Skipping Strategies for High Definition H.264/AVC Real Time Video Coding

Susanna Spinsante, Ennio Gambi, and Franco Chiaraluce

Original scientific paper

Abstract— **This paper discusses the feasibility of High Definition H.264/AVC video coding in the Baseline profile, for real time applications like videoconferencing. By means of suited SKIP strategies and exploiting the higher resolution, redundancy and visual information amount of the High Definition formats, with respect to Standard Definition, it is possible to speed up significantly the Mode Decision procedure, without affecting the final perceived quality.**

Index Terms— **H.264/AVC, Profile, High Definition, SKIP, Fast Mode Decision**

I. Introduction

An H.264/AVC [1] video encoder carries out a number of encoding processes, including motion estimation, motion compensation, transform, quantization, and entropy coding. Among them, the mode decision process, to select the most suitable coding mode for each macroblock (MB) consisting of 16×16 luma samples and associated chroma samples, can be computationally intensive. The encoder may choose to encode a MB using Intra prediction from neighboring samples in the same frame or field, or using Inter (motion compensated) prediction from samples in previously coded frames or fields. Multiple Intra coding and motion compensated modes, and multiple reference frames are supported by the standard: this flexible choice of coding modes leads to an efficient compression, but at the expense of an increase in the computational complexity.

In order to reduce the encoder computational burden during the mode decision process, low-complexity (or *fast*) algorithms have been proposed, that aim at finding the prediction MB without the need of an exhaustive evaluation of all the available options. As a matter of fact, examining all possible modes takes most of the time out of the total encoding time. The final MB selected by a fast algorithm may be not the optimum MB: a trade-off between distortion and computational time reduction must be accepted. A vast body of literature exists on fast mode decision methods (see [2], [3], [4] for example, where further references can be found), to be applied in the framework of the ITU H.264/AVC video coding standard. However, most of these techniques, aimed at reducing the computational cost beared by an encoder to find the best coding mode for a given macroblock, are still proposed for CIF or QCIF sequences, whereas not so many examples of fast techniques specifically conceived for usage in real time High Definition (HD) systems can be found in the literature. On the other hand, at the authors' best knowledge, numerical results about the performances provided by these techniques when applied to HD video sequences in real time services are difficult to find as well.

Since a long time, our interest has been focused on videoconference applications, whose commercial and technical impact has become more and more important. At present, HD represents the evolution of videoconferencing, being able to enhance the overall viewing experience, by providing more vibrant and realistic colors, sharp and smooth movements. The H.264/AVC standard has been proven ideal for videoconferencing: although it requires more processing power than the previous H.26*x* algorithms, it provides good video transfer and low latency encoding and decoding, that result in smoother and more natural video flow.

Among the options available in the standard, the so-called Fidelity Range Extension (FREXT) profile, conceived to support specific applications like content contribution, studio editing and post processing, should be the most proper one to manage, in the best way, the HD coding. Really, this profile cannot be used in constrained environments (real time services, like videoconferencing) because of the high processing times it involves, which are in contrast with real time transmission requirements. As a matter of fact, the Baseline profile, originally conceived to support real time conversational services, such as videoconference and video telephony, seems the only one able to meet constraints on the time delay.

According with these considerations, it seems interesting to test the usage of the H.264 Baseline profile for HD coding in real time services, and, more specifically, in videoconferencing. If Baseline is preferred to FREXT, there is the need to evaluate its impact on performance, as Baseline does not provide some enhanced coding features, like Bi-predicted frames (B frames) or CABAC (Context Adaptive Binary Arithmetic Coding), that are usually considered as crucial for a satisfactory video quality.

Based on such premises, the object of this paper is twofold: on one hand we show that the Baseline profile can be actually used for HD encoding, instead of the FREXT profile, without significant quality degradation but with savings in encoding time; on the other hand, we discuss two novel fast mode decision strategies, to be included in the Baseline profile, that permit further reduction in processing time, thus making real time HD coding more feasible, without significantly affecting the final video quality. The fast mode decision strategies

Manuscript received September 5, 2006 and revised June 7, 2007.

