IMPROVING ENCODING EFFICIENCY OF ENDOSCOPIC VIDEOS BY USING CIRCLE DETECTION BASED BORDER OVERLAYS

Bernd Münzer, Klaus Schoeffmann, Laszlo Böszörmenyi

Institute of Information Technology, Klagenfurt University, Austria
{bernd,ks,laszlo}@itec.aau.at

ABSTRACT

Videos of endoscopic procedures typically feature a circular content area in the image center. This area is surrounded by a dark border that carries no relevant information but is subject to noise. Thus, a considerable proportion of the available bitrate has to be wasted to encode the border regions. We propose to superimpose the border regions with a homogenous black mask so that it can be encoded efficiently with skipped macroblocks. To determine the exact position and size of the circular content area we use an efficient circle detection algorithm. Through an evaluation with 138 videos we show that the border overlay can significantly reduce the bitrate without degrading the visual quality of the content area.

Index Terms— Content analysis, endoscopy, video coding

1. INTRODUCTION

Medical endoscopy is a minimally invasive method for diagnostic as well as therapeutic interventions in a variety of human body regions (e.g., abdomen, colon, joints). The endoscopist is guided by a video signal produced by a small camera that is inserted into hollow organs or cavities via a natural or artificial orifice. The video signal is routinely captured and archived for documentation purposes and retrospective analysis. This leads to large collections of endoscopic videos that consume a lot of storage space.

A promising approach to reduce the enormous storage requirements is to use automatic content-based analysis methods and exploit domain specific characteristics, like the following: Due to the very restricted field of view the image sensor is built in such a way that it captures the entire light that is projected by the camera lens. The area of this light projection inherently has a circular shape but common video codecs only support rectangular formats. Hence, the actual content is limited to a circular content area in the center of the image which is surrounded by an irrelevant dark border (unless an optional zoom function is used). A typical example of an endoscopic video frame is illustrated in Figure 2(a). About half of the pixels belong to the dark border. This ratio is especially high for videos with a widescreen aspect ratio (16:9), which is the common format for HD (High Definition) videos.

Unfortunately, the border does not form a perfectly homogeneous black area but is subject to a considerable amount of noise due to image sensor imperfections. Figure 1 depicts example patches and the corresponding histograms to illustrate the intensity of the noise. If a video is recorded, the encoder misinterprets the noise as detail information and tries to preserve it. This causes a certain ratio of the available bitrate to be wasted to encode completely irrelevant information caused by sensor errors. Figure 2(b) illustrates the macroblock (MB) types chosen by an H.264/AVC encoder for an example frame which actually has a rather low noise intensity. Nevertheless, many of the MBs in the border regions are coded as intra MB or predictive MB which implies the useless coding of coefficients and prediction information.

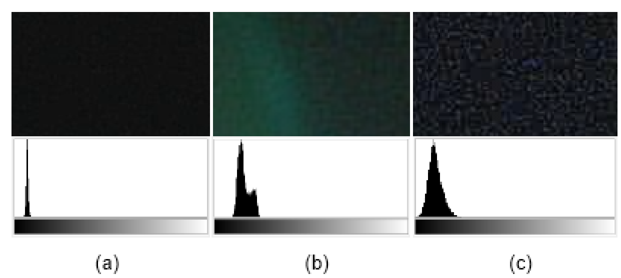
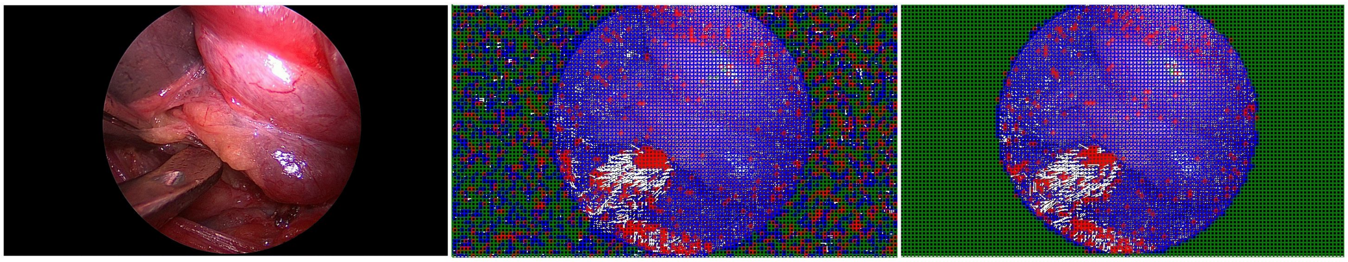


Fig. 1. Example patches from border regions with corresponding histograms. (a) in-patient (HD), (b) in-patient (SD), (c) out-of-patient (HD)

In order to increase the encoding efficiency of endoscopic videos, we propose to superimpose the noisy border regions with a pure black mask which we call border overlay. The underlying conjecture is that this modification causes H.264/AVC encoders to choose skipped MBs instead of intracoded and predictive MBs for the border regions. Thus, no image sample information, residual coefficients, MB partitions etc. have to be coded and a lower overall bitrate can be achieved for the same visual quality of the content area.

