A VIDEO TRANSCODING SCHEME FOR E-LEARNING MULTIMEDIA APPLICATIONS

Nuno Santos, Pedro A. Amado Assunção
Polytechnic Institute of Leiria / ESTG – Institute of Telecommunications
Morro do Lena – AltoVieiro, 2401-951 Leiria, Portugal

Keywords: Transcoding, multimedia, MPEG, video objects.

Abstract: In this paper, we propose a segmentation based transcoding scheme for adapting MPEG-2 e-learning visual contents to heterogeneous environments. This is achieved by converting MPEG-2 video into MPEG-4 video objects with arbitrary shape and different semantic value in e-learning context. The transcoding scheme is based on a hybrid segmentation method, which employs both compressed and pixel domain techniques, for extraction of two video objects from MPEG-2 streams. The objective is two-fold: i) to enable individual object coding and manipulation; ii) to increase the scene coding efficiency. The results show that our hybrid segmentation method is capable of identifying the video objects of interest with good accuracy. Moreover, the transcoding efficiency of the proposed scheme is better than straightforward conversion from MPEG-2 to MPEG-4.

1 INTRODUCTION

Digital multimedia is part of an ever increasing field of applications where visual information plays the major role. In this context, the MPEG family of coding standards have undoubtedly contributed to the widespread use of compressed multimedia in many different domains. This is particularly true for the case of MPEG-2 (ISO/IEC, 1999) and MPEG-4 (ISO/IEC, 1999), since these two are among the most used in current multimedia services and applications. However, multimedia content delivery to different user contexts through diverse access networks, requires specific adaptation tools for providing Universal Multimedia Access (UMA) (Pereira and Burnet, 2003). For this purpose, several authors have proposed different types of transcoding as the best solution for dealing with adaptation problems in heterogeneous communication scenarios (Assuncao and Ghanbari, 1998; Reibman *et al.*, 2000; Shanableh and Ghanbari, 2000; Sun *et al.*, 2002).

In the case of transcoding from MPEG-2 to MPEG-4, several aspects have been addressed in previous work. For example, in (Takahashi *et al.*, 2001) a solution for the problem of motion vector reuse was proposed and in (Guo *et al* 2001; Xie *et al* 2003) different efficient methods are addressed for transcoding from MPEG-2 to MPEG-4. A common aspect of these transcoding methods is that of dealing with video frames or, using MPEG-4 terminology, rectangular video objects. We address a different type transcoding where MPEG-2 video is converted into MPEG-4 video objects with arbitrary shape.

Distance education and e-learning are increasingly important application domains where multimedia technology plays a relevant role. The particular characteristics of such environments give rise to new types of heterogeneous transcoding and media adaptation schemes for efficient representation and manipulation (Dorai *et al*, 2003).

In this paper we propose a transcoding scheme for adapting MPEG-2 coded video into MPEG-4 video objects of arbitrary shape for e-learning applications. The video scenes to be transcoded are constrained by the application specific context. They comprise two objects of interest: the whiteboard and the lecturer. In the proposed scheme, we exploit the fact that both the whiteboard and the lecturer might be encoded as independent video objects. Then, by taking advantage from the MPEG-4 coding tools, we propose a transcoding mechanism for matching MPEG-2 coded signals into separate MPEG-4 video objects. The proposed transcoding scheme relies on a hybrid domain spatio-temporal segmentation algorithm. The two objects referred to above are extracted from the coded video frames and then

Assuncao P. and Santos N. (2004).
A VIDEO TRANSCODING SCHEME FOR E-LEARNING MULTIMEDIA APPLICATIONS.
In *Proceedings of the First International Conference on E-Business and Telecommunication Networks*, pages 278-283
DOI: 10.5220/0001397702780283