Authors are with the Dipartimento di Elettronica, Intelligenza Artificiale e Telecomunicazioni (DEIT), Università Politecnica delle Marche, via Brecce Bianche, 60131 Ancona, Italy, (email: {s.spinsante, e.gambi, f.chiaraluce}@univpm.it).

discussed in this paper rely on the proper definition of a threshold used to control the amount of SKIP MBs in each encoded frame. Such a threshold has a fundamental role, both in the CIF and HD coding, and will be subjected to a thorough analysis, aimed at optimizing its effects on the encoder performance.

The paper is organized as follows: Section II provides some results on the use of the Baseline profile for HD coding, to confirm the feasibility of this choice. Section III reviews the main issues related to fast mode decision techniques in H.264/AVC, that are introductory to the novel schemes presented in Section IV. The performance achieved is reported, with some examples, in Section V. Finally, Section VI concludes the paper.

II. Applicability of the Baseline profile for HD coding

A comparison between the Baseline (66,30) and the FREXT High (100,40) profiles has been developed, by applying the two profiles to a series of High Definition video sequences. The profiles are described in detail in the standard [1] and the related documents; in the following, we will limit to remind the features of interest to the present study. The test HD video sequences (Car, Park Run, and Shields) have been selected over a wide set, as representative of different motion and texture features. Single frames extracted from each sequence are shown in Fig. 1.

Comparison is based on the evaluation of the Peak Signal to Noise Ratio (PSNR) of the Luma (Y) component averaged over all the frames, the resulting bit rate, and the total encoding time. As the latter may depend on the test platform adopted, it has been expressed in relative units, r.u.. Examples are reported in Table I for the Car, Park Run, and Shields HD sequences, when the following encoder configuration is selected:

- Total number of frames: 100
- Frame rate: 30 frames per second (fps)
- Hadamard transform: disabled (Baseline), enabled (FREXT)
- Image format: 1280×720 progressive
- Max search range: 64
- Number of reference frames: 5
- P-slice reference: 1
- Quantization Parameter (QP): 24
- Entropy coding: CAVLC (Baseline), CABAC (FREXT)
- RD Optimization: enabled
- Number of B frames: 0 (Baseline), 49 (FREXT)

For each sequence, the PSNR values obtained with the two profiles differ by less than 1 dB, but the computational time required by FREXT is more than twice that required by Baseline, for Car and Park Run sequences, whereas it is almost twice for Shields. Consequently, it seems preferable to assume Baseline also for HD coding. In regard to the quality issue, it is known that objective evaluations (based on the PSNR) must be combined with subjective evaluations. We have made tests of this type, too, considering other sequences as well besides those discussed in the table, in order to evaluate different motion features, by reaching quite similar

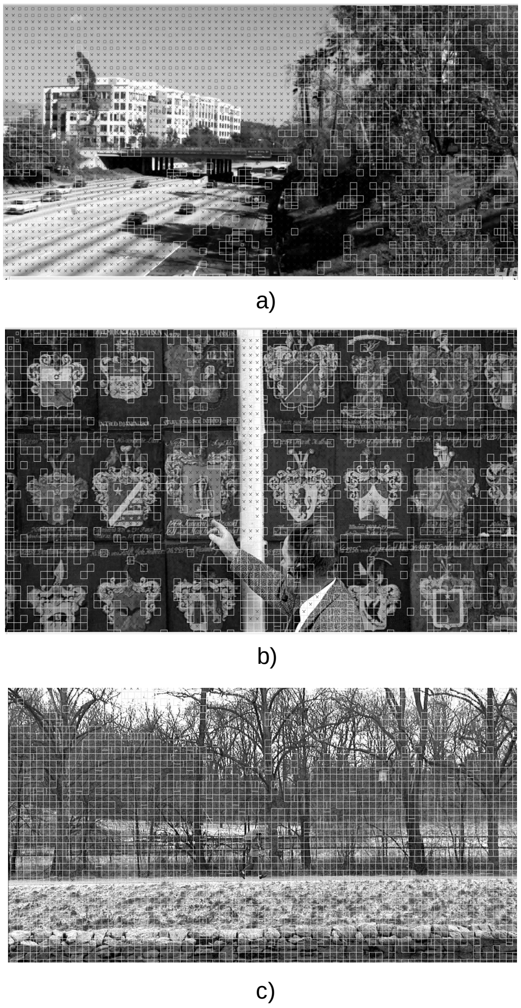


Fig. 1. Frames of the test HD sequences: a) Car, b) Shields, c) Park Run

TABLE I
QUALITY, BIT RATE, AND ENCODING TIME COMPARISON BETWEEN BASELINE AND FREXT PROFILES, FOR DIFFERENT HD VIDEO SEQUENCES (QP = 24)

