David Zhang (Ed.)

LNCS 4901

Medical Biometrics

First International Conference, ICMB 2008
Hong Kong, China, January 2008
Proceedings

Springer

David Zhang (Ed.)

Medical Biometrics

First International Conference, ICMB 2008
Hong Kong, China, January 4-5, 2008
Proceedings

Springer

Volume Editor

David Zhang
The Hong Kong Polytechnic University
Department of Computing
Hung Hom, Kowloon, Hong Kong
E-mail: csdzhang@comp.polyu.edu.hk

Library of Congress Control Number: 2007941380

CR Subject Classification (1998): I.4, I.5, I.3, I.2.10, I.2.1, J.3

LNCS Sublibrary: SL 6 – Image Processing, Computer Vision, Pattern Recognition, and Graphics

ISSN 0302-9743
ISBN-10 3-540-77410-6 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-77410-5 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12208487 06/3180 5 4 3 2 1 0

Lecture Notes in Computer Science 4901

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

Preface

The past decade has seen a rapid growth in the demand for the usage of behavioral and physiological characteristics of humans for e-*security and forensics* applications. The significant advances in the biometrics techniques are increasingly incorporated in a number of other applications. In this context, medicine has emerged as the largest and most promising area. Medical biometrics primarily refers to the usage of behavioral and physiological characteristics of humans for *medical diagnosis and body care*. Thus the goal of medical biometrics is to explore solutions to the open problems in medicine using biometric measurements, technologies and systems.

The International Conference on Medical Biometrics (ICMB 2008) was the first major gathering in the world devoted to facilitating this interaction. We are pleased that this conference attracted a large number of high-quality research papers that will benefit international medical biometrics research. After a careful review process, 40 papers were accepted either for oral (17) or poster (23) presentations. In addition to these technical presentations, this conference also held a workshop on Computerized Traditional Chinese Medicine. Our efforts are focused to generating awareness among researchers, organizing researchers to foster research in this area, and providing a common platform for discussion and investigation. This conference provided a forum for discussing practical experiences in applying the state-of-the-art biometric tecnologies for medical diagnosis and body care, which will further stimulate research in medical biometrics.

We are grateful to Max A. Viergever, Heinz-Otto Peitgen, and Tadashi Watsuji for accepting our invitation to give keynote talks at ICMB 2008. In addition, we would like to express our gratitude to all the contributors, reviewers, Program Committee and Organizing Committee members who made this a very successful conference. We also wish to acknowledge the International Association of Pattern Recognition (IAPR), IEEE Computational Intelligence Society (IEEE-CIS), National Natural Science Foundation in China (NSFC), and Springer for sponsoring this conference. Special thanks are due to Jane You, Milan Sonka, Xiaoyi Jiang, Ajay Kumar, Lei Zhang, Jing Li, Vivek K., Zhenhua Guo and Liu Li for their dedication and hard work in various aspects of the conference organization. We thankfully acknowledge the partial support from the National Natural Science Foundation in China (NSFC) Key Overseas Project (60620160097).

We hope that the fruitful technical interactions during this conference will benefit your further research and development efforts in medical biometrics.

October 2007

David Zhang

Organization

General Chair

David Zhang (The Hong Kong Polytechnic University, Hong Kong)

Program Chairs

Milan Sonka (The University of Iowa, USA)
Jane You (The Hong Kong Polytechnic University, Hong Kong)
Xiaoyi Jiang (University of Münster, Germany)

