Advances in Intelligent Systems and Computing 1407

Vikrant Bhateja
Suresh Chandra Satapathy
Carlos M. Travieso-González
V. N. Manjunath Aradhya *Editors*

Data Engineering and Intelligent Computing

Proceedings of ICICC 2020

Springer

Advances in Intelligent Systems and Computing

Volume 1407

The series “Advances in Intelligent Systems and Computing” contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing such as: computational intelligence, soft computing including neural networks, fuzzy systems, evolutionary computing and the fusion of these paradigms, social intelligence, ambient intelligence, computational neuroscience, artificial life, virtual worlds and society, cognitive science and systems, Perception and Vision, DNA and immune based systems, self-organizing and adaptive systems, e-Learning and teaching, human-centered and human-centric computing, recommender systems, intelligent control, robotics and mechatronics including human-machine teaming, knowledge-based paradigms, learning paradigms, machine ethics, intelligent data analysis, knowledge management, intelligent agents, intelligent decision making and support, intelligent network security, trust management, interactive entertainment, Web intelligence and multimedia.

The publications within “Advances in Intelligent Systems and Computing” are primarily proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Indexed by DBLP, EI Compendex, INSPEC, WTI Frankfurt eG, zbMATH, Japanese Science and Technology Agency (JST).
All books published in the series are submitted for consideration in Web of Science.

More information about this series at http://www.springer.com/series/11156

Vikrant Bhateja · Suresh Chandra Satapathy ·
Carlos M. Travieso-González ·
V. N. Manjunath Aradhya
Editors

Data Engineering and Intelligent Computing

Proceedings of ICICC 2020

Springer

Editors
Vikrant Bhateja
Department of Electronics
and Communication Engineering
Shri Ramswaroop Memorial Group
of Professional Colleges (SRMGPC)
Lucknow, Uttar Pradesh, India

Dr. A.P.J. Abdul Kalam Technical
University
Lucknow, Uttar Pradesh, India

Carlos M. Travieso-González
Department of Signals
and Communications
Institute for Technological Development
Las Palmas de Gran Canaria, Spain

Suresh Chandra Satapathy
School of Computer Engineering
Kalinga Institute of Industrial Technology
Bhubaneswar, India

V. N. Manjunath Aradhya
Department of Computer Applications
JSS Science and Technology University
Mysuru, India

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-981-16-0170-5 ISBN 978-981-16-0171-2 (eBook)
https://doi.org/10.1007/978-981-16-0171-2

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Conference Organisation Committees

Chief Patrons
Dr. D. Hemachandra Sagar, Chancellor, DSU, Chairman DSI
Dr. D. Premachandra Sagar, Pro Chancellor, DSU, Chairman DSI

Patrons
Shri Galiswamy, Secretary, DSI
Dr. K. N. B. Murthy, Vice Chancellor, DSU
Shri. R. Janardhan, Pro-Vice Chancellor, DSU
Dr. Puttamadappa C., Registrar, DSU

Honorary Chairs
H. P. Khincha, Evangelist, DSU
K. Jairaj, Advisor, DSU

Honorary General Chair
Dr. Hao Ying, Wayne State University, Michigan, USA

General Chair
Dr. S. C. Satapathy, KIIT, Bhubaneshwar, Odisha, India

Publication Chair
Dr. Vikrant Bhateja, SRMGPC, Lucknow, U.P., India
Dr. Carlos M. Travieso-Gonzalez, University of Las Palmas de Gran Canaria (ULPGC), Spain

Organizing Chairs
Dr. M. K. Banga, Dean Research, DSU
Dr. A Srinivas, Dean SoE, DSU
Dr. Sanjay Chitnis, Chairman CSE, DSU

Technical Program Chairs
Dr. V. N. Manjunatha Aradhya, SJCE, Mysore, Karnataka, India
Dr. Wendy Flores-Fuentes, Autonomous University of Baja California, Mexico

Program Chairs
Dr. Rajesh T. M., Dept. of CSE, DSU
Dr. Jasma Balasangameshwara, Dept. of CST, DSU

Organizing Secretaries
Dr. Shaila S. G., Dept. of CSE, DSU
Dr. Girisha G. S., Dept. of CSE, DSU
Dr. Mallanagouda Patil, Dept. of CSE, DSU

Advisory Committee
Dr. Aime' Lay-Ekuakille, University of Salento, Lecce, Italy
Dr. Amira Ashour, Tanta University, Egypt
Dr. Aynur Unal, Standford University, USA
Dr. Bansidhar Majhi, IIIT Kancheepuram, Tamil Nadu, India
Dr. Dilip Kumar Sharma, Vice-Chairman, IEEE U.P. Section
Dr. Ganpati Panda, IIT Bhubaneswar, Odisha, India
Dr. Jagdish Chand Bansal, South Asian University, New Delhi, India
Dr. João Manuel R. S. Tavares, Universidade do Porto (FEUP), Porto, Portugal
Dr. Jyotsana Kumar Mandal, University of Kalyani, West Bengal, India
Dr. K.C. Santosh, University of South Dakota, USA
Dr. Le Hoang Son, Vietnam National University, Hanoi, Vietnam
Dr. Naeem Hanoon, Multimedia University, Cyberjaya, Malaysia
Dr. Nilanjan Dey, TIET, Kolkatta, India
Dr. Noor Zaman, Universiti Tecknologi, PETRONAS, Malaysia
Dr. Roman Senkerik, Tomas Bata University, Zlin, Czech Republic
Dr. Swagatam Das, Indian Statistical Institute, Kolkatta, India
Dr. Siba K. Udgata, University of Hyderabad, Telangana, India
Dr. Shri Nivas Singh, MMMUT, Gorakhpur, U.P., India
Dr. Steven L. Fernandez, The University of Alabama at Birmingham, USA.
Dr. Tai Kang, Nanyang Technological University, Singapore
Dr. Valentina Balas, Aurel Vlaicu University of Arad, Romania

