa - \sum_{i=1}^{m}(p_{i+} + p_{+i})}{oa - \sum_{i=1}^{m}(p_{i+} + p_{+i})}$	$]-\infty, 1]$
	Modified kappa [12]	$mkap = (oa - \frac{1}{m})/(1 - \frac{1}{m})$	$]-\infty, 1]$
	Average mutual information [14]	$ami = \frac{1}{m}\sum_{i,j=1}^{m} p_{ij}\frac{p_{ii}}{p_{i+} + p_{+i} - p_{ii}}$	$[-\infty, 1[$

TABLE 2: Category-level and Map-level Accuracy Measures [6].

The overall accuracy is the sum of the total correct pixels divided by the total number of pixels, since the accuracies of individual classes are calculated by the producer accuracy. The producer accuracy gives the probability of a pixel being correctly classified. On the other hand, the user accuracy indicates the probability that a pixel on the map represents the same class on the ground [11]. In [15], the averages of the user and producer accuracy are proposed.

The Kappa coefficient, introduced by [7], is a widely used accuracy measure for classification assessment. Kappa may be used as a measure of agreement between prediction and reality, or to determine if the values in a confusion matrix represent a significantly better than a randomly obtained one [16]. Several modifications on Kappa coefficient, for category and map-level assessment, have been suggested. The Conditional Kappa [9] and the Modified Conditional Kappa [10] have been presented as the additional means of incorporating an adjustment for hypothetical chance agreement into the category-level accuracy assessment. Among many other map-level measures, the Average Accuracy of the user, the producer, and the combination of both perspectives [11], Average of Short Mapping Accuracy Index [12], Hellden's Mean Accuracy Index [13] and the Average Mutual Information [14] may be mentioned.

In [6], it is conducted a consistent analysis of accuracy measures using Kendall's Tau (K_τ), which is the difference between percentage of concordance (PC) and the percentage of discordance (PD) of two accuracy measures computed on a collection of confusion matrices. Formally, given n confusion matrices ($M_1, M_2, \dots, M_n$) and two accuracy measures (a and b), a series of pairs is determined from these *n* confusion matrices, that is: (a_1, b_1),(a_2, b_2),...,(a_n, b_n). For two specific confusion matrices M_i and M_j, if either $a_i > a_j$ and $b_i > b_j$ or $a_i < a_j$ and $b_i < b_j$, there is a *concordant pair*; if either $a_i > a_j$ and $b_i < b_j$ or $a_i < a_j$ and $b_i > b_j$, there is a *discordant pair*. For n confusion matrices, $T = n(n-1)/2$, for $1 \leq i, j \leq n$, is the number of pairs of possible matrix combinations. Denoting C and D as the number of concordance and discordance pairs, PC and PD are the concordance and discordance percentage, written respectively, as C/T and D/T.

3. METHODOLOGY

The adopted methodology consists of four steps: definition of a classification problem; attributes extraction from an adopted study image; classification of the selected image; calculation and comparison of accuracy measures. The image and the adopted study area are shown in Subsection 3.1, Subsection 3.2 presents the adopted image classification method and the attributes extracted from the study image to process the classifications are mentioned in Subsection 3.3. For each classification result, the different accuracy measures presented in Table 2 are computed with basis on validation samples (Table 3). It is worth mentioning that the aim of this study is to investigate the assessment measures previously discussed and not the obtained classification results.

It is worth mention that in this study, different from [6] where confusion matrices were observed from previously studies through diverse images and classification methods, the confusion matrices were obtained using a particular image classification method, various features extracted from a selected image and considering distinct classes (classification problems).

3.1 Data and Study Area

The study area corresponds to a region near to Tapajós National Forest (FLONA), Brazil. The FLONA is a federal conservation unit having approximately 544 thousand hectares. This unit is bounded by the urban area of Belterra at North, by the BR-163 highway at East, by the Tapajós River at West and by the Tinga and Cupuari rivers at South. It is also characterized by wood extraction that occurred over the past two decades, provoking significant formation of secondary forest areas, concentrated mainly along the BR-163 highway. Additionally, this area has received governmental incentive for mechanized grain production, aiming to export in the first decade of 2000 [17].

Examples of land cover samples collected on the study area are shown in Figure 1. These samples were verified in a field work performed during September 2009. Fortunately, the IRS LISS-3 multispectral sensor (23.5 meters of resolution) was able to acquire images over the FLONA area with low cloud coverage in September 10th, 2009. The collected samples were used to train the classification method and to validate the results. For each class, two sets of samples were randomly defined, where the first one is responsible for training and another for validation. The training set is approximately twice larger than the validation one. Table 3 summarizes the quantity of pixels adopted to train the classification method and to validate its results.

Based on the field work information, a hierarchical class tree was organized, as shown in Figure 2. This arrangement allows to define four scenarios, that is, four ways to partitionate the set of classes. The first scenario is the most specific, since it uses all the tree leaves. The complete tree presents 15 classes, which are: cloud, cloud shadow, dirty pasture, clean pasture, agriculture with high grass, abandoned agriculture, primary forest, secondary forest with 1 to 5 years, secondary forest with 6 to 12 years, secondary forest with 13 to 30 years, urban area, soil with rest of agriculture, prepared soil for agricultural use, bare soil and water (river). Superclasses were used to generate the second scenario, which are pasture, agriculture use and secondary forest, together with cloud, cloud shadow, bare soil, primary forest, urban area, agriculture with high grass and water, completing 10 classes. The term “superclass” determines that a class is composed by the union of similar classes. The third scenario is composed by 8 superclasses which are cloud, cloud shadow, urban area, water, soil, pasture, agriculture and forest. The fourth scenario gathers superclasses cloud, cloud shadow, urban area, water, soil and vegetation.

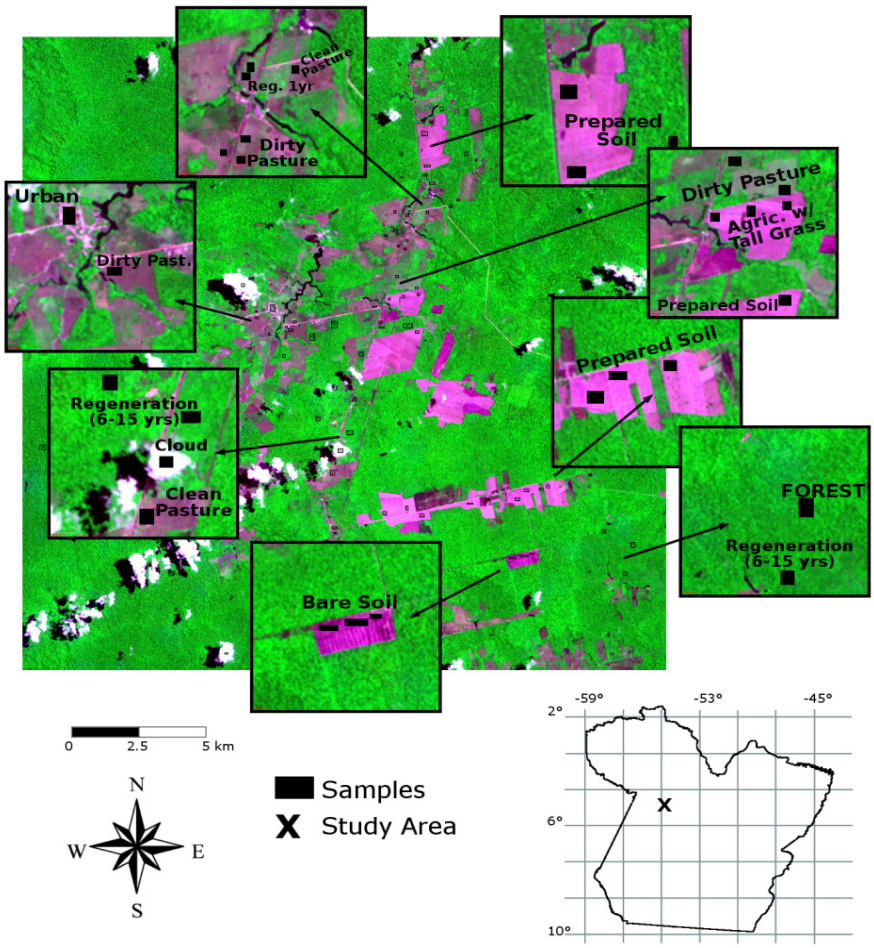


FIGURE 1: Study Area and Land Cover Samples.

Classes	Ground Truth Pixels	
	Training	Validation
Cloud	205	96
Cloud Shadow	117	53
Dirty Pasture	246	126
Clean Pasture	397	185
Agriculture with High Grass	34	14
Abandoned Agriculture	60	28
Primary Forest	115	61
Secondary Forest (1 to 5 years)	80	37
Secondary Forest (6 to 12 years)	191	99
Secondary Forest (13 to 30 years)	23	9
Urban Area	81	40
Soil with Rest of Agriculture	59	30
Prepared Soil for Agriculture	37	16
Bare Soil	26	12
Water (River)	28	12

TABLE 3: A summary about the quantity of ground truth pixels used to train the classification method and to validate the classification results.

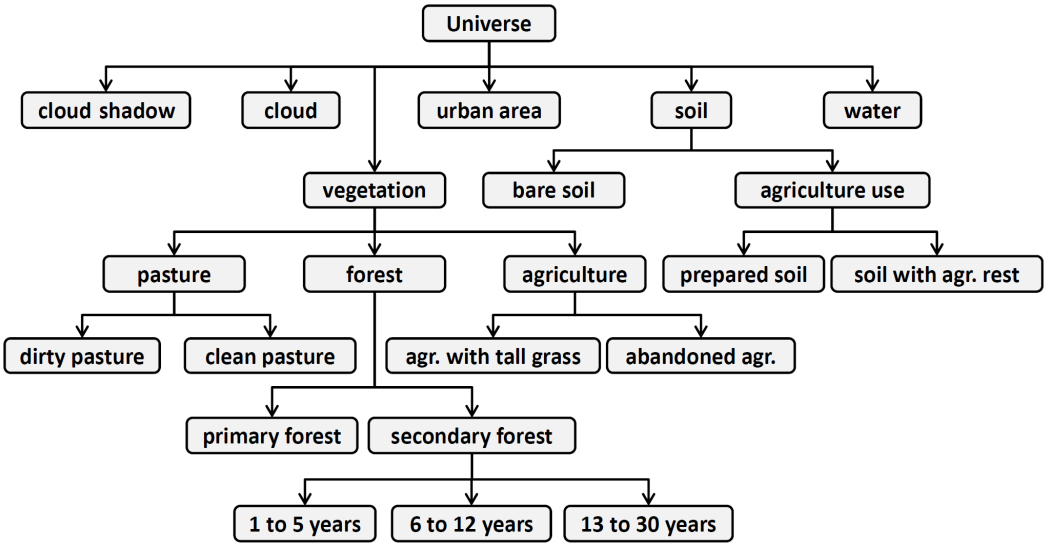


FIGURE 2: Hierarchical Class Tree Used In This Study.