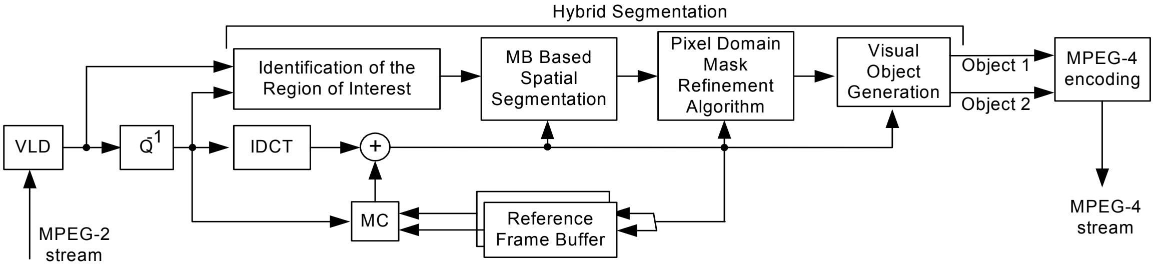


Figure 1: MPEG2 – MPEG4 Segmentation Based Transcoding

independently encoded by using different coding parameters, according to their inherent characteristics. The results show a good performance taking into account both objective and subjective quality as well as coding efficiency.

This paper is organised as follows. In the next section we address the specific characteristics of the visual contents that we are dealing with in this work. In section 3 we describe the hybrid segmentation algorithm used for fast extraction of video objects. The experimental results are presented in section 4 and finally section 5 concludes the paper.

2 THE VISUAL CONTENT

The type of multimedia contents that we deal with in this work involves recorded MPEG-2 video originally captured from a typical classroom scene. This consists of a teacher speaking, writing on a whiteboard and moving in front of the whiteboard area. In the MPEG-4 context this scene contains two video objects with different semantic value for the human observer: the lecturer and the whiteboard. Figure 1 shows one picture of the visual content used in our experiments.

In regard to subjective quality it should be stressed that motion smoothness is more important than texture accuracy in the case of the lecturer, while texture is much more important than motion in the case of the whiteboard. This means that different requirements should be defined for the temporal and spatial quality of each video object in order to achieve better transcoding efficiency. Note that the most active periods correspond to different types of motion, such as walking, writing on the whiteboard, gesture and speaking.

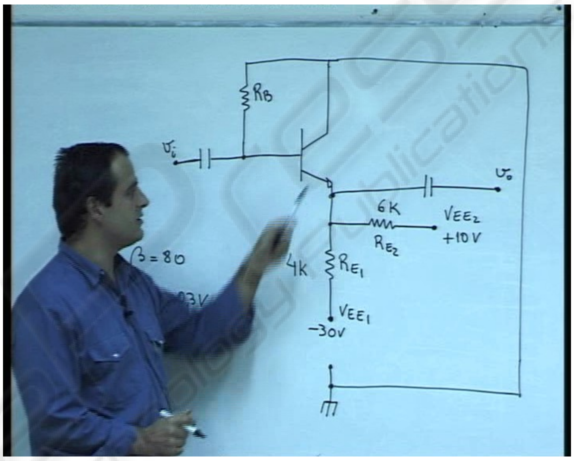


Figure 2: A typical image from a video sequence used in e-learning environment

The whiteboard is where the lecturer writes down pedagogical contents for supporting and complementing the oral explanations. The main characteristic of this video object consists of its relatively slow motion and high texture detail. The slow motion results from the human writing speed on such type of board whereas the high spatial detail is a consequence its specific visual contents, i.e., characters and diagrams written with a marker. Therefore, this video object can be efficiently encoded at reduced temporal rates such that more bits are allocated to encode the texture information, i.e., higher spatial quality.

3 SEGMENTATION BASED TRANSCODING

The segmentation based transcoding architecture proposed in this paper is depicted in Figure 2. It is comprised of a modified MPEG-2 decoder which includes a hybrid video segmentation algorithm and a MPEG-4 visual encoder. The hybrid segmentation process operates in both the DCT and pixel domains (Kim *et al.*, 1999; Yu *et al.*, 2003). As it can be seen in the figure, the input MPEG-2 video stream is transcoded into two MPEG-4 video objects by using segmentation before the MPEG-4 encoding stage.