Technical Program Committee
Prof. H R Vishwakarma, VIT, Vellore, India
Prof. Dan Boneh, Stanford University, California, USA
Prof. Alexander Christea, University of Warwick, London UK
Prof. Ahmad Al- Khasawneh, The Hashemite University, Jordan
Dr. Bharat Singh Deora, JRNRV University, India
Prof. Jean Michel Bruel, Department Informatique IUT de Blagnac, Blagnac, France
Prof. Ngai-Man Cheung, University of Technology and Design, Singapore
Prof. Yun-Bae Kim, Sung Kyun Kwan University, South Korea

Prof. Ting-Peng Liang, National Chengchi University Taipei, Taiwan
Prof. Sami Mnasri, IRIT Laboratory Toulouse, France
Prof. Lorne Olfman, Claremont, California, USA
Prof. Brent Waters, University of Texas, Austin, Texas, United States
Prof. Philip Yang, Price Water House Coopers, Beijing, China
Prof. R K Bayal, Rajasthan Technical University, Kota, Rajasthan, India
Prof. Martin Everett, University of Manchester, England
Prof. Feng Jiang, Harbin Institute of Technology, China
Prof. Prasun Sinha, Ohio State University Columbus, Columbus, OH, United States
Dr. Savita Gandhi, Gujarat University, Ahmedabad, India
Prof. Xiaoyi Yu, Chinese Academy of Sciences, Beijing, China
Prof. Gengshen Zhong, Jinan, Shandong, China
Dr. Mukesh Shrimali, Pacific University, Udaipur, India
Prof. Komal Bhatia, YMCA University, Faridabad, Haryana, India
Prof. S. R. Biradar, College of Engineering & Technology, Dharwad, Karnataka
Dr. Soura Dasgupta, SRM University, Chennai, India
Dr Sushil Kumar, Jawaharlal Nehru University, New Delhi, India
Dr. Amioy Kumar, IIT Delhi, India

Preface

This book is a collection of high-quality peer-reviewed research papers presented at the '4th International Conference on Intelligent Computing and Communication (ICICC-2020)' held at the School of Engineering (SoE), Dayananda Sagar University (DSU), Bengaluru, Karnataka, India, during 18–20 September 2020.

After the success of the past three editions of ICICC held in 2016 at the University of Kalyani, West Bengal, India, in 2017 at MAEER's MIT College of Engineering, Pune, India, and in 2019 at SoE, DSU, Bengaluru, India, the 4th edition of ICCC was again organized by SoE, DSU, Bengaluru, India. All papers of the past ICICC editions are published by Springer AISC series. Presently, ICICC-2020 provided a platform for academicians, researchers, scientists, professionals and students to share their knowledge and expertise in the diverse domain of intelligent computing and communication.

ICICC-2020 had received a number of submissions from the field of ICT, intelligent computing and its prospective applications in different spheres of engineering. The papers received have undergone a rigorous peer-review process with the help of the technical programme committee members of the conference from various parts of the country as well as abroad. The review process has been crucial with the minimum of two to three reviews each along with due checks on similarity and content overlap. This conference has featured seven exclusive theme-based special sessions as enlisted below along with the main track and hence witnessed more than 400 submissions in total.

S1: Application of Artificial Intelligence in Mobile Apps: A Machine and Deep Learning Techniques Perspective
S2: Deep Learning for Intelligent Systems
S3: Computer Vision for Multimedia Security and Forensics
S4: Emerging Trends in Machine Learning Methodologies and Applications in Artificial Intelligence
S5: Artificial Intelligence and Nature-Inspired Algorithms
S6: Recent Innovations in Artificial Intelligence Research and Computing Technologies
S7: Artificial Intelligence and Machine Learning Applications (AIML)

Out of the above pool, only quality papers were finally presented in 10 parallel tracks/day (of virtual presentation sessions chaired by esteemed professors from premier institutes of the country) and compiled in this volume for publication. This volume consists of these selected papers focussing on the theme of 'Data Engineering and Intelligent Computing'.

The conference featured many distinguished keynote addresses by eminent speakers like Mr. Upendra M Kulkarni (Vice President and G M, Firmware and Software (Non-volatile Memory Solutions Group) at Intel Corporation, California, USA, who delivered a lecture on the Role of High Performance and High-Density Storage in Intelligent Computing; Mr. Andrew Haschka (APAC Technology Practice Lead, Application Modernization, Google Cloud at Google, Australia), who delivered a lecture on Building on Application Platform; Dr. Dinakar Sitaram (Chairperson of the Cloud Computing Innovation Council of India working group, ex-CTO HP System Technology, Software Division, India), who delivered a lecture on Speech Recognition for Indian and Low Resource Languages; Mr. Swaminathan Venkataraman (Worldwide Product and Solution Architect, Communications and Media Solutions (CMS), Hewlett Packard Enterprise, India), who delivered a lecture on Cloud Native 5G.

These keynote lectures/talks embraced a huge toll of audience of students, faculties, budding researchers as well as delegates. The editors thank the General Chair, TPC Chair and the Organizing Chair of the conference for providing valuable guidelines and inspirations to overcome various difficulties in the process of organizing this conference. The editors also thank SoE, DSU, Bengaluru, for their whole-hearted support in organizing this conference.

