Perceptually Lossless Ultrasound Video Coding for Telemedicine Applications

Alessandro Neri, Federica Battisti, Marco Carli, Marina Salatino, Michela Goffredo, and Tommaso D’Alessio

University of Roma "Roma TRE", Roma, Italy
Via della Vasca Navale, 84
00146 Roma Italy

ABSTRACT

The paper reports the results of a subjective test carried out to investigate the perceptual effects of different modalities and coding parameters of H.264/AVC (e.g. enabling/disabling the in-loop de-blocking filter) when applied to echocardiographic and echodoppler sequences. The experimental results show that adoption of last generation codecs can preserve the diagnostic effectiveness of the ultrasound imaging sequences.

1. INTRODUCTION

Echographic systems encounter an increasing interest because they allow studying both the morphology of the internal organs and their functional characteristics, in a safe, non invasive, portable, and low cost way. For instance, in cardiology, they provide real-time dynamical information on anatomy and dimension of heart and great vessels, as well as blood flow velocity that can be effectively employed to assess the functionality of cardiac valves and chambers, and the presence of stenoses together with their clinical relevance.

As the complexity of modern echographic systems has been continually increased, many efforts have been made in order to improve friendliness in the use of the equipment, in the setting and managing of acquisition parameters, in processing and rendering the images, both in real time and/or off-line.

However, less attention has been devoted to the insertion of these equipments in a telemedicine context. In fact, literature is mainly focused on the encoding of single frames by means of DCT and wavelet transforms and perceptual aspects have been mainly accounted for in the optimization of the visibility matrix [1]-[5].

When storing single images, the standard format used is the DICOM one, with no compression or at most with lossless compression. However, when storing sequences (cine loop) a lossy digital compression is automatically introduced by off-the-shelf equipments. The characteristics of the coding algorithms are often undeclared, the compression level can not be adjusted and no control on the resulting quality of the compressed sequence is available.

State of the art and next generation codecs are providing substantial compression efficiency for transmission and storage of digital videos. Together with transmission technologies offering larger bandwidths over DSL, wireless and satellite networks, the capability of delivering high quality video services for medical applications is now a reality.

However, the tuning of the parameters that control encoder performance must be driven by the diagnostic effectiveness instead of aesthetic criteria. Basically, video quality metrics for medical imaging should consider degradations as: accuracy of measures of areas or volumes, the Receiver Operating Characteristics (ROC) curves related to the detection of presence or absence of abnormalities, identification, and localization of abnormality regions.

Obviously the amount of compression that can be applied while maintaining an adequate diagnostic capability depends also on the initial level of quality and resolution of the image provided by the equipment itself. In this paper the aspects concerning the basic performance of the echographic equipment are not considered: the goal of the work is the measurement of the quality resulting after the video coding.

In this case, there are no test or reference objects (phantoms) available as it is the case for quality control and for evaluating the resolution in depth and/or lateral and/or in density, but it is necessary to define new tools and metrics useful to evaluate the effectiveness of compression methods. To this purpose, it is useful to remember that presently the ultimate way of assessing the quality provided by compression algorithms is through panels of specialized medical personnel who could
evaluate if and/or to what extent a compressed video is adequate for a correct diagnosis.

Even if in literature there are many efforts aimed at defining “objective criteria” that could overcome the need to have doctors’ panels available, the automatic evaluation of what could be considered as a “perceptually lossless compressed image” is still an open field of work. In fact, previous investigations have been mainly focused on still images [6-12], while sequences have received less attention [13-15].

For this purpose, we encoded the echocardiographic videos using the latest video coding standard, i.e. H.264/AVC standard. It provides high compression rates while providing a good perceptual quality. The ability to recognize particular patterns in the video is crucial to promote such a system for distance monitoring and medical surveillance.

The paper reports the results of subjective tests carried out to investigate the perceptual effects of different coding parameters of H.264/AVC (e.g. enabling/disabling the in-loop de-blocking filter) when applied to echocardiographic and echodoppler sequences.

The remaining part of the paper is organized as follows: in Section 2 the basic cardiographic signal is described. In Section 3 some results of the performed experiment are presented; finally, in Section 4 the conclusions are drawn.