Program Committee

Hong Yan (City University of Hong Kong, Hong Kong)
Prabir Bhattacharya (Concordia University, Canada)
Alejandro Frangi (Pompeu Fabra University, Spain)
Bernadette Dorizzi (Institut National des Télécommunications, France)
Chuang-Chien Chiu (Feng Chia University, Taiwan)
Dinggang Shen (University of Pennsylvania School of Medicine, USA)
Heang-Ping Chan (University of Michigan, USA)
Changen Wang (National Natural Science Foundation of China, PR China)
Jeffery Huang (Raytheon Company, USA)
Jerome Zhengrong Liang (State University of New York at Stony Brook, USA)
Kwang Suk Park (Seoul National University, Korea)
Kwang-Sup Soh (Seoul National University, Korea)
Liang-Yu Shyu (Chung Yuan Christian University, Taiwan)
Mateo Aboy (Oregon Institute of Technology, USA)
Michael Abrmoff (University of Iowa, USA)
Naimin Li (No.211 Hospital of People Liberation Army, PR China)
Nicholas Ayache (The French Research Institute of Computer Science and Automatic Control, France)
Opas Chutatape (Nanyang Technological University, Singapore)
Pong Chi Yuen (The Hong Kong Baptist University, Hong Kong)
Yazhu Chen (Shanghai Jiao Tong University, PR China)
Stanislaw Osowski (Warsaw University of Technology, Poland)
Ruwei Dai (Chinese Academy of Science, PR China)
Witold Pedrycz (University of Alberta, Canada)
Yanda Li (Tsinghua University, PR China)
Yang Cai (Carnegie Mellon University, USA)
Yoshiharu Yonezawa (Hiroshima Institute of Technology, Japan)
Ali Can (Woods Hole Oceanographic Institution, USA)

Aly A. Farag (University of Louisville, USA)
Badrinath Roysam (Rensselaer Polytechnic Institute, USA)
Denis Noble (University of Oxford, UK)
Greg Goddard (University of California San Francisco, USA)
Howard L. Tanenbaum (Albany Medical College, USA)
Joao Paulo Silva Cunha (Universidade de Aveiro, Portugal)
Joes Staal (University of Canterbury, Netherlands)
John J. Shynk (University of California, Santa Barbara, USA)
Leif Sörnmo (Lund University, Sweden)
May Dongmei Wang (Georgia Institute of Technology, USA)
Richard B. Reilly (University College Dublin, Ireland)
T. Walter (Paris School of Mines, France)
Hongzhi Yang (Sun Yat-sen University, PR China)
Tingdong Zhang (Harbin Medical University The 2nd Affiliated Hospital, PR China)
Jihua Qin (Shandong University of Traditional Chinese Medicine, PR China)
Rong Liang (Beijing University of Chinese Medicine, PR China)
Weitong Huang (Tsinghua University, PR China)
Yue Zhou (Shanghai Jiao Tong University, PR China)
Kuanquan Wang (Harbin Institute of Technology, PR China)
Joachim Hornegger (University of Erlangen, Germany)
Jianzhou Zhang (University of Sichuan, PR China)
Yung-Fu Chen (China Medical University, Taiwan)
Ajay Kumar (Indian Institute of Technology Delhi, India)
Yongping Zheng (The Hong Kong Polytechnic University, Hong Kong)
James Gee (University of Pennsylvania School of Medicine, USA)

Lecture Notes in Computer Science

Sublibrary 6: Image Processing, Computer Vision, Pattern Recognition, and Graphics

Vol. 4901: D. Zhang (Ed.), Medical Biometrics. XII, 324 pages. 2007.

Vol. 4844: Y. Yagi, S.B. Kang, I.S. Kweon, H. Zha (Eds.), Computer Vision – ACCV 2007, Part II. XXVIII, 915 pages. 2007.

Vol. 4843: Y. Yagi, S.B. Kang, I.S. Kweon, H. Zha (Eds.), Computer Vision – ACCV 2007, Part I. XXVIII, 969 pages. 2007.

Vol. 4842: G. Bebis, R. Boyle, B. Parvin, D. Koracin, N. Paragios, S.-M. Tanveer, T. Ju, Z. Liu, S. Coquillart, C. Cruz-Neira, T. Müller, T. Malzbender (Eds.), Advances in Visual Computing, Part II. XXXIII, 827 pages. 2007.

