Comparison Between JPEG2000, AVC and HEVC in Compressing Scanned Sheet Music Images

Alexandre Zaghetto[1], Mateus Mendelson[2], Eustaquio Grilo[3]

Abstract— **This paper evaluates the performance of different image and video coders in compressing scanned sheet music images. For that purpose, the state of the art still image coder JPEG2000 and the video coders AVC and HEVC are used. First, each page of the scanned musical piece is treated as a still image and compressed independently by JPEG2000, AVC-INTRA and HEVC-INTRA. Then, the scanned pages are interpreted as frames of a video sequence and encoded by AVC-INTER or HEVC-INTER. By doing so, interframe prediction may be used as a pattern matcher. Since sheet music has a well behaved structure of symbols, it is expected that interframe prediction will easily find patterns on reference frames that are very similar to those being currently encoded. In other words, present frames use previously encoded frames as a dictionary. The pattern matching algorithm (motion estimation and compensation) generates residual data that can be more efficiently compressed. Results show that HEVC consistently outperforms AVC and JPEG2000. Moreover, the proposed experiments indicate that HEVC-INTER, in average, outperforms HEVC-INTRA when used to compress sheet music images.**

Keywords— **HEVC, AVC, JPEG2000, Sheet Music Compression.**

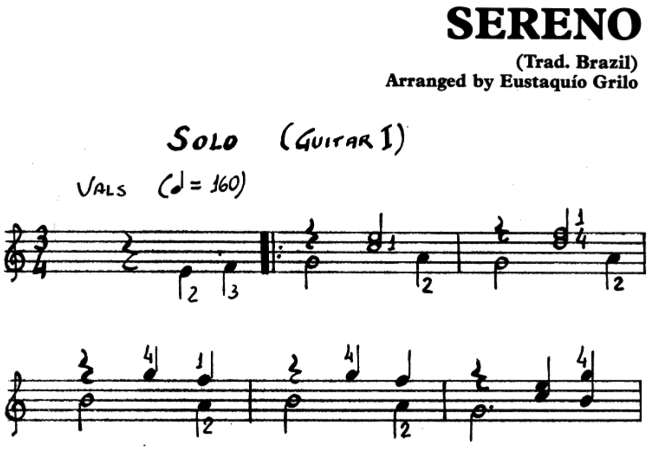


Fig. 1. Beginning of “Sereno”, a traditional Brazilian piece, arranged by Prof. Eustaquio Grilo. The recurrence of similar symbols may be observed.

I. Introduction

Printed music is nothing more than a stylized Cartesian plane, with the notehead itself called *punto*, the Italian word for “dot” or “point”. Roughly speaking, each *punto* is placed on the Cartesian plane according to a discrete function $f[n]$, where the independent variable n denotes the position of a note in a sequence and $f[n]$ denotes the frequency of the sound that must be produced by an instrument in that position. Musical notation enables simultaneous reading and playing, plus storing and spreading music as scores. The consequence is that we may have access to the great musical legacy from the past. As an example, Figure 1 shows the beginning of “Sereno”, a traditional Brazilian piece.

Music has been omni-present throughout human cultures and for a long period it was passed from generation to generation through oral tradition. The Italian conductor Guido d’Arezzo began to formalize its modern structure [1] in the early 11th century by proposing the use of five parallel lines, which compose a *staff*, and naming the musical notes as they became presently known. From this point forward, the written registry of music began to be standardized and gained notoriety and importance.

Examples of symbols that compose a musical score are lines, clefs, notes, rests, breaks, accidentals, key signatures, time signatures, note relationships, dynamics, articulation marks, ornaments, octave signs, repetitions and codas [2]. The recurrence of symbols in a musical score is explored in two of some of the encoding methods evaluated in this paper.

Since the latter half of the nineteenth century, an academical approach has been necessary in the research of past music in all aspects, from the discovery of manuscripts and early editions, up to references for the final performances. For music professors and students this means a considerable amount of documents to deal with. Nowadays, computers enable digital storage. However, due to the huge amount of data, compression is needed. Although music optical character recognition can be used, this work considers that the preservation of the original musical documents is also of great interest. In other words, not only the content is important, but also the visual characteristics.