3.2 Support Vector Machine

The Support Vector Machine (SVM) method was adopted in this study. SVM is a recent technique which has received great attention in recent years due to its excellent generalization ability, data distribution independence and its robustness on the Hughes phenomena.

This method consists of finding a separation hyperplane between the training samples with the larger margins. The separating hyperplane is the geometric place where the following linear function is zero:

$$f(\mathbf{x}) = \langle \mathbf{x}, \mathbf{w} \rangle + b \tag{1}$$

where $\mathbf{w}$ represents the orthogonal vector to the hyperplane $f(\mathbf{x}) = 0$; $b/\|\mathbf{w}\|$ is the distance from the hyperplane to the origin and $\langle\cdot,\cdot\rangle$ denotes the inner product. The parameters of (1) are obtained from the following quadratic optimization problem [18]:

$$\sum_{i=1}^{m}\lambda_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j y_i y_j \langle\varphi(\mathbf{x}_i),\varphi(\mathbf{x}_j)\rangle$$
$$subject\ to: \begin{cases} 0 \le \lambda_i \le C;\ \ i = 1,\dots,m \\ \sum_{i=1}^{m}\lambda_i y_i = 0 \end{cases} \tag{2}$$

where λ_i are Lagrange multipliers, $y_i = \{-1, +1\}$ define the class of $\mathbf{x}_i$, since SVM is a binary classifier, C acts as an upper bound of λ values and $\varphi(\mathbf{x})$ is a function adopted to remap the input vectors into a higher dimensionality space. The inner product $\langle\varphi(\mathbf{x}_i),\varphi(\mathbf{x}_j)\rangle$ is known as Kernel function. A popular example of Kernel is the Radial Basis Function (RBF), expressed by $\langle\varphi(\mathbf{x}_i),\varphi(\mathbf{x}_j)\rangle = \exp\left(\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma^2\right)$, $\sigma \in \mathbb{R}_+$, which was adopted in this study. The parameters C and σ tunning was performed according to a Grid Search procedure considering as values $C = \{1, 10, 100, 1000\}$ and $\sigma = \{0.5, 1.0, 1.5, 2.0,\ 2.5\}$.

The optimization problem (2) is solved considering a training set $\mathcal{D} = \{(\mathbf{x}_i, y_i): i = 1,\dots,l\}$, where $\mathbf{x}_i \in \mathbb{R}^d$. Let $SV = \{\mathbf{x}_i :\ \lambda_i \neq 0; i = 1,\dots,l\}$, known as support vector set. The parameters $\mathbf{w}$ and b are computed by:

$$\mathbf{w} = \sum_{\mathbf{x}_i \in SV}\lambda_i y_i \varphi(\mathbf{x}_i) \tag{3}$$

$$b = \frac{1}{\#SV}\sum_{\mathbf{x}_i \in SV} y_i + \sum_{i=1}^{l}\sum_{j=1}^{l}\lambda_i\lambda_j y_i y_j \langle\varphi(\mathbf{x}_i),\varphi(\mathbf{x}_j)\rangle \tag{4}$$

To apply this method to a multiclass problem (problems with more than two classes), the adoption of a multiclass strategy is necessary, such as “one-against-all” or “one-against-one”. Here, the “one-against-all” strategy was adopted. For more details about multiclass strategies, please refer to [19]. The ENVI 4.7 software was used to perform the SVM classifications.

3.3 Feature Extraction and Classification

Several features were extracted from the multispectral IRS LISS-3 image using the ENVI software. These features were obtained using the LISS-3 Red (R), Green (G), Blue (B) and Near Infrared (IRNear) spectral bands, the Normalized Difference Vegetation Index (NDVI) [20] and the Haralick’s Texture Features [21] such as mean (M), variance (Var), homogeneity (H), contrast (C), dissimilarity (D), entropy (E), second moment (SM), and correlation (CR).

The Principal Component Analysis (PCA) [22] was applied to extract features optimally uncorrelated, and the Minimum Noise Fraction technique (MNF) [23] was applied to reduce the data dimensionality and isolate the noise. Combinations of these features were used on SVM to classify each set of classes mentioned in the Subsection 3.1. Finally, the consistency of the accuracy measures were computed and then their consistency were investigated.

The next step was to investigate the consistency of category-level and map-level accuracy measures through statistical correlation analysis. Figure 3 shows the activities flow chart.

Initially, the SVM was applied on the Red, Green, Blue and Near Infrared (R+G+B+IRNear) bands. The NDVI information was added to Red, Green, Blue and Near Infrared bands (NDVI+R+G+B+IRNear) for the second classification.

The PCA was applied in the Red, Green, Blue, Near Infrared and NDVI features and the first three principal components (PCA 1, 2 and 3 respectively) were used on the third classification. The forth classification uses all features together, that is, the spectral, NDVI and texture (denoted by NDVI+R+G+B+IRNear+M+Var+H+C+E+D+SM+C). The PCA was applied on the eight texture bands (i.e., M+Var+H+C+D+E+SM+CR) and the first three PCA components were used for classification. The three first PCAs for NDVI+R+G+B+IRnear+E+D+SM were also classified.

The features entropy (E), dissimilarity (D) and second moment (SM) were chosen through visual analysis, since they presented more information in the contrast between different classes. The NDVI+R+IRNear bands are selected as a color composition to generate a HSV color transformation [22]. This HSV transformation was used for classification. Finally, the MNF transform was applied on the NDVI, R, G, B and IRNear features and the first three produced features (denoted by MNF1, 2 and 3) were used for classification.

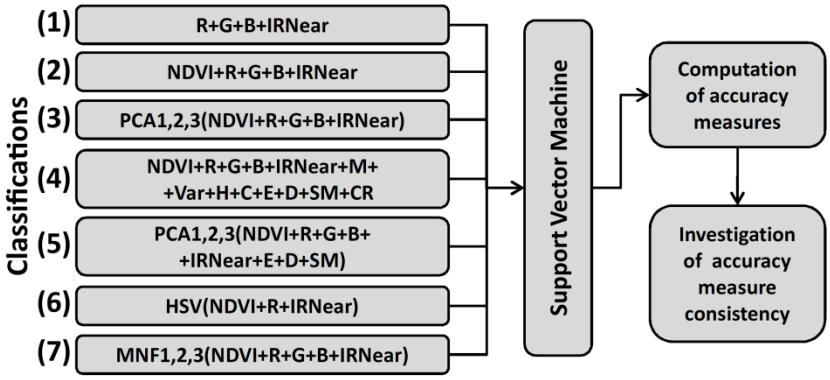


FIGURE 3: Activities Flow Chart.

Each combination of features, illustrated in Figure 3, was used to classify each of the four set of classes. As a result, 28 confusion matrices were obtained. Afterwards, these matrices were used to compute the accuracy measures.

4. RESULTS

For each of the four scenarios defined on Subsection 3.1, SVM classifications using each one of the seven feature combinations determined in Subsection 3.3 were produced. So, 28 classification results and its respective confusion matrices were obtained. From the confusion matrices, the accuracy measures presented in Table 2 were calculated (6 category-level and 10 map-level measures). To summarize a comparison among category-level measures, the PC and PD (percentage of concordance and discordance) were calculated over all classes of each scenario, and then K_τ was calculated. These results are shown in Table 4. Table 5 presents the K_τ between each pair of map-level accuracy measure.

Analyzing the results, the K_τ varied between 60% to 100% on Scenario 1, 34% to 100% on Scenario 2, and 48% to 100% on Scenarios 3 and 4. For the pairs $ua - uck$, $ua - muck$ and $uck - muck$, K_τ were kept equal to 100% in all Scenarios, which means that every user measure may be used equivalently. For the producer measures the same result was observed only for $pa - mpck$. The observed K_τ for the other producer measures was above 82%.

As expected, K_τ was not so different when users and producers measures are compared. Observing the results on different scenarios, most of the category-level measures showed less correlation for the second scenario (10 classes).

As can be seen in the map-level measures (Table 5), ami presents higher percentage of discordance compared with other measures. The ami uses all the elements of confusion matrix,

differently from other measures that use just the class (p_{ii}), producer (p_{i+}) and user (p_{+i}) accuracies, which explains this behavior. It may also observed that oa, cap, kap and $mkap$ are perfectly concordant. Except for ami, the measures present more than 90% of concordance.

The map-level measures, computed from each one of the 28 confusion matrices, are represented on the line graphic of Figure 4. The confusion matrices are organized as follows: 1, 8, 15 and 22 refers for classification 1 of the scenarios 1, 2, 3 and 4, respectively; 2, 9, 16 and 23 for classification 2; 3, 10, 17 and 24 for classification 3; 4, 11, 18 and 25 for classification 4; 5, 12, 19 and 26 for classification 5; 6, 13, 20 and 27 for classification 6; and 7, 14, 21 and 28 for classification 7.

Based on the accuracy levels, it was possible to verify that the analyzed map-level measures have very similar results. The fourth set of classes (22nd to 28th confusion matrices) presented an increase in the accuracy measures in comparison with the other sets. This is explained by the class configuration, since a less specific scenario tends to provide less complexity on classification problems.

Scenario 1	pa	uck	pck	$muck$	$mpck$
ua	.68	1.0	.60	1.0	.68
pa		.68	.82	.68	1.0
uck			.60	1.0	.68
pck				.60	.82
$muck$					.68

Scenario 2	pa	uck	pck	$muck$	$mpck$
ua	.41	1.0	.34	1.0	.41
pa		.41	.84	.41	1.0
uck			.34	1.0	.41
pck				.34	.84
$muck$					.41

Scenario 3	pa	uck	pck	$muck$	$mpck$
ua	.48	1.0	.49	1.0	.48
pa		.48	.94	.48	1.0
uck			.49	1.0	.48
pck				.49	.94
$muck$					.48

Scenario 4	pa	uck	pck	$muck$	$mpck$
ua	.48	1.0	.56	1.0	.48
pa		.48	.92	.48	1.0
uck			.56	1.0	.48
pck				.56	.92
$muck$					.48

TABLE 4: K_τ between each category-level accuracy measure pair: Scenario 1 (15 classes), Scenario 2 (10 classes), Scenario 3 (8 classes) and Scenario 4 (6 classes).

	aau	aap	$daup$	$amah$	ams	$caup$	kap	$mkap$	ami
oa	.889	.778	.836	.815	.825	.915	.968	1.00	-.640
aau		.825	.894	.884	.873	.921	.857	.889	-.593
aap			.931	.942	.931	.862	.767	.778	-.450
$daup$				.979	.968	.921	.804	.836	-.508
$amah$					.989	.899	.794	.815	-.497
ams						.910	.804	.825	-.508
$caup$							.884	.915	-.556
kap								.968	-.619
$mkap$									-.640

TABLE 5: Results of K_τ between each map-level accuracy measure pair.

5. CONCLUSIONS

The objective of this study was to evaluate the consistency of measures used to quantify the accuracy of image classification results. The SVM method was applied in different classification scenarios using different features extracted from the IRS multi-spectral images. Results showed that measures, for category or map-level accuracy, are consistent with every classification scenario and the adopted features. This consistency is expected since all analyzed measures are computed from the same information source, which is the confusion matrix.