3.1 DCT domain coarse segmentation

The DCT domain algorithm operates on predefined temporal window where a coarse spatial region is identified as the location of the lecturer. Since this is a low motion sequence the boundaries of the temporal window were set at the I pictures of the MPEG-2 stream. This dynamic region is found by a fast algorithm that evaluates the DC distance measure between two MB located in the same spatial position of consecutive I pictures. The DC distance $d_{i,n,m}$ between two MB in pictures n and $n+t$, $t \in N$, both with MB address i, is determined as follows:

$$d_{i,n,t} = \left| a_{i,n} - a_{i,n+t} \right|$$

and $a_{i,n}$ is given by,

$$a_{i,n} = \left(\sum_{k=0}^{3} Y_{i,n}(k) \right) / 4 + Cb_{i,n} + Cr_{i,n}$$

where $Y_{i,n}(k)$ is the DC coefficient of the luminance block k in MB i of picture n, and $Cb_{i,n}$, $Cr_{i,n}$ are the DC coefficients of the corresponding chrominance blocks. Part of the dynamic region is comprised of those MB whose DC distance is greater than a threshold *Th*. The complete region of interest is then found after a further processing step that fills in the "holes" left by the previous one, i.e., those MB that belong to the inside part of the moving region and were not identified because they have similar texture to that of their neighbours. This DCT domain processing allows identification of the slow moving region in a temporal segment of the sequence. Note that the spatial region found through this process contains more MB than actually those which belong to the object of interest. However, this is an extremely fast and efficient method for identifying the coarse region within which the lecturer can be found.

The result is a set of MB addresses that define a much smaller region containing the video object. From the experimental results we found that, in the type of sequence we are using, the size of this region is about 25% - 43% of the total image size. Therefore the pixel domain mask refinement that follows this stage operates on a much smaller set of data.

3.2 Pixel domain mask refinement

The actual size of the moving video object is smaller than that of the DCT domain mask obtained in the previous step because the spatial region is identified in a GOP basis. Thus, the resulting mask includes, not only the object of interest, but also the surrounding area where it moves between two I pictures.

One of the tasks of the pixel domain refinement algorithm is to shrink the DCT mask up to the actual boundaries of the object. As a consequence, the object shape is also refined up to the pixel level. This also improves the quality of the segmentation because the first mask obtained in the DCT domain is coincident with the MB structure of the original picture, hence it is a stepwise boundary.

The pixel domain algorithm implemented for obtaining the refined masks involves 4 steps: 1- median filtering the region of interest to eliminate noise (Yin *et al.*, 1996); 2- histogram analysis of the spatial region of interest to determine the best threshold for splitting the pixels into two sets; 3- obtaining a segmentation mask based on the histogram and 4- post-processing to eliminate isolated small groups of pixels which do not belong to the main area (characters, lines, etc).

4 EXPERIMENTAL RESULTS

In order to evaluate the performance of the segmentation based transcoding scheme, we have used MPEG-2 video streams, which are currently being used for e-learning purposes within the intranet of our campus. These are of the type shown in Figure 1. We have carried out several subjective tests in order to find out the minimum bandwidth that achieves a good subjective quality for this type of application since poor picture quality is not acceptable because it may lead to additional learning difficulties. From these tests we have found that 2 Mbps provides an acceptable visual quality. Then the MPEG-2 stream was transcoded into an MPEG-4 visual stream combining two video objects. The objective is two-fold, *i.e.*, to enable individual object coding and manipulation as well as to increase the scene coding efficiency. While the former is achieved by proper segmentation, the latter greatly depends on both the efficiency of the coding algorithm and the set of coding parameters.

4.1 Segmentation

The hybrid segmentation algorithm described in Section 3 was used to produce the segmentation mask for extracting the two objects of interest from the MPEG-2 video stream. In order to ease the comparison we show the results for the picture shown in Figure 1. In Figure 3 we show the coarse region obtained from the DCT domain algorithm and in Figure 4 the corresponding spatial region. As we have pointed out before, this region is greater than the actual moving object (the lecturer). As we pointed out before, the mask precision is limited to MB level because this is the processing data unit in the DCT domain. In Figure 5 we show the region identified by the histogram based algorithm and post processing and Figure 6 shows the video object identified through this process.

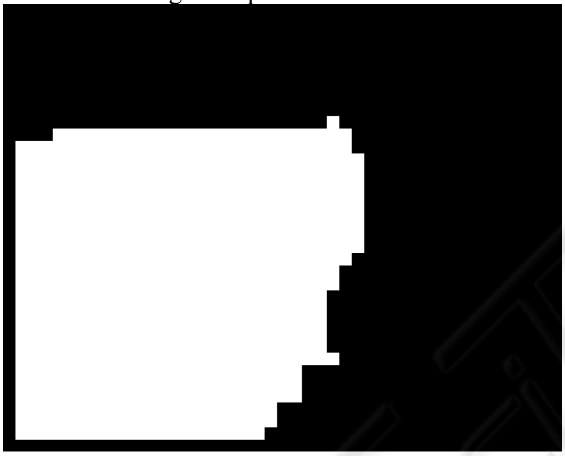


Figure 3: Coarse mask obtained in the DCT domain.