The editorial board take this opportunity to thank the authors of all the submitted papers for their hard work, adherence to the deadlines and patience during the review process. The quality of a refereed volume depends mainly on the expertise and dedication of the reviewers. We are indebted to the TPC members who not only produced excellent reviews but also did these in short time frames.

Lucknow, India	Dr. Vikrant Bhateja
Bhubaneswar, India	Dr. Suresh Chandra Satapathy
Las Palmas de Gran Canaria, Spain	Dr. Carlos M. Travieso-González
Mysuru, India	Dr. V. N. Manjunath Aradhya
October 2020	

The original version of the book was revised: Book volume number has been corrected as 1407. The correction to the book is available at https://doi.org/10.1007/978-981-16-0171-2_61

Contents

About the Editors

Vikrant Bhateja is Associate Professor, Department of Electronics and Communication Engineering, Shri Ramswaroop Memorial Group of Professional Colleges (SRMGPC), Lucknow, and also Head (Academics and Quality Control) in the same college. He completed his doctorate in Bio-Medical Imaging and Signal Processing and has a total academic teaching experience of 16 years with around 125 publications in reputed international conferences, journals and online book chapter contributions. His areas of research include digital image and video processing, computer vision, medical imaging, machine learning, pattern analysis and recognition. Dr. Vikrant has edited 15 proceeding books/editorial volumes with Springer Nature. He is Editor-in-Chief of *IGI Global—International Journal of Natural Computing and Research* (IJNCR). He is also Associate Editor in *International Journal of Synthetic Emotions* (IJSE) and *International Journal of Ambient Computing and Intelligence* (IJACI) under IGI Global press. He is Guest Editor of Special Issues in reputed Scopus/SCIE indexed journals under Springer Nature: *Evolutionary Intelligence* and *Arabian Journal of Science and Engineering.*

Suresh Chandra Satapathy is a Ph.D. in Computer Science, currently working as Professor and at KIIT (Deemed to be University), Bhubaneshwar, Odisha, India. He held the position of the National Chairman Div-V (Educational and Research) of Computer Society of India and is also Senior Member of IEEE. He has been instrumental in organizing more than 20 International Conferences in India as Organizing Chair and edited more than 30 Book Volumes from Springer LNCS, AISC, LNEE and SIST Series as Corresponding Editor. He is quite active in research in the areas of swarm intelligence, machine learning and data mining. He has developed a new optimization algorithm known as Social Group Optimization (SGO) published in Springer Journal. He has delivered number of keynote address and tutorials in his areas of expertise in various events in India. He has more than 100 publications in reputed journals and conference proceedings. Dr. Suresh is in Editorial Board of IGI Global, Inderscience, Growing Science journals and also Guest Editor for *Arabian Journal of Science and Engineering* published by Springer.

Carlos M. Travieso-González received the M.Sc. degree in 1997 in Telecommunication Engineering at Polytechnic University of Catalonia (UPC), Spain, and Ph.D. degree in 2002 at the University of Las Palmas de Gran Canaria (ULPGC-Spain). He is Full Professor on Signal Processing and Pattern Recognition and Head of Signals and Communications Department at ULPGC, teaching from 2001 on subjects on signal processing and learning theory. His research lines are biometrics, biomedical signals and images, data mining, classification system, signal and image processing, machine learning, and environmental intelligence. He has researched in 50 International and Spanish Research Projects, some of them as Head Researcher. He has 440 papers published in international journals and conferences. He has published 7 patents in Spanish Patent and Trademark Office. He is Evaluator of project proposals for European Union (H2020), Medical Research Council (MRC—UK), Spanish Government (ANECA—Spain), Research National Agency (ANR—France), DAAD (Germany), and other Institutions. He has been General Chair in 16 international conferences, mainly sponsored by IEEE and ACM. He is Associate Editor in CIN and Entropy (Q2-WoS journals). He has been awarded in "Catedra Telefonica" Awards in editions 2017, 2018 and 2019 on Knowledge Transfer Modality.

V. N. Manjunath Aradhya is currently working as Associate Professor in the Department of MCA, Sri Jayachamarajendra College of Engineering, Mysuru. He received the M.S. and Ph.D. degrees in Computer Science from the University of Mysore, Mysuru, India, in 2004 and 2007, respectively. He is a recipient of "Young Indian Research Scientist" from Italian Ministry of Education, University and Research, Italy, during 2009–2010. His professional recognition includes as Technical Editor for *Journal of Convergence Information Technology* (JCIT), Editor Board in *Journal of Intelligent Systems*, a reviewer for *IEEE Transactions on System, Man and Cybernetics—Part B*, *Pattern Recognition* (PR), *Pattern Recognition Letters* (PRL), etc. Also he is member of program committees of over 45 international conferences. His research interest includes pattern recognition, image processing, document image analysis, computer vision, machine intelligence, applications of linear algebra for the solution of engineering problems, biclustering of gene expression data and web data analysis and understanding. He has published more than 135 research papers in reputed international journals, conference proceedings and edited books.

Heuristic Approach to Detect Pectoral Muscle and Nipple in Mammogram for Computer-Aided Diagnosis

Vibha Bafna Bora, Ashwin Kothari, and Avinash Keskar

Abstract In computer-aided diagnosis algorithms for mammograms, many preprocessing steps like pectoral muscle detection and the coordinates of the nipple are done automatically before the actual detection and diagnosis process begin. Normally, in medio-lateral oblique (MLO) views, the pectoral muscle should be seen to the tip of the nipple or below. Consequently, straight line estimate of pectoral muscle is parallel to the tangent drawn at nipple. Considering this heuristic anatomical relation, in the proposed work, hybrid combination of textured Hough transform with distant surround suppression was used to detect straight or curved pectoral muscle on MLO view followed by orthogonal Radon transform to detect the nipple position. The proposed algorithm was tested on 266 MLO views from miniMIAS and 74CC views from local hospital. The pectoral muscle was detected with accuracy of 97%, and detected nipple position was acceptable in 93.67% of mammograms with average error of 5.755 mm with location annotated by a radiologist.