User measures are equivalent for category-level assessment, since they present 100% of concordance. However, the producer measures are not equivalent, but show a high level of concordance (at least 82%). In general, map-level accuracies are highly correlated. Lower levels of concordance are verified between the Average Mutual Information measure which every other measures.

Unlike most of the measures based on class accuracy and the marginal user and producer, Average Mutual Information uses all confusion matrix information.

According to the results, the analyzed accuracy measures reflect a common behavior for most situations. None of them can be used to mask or highlight certain results, and they do not lead to divergent conclusions.

As future perspectives, this investigation must be reproduced using others methodologies for scenario definition, instead use an intuitive scenario definition, as performed in this study. Additionally, different features and other image classification methods, more/less robust than SVM, must be considered too.

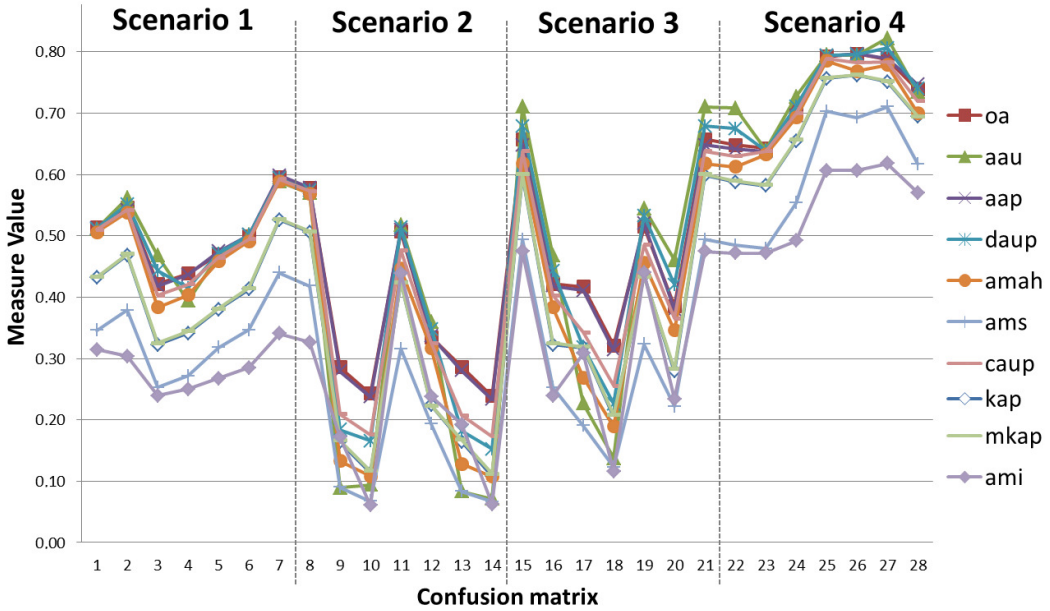


FIGURE 4: Map-level measures computed for each confusion matrix.

6. Acknowledgments

The authors would like to thanks CAPES, CNPQ (3284469/2006-0) and FAPESP (2014/14830-8) for the financial support.

7. REFERENCES

[1] Y. E. Shimabukuro, R. Almeida-Filho, T. M. Kuplich, R. M. Freitas, “Mapping and monitoring land cover in Corumbiara area, BrazilianAmazonia, using JERS-1 SAR multitemporal data”, IEEE International Geoscience and Remote Sensing Symposium, Barcelona, 2007.

[2] A. Niedermeier, S. Lehner, J. Sander, “Monitoring big river estuaries using SAR images”, IEEE International Geoscience and Remote Sensing Symposium, Sydney, 2001.

[3] K. Whitehead, B. Moorman, P. Wainstein, “Determination of variations in glacier surface movements through high resolution interferometry: Bylot Island, Canada”, IEEE International Geoscience and Remote Sensing Symposium, Cape Town, 2009.

[4] A. Chesnel, R. Binet, L.Wald, “Object oriented assessment of damage due to natural disaster using very high resolution images”, IEEE International Geoscience and Remote Sensing Symposium, Barcelona, 2007.

[5] M. Kasanko, V. Sagris, C. Lavalle, J. I. Barredo, L. Petrov, K. Steinnocher, W. Loibl, C. Hoffmann, “GEOLAND spatial planning observatory: How remote sensing data can serve the needs of urban and regional planning”, Urban Remote Sensing Joint Event, Paris, 2007.

[6] C. Liu, P. Fraizer, L. Kuman, “Comparative assessment of the measures of thematic classification accuracy,” Remote Sensing of Environment. vol. 107, pp. 606-616, 2007.

[7] J. A. Cohen, “Coefficient of agreement of nominal scales,” Educational and Psychological Measurement. vol. 20, pp. 37-46, 1960.

[8] M. Story, R. G. Congalton. “Accuracy assessment: a user’s perspective,” Photogrammetric Engineering and Remote Sensing. vol. 52, pp. 397-399, 1986.

[9] G. H. Rosenfield, K. A.Fitzpatrick-Lins, “Coefficient of agreement as a measure of thematic classification accuracy,” Photogrammetric Engineering and Remote Sensing. vol. 52, pp. 223-227, 1986.

[10] S. V. Stehman, “Selecting and interpreting measures of thematic classification accuracy,” Remote Sensing of Environment. vol. 62, pp. 77-89, 1997.

[11] T. Fung, E. Ledrew, “The determination of optimal threshold levels for change detection using various accuracy indices,” Photogrammetric Engineering and Remote Sensing. vol. 54, pp. 1449-1454, 1988.

[12] N. M. Short, “The LANDSAT tutorial workbook: basics of satellite Remote Sensing”. National Aeronautics and Space Administration, Scientific and Technical Information Branch, 1982.

[13] U. A. Hellden, “Test of LANDSAT-2 imagery and digital data for thematic mapping illustrated by an environmental study in northern Kenya”, Lund University Natural Geography Institute Report, 1980.

[14] J. T. Finn, “Use of the average mutual information index in evaluating classification error and consistency,” International Journal of Geographical Information Systems. vol. 7, pp. 349-366, 1993.

[15] R. G. Congalton, “A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data,” Remote Sensing of Environment. vol. 37, pp. 35-46, 1991.

[16] J. R. Jansen, Introductory Digital Image Processing: A Remote Sensing Perspective, Pearson Prentice Hall, Upper Saddle River, 2005.

[17] ZEE BR-163, 2011. Zoneamento ecológico-econômico da rodovia BR-163, Access in December 2011. http://zeebr163.cpatu.embrapa.br.

[18] S. Theodoridis, K. Koutrombas, Pattern Recognition, Academic Press, San Diego, 2006.

[19] A. R. Webb. Statistical Pattern Recognition, Jhon Wiley and Sons, Chichester, 2002.

[20] J. G. P. W. Clevers, “The derivation of a simplified reflectance model for the estimation of leaf area index,” Remote Sensing of Environment. vol. 35, pp. 53-70, 1988.

[21] R. M. Haralick, K. Shanmugam, I. Dinsten, “Texture features for image classification,” IEEE Transactions on Systems, Manchine and Cybernetics. vol. 3, pp. 610-622, 1973.

[22] R. C. Gonzales, R. E. Woods. Digital Image Processing, Prentice Hall, California, 2007.

[23] A. A. Green, M. Berman, P. Switzer, M. D. Craig, “A transformation for ordering multispectral data in terms of image quality with implications for noise removal,” IEEE Transactions on Geoscience and Remote Sensing. vol. 26, pp. 65-74, 1998.

Contourlet Transform Based Method For Medical Image Denoising

Abbas H. Hassin AlAsadi *abbas.hassin@uobasrah.edu.iq*
Faculty of Science /Department of Computer Science
Basra University
Basra, Iraq

Abstract

Noise is an important factor of the medical image quality, because the high noise of medical imaging will not give us the useful information of the medical diagnosis. Basically, medical diagnosis is based on normal or abnormal information provided diagnose conclusion. In this paper, we proposed a denoising algorithm based on Contourlet transform for medical images. Contourlet transform is an extension of the wavelet transform in two dimensions using the multiscale and directional filter banks. The Contourlet transform has the advantages of multiscale and time-frequency-localization properties of wavelets, but also provides a high degree of directionality. For verifying the denoising performance of the Contourlet transform, two kinds of noise are added into our samples; Gaussian noise and speckle noise. Soft thresholding value for the Contourlet coefficients of noisy image is computed. Finally, the experimental results of proposed algorithm are compared with the results of wavelet transform. We found that the proposed algorithm has achieved acceptable results compared with those achieved by wavelet transform.

Keywords: Medical Image, Denoising, Wavelet Transform, Contourlet Transform.

1. INTRODUCTION

In the past decades, several non-invasive new imaging techniques have been discovered such as computerized tomography (CT) scan, single-photon emission tomography (SPET), ultrasound, digital radiography, magnetic resonance imaging (MRI), spectroscopy and others. These techniques have provided the physicians with new information about the interior of the human body that has never been available before, but for various reasons, the use of the information is very limited, and requires the use of computer technology, advanced instruments, image processing techniques, such as the elimination of the noise generated during the acquisition or transmission; enhanced contrast image; showing detail image; and so on.

Noise is one of the medical image quality important factors. High noise of medical images may inaccuracy in the diagnosis of diseases, especially cancer diseases. It is well known that the medical diagnostic process is very important in this kind of diseases; it is mainly based normal or abnormal information provided by medical imaging to diagnose conclusion. High quality of medical images is considered the first step in the correct diagnosis, so the need to minimize the impact of noise in this kind of images.

Image Denoising is a central pre-processing step in image processing to eliminate the noise in order to strengthen and recover small details that may be hidden in the data.

The use of signal processing techniques has been recently reported by several researchers with satisfactory results. These approaches take into account the signal and noise properties in different ways.

In spite of the fact that, the Discrete Wavelet Transform (DWT) has been successfully applied for a wide range of image analysis problems. With these preferences in use, but it is recorded two observations [1]: (1) ignoring the smoothness along contours ;(2) providing only limited directional information which is an important feature of multidimensional signals [2].

Partially, these two problems have been solved by the Contourlet Transform (CT) which can efficiently approximate a smooth contour at multiple resolutions. Additionally in the frequency domain, the CT offers a multiscale and directional decomposition, providing anisotropy and directionality, features missing from the DWT [3][4] (see Figure 1). The CT has been practically used in a variety of applications, such as image denoising [5], image classification [6], image compression [7] CBIR [8], etc.

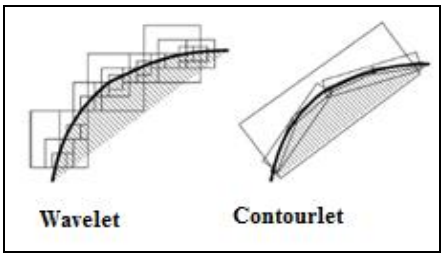


FIGURE 1: A major difference between Wavelet and Contourlet on contour representation.