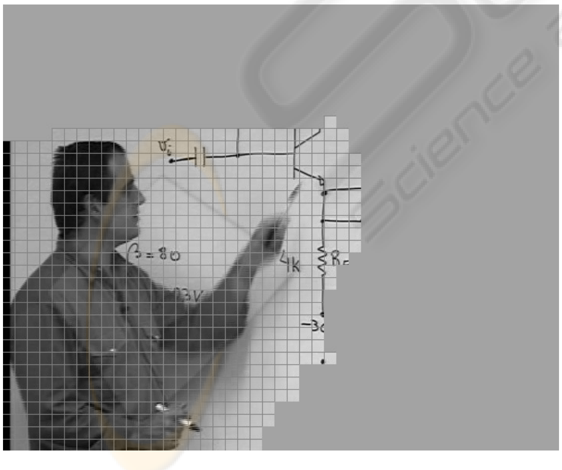


Figure 4: Coarse spatial region obtained from the corresponding mask

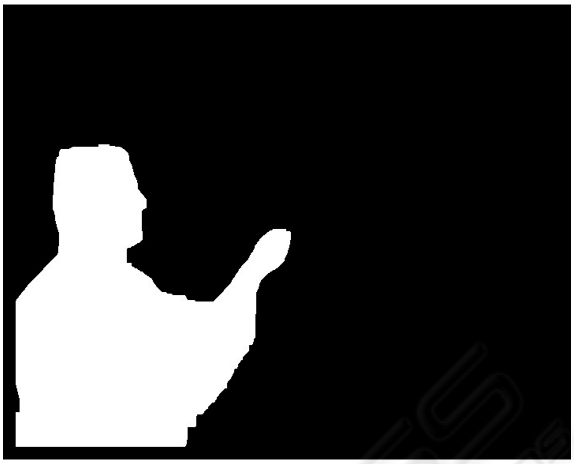


Figure 5: Refined mask

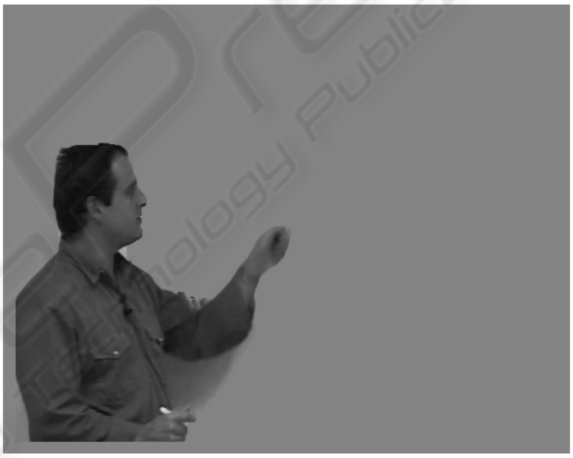


Figure 6: Video object

4.2 Transcoding efficiency

In order to evaluate the picture quality under a significant transcoding ratio, we have set the output bit rate to 500 Kbps and we compare three different transcoding schemes. In all cases we have used the same input video sequence, which was available in the server at 2 Mbps.

For reference and comparison with the proposed scheme we have used straightforward transcoding from MPEG-2 to MPEG-2 and from MPEG-2 to MPEG-4, using a single rectangular object. In the case of the proposed scheme, we have taken into account the inherent characteristics of each visual object, as pointed out before. Then for each video object we have set the same output bit rate of 250 kbps but different temporal rates. The lecturer was encoded at 25Hz whereas the whiteboard was encoded at 6.25Hz. For comparison with the video frames, after decoding the two objects these were combined to form frames again.