When it comes to still image and video compression, one may refer to JPEG2000, AVC and HEVC as the state of the art. JPEG2000 was conceived for image compression and consistently outperforms its predecessor JPEG. HEVC and AVC are video coders, but can also be used as still image compressors. When referring to HEVC and AVC as video encoders the acronyms HEVC-INTER and AVC-INTER will be used. HEVC-INTRA and AVC-INTRA are their image encoder versions. In this paper we evaluated the use of these five encoding approaches in compressing sheet music images.

[1]Dept. of Computer Science,[2]Dept. of Software Engineering and [3]Dept. of Music, University of Brasília (UnB), Brasília-DF, Brazil, E-mails: zaghetto@unb.br, {mendelson.mateus, eustaquio.grilo}@gmail.com

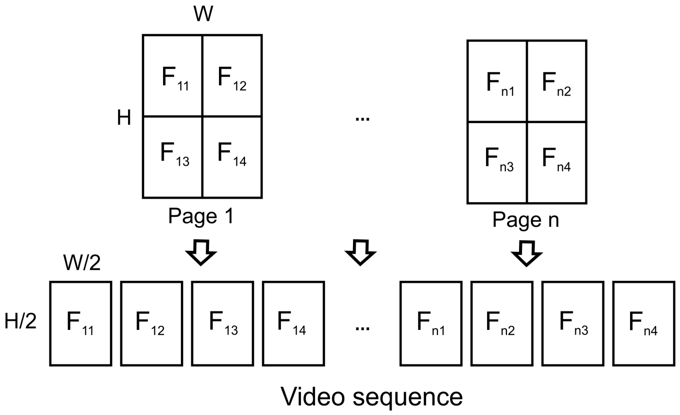


Fig. 2. Proposed page processing algorithm. Each scanned $H \times W$ pixels page is segmented into four $H/2 \times W/2$ pixels sub-pages. Then, these sub-pages are used to build a video sequence.

II. BACKGROUND

Scanned musical scores may be compressed as grayscale images or may be binarized before compression. For binary images, a bi-level compression algorithm, such as JBIG [3] and JBIG2 [4] may be applied. Alternatively, scanned images may be processed by an optical music recognition (OMR) software. Binarization may cause strong degradation to symbol contours and textures. In optical music recognition, the original aspect of the printed music is completely lost. These two approaches, binarization/bi-level compression or optical music recognition, are not the best approach if one is interested in compressing a music document while preserving as much as possible of its original *aesthetic* value. Hence, whenever possible, continuous-tone compression is preferred.

Examples of continuous-tone image compression algorithms are JPEG [5] and JPEG2000 [6]. The video compression standard HEVC [7], [8] operating in pure intra mode is also a very efficient compressor for still images, as well as its predecessor, AVC [9]. Multilayer approaches such as the mixed raster content (MRC) imaging model [10] may also be used. But in this case, compression is challenged by soft edges and often requires pre- and post-processing [11].

Musical symbols along sheet music presents a repetitive structure such that dictionary-based compression methods become very efficient [12]. For continuous-tone sheet music images, the recurrence of similar symbols may be observed in Figure 1. Nevertheless, the development of an efficient dictionary-based encoder relying on continuous-tone pattern matching for high resolution images is a challenging problem. Here, in addition to the most successful still image approaches (JPEG2000, AVC-INTRA and HEVC-INTRA) we suggest the use of encoders based on page processing procedure, pattern matching predictors and efficient transform encoding of the residual data [13], [14] that explores the recurrence of symbols throughout the musical score (AVC-INTER and HEVC-INTER). In the next section we describe the proposed coding scheme.

III. PROPOSED CODING SCHEME

It is known that HEVC [7], [8] performs significantly better than its predecessor AVC [15], [9]. Therefore, HEVC is considered in this work as the higher bound of the possible compression performance. Among the many improvements brought into HEVC, we may mention a very efficient pattern matching algorithm with variable block size implemented by a coding structure that includes the concepts of *coding units* (CU), *prediction units* (PU) and *transform units* (TU).

