Inter/intra-frame Compression of Video Sequences of Dynamic Signatures with Hierarchical Decomposition

R. Kountchev[1], M. Milanova[2], Vl. Todorov[3], R. Kountcheva[4]

Abstract: **In the paper is presented one new method for efficient compression of TV sequences representing dynamic signatures (each signature is a single sequence). The method is based on inter/intra-frame compression of the consecutive TV frames using Hierarchical decomposition. The compression starts with the processing of the reference frame in the sequence. The coarse approximation of the reference frame is first calculated and it is then used as a basis for the whole sequence. The remaining TV frames are coded sequentially one by one, using the approximation for the first frame and correcting it in correspondence with the currently processed frame. The information, which represents the details for every frame is added to the coded video sequence data.**

Key-Words: **Dynamic signatures archiving, intra-frame compression, Hierarchical decomposition**

I. Introduction

The security threats nowadays require new advanced security concepts to be elaborated and implemented as conceptual and feasible responses to these challenges. The solution of the problem requires more sophisticated and reliable means for information security, tracing of people and goods, and data access based on person's identification or authentication. The visual comparison of an individual's signature has been used for many years as a method of identification and is now popular for electronic comparison in which signature dynamics such as pressure, speed, and pen stroke direction are measured. Signature is not classified as characteristic, pertaining to the individual, because this data may change as the user ages and such systems require regular updates of the used database. One of the widely used solutions for the processing of dynamic signatures is based on the use of famous tools such as table-pads, which capture information about the signatures' dynamic characteristics [1].

One contemporary approach is the use of TV records of signatures, taken in the process of signing. This provides a variety of visual information which permits the extraction of characteristics, concerning the way the pen is hold, the way of signing (with stops or without stops, the way the pen is lifted, etc.), i.e. huge amount of information which else could not be sensed or represented mathematically.

The main problems, related with the security system implementation based on using biometric information, usually result from the large amounts of visual data, which have to be sent, received, stored and processed. Similar problems exist in mobile communications, providing services, which require exact and reliable personalization and authentication (bank transactions, etc.). The usual approach is to archive the related information after compression based on one of the well known video compression standards MPEG-1, MPEG-2, MPEG-4 etc. [2,3,4]. They all ensure high compression and good quality of the restored video. All these standards use intra-frame coding based on DCT (MPEG-1, MPEG-2) or wavelet transforms (MPEG-4), inter-frame prediction and inter-frame interpolation with motion compensation. The main disadvantages of the MPEG standards are the low quality of the fast moving objects (as a result from the inter-frame processing) and the relatively high computational complexity. The Motion JPEG 2000 (M-JPEG) standard [5] is aimed at the intra-frame compression of the consecutive TV frames and is based on wavelet transforms. It offers high quality, but the compression ratio is lower than that of the MPEG standards.

Specific feature of the signatures' TV records is that they are produced using a fixed camera. In this case the motion compensation is not necessary and the video processing could be significantly simplified. For such applications is expected that the inter/intra-frame compression will give relatively good results for lower computational complexity.

In this paper is offered one new approach for compression of video sequences, based on inter/intra-frame processing with 2D orthogonal transforms. The new point is to process the TV records of dynamic signatures with specially developed compression method, based on a Hierarchical decomposition, with Inverse Difference Pyramid (IDP). This decomposition permits the first (coarse) approximation for the lowest decomposition layer of one (reference) frame in a group of TV frames to be used for the processing of the remaining frames in the group and the high-quality approximations for each frame to be then calculated individually. As a result is obtained good quality for the fast moving objects in the restored video sequence.

The paper is arranged as follows: Section 2 introduces in brief the principles of the Hierarchical decomposition; Section 3 presents the new method for inter/intra-frame processing of video sequences; Section 4 gives some experimental results obtained and Section 5 (the Conclusion) focuses on the future development of the method.

II. Hierarchical Decomposition

The detailed description of the IDP had been presented in earlier publications of the authors [6,7]. The main operations comprising the Hierarchical decomposition (HD) method which is a simplified version of the IDP are presented in brief below.