2. CONTOURLET TRANSFORM BACKGROUND

The Contourlet transform has been developed to overcome the limitations of the wavelets transform [9]. It permits different and elastic number of directions at each scale, while achieving nearly critical sampling.

The Contourlet transform can be worked into two basic steps: Laplacian pyramid decomposition and directional filter banks. Firstly, the Laplacian pyramid (LP) is used to decompose the given image into a number of radial subbands, and the directional filter banks (DFB) decompose each LP detail subband into a number of directional subbands. The band pass images from the LP are fed into a DFB so that directional information can be captured. The scheme can be iterated on the coarse image. Figure 2 shows a schematic diagram of a multilayer decomposition Contourlet.

The combination of the LP and the DFB is a double filter bank named Pyramidal Directional Filter Bank (PDFB), which decomposes images into directional subbands at multiple scales.

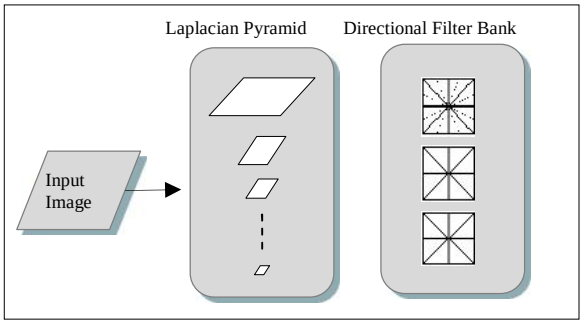


FIGURE 2: A schematic diagram of a multilayer decomposition Contourlet [9].

There are many research works have used CT in different applications, especially in the field of denoising and distortions of the images. Bhateja *et al.* [10] have presented a Contourlet based speckle reduction method for denoising ultrasound images of breast. In [11], authors proposed a novel method for denoising medical ultrasound images, by considering image noise content as combination of speckle noise and Gaussian noise. Fayed *et al.* [12] have presented a method for

extracting the image features using Contourlet Harris detector that is applied for medical image retrieval. Song *et al*. [13] have used scale adaptive threshold for medical ultrasound image, where in the subband Contourlet coefficients of the ultrasound images after logarithmic transform are modeled as generalized Gaussian distribution. Hiremath *et al*. [14] have proposed a method to determine the number of levels of Laplacian pyramidal decomposition, the number of directional decompositions to perform on each pyramidal level and thresholding schemes which yields optimal despeckling of medical ultrasound images, in particular. This method consists of the log transformed original ultrasound image being subjected to Contourlet transform, to obtain Contourlet coefficients. The transformed image is denoised by applying thresholding techniques on individual band pass sub bands using a Bayes shrinkage rule.

3. PROPOSED ALGORITHM

It is known that, the most common technique to remove noise from images is to transform the noisy image from the spatial domain into the frequency domain, such techniques as the Wavelet, Curvelet, and Contourlet transforms, and then compare the transform coefficients with a fixed threshold.

Typically, the low frequencies contain most of the information, which is commonly seen as a peak of data within the time-frequency domain. While, the information at the high frequencies is usually noise. The image can easily be altered within the time-frequency domain to remove the noise. Therefore, our proposed algorithm defines a new threshold value for the Contourlet coefficient to eliminate the unwanted pixels.

The Contourlet transform expression is given by,

$$C_{j,k}^{(l)}(t) = \sum_{i=0}^{3} \sum_{n \in Z^2} d_k^{(l)}(2n - k_i) \left[\sum_{m \in Z^2} f_i(m) \delta_{j-1,2n-m} \right] \quad \cdots(1)$$

Where $C_{j,k}^{(l)}(t)$ represents the Contourlet transform of the image. The $d_k^{(l)}$ and $f_i[m]$ represents the directional filter and the band pass filter in the equation. Thus *j*, *k* and n represent the scale direction and location. Therefore *l* represents the number of directional filter bank decomposition levels at different scales *j*. Thus the output of Contourlet transform is a decomposed image coefficients.

The Laplacian pyramid at each level generates a Low pass output (LL) and a Band pass output (LH, HL, and HH). The Band pass output is then passed into directional filter bank, which results in Contourlet coefficients [15]. The Low pass output is again passed through the Laplacian pyramid [16] to obtain more coefficients and this is done till the fine details of the image are obtained. Figure 3 shows the decomposition of a given image.

3.1 Estimation of Parameters

In this section, we defined some parameters that help to determine the degree of adequacy of the proposed algorithm.

3.1.1 Optimal Threshold

Selecting the optimal threshold is a key problem for the denoising algorithms based on the threshold. The soft threshold method is selected [17]. This method is fit for image denoising based on the CT since the threshold is different for each direction of each scale. It can be described as:

$$Sthr_{j,k} = \pm_{j,k} * \sqrt{\frac{2\log(j+1)}{j}} \quad \cdots(2)$$

$$\pm_{j,k} = \left(\frac{Median\left(C_{i,j}\right)}{0.6745} \right) \quad \cdots(3)$$

Where $Sthr_{j,k}$ is the threshold of k^{th} direction of j^{th} scale; $\pm_{j,k}$ is the standard deviation of the noisy image; $C_{i,j}$ is the Contourlet coefficient of noisy image.

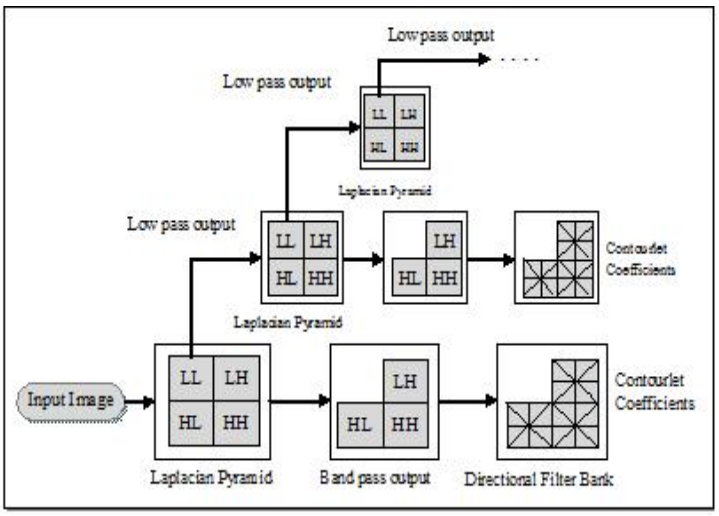


FIGURE 3: Decomposition of Contourlet Transform.

3.1.2 Noise Model

For verifying the denoising performance of the CT, two kinds of noise are added into our sample of the medical images [18]: the first is Gaussian noise; the second is speckle noise.

Gaussian noise is most commonly used as additive white noise. It is Gaussian distribution, which has a bell shaped probability distribution function given by:

$$f(x) = \frac{1}{\sqrt{2®\pm^2}} e^{-(x-«)^2/2\pm^2} \quad \cdots(4)$$

where x represents the gray level, « is the mean of the function, and $\pm^2$ is the standard deviation of the noise.

Speckle noise is a multiplicative noise i.e. it is direct proportion to the local grey level in any area. Speckle noise follows a gamma distribution and is given as:

$$f(x) = \frac{x^{\sim-1}}{(\sim-1)!\pm^{\sim}} e^{-\frac{x}{\pm}} \quad \cdots(5)$$

where x represents the gray level, $\pm^2$ is the standard deviation of the noise, and $\propto$ is the shape parameter of gamma distribution.

3.1.3 Performance Criterions

The parameters which are used in estimation of performance are Signal to Noise Ratio (SNR), Mean Square Error (MSE), and Peak Signal to Noise Ratio (PSNR) [19].

Signal to Noise Ratio compares the level of desired signal to the level of background noise. The higher SNR is the lesser the noise in the image and vice versa:

$$SNR = 10\log\left(\frac{\pm^{2}_{Org}}{\pm^{2}_{Denoised}}\right) \quad \cdots(6)$$

Where, $\pm^{2}_{Org}$ is the variance of the original image and $\pm^{2}_{Denoised}$ is the variance of error between the original and denoised image.

Mean square error is given by :

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[Org(i,j) - Denoised(i,j)\right]^{2} \quad \cdots(7)$$

Where, $Org(i,j)$ is the original image and $Denoised(i,j)$ is the image denoised with some filter and mn is the size of the image.

PSNR gives the ratio between possible power of a signal and the power of corrupting noise present in the image.

$$PSNR = 10\log_{10}\frac{255^{2}}{MSE} \quad \cdots(8)$$

Higher the PSNR gives lower the noise in the image.

3.2 Algorithm Description

The block diagram of proposed algorithm is shown in Figure 4.

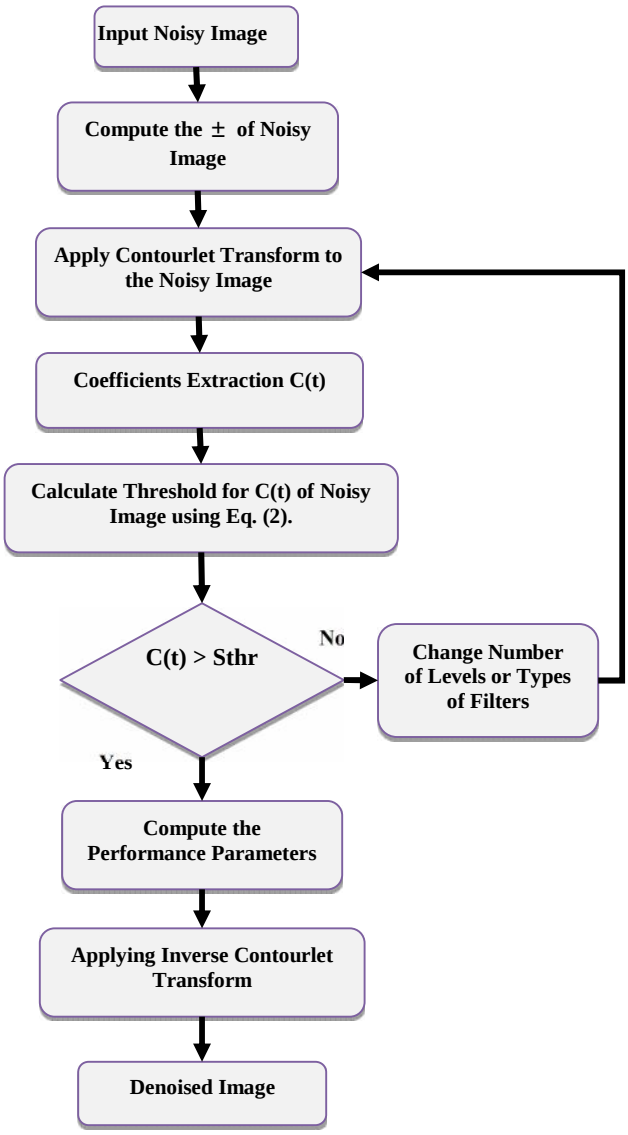


FIGURE 4: Block Diagram of Proposed Algorithm.