Keywords Pectoral muscle detection · Nipple detection · Computer aided diagnosis · Hough transform · Radon transform · Distant surround suppression

1 Introduction

Mammograms are X-ray images of the breast with low dose and are used for screening asymptomatic women for early finding of breast cancer. Breast cancer is one of the wide spread diseases in women, and a lot of effort has been put in avoidance through

V. B. Bora (✉)
G H Raisoni College of Engineering, Nagpur, India
e-mail: vibha.bora@raisoni.net

A. Kothari · A. Keskar
Visvesvaraya National Institute of Technology, Nagpur, India
e-mail: ashwinkothari@ece.vnit.ac.in

A. Keskar
e-mail: avinashkeskar@ece.vnit.ac.in

V. Bhateja et al. (eds.), *Data Engineering and Intelligent Computing*,
Advances in Intelligent Systems and Computing 1407,
https://doi.org/10.1007/978-981-16-0171-2_1

different types of screening processes. Such screening will necessarily generate a huge number of mammograms which must be sighted and read by a limited number of professional radiologists. Hence, usage of computer-assisted interpretation of the results can reduce the workload on radiologists [1]. These systems are usually used as a second adviser, with the final conclusion regarding the presence of a cancer left to the radiologist.

Pectoral muscle and nipple location are normally the key points for mammogram registration in computer-aided diagnosis (CAD). Removal of pectoral muscle is also vital in following cases:

1. Pectoral muscle must be removed completely in an automated breast tissue density quantification method, otherwise, in case of fatty tissues, it will cause an over evaluation of the breast density. On the contrary, mammogram density may be underestimated, if the triangular pectoral region accidentally includes some of the dense glandular area adjacent to the pectoral region [2].
2. Due to resemblance between mammographic glandular parenchyma and pectoral muscles, higher number of false positives occurs in mass detection method [3].

Similarly, location of nipple also plays crucial role in following cases:

1. Nipple location is one of the significant landmarks for measuring possible asymmetry between the right and left breast of a patient in computer-aided detection (CAD) analysis [4].
2. Radiologists normally pay more attention to the nipple while examination, as cancer can arise in glandular tissues which converge onto nipple [5, 6].
3. Coordinates of nipple are important in finding the architectural distortion of the mammogram [7].
4. In CAD, nipple is an alignment pivot [8].

Proper detection of pectoral muscle and nipple is being so important that these have been a main support to the researchers working on CAD. The methods [9–12] in which muscle was assumed to be straight give moderate accuracy since majority of the muscle follows curvature. As reported by Kinoshita et al. [12], images with straight line edges had enhanced performance with and inferior performance with other types of edges of the pectoral muscle region. The method by Ferrari et al. [13], based on Gabor transform, was reported to be very time-consuming, and only 84 miniMIAS images were taken which excluded very challenging images like mdb053l, mdb061l, mdb179l, mdb240r, etc. The performance of Kwok et al. [14] was visually reviewed by a five point index only, but no pixel-wise quantitative consideration was performed. Also method Raba et al. [15] results into excess of segmentation for mammograms when the dense glandular tissues are close to the pectoral muscle and the performance was only visually evaluated. In texture field method [16], quantitative assessment was performed but for mammogram with BI-RADS type-4; the breast that contains glandular and fibrous tissue greater than 75% and skin folded in upper part covers 20–30% of the breast region near to muscle as a pectoral muscle. Thus, the registration of pectoral muscle is a challenging task due to several factors like overlapping of muscle by fibro glandular tissue, muscle periphery masked by artifacts such as sticky tapes and axillary fold due to ill positioning.

The method [6, 8] uses the breast boundary technique to obtain nipple location and neither of the two methods reported whether the nipple was in silhouette on the mammogram. In [5] also, breast boundary was extracted, and the performance of the method was evaluated using 24 images; a very limited data set with 16 medio-lateral (MLO) views and eight cranio-caudal (CC) views. As reported in [12], the nipple is detected only if nipple is located on breast surface as bright regions. Thus, nipple registration is also a challenging task because of the nipple visibility or invisibility in mammogram, variations in image quality and patient positioning.

The aim of the proposed work is to develop a robust method for pectoral muscle detection in MLO view which must consider overlapping of muscle with dense tissues as well as curvature followed by nipple detection.

In the proposed method, textured Hough transform with distant surround suppression technique is developed to delineate straight or curved pectoral muscle accurately. Further, using the slope of pectoral muscle, two orthogonal Radon transforms are taken to register nipple (visible or invisible) on mammogram.

2 Materials and Methods

2.1 Database

The database consists of total 340 mammograms, 266 are MLO views from a miniMIAS database [17], and 74 are CC views locally procured from (Central India Cancer Research Institute (CICRI), Nagpur, Maharashtra, India). The original image of miniMIAS database, digitized at 50 micron pixel edge, has been condensed to 200 micron pixel edge and padded so that single image is 1024 × 1024 pixels. CICRI mammograms were digitized with GE Senographe600T scanner with a pixel edge of 100 micron. Each original image is thus 2370 × 1770 pixels and a gray value resolution of 12 bit. These images are down sampled, without loss of significant information to 1185 × 885 pixels with a resolution of 200 micron pixel and a gray value resolution of 8 bit. The proposed method was executed on MATLAB ver. R2016.