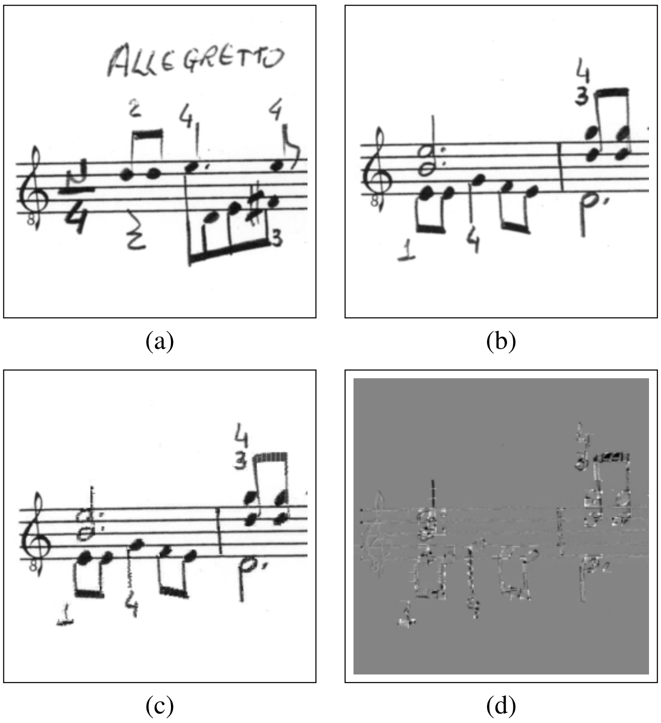
In HEVC, the coding unit is the fundamental unit of the region splitting, with sizes that varies from 8×8 to 64×64 in a quadtree structure. Each coding unit can be further symmetrically or asymmetrically partitioned into prediction units. The prediction units can be coded using one of the 35 intra prediction modes, or using interframe prediction.

Intra CUs have two types of PUs ($2N \times 2N$ and $N \times N$) and inter CUs have four types of PUs ($2N \times 2N$, $2N \times N$, $N \times 2N$ and $N \times N$). Hence, HEVC implements 7 partition modes, including SKIP. Finally, HEVC also presents a quadtree structure transform coding with block sizes varying from 4×4 to 32×32. The best block partitions, predictions and transform unit sizes are determined in a rate-distortion sense.

Given that the music is also to be compressed by video coders, the proposed encoding scheme organizes the scanned pages in such a way that interframe prediction may find on previously encoded coding units symbol parts that are similar to those on coding units currently being encoded. Figure 2 illustrates the proposed page processing algorithm. First, each scanned $H \times W$ pixels page is divided into four $H/2 \times W/2$ pixels sub-pages, which are further organized as frames of a video sequence and then encoded through the proposed encoder. The reason that page subdivision is used in the multi-page compression is that in some cases similar symbols are more likely to be found on the same page rather than on different pages of the same document. If the symbols aspect is constant throughout the whole musical piece, each page may be converted into one single frame, the motion estimation search range may be increased and segmentation may be skipped. The final step is to compress the resulting video using AVC-INTER or HEVC-INTER.

The basic idea of the interframe prediction is to exploit similarities between video frames in order to reduce the amount of information to be encoded. Based on previously encoded units, it first constructs a prediction of the current frame and then creates a residual frame by subtracting the prediction from the current frame. Figure 3 illustrates the effect of using interframe prediction as an approximate pattern matching algorithm. Figures 3 (a) and (b) show examples of a reference and a current music part, respectively. Figures 3 (c) and (d) represent the prediction of current musical symbols using 4×4 block partitions and the corresponding residual data, respectively. Note that although the reference and the current image represent different parts of a musical piece, the symbols are very similar, enabling efficient prediction.

It is noteworthy that 4×4 prediction generates a lower-energy residual, when compared with the 8×8 and 16×16



(a) (b)

(c) (d)

Fig. 3. Approximate pattern matching using interframe prediction: (a) reference frame; (b) current frame; (c) predicted symbols (block size: 4 × 4 pixels); and (d) prediction residue.

prediction, for instance. However, smaller partitions require a larger number of bits to encode the motion vectors. This implies that prediction unit size selection has a major impact on compression performance and must be dealt with by a rate-distortion optimization algorithm.