4. RESULTS AND DISCUSSION

The proposed algorithm is applied in different medical images datasets, such as MRI, X-ray, CT scan, and ultrasound images. All images have the same size of 512×512 pixel, with 256-level grayscale.

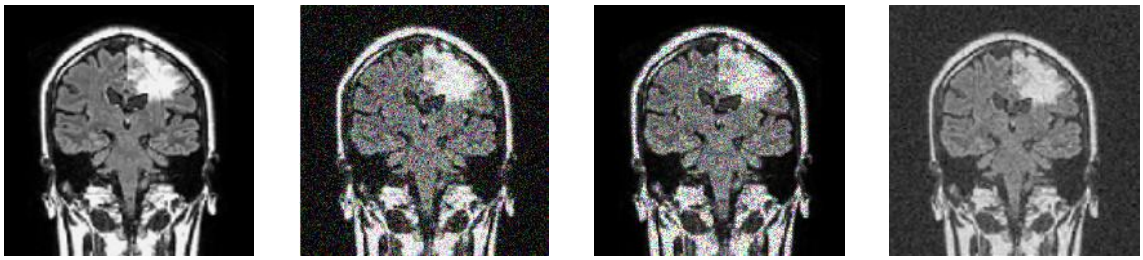
For verifying the performance of the proposed algorithm, two types of noise models are added to these images. One is an additive noise such as Gaussian noise which is given by Eq. (4); the other is a multiplicative noise i.e., speckle noise which is given by Eq. (5).

In the experimental results, the Gaussian noise with mean =0 and variance=0.03 is added to given images, while the speckle noise with noise = 0.1 is also added to the same images. For the LP stage, the 9-7 filter is used to decompose the image into 4 scales; for the DFB stage, direction

is partitioned into 3, 4, 8 and 16 directional subbands from coarse to fine scales respectively. Threshold selection is based on Eq. (2) and Eq. (3).

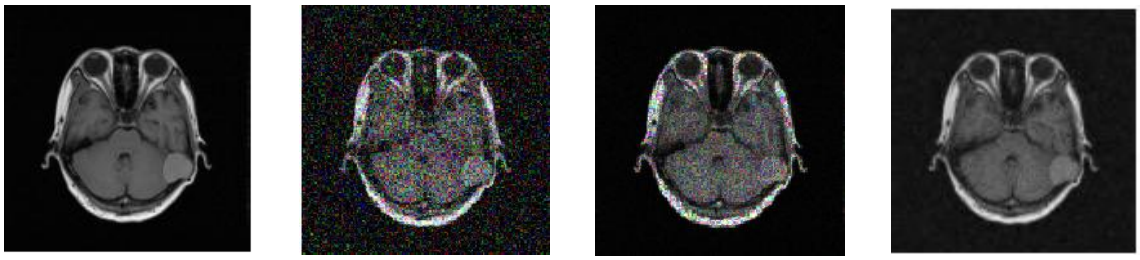
Figure 5 to Figure 9 show the visual results of brain MRI, CT scan, tumor MRI, ultrasound, and x-ray images after applying the proposed algorithm respectively. The performances of the proposed algorithm using PSNRs and SNRs are shown in Table 1 and Table 2 respectively.

As a final point, for more judgments on the proposed algorithm in high noise levels, it is compared with the wavelet methods. The results of the comparison using PSNR and SNR are shown in Table 3 and Table 4 respectively.



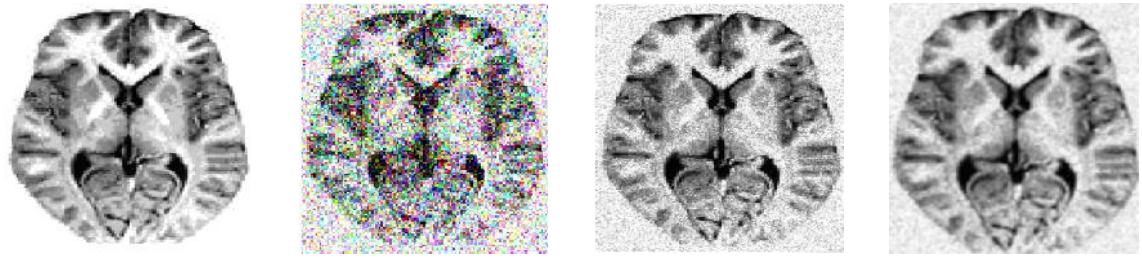
a b c d

FIGURE 5: The visual results of brain MRI image. (a) Orignal image (b) Noisy image by Gaussian noise (c) Noisy image by speckle noise (d) Denoised image using CT.



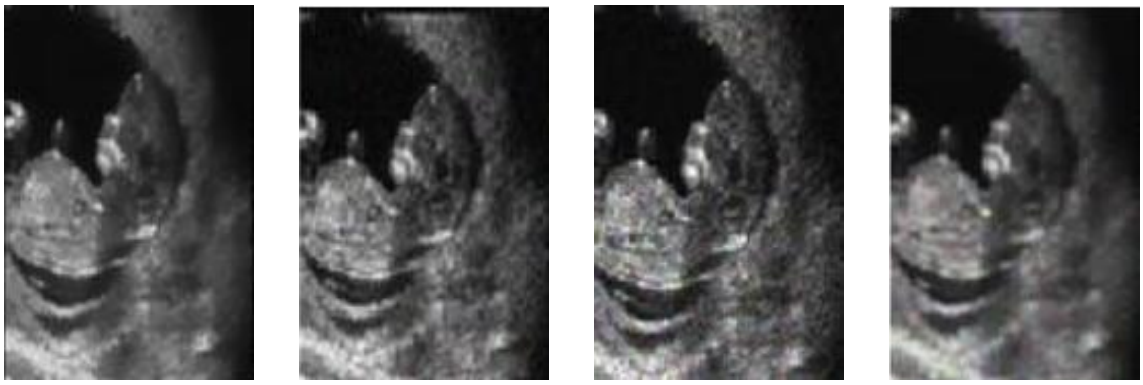
a b c d

FIGURE 6: The visual results of CT scan image. (a) Orignal image (b) Noisy image by Gaussian noise (c) Noisy image by speckle noise (d) Denoised image using CT.



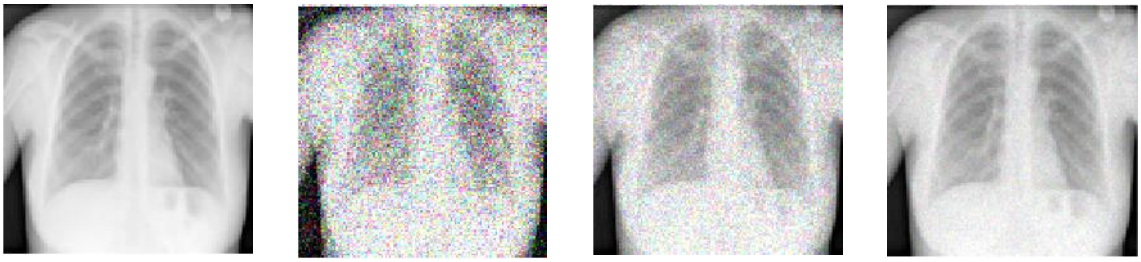
a b c d

FIGURE 7: The visual results of tumor MRI image. (a) Orignal image (b) Noisy image by Gaussian noise (c) Noisy image by speckle noise (d) Denoised image using CT.



a b c d

FIGURE 8: The visual results of ultrasound image. (a) Orignal image (b) Noisy image by Gaussian noise (c) Noisy image by speckle noise (d) Denoised image using CT.



a b c d

FIGURE 9: The visual results of x-ray image. (a) Orignal image (b) Noisy image by Gaussian noise (c) Noisy image by speckle noise (d) Denoised image using CT.

Image	Noisy Image (dB)		Denoised Image by CT (dB)	
	Gaussian	Speckle	Gaussian	Speckle
Brain MRI	16.85	21.27	22.10	28.44
CT scan	14.57	17.19	19.79	26.10
Tumor MRI	7.82	12.36	13.52	18.01
Ultrasound	11.85	15.07	16.60	19.60
X-ray	16.35	18.30	20.66	22.37

TABLE 1: The PSNR value of noised and denoised images.

Image	Noisy Image (dB)		Denoised Image by CT (dB)	
	Gaussian	Speckle	Gaussian	Speckle
Brain MRI	**8.65**	**13.96**	**15.79**	**19.22**
CT scan	**4.36**	**9.85**	**8.19**	**13.27**
MRI Tumor	**2.37**	**13.70**	**6.83**	**18.76**
Ultrasound	**2.26**	**5.83**	**7.37**	**9.52**
X-ray	**14.00**	**10.42**	**18.98**	**16.78**

TABLE 2: The SNR value of noised and denoised images.

Image	Noisy Image (dB)		Denoised Image by WT (dB)		Denoised Image by CT (dB)	
	Gaussian	Speckle	Gaussian	Speckle	Gaussian	Speckle
Brain MRI	**16.85**	**21.27**	**20.78**	**26.31**	**22.10**	**28.44**
CT scan	**14.57**	**17.19**	**18.49**	**23.97**	**19.79**	**26.10**
Tumor MRI	**7.82**	**12.36**	**12.21**	**15.98**	**13.52**	**18.01**
Ultrasound	**11.85**	**15.07**	**15.29**	**17.57**	**16.60**	**19.60**
X-ray	**16.35**	**18.30**	**19.34**	**20.29**	**20.66**	**22.37**

TABLE 3: The PSNR values of the comparison between the proposed algorithm and the wavelet method.

Image	Noisy Image (dB)		Denoised Image by WT (dB)		Denoised Image by CT (dB)	
	Gaussian	Speckle	Gaussian	Speckle	Gaussian	Speckle
Brain MRI	8.65	13.96	14.47	17.09	15.79	19.22
CT scan	4.36	9.85	11.95	11.26	8.19	13.27
Tumor MRI	2.37	13.70	17.45	16.75	6.83	18.76
Ultrasound	2.26	5.83	8.22	7.74	7.37	9.52
X-ray	14.00	10.42	15.46	14.65	18.98	16.78

TABLE 4: The SNR values of the comparison between the proposed algorithm and the wavelet method.

5. CONCLUSION

High quality of medical images is considered the first step in the correct diagnosis, so the need to minimize the impact of noise in this kind of images. In this paper, an algorithm of medical image denoising based on Contourlet transform is proposed. The Contourlet transform is chosen because it is suitable for processing two-dimensional images, and also uses more directions in the transformation and can removes the noise pretty well in the smooth regions and also along the edges. We applied the algorithm in different medical images datasets. The experimental results show that this proposed algorithm performs better than the wavelet methods in both visually and statistically.

Finally, design a new filter for the directional filter banks is mentioned as a future work.

6. REFERENCES

[1] Do, Minh N., and Martin Vetterli. "Contourlets: a directional multiresolution image representation." *Image Processing. 2002. Proceedings. 2002 International Conference on*. Vol. 1. IEEE, 2002.