Vol. 4841: G. Bebis, R. Boyle, B. Parvin, D. Koracin, N. Paragios, S.-M. Tanveer, T. Ju, Z. Liu, S. Coquillart, C. Cruz-Neira, T. Müller, T. Malzbender (Eds.), Advances in Visual Computing, Part I. XXXIII, 831 pages. 2007.

Vol. 4815: A. Ghosh, R.K. De, S.K. Pal (Eds.), Pattern Recognition and Machine Intelligence. XIX, 677 pages. 2007.

Vol. 4814: A. Elgammal, B. Rosenhahn, R. Klette (Eds.), Human Motion – Understanding, Modeling, Capture and Animation. X, 329 pages. 2007.

Vol. 4792: N. Ayache, S. Ourselin, A. Maeder (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2007, Part II. XLVI, 988 pages. 2007.

Vol. 4791: N. Ayache, S. Ourselin, A. Maeder (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2007, Part I. XLVI, 1012 pages. 2007.

Vol. 4781: G. Qiu, C. Leung, X. Xue, R. Laurini (Eds.), Advances in Visual Information Systems. XIII, 582 pages. 2007.

Vol. 4778: S.K. Zhou, W. Zhao, X. Tang, S. Gong (Eds.), Analysis and Modeling of Faces and Gestures. X, 305 pages. 2007.

Vol. 4756: L. Rueda, D. Mery, J. Kittler (Eds.), Progress in Pattern Recognition, Image Analysis and Applications. XXI, 989 pages. 2007.

Vol. 4738: A. Paiva, R. Prada, R.W. Picard (Eds.), Affective Computing and Intelligent Interaction. XVIII, 781 pages. 2007.

Vol. 4729: F. Mele, G. Ramella, S. Santillo, F. Ventriglia (Eds.), Advances in Brain, Vision, and Artificial Intelligence. XVI, 618 pages. 2007.

Vol. 4713: F.A. Hamprecht, C. Schnörr, B. Jähne (Eds.), Pattern Recognition. XIII, 560 pages. 2007.

Vol. 4679: A.L. Yuille, S.-C. Zhu, D. Cremers, Y. Wang (Eds.), Energy Minimization Methods in Computer Vision and Pattern Recognition. XII, 494 pages. 2007.

Vol. 4678: J. Blanc-Talon, W. Philips, D. Popescu, P. Scheunders (Eds.), Advanced Concepts for Intelligent Vision Systems. XXIII, 1100 pages. 2007.

Vol. 4673: W.G. Kropatsch, M. Kampel, A. Hanbury (Eds.), Computer Analysis of Images and Patterns. XX, 1006 pages. 2007.

Vol. 4642: S.-W. Lee, S.Z. Li (Eds.), Advances in Biometrics. XX, 1216 pages. 2007.

Vol. 4633: M. Kamel, A. Campilho (Eds.), Image Analysis and Recognition. XII, 1312 pages. 2007.

Vol. 4584: N. Karssemeijer, B. Lelieveldt (Eds.), Information Processing in Medical Imaging. XX, 777 pages. 2007.

Vol. 4569: A. Butz, B. Fisher, A. Krüger, P. Olivier, S. Owada (Eds.), Smart Graphics. IX, 237 pages. 2007.

Vol. 4538: F. Escolano, M. Vento (Eds.), Graph-Based Representations in Pattern Recognition. XII, 416 pages. 2007.

Vol. 4522: B.K. Ersbøll, K.S. Pedersen (Eds.), Image Analysis. XVIII, 989 pages. 2007.

Vol. 4485: F. Sgallari, A. Murli, N. Paragios (Eds.), Scale Space and Variational Methods in Computer Vision. XV, 931 pages. 2007.

Vol. 4478: J. Martí, J.M. Benedí, A.M. Mendonça, J. Serrat (Eds.), Pattern Recognition and Image Analysis, Part II. XXVII, 657 pages. 2007.