The example shown in Figure 3 suggests that previously encoded symbols (reference frames) may be seen as a dictionary used by the pattern matching algorithm (interframe prediction). The dictionary is updated in parallel with the encoding process, since new reference frames become constantly available. Furthermore, a rate-distortion optimization algorithm estimates which combination of block partitions, predictions and transform unit sizes should be applied. Once the residual data is available, HEVC uses an integer transform with similar properties as the DCT (discrete cosine transform) and the resulting transformed coefficients are scaled, quantized and entropically encoded using CABAC (Context-adaptive Binary Arithmetic Coding).

The use of JPEG2000, AVC-INTRA and HEVC-INTRA is straightforward. Each page of a musical piece is encoded separately as a still image.

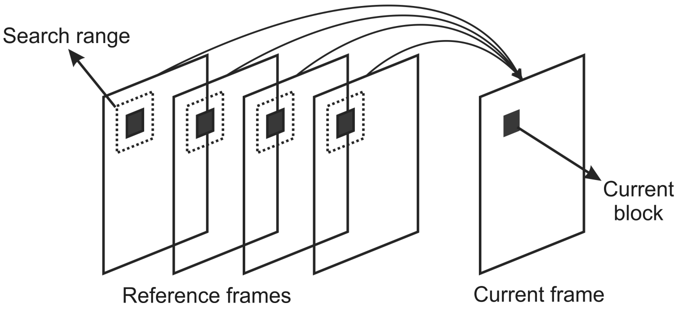
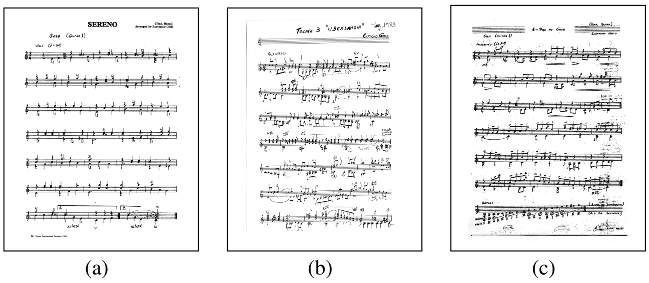


Fig. 4. Configuration parameters that have greater influence on the encoder performance: Rf (number of reference frames) and Sr (search range). In our experiments, Rf and Sr are set to 4 and 64, respectively.



(a) (b) (c)

Fig. 5. First pages of typical musical scores used in our test set: (a) “Pau no Gato” (number of pages: 2, size: 1088 × 800); (b) “Sereno” (number of pages: 2, size: 1088 × 800 pixels); and (c) “Tocata” (number of pages: 4, size: 1088 × 800 pixels). Compositions/arrangements by Prof. Eustaquio Grilo.

