OPEN ACCESS

International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-4, November 2019

Automatic Vertebral Body Segmentation using Semantic Segmentation

Adela Arpitha, Lalitha Rangarajan

Abstract*: *Segmentation of vertebral bodies (VB) is a preliminary and useful step for the diagnosis of spine pathologies, deformations and fractures caused due to various reasons. We present a method to address this challenging problem of VB segmentation using a trending method – Semantic Segmentation (SS). The objective of semantic segmentation of images usually consisting of three main components - convolutions, downsampling, and upsampling layers is to mark every pixel of an image with a correlating class of what is being described. In this study, we developed a unique automatic semantic segmentation architecture to segment the VB from Computed Tomography (CT) images, and we compared our segmentation results with reference segmentations obtained by the experts. We evaluated the proposed method on a publicly available dataset and achieved an average accuracy of 94.16% and an average Dice Similarity Coefficient (DSC) of 93.51% for VB segmentation.

Keywords: Automatic Segmentation; Vertebral Body; Semantic Segmentation; CT.

I. INTRODUCTION

There has been a significant increase during the past few decades in the number of people suffering from spine problems. The significance of medical images has significantly risen as it is considered as a vital tool to visualize and uncover the internal body and obtain information about the human body for a plausible diagnosis and treatment if required. [1]. The medical image processing has attracted more and more attention from computer science experts as it has many purposes such as understanding image content, extracting the important features of the image, searching, visualizing, mining, detecting boundaries within medical images, segmentation, registration, classification and so on to obtain further diagnoses insight.

Image segmentation is necessary to extract information from an image. It is a computer vision task in which we label specific regions of an image according to what's being shown. Based on the problem to be solved, segmentation partitions an image into its fundamental parts or objects. When the region of interest has been confined in a specific application, segmentation stops. Spine image segmentation is important for automated localization, registration between two images, disease diagnosis, fracture detection and classification, surgical planning, performing surgery with computer-assisted surgery systems, and post-operative assessment.

One of the most difficult tasks in the process of segmentation is making it completely automatic. Today, medical imaging modalities generate large number of images in high resolutions that is tedious to be manually examined. Due to the requested high accuracy, the medical image segmentation is considered a challenging task. The disadvantage of manual segmentation is that, it is time consuming and the results are not really reproducible because the image interpretations by humans may vary significantly across interpreters [2]. Automated image segmentation could increase precision by eliminating the subjectivity of the clinician [3]. This drives the development of more efficient and robust problem-tailored image analysis methods for medical imaging.

In the human body, for the anatomical representation, spine is considered one of the most solid markers. It can readily be used as a landmark to the other organs in chest and abdomen. The spine, with its complex anatomy, consists of the spinal cord, 33 vertebrae, 23 intervertebral disks, and connecting ribs and nerves. The vertebrae are divided into 5 sections namely cervical (7 vertebrae), thoracic (12 vertebrae), lumbar (5 vertebrae), sacrum (5 fused vertebrae) and coccyx (4 fused vertebrae) that are stacked on top of one another with flexible intervertebral discs in-between each vertebrae. A vertebra is made up of two sections: an anterior segment that composes the VB and a posterior part or the vertebral arc that consists of spinous processes, pedicles, vertebral forearm and other anatomical regions [4] as shown in Fig.1. The VBs give stability to the spine and protects the spinal cord. For medical applications such as neurosurgery or oncology, detection and segmentation of VB is crucial. Thus, a robust algorithm that segments the spine into its constituent parts and creates a model (preferably 3D or 4D) of it which is naked eye perceivable, is need of the hour. Following segmentation, the shape and status of individual vertebrae can be determined which can promote further analysis.

There are different types of medical imaging modalities used for spine imaging such as plain radiographs, CT, Magnetic Resonance Images (MRI), Dual-Energy X-Ray Absorptiometry (DXA), Positron emission tomography (PET), bone scans, ultrasound (US), and nuclear medicine and each of them used for specific purpose. CT is preferable to all other imaging modalities in the detection of vertebral fractures and unstable injuries as it visibly enhances the bony structures required for such kind of diagnosis.