Car		
	Baseline	FREXT
Avg PSNR Y (dB)	41.15	40.3
Bit rate (Mbit/s)	2.9	3.4
Encoding time (r.u.)	2.79	6.12

Park Run		
	Baseline	FREXT
Avg PSNR Y (dB)	65.77	64.9
Bit rate (Mbit/s)	18.1	18.4
Encoding time (r.u.)	2.41	5.23

Shields		
	Baseline	FREXT
Avg PSNR Y (dB)	69.25	68.4
Bit rate (Mbit/s)	3.5	3.6
Encoding time (r.u.)	2.91	5.12

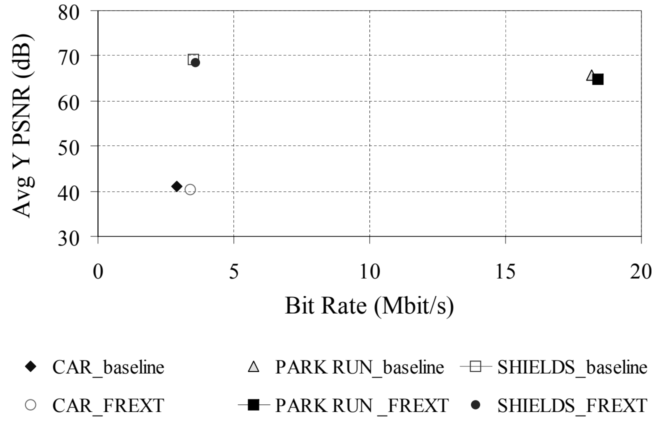


Fig. 2. Rate-distortion performances for the three HD sequences, QP = 24

conclusions throughout. For better evidence, Fig. 2 shows the rate-distortion results for each sequence, and for each of the encoder configurations tested. FREXT requires slightly higher bit rates, but does not provide as much higher quality than Baseline. As the performances of the two profiles are similar, the required encoding time can be assumed as the most relevant figure according to which Baseline can be preferred to FREXT, for HD real time coding.

The above comparison addresses the choice within the standard, but the total time required by the Baseline profile to encode an HD sequence remains, necessarily, more than two times larger than that required for encoding its corresponding CIF version, given the different formats. Thus, with the aim of applying HD coding in real time services, it should be advisable to design new fast mode decision strategies, able to take into account the peculiarities of the HD format (higher resolution and redundancy, up to four times more visual information) and the limitation of the human visual perceptivity. Our proposal is to stress some of the ideas exploited by fast mode decision algorithms, previously applied to the CIF and QCIF formats only, according with the procedures described next.

III. Fast Mode Decision Issues

One of the key aspects affecting the total encoding time required by an H.264/AVC encoder is the choice of the best coding mode for each MB of a frame, i.e. the mode decision process.

Under the variable block-size motion compensation, a MB can be divided into 16×16, 16×8, 8×16, and 8×8 subpartitions, and each 8×8 subblock can be divided into 8×8, 8×4, 4×8, and 4×4 block sizes. Moreover, H.264/AVC employs a new Intra coding method, called spatial-domain intra prediction, performed in units of 4×4 or 16×16. Consequently, the coding mode for each MB must be selected in a set of potential modes that includes, in the order tested by the encoder, Inter(16×16, 16×8, 8×16, 8×8, 8×4, 4×8, 4×4) and Intra(4×4, 8×8, 16×16) modes. Besides all these modes, the encoder has to evaluate the SKIP option too. The SKIP mode represents the case in which the block size is 16×16, but no motion and no residual information are coded. The encoder outputs just a SKIP indicator, without transmitting data as motion vectors, reference frame number, segmented block information, and so on. Except for SKIP and Intra modes, each Inter mode decision requires a motion estimation process.