2.2 Artifact Removal, Orientation Correction and Enhancement

In scanned or digital mammograms labels, tags and scanning artifacts were present, which make detection of key point more challenging. Artifacts were removed using morphological process and connected component algorithm [18]. There is no orientation standard in mammography, and as reported by Mustra et al. [19] compared to unknown position in the image, known orientation makes detection of muscle somewhat easier, so muscles are correctly oriented with the help of heuristic approach [20]. Hence, each mammogram had pectoral muscle in the top left corner before muscle and nipple detection algorithm is applied. Next breast background is cropped and contrast enhanced [18, 21].

2.3 Pectoral Muscle Detection

The pectoral region is the high intensity triangular section in the top left corner of the mammogram. Let the mammogram image be indicated by $I(x, y)$. The top left corner of the image is the origin (1, 1), where X is defined to be the vertical axis and Y to be the horizontal axis. M: no. of rows and N: no. of columns. Hough transform can be used to approximate the muscle edge with straight lines [18]. But in many mammograms, pectoral muscles are heavily textured and occluded by dense tissues. Hence, in proposed work, instead of applying Hough transform to an intensity mammogram, it was applied to textured plot of mammogram to obtain straight line approximation of muscle edge accurately.

2.3.1 Textured Hough Transform, Region of Interest and Straight Line Approximation

The textured plot of mammogram (TPM) can be plotted using texture feature such as local entropy, phase gradient and gray intensities [22]. Local entropy is a statistical measure of randomness in the subimage. It achieves maximum value when all elements of the image are maximally random and minimum for constant image. As muscle boundary is characterized by high intensity change, hence in TPM, muscle edge will have high values of local entropies. Since orientation correction is done, muscle is always located in upper left corner. Hence geometrically, it is always situated in first quadrant from 30° to 90°. The phase gradient at each pixel location (x, y) in the mammogram is obtained using first derivative Sobel operator [18]. Hence, TPM is the set of those pixels having orientation in the range of 30°–90°, local entropies greater than mean entropy of the mammogram and gray intensities greater than 0.5 of I_{max}, where I_{max} is maximum intensity present in mammogram. Since muscle is high intensity region, pixels present in muscle are always greater than 0.5 of I_{max}.

During image acquisition, depending on anatomy and position of patient, half of the breast region or near about of it could be occupied by the pectoral muscle [13]. Hence, from the TPM, initial rectangle shaped region of interest (ROI) is considered for straight line approximation of pectoral edge and is represented as.

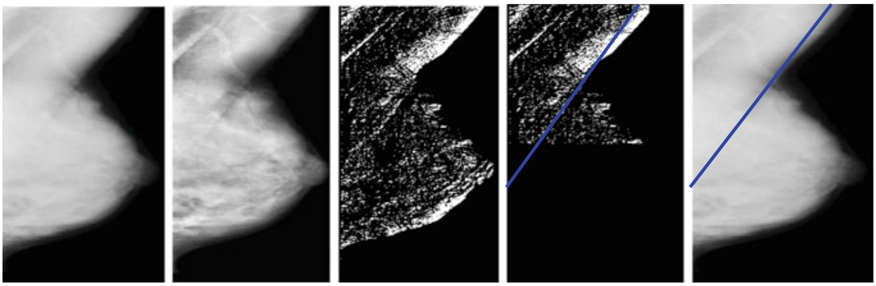


Fig. 1 Straight line approximation of pectoral muscle for mdb053**l**. **a** Original images mdb053**l**, **b** mammogram after contrast enhancement, **c** textured plot of mammogram; TPM, **d** region of interest and approximated straight edge in blue color using Hough transform, **e** approximated straight pectoral edge overlapped on original mammogram

$$\text{ROI} = \{\text{TPM}(x, y) : 1 \leq x \leq M/2 \text{ and } 1 \leq y \leq 3 * N/4\} \quad (1)$$

To the ROI of Eq. (1), Hough transform is applied to obtain strongest edge [18, 22]. If the strongest straight line intersects top and left edge of the mammogram, then it is a valid straight line, otherwise ROI is tuned by about 20%.

Figure 1a–e shows the entire process of straight line approximation of pectoral muscle for the case mdb053**l**.

2.3.2 Pectoral Edge Refinement Using Distant Surround Suppression

The anatomical shape of the pectoral muscle is either straight, concave or convex or mixture of both. Using Hough transform on TPM, pectoral edge was approximated as a straight line PQ as shown in Fig. 2a. Let the coordinates of the approximated straight pectoral boundary be P (1, P_{wt}) and Q (P_{ht}, 1), where P_{wt} and P_{ht} are the distance from the upper left corner of the mammogram to the top and bottom edge of the approximate pectoral muscle, respectively. To obtain the actual curvature of the muscle, the region near the straight edge is considered by shifting straight line toward left and right by value Px from the approximated edge PQ as shown in Fig. 2a. The left straight boundary AB is plotted from $A(1, P_{wt}\text{-}Px)$ to $B(P_{ht}\text{-}Px, 1)$, whereas right straight boundary CD is plotted from $C(1, P_{wt} + Px)$ to $D(P_{ht} + Px, 1)$, and the selected textured plot of mammogram (STPM), i.e., APCDQBA for mdb053l is shown in Fig. 2a. Further STPM is $n \times n$ (here 5×5 empirical value taken) block averaged so that the pixels which are least involved in pectoral edge representation are removed as seen in Fig. 2b. The averaged STPM is obtained as

$$\text{Avg\_STPM}[x : x + (n-1), y : y + (n-1)] = \frac{\sum_{s=0}^{(n-1)} \sum_{t=0}^{(n-1)} \text{ROI}(x + s, y + t)}{n^2} \quad (2)$$