Manuscript published on November 30, 2019.
* Correspondence Author

Adela Arpitha*, Department of Studies in Computer Science, University of Mysore, Mysuru, India. Email: adelaarpitha23@gmail.com

Lalitha Rangarajan, Department of Studies in Computer Science, University of Mysore, Mysuru, India.
Email: lalitar@compsci.uni-mysore.ac.in

Even though CT images offer a high contrast between bone and soft tissues, automatic vertebra segmentation remains difficult. Factors like complex vertebral structure, images taken with restricted field of view causing lack of broader contextual information [1], unclear boundaries of the vertebrae, abnormal spine curves like scoliosis, adjacent bone regions, aging, degenerative joint disease alterations as well as the variety of pathological cases like osteoporosis encountered in an aging population, tumor, deformations, trauma, and fractures make automatic segmentation challenging [5]. In addition, the output of segmentation algorithm is affected due to intensity inhomogeneity and partial volume effect.

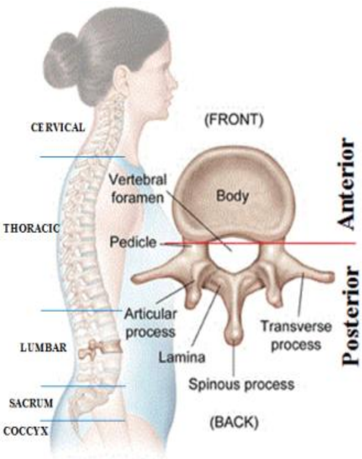

Fig.1. Structure of the spine and a representative vertebra in axial view with its named parts and sections

II. LITERATURE

For vertebra segmentation, there exist automatic and semiautomatic methods for both two-dimensional (2D) and three-dimensional (3D) image datasets. With the mode of data acquisition gradually shifting from 2D to 3D imaging, the volume of data per patient is largely increasing, making computer-aided detection desirable as an aide to the radiologist. Segmentation techniques based on statistical shape and appearance models, atlas based methods, deformable models, graph theory based methods, fuzzy logic based models, level sets, template based models and many more have been employed based on their own merits. A multiple center milestone study of clinical vertebra segmentation is presented in [6] with the objective to evaluate the performance of several state-of-the-art vertebra segmentation algorithms on CT images.

In [7], correlation is computed on an identified single Region Of Interest (ROI) that included a VB and a correlation algorithm automatically identifies ROIs consisting of VBs in the remaining images. The detected ROIs are adjusted to facilitate the segmentation process performed via graph based and line based algorithms. With minimal user assisted initialization, an intensity correction step is started in [8] to deal with the bias field artifacts. Appearance based VB probability maps guided the subsequent hybrid level set segmentation. In [9], a close triangular mesh targeted the VB surface that guaranteed the enclosure. Determining the location of intersection of each radial trajectory with the target surface, accomplished the segmentation. A novel statistical shape modeling framework for VB segmentation is developed by [10]. At first, a shape model using 2D Principal Component Analysis (PCA) is constructed. Secondly, a matched filter is used to detect the VB region. Next, graph cuts which integrates intensity and spatial interaction models is used for initial segmentation and finally initially segmented region is registered with the shape prior to obtain the final segmentation.

Superpixels have the ability to adapt themselves to the image structures, once their formation law follows the homogeneity of the image regions. This nature of superpixels have been utilized in the following few works. A multi-parametric ensemble learning technique using superpixels to automatically detect and segment lumbar vertebral VB is presented by [11]. Their method is designed with special emphasis given to robustness so that it performed well despite the inherent variation in scanning protocols. [12] took advantage of superpixels to reduce the image complexity for easy detection of each VB contour. As the boundary of each superpixel does not fit very well in the vertebra contour for some diseases or abnormalities, the Otsu's method as a post-segmentation step is proposed to divide the superpixels into smaller ones. The final segmentation is obtained through a region growing approach using manual points selection which produced masks of vertebrae. In [13], initially each vertebra with 3 point clicks in the mid-sagittal cross-section is initialized that enabled approximation of VB size and center. Next, a hybrid level-set algorithm combined a thresholded pre-segmentation within a cylindrical shape with intensity and gradient features resulting in a 3D contour, which could be visualized as an overlay in each cross-section or as a segmented 3D object.