[2] Zhou, Z-F., and P-L. Shui. "Contourlet-based image denoising algorithm using directional windows." *Electronics Letters* 43.2 (2007): 92-93.

[3] Do, Minh N., and Martin Vetterli. "The contourlet transform: an efficient directional multiresolution image representation." *Image Processing, IEEE Transactions on* 14.12 (2005): 2091-2106.

[4] Po, DD-Y., and Minh N. Do. "Directional multiscale modeling of images using the contourlet transform." *Image Processing, IEEE Transactions on* 15.6 (2006): 1610-1620.

[5] Tsakanikas, Panagiotis, and Elias S. Manolakos. "Improving 2-DE gel image denoising using contourlets." *Proteomics* 9.15 (2009): 3877-3888.

[6] Liu, Zhe. "Minimum distance texture classification of SAR images in contourlet domain." *Computer Science and Software Engineering, 2008 International Conference on*. Vol. 1. IEEE, 2008.

[7] Esakkirajan, S., et al. "Image compression using contourlet transform and multistage vector quantization." *GVIP J* 6.1 (2006): 19-28.

[8] Rao, Ch Srinivasa, S. Srinivas Kumar, and B. N. Chatterji. "Content based image retrieval using contourlet transform." *ICGST-GVIP Journal* 7.3 (2007): 9-15.

[9] Minh, N. "*Directional Multiresolution Image Representations,"* (Doctoral dissertation of Engineering), Computer Engineering, University of Canberra, Australia, 2002.

[10] Bhateja, Vikrant, et al. "A modified speckle suppression algorithm for breast ultrasound images using directional filters." *ICT and Critical Infrastructure: Proceedings of the 48th*

Annual Convention of Computer Society of India-Vol II. Springer International Publishing, 2014.

[11] Hiremath, P. S., Prema T. Akkasaligar, and Sharan Badiger. "Removal of Gaussian Noise in Despeckling Medical Ultrasound Images." *The International Journal of Computer Science & Applications (TIJCA)* 1.5 (2012).

[12] Fayed, Hassan, Mohamed Rizk, and Ahmed Aboul Seoud. "Improved Medical Image Retrieval using Contourlet Techniques Based Interest Points Detector." *Canadian Journal on Image Processing and Computer Vision* 4.2 (2013).

[13] Song, Xiao-yang, et al. "Speckle reduction based on contourlet transform using scale adaptive threshold for medical ultrasound image." *Journal of Shanghai Jiaotong University (Science)* 13 (2008): 553-558.

[14] Hiremath, P. S., Prema T. Akkasaligar, and Sharan Badiger. "Speckle reducing contourlet transform for medical ultrasound images." *Int J Compt Inf Engg* 4.4 (2010): 284-291.

[15] Sivakumar, R., et al. "Image denoising using contourlet transform." *Computer and Electrical Engineering, 2009. ICCEE'09. Second International Conference on*. Vol. 1. IEEE, 2009.

[16] Burt, Peter J., and Edward H. Adelson. "The Laplacian pyramid as a compact image code." *Communications, IEEE Transactions on* 31.4 (1983): 532-540.

[17] Buades, Antoni, Bartomeu Coll, and Jean-Michel Morel. "A review of image denoising algorithms, with a new one." *Multiscale Modeling & Simulation* 4.2 (2005): 490-530.

[18] Saxena, Chandrika, and Deepak Kourav. "Noises and Image Denoising Techniques: A Brief Survey."*International Journal of Emerging Technology and Advanced Engineering*, 4.3 (2014):878-885.

[19] Wang, Zhou, et al. "Image quality assessment: from error visibility to structural similarity." *Image Processing, IEEE Transactions on* 13.4 (2004): 600-612.

SVM Based Recognition of Facial Expressions Used In Indian Sign Language

Daleesha M Viswanathan *daleesha_mv@address.com*
Department of Computer Science
Cochin University of Science and Technology
Kochi-22, Kerala, India

Sumam Mary Idicula *sumam@cusat.ac.in*
Department of Computer Science
Cochin University of Science and Technology
Kochi-22, Kerala, India

Abstract

In sign language systems, facial expressions are an intrinsic component that usually accompanies hand gestures. The facial expressions would modify or change the meaning of hand gesture into a statement, a question or improve the meaning and understanding of hand gestures. The scientific literature available in Indian Sign Language (ISL) on facial expression recognition is scanty. Contrary to American Sign Language (ASL), head movements are less conspicuous in ISL and the answers to questions such as yes or no are signed by hand. Purpose of this paper is to present our work in recognizing facial expression changes in isolated ISL sentences. Facial gesture pattern results in the change of skin textures by forming wrinkles and furrows. Gabor wavelet method is well-known for capturing subtle textural changes on surfaces. Therefore, a unique approach was developed to model facial expression changes with Gabor wavelet parameters that were chosen from partitioned face areas. These parameters were incorporated with Euclidian distance measure. Multi class SVM classifier was used in this recognition system to identify facial expressions in an isolated facial expression sequences in ISL. An accuracy of 92.12 % was achieved by our proposed system.

Keywords: Indian Sign Language, Facial Expression, Gabor Wavelet, Euclidian Distance, SVM.

1. INTRODUCTION

Facial expressions convey important synaptic information in sign language. For example, facial expression would modify the meaning of hand gesture sign for “you eat” in to “please eat”, “did you eat?” and “can you eat?” In ASL, the facial expressions and head movements are generally classified into expression of emotions, conversation regulator, expressions that modify the quality and quantity of a sign [1]. Similar to ASL, there are many occasions in ISL where facial expressions are used in sign language recognition. Whether a sentence is interrogative or simple sentence is decided by the associated facial expression [1]. As an example, “where are you going?” in ISL following “subject object verb” pattern is “you going where”, of which, ‘you’ and ‘going’ are signed by hand gesture but the question “where” is conveyed by facial expression [2]. In addition, the emotions such as anger, happiness, surprise and sadness are signed by facial expression in ISL [3]. In ASL, facial expression and head movement recognition systems have been standardized (Google Scholar and Web of Science citations) [1]. However, these works would not be directly applicable in ISL due to many reasons. Inherent difference exist between ASL and ISL [2]. Contrary to ASL, head movements are less conspicuous in ISL and there is clear variation in facial expressions between the two systems [3]. In addition, the answers to questions, such as yes or no are signed by hand in ISL

This paper deals with a system which can recognize facial changes in an isolated facial expression ISL sentences. The facial feature changes were detected by the appearance based approach [5]. Prior to analysis and feature extraction, the region of interest on which features has to be extracted would be identified and cropped using image processing methods. The facial expressions were modeled with the help of Gabor wavelet parameters and these parameters were incorporated by Euclidean distance measure. Multi class SVM classifier was used in this recognition system to identify facial expression in ISL. This paper is organized as follows: After a brief survey of related works in Section 2, facial expression recognition framework is presented in Section 3. Section 4 deals with result and discussion and Section 5 summarize the paper.

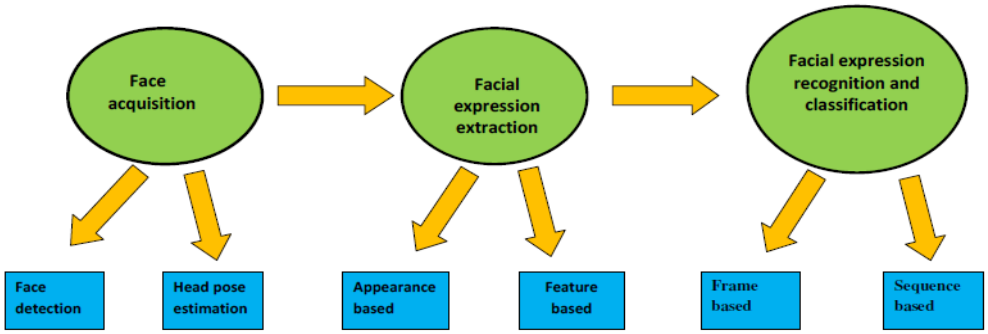


FIGURE 1: Basic steps in facial expression analysis.

2. RELATED WORKS

The accurate analysis of facial expression along with hand gesture is highly crucial in deciding the overall success of a recognition system. Accurately detecting the changes due to facial expression, its representation and classification are the major tasks of a recognition system (Figure 1). There are many challenges in facial expression features due to the dynamic nature of the signals that transmit information over time. Even the best classifiers fail to achieve accurate recognition rate due to inadequate features [4]. In facial feature extraction, for expression analysis, there are mainly three types of approaches. The holistic approach based on appearance, the geometric feature based method, and hybrid method based on the appearance and geometry [5].

In the Geometric method, the location of key facial components such as mouth, eyes, eyebrow and nose are being tracked and any variation due to expression on these parts are being targeted [4]. Subsequently, the feature vector transmits the extracted facial components at these key geometric regions on face [6]. This analysis has wide applicability in exploiting facial representation. In an approach, Candida wire frame model and active appearance algorithm for tracking and SVM for classification are used on image sequences [7].

The appearance based method, on the other hand, focuses on the whole face or specific region on face to frame the feature vector [4] [5]. The appearance based features method targets the textural changes on face such as wrinkles and furrows on face. Well known Holistic approaches are based on Principal component analysis [8]. Linear Discriminate analysis (LDA) [9], Independent component analysis (ICA) [10] and Gabor wavelet analysis [11] which were applied to either the whole face or specific face regions to extract the facial appearance changes [5]. In a study, Affine Moment Invariant was used as feature vector to compute the changes in eyebrows, nose, lips and mouth area.[12] Hybrid feature were also used to recognize facial expressions. In an approach, the face was partitioned and on the upper part, the Gabor wavelet method was applied and Active Appearance Model to the whole face [13].

One of the major applications of facial expression recognition system is the Sign language recognition process [5]. Without facial expression, signer cannot convey the complete intended meaning of the signs. In these situations, transition of expressions from neutral stage to the peak expressive state would be the major change takes place [14]. These expressions and changes can vary with the signer. Very few works were done in this area accounting this complexity. The percentage of works on facial expression over the total works in sign language computing is very low. In ISL, a thorough literature search in Google scholar and web of science (Technology) revealed lack of published research in this area. Therefore research efforts are required in this area to improve the ISL based recognition systems. In ASL, head movements along with facial expression and hand gestures are a major component in sign language representation.

Gabor wavelet representation has been successfully adopted in facial expression analysis [5]. In our work, we adopt, appearance based feature expression using Gabor Wavelet as the basic method in the facial expression analysis. Facial expressions are key components of ISL and facial gestures consistently lead to change in skin textures by forming wrinkles and furrows [4]. Gabor wavelet is well-known for capturing subtle textural changes on surface [4]. Therefore, in this analysis, Gabor wavelet parameters with Euclidian distance measure and Multi class SVM classifier were used to identify facial expressions in ISL.

3. RECOGNITION FRAMEWORK

Facial expressions in ISL are described using facial feature changes. A simplified description of six main expressional changes considered in this paper are represented in (Figure 2).Our recognition system uses the changes that appears in the upper and lower face areas of the signer to classify the expressions during the information exchange (Fig.2). The four main phases in this recognition system are presented in the block diagram (Fig.3).