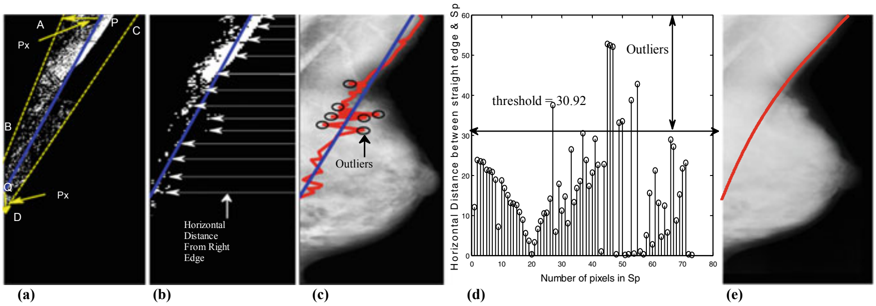


Fig. 2 Pectoral muscle detection using distant surround suppression. **a** Selected textured plot of mammogram (STPM) for mdb053**l**, **b** 5 × 5 block averaged STPM, scanned horizontally to obtain S_P; a set of first nonzero pixels from right edge of mammogram. **c** S_p plotted as a rough pectoral edge; black circle is indicated as some outliers. **d** Absolute distance plot between pixels in S_p and their horizontal distance between straight pectoral edge and S_P. **e** Smoothened pectoral muscle using cubical polynomial on original mammogram

where $x = 1, n, 2n, \ldots, M/2$ and $y = 1, n, 2n, \ldots, N/4$. Blocks with average value of 0.51 or higher are considered as 1 (white) otherwise to 0 (Black). The size of Px is optimally chosen to wrap pectoral boundary variation for vastly concave or convex shaped muscle. If the smaller value of Px is chosen, then STPM will miss pixels which are significant on pectoral edge to give the accurate shape. Contrary if larger value Px is chosen, then unwanted nearby pixels may get perceived to form imprecise shape of the muscle. In this work, the experimental value of $Px = 80$ pixels is taken.

In next step, binary Avg_STPM image is scanned horizontally from right vertical edge to pick the first nonzero pixel as shown in Fig. 2b. As a result, a string of pixels position (S_P) is obtained and considered as rough pectoral edge as shown in Fig. 2c, out of which majority of pixels lies on the pectoral edge. The outliers in S_P may be due to heavy textured pattern nearby muscle edge or due to axillary folds. To remove the outliers, distant surround suppression (DSS) technique is applied. In DSS technique, horizontal distance between approximated straight pectoral edge PQ and S_P is calculated as follows

The Cartesian slope intercept form of approximated straight pectoral edge is

$$X = m_y Y + b \tag{3}$$

where m is the slope of the line PQ with respect to horizontal y-axis and b as the x intercept. The two end points are $P\,(1, P_{wt})$ and $Q(P_{ht}, 1)$, and then, the slope m_y and x intercept b is given as

$$m_y = \frac{(1 - P_{ht})}{(P_{wt} - 1)} \tag{4}$$

$$b = 1 - m_y P_{wt} \tag{5}$$

The horizontal coordinate of points satisfying the straight pectoral muscle is obtained using the equation

$$Y(y) = \frac{X(x) - b}{m_y} \tag{6}$$

Then, the horizontal distance between points on PQ and S_P is

$$\mathrm{dst}(x) = |Y(y) - S_P(y)| \tag{7}$$

where S_P is the set of position of first nonzero pixels from right edge of mammogram when Avg_STPM is scanned horizontally and the plot is drawn in Fig. 2d.

Now to remove outliers, empirical threshold is calculated as

$$\mathrm{threshold} = \mathrm{med}(\mathrm{dst}) + \mathrm{IQR}(\mathrm{dst}) \tag{8}$$

where med is the median value and IQR is the interquartile range of dst.

If the horizontal distance dst(x) for pixel of S_P is greater than threshold then it will be treated as outliers as shown in Fig. 2c and removed from S_P. The filtered S_P is further smoothened using cubical polynomial and is plotted on original mammograms shown in Fig. 2e. (The coordinates of point P and Q are concatenated with S_P before polynomial modeling).

2.4 Nipple Detection Using Orthogonal Radon Transform

Mathematically, Radon transform of mammogram $I(x, y)$ is represented [18, 23] as

$$I(\rho, \phi) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(x, y)\delta(x\cos\phi + y\sin\phi - \rho)\mathrm{d}x\mathrm{d}y$$
$$-\infty < \rho < \infty, 0 \le \phi \le \pi \tag{9}$$

where $\rho = x\sin\phi + y\cos\phi$ is the straight line equation in polar form. If orthogonal Radon transform $g(\rho, \phi)$ and $g(\rho, \phi + 90°)$ is drawn, then the intersection of two Radon projections at some specific location has the ability to detect the position of nipple on mammogram. In the proposed method of nipple detection, in MLO view, the first step is to obtain the slope of the pectoral muscle m_x with respect to vertical x-axis, and angle of pectoral muscle with respect to vertical x-axis is given by

$$m_x = \frac{(P_{wt} - 1)}{(1 - P_{ht})} \tag{10a}$$

$$\phi = \tan^{-1}(m_x) \tag{10b}$$

Then, the two Radon transforms are taken at an angle equal to ϕ and $\phi + 90°$. The first Radon transform taken at the angle ϕ gives the location A (Loc(A)) and is marked as the last nonzero intensity on the projection as shown in Fig. 3a, b. Mathematically denoted as

$$\text{Loc}(A) = \{\text{lst}(g(x \sin\phi + y \cos\phi, \phi))|\forall g(\rho, \phi) \sim 0\} \tag{11}$$

where lst indicates the last element of the set having nonzero value on the projection. The tangent at point A is parallel to the approximated straight pectoral muscle edge. The equation of line passing Loc(A) $[x_1, y_1]$ is

$$(x - x_1) = m_a(y - y_1) \tag{12}$$

where $m_a = m_y$ (refer Eq. 4) is slope of the pectoral edge with respect to horizontal y-axis.

