VOL. 10, NO. 18, OCTOBER 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences
©2006-2015 Asian Research Publishing Network (ARPN). All rights reserved.
www.arpnjournals.com

IMAGE BASED HEART MURMUR CLASSIFICATION USING HAMMING DISTANCE MEASURE

Suresh. D[1], B. Ramamurthy[2], S. Ravi[3] and B. C. Premkumar[3]
[1]Department of Electronics, S. K. University, Anantapur, India
[2]Department of Electronics, S. K. University, Anantapur, India
[3]Department of ECE, M. G. R. Educational and Research Institute, Chennai, India
E-Mail:suresh_aps@yahoo.com

ABSTRACT

The heart murmur signal is an acoustic signal and presents artefacts when its path is tracked in a non-optimal manner. The sound signal is very similar to the light wave and takes the shortest path when travelling from one region to another. The heart murmur occurs inside the heart chamber (mechanical model of the heart) but is sensed external to it using a microphone or stethoscope. This difference in the distance between the occurrence point and the sensing point itself is an artefact at the sensing point. Thus, further transforms acting on the sensed signal to extract the features out of it is like an invasive mathematical operation on the data and increases the artefacts. In this paper, non-invasive mathematical approach for analysing the data from the sensing point using image based approach is presented. Our approach of content extraction in this paper searches for the content in the input rather following traditional text mining techniques and translational issues.

Keywords: content extraction, heart murmur, hamming distance, image processing and phonocardiogram.

INTRODUCTION

The requirement of preserving the richness of the data from one node to another requires a non-invasive model to handle the data. This implies just observing the data, but not uses lossy mathematical transformations to it. The heart murmur signal is one category, where feeble but rich information is available from the heart chamber to the sensing unit. From basic reflection theory, the richness in data is maximized if the shortest path theory by default travelled by the acoustic wave from the originating to the sensing node. Any lossy mathematical operation on this data, to extract the feature vectors is equivalent to bending the acoustic wave and making it to reach the sensing node in multiple non-optimal ways. In terms of accuracy, this implies unintentional addition of artifacts. In this context, the total distance finding problem is the problem of looking for short segments that are over-represented among a set of long sequences. Deterministic Combinatorial approaches based on word statistics and probabilistic approaches based on local multiple sequence alignment are the two known approaches for finding similar patterns. The total Hamming distance is the sum of the Hamming distance between the original string and each closest match. Then this process is repeated for every possible distinct string of length *l* (which is 4_ strings) to find the string that has the minimum total Hamming Distance. The string with the minimum total Hamming distance is the "Median String," which, interestingly enough, is the same as the consensus string. So, once the Median String has been found, the original method using the profile matrix can be run backwards to find the closest matching string in each image sequence-the closest match being the one with the smallest Hamming distance as compared to the Median String.

PCG HEART SIGNAL

Many diseases of the heart cause changes in the heart sounds and additional murmurs before other signs and symptoms appear. Heart sound analysis by auscultation is the primary test conducted by physicians or cardiologists to assess the condition of the heart. A classical example of acoustic signals is phonocardiography (PCG) signal - an acoustic recording and plot of heart sound and murmur signals as function of time. These sounds provide an indication of the heart rate, rhythmicity, noise-like features, and efficiency of heart valves and provide valuable clinical information. They also give clinical information regarding effectiveness of blood pumping and heart valve action. Auscultation is a process of listening to heart sounds and murmurs using acoustical stethoscope and making clinical decisions related to heart diseases, in particular the valvular diseases. Sample PCG signals are shown in Figure-1.

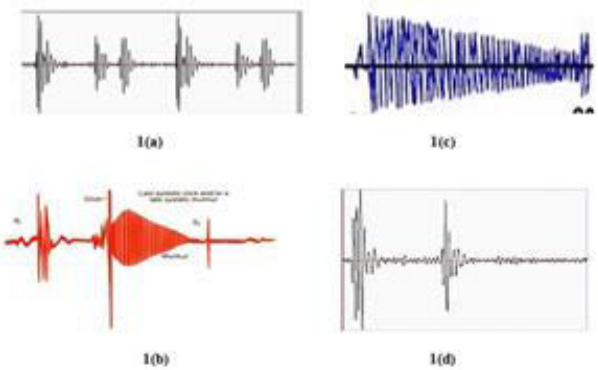


Figure-1. Sample of PCG signals.