Some of the drawbacks of machine learning are computational time incurred in processing high dimensional data, selecting optimal feature extraction technique depending on the data and more. These limitations have led to the advancement of deep learning era. Deep learning addresses these issues intelligently by focusing on the right features extracted directly from raw images without the need of any preprocessing, but with a little guidance from the developer. In the work of [14], vertebra segmentation has been approached as an instance segmentation problem. A fully convolutional neural network is combined with an instance memory that retained information of already segmented vertebrae. This network, using the instance memory, iteratively analyzes image patches to search and segment the first not yet segmented vertebra and concurrently each vertebra is classified as being partially or completely visible, thus excluding further analysis of partially visible vertebrae. [15] proposed a deep learning approach for simultaneous localization and identification of vertebrae. Features based on cubic intensity are drawn from image voxels. Localized points are polished from the detected region of vertebral column by thresholding locally continued by initializing a statistical multi-vertebrae model on the localized vertebrae. The VB of the model is registered to the image edges to obtain a segmentation using an iterative expectation maximization technique. Based on cascaded 3D Fully Convolutional Networks (FCNs) labels, [2] trained a regression 3D FCN to localize the lumbar region by fitting a bounding box. Later, a 3D U-net like FCN is developed that performs pixel-wise multi-class segmentation by mapping the localized vertebral region to its labels.

Though the performance of deep neural network is better than machine learning algorithms, their requirement of large amount of data and memory for training is still a key limitation. In order to overcome the limitations to some extent, we employed semantic segmentation to perform VB segmentation that does not require 1000s of data to learn. In a semantic segmentation network, each pixel in an image is classified, that results in a class-wise segmented image.

This paper is organized as follows. Section 1 introduces the need for VB segmentation along with a briefing of its anatomy. Section 2 attempts to identify some of the recent works on this domain. Section 3 explains the idea and adaptation of SS to our task. The customized model architecture is explained in Section 4 followed by discussions on experimental design and results in Section 5. Conclusion and future scope wraps up the paper.

III. Methodology

In this paper, we address automatic segmentation of VB from CT images in axial view. The dataset contains vertebral regions from thoracic to sacrum with varying Field Of View (FOV) and we consider all CT data for our study. As our task involves segmenting only the VB of the spine that is located in a certain area of a frame, there is no need to process the entire frame. Localizing the vertebral region will enhance the performance of segmentation. In medical images, most structures of clinical interest have a characteristic shape and anatomical location relative to other structures. Thus, for computational simplicity, based on the prior knowledge of the location of the vertebra when the spine is imaged in axial view and also keeping in par with the varying FOV in the dataset, a rectangular area of fixed size around the area of the spine in each training image is cropped automatically. The cropping area is selected such that it is big enough to tolerate the spatial variation of the spine structure across all the images and also avoids considering other bony areas around the vertebra. We then use semantic segmentation to perform VB segmentation. The idea of SS is to recognize and understand what's in the image at pixel level by associating a label or category with every pixel in an image. It allows changing the number of categories to classify the contents of an image. It can be 2 class or multiple classes. The reason for employing SS is that it cleanly detects objects that are irregularly shaped as is in the case of vertebrae.

In semantic segmentation, one of the common approaches is to create a SegNet that is based on Convolution Neural Network architecture (CNN). The segmentation networks are usually made up of three main components - convolutions, downsampling, and upsampling layers. The given input image is classified into one of the many predefined categories by the CNN. Our SS network takes input images and goes through convolution, batch normalization, Rectified Linear Unit (ReLU), maxpooling and then the image gets downsampled. It goes through the same processes and is downsampled again. Contrary to classification where the final outcome of the very deep network is the main significant thing, SS not only requires discrimination at pixel level but also a mechanism to project the discriminative features learnt at different stages of the encoder onto the pixel space for segmentation [16]. In order to do so we appended a reverse implementation of CNN. To ensure that the input and final images are of the same size, the upsampling and downsampling processes are performed the same number of times. In the end, a pixel classification output layer maps every pixel to a particular class. As a result, both our input and output are images. This structures an encoder-decoder architecture, which empowers semantic segmentation.

IV. Model Architecture

The proposed SS network architecture is as shown in Fig. 2. We implemented semantic segmentation architecture with the training done from scratch. Extensive experimentation is performed in working out the right number of filters, batch size, epochs and learning rate to be fixed for this particular data. Generalization of these values is not possible as it depends on the nature of data.