In principle, for each MB of a P (inter) frame, all these modes should be tested and their performance compared, to find the best coding mode, i.e. the mode giving the minimum rate-distortion (RD) cost. This can be defined according with a SAD (Sum of Absolute Differences) or SSD (Sum of Squared Differences) criterion.

An exhaustive search can be a big burden for the encoder, quite unpractical to implement in case of HD video sequences. On the other hand, empirical evidence on the mode distribution, widely accepted in the literature, shows that, typically, more than half the MBs in real-world video sequences should be encoded with the SKIP mode [3].

Moreover, the SKIP mode has the lowest computational complexity, as discussed above. For these reasons, optimization of the mode decision procedure, also including the handling of the SKIP option, seems to be the right way for obtaining a significant reduction in the encoding time: by increasing the usage of the SKIP mode, that is assigned the highest priority in the mode selection phase, the coding time is reduced. In case of CIF and QCIF formats, however, this is generally unacceptable because of the impact on the quality; conversely, we have verified that the situation is different for HD sequences, where the larger amount of information available permits to compensate the risk of quality loss. These considerations, in turn, have inspired the new algorithms we propose. The latter are presented in the following section: first, preliminary studies performed on CIF sequences are discussed, as they motivate the strategies that are subsequently applied to HD sequences.

IV. The proposed algorithms

The fast mode decision algorithms proposed in this paper combine the strategy presented in [5], with an alternative handling of the SKIP option, as discussed in [6]. They are integrated in the Baseline profile of the Reference Software ver. JM9.2, and tested for use with HD coding. As mentioned, both the solutions found in the literature were originally conceived for CIF video sequences and applied to them only. Consequently, preliminary analyses have been performed on the CIF format, then the resulting modified algorithms have been applied to HD sequences, and evaluated for use in real time services.

In [5], the SKIP mode is assigned the highest priority, which means it is selected first. Its RD cost is compared against a threshold T_1 defined as the product between the minimum number of bits required for non SKIP inter modes and a parameter λ_{Mode} that has the meaning of Langrangian multiplier in the cost minimization: $T_1 = N_{bits} \cdot \lambda_{Mode}$.

Such a threshold, that equals zero for the first frame, is continuously updated as the encoding of inter frames proceeds.

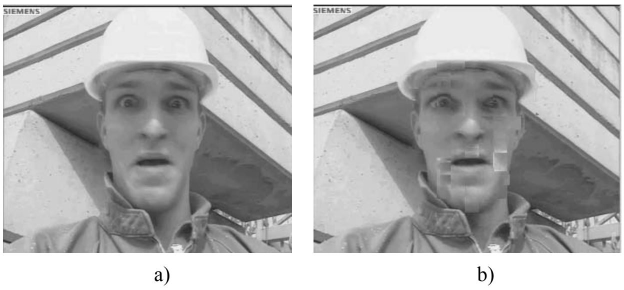


Fig. 3. Quality degradation due to a factor 100 applied to T_1, Foreman CIF: a) original encoder, b) modified encoder

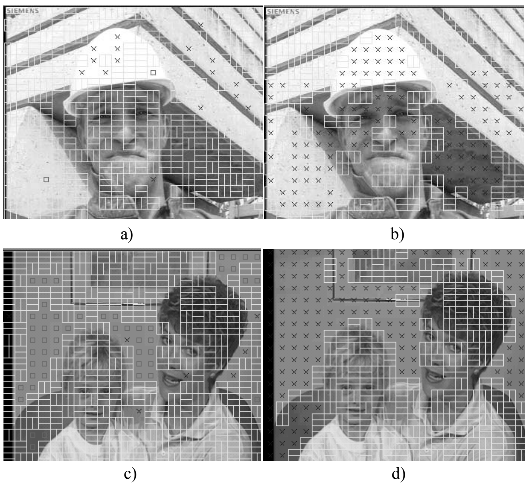


Fig. 4. Comparison between the coding strategies performed by [5] and [6] in the case of: Foreman (a - b) and Mother & Daughter (c - d) CIF sequences