©2006-2015 Asian Research Publishing Network (ARPN). All rights reserved.

METHODOLOGY

- The low frequency heart murmur signal is first considered as overlapping frames of variable length.
- These frames are trained to maximize the mutual information (MMI) and minimize false classification by maximizing the Hamming distance among the vectors.

CONTENT EXTRACTION STUDIES

Pattern matching

Pattern matching is used to compute similarities between images and prototypes, which can be treated with transformations like translation and rotation. Pattern matching techniques are vital in analyzing document images and web documents. Documents in general are scanned into a binary image, and then any operations are performed, like page segmentation, character identification, erosion, dilation of binary image morphology. They can be used for extracting pixel aggregations. We can use them in image processing and for subsequent analysis. These are in general translational invariant. We can use them for treating both foreground and background simultaneously irrespective of connected component analysis. Further, there exist a variety of methods for controlling the noise immunity of these operations. One of the most important uses of pattern matching is in the analysis of character shapes. For binary input, the result of image processing can be either binary or gray scale images. Binary results are much faster to compute, but they contain less information. Even for binary output, the internal operations can be integer or Boolean. For integer operations, such as convolution and threshold convolution, noise immunity in general is attained in performing costly arithmetic operations on each of the pixel.

Distance measures and alignments

The Hamming distance calculation rigidly assumes that the i^{th} symbol of one sequence is already aligned against the i^{th} symbol of the other. However, it is often the case that the i^{th} symbol in one sequence corresponds to a symbol at a different-and unknown-position in the other. To differentiate the falling and sitting scenarios from the captured video frame, the concept of total distance is used. It is appropriate to use total distance and make use of the three operations delete, insertion and substitution to map the received frame with the reference frame.

Methodology

a) Measure the total distance of the received frame with respect to the reference frame. This is defined as the number of operations required to match the pixels of the received frame with the reference frame within a threshold value.

b) Compute the minimum of the total distance ed1 and ed2 and perform decision.

$$\therefore ed_1 = \sum_{i=1}^{3}\sum_{j=1}^{3}\left[P_{i/p}(i,j) - P_{ref_1}(i,j)\right]$$

and

$$\therefore ed_2 = \sum_{i=1}^{3}\sum_{j=1}^{3}\left[P_{i/p}(i,j) - P_{ref_2}(i,j)\right]$$

and $\min(e_{d_1}, e_{d_2})$Vsub-blocks decide the closest match of the captured frame with the individual reference frames.

Illustrations

Let the (i,j) sub blocks of k^{th} frame is given below

128	**127**	**132**
119	**118**	**128**
121	**122**	**126**

3*3 sub block of k^{th} frame
Let the sub block of k^{th} reference frame for scenario-1is given below:

128	**127**	**131**
119	**118**	**127**
121	**122**	**12**

Similarly, for reference frame -II, the pixel values for the sub block is given below:

112	**125**	**119**
122	**127**	**117**
121	**118**	**126**

Algorithm

For k^{th} scanned image

1) Read i^{th} block values
2) Compute Hamming distance for median string of length 'l'
3) Compute the minimum Hamming distance of (2)
4) Repeat (1) to (3) for $(i+1)^{th}$ block to cover all sub-blocks of the scanned image
5) Obtain the vector representing the minimum Hamming distance of all the NxN sub-blocks of the kth scanned image
6) Repeat steps (1) to (5) for other images representing the same context.

DECISION ALGORITHM

From the individual vectors for the scanned images obtained using the Total distance finding algorithm, a decision threshold is chosen to determine a correlated consensus value. As mentioned in Table-2(Refer Annexure), Poor correlation indicates identical number of zeroes in the sub blocks between two pixel-maps. For example, using these different classes of PCG signals are packed in separate bins.

©2006-2015 Asian Research Publishing Network (ARPN). All rights reserved.