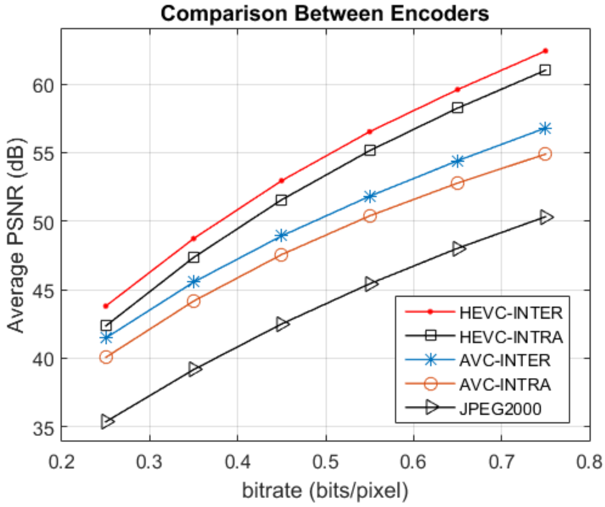
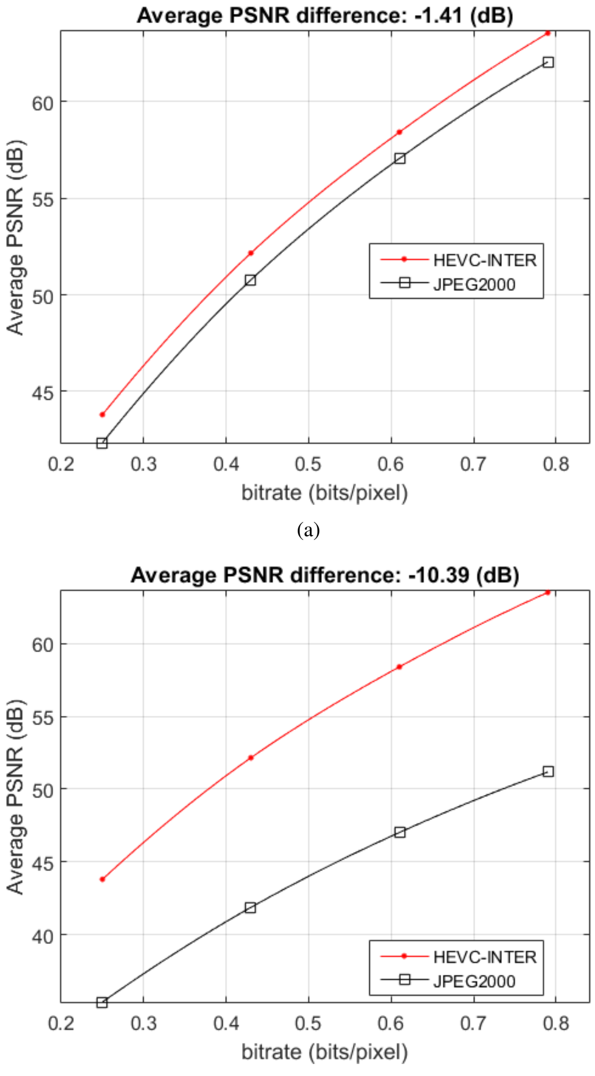


Fig. 6. Average PSNR plots for all documents in the test set. Results show that HEVC with page processing (HEVC-INTER) outperforms HEVC-INTRA, AVC-INTRA, AVC-INTER and JPEG2000.

IV. Experimental Results

Two configuration parameters have greater influence on the AVC-INTER and HEVC-INTER encoders. One is the number of reference frames, Rf, the other is the search range, Sr, as illustrated in Figure 4. In our tests, different page sets are compressed using JPEG2000, HEVC-INTRA and AVC-INTRA (HEVC and AVC operating in pure intra mode), HEVC-INTER and AVC-INTER (HEVC and AVC using also interframe prediction). In JPEG2000, HEVC-INTRA and AVC-INTRA compression, pages are encoded separately and only intraframe prediction modes are used. As for HEVC-INTER and AVC-INTER, the first frame of the sequence is encoded as an I-frame (intra frame) and all the remaining frames are encoded as P-frames (past frames are used as reference frames).

We evaluated the performance of our method adjusting Sr



(a)

(b)

Fig. 7. PSNR plots comparing HEVC-INTER with HEVC-INTRA and JPEG2000. In average, HEVC-INTER outperforms HEVC-INTRA (next best performance) and JPEG2000 (worst performance) by 0.6 and 8.86 dB, respectively.

to 64 and Rf to 4. For the sake of illustration, Figures 5 (a), (b) and (c) show the first page of three test sequences: “Pau no Gato”, “Tocata” and “Sereno”. Six other musical pieces also compose the test set. These documents are described in Table I.

Figure 6 show the average PSNR (peak signal-to-noise ratio) plot for the nine sequences from the test set. Results show that, in average, the HEVC compressor outperforms all other encoders. AVC performs better than JPEG2000. In HEVC and AVC compression, the versions that use inter prediction present better results than the intra-only configurations. Figures 7 (a) and (b) show PSNR plots comparing HEVC-INTER with HEVC-INTRA and JPEG2000. In average, HEVC-INTER outperforms HEVC-INTRA (next best performance) and JPEG2000 (worst performance) by 1.41 and 10.39 dB, respectively. Table II compares the average PSNR achieved by each encoder at 0.5, 0.75 and 1.0 bits/pixel.