[1]Roumen K. Kountchev is with the Faculty of Communications and Communications Technologies, Technical University, Kl. Ohridsky 8, 1000 Sofia, Bulgaria, E-mail: rkountch@tu-sofia.bg

[2]Mariofanna G. Milanova is with the Department of Computer Science, UALR, USA, E-mail: mgmilanova@ualr.edu

[3]Vladimir T. Todorov is with T&K Engineering, Mladost 3, POB 12, 1712 Sofia, Bulgaria, E-mail: todorov_vl@yahoo.com

[4]Roumiana At. Kountcheva is with T&K Engineering, Mladost 3, POB 12, 1712 Sofia, Bulgaria, E-mail: kountcheva_r@yahoo.com

The initial presumption is that the matrix [X] of the original image is divided into sub-images of size $2^n \times 2^n$ and each is processed with a two-dimensional (2D) orthogonal transform using only the low-frequency spectrum coefficients (i.e. the transform is "truncated"). The values of the transform coefficients are then processed with the corresponding inverse orthogonal transform. In result is obtained the coarse approximation of the processed image. This approximation is subtracted pixel by pixel from the original and in result is obtained the difference image with elements $e_p(i,k)$. The coarse approximation p is defined as:

$$e_p(i,k)=\begin{cases} x(i,k)-\tilde{x}(i,k) & \text{for} \quad p=0; \\ e_{p-1}(i,k)-\tilde{e}_{p-1}(i,k) & \text{for} \quad p=1,2,..,P, \end{cases} \quad (1)$$

where x(i,k) is the pixel (i,k) in the sub-image of size $2^n \times 2^n$ from the input image [X]; $\tilde{x}(i,k)$ and $\tilde{e}_{p-1}(i,k)$ are correspondingly the pixels of the restored and of the difference sub-images in the coarse approximation p. Every difference sub-image is divided into 4 sub-images of size $2^{n-1} \times 2^{n-1}$ and each is then processed with the 2D orthogonal transform again. Using the values of the calculated transform coefficients the image is restored and the second approximation and the corresponding difference image are calculated. The process continues in similar way until the approximation with the needed visual quality is obtained. The processing usually does not require all possible approximations to be calculated, because the needed image quality is obtained earlier, depending on the application.

The approximation models of the input or difference image (p) are represented by the relations:

$$\tilde{x}(i,k)/\tilde{e}_{p-1}(i,k)=IT[y_p(u,v)], \quad (2)$$
$$y_p(u,v)=T[x(i,k)/e_{p-1}(i,k)]$$

where T[•] is the operator for the "truncated" direct 2D transform applied on the input block of size $2^n \times 2^n$, or on the difference sub-image of size $2^{n-p} \times 2^{n-p}$ for the approximation p=1,2,..,P; IT[•] is the operator for the inverse 2D transform of the spectrum coefficients $y_p(u,v)$ from the approximation p of the "truncated" transform $2^{n-p} \times 2^{n-p}$, obtained in result of the transformation of every ¼ part of the difference sub-image, $e_{p-1}(i,k)$.

The set of transform coefficients, chosen for every approximation, can be different. The coefficients obtained in result of the 2D transform from all consecutive approximations are sorted in accordance with their frequency, scanned sequentially and losslessly compressed. The image restoration (decoding) is performed in reverse order. In the case when this decomposition is used for color images, each of the image components is processed individually. For example, for the YCrCb format 3 pyramids are built.

III. INTER/INTRA-FRAME CODING OF VIDEO SEQUENCES

The HD-based inter/intra-frame coding of video sequences is similar with that for single still image processing, presented above. The difference is that the original TV sequence is divided into groups of consecutive TV frames, processed together. The optimum length of the group is in the range of 12 – 15 frames, which corresponds to the statistical correlation [8] in the TV sequences. The middle or the first frame in the group is used as a reference one and for the whole group is built one common pyramid. For this, the first approximation is calculated for the reference frame only. The difference images for every TV frame in the group are calculated by subtracting the approximation obtained for the reference frame from the corresponding frame. The processing of the sequence continues in similar way for the next decomposition layer. For this, the two parts of the group are divided into smaller groups of 5 TV frames each. The middle frame in the sub-groups is used as a reference one and the next decomposition layer is prepared. The block diagram of the algorithm for 3-layer inter/intra-frame decomposition of 15 consecutive TV frames is given in Fig. 1.

IV. EXPERIMENTAL RESULTS

The experiments were performed with grayscale video sequences (color information in this case is not of high importance). The video camera was fixed. For the experiments were used groups of consecutive TV frames consisting of 15 and 25 frames. The middle frame was used as a reference one. The frames were extracted from DV format with Pinnacle 11. The size of each TV frame was 768 x 576 pixels, 8 bpp. Example single frames are shown in Fig. 2.a,b.

