Homogenous Video Transcoding of H.264/AVC Intra Coded Frames

Jarosław Marek

Poznan University of Technology
Chair of Multimedia Telecommunications and Microelectronics
ul. Polanka 3, 60-965 Poznań, Poland
jmarek@multimedia.edu.pl

Abstract. The main goal of transcoding is to change bit rate of video sequence. This can be done by cascaded connection of decoder and encoder, known as Cascaded Pixel Domain Transcoder (CPDT). Decoding and re-encoding video bit stream always gives lower image quality than encoding original sequence. This paper presents a new technique of video transcoding that is able to deliver image quality superior to CPDT and has lower computational complexity. The technique is restricted to homogenous (within the same bit stream format) transcoding of bit streams encoded according to H.264/AVC(MPEG 4) standard specification. The standard defines different types of encoded frames but proposed technique is designed for I(ntra) type frames only.

Keywords: AVC, MPEG 4, H.264, video transcoding, video encoding, re-quantization, bit rate reduction.

1 Introduction

Encoded video sequences can be transmitted with various bit stream formats and bit rates. For transmission over heterogeneous networks, changing of sequence's bit rate is sometimes required to accommodate to channel's throughput. This can be simply done when video bit stream is scalable. When it is not, input bit stream should be transcoded to lower bit rate.

Transcoding operation can be performed in various ways [1-3]. One of them is to connect decoder which decodes input bit stream with encoder which forms a new bit stream. Such a scheme is depicted in Fig. 1 and it is usually referred as Cascaded Pixel Domain Transcoder (CPDT) [2, 3]. The greatest advantage of this architecture is that it is simple to implement. There are good decoders and encoders available which can be used. Unfortunately, computational complexity of this architecture is very high. Moreover, it does not ensure the best image quality that can be achieved.

Another approach is to integrate decoding and encoding processes into a single transcoding process. This can reduce computational complexity of the solution, as not all encoding/decoding operations are required for transcoding. Moreover, knowledge of an input bit stream in the encoding part, can be used to achieve higher image quality. Solution of this type is presented in this paper. Proposed transcoding technique is able to transform input video bit stream into another video bit stream without image reconstruction and encoding processes. This is known as transcoding in frequency domain. The technique allows for homogenous transcoding [1], that is transcoding of bit streams within the same format. Generally, all types of frames, encoded in a given format, can be transcoded. However, this technique is designed for transcoding of (I)ntra type frames, encoded according to H.264/AVC(MPEG 4) standard specification [4]. In the following sections one can find grounds for a design of a new transcoder architecture, presentation of used techniques and achieved results.

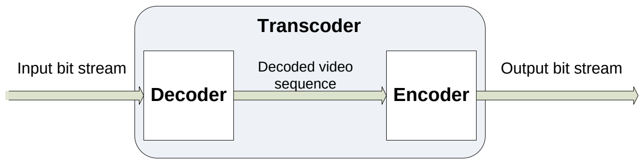

Fig. 1 Cascaded Pixel Domain Transcoder (CPDT).

2 Quality Loss Caused by CPDT

For a given bit-rate, quality of transcoded (and decoded) sequence is always lower than quality of sequence decoded from a bit stream achieved by encoding original sequence. The reason for this is that in CPDT there is no information about original image and encoding base on distorted image (result of first time coding). One can find in a literature references to CPDT coding efficiency [5, 6] but there are no comprehensive results of research. This is why quality loss tests were performed for this kind of transcoder. Results of these tests allow to identify weak points of this solution as well as allow to compare them with a newly proposed techniques. In Fig. **2** one can find an interpretation of quality loss value used in this paper.

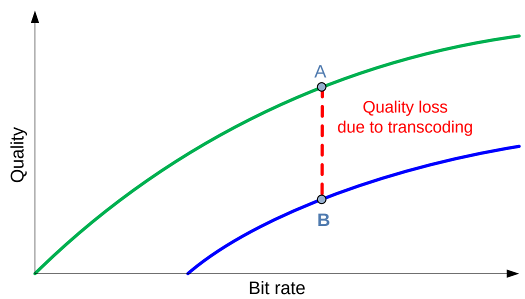

Fig. 2 Quality loss caused by transcoding.