TABLE I
THE TEST SET IS COMPOSED BY 9 MUSICAL PIECES. HEIGHT AND WIDTH ARE GIVEN IN PIXELS.

Musical piece	# of pages	height	width
Pau no gato	2	1088	800
Suite brasileira	4	1088	800
Sereno	2	1088	800
Assum preto	1	1088	800
Escalas	3	1088	800
A canoa virou	1	1088	800
Suite quabra-dedos	2	1088	800
Exercício polivalente	2	1088	800
Tocata Uberlândia	4	1088	800

TABLE II
AVERAGE PSNR AT 0.5, 0.75 AND 1.0 BITS/PIXEL.

	Bitrate (bits/pixel)			
Encoder	0.25	0.43	0.61	0.79
JPEG2000	35.36	41.87	47.02	51.17
AVC-INTRA	40.06	46.93	51.86	55.68
AVC-INTER	41.47	48.29	53.4	57.71
HEVC-INTRA	42.35	50.77	57.05	62.05
HEVC-INTER	43.81	52.16	58.41	63.53

V. CONCLUSIONS

In this paper the evaluation of different image and video coders in compressing scanned sheet music images was performed. Among the evaluated coding schemes, HEVC-INTER presented the best average performance. Its coding structures, including coding units, prediction units and transform units, with a rate-distortion optimization algorithm, indirectly implement a very efficient pattern matcher. In addition, the intraframe prediction, the DCT-based transformation and CABAC also contribute to improve the encoding efficiency. Results show that HEVC-INTER objectively outperforms HEVC-INTRA, AVC-INTRA, AVC and JPEG2000.

REFERENCES

[1] J. extensions, “Guido of Arezzo and His Influence on Music Learning,” *Musical Offerings, Cedarville University*, vol. 3, no. 1, pp. 37-59, Spring 2012.

[2] G. Heussenstamm, *The Norton Manual of Music Notation*, 1st ed. EUA: W. W. Norton and Company, 1987.

[3] JBIG, “Information Technology - Coded Representation of Picture and Audio Information - Progressive Bi-level Image Compression. ITU-T Recommendation T.82,” March 1993.

[4] JBIG2, “Information Technology - Coded Representation of Picture and Audio Information - Lossy/Lossless Coding of Bi-level Images. ITU-T Recommendation T.88,” March 2000.

[5] W. B. Pennebaker and J. L. Mitchell, *JPEG Still Image Data Compression Standard*. Chapman and Hall, 1993.

[6] D. S. Taubman and M. W. Marcellin, *JPEG 2000: Imagem Compression Fundamentals, Standards and Practice*. EUA: Kluwer Academic, 2002.

[7] G. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (hevc) standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1649–1668, Dec 2012.

[8] ITU-T, "High Efficiency Video Coding. Recommendation ITU-T H.265," April 2015.

[9] ——, "Advanced Video Coding for Generic Audiovisual Services. Recommendation ITU-T H.264," February 2014.

[10] MRC, "Mixed Raster Content (MRC). ITU-T Recommendation T.44." 1999.

[11] A. Zaghetto and R. L. de Queiroz, "Pre- and postprocessing for multilayer compression of scanned documents," *Journal of Electronic Imaging*, no. 20, p. 043005, Oct. 2011.

[12] N. Francisco, N. Rodrigues, E. da Silva, M. de Carvalho, S. de Faria, and V. da Silva, "Scanned compound document encoding using multiscale recurrent patterns," *IEEE Transactions on Image Processing*, vol. 19, no. 10, pp. 2712–2724, Apr. 2010.

[13] A. Zaghetto and R. de Queiroz, "High quality scanned book compression using pattern matching," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, Sept. 2010, pp. 2165–2168.

[14] A. Zaghetto, B. Macchiavello, and R. de Queiroz, "HEVC-based anned document compression," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*, Sept 2012, pp. 821–824.

[15] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[16] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," *Proc. 13th VCEG-M33 Meeting*, Apr. 2001.