Digitization of ECG Paper Records using MATLAB

Priyanka R. K. Shrivastava, Shraddha Panbude, Geeta Narayanan

Abstract— Electrocardiogram (ECG) is the most important and widely used method to study the heart related diseases. The detailed study of ECG graph by the medical practitioner helps him to understand and identify the condition of the heart. Based on the information retrieved from the ECG graph the patient can be given proper treatment. The person having a medical history of heart ailments will have to maintain a record of all the ECG papers for timely analysis and diagnosis of the diseases. This process requires large storage space and extensive manual effort. The conventional technique of visual analysis to inspect the ECG signals by doctors or physicians are not effective and time consuming. Therefore, an automatic system which involves digital signal integration and analysis is required. In this study a MATLAB-based tool is being designed to convert electrocardiography (ECG) information from paper charts into digital ECG signals. Here we develop a method that involves processing of ECG paper records by an efficient and iterative set of digital image processing techniques for the conversion of ECG paper image data to time series digitized signal form, resulting in convenient storage and retrieval of ECG information. In addition, this tool can be used to potentially integrate digitized ECG information with digital ECG analysis programs and with the patient's electronic medical record.

Keywords- ECG, Compression, Segmentation, image retrieval, Digitization. Laplacian filtering.

I. INTRODUCTION

Electrocardiography (ECG) has existed for more than a century, and digital ECG has existed for more than two decades [1]. The electrocardiogram (ECG) is a graphical representation of the electrical activity of the heart and is obtained by connecting specially designed electrodes to the surface of the body. Normally, ECG has to be printed on a thermal paper for further physical inspection by medical practitioner [4]. The measurement and evaluation of 12-lead electrocardiograms (ECGs) printed on paper is an integral part of modern clinical trials [2]. This leads to large volume of ECG reports that requires a tedious and error prone process of examining and retrieval of the same [4]. Digital signals allow very high signal processing capabilities, easy storage, transmission and retrieval of information [5]. Hence Digital ECG offers the opportunity to rapidly calculate ECG attributes such as vector angles and waveform intervals previously laborious and imprecise with paper tracings [1].

In addition, the image processing techniques are able to further transfer the ECG paper charts, web-based and DICOM-based ECG screenshots, and PDF files to the digital time database for data extraction, compression, archive, and storage [3]. However, most long-term retrospective cardiac patient databases, which span over a duration of two decades or more have paper ECG data. Therefore, a system that can re-digitize to a universal format will be an added utility. Furthermore, such tools would allow for comparison of a patient's current ECG with a baseline ECG, enhancing confidence in the diagnosis of acute changes. In this research work, we will design a Matlab (MathWorks Inc., Natick, MA)-based tool for digitization of paper-ECG data. It will utilize various image processing techniques involves scanning of a paper ECG to an image, thresholding for graphic grid removal, and detecting the ECG signal contour followed by digitization. The motivation for this work is that this system will allow rapid calculation of ECG features. Apart from rapid calculations, digitizing paper ECG can allow the integration of ECG data with the patient's Electronic Medical Record (EMR) [1]. Academic research will also indirectly benefit from these efforts; indeed, there still exist today many large databases collected on paper ECGs that must be converted to digital format to be efficiently archived [6].

II. REVIEW OF RELATED WORK

The 12-lead electrocardiogram (ECG) is widely used in clinical research to determine primary and secondary end points and to assess the effect of drug-induced relevant changes. Despite a technology that permits storage of raw data in digital format in many years, many ECG analysis are still performed on paper printouts, either directly on paper copies or with the support of on-screen caliper methods applied to the scanned printout converted to a digital image format [1]. Recently, the Food and Drug Administration has launched an initiative intended to recommend to the pharmaceutical industry the use of digital annotated ECGs for the submission of new drug applications [2]. Although intended to provide a more efficient way to review submitted data, this effort has also triggered the interest toward algorithm-based and semi automated ECG analysis. The digital ECG format chosen by the Food and Drug Administration has been already adopted by several manufacturers, thus facilitating the exchange and the usage of digital ECGs in the pharmaceutical arena. More importantly, the new regulatory guidelines will magnify the need for efficient and widely applicable tools for the conversion of paper ECGs to digital ECGs, particularly for retrospective studies.

Academic research will also indirectly benefit from these efforts; indeed, there still exist today many large databases collected on paper ECGs that must be converted to digital format to be efficiently archived. The task of actually deriving a digital ECG from a paper printout has been already approached, although it was generally limited to research applications [3,4] and often specifically designed to the execution of well-defined studies [5-7]. Moreover, the methods used were often based on commercial products the main objectives and contexts of which were not in the ECG arena and that could not be optimized and tuned for the task of deriving an ECG waveform from a grid-supported image.