Vol. 4477: J. Martí, J.M. Benedí, A.M. Mendonça, J. Serrat (Eds.), Pattern Recognition and Image Analysis, Part I. XXVII, 625 pages. 2007.

Vol. 4472: M. Haindl, J. Kittler, F. Roli (Eds.), Multiple Classifier Systems. XI, 524 pages. 2007.

Vol. 4466: F.B. Sachse, G. Seemann (Eds.), Functional Imaging and Modeling of the Heart. XV, 486 pages. 2007.

Vol. 4418: A. Gagalowicz, W. Philips (Eds.), Computer Vision/Computer Graphics Collaboration Techniques. XV, 620 pages. 2007.

Vol. 4417: A. Kerren, A. Ebert, J. Meyer (Eds.), Human-Centered Visualization Environments. XIX, 403 pages. 2007.

Vol. 4391: Y. Stylianou, M. Faundez-Zanuy, A. Esposito (Eds.), Progress in Nonlinear Speech Processing. XII, 269 pages. 2007.

Vol. 4370: P.P. Lévy, B. Le Grand, F. Poulet, M. Soto, L. Darago, L. Toubiana, J.-F. Vibert (Eds.), Pixelization Paradigm. XV, 279 pages. 2007.

Vol. 4358: R. Vidal, A. Heyden, Y. Ma (Eds.), Dynamical Vision. IX, 329 pages. 2007.

Vol. 4338: P.K. Kalra, S. Peleg (Eds.), Computer Vision, Graphics and Image Processing. XV, 965 pages. 2006.

Vol. 4319: L.-W. Chang, W.-N. Lie (Eds.), Advances in Image and Video Technology. XXVI, 1347 pages. 2006.

Vol. 4292: G. Bebis, R. Boyle, B. Parvin, D. Koracin, P. Remagnino, A. Nefian, G. Meenakshisundaram, V. Pascucci, J. Zara, J. Molineros, H. Theisel, T. Malzbender (Eds.), Advances in Visual Computing, Part II. XXXII, 906 pages. 2006.

Vol. 4291: G. Bebis, R. Boyle, B. Parvin, D. Koracin, P. Remagnino, A. Nefian, G. Meenakshisundaram, V. Pascucci, J. Zara, J. Molineros, H. Theisel, T. Malzbender (Eds.), Advances in Visual Computing, Part I. XXXI, 916 pages. 2006.

Vol. 4245: A. Kuba, L.G. Nyúl, K. Palágyi (Eds.), Discrete Geometry for Computer Imagery. XIII, 688 pages. 2006.

Vol. 4241: R.R. Beichel, M. Sonka (Eds.), Computer Vision Approaches to Medical Image Analysis. XI, 262 pages. 2006.

Vol. 4225: J.F. Martínez-Trinidad, J.A. Carrasco Ochoa, J. Kittler (Eds.), Progress in Pattern Recognition, Image Analysis and Applications. XIX, 995 pages. 2006.

Vol. 4191: R. Larsen, M. Nielsen, J. Sporring (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006, Part II. XXXVIII, 981 pages. 2006.

Vol. 4190: R. Larsen, M. Nielsen, J. Sporring (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006, Part I. XXXVVIII, 949 pages. 2006.

Vol. 4179: J. Blanc-Talon, W. Philips, D. Popescu, P. Scheunders (Eds.), Advanced Concepts for Intelligent Vision Systems. XXIV, 1224 pages. 2006.

Vol. 4174: K. Franke, K.-R. Müller, B. Nickolay, R. Schäfer (Eds.), Pattern Recognition. XX, 773 pages. 2006.

Vol. 4170: J. Ponce, M. Hebert, C. Schmid, A. Zisserman (Eds.), Toward Category-Level Object Recognition. XI, 618 pages. 2006.

Vol. 4153: N. Zheng, X. Jiang, X. Lan (Eds.), Advances in Machine Vision, Image Processing, and Pattern Analysis. XIII, 506 pages. 2006.