2. ECHOCARDIOGRAPHY

A typical telemedicine set up could involve at least two simultaneous video bit streams with different quality levels: a low bit rate stream for video communications and a good quality video stream for diagnostics. Even if up to now in the storage and/or transmission of medical images lossless algorithms are deemed necessary (so that only compression rates up to 2 - 3 to 1 have been used), there is an increased acceptance of compression algorithms which could be considered as "perceptually lossless". In this case, the perceptual quality of the content delivered over communication networks is crucial in ensuring a top-class doctor’s experience. In particular, the echocardiographer’s task is the assessment of the presence of significant deviations from normality of patient’s cardiac functions, the evolution of pathology over time, as well as the effectiveness of a prescribed therapy [8].

More in detail, the assessment of ventricular function is performed by comparing end-diastole and end-systole frames in order to evaluate global indicators of the change of the ventricular dimension during systole (e.g. ejection fraction) as well as regional data such as wall thickness and wall motion. In addition to the estimation of the size of ventricles and atria, the evaluation of valve diseases requires the assessment of their opening and closing ability and the characterization of the blood flow through them. In doppler echocardiography direction and velocity of blood flow across the valves is estimated on the basis of the carrier frequency shift of the ultrasound wave. The velocity field is then combined with the amplitude of the backscattered signal for the production of a colored map of blood pressure distribution, which allows, for example, direct assessment of both pressure and volume overloads. By convention, red color represents blood flow toward the transducer, while blue color is associated to blood flow away from it.

Among the others, as critical elements that determine the usefulness of echocardiographic sequences for diagnosis we cite the ability to extract endocardial and epicardial borders. In fact the location of the endocardial edge allows estimating chamber area and wall motion, while joint processing of the endocardial and epicardial edges allows to estimate wall thickness and ventricular mass. However, due to the characteristics of the medium, the penetration characteristics of ultrasound waves are, in general, lower then those exhibited by other imaging technologies. In addition, the acquired video is often affected by dropout, speckle, artifacts, with high degradation in the far field of the signal to noise ratio produced by the attenuation of the ultrasound wave.

Speckle is produced by the interaction between the ultrasound wave and independent scatters within a resolution cell and appears as a granular, multiplicative noise. Moreover, since the acoustic impedances of blood and endocardium are relatively small, the endocardial edge is often poorly defined. In addition the velocity of the endocardium boundary fairly exceeds the velocity of the epicardium boundary. Overall gain, time gain compensation and depth setting can also affect the resolution of the imaging system as well as the signal-to-noise ratio.

Thus analysis of echocardiographic images can be particularly challenging even when lossless data are available. In synthesis, echocardiography is an effective, non invasive, real-time tool for providing dynamical information on anatomy and dimension of heart and great vessels, and blood flow velocity, its use is widespread, and its use in telemedicine contexts can greatly benefit patients during exams and follow-up. Moreover, ultrasound videos represent an interesting benchmark as US examinations imply the use of real time imaging modalities that must be maintained in Telemedicine applications.

3. SUBJECTIVE EXPERIMENT AND EXPERIMENTAL RESULTS

The process of compressing a video sequence can introduce distortions or impairments that reduce the
perceived quality of the video. To evaluate the lowest compression rate that allows considering a video sequence suitable for diagnostic purposes, a subjective experiment has been performed. In Figures 2 and 3 examples of original and compressed sequences are shown.

In our test 6 subjects have been chosen from a pool of cardiologists from the Department of “Scienze Cardiovascolari, Respiratorie e Morfologiche” in the ‘Policlinico Umberto I’ Hospital in Rome. The doctors have an experience in the cardiologic field and in echography of at least five years. The authors are aware that the small number of subjects can only give an indication about the quality assessment; nevertheless the specific application considered requires highly trained experts whose judgement and hints are of great value and indicative of the results of larger experiments.

The doctors were asked to wear any vision correcting devices (glasses or contact) they usually wear when watching television. The experiment was performed on each subject separately and it lasted approximately 10 minutes for each doctor.

Each subject is seated straight ahead of the monitor, located at or slightly below eye height for most subjects. The subjects are positioned at a distance of 70 cm from the screen, as shown in Fig. 1.