III. PROBLEM DEFINITION AND PROPOSED WORK

Normally, ECG has to be printed on a thermal paper for further physical inspection by medical practitioner. This leads to large volume of ECG reports that results in a tedious and error prone process of examining and retrieval of the same. A cardiac patient with history of heart ailments will always have to carry and maintain a bulk of such ECG reports while visiting a physician for consultation. Further these bulk ECG records are very hard to be shared among doctors. Even in hospitals due to lack of space and difficulty in maintaining records a huge amount of data is discarded time to time. Our goal is to design and evaluate a system that transforms a paper 12-lead ECG into a stream of digital data that can be interpreted and measured by a commercially available automated ECG analysis program. It is to be noted that printed ECG reports includes some characters as well as annotations, while taking an image of such ECG strips it is desirable to remove these printed characters. Efficient noise removal and image enhancing algorithms are also required so as to increase overall digitization process accuracy. Digitizing the ECG paper signal will help to solve storage problem. It is also cost effective when comes to sharing and diagnosis. The overall algorithm is shown below.
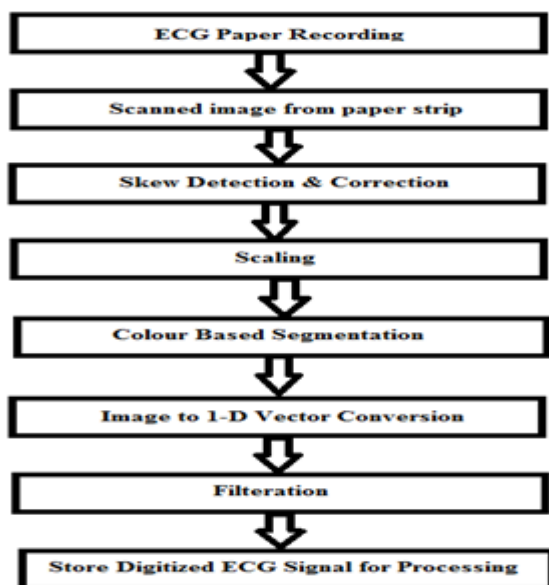

Fig. 1 Algorithm

A) Scanning

ECG paper recordings need to be scanned. Scanning resolution can be 600/300/200 dpi (dots per inch). Preferred algorithm for the image compression is JPEG. The image can also be obtained from a digital camera used in mobile phones. Figure 2 represents a scanned ECG paper recording.

B) De-Skewing

Skew is a common phenomenon that can appear in a scanned image. Skewing rotates the image to an angle, resulting in a rotated image. Hough Transform has been used to de-skew image. Skew angle is calculated using the background grid lines of scanned ECG image.

C) Image Enhancement

This step enhances the ECG image by making the signal lines sharper. Laplacian filtering is applied for making background noise lighter than the main ECG signal. This is followed by setting a threshold. Threshold value is chosen by comparing between noise pixels and pixels representing actual ECG signal. If ECG signal pixel values are close to the threshold then pixels will be made darker by subtracting a fix value. Whereas the noise pixels values, close to the threshold, will be made lighter by adding a fix value. Therefore, the resulting image will contain distinct ECG signal in the image.

D) Color Based Segmentation

This step is applied to remove the background grid of an ECG. Background grid is usually of lighter shade of color than the actual signal waveform. In this step ECG image is processed column by column. The darkest pixels are extracted in each column and are replaced by pure black pixel and rest of the pixels as pure white. It produces a binary image. However this may lead to extraction of undesired printed character as well. Figure 2 (b) shows the output of color based segmentation; it has some printed characters also.

E) Region Based Segmentation

This step removes the isolated pixels that do not represent the signal (printed characters). This step comprises of the following steps

Step 1: Remove the frame of the image that is border of the input image.

Step 2: Scan the input ECG image column by column.

Step 3: Repeat for each column:

a. Extract the black pixels (pixels that represent the signal).

b. In case of the presence of isolated pixels (not likely to happen with the color based segmentation), delete them.

c. If there is more than one black pixel:

i . Algorithm first checks the previous and the next columns, if they contain only one pixel, then for the current column select the pixel that likely goes with the flow of the signal pattern.

ii. If the next column contains the same problem of having more than one black pixel, then for the current column select the middle pixel that lies exactly in the middle of the black pixel pattern.

Step 4: Save the index of the picked pixel. The process of choosing the middle value in step 2 is performed by the following algorithm:

Step 4.1: get the centers of different regions

Step 4.2: separate regions and compute the centroid of each region

1a: Set threshold value of region separation If current_point > previous_asmple + threshold_value (if true, then current region is separated)

add center(average of first and last point in the region)

else (we need to check whether it is a last region)

add center(average of first and last point in the region)

Step 5: Pick centre point value which is the closest to the previous column pixel index. For this centres calculated in step 2 are used. We find the center that has the minimum difference with the previous column pixel index and that is our next center.