The proposed novel architecture for VB segmentation is as follows: during down sampling, the original localized images are re-sized to $32 \times 32 \times 1$ and two rounds of convolution, batch normalization and ReLU are performed. Finally max pooling is done. The convolution layer has 3x3 kernel size and padding set to 1. The batch normalization is a step between each layer where the output of the previous layer is normalized first before the next layer. The ReLU activation function is set which retains the positive part of its arguments. A second round of convolution, batch normalization and ReLU is performed again. Max pooling is performed with 2×2 size and stride set to 2 before being upsampled. In the upsampling process, the transposed convolution is used which performs the upsampling and filtering at the same time with cropping parameter set to 1. It is upsampled by 4. Next is the Softmax function that performs categorical probability distribution by analyzing and predicting the segmentation problem in identifying to which class the given image belongs to during testing, followed by the last pixel classification layer. These final set of layers as a combination, are responsible for making pixel classifications to predict the categorical label for each image pixel.

In training, input images are normalized and shuffled by every epoch. We set the epoch size to 100, learning rate to 1e-3, batch size to 16 and used stochastic gradient descent with momentum optimizer. All the above layers are stacked to complete the semantic segmentation network.

A. Implementation Details

The proposed work is implemented in MATLAB 2017a, trained on a 64 bit laptop with 2.4 GHz Intel(R) Core(TM) i7 CPU with 8GB CPU memory. For CPU implementation, the training stage is tedious.

V. Experimental Design, Results And Discussion

We trained and tested our method on the public dataset [17] obtained from the SpineWeb website. All scans of CT and ground truths are of size 512x512 pixel. In some data, if a VB is not fully scanned, manual segmentation of those half scans is not made available.

The dataset is divided into training set and testing set in the ratios 6:4, 7:3, and 8:2. To ensure fairness of the analysis results, the testing dataset that we used is different from the dataset that is used in the training set.

During training with parameters mentioned in Section 5, though the accuracies appeared to be going well, the network failed to properly segment the VB. This is because the spatial area of the VB in a given frame is small compared to the area of the

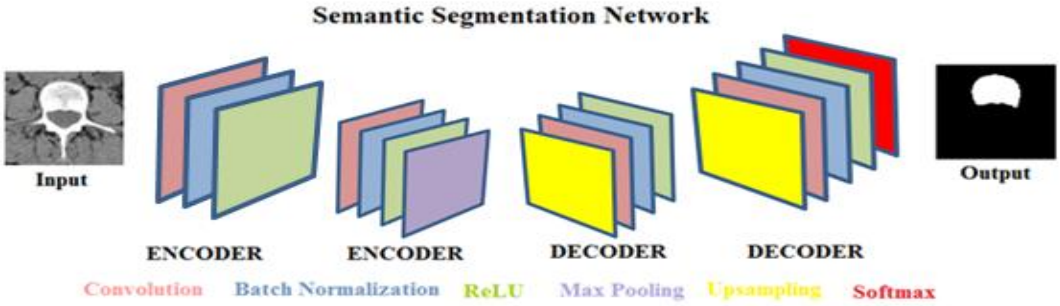

Fig. 2. Illustration of VB semantic segmentation architecture as an encoder-decoder. Downsampling is made up of convolutions, batch normalization, ReLU and maxpooling layers. Transpose convolution does the upsampling. The final decoder output is fed to softmax classifier for pixel wise classification. There are no fully connected layers.

TABLE I. Accuracy and DSC for different ratios of training and testing sets to segment VB and background classes