Table-2 Comparison of Δx with x

From Table-2 (Refer Annexure), the complementary total distance $\widetilde{e_d}$ is computed and the image with $\widetilde{e_d}$ above a set threshold (in this work 5) is grouped into the same bin. The sub-block-I 'x' values of all three images are shown below:

$Image_1$Subblock – I 'x' value = $\{0,0,-3,-176,0,0,0,32\}$
$Image_2$Subblock - I 'x' value = $\{0,-163,57,-141,49,45,89,89\}$
$Image_3$Subblock – I 'x' value = $\{0,0,-3,-140,0,0,7,140\}$

The calculated complementary edit distances are

$\widetilde{e_d}$ Between $Image_1$& $Image_2$ = 7
$\widetilde{e_d}$ Between $Image_2$ & $Image_3$ = 3

IMPLEMENTATION OF THE WORK

Multiple scanned images of uniform size with either matching or non-matching content are given as input. In the first approach, the pixel values of individual images are treated as pair wise distance between points and the positions of those points are reconstructed using the partial digest problem solution approach. From the knowledge of set of distances between every pair in a sub block of kxk pixels [in this case 'k' is chosen as 5], the actual distance is extracted i.e., 'x' from Δxin the second approach, the matching based position vector extraction is done and using the total distance metric, the decision with respect to the content, bin packing is done. Both the approaches show striking similarity and the elements of the bin are formed with high accuracy.

Partial digest algorithm

Using PDA, one set 'x' is derived for a'Δx 'of length 'L' for each of the sub-block of size K x K, for every image. The methodology is given below:

Image
↓
Extract Sub-blocks
↓
For every Sub-block extract Δx
↓
For every Δx obtain 'x'

For a given Δx of length 'L', the size of 'x' is determined from $\frac{\lceil size(x) size(x-1) \rceil}{2} = L$.

In this work, 'L' is chosen as 25, so that $\lceil size(x) \rceil = 8$ where $\lceil y \rceil$ is the nearest integer exceeding 'y'

PARTIAL DIGEST (L)
Width ← Max element in L
Remove (width, L)
X ← {0, which}
Insert (L, X)

Insert (L, X)
if L is NULL print X
return
y← Max element in L
if$\Delta(y, X) \subseteq L$
Add y to X and remove lengths $\Delta(y, X)$ from L

Insert (L, X)
Remove y from X and add length $\Delta(y, X)$ to L
if$\Delta(width - y, X) \subseteq L$
Add width – y to X and remove length $\Delta(width - y, X)$ from L
Insert(L, X)
Remove width -y from X and add length $\Delta(width - y, X)$ from L
Return.

Median string problem

Given a set of sub-blocks and a search/query median string

Input: At x n matrix sub-block, and l, the length of the pattern to find.

Output: At string v and l that minimizes $TotalDistance\ (v, image)$over all strings of that length.

Heart murmur classifier algorithm

Let 'L' be the length of the Median String and 'M' be the size of the sub-block chosen. [More L=8, M=25] Let the positions of the elements in sub-block be M_0M_1

a) Find Hamming distance between the median string size of 'L' and the elements of 'M' with starting element M_0
b) Repeat step (a) for elements of 'M' but with starting element = M_1. Repeat the steps above for successive starting position of 'M'
c) Obtain the position value corresponding to the minimum Hamming distance out of the above

For illustration:
Let Median-string values be $X_0, X_1, \ldots, X_7$
Then
Step 1 Compute HD (X_0,M_0;…….X_7,M_7)
Step 2 Compute HD (X_0,M_1;…….X_7,M_8)
Step 3 Compute HD ( )
.
.
Compute HD (X_0,M_{18};…….X_7,M_{25})

RESULTS AND DISCUSSIONS

Classified output for first heart murmur signal

The acquired vector length is 17 and for the seventeen positions frame wise the Hamming distance is computed. The non-matching positions with the different feature vectors is shown in Figure-2.

©2006-2015 Asian Research Publishing Network (ARPN). All rights reserved.