Figure 2 (c) shows the output of this step, the red line is actual detected signal. It is to be noted here that the printed characters (patient id….) are still there but do not have any significance over the actual detected signal (red line).
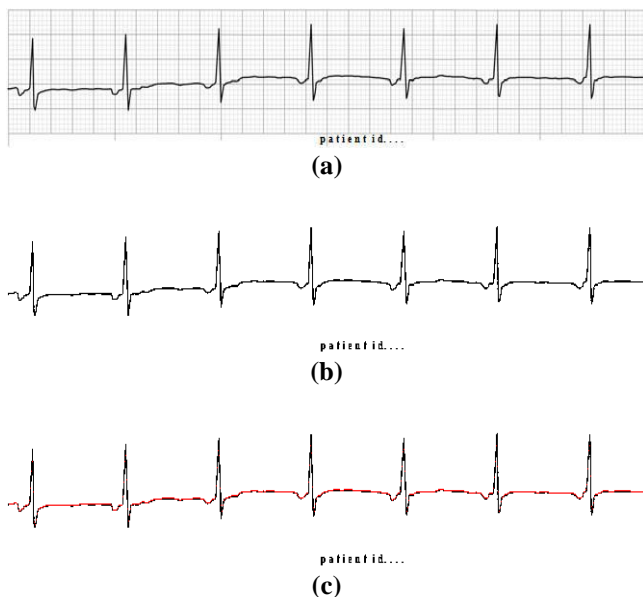

(a)

(b)

(c)

Fig. 2 Steps in Algorithm

(F) Signal Representation

Since each column is governed by a single pixel so the signal is represented by the corresponding column pixel index. Finally the signal is saved as a txt file

(G) Median Filtration

Actual ECG signal recorded by an ECG machine can be degraded due to presence of noise. Various sources of noise can be [10]:

☐ Item Power Line Interface (50Hz or 60Hz noise from power lines)

☐ Baseline wander (low frequency noise)

☐ Muscle noise

☐ Other interference (i.e., radio frequency noise from other equipment)

Noisy signal can result into misdiagnosis so it is desirable to remove noise from the actual ECG signal as much as possible. A noisy ECG signal image is taken and the performance of various filters including median filter, FIR or IIR type of filters are compared and it is found that median filter has best result both by visual inspection and by quantitative measurement (by Signal to Noise Ratio).

IV. CONCLUSION AND FUTURE WORK

An efficient method for extraction and digitization of ECG signal from various sources such as thermal ECG printouts, scanned ECG and captured ECG images from devices is proposed. The methodology produced a reasonably accurate waveform that is free from printed character and as tested through heart rate, qrs width and stability calculations. Further work is in progress to enhance the overall efficiency of the method and also developing a fuzzy based expert system that can assist a doctor in diagnosis and generating automatic diagnosis reports as well.

REFERENCES

1. "Novel Tool for Complete Digitization of Paper Electrocardiography Data" Lakshminarayan Ravichandran, Chris Harless, Amit J. Shah3, Carson A. Wick , James H. Mcclellan, And Srini Tridandapani.
2. "New Method for Digitization and Computerized Analysis of Paper Recordings of Standard 12-Lead Electrocardiograms" WT Lawson, GS Wagner, RS Startt-Selvester, GA Ybarra Department of Cardiology.
3. "Image Processing on ECG Chart for ECG Signal Recovery", TW Shen, TF Laio Tzu Chi University, Hualien, Taiwan, Computers in Cardiology 2009;36:725−728.
4. "ECG Paper Records Digitization through Image Processing Techniques", Deepak Kumar Garg, Diksha Thakur, Seema Sharma, Shweta Bhardwaj, International Journal of Computer Applications (0975 – 888) Volume 48– No.13, June 2012.
5. "Converting ECG and Other Paper Legated Biomedical Maps into Digital Signals", A.R.Gomes e Silva, H.M. de Oliveira, and R.D. Lins, Federal University of Pernambuco- UFPE, Signal Processing Group.
6. "ECGScan: a method for conversion of paper electrocardiographic printouts to digital electrocardiographic files", Fabio Badilini, PhD,T, TanjuErdem, PhD, WojciechZareba, MD, Arthur J. Moss, F. Badilini et al. / Journal of Electrocardiology 38 (2005) 310– 318.
7. "Electrocardiogram Display Data Capturing and Digitization Based on Image Processing Techniques" Lai Khin Wee, Eko Supriyanto, New Aspects Of Applied Informatics, Biomedical Electronics & Informatics And Communications, ISBN: 978-960-474-216-5.
8. Gonzalez RC, Woods R.E, "Digital Image Processing", 3rd Ed., Addison-Wesley, Reading, MA, Longman, New York, 2008.
9. Gonzalez RC, Woods R.E, "Digital Image Processing using Matlab", 3rd Ed., Addison-Wesley, Reading, MA, Longman, New York, 2008.

Priyanka R. K. Shrivastava, pursuing ME(EXTC) from Vidyalankar Institute of Technology, completed BE(Biomedical) in 2010 from Mumbai University. Research interests include the areas of development of algorithms of signal and image processing, Biomedical Instrumentation, WBAN and Network Security .

Shraddha Panbude, Assistant.Professor, EXTC at Vidyalankar Institute of Technology. Research interests include the areas of Image Processing and Wireless Communication.

Geeta Narayanan, Associate Professor, ETRX at Vidyalankar Institute of Technology. Research interests include the areas of Power Electronics, Control Systems,Robotics and Instrumentation.