The higher curve refers to first-time coding. It is the highest image quality that can be obtained for a given bit rates and for a given bit stream format. The lower curve refers to transcoding, which introduce some quality degradation. Point B shows quality and bit rate of re-encoded sequence. Point A indicates quality of first time coded sequence with the same bit rate as for point B. Loss in quality can be calculated as a distance between these two points.

All tests were conducted with H.264/AVC(MPEG 4) reference software (JM version 13.2). There were used 11 high quality sequences, namely: bluesky, city, crew, harbour, ice, pedestrian, riverbed, rushour, soccer, station, sunflower, station. Each sequence had 100 frames and 704x576 resolution. Some of these sequences originally have 1280x720 resolution. Smaller versions were achieved by cropping original frames to desired resolution (only images' centres were used).

At first, original sequences were encoded with a set of ten different QP_F values. 'QP' is a parameter defined by H.264/AVC(MPEG 4) standard determining the quantizer that should be used. The higher value QP parameter has, the wider quantization step size is. 'QP_F' refers to a value of QP used for first-time coding. After encoding process all achieved bit streams were decoded. Each decoded sequence has been re-encoded with consecutive QP_T values, starting from $QP_T=QP_F$. 'QP_T' refers to the value of QP used for transcoding. In a final step quality losses were calculated as presented in Fig. 2. This gives a set of curves for every sequence. In Fig. 3 there is an example of a single curve that can be obtained for a sequence firstly encoded with QP_F parameter and then transcoded with eleven QP_T values, as described above.

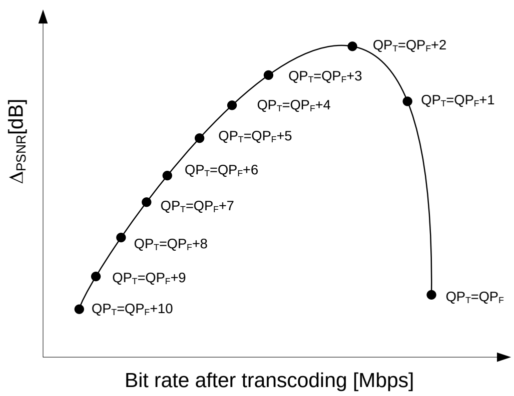

Fig. 3 Example of quality loss (Δ_{PSNR}) caused by transcoding. Points on a curve indicate re-encoding with different QP_T values from a single sequence encoded with QP_F value. The absolute difference in QP_T between neighbouring points laying on the same curve is 1. First point (from right) on the curve indicates transcoding with QP_T equal to QP_F.

In this paper, results for only one sequence is presented. They can be found in Fig. **4**. It has to be stated that these results are representative for all other sequences. For all the data gathered, when QP_T equals QP_F, quality loss can be as high as 0.7dB (sequence: sunflower, $QP_T=QP_F=28$). This is a huge image quality degradation, especially while the encoding parameters are not being changed from first encoding to transcoding. There are two reasons for this situation. First of all, coding of DC coefficients in Intra16x16 prediction mode (defined by H.264/AVC(MPEG 4) standard) is not always reversible. There are combinations of DC coefficients for which decoding and re-encoding them with the same QP will result in changed coefficients. Second reason is that encoder uses Lagrangian optimization (which balances between quality degradation and bits generated) for block and macroblock mode decisions. There are situations when encoder has to choose whether to use block/macroblock mode with low quality degradation and few bits needed or mode with no quality degradation and more bits needed for encoding. Choosing first option in case of transcoding with the same QP results in additional image degradation.

The highest loss in decoded sequence quality when transcoded is when QP_T-QP_F difference is from 1 to 3. This loss can be as high as 2dB and generally decreases with the increase of QP_F. It can be seen that when difference between QP_T and QP_F values raises, quality loss quickly decreases. The above suggests that there is a need for technique that could deliver lower quality loss for small bit rate reduction.