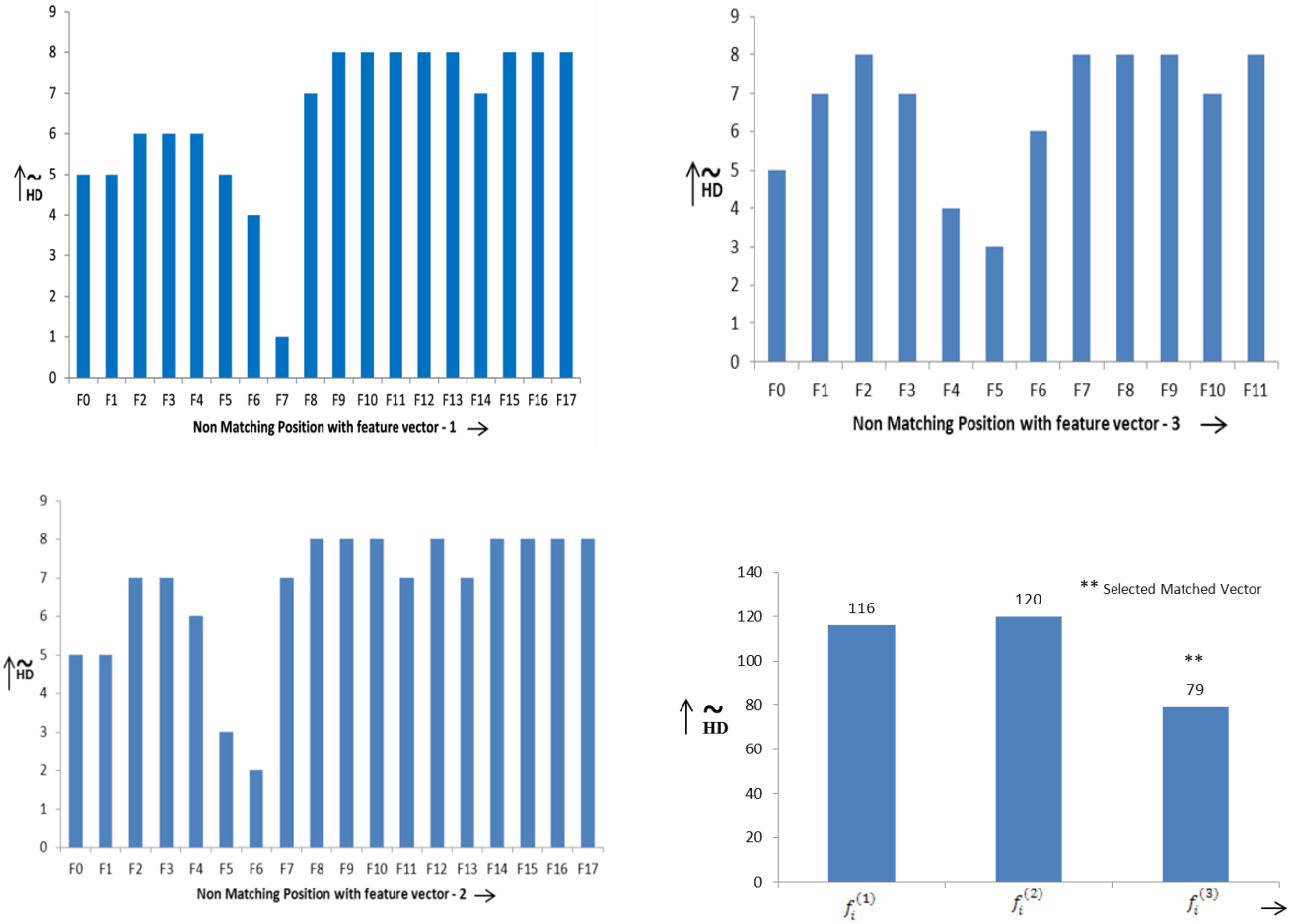


Figure-2. The non-matching positions with the different feature vectors.

Table-1. Feature vector variations in sub blocks.

PCG-1

24	21	20	33	166	21	20	33	166	229
9	51	32	14	187	51	32	14	187	245
3	42	27	31	180	42	27	31	180	243
82	30	25	8	35	30	25	8	35	150
139	245	203	208	32	245	203	208	32	61
20	33	166	229	237					
32	14	187	245	244					
27	31	180	243	250					
25	8	35	150	244					
203	208	32	61	205					

©2006-2015 Asian Research Publishing Network (ARPN). All rights reserved.