Vol. 4142: A. Campilho, M. Kamel (Eds.), Image Analysis and Recognition, Part II. XXVII, 923 pages. 2006.

Vol. 4141: A. Campilho, M. Kamel (Eds.), Image Analysis and Recognition, Part I. XXVIII, 939 pages. 2006.

Vol. 4122: R. Stiefelhagen, J.S. Garofolo (Eds.), Multimodal Technologies for Perception of Humans. XII, 360 pages. 2007.

Vol. 4109: D.-Y. Yeung, J.T. Kwok, A. Fred, F. Roli, D. de Ridder (Eds.), Structural, Syntactic, and Statistical Pattern Recognition. XXI, 939 pages. 2006.

Vol. 4091: G.-Z. Yang, T. Jiang, D. Shen, L. Gu, J. Yang (Eds.), Medical Imaging and Augmented Reality. XIII, 399 pages. 2006.

Vol. 4073: A. Butz, B. Fisher, A. Krüger, P. Olivier (Eds.), Smart Graphics. XI, 263 pages. 2006.

Vol. 4069: F.J. Perales, R.B. Fisher (Eds.), Articulated Motion and Deformable Objects. XV, 526 pages. 2006.

Vol. 4057: J.P.W. Pluim, B. Likar, F.A. Gerritsen (Eds.), Biomedical Image Registration. XII, 324 pages. 2006.

Vol. 4046: S.M. Astley, M. Brady, C. Rose, R. Zwiggelaar (Eds.), Digital Mammography. XVI, 654 pages. 2006.

Vol. 4040: R. Reulke, U. Eckardt, B. Flach, U. Knauer, K. Polthier (Eds.), Combinatorial Image Analysis. XII, 482 pages. 2006.

Vol. 4035: T. Nishita, Q. Peng, H.-P. Seidel (Eds.), Advances in Computer Graphics. XX, 771 pages. 2006.

Vol. 3979: T.S. Huang, N. Sebe, M. Lew, V. Pavlović, M. Kölsch, A. Galata, B. Kisačanin (Eds.), Computer Vision in Human-Computer Interaction. XII, 121 pages. 2006.

Vol. 3954: A. Leonardis, H. Bischof, A. Pinz (Eds.), Computer Vision – ECCV 2006, Part IV. XVII, 613 pages. 2006.

Vol. 3953: A. Leonardis, H. Bischof, A. Pinz (Eds.), Computer Vision – ECCV 2006, Part III. XVII, 649 pages. 2006.

Vol. 3952: A. Leonardis, H. Bischof, A. Pinz (Eds.), Computer Vision – ECCV 2006, Part II. XVII, 661 pages. 2006.

Vol. 3951: A. Leonardis, H. Bischof, A. Pinz (Eds.), Computer Vision – ECCV 2006, Part I. XXXV, 639 pages. 2006.

Vol. 3948: H.I. Christensen, H.-H. Nagel (Eds.), Cognitive Vision Systems. VIII, 367 pages. 2006.

Vol. 3926: W. Liu, J. Lladós (Eds.), Graphics Recognition. XII, 428 pages. 2006.

Vol. 3872: H. Bunke, A.L. Spitz (Eds.), Document Analysis Systems VII. XIII, 630 pages. 2006.

Vol. 3852: P.J. Narayanan, S.K. Nayar, H.-Y. Shum (Eds.), Computer Vision – ACCV 2006, Part II. XXXI, 977 pages. 2006.

Vol. 3851: P.J. Narayanan, S.K. Nayar, H.-Y. Shum (Eds.), Computer Vision – ACCV 2006, Part I. XXXI, 973 pages. 2006.

Vol. 3832: D. Zhang, A.K. Jain (Eds.), Advances in Biometrics. XX, 796 pages. 2005.