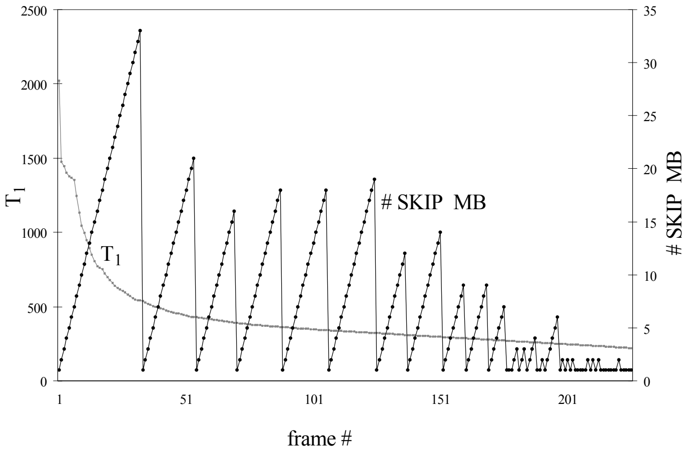


Fig. 5. Threshold T_1 and number of SKIP MBs, for the Foreman CIF sequence, algorithm [6] applied

Updating T_1 does not require any additional computation, as its factors are determined during the encoding process. If the cost of the SKIP mode is smaller than the threshold, SKIP is selected as the best mode, thus avoiding to test the other modes. The limit of this procedure, we have proved through simulation, is that a threshold so computed is often very small, and this prevents the possibility that the SKIP mode is chosen as many times as theoretically admissible on the basis of the sequence features. Heuristic solutions, consisting for example in multiplying T_1 by a fixed factor, must be accurately controlled: the number of skipped MBs increases, but often at the expense of a significant drop in quality; an example is shown in Fig. 3 for the Foreman CIF sequence, when a factor 100 is used.

A more effective solution, particularly for HD sequences, where the number of skipped MBs can have a strong effect on the reduction of coding time, can be found by resorting to a different threshold definition. In Ref. [6], dealing again with CIF sequences, a threshold having the same role of T_1 was set equal to the average RD cost of all the MBs that have been previously encoded with the SKIP mode. Following [6], at first this solution has been applied to standard test video sequences in CIF format; among them, Foreman and Mother & Daughter have been selected as representatives of opposite motion properties. In both the cases a remarkable reduction in the computational time has been obtained. It is justified by the increased number of skipped MBs, that is more evident in the case of Mother & Daughter video sequence, due to its low motion. More specifically, the skipping rate is 43% higher in the case of Foreman, and 95% in the case of Mother & Daughter, with respect to the previous scheme. An example is shown in Fig. 4, where a cross means a skipped MB, a square means an INTRA MB and the other borders show the INTER partitions selected during the coding process.

We will show in Section V that this fast decision strategy permits significant reductions of the coding time even when applied to HD videos with different features. In the case of sequences with limited motion, however, margins exist for further improvements. This is because even a threshold computed according to the strategy in [6] tends to decrease, thus making the choice of the SKIP mode less and less probable. This is clearly shown in Figs. 5 and 6, for the Foreman CIF and Park Run HD sequences respectively. The number of skipped MBs becomes smaller and smaller when the frame number increases, as the threshold T_1 is progressively reduced. It is easy to foresee that a sort of critical value exists (not shown in the figures) above which the SKIP mode is no longer chosen for encoding. Obviously, the critical value affecting the algorithm performance varies from sequence to sequence, depending on its format also.

An estimate for the critical value of the threshold T_1 that avoids further selection of the SKIP mode has been determined empirically, by means of a classic approach, which consists in averaging a large number of critical values that result from the analysis of several video sequences featuring different motion and texture properties. The critical values so found range from 600 to 900. More specifically, we have found that the number of skipped MBs in low motion sequences decreases,

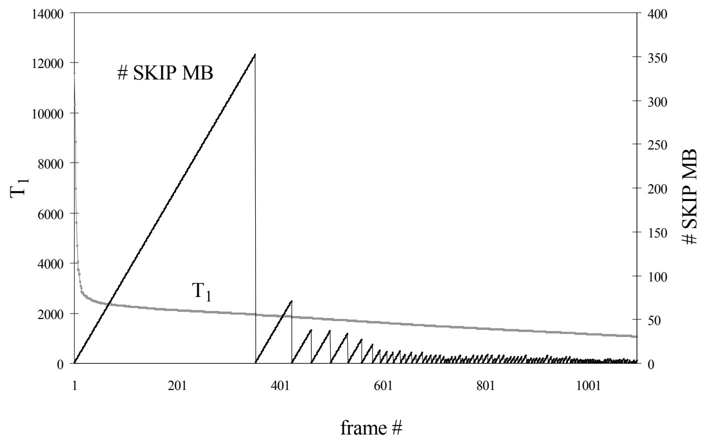


Fig. 6. Threshold T_1 and number of SKIP MBs, for the Park Run HD sequence, algorithm [6] applied