PCG-2

252	253	253	253	245	253	253	253	245	255
253	253	253	253	255	253	253	253	255	245
255	245	254	225	251	245	254	225	251	255
229	249	248	248	254	249	248	248	254	227
180	25	0	0	219	25	0	0	219	231
253	253	245	255	248					
253	253	255	245	255					
254	225	251	255	241					
248	248	254	227	255					
0	0	219	231	237					

PCG-3

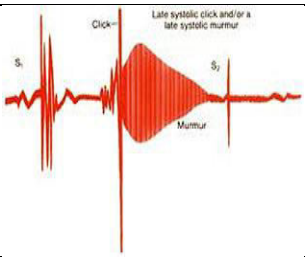
122	0	240	191	115	0	240	191	115	250
0	176	210	6	5	176	210	6	5	71
0	250	18	77	255	250	18	77	255	62
30	255	0	25	208	255	0	25	208	147
117	202	29	122	87	202	29	122	87	225
240	191	115	250	171					
210	6	5	71	255					
18	77	255	62	124					
0	25	208	147	85					
29	122	87	225	56					

Table-2. Comparison of Δx with x.

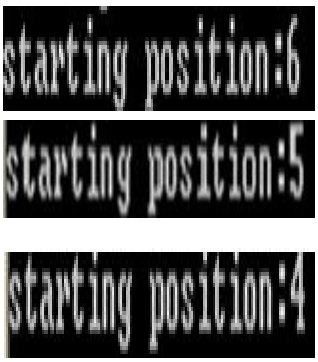
Input		Δx	x
PCG1	SB-1	{24,21,20,33,166,9,51,32,14,187, 3,42,27, 31,180,82,30, 25, 8,35,139, 245,203,208, 32}	[0,0,-3,-176, 0,0,0,32]
	SB-2	{21,20,33,166,229,51,32,14, 187,245,42,27,31,180,243,30, 25,8,35,150, 245,203,208, 32,61}	[0,29,-147,-142, -184,-89,26,61]
	SB-3	{20,33,166,229,237,32,14,187, 245, 244, 27,31,180, 243,250, 25, 8,35,150,244, 203, 208, 32,61,205}	[0,29,-147,-142, -184,-89,26,61]
PCG2	SB-1	{243,243,243,243,243,243,243,243,243,242, 230,255,190,0,78,155,255,61,0,44,40,230,32 ,252,89}	[0,-163,57,-141, 49,45,89,89]
	SB-2	{243,243,243,243,241,243,243, 243,242,240,255,190,0,78,255, 255,61,0,44,233,230,32,252,89,80}	[0,-9,-172, 48, -150, -153, 36, 80]
	SB-3	{243,243,243,241,243,243,243,242,240,242, 190,0,78,255,231,61,0,44,233,255,32,252,89 ,80,236}	[0,156,147,-16, 204,-19,3,236]
PCG3	SB-1	{30,7,0,0,140,21,11,3,0,145,22,10,1,0,133,45 ,8,0,0,143,21,7,8,0,140}	[0,0,-3, 140, 0,0, 7,140]
	SB-2	{7,0,0,140,224,11, 3,0,145,228,10,1,0,133, 230,8,0,0,143,211,7,8,0,140,217}	[0,77,217,209, 210,6,74,217]
	SB-3	{0,0,140,224,255,3,0, 145,228,255,1,0,133, 230,249,0,0,143,211,255,8,0,140,217,244}	[0,27,104,244, 236,-11,33,244]

©2006-2015 Asian Research Publishing Network (ARPN). All rights reserved.

Table-3. Content extraction results using Median string.