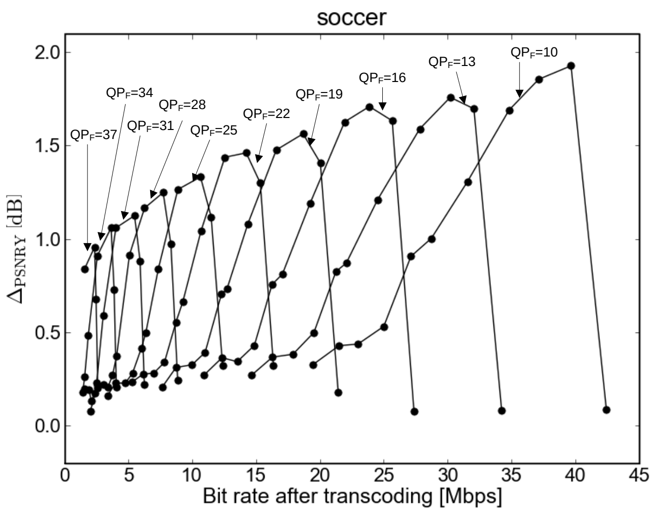

Fig. 4 Quality loss caused by transcoding of 'soccer' sequence. Points on each curve indicate re-encoding with different QP_T values from a single sequence encoded with QP_F value. The absolute difference in QP_T between neighbouring points laying on the same curve is 1. First point (from right) on each curve indicate transcoding with QP_T equal to QP_F.

3 Proposed Requantization Technique

H.264/AVC(MPEG 4) is different from previous video coding standards because it engages prediction of image samples for intra coded macroblocks instead of DC transform coefficients prediction. Modification of image samples in a single macroblock or block within a macroblock can change prediction of neighbouring macroblocks. This prediction is used for neighboring macroblocks reconstruction. If its change is not compensated by transform coefficients modification, there will be a decoding error. If there will be the decoding error, prediction for the next macroblocks will be also wrong and they will be wrongly decoded. And so on. This leads to the conclusion that when one block or macroblock is being changed, all next blocks/macroblocks have to be also changed to prevent error propagation.

Requantization of macroblock results in a modification of decoded image. Basing on information from previous paragraph, it may be stated that requantization of single macroblock forces requantization of all next macroblocks. It is because prediction for neighbouring macroblocks changes and these changes have to be compensated.

Similar situation happens when during transcoding macroblock mode is being chosen different than macroblock mode used for encoding original sequence. As a result, decoded image samples are being changed and all the above applies here. Prediction for following macroblocks is being changed and finally, requantization for all of them is required.

Macroblock mode selection has a great impact on coding efficiency. Moreover, macroblock mode decision depends on target bit rate [7]. For these reasons, when re-encoding video, new macroblocks' modes should be chosen. However, it can be assumed that for transcoding with small differences in bit rate, quality loss caused by wrongly chosen block/macroblock modes is also small.

The main idea for minimizing quality loss caused by transcoding is to avoid re-quantization error. This can be done when:

1. Prediction image for following blocks after requantization remains unchanged,
2. Transcoded sequence uses the same macroblock modes as sequence coded for the first time.

When transcoder does not change macroblock modes, all of them are known for entire image prior to re-encoding. This gives an information which of coded pixels will be used to form prediction for neighbouring blocks. Furthermore, knowledge about neighbouring blocks modes allows to determine whether given pixel will be used directly to form prediction image or its value will be used to find average of a few pixels (this average value will be used for prediction image preparation). When pixels will be used directly, their values cannot change after requantization. When latter applies, pixel values can be modified after requantization, but sum of pixels used for prediction have to stay unchanged. The same applies for luma as well as for chroma components.

Finally, there are blocks not used for prediction. They can be modified freely as their modification has no impact on the rest of the image. The most obvious examples are blocks of Intra16x16 mode coded macroblocks, excluding the lowest row and the most right column. These are never used for prediction as it is a rule defined by H.264/AVC(MPEG 4) standard.

Two schemes for coefficient modification are proposed for requantization. In both of them, consecutively, in a reverse zig-zag order (defined by H.264/AVC(MPEG 4) standard), non zero DCT coefficients are checked, if they can be modified according to rules presented above. There are two methods of modification:

1. The coefficient value is set to 0, if its amplitude is not greater than a given threshold;
2. Amplitude of the coefficient is reduced by no more than a threshold value.

These methods are supplementary, can be used together or separately, and allow for a bit rate control.

4 Experiments

Experiments were conducted with the software and sequences described in a Section 2. All tests were performed according to following procedure:

1. Take original sequence and encode it with a set of ten different QP_F values.
2. Decode all bit streams resulting from a step 1.
3. Re-encode each sequence achieved in a step 2 with ten consecutive QP_T values starting from QP_F.
4. Take each bit stream achieved in step 1 and transcode it using proposed technique with two proposed schemes and thresholds ranging from 1 to 4. This gives 8 transcoded bit streams for each input bit stream.