Vol. 3736: S. Bres, R. Laurini (Eds.), Visual Information and Information Systems. XI, 291 pages. 2006.

Vol. 3667: W.J. MacLean (Ed.), Spatial Coherence for Visual Motion Analysis. IX, 141 pages. 2006.

Vol. 3417: B. Jähne, R. Mester, E. Barth, H. Scharr (Eds.), Complex Motion. X, 235 pages. 2007.

Vol. 2396: T.M. Caelli, A. Amin, R.P.W. Duin, M.S. Kamel, D. de Ridder (Eds.), Structural, Syntactic, and Statistical Pattern Recognition. XVI, 863 pages. 2002.

Vol. 1679: C. Taylor, A. Colchester (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI'99. XXI, 1240 pages. 1999.

₹424.00

Table of Contents

Feature Extraction and Classification

Health Care

Medical Diagnosis

Medical Image Processing and Registration

A New Feature Selection and Classification Scheme for Screening of Oral Cancer Using Laser Induced Fluorescence

Lalit Gupta, Sarif Kumar Naik, and Srinivasan Balakrishnan

Philips Research Asia - Bangalore,
Bangalore - 500045, India
{lalit.gupta,sarif.naik,srinivasan.balakrishnan}@philips.com

Abstract. Screening for oral cancer in its early stage is of utmost importance for improving the survival rate of oral cancer patients. The current method of visual examination followed by biopsy of suspected cases is subjective with inter and intra personal variations. With a low ratio of oral-cancer experts to patients compounded by the reluctance of patients to undergo biopsy in rural India. The situation cries out for automatic screening device for oral cancer. In this context, optical spectroscopy based on Laser Induced Fluorescence (LIF) has been shown to be a promising technique in distinguishing between cancerous and benign lesions in the mouth. However, it has been observed that it is very difficult to distinguish pre-malignant spectra from malignant and normal spectra recorded *in-vivo*. Hence, obtaining the most discriminating features from the spectra becomes important. In this article a new method of feature selection is proposed using mean-shift and Recursive Feature Elimination (RFE) techniques to increase discrimination ability of the feature vectors. Performance of the algorithm is evaluated on a *in-vivo* recorded LIF data set consisting of spectra from normal, malignant and pre-malignant patients. Sensitivity of above 95% and specificity of above 99% towards malignancy are obtained using the proposed method.

1 Introduction

Oral cancer refers to the cancer of the oral cavity and pharynx. It is the sixth most common form of cancer in the world. It poses a major public health problem. The number of new cases are estimated to be 274,289 [1] per year and about two-thirds of them arise in developing countries. The Indian sub-continent accounts for more than one-third of the total number oral cancer cases in the world [2]. The incidence of oral cancer is rising in several other regions of the world such as Europe, Taiwan, Japan and Australia. Oral cancer is most often found by physicians at a later stage as a persistent ulcer with raised edges and an indurated base. Oral cancer may be preceded by pre-malignant lesions such as leukoplakia, erythroplakia, oral sub-mucous fibrosis (OSMF) etc. Some of these lesions may transform into cancer over a period of time. The cancerous lesions of oral cavity are staged depending on the size, location, the clinical aspects and

D. Zhang (Ed.): ICMB 2008, LNCS 4901, pp. 1–8, 2007.

the histopathological features of biopsy. The 5-year survival rate in oral cancer is less than 50%. However, if diagnosed at early pre-cancerous stage, the chances of survival could be greatly improved with minimal intervention.

The current method of visual examination followed by biopsy of suspected cases is subjective. With a low ratio of oral-cancer experts to patients compounded by the reluctance of patients to undergo biopsy in rural India. The situation cries out for automatic screening device for oral cancer. In recent years, various optical spectroscopy methods have been studied for the screening of oral cancer [3,4]. In general LIF spectroscopy in particular has been shown to be a promising technique for classifying oral lesions as cancerous and non-cancerous [3,4]. In the present study LIF technology has been used for screening of oral cancer. The scope of this article includes automatic classification of LIF spectra into three classes malignant, pre-malignant and normal.