As we can observe in Figure 7, the proposed scheme achieves a good performance comparing with both references. By using different coding parameters for each video object, the transcoded pictures have better spatial quality in the whiteboard area, mainly because of its reduced temporal rate which allows more bits to encode the texture. The composition problem that arises when different video objects are displayed at different frame rates may be overcome by filling in the missing areas with pixels from the surrounding area. However, this issue is not addressed in this paper. The same behaviour as shown in Figure 7 is obtained for other transcoding ratios.

5 CONCLUSION

The transcoding scheme proposed in this paper is suitable for e-learning applications where MPEG-2 to MPEG-4 conversion might be useful. The experimental results show that a good performance is achieved by choosing different temporal rates for video objects according to their specific characteristics.

A possible application of this type of transcoding is in wireless access where the user may receive the audio and only the whiteboard visual information at much lower bit rates but still with an acceptable quality of service.

REFERENCES

Assuncao P. and Ghanbari M, 1998. A frequency domain transcoder for dynamic bit rate reduction MPEG-2 bit streams, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 8, No 8, pp. 953-967.

Dorai, C., Oria V. and Neelavalli V., 2003. Structuralizing Educational Videos Based on Presentation Content, *IEEE International Conference on Image Processing*, Barcelona-Spain.

Guo W., Lin L., Zheng W and Zheng W., 2001. Mismatched MB Retrieval from MPEG-2 to MPEG-4 Transcoding. IEEE Pacific Rim Conference on Multimedia, Beijing, China.

ISO/IEC 13818-2, 1995. Generic Coding of Moving Pictures and Associated Audio - Part 2: Video.

ISO/IEC 14496-2, 1999. Information Technology - Generic Coding of Audio-Visual Objects – Part 2: Visual, Vancouver.

Kim M., Choi J. G., Kim D., Lee H., Lee M. H., Ahn C. and Ho Y-S., 1999. A VOP Generation Tool: Automatic Segmentation of Moving Objects in Image Sequences Based on Spatio-Temporal Information, *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1216-1226, Vol. 9, No 8.

Pereira F., Burnet I., 2003. Universal Multimedia Experiences for Tomorrow, *IEEE Signal Processing Magazine*, vol. 20, No. 2.

Reyes G., Reibman A., Chang S-F., Chuang J., 2000. Error Resilient Transcoding for Video over Wireless Channels*, IEEE Journal on Selected Areas in Communications*, Vol. 18, No. 6.

Shanableh T., and Ghanbari M., 2000. Heterogeneous video transcoding to lower spatio-temporal resolutions and different encoding formats, *IEEE Transactions on Multimedia*, Vol. 2, No 2, pp. 101-110.

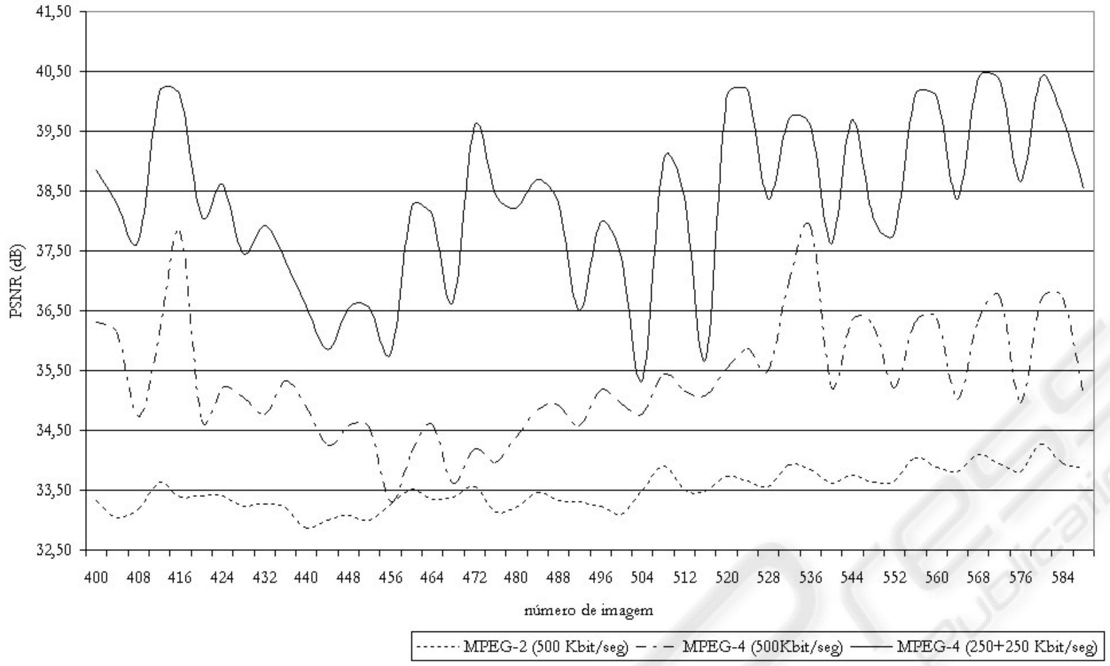
Takahashi K., Satoh K., Suzuki T., Yagasaki Y., 2001. Motion Vector Synthesis Algorithm for MPEG2-to-MPEG4 transcoder, Visual Communications and Image Processing, Proceedings of SPIE, Vol. 4310, pp. 872-882.