To determine the circular content area and create the border overlay we use our circle detection algorithm presented in [1]. For a detailed introduction to the domain of endoscopy and a survey of related work in the field of content-based processing of endoscopic videos please refer to [2].



(a) (b) (c)

Fig. 2. Visualization of the macroblock analysis of a representative example frame. (a) shows the original image, (b) shows the macroblock types of the unmodified version, (c) shows the macroblock types of the overlay version. The color codes are as follows: green = skipped MB (very low bitrate consumption), red = intra-coded MB (high bitrate consumption), blue = predicted MB (medium bitrate consumption), white = motion vector

2. BLACK BORDER OVERLAY

In order to create the black border overlay mask we first have to determine the exact position and size of the circular content area. For that purpose we employ an efficient circle detection algorithm that we recently proposed in [1]. This algorithm has specifically been designed for the domain of endoscopic videos and exploits a number of domain specific characteristics. This renders it superior to generic approaches like the generalized Hough transform [3] and its many derivatives in terms of accuracy and runtime.

The result of the circle detection algorithm is a sequence of circle parameter tuples in the form (x_c, y_c, r) where x_c and y_c represent the center coordinates of the circle and r represents the radius. Each tuple corresponds to one frame of the video. This is necessary because neither the position nor the size of the circular content area in the image are stable but can change at any time because of strong camera motion (especially rotation) and/or zoom (further technical details of endoscopes and endoscopic videos can be found in [4]). We create the border overlay by modifying each individual frame according to equation 1.

$$I(x,y) = \begin{cases} 0 & \text{if } x \leq x_c - b(y) \vee x \geq x_c + b(y) \\ I(x,y) & \text{otherwise} \end{cases} \quad (1)$$

with

$$b(y) = \begin{cases} \sqrt{(r * \varepsilon)^2 - (y_c - y)^2} & \text{if } |y_c - y| < r * \varepsilon \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where $I(x,y)$ is the pixel at position (x,y) of the frame and ε is the tolerance margin coefficient which can be used to slightly extend the content area. This makes sense because the viewing direction of the endoscope is indicated by a small spike at the border of the circular content area and may be preserved. In the following experiments we used an ε of 1.

3. EVALUATION

To evaluate the impact of the black border overlay on encoding efficiency, we used four representative video data sets with a total length of 224 minutes consisting of 138 video files that have been randomly obtained from cooperating hospitals. All videos have originally been encoded with the MPEG-2 video codec in constant bitrate mode (20 Mb/s for 1920x1080, 12 Mb/s for 1280x720 and 7 Mb/s for SD resolutions, each with a framerate of 25 fps). This configuration is commonly used in many endoscopic video capturing systems for historical reasons. The data sets are summarized in Table 1.

Data set	Resolution	Files	Length
ip (HD)	1920x1080, 1280x720	20	74 min.
ip (SD)	720x576, 720x544	69	53 min.
ip (CIF)	384x288	31	27 min.
oop (HD)	1920x1080, 1280x720	18	70 min.

Table 1. Video data sets

The data sets contain typical endoscopic videos in different resolutions. Our main data set consists of HD (High Definition) videos because HD systems are currently establishing as common practice and presumably will be widely used in the near future. Nevertheless, even if HD is used during surgery, due to economic restrictions, for recording most surgeons use SD (Standard Definition) or even CIF (Common Interchange Format) resolution. Note that currently no international standard exists that describes the quality requirements of endoscopic archives. Recordings in SD quality are particularly relevant for this evaluation because existing endoscopic video archives mainly consist of such videos. They feature noticeably more border noise than HD videos. Moreover, the border area sometimes has a green or red tinge, which wastes bits that could better be used for the content area.

The first three data sets show actual operation scenes which take place inside the patient (ip). However, as our data set has been obtained randomly from a cooperating hospital it quite authentically reflects the way endoscopic videos are recorded and stored in practice. We observed that it also contains a considerable proportion of frames showing the environment outside the patient (oop) instead of an actual intervention. Out-of-patient frames are extremely noisy, apparently due to the ambient light of the operation room. They carry no relevant information and in a perfect world should be filtered out completely (we will address this issue in future work). Nevertheless, they are often captured unintentionally and hence are currently included in many stored recordings. Therefore, we split up the ip and oop components of the HD data set and evaluated the encoding efficiency separately. The difference in terms of border noise is illustrated in Figure 1.