but remains adequate, even when T_1 gets low, whereas, in the case of panning or high motion sequences, the reduction of the number of skipped MBs becomes unacceptable.

Then, in our algorithm focusing on HD applications, the current value of T_1 is doubled any time it falls down an average critical value set to 800. This average value has been obtained by the statistical analysis of video sequences with different motion properties; should the algorithm be applied to a particular subset of them (e.g. sequences presenting a lot of scene change events, with a dominant amount of intra MBs per frame), the average critical value should be selected accordingly. The remainder of the procedure, however, remains the same. A smart method to determine adaptively the threshold value for an arbitrary sequence is another interesting issue, to be tackled in a future work.

Through these corrections, we have verified that the number of skipped MBs increases significantly, most of all in the case of sequences with relatively low motion, but with negligible quality degradation. In practice, the proposed algorithm modification, noted by SKIP2 in the following, is able to compensate some limits of the fast decision mode in [5], where the intra coding option is often selected also for MBs in the background, that instead could be skipped without serious impact on the HD quality.

When the SKIP mode is not selected as the best one, it is necessary to proceed with the computation of the costs of the other modes and their comparison. Just in [6], it was suggested to omit testing of Intra 4×4 and Intra 16×16 because of their very limited statistical incidence. Now we add the further experimental observation, derived over a wide set of different video sequences, that 8×8 and 4×4 partitions are used by the Reference Software in a very limited number of cases, and only for those parts of the image which are very rich in details. Therefore, they can be omitted in the selection, this way defining a further strategy, named SKIP3 afterwards, that in many cases permits an additional encoding time reduction. Once again, this strategy fits well to HD sequences, where the consequent loss of details is quite negligible, thanks to their higher resolution. Another novelty of SKIP3 is that the 16×16 coding option is now evaluated against the cost of the SKIP mode only. A flow-chart representation of the SKIP3 algorithm and its main functions is presented in Fig. 7.

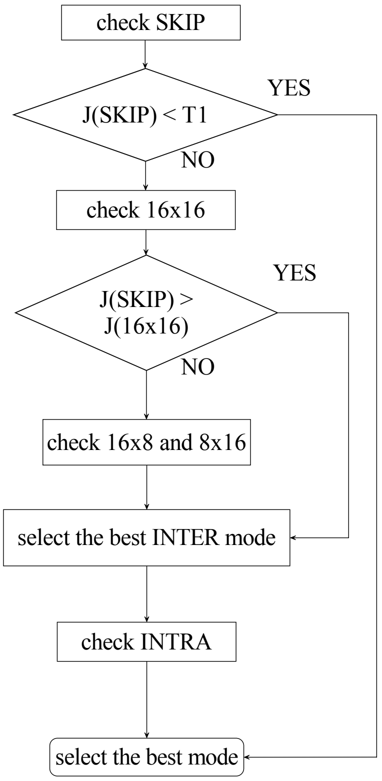


Fig. 7. Block diagram representation of the SKIP3 algorithm

V. Results

In order to compare the performance of the different algorithms described in the previous sections, several HD test sequences have been selected, namely Park Run, Mobile Calendar, Shields, Car, and Golf, as representatives of a wider set of sequences, because of their specific motion and texture features. More precisely, the Car and Golf sequences are good representatives of videos showing an almost static background, and some moving elements in the foreground; Mobile Calendar and Park Run represent high motion sequences featuring objects and camera movements; finally, Mobile Calendar and Shields are a reference video for sequences of high detail and complex texture.