60:40	Accuracy (%)	DSC (%)
VB	92.2	91.92
Background	91.6	91.87

70:30	Accuracy (%)	DSC (%)
VB	93.7	93.32
Background	92.9	93.27

80:20	Accuracy (%)	DSC (%)
VB	96.6	95.26
Background	93.8	95.13

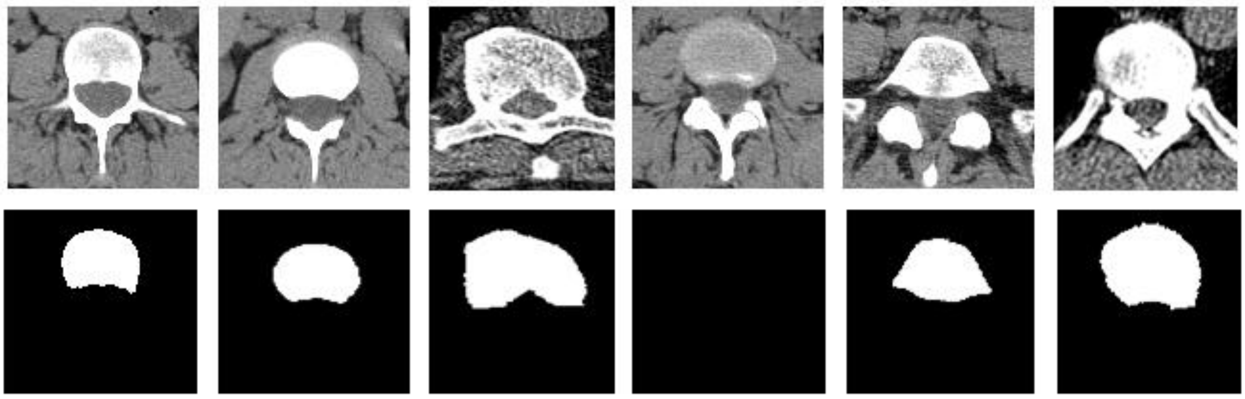

Fig. 3. Representative images from the test set (row 1) and their successful VB segmentation outputs (row 2)

background and hence the network eventually learned to classify the background class better than the VB itself. The poor results are due to class imbalance. Class imbalance favors the dominant class in the learning procedure. To fix this, class weighting to balance the classes are used which fared better than increasing the number of epochs used for training or modifying the network. To increase the weight given to under-represented classes, inverse frequency weighting is used where the class weights are the inverse of the class frequencies. The last layer (pixel classification layer) is updated to use inverse class weights and the network is trained again.

For quantitative analysis, means of accuracy, and DSC for each class in the data set is chosen. The results of three different ratios of training and testing sets to segment VB and background classes are as shown in Table 1.

A few representative images are selected from the results of the testing set. The manually labeled images of each subject by the experts are used as gold standards. Fig. 3 presents the results obtained.

During localization, though most of the surrounding internal organ structures and scanner artifacts are excluded, some parts of the ribs in the thoracic region and parts of wings of illium in the lower lumbar and sacral regions were still included in the ROI frame. Our method partially failed to completely segment VB in cases when proximity of ribs and wings of illium is too close to the vertebra. Multi class semantic segmentation of the surrounding bony structures would further improve the performance of the model.

Retrieval Number: D8584118419/2019©BEIESP
DOI:10.35940/ijrte.D8584.118419
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

VI. CONCLUSION AND FUTURE WORK

A novel semantic segmentation network to segment only the VB is proposed and the model performance is evaluated by measuring accuracy and DSC with varying sizes of training and testing sets. The model has shown robust and good accuracy, in spite of the various clinical conditions and varying FOV present in the dataset. It distinguishes neighboring bone structures, varying VB shapes from thoracic to sacrum regions and could be extended to other segmentation tasks that have difficulties with repeated similar shapes or close distances between structures. No pre-trained networks have been used for preliminary steps of training. In future research, we intend to assess our method using a deeper network and broad training data, to improve the model performance. We also intend to use these segmented VB outputs to further automatically classify VB fracture types and extend to analyze VB in 3D.

ACKNOWLEDGMENT

The first author would like to thank University Grants Commission’s (UGC) for supporting this research work through Maulana Azad National Fellowship (MANF).