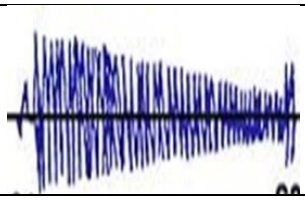
PCG1

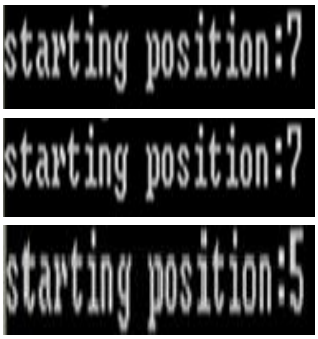
24	21	20	33	166
9	51	32	14	187
3	42	27	31	180
82	30	25	8	35
139	245	203	208	32



PCG2

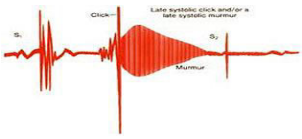
252	253	253	253	245
253	253	253	253	255
255	245	254	225	251
229	249	248	248	254
180	25	0	0	219





PCG3

122	0	240	191	115
0	176	210	6	5
0	250	18	77	255
30	255	0	25	208
117	202	29	122	87



CONCLUSIONS

This paper reports noninvasive signal analysis of heart sounds and murmurs and guarantees the extraction of an optimal set of nonlinear features. The classification of innocent murmurs from murmurs caused by various heart valve pathologies is presented using the minimum value of the nonmatched positions measure calculated across different image frames using Hamming distance measure.

REFERENCES

[1] P. Pevzner and S. H. Sze. 2000. Combinatorial approaches to finding subtle signals in DNA sequences. In Proc. of the Eighth International Conference on Intelligent Systems for Molecular Biology. pp. 269-278.

[2] Falah, Tarek El, Ghnimi Maroua, Elloumi Mourad. 2014. A Consensus Algorithm for Simple Motifs Finding. DEXA. pp. 33-37.

©2006-2015 Asian Research Publishing Network (ARPN). All rights reserved.

[3] Abbas M. M., Malluhi Q. M., Balakrishnan P. 2014. Scalable Multi-core Implementation for Motif Finding Problem. ISPDC. pp. 178-183.

[4] Tanaka S. 2014. Improved Exact Enumerative Algorithms for the Planted (,)-Motif Search Problem. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 11(2): 361-374.

[5] Daniel Hillel Schonberg. 2007. Practical Distributed Source Coding and its application to the compression of Encrypted data. Technical Report No. UCB/EECS-2007-93.

[6] Montserrat Ros and Peret. 2004. A Hamming distance based VLIW/ EPIC code compression technique. Proceeding of International conference on compilers, Architecture and synthesis for Embedded system.

[7] Seon-Won Seong and P. Mishra. 2008. Bitmask based code compression for embedded system. IEEE transaction on computer aided design of integrated circuit and system. 27(4): 673-685.

[8] N. R. Shin and K. Hong. 2011. A Study on the Development of Multiple Objects Extraction Method Using Difference Image Edge Information. International Journal of Energy, Information and Comm. 2(2): 108-120.

[9] H. Kim, R. Sakamoto, I. Kitahara, T. Toriyama and K. Kogure. 2007. Robust Foreground Extraction Technique Using Background Subtraction with Multiple Thresholds. Opt. Eng. 46(9): 097 004-1-097 004-12.

[10] Kiran Kumari Patil, Dr. B. S. Nagbhushan and Dr. Vijaya Kumar B. P. 2011. Psychoacoustic models for heart sounds. In: Proceedings of Advances in Networks and Communications, communications in computer and Information Science. 132(Part 2): 555-556.

[11] Adrian D. C. Chan, Mohyeldin M. Hamdy, Armin Badre, and Vesal Badee. 2008. Wavelet Distance Measure for Person Identification Using Electrocardiograms. IEEE transactions on instrumentation and measurement. 57(2).

[12] S. M. Debbal and F. Bereksi. 2007. Time-Frequency Analysis of the First and Second Heart Beat Sounds. Applied Mathematics and Computation. 184(2): 1041-1052. doi:10.1016/j.amc.2006.07.005.

[13] S. Molau, M. Pitz, R. Schluter and H. Ney. 2001. Computing Mel-Frequency Cepstral Coefficients on the Power Spectrum. IEEE International Conference on Acoustics, Speech, and Signal Processing, Utah. pp. 73-76.

[14] Wenjie Fu, Xinghai Yang, Yutai Wang. 2010. Heart Sound Diagnosis Based on DTW and MFCC. 3rd International Congress on Image and Signal Processing (CISP2010).