Quality degradation and coding time reduction are the two main performance figures of interest, when dealing with fast mode decision algorithms. Quality can be evaluated both in objective and subjective terms. An objective quality measure can be defined as the average PSNR of the Y (Luma) component over a number of video frames. Fig. 8 shows the quality reduction due to the application of the SKIP2 algorithm to the encoding of the Car HD sequence, with respect to the Reference Software. The maximum penalty revealed does not exceed 0.1 dB, that, also looking at the Joint Video Team (JVT) documents, seems to be a good achievement. Similar

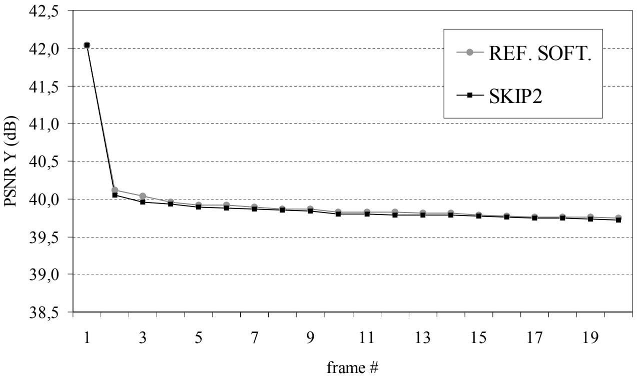


Fig. 8. Quality comparison for the Car video sequence, when using the Reference Software and SKIP2

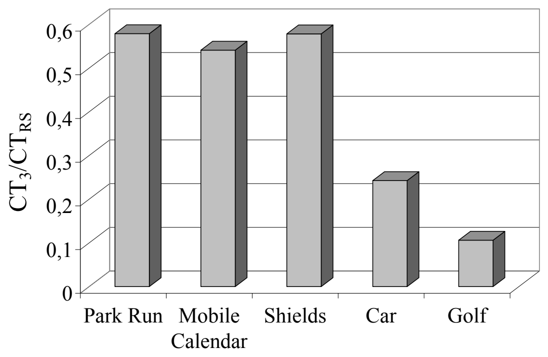


Fig. 9. Coding time required by encode_one_macroblock function of the SKIP3 algorithm with respect to the Reference Software

results are provided by algorithm SKIP3 too, and confirmed by subjective comparisons over a number of diffferent sequences.

The latter figure of merit in the performance evaluation is the coding time reduction allowed by the fast mode decision procedures. Fig. 9 shows the ratio between the coding time required by SKIP3 and that required by the Reference Software, for several HD sequences, measured on the execution of the core encode_one_macroblock function. In the case of Car and Golf sequences, the SKIP3 algorithm exhibits the greatest reduction, with a required time falling down to a 10% of the time needed by the Reference Software.

Finally, Fig. 10 reports the ratio between the total coding time required by SKIP3 and that required by SKIP2. From the figure we see that our last proposal is particularly effective when applied to the Mobile Calendar and Car sequences, for which SKIP3 permits to reduce by almost 20% the encoding time needed when SKIP2 is adopted. It is possible to argue that when sequences rich in details are to be encoded, SKIP3 provides a remarkable reduction of the coding time, being designed to avoid the evaluation of the smallest MB partitions.

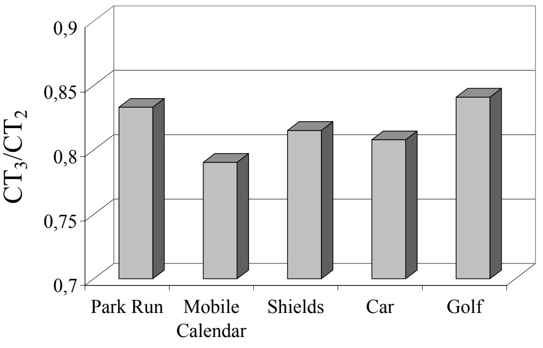


Fig. 10. Reduction of the total encoding time required by the SKIP3 algorithm with respect to SKIP2

In the case of Park Run, Shields, and Golf sequences, SKIP3 shows good performance: basically, it exhibits a good behavior on almost any kind of video sequence.