REFERENCES

1. Alaa Atef Darwish, Mohammed A-M. Salem, Doaa Hegazy, and Hala M. Ebeid, "Vertebrae segmentation techniques for spinal medical images," *In Intelligent Computing and Information Systems (ICICIS)*, 2015 IEEE Seventh International Conference on, pp. 110-115. IEEE, 2015.
2. Rens Janssens, Guodong Zeng, and Guoyan Zheng, "Fully automatic segmentation of lumbar vertebrae from CT images using cascaded 3D fully convolutional networks," In *Biomedical Imaging* (ISBI 2018), 2018 IEEE 15th International Symposium on, pp. 893-897. IEEE, 2018.
3. Malinda Vania, Dawit Mureja, and Deukhee Lee, "Automatic Spine Segmentation using Convolutional Neural Network via Redundant Generation of Class Labels for 3D Spine Modeling," *arXiv preprint* arXiv:1712.01640 (2017).
4. Adela Arpitha, Lalitha Rangarajan, and K. S. Raghunandan, "Automatic Vertebral Body Fracture Classification through PCA based Angle Estimation," *International Journal of Computer Application* (2250-1797),Volume 8–volume 1, 2018
5. Jean-Baptiste Courbot, Edmond Rust, Emmanuel Monfrini, and Christophe Collet, "Vertebra segmentation based on two-step refinement," *Journal of computational surgery* 4, no. 1 (2016): 1.
6. Jianhua Yao, Joseph E. Burns, Daniel Forsberg, Alexander Seitel, Abtin Rasoulian, Purang Abolmaesumi, Kerstin Hammernik, Martin Urschler, Bulat Ibragimov, Robert Korez, Toma ˇz Vrtovec, Isaac Castro-Mateos, Jose M. Pozo,Alejandro F. Frangi, Ronald M. Summers, Shuo Li, "A multi-center milestone study of clinical vertebral CT segmentation." *Computerized medical imaging and graphics* 49 (2016): 16-28.
7. Sewon Kim, Won Bae, Koichi Masuda, Christine Chung, and Dosik Hwang, "Semi-Automatic Segmentation of Vertebral Bodies in MR Images of Human Lumbar Spines," *Applied Sciences* 8, no. 9 (2018): 1586.
8. Georg Hille, Sylvia Saalfeld, Steffen Serowy, and Klaus Tönnies, "Vertebral body segmentation in wide range clinical routine spine MRI data," *Computer methods and programs in biomedicine* 155 (2018): 93-99.
9. Shuang Liu, Yiting Xie, and Anthony P. Reeves, "Automated 3D closed surface segmentation: application to vertebral body segmentation in CT images," *International journal of computer assisted radiology and surgery* 11, no. 5 (2016): 789-801.
10. Ahmed Shalaby, Melih Aslan, Hossam Abdelmunim, and AlyA Farag, "2d pca-based shape prior for level sets segmentation framework of the vertebral body," *In Biomedical Engineering Conference* (CIBEC), 2012 Cairo International, pp. 134-137. IEEE, 2012.
11. Bilwaj Gaonkar, Yihao Xia, Diane S. Villaroman, Allison Ko, Mark Attiah, Joel S. Beckett, and Luke Macyszyn, "Multi-parameter ensemble learning for automated vertebral body segmentation in heterogeneously acquired clinical MR images," *IEEE journal of translational engineering in health and medicine* 5 (2017): 1-12.
12. Paulo Duarte Barbieri, Glauco Vitor Pedrosa, Agma Juci Machado Traina, and Marcello Henrique Nogueira-Barbosa, "Vertebral body segmentation of spine MR images using superpixels," In *International Symposium on Computer-Based Medical Systems*, 28th. Institute of Electrical and Electronics Engineers–IEEE, 2015.
13. Georg Hille, Sylvia Glaßer, and Klaus Tönnies, "Hybrid level-sets for vertebral body segmentation in Clinical Spine MRI," *Procedia Computer Science* 90 (2016): 22-27.
14. Nikolas Lessmann, Bram van Ginneken, Pim A. de Jong, and Ivana Išgum, "Iterative fully convolutional neural networks for automatic vertebra segmentation," *arXiv preprint* arXiv:1804.04383 (2018).
15. Amin Suzani, Abtin Rasoulian, Alexander Seitel, Sidney Fels, Robert N. Rohling, and Purang Abolmaesumi, "Deep learning for automatic localization, identification, and segmentation of vertebral bodies in volumetric MR images," In *Medical Imaging 2015: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 9415, p. 941514. International Society for Optics and Photonics, 2015.
16. https://medium.com/nanonets/how-to-do-image-segmentation-using-deep-learning-c673cc5862ef Retrieved on August 22, 2018
17. Melih S. Aslan, Ahmed Shalaby, and Aly A. Farag, "Clinically desired segmentation method for vertebral bodies," In *Biomedical Imaging* (ISBI), 2013 IEEE 10th International Symposium on, pp. 840-843. IEEE, 2013