Values of bit rate and PSNR obtained during realization of steps 3 and 4 were used to form rate-quality curves for CPDT and proposed technique respectively. Next, differences in PSNR (distance from rate-quality curves as illustrated in Fig. 2) were calculated for bit rates equal to these achieved in step 4. As there are no samples for CPDT with exactly the same bit rates, values of PSNR measure were achieved by using spline interpolation. Example results for 'soccer' and 'sunflower' sequences can be found in Fig. 5 and Fig. 6. Proposed technique is able to deliver better results than CPDT for all bit rates. However, the results depend on video sequence and bit rate. Generally, for high values of QP_F, proposed technique is better than CPDT for both schemes and all thresholds, and with the increase of QP_F some scheme and threshold combinations can give worse results.

Fig. 7 depicts the highest gain in PSNR for a proposed technique comparing to CPDT, for all analyzed sequences. Supplementary to the above is Fig. 9 which depicts bit rate reduction in relation to bit rate of firstly coded sequence (not transcoded). Additionally, Fig. 8 depicts the lowest gain (which can be at some points considered as the highest loss) in PSNR for a proposed technique comparing to CPDT. Experiments showed that the above is connected with the highest bit rate reductions achieved. Values of these reductions, comparing to firstly coded sequences, can be found in Fig. 10. Proposed technique is designed for small bit rate reduction and as can be seen in Fig. 10 it is able to make bit stream over 18% smaller. However, this reduction is not guaranteed and can be equal to 4% for some sequences.

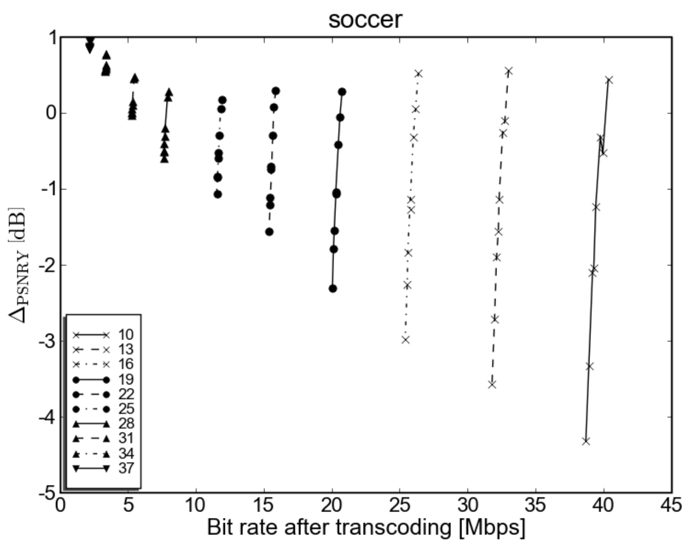

Fig. 5 Gain in PSNR for luma component achieved by using proposed technique instead of CPDT. Each curve represent transcoding of a single sequence encoded with a QP_F given in a legend.

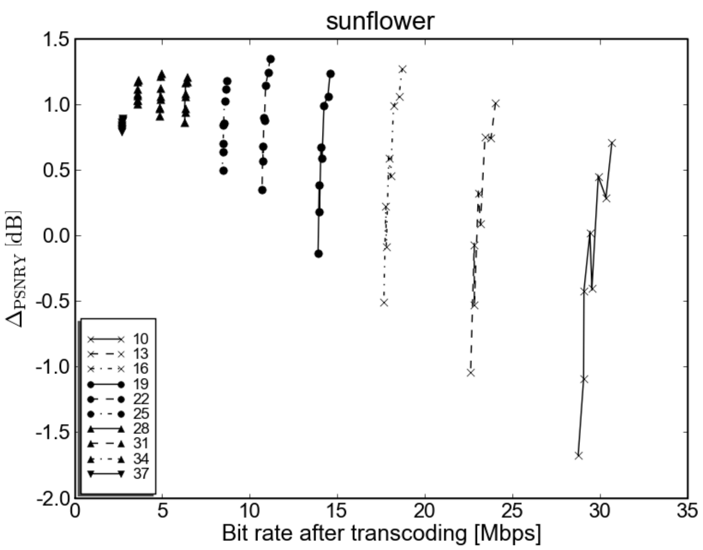

Fig. 6 Gain in PSNR for luma component achieved by using proposed technique instead of CPDT. Each curve represent transcoding of a single sequence encoded with a QP_F given in a legend.