We used the well-established x264 encoder (version 0.118.x) in conjunction with the ffmpeg API (version 0.7.13) to transcode the original MPEG-2 videos of our test sets to H.264/AVC. We chose this codec because it is one of the most popular state-of-the-art codecs and capable of encoding homogenous areas very efficiently with skipped macroblocks. As rate control mechanism we chose crf (constant rate factor) because it is the best suited mechanism for our application as we are not interested in a constant bitrate or fixed filesize but in constant quality. The crf rate control accepts a crf value in the range from 0 to 51 where a lower value involves a lower quantization and thus yields a higher quality [5].

For each original MPEG-2 video file, we used x264 (Main profile with default settings) to encode one H.264 version without modification and one with a black border overlay. We performed this step with crf values from 18 to 28 which is a "subjectively sane range" according to the ffmpeg encoding guide[1]. We exploit the fact that the crf rate control provides the same visual quality of the content area for both versions of each crf value, but uses a different overall bitrate due to the modified border regions in the overlay version.

To ensure that this assumption holds and a file size reduction does not involve a quality degradation we measured the PSNR (Peak signal to noise ratio) of the two transcoded versions. As reference we used a H.264 version encoded with crf 1 which is practically lossless compared to the original MPEG-2 video. For the PSNR calculation we only considered the circular content area because otherwise the overlay would be judged as quality degradation. The content areas can not be expected to be completely identical because the rate control mechanism behaves slightly different due to the modified global image content. This is also the reason for minor differences in the MB decisions inside the content area (see Figure 2). We found that the average PSNR difference between the two versions is 0.17 dB (between 0.14 dB for crf 28 and 0.22 dB for crf 18). This slight difference is neglectable so we can state that the content areas of both versions have the same visual quality. Based on that, it is straightforward to measure the gain in coding efficiency by looking at the filesize difference of the two versions for each crf value.

[1] http://ffmpeg.org/trac/ffmpeg/wiki/x264EncodingGuide

4. EXPERIMENTAL RESULTS

The detailed results of our experiments are summarized in Table 2 and illustrated in Figure 3. The values in Table 2 denote the weighted average filesize (or bitrate) reduction between the unmodified and the overlay version of all videos of the respective data set for a given crf value, i.e., for example the value of 6.74 means that the filesize has been reduced by 6.74%. Formally, they have been calculated according to equation (3).

$$R(d,q) = \frac{\sum_{i=1}^{n_d} (1 - S'_{d,i,q}/S_{d,i,q}) * L_{d,i}}{\sum_{i=1}^{n_d} L_{d,i}} \tag{3}$$

$R(d,q)$ is the bitrate reduction for data set d and crf value (quality) q,
n_d is the number of video files of data set d,
$S_{d,i,q}$ is the filesize of video i of data set d with crf q encoded without modification,
$S'_{d,i,q}$ is the filesize of video i of data set d with crf q encoded with a black border overlay,
$L_{d,i}$ is the (temporal) length of video i of data set d.

	Average bitrate reduction			
crf	ip (HD)	ip (SD)	ip (CIF)	oop (HD)
18	6.74	26.28	33.05	48.20
20	6.61	22.90	22.32	48.04
22	6.16	16.32	10.92	48.02
24	5.57	9.64	4.25	47.82
26	4.98	4.87	0.43	47.02
28	3.79	2.03	-1.15	44.27

Table 2. Coding efficiency improvement

For the HD in-patient videos only a modest gain in encoding efficiency of about 6% can be observed. There is no significant difference between the two different HD resolutions (1920x1080 and 1280x720). On the contrary, for the out-of-patient videos the border overlay tremendously reduces the bitrate by nearly 50%. For both data sets we can only observe a slight increase for lower crf values. The result for the HD in-patient videos can be explained by the fact that their original quality is already high. On the other side, the extremely high reduction rate for HD out-of-patient videos correlates to the ratio of border pixels to total pixels, which is in the range from 48-53% in our testset. This means that without a border overlay the encoder uses roughly the same number of bits for the noisy border region as for the content area.

For the low resolution videos the result is very different. Both SD and CIF resolution show a strong increase of encoding efficiency with higher quality. For crf 18 (which produces high quality with practically no visual loss compared to the original signal) the gain is about one third of the filesize, which is a substantial improvement. This result is a consequence of the lower quality compared to the HD in-patient videos which leads to a higher amount of border noise. What makes this result even more remarkable is the fact that the low resolution videos have an aspect ratio of 5:4 or 4:3 while the HD videos have a 16:9 widescreen aspect ratio. That means that the ratio of border pixels to total pixels is only about 30-40% while for the HD videos it is about 50%.