VI. Conclusions

Our study has demonstrated applicability of the H.264/AVC Baseline profile for the encoding of HD video sequences. In particular, just using some clever adaptations of the SKIP strategy, already presented in previous literature, a significant reduction of the encoding time, as required by real time applications, is possible, while preserving the image quality. We have proposed two new algorithms, named SKIP2 and SKIP3, that are particularly effective in the case of videos with dominant static background and a few moving elements, like the Car and Golf sequences, where they permit to have an almost halved encoding time, with respect to the Reference Software. The effectiveness of the proposed solutions strongly relies on the correct definition of the threshold used to drive the skipping selection; this motivated the in-depth analysis of alternative options, their implications on the encoding performance, and the impact of different updating rules, that have been presented in the previous sections. Another relevant aspect is that such a result can be achieved even with a reduction in the bit rate. About this point, however, it must be said that all tests described in this paper have been performed by maintaining a fixed QP equal to 24. As well known, the QP can be varied to adjust the final bandwidth required by the encoded signal: increasing QP gives a reduction of the required bandwidth but also a stronger degradation of the final quality. QP varies in the range [0, 51], so that QP equal to 24 is a relatively low value, which gives a wide required bandwidth. In some cases, this bandwidth can be too large for practical applications and the QP value should be increased. Actually, real implementations usually adopt a variable QP. To investigate performance of the fast mode decision strategies in such a more dynamical scenario could be a next development of the study.

ACKNOWLEDGMENT

The authors wish to thank D. Beccacece and C. Falasconi, of Aethra Telecomunicazioni, and L. Ciccarelli for their theoretical and practical contribution to the development of this project.

REFERENCES

[1] ITU, Ed., *Advanced video coding for generic audiovisual services*, ser. ITU-T Recommendation H.264. Telecom. Standardization Sector of ITU, 2005.

[2] C. Grecos and Y. Ming, "Fast inter mode prediction for P slices in the H.264 video coding standard," *IEEE Trans. Broadcast.*, vol. 51, no. 2, pp. 256–263, June 2005.

[3] R. Song-Hak and J. Ostermann, "Fast mode decision for H.264/AVC using mode and RD cost prediction," in *Proc. 1st International Conference on Communication and Electronics*, Las Vegas, NV, Oct. 2006, pp. 264–269.

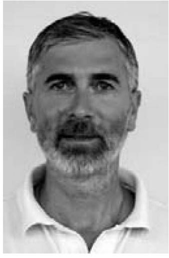
[4] I. Choi, J. Lee, and B. Jeon, "Fast coding mode selection with rate-distortion optimization for MPEG-4 part-10 AVC/H.264," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 12, pp. 1557–1561, Dec. 2006.

[5] P. Yin, H. Tourapis, A. Tourapis, and J. Boyce, "Fast mode decision and motion estimation for JVT/H.264," in *Proc. International Conference on Image Processing, ICIP 03*, vol. 3, Barcelona, Spain, Sept. 2003, pp. 853–856.

[6] C. Y. Chang, C. H. Pan, and H. Chen, "Fast mode decision for P-frames in H.264," in *Proc. Picture Coding Symposium, PCS 04*, San Francisco, CA, Dec. 2004.

Susanna Spinsante received her Laurea degree (summa cum laude) in Electronic Engineering in 2002 from the Università di Ancona, Italy and, in 2005, the PhD in Electronic Engineering and Telecommunications at the Università Politecnica delle Marche. Her main research interests are related to security and encryption aspects in communication networks, multimedia applications over IP, coding and audio/video applications. She is a member of IEEE.

Ennio Gambi was born in Ancona, Italy in 1961. He received his Laurea degree in Electronic Engineering from the Università di Ancona, Italy, in 1986. In 1992 he joined, as a researcher, the Dipartimento di Elettronica ed Automatica of the Università di Ancona, where he is currently the lecturer for the course of Telecommunication Systems. He is presently working on radio communication systems, multiple access techniques, video signal processing.



Franco Chiaraluce was born in Ancona, Italy, in 1960. He received the Laurea in Ingegneria Elettronica (summa cum laude) from the Università di Ancona in 1985. Since 1987 he joined the Dipartimento di Elettronica ed Automatica of the same university. At present, he is an Associate Professor at the Università Politecnica delle Marche. His main research interests involve various aspects of communication systems theory and design, with special emphasis on coding, cryptography and multiple access techniques. He is co-author of more than 150 scientific papers and two books. He is member of IEEE and IEICE.