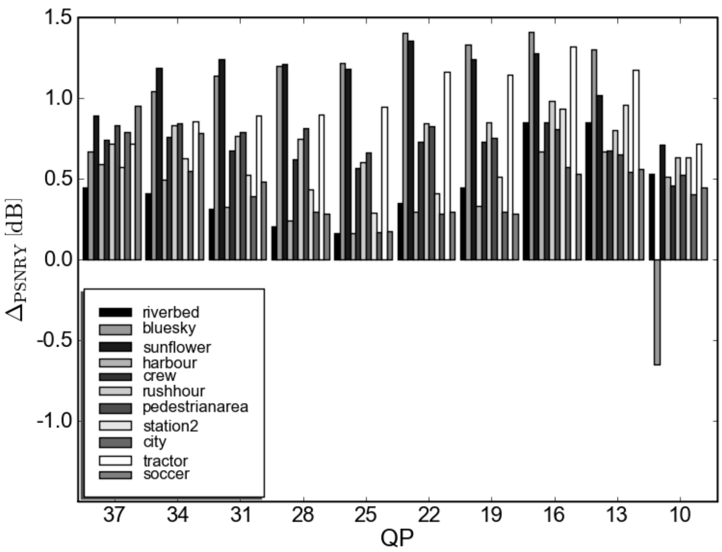

Fig. 7 Maximum gain in PSNR of luma component achieved by using proposed technique instead of CPDT.

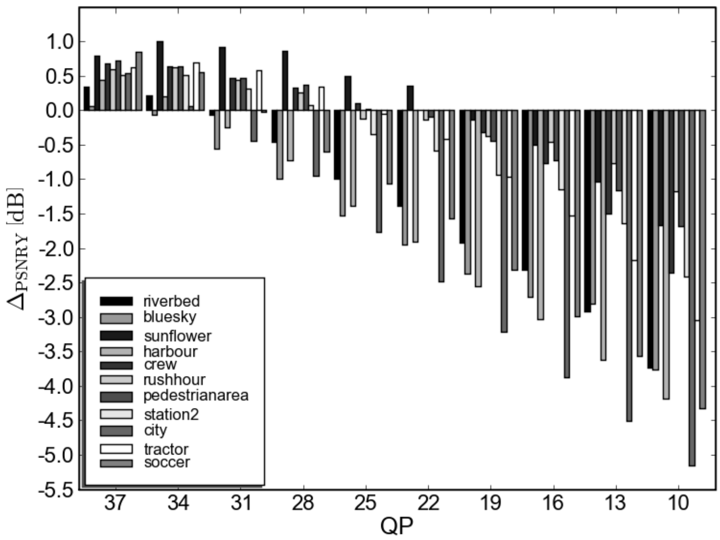

Fig. 8 Minimum gain in PSNR of luma component achieved by using proposed technique instead of CPDT.

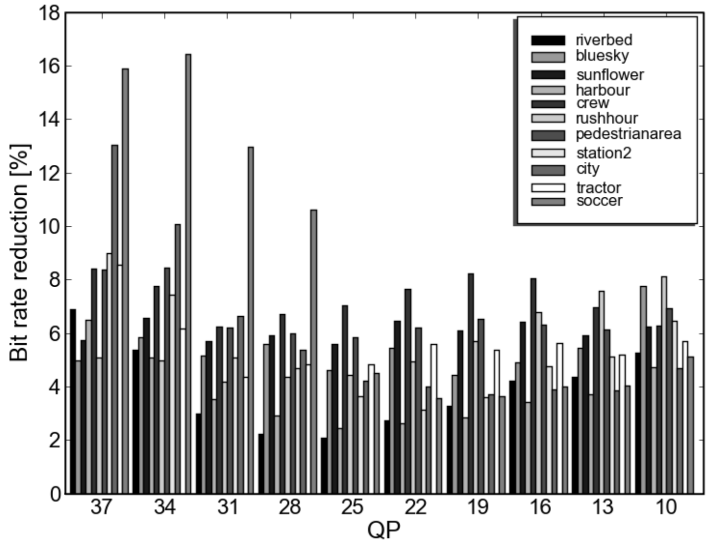

Fig. 9 Bit rate reduction for proposed transcoding technique, achieved for the highest gain in PSNR luma.