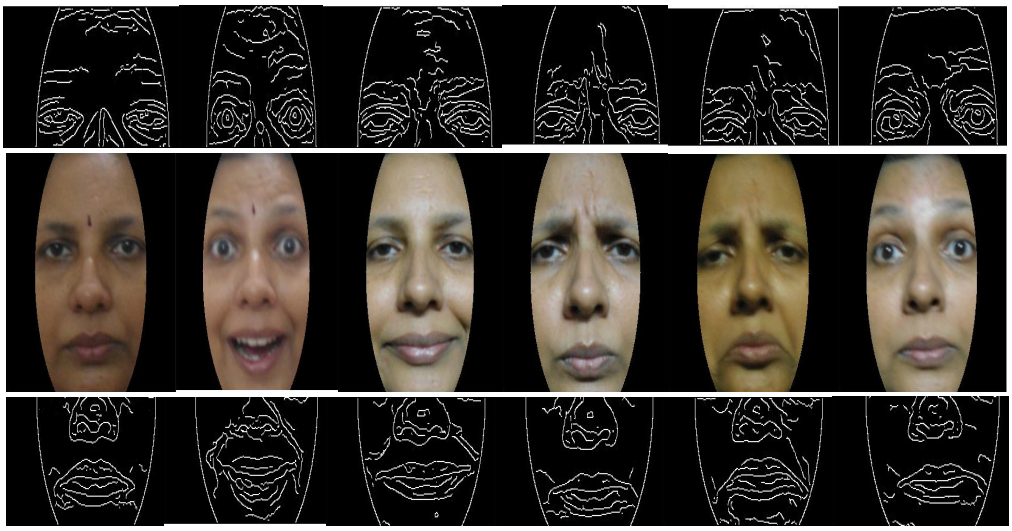


FIGURE 2: Row (b) shows the expressions neutral, surprise, happy, angry, sad and WH grammatical Marker (question). Row (a) and (c) shows the wrinkle changes in the upper and lower area of the face during signing.

3.1 Image Acquisition

Videos of isolated expression sequences, representing a short ISL sentence were recorded. In all the videos, signers' facial expression starts from neutral expression state and evolved into apex or peak point of expression. The length of the sequences varies depending on the facial expression and the signer/subject. The capturing is done using camera with frontal view of the person.

3.2 Pre-processing

This phase of the system is to extract the expressed faces from the frames of the video for recognition process. The two sub tasks involved in this pre-processing phase are the detection of the face from the frames and the extraction of frames with apex expressions. The cropped face area alone is used in the later phases of our recognition system.

3.2.1 Frame Extraction

In a video, there are two phases of interest. The neutral phase and the apex phase where maximum variation from the neutral phase occurs. The most relevant information is present in the final phase or apex phase and is crucial in facial expression recognition [14]. Following this methodology, the first frame representing neutral expression and last frame with peak expression are extracted for this study.

3.2.2 Face Detection and Extraction

Aim of the module is to extract the face area from extracted frames. Viola and Jones method based on Haar-like features and the AdaBoost learning algorithm [15] was used to detect the face part from the frames. Detected face part was cropped as a square, resized and masked into an elliptical shape as shown in Figure 4 to get the exact area of facial expression. Elliptically shaped cropped face would be partitioned horizontally along the elliptical centre into upper and lower face regions. The RGB segmented face areas were then converted into a normalized gray scale image. Figure 3 shows the example of the facial extraction process.

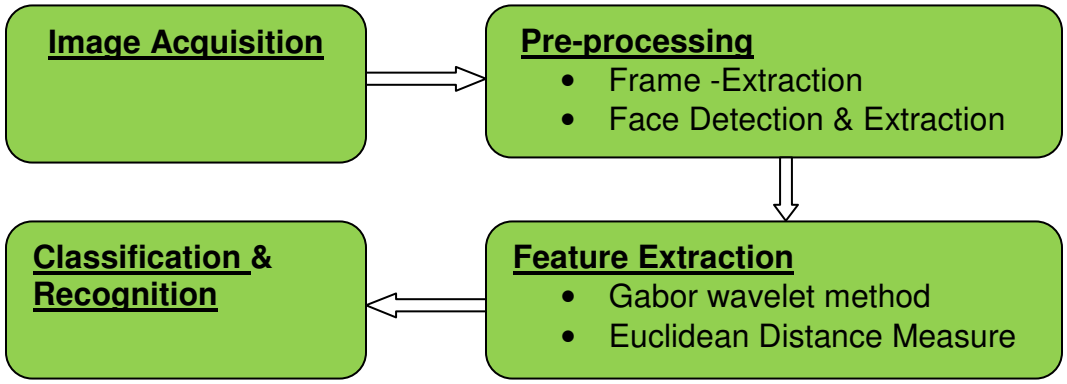


FIGURE 3: The four main phases in recognition system.

FIGURE 4: Steps in facial region extraction.

Function of facial expression system as shown in block diagram (Fig. 3) is summarized as follows.

3.3 Feature Extraction

The Figure 2 clearly indicates the importance of wrinkles on facial expression recognition. This motivates us to apply the well known textural analysis feature extraction method to the recognition process. Gabor wavelet method, which is well known for textural analysis and facial feature recognition, was chosen in our recognition system to represent the facial expressions. The signers' expression intensity and duration variations are not unique. Therefore, some common measuring criteria are required for modelling facial expression changes in an image

sequence. Facial expressions were described with the help of Gabor wavelet parameters chosen from the partitioned facial areas. These feature parameters were incorporated with Euclidean distance measure to represent the facial expression changes in an image sequence.

3.3.1 Gabor Wavelet Feature Representation

Gabor features were calculated by convolution of input image with Gabor filter bank [16] [17]. Gabor filter works as a band pass filter for the local spatial frequency distribution thereby achieving an optimal resolution in both the spatial and frequency domain. The 2D Gabor filter $\psi(x,y,f,\theta)$ can be represented as a complex sinusoidal signal, modulated by a Gaussian kernel function as in Eq (1).

$$\psi(x,y,f,\theta) = [1/2\pi\sigma^2]\,[\exp\{-(x_1^2 + y_1^2)/2\sigma^2\}]\,[\exp(2\pi f x_1)] \quad (1)$$

where $x_1 = x\cos\theta + y\sin\theta$

$y_1 = -x\sin\theta + y\cos\theta$

σ is the standard deviation of Gaussian envelop along the x, y dimension, f is the central frequency of the sinusoidal plane wave, θ is the orientation of gabor filter.

Feature extraction procedure can then be written as the convolution of gray scale facial expression image $I(x,y)$, with the Gabor filter $\psi(x,y,f,\theta)$ as in Eq (2).

$$G_{(u,v)}(x,y) = I(x,y) * \psi(x,y,f,\theta) \quad (2)$$

In Eq (2), $G_{(u,v)}(x,y)$ represent the complex convolution output which can decomposed into real and imaginary part as follows:

$E_{(u,v)}(x,y) = Re[G_{(u,v)}(x,y)]$ and $O_{(u,v)}(x,y) = I_m[G_{(u,v)}(x,y)]$.

Based on these result, both the phase as well as the magnitude response of the filter can be computed. In our work Gabor feature representation was based only on the magnitude response of the Gabor filter by neglecting the phase information. Small spatial displacement causes significant variation in phase value. Due to this variation, the two Gabor features could not be directly compared. Magnitude response $A_{(u,v)}(x,y)$ of the filter can be computed as in Eq (3).

$$A_{(u,v)}(x,y) = \sqrt{E^2 + O^2} \quad \text{where } E = E_{(u,v)}(x,y) \text{ and } O = O_{(u,v)}(x,y) \quad (3)$$

A Gabor filter bank with 5 frequencies and 8 orientations was used to extract Gabor features in our work. Down sampling was done on all magnitude response, which were then normalized and concatenated into Gabor Feature Vector.

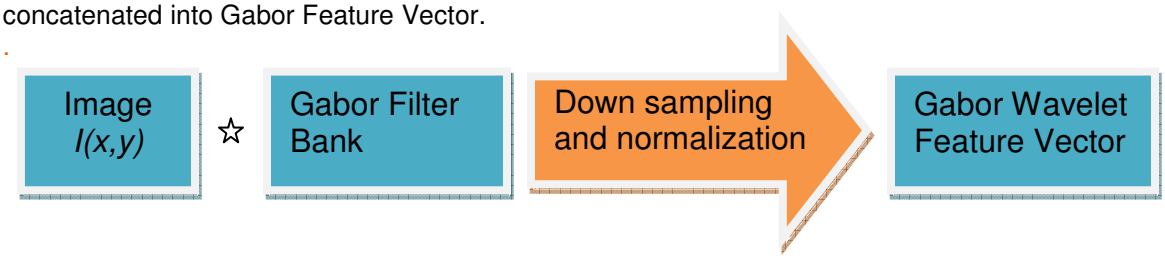


FIGURE 5: Gabor wavelet feature extraction sequence.

3.3.2 Distance Measure On Gabor Feature Vector

Gabor wavelet feature vector representing magnitude information in 2D real matrix form was converted to one dimensional matrix without any loss in information. For a facial expression corresponding to a frame, feature vectors were generated for upper and lower face regions which was partitioned horizontally alone the elliptical centre. In our work we concentrated only on first and last frames in the videos corresponding to isolated expression sentences. By this processing

method, four feature vectors were extracted from a video where two feature vectors were representing the upper part of the faces in the first and last frame and another two vectors representing the lower parts. Later, the Euclidean distance measure was applied between feature vectors corresponding to the upper parts and lower part feature vectors. In addition, the percentage change that occurs in the upper and lower face areas corresponding to the total change between the neutral and peak expression vectors were calculated. All these four measures later act as the feature descriptors for the facial expression changes.

3.4 Expression Classification using SVM

Support Vector Machine (SVM) classifier exhibit high classification accuracy for small training sets and good generalization performance on difficult to separate data set. Because of the superior classification performance, SVM was chosen for this study for facial expression changes. Displacement measure vectors corresponding to the facial expression changes, extracted in the previous phase were fed to the SVM for classification training.

SVM originally developed as a linear binary classifier [18] based on optimal separating hyper plane [Eq (4)] in the feature space and decision function [Eq (5)]. For a given training data pair (x_i , y_i), $y_i \in \{+1, \_1\}$.

$(w . x) + b = 0$, $w \in R^n$,$b \in R$, weight vector w, bias b (4)

$f(x) = (\Sigma_i \alpha_i y_i . k(x_i . x) + b)$ (5)

The SVM map the original input space into a higher dimensional feature space in order to achieve a linear solution [19] [20]. This mapping is done using kernel function. In Eq (5) $k (x_i . x)$ is the kernel transformation function. This can be a linear or non-linear kernel. Lagrange co-efficient α_i chosen from training samples are called support vectors and the decision function is defined by these support vectors. SVM with Radial Basis Function (RBF), a non –linear kernel function was used in this work for the analysis of facial expression changes.