The second projection taken at $\phi + 90°$ gives the location of point B which is the highest peak point on the radon projection as shown in Fig. 3c, d.

$$\begin{aligned}\text{Loc}(B) = \{&g(x \sin((\phi + 90) + y \cos(\phi + 90), (\phi \\ &+ 90))|\forall g(\rho, \phi + 90°) = \max(g(\rho, \phi + 90°)\}\end{aligned} \tag{13}$$

The equation of line passing from Loc(B) $[x_2, y_2]$ is

$$(x - x_2) = m_b(y - y_2) \text{ where } m_b = \tan(\tan^{-1}(m_y) + 90°) \tag{14}$$

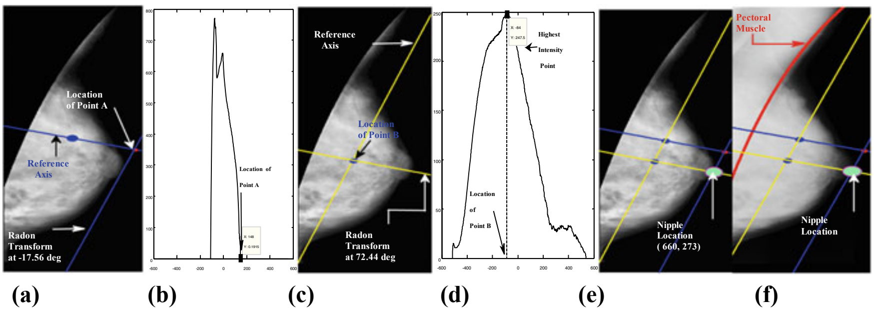


Fig. 3 Nipple detection on mdb053l with pectoral muscle delineated. **a** Radon transform at angle $\phi = -17.56°$ indicating location of Point A, **b** radon projection at angle ϕ indicating the last nonzero intensity as location of Point A. **c** radon transform at angle ϕ + 90° (i.e., 72.44°) indicating Point B on the reference axis. **d** Radon projection at ϕ + 90° indicating Point B as the highest intensity point. **e** Nipple position located on intersection of two orthogonal Radon transforms. **f** Registration of pectoral muscle and nipple position on original mammogram

The intersection of line drawn from location A and B is the location of nipple as shown in Fig. 3e, f. Similarly, for CC view, Radon transform is applied at angle 0° and 90° to find the location of the nipple, which is in majority of the cases at pinnacle of the breast.

2.5 Performance Evaluation

2.5.1 Performance Metrics for Pectoral Muscle

The accuracy of the proposed technique was estimated through comparison of segmented pectoral "mask" $P = (p_1, p_2, \ldots, p_x)$ with its corresponding "gold standard" $Q = (\mathrm{q}_1, \mathrm{q}_2, \ldots, q_y)$ created by experienced radiologist using quantitative measures. For each MLO view mammogram, the accuracy of proposed method is evaluated by three performance metrics and qualitatively by pectoral score index (PSI).

1. **Pixel Percentage of False Positive (FP)**:

A pixel outside the reference region but inside the acquired pectoral edge. Mathematically,

$$\text{Pixel Percentage of FP} = \frac{|P \cup Q| - |Q|}{|Q|} \times 100\% \quad (15)$$

where $P \in$ attained pectoral muscle region, $Q \in$ gold standard pectoral muscle.

2. **Pixel Percentage of False Negative (FN)**:

A pixel inside the reference region but outside the attained pectoral edge. Mathematically,

$$\text{Pixel Percentage of FN} = \frac{|P \cup Q| - |P|}{|Q|} \times 100\% \quad (16)$$

3. **Hausdorff Distance (HD)**:

The Hausdorff distance is defined by as

$$\mathrm{HD}(P, Q) = \max(h(P, Q), h(Q, P)) \quad (17)$$

where $h(P, Q) = \max_{p \in P} \min_{q \in Q} \|p - q\|$ and $\|.\|$ is the Euclidean distance.

4. **Pectoral Score Index (PSI)**:

Radiologist qualitatively determines accuracy of pectoral muscle using four point pectoral score index (PSI), as follows. If FP and FN > 15%: Not acceptable (1), if FP and FN < 15%: Tolerable (2), if 5% < FP and FN < 10%: Acceptable (3) and if FP and FN < 5%: Accurate (4). An index of 3 or more indicates a sufficient muscle registration. But pectoral muscle registration with tolerable limit can be used for nipple registration.

2.5.2 Performance Metrics for Nipple Registration

With respect to the visibility of the nipple as a "gold standard", radiologists interpret all the images in the database to evaluate the performance of the nipple detection algorithm. Let $N_m(x, y)$ be the result of automated nipple detection method and $N_r(c, d)$ be nipple location annotated by radiologist. The accuracy of algorithm is quantitatively evaluated by calculating the chess board distance between N_m and N_r.