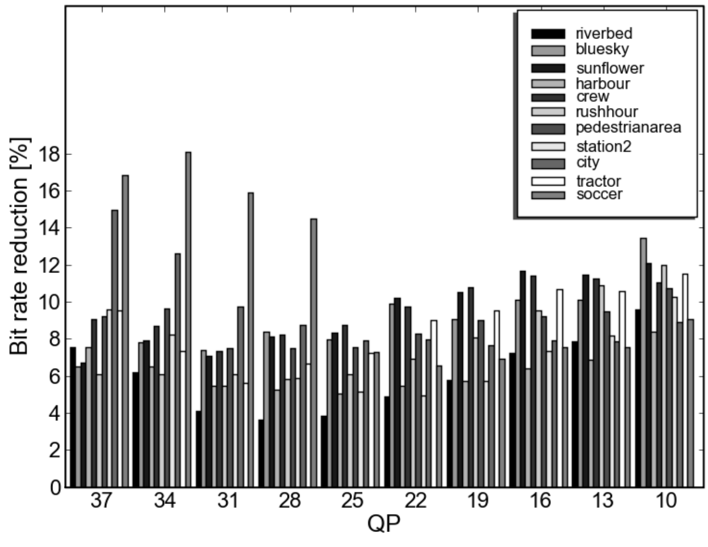

Fig. 10 Maximum bit rate reduction for proposed transcoding technique achieved in tests.

5 Conclusions

In this paper, new technique of H.264/AVC(MPEG 4) bit streams transcoding is proposed. It operates on intra coded frames exclusively and it is performed in the transform domain. It does not require image decoding prior to new bit stream formatting as well as very computationally intensive task of macroblock mode decision. As a result proposed technique requires much less operations than CPDT.

The technique is suitable for small bit rate reduction. For 11 test sequences, the highest bit rate reduction was 18,1%. In this case image quality was not as good as the one achieved by CPDT. Putting a criterion of supporting better quality than CPDT, 16.43% bit rate reduction can be achieved.

By using proposed technique, one can achieve as much as 1.4dB better results in PSNR, than employing CPDT scheme. However, it should be mentioned, that for high bit rates (low QP_F values) there are situations when using proposed technique can result in worse image quality.

Acknowledgments. This work was supported by the public funds as a research project in years 2007-2008.

6 References

1. Ahmad, I., Wei, X., Sun, Y., Zhang, Y.-Q.: Video Transcoding: an Overview of Various Techniques and Research Issues. IEEE Transactions on Multimedia, vol. 7, pp. 793-804 (2005)
2. Vetro, A., Christopoulos, C., Sun, H.: Video Transcoding Architectures and Techniques: an Overview. Signal Processing Magazine, IEEE vol. 20, pp. 18-29 (2003)
3. Xin, J., Lin, C.-W., Sun, M.-T.: Digital Video Transcoding. Proceedings of the IEEE, vol. 93, pp. 84-97 (2005)
4. Advanced Video Coding for Generic Audiovisual Services. Version 8 (including SVC extension). ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG-4 AVC), ITU-T and ISO/IEC JTC 1 (2007)
5. Zhang, P., Lu, Y., Huang, Q., Gao, W.: Mode Mapping Method for H.264/AVC Spatial Downscaling Transcoding. In: Proc. IEEE Int. Conference on Image Processing (ICIP), vol. 4, pp. 2781- 2784, Singapore (2004)
6. Notebaert, S., De Cock, J., Van de Walle, R.: Improved H.264/AVC Requantization Transcoding Using Low-Complexity Interpolation Filters for 1/4-Pixel Motion Compensation. In: Computational Intelligence in Image and Signal Processing (CIISP), IEEE Symposium on, pp. 307-312, Honolulu (2007)
7. Lefol, D., Bull, D.R.: Mode Refinement Algorithm for H.264 Inter Frame Requantization. In: Proc. IEEE Int. Conference on Image Processing (ICIP), pp. 845-848, Atlanta, USA (2006)