4. RESULT AND DISCUSSION

The accuracy and performance of the proposed system were evaluated by using a set of videos that involve some common facial expression patterns used in the construction of ISL sentences.

4.1 Data Set

Ten video frames of isolated facial expression for happiness, sadness, angry, surprise and question were taken from five different persons at different time and location. The testing data set was chosen from different signers (person independent) and also from the available video sequences from internet. All the extracted frames for training and test were converted into normalized gray scale image of resolution 320 x 240 before processing.

4.2. Performance Analysis

For performance analysis, the facial expression was partitioned into upper and lower halves. The Gabor wavelet methodology was applied to both halves at the neutral frame and peak frame. The variation between these frames was computed by using Euclidian distance measures. In addition, the percent variation of each half between the neutral and peak frame was computed over the whole face variation. The student paired t test was performed to test any significant changes between the upper and lower halves for each facial expression category. The results indicated significant difference between both the halves ($p < 0.001$). This clearly justifies the partitioning approach, as this proposed scheme improve the recognition accuracy compared to facial gestures as whole face. Subsequently, a training data set is created by this approach. Displacement measure vectors corresponding to the facial expression changes taken by training were fed to SVM for classification of each expression variation.

The testing phase, output were tested with statistical measures such as sensitivity (recall), specificity, precision, f-measure and accuracy. Accuracy which shows the overall correctness of a model alone is not sufficient measure to make decision on the performance of a recognition system. Along with accuracy, we need to calculate sensitivity (recall) which gives how good a test is at detecting the correct expression, whereas specificity gives an estimate of how good a method in identifying negative expressions correctly, precision is a measure of exactness and F-measure is the harmonic mean of precision and recall. Score of all these measure reaches its best values at 1 and worst score at 0. A concomitant increase or decrease in values of specificity and sensitivity is required to consider a method to be superior or inferior. Similarly, a higher value for precision along with recall is required to rate a method as superior.

In this study lowest value for all the statistical measures were more than 0.7 (Fig 6 & 7). Generally values more than 0.5 were considered to be an acceptable and successful system for recognition [21]. Confusion matrix for recognizing facial expression changes is presented (Table 1). Table 2 shows the recognition accuracy of each expressional change from neutral to peak phases. The result indicated an average accuracy of 92.12 %. In addition, the performances of other statistical measures were superior (Figure 6 and 7).

TABLE 1: Confusion matrix for recognizing facial expression changes.

Expr.	Happy	Angry	Sad	Surprise	WH_expr	Neutral
Happy	**0.9**	0.0	0.0	0.0	0.0	0.0
Angry	0.0	**0.9**	0.0	0.0	0.1	0.0
Sad	0.0	0.0	**1.0**	0.0	0.0	0.0
Surprise	0.0	0.0	0.1	**0.8**	0.1	0.0
WH_expr	0.0	0.1	0.1	0.1	**0.7**	0.0
Neutral	0.0	0.0	0.0	0.0	0.0	**1.0**

TABLE 2: Expression Recognition Accuracy.

Expression	Happy	Angry	Sad	Surprise	WH_expr	Neutral
Accuracy	96%	93%	90%	93%	87%	100%

5. CONCLUSION AND FUTURE WORKS

We proposed a method to recognize the facial expressional changes in isolated ISL sentences. Six common facial expression categories of ISL were tested with Gabor wavelet methodology. Isolated facial expression sentences in ISL were taken for training and testing. Gabor wavelet parameters from the partitioned face areas, Euclidian distance measure and Multi class SVM classifier with RBF kernel function were used in this recognition system. There was significant difference between the upper and lower halves for their variability. Statistical measures were computed and the results indicated an overall accuracy of 92.12 % for the proposed system. Also found that the values for all the statistical measures were more than 0.7. The performance analysis revealed that the method followed in this study as highly promising and this could be used for the further up gradation of facial expression recognition systems in ISL. Since there is no commonly used dataset for Indian Sign language recognition, it is very difficult to compare different methods quantitatively. The scientific literature available in Indian Sign Language (ISL) on facial expression recognition is also lacking, which further complicates the comparison. The proposed method can be further extended for the recognition of multiple facial expressional changes in ISL sentences.

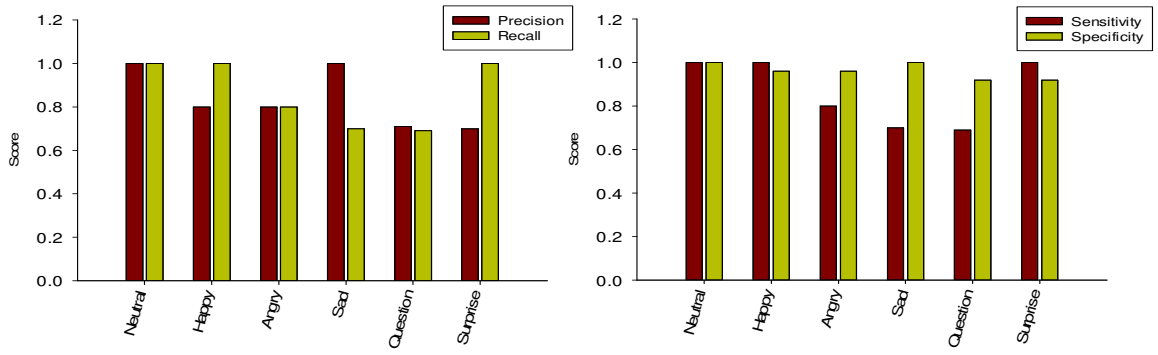


FIGURE: 6 The statistical measures precision, recall, specificity and sensitivity for 6 facial expression categories.

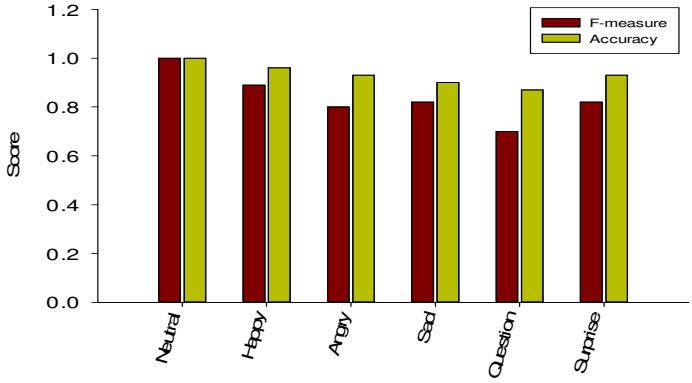


FIGURE 7: The F measure and accuracy for six facial expression categories.

6. REFERENCES

1. N. Tan Dat and S. Ranganath, "Facial expressions in American sign language: Tracking and recognition", Pattern Recognition, vol 45, pp. 1877-1891, 2012.

2. M.W. Morgan,"Topology of Indian Sign Language verbs from a comparative perspective", An annual review of South Asian languages and linguistics,vol 222, pp. 103-131, 2009.

3. B. Bridges and M. Metzger," Deaf tend your", Calliope press. Silver Spring, 1996, pp 27-38 .

4. Y.L. Tiam, T. Kanade and J.F. Cohn, " Facial expression analysis", Springer New York, 2005, pp. 247-275.

5. C. Shan, S. Gong and P.W. McOwan," Facial expression recognition based on local binary patterns: A comprehensive study", Image and Vision Computing, vol. 27, pp. 803-816, 2009.

6. D. Das,"Human's facil parts extraction to recognize facial expression", International journal on information theory, vol. 3, pp. 65-72. 2014.

7. R.A, Patil, V. Sahula and A.S. Mandal,"Facial expression recognition in image sequences using active shape model and SVM", Uksim Fifth European Modelling Symposium on Computer Modelling and Simulation,, 2011, pp. 168-173.

8. M. Turk and A.P. Pentlanad,"Face recognition using Eigen face", IEEE Conference on computer vision and pattern recognition, 1991.

9. P.N. Belhumeur, J.P. Hespanha and D.J. Kriegman,"Eigen faces versus Fisher faces: Recognition using class specific linear projection", IEEE Transactions on pattern analysis and machine intelligence, pp. 711-720, 1997.

10. Bartlett, M.S., J.R. Movellan and T.J. Sejnowski," Face recognition by independent component analysis", IEEE Transactions on Neural Networks, pp. 1450-1464, 2002.

11. M.J. Lyions, J. Budynek and S. Akametsu,"Automatic classification of single facial image", IEEE Transactions on pattern analysis and machine intelligence, pp. 1357-1362, 1999.

12. R. Londhe and V. Pawar,"Facial expression recognition based on affine moment invariants", International Journal of Computer Science, vol. 9, pp. 382-392, 2012.

13. Y. Zhan and G. Zhou,"Facial expression recognition based on hybrid features and fusing discrete HMMs", Virtual reality-Book chapter, Springer 2007, pp. 408-417.

14. J.M. Gold,"Efficiency of dynamic and static facial expression",Journal of vision, pp. 1-12, 2013.

15. P. Viola and M. Jones,"Rapid object detection using a bossted cascade of simple features", Computer visison and pattern recognition, 2001.

16. C. Liu and H. Wechsler," Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition", IEEE Transactions on image processing, 2002, pp. 467-476.

17. V. Struc and N. Vavesic,"From Gabor magnitude to Gabor face features: Tackling the problem of face recognition under severe illumination changes",Face recognition-Chaper 12. www.intechopen.com, 2011, p. 215-238.

18. V. Vapnik,"The nature of statistical learning theory",Springer, 2000.

19. S. Sindhumol, K. Anil and B. Kannan,"Abnormality detection from Multispectoral brain MRI using multiresolution independent component analysis", International journal of signal processing, image processing and pattern recognition, pp. 177-190, 2013.

20. I. Ari and L. Akarun,"Facial Feature Tracking and Expression Recognition for Sign Language",2009 IEEE 17th Signal processing and communications applications conference, 2009, pp. 479-482.

21. S.S.K, Nair, N.V.S. Reddy and K.S. Hareesha," Explointing heterogenous features to improve insillico prediction of peptide status of amyloidogenic or non-amyloidogenic", BMC bioinformatics, pp. 2-9. 2011.

INSTRUCTIONS TO CONTRIBUTORS

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

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

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

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

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

LIST OF TOPICS

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

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

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

CALL FOR PAPERS

Volume: 9 - **Issue: 3**

i. Submission Deadline : April 30, 2015 **ii. Author Notification:** May 31, 2015

iii. Issue Publication: June 2015

CONTACT INFORMATION

Computer Science Journals Sdn BhD

B-5-8 Plaza Mont Kiara, Mont Kiara

50480, Kuala Lumpur, MALAYSIA

Phone: 006 03 6204 5627

Fax: 006 03 6204 5628

Email: cscpress@cscjournals.org

CSC PUBLISHERS © 2015
COMPUTER SCIENCE JOURNALS SDN BHD
B-5-8 PLAZA MONT KIARA
MONT KIARA
50480, KUALA LUMPUR
MALAYSIA

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