Most of the algorithms reported for the classification of LIF data use traditional multi-variate statistical techniques such as principal component analysis (PCA), singular value decomposition (SVD) etc. Nayak *et al.*[3] used PCA for feature extraction and match no-match scheme for classification of LIF Spectra obtained *in-vitro*. They used PCA to obtain a feature vector for each samples. Each of these feature vectors consists of the scores from the first four principal components and the sum of squared spectral residual for each spectrum. Match no-match test is performed using Mohalanobis distance and spectral residual. Sensitivity of 99% and specificity of 95% is reported using this methods. Kamath *et al.*[5] have worked on the same data using K-means clustering and similar results are reported. Majumder *et al.*[6] reported an algorithm using support vector machine (SVM) for the development of algorithms for optical diagnosis of cancer. Recursive feature elimination (RFE) is proposed for feature selection. SVM is used for classification stage as well as for feature selection. Better results are claimed compared to PCA based feature extraction algorithms and Fishers linear discriminant (FLD)- based algorithms. Using the proposed scheme sensitivity of 95% and specificity of 96% toward cancer are reported on the training set and sensitivity of 93% and specificity of 97% toward cancer are reported on the independent validation set. In the pre-processing step of the algorithm, variance of all the spectra (training as well as testing) in the data is used for normalization. However, variation in test samples should not be considered to normalize the training samples because test samples are supposed to be blind from the training samples. Hence, the reported accuracy in sensitivity and specificity likely be far away from the actual.

Naik *et al.*[7] performed a comparative study of different feature extraction techniques for the classification of LIF spectra recorded from oral tissue. It is reported that discrete wavelet transform (DWT) features give better classification accuracy than other features such as PCA, linear discriminant analysis(LDA) or independent component analysis(ICA). Classification is performed using support vector machines. Apart from [3,5] and [6] other work only classify oral cancer into two classes of malignant and pre-malignant. Only Majumdar *et al.*[6] and [7] worked on *in-vivo* recorded spectra and all other work are on *in-vitro* recorded spectra.

In this paper, a system for classifying LIF data is proposed. In the first step features are extracted from LIF spectra using DWT [7] and then the most relevant features are selected from all the features using a new feature selection technique. Finally, classification is done using support vector machines. The rest of the article is organized as follows: A brief description of the LIF data set used for classification is discussed in Section 2. The algorithm proposed for feature selection is explained in Section 3. Results obtained using proposed algorithm are discussed in Section 4. Concluding remarks on the proposed algorithm are given in Section 5.

2 Overview of LIF Spectra

The details of LIF system used for collecting data is given in [7]. Three examples of spectra corresponding to normal, pre-malignant and malignant condition and recorded *in-vitro* are shown in Fig. 1(a). It is observed that the malignant spectrum contains only one peak due to Nicotinamide Adenine Dinucleotide (NADH) at 440nm and the normal spectrum contain two peaks one due to collagen at 400nm and the other due NADH at 440nm. Two peaks can also be seen in pre-malignant spectra. However, the difference between normal spectra and pre-malignant spectra is that he relative strength of the two peaks in case of pre-malignant spectrum is similar. Hence, to a great extent, it is possible to distinguish between normal, pre-malignant and malignant spectra easily by visual observation.