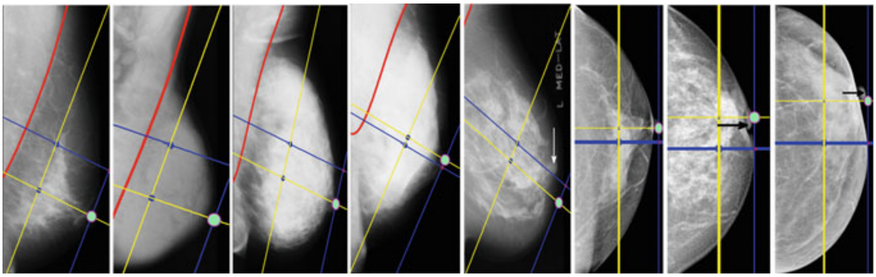
$$C_{\text{dist}}(N_m, N_r) = \max(|x - c|, |y - d|) \tag{18}$$

Radiologist determines accuracy of nipple registration using three point nipple score index (NSI). If C_{dist}> 10 mm: Not Acceptable (1), if 5 mm < C_{dist}< 10 mm: Acceptable (2) and if C_{dist}< 5 mm: Accurate (3).

3 Results and Discussion

3.1 Accuracy of Pectoral Muscle and Nipple Detection on MLO and CC Views

The method for the detection of pectoral muscle and nipple was applied to 266 MLO views from miniMIAS which includes many challenging images and 74 CC views from CICRI database. The algorithm seems to be very effective for pectoral muscle overlapped by dense tissues or muscles having very low distinction with breast tissues. For mammograms having large and small pectoral muscle as shown in Fig. 4a–e, the proposed method gives accurate registration of pectoral muscle and nipple except in mdb031**l**, and nipple is not detected correctly (distance between automated nipple and annotated nipple, C_{dist}> 12 mm). As many reported methods are unable to segment mammograms having muscle with heavy texture and edge obscured by highly dense breast tissues, for example, mdb062r (see Fig. 4b), but the proposed method can adequately segment it with PSI of 4. Here, the pectoral muscle is estimated with FN = 6.3% and FP = 2.8%. As mentioned by Kwok et al. [14], no muscle segmentation was produced for mdb179**l** due to fuzzy pectoral edge. With proposed textured-based surround suppression method as shown in Fig. 4d, pectoral muscle is registered with FP = 7.82% and FN = 10.37% and can be used



(a) mdb233**l** [4,4] **(b)** mdb062**r** [4,4] **(c)** mdb240**r** [3,4] **(d)**mdb179**l** [2,4] **(e)** mdb031**l** [4,1] **(f) cicri01** [3] **[g] cicri02** [2] **(h) cicri03** [2]

Fig. 4 Results of pectoral muscle and nipple detection on original MLO views from the miniMIAS and nipple detection on CC views from CICRI database. Red line indicates the muscle delineated using proposed method. Nipple detected automatically using proposed method is indicated by circle (light green face color). The qualitative indices are shown as [PSI, NSI] for MLO and [NSI] for CC views

to calculate the nipple position. In Hough [11] and Gabor [13], edge segmentation is unacceptable if FP and FN > 10%, while in Radon [12], it is unacceptable if FP and FN > 15%. For the proposed method, pectoral muscle detection is unacceptable if FP and FN > 10%.

The three performance metrics for pectoral muscle detection namely HD, FP and FN are tabulated in Table 1. From Table 1, proposed method outperforms Hough [11], Kinoshita et al. [12], Gabor [13] method, all of which were implemented on the same miniMIAS platform with different number of images. Table 2 shows range of chess board distance C_{dist} in terms of NSI, accuracy and average error for proposed method of nipple registration on both the database and is compared with Kinoshita et al. [12].

The proposed method fails to register the vertical pectoral muscle. Because a valid strongest straight line should intersect top and left edge of the oriented mammogram, the vertical pectoral muscle cannot satisfy the validation condition. For example, in mdb151**l**.

4 Conclusion

Precise detection of pectoral muscle and nipple registration on mammograms is demanding because of muscle texture, dense tissue overlap on muscle edge, image quality and nipple projections. A new accurate texture-based distance surround suppression method for automatic detection of the pectoral muscle on the MLO views and nipple registration using orthogonal Radon transform is proposed on the both the views. In two stage pectoral registration technique, textured Hough transform was applied to obtain straight line approximation of muscle followed by distant surround suppression and polynomial modeling to refine it to a curve. Without breast

Table 1 Comparison of proposed method with three reported methods

Author/methods	Database/(No. of images)	HD (in mm)	FP (%)	FN (%)	Images with (FP and FN) < 5%	Images with 5% < (FP&FN) < 10%	Images with 10% < (FP&FN) < 15%	Images with (FP and FN) > 15%	Accuracy (%)
Proposed/Textu-redHough + distant surround suppression	miniMIAS (266)	3.22 ± 1.66	1.54 ± 3.86	3.7 ± 3.88	214 (80.45%)	44 (16.55%)	6 (2.25%)	2 (0.75%)	97
Ferrari et al. [11] /Hough	miniMIAS (84)	7.08 ± 5.26	1.98 ± 6.09	25.19 ± 19.14	11.90%	9.52%	78.57%	NA	21.42
Ferrari et al. [13] /Gabor	miniMIAS (84)	3.84 ± 1.73	0.58 ± 4.11	5.77 ± 4.83	53.57%	26.19%	20.24%	NA	79.76
Kinoshita et al. [12] /radon	Local (540)	12.45 ± 22.96	8.99 ± 38.72	9.13 ± 11.87	28.9%	40.7%		30.4%	69.6