The experimental session is composed of three steps:

- Instructions: the subject receives verbal instructions. The instructions are also displayed so that at any stage of the experiment it is possible to get the information needed.

- Training: the original video and the most compressed version of it (presenting the worst visual quality) are shown. These sequences represent the extremes for the experiment and are used to establish the acceptable value range for diagnostic accuracy. The least useful videos in the training stage should be assigned a value of ‘1’.

- Test: it is performed with the complete set of test sequences presented in a random order.

The subject task consisted in giving a numerical judgment on the video quality after each test sequence on a scale from ‘1’ (worst quality) to ‘5’ (best quality). When the experiment is complete, the test subjects are asked a few questions before they leave. These questions gather interesting information about the evaluation process that cannot be gathered during the experiment. This is important to understand the most important part to be considered in a diagnosis.

The original video considered in this first experiment represents a cardiac cycle of a patient affected by a cardiac pathology. The sequence is 2 seconds long and has been selected by a specialist in such a kind of diseases. It is important to notice that the length is considered valid and appropriate for diagnostic purposes.

The format of the original video is uncompressed AVI, with a size of 28 MB. This sequence was compressed using an H.264/AVC video encoder, with 8 different bit rates as shown in Tab. I. As it can be noticed, the reduced file size allows the delivering of this video by using the UMTS/CDMA2000 generation of mobile communication. This is a first fundamental step for a wider diffusion of second opinion and/or telediagnosis in this field.

<table>
<thead>
<tr>
<th>Bit/rate</th>
<th>Compressed file size</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 Mb/s</td>
<td>1400 kB</td>
</tr>
<tr>
<td>4 Mb/s</td>
<td>1200 kB</td>
</tr>
<tr>
<td>3 Mb/s</td>
<td>900 kB</td>
</tr>
<tr>
<td>2 Mb/s</td>
<td>600 kB</td>
</tr>
<tr>
<td>1 Mb/s</td>
<td>300 kB</td>
</tr>
<tr>
<td>0.9 Mb/s</td>
<td>270 kB</td>
</tr>
<tr>
<td>0.8 Mb/s</td>
<td>240 kB</td>
</tr>
<tr>
<td>0.7 Mb/s</td>
<td>210 kB</td>
</tr>
</tbody>
</table>

After the experiment, a screening procedure has been carried out to verify the ‘goodness’ of the subjective answers. No outliers have been identified.
As can be noticed in Figure 4, increasing the compression ratio causes an increase in the annoyance perceived by the user until a threshold (in Mb/s) that can be considered as the usability limit for medical purposes is reached. In the performed test, a bit rate of 3Mb/s has been judged as a quasi-optimal quality threshold.

An interesting outcome of this experiment, is the possibility to compress echocardiographic and echodoppler sequences while preserving the effectiveness for diagnostic purposes.

4. CONCLUSIONS

The results of the subjective test on the perceptual effects of different modalities and coding parameters of H.264/AVC, (e.g. enabling/disabling the in-loop de-blocking filter) show that channel capacity of about 3 Mb/s is required to preserve the diagnostic effectiveness of the ultrasound imaging sequences. However, due to limitations of the ultrasound imaging system employed for the test execution, we were unable to determine the minimum rate associated to the separate coding of the various elements.
composing the rendered video (text, ECG, ultrasonic image).

On the other hand, the object handling capability of the MPEG4 codec family is particularly suited for cardiographic videos that combine textual information with one dimensional graphs and 2D or 3D videos that have to be rendered in an isochronous way. In fact, direct coding of the mixed real-time video could be either highly ineffective if text and graph readability has to be preserved, or highly unsatisfactory because blocking and blurring in the text and graph areas induced by low bit rate coding could influence the overall quality evaluation.

Separate coding of the different video-objects is also recommended in doppler echocardiography where two scalar fields representing the echo amplitude and the radial velocity of a voxel are combined by the rendering device into a colored image. In fact, in this way we can achieve better compression together with better resilience of chroma distortions.

Acknowledgements: The cooperation of cardiologists from the Department of “Scienze Cardiovascolari, Respiratorie e Morfologiche” in the ‘Policlinico Umberto I’ Hospital in Rome is gratefully acknowledged.

REFERENCES