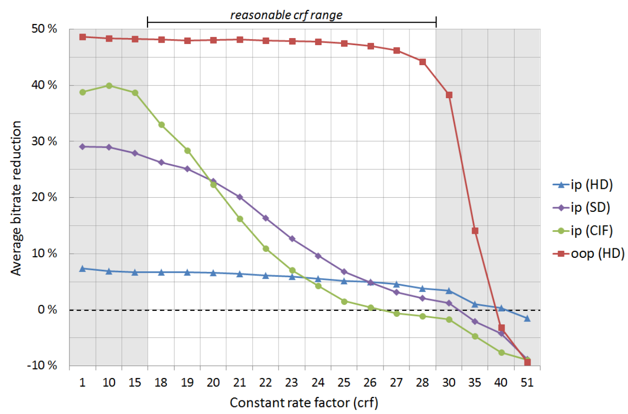


Fig. 3. Bitrate reduction for different crf values

The gain in encoding efficiency highly depends on the crf factor. The reason is that the higher the crf value, the more noise is filtered out by the stronger quantization itself, so the difference between the overlay version and the non-overlay version becomes smaller. Thus, the tradeoff of a low crf value that produces high quality for the content area is that it also retains the border noise if no overlay is used.

In Figure 3 we also illustrate the filesize difference for crf values outside the reasonable range (with a gray background and with larger steps). It shows that crf values below 18 do not give a much better result, although they produce much larger files. Interestingly, with the highest possible crf value of 51 all data sets yield a negative reduction value. That means that the version with the black border overlay is larger than the unmodified version. However, this counterintuitive result has no practical relevance because the video quality in this range is anyway far from being acceptable in practice.

5. CONCLUSIONS AND FUTURE WORK

In this paper we evaluated the impact on coding efficiency of superimposing the noisy border regions of endoscopic videos by a homogenous (black) border overlay. We demonstrated that this modification causes encoders to predominantly chose skipped macroblocks for the border regions. Thus, the overall bitrate can be reduced considerably without degrading the quality of the circular content area. In case of a limited constant bitrate the visual quality of the content area could even be enhanced because a higher ratio of the bitrate is available for the content area. As a convenient side effect, the decoding time is reduced as well because for the skipped MBs no coefficients and no motion vectors have to be decoded. Moreover, denoised borders have a more pleasing visual appearance and are less distracting. However, the extent of the bitrate reduction depends on quality-related encoding parameters. With a higher quantization, the amount of reduction decreases because the noise is subject to quantization as well.

Our results show that the impact of the border overlay on encoding efficiency also depends on the original quality of the processed videos. While for HD videos the bitrate reduction is rather modest (around 6%), a substantial reduction of up to 33% can be achieved for SD and CIF videos because their original quality is lower and the border noise is more intense. The most significant gain in encoding efficiency (around 50%) was observed for out-of-patient video segments which typically feature very heavy border noise. These irrelevant segments can congest medical video archives because the heavy border noise is misinterpreted as detail and therefore encoded with a high bitrate. In our future research, we plan to address the problem of automatic out-of-patient detection separately and more in-depth.

Additional future work includes user studies with surgeons on the necessary minimal required visual quality for endoscopic archives. A further interesting question is the gain in content based analysis, if a previous black-border and out-of-patient filtering has been performed.

6. REFERENCES

[1] Bernd Münzer, Klaus Schoeffmann, and Laszlo Böszörmenyi, "Detection of circular content area in endoscopic videos," in *Proceedings of the IEEE International Symposium on Computer-Based Medical Systems (CBMS'13)*, 2013.

[2] Bernd Münzer, "Requirements and prototypes for a content-based endoscopic video management system," M.S. thesis, Alpen-Adria University Klagenfurt, 2011, [online] Available: `http://ubdocs.uni-klu.ac.at/open/hssvoll/AC08958469.pdf`.

[3] J. Illingworth and J. Kittler, "A survey of the hough transform," *Computer Vision, Graphics, and Image Processing*, vol. 44, no. 1, pp. 87–116, Oct. 1988.

[4] Mathias Lux, Oge Marques, Klaus Schöffmann, Laszlo Böszörmenyi, and Georg Lajtai, "A novel tool for summarization of arthroscopic videos," *Multimedia Tools and Applications*, vol. 46, no. 2-3, pp. 521–544, Sept. 2009.

[5] L. Merritt and R. Vanam, "Improved rate control and motion estimation for h.264 encoder," in *IEEE International Conference on Image Processing, 2007. ICIP 2007*, Oct. 2007, vol. 5, pp. V –309 –V –312.