Xie R., Liu J. and Wang X. 2003. Efficient MPEG-2 to MPEG-4 Compressed Video Transcoding, Visual Communications and Image Processing, Proceedings of SPIE Vol. 4671, pp. 192-201, Lugano-Switzerland.

Xin J., Sun M-T., Choi B-S., Chun K-W., 2002. An HDTV to SDTV Spatial Transcoder, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 12, No 11.

Yin L., Yang R., Gabbouj M., Neuvo Y., 1996. Weighted Median Filters: A Tutorial, *IEEE Trans. on Circuits and Systems*, vol. 43, n 3, pp. 157-192.

Yu X-D., Duan L-Y. and Tian Q., 2003. Robust Moving Video Object Segmentation in the MPEG Compressed Domain, *IEEE International Conference on Image Processing*, Barcelona-Spain.



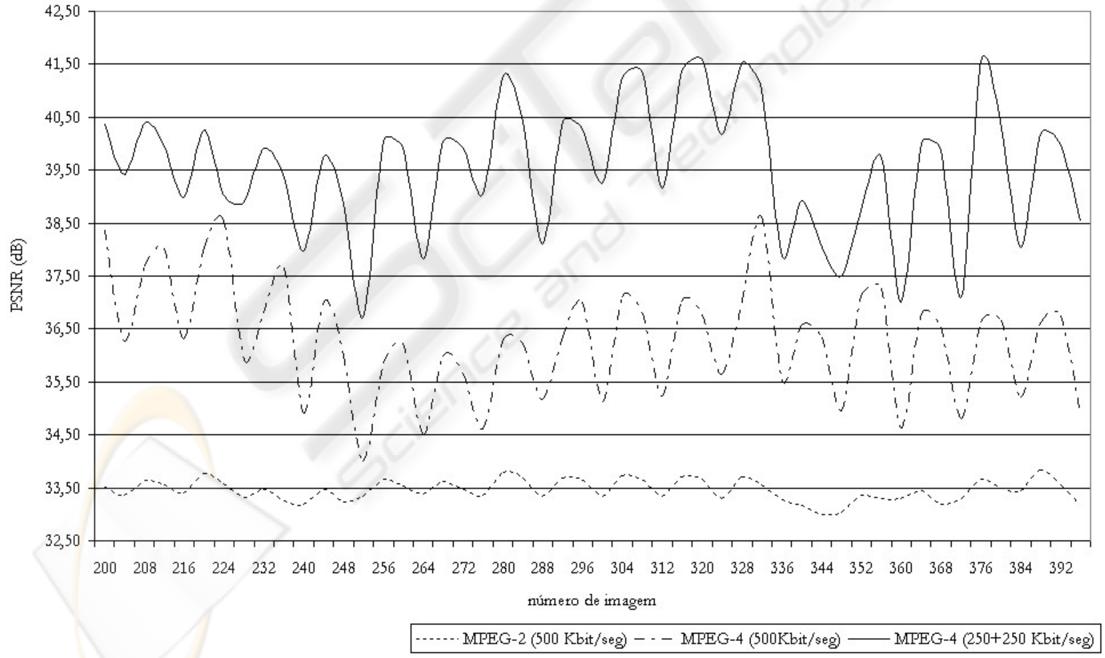


Figure 7: PSNR of the transcoded sequence