The aim of the current study is to screen oral cancer non-invasively in real time. Hence, for this study all the spectra are recorded *in-vivo*. Some examples of spectra recorded using the LIF system are shown in Fig. 1(b)-(c). However, unlike *in-vitro* data *in-vivo* data has huge variations. This can be observed from the spectra shown in Fig. 1(b)-(c). It can be observed from Fig. 1(b) that pre-malignant spectrum look like a weighted average spectrum of a normal and pre-malignant spectra, i.e., there is a peak at 440nm for all types of spectra and the peak strength at 400nm increases from malignant to normal to premalignant. Fig. 1(c) shows a set of spectra, in which it is almost impossible to differentiate between malignant and pre-malignant spectra by visual observation. In a few rare cases malignant spectrum also shows a valley at 400nm. In these cases it becomes difficult to visually differentiate between malignant and normal spectra. The

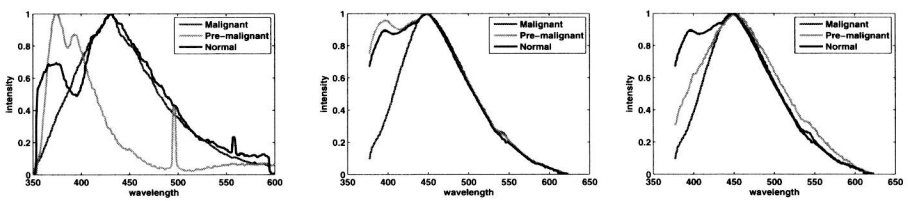

Fig. 1. (a) Spectrum showing LIF data recorded *in-vitro*. (b)-(c) Spectra recorded using the LIF system *in-vivo*.

possible sources of these variations are (1) subjective evaluation of malignancy of lesions by clinicians and (2) variation in thickness of the layer of tissue in different stages of malignancy.

3 Proposed Methodology

Each LIF spectra consists of 1024 sample points recorded from the wavelength of 375nm to 600nm. The steps followed in the proposed algorithm are shown in Fig. 2. The complete algorithm comprises of three steps, and each of the three steps are described below:

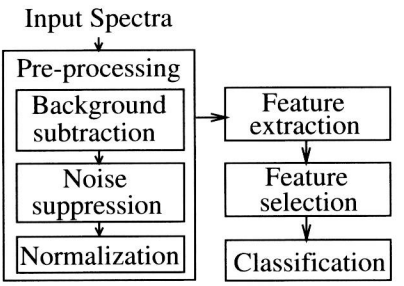

Fig. 2. Steps in the algorithm

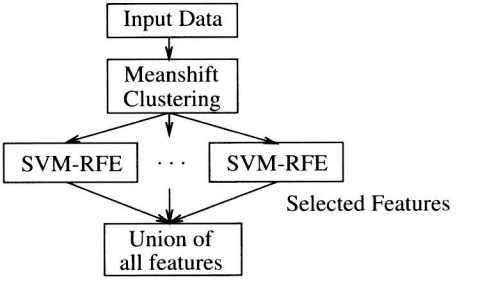

Fig. 3. Steps in feature selection

3.1 Pre-processing

The first step is pre-processing. It contains three sub-steps, background subtraction, filtering and normalization. First, we subtract the background spectra to get the actual spectra information. Due to the variations in recording environment, the LIF system can induce some unwanted signal (background noise) to the spectrum. This is commonly known as plasma lines included to the original signal due to back scattering properties of the quartz window, used in the LIF system. Hence, a background spectrum is captured before each set of recordings from a patient, by exposing the probe to the blank space. By subtracting

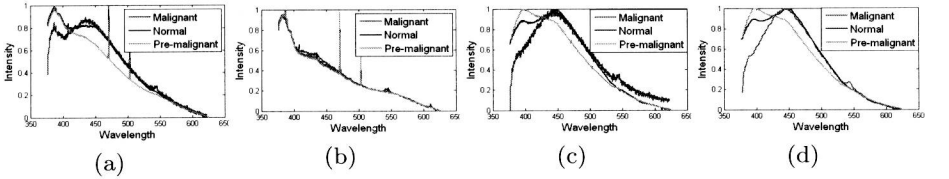

Fig. 4. (a) Raw spectra obtained from normal, malignant and premalignant sites. (b) Background spectra corresponding to spectra in (a); (c) Background subtracted spectra corresponding to spectra in (a); (d) Resultant spectra obtained by filtering the spectra in (c).