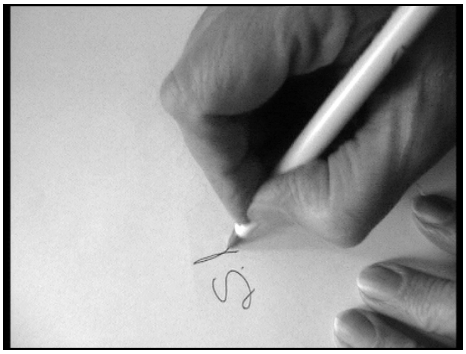

Fig. 2.a. The initial frame in a sequence of 15 TV frames

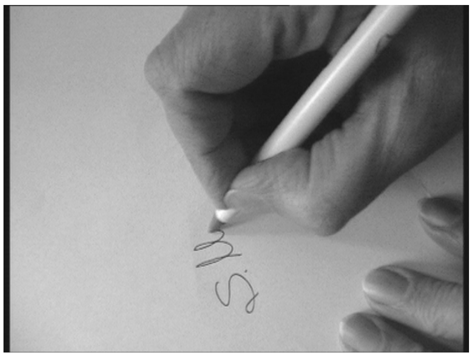

Fig. 2.b. The final frame in a sequence of 15 TV frames

For the experiments was used a 2-layer decomposition. The size of the initial (lower) sub-block was 8x8 pixels and that of the upper-layer sub-block, 4x4 pixels. The coefficients used for the creation of the decomposition approximations were 4 for the lower layer and 3 – for the next one. The results obtained for a sequence of 20 frames are given in Table 1 below. Here is given the information about the compression ratio (CR); the size of the coded approximating images of the consecutive TV frames for layer 2 (Size L2) and the quality of the restored images (PSNR). The size of the approximating file for Layer 1 of the reference TV frame is 12 518 B.

TABLE 1

RESULTS OBTAINED FOR A SEQUENCE OF 20 TV FRAMES

Frame №	Size L2 [B]	File size [B]	PSNR [dB]	CR
10 (ref.)	6 428	442368	38,19	18,35
1	9 659	884736	34,08	20,59
2	10 802	1327104	34,39	25,94
3	10 939	1769472	34,80	29,23
4	12 294	2211840	35,30	28,01
5	12 586	2654208	35,93	33,62
6	12 579	3096576	35,62	39,22
7	12 340	3538944	35,97	44,82
8	13 674	3981312	36,36	50,43
9	13 489	4423680	37,45	56,03
11	14 656	4866048	37,04	61,63
12	16 601	5308416	36,26	67,24
13	16 915	5750784	35,78	72,84
14	17 078	6193152	35,51	78,44
15	17 287	6635520	34,47	84,04
16	17 294	7077888	33,98	89,65
17	17 433	7520256	33,37	95,25
18	17 544	7962624	34,10	100,85
19	17 261	8404992	33,76	106,46
20	17 372	8847360	34,09	112,06
Mean PSNR = 35,3 dB; Coded file size: 225 691 [B]; CR = 112				

The number of participating coefficients in the second layer was defined by taking into account that only ¾ of the coefficients necessary for the presentation of the second-layer approximation are used, as a result from the existing relations between transform coefficients in the consecutive decomposition layers [7]. The column “File size” gives the number of Bytes for the sum of the processed TV frames for the corresponding row in the table. The column “Size L2” gives the size of the approximating file for the second decomposition layer obtained after lossless compression of the coefficients’ values. The PSNR (in dB) was calculated individually for each consecutive frame.

Similar investigations were performed for sequences of length 15 and 25 frames. The results obtained show that the compression ratio is lower for sequences longer than 15-18 frames. Naturally, this relation depends on the visual content to a high degree.

The place of the reference frame is of high importance as well. In the example it was chosen to be in the middle of the processed sequence. For the case, when the reference frame was placed in the beginning of the group the results for the compression ratio and the quality of restored images were a little worse. In the first case (reference frame in the middle) the correlation between the consecutive frames was used better and as a result, the size of the approximating images in Layer 2 was smaller. The basic disadvantage is that the process needs the first half of the frames to be saved in a buffer and processed when the reference frame is obtained.

The experimental results for the processed video sequences were compared with those for MJPEG and MPEG-2. The MJPEG is the closest tool, because the compression there is intra-frame. The experimental results show that for similar compression (277 KB for the proposed method and 367 KB for MJPEG) the mean PSNR is correspondingly 35,3dB and 34,2 dB. The comparison results with MPEG are as follows: in general, the bit-rate and the quality of the restored sequence of TV frames are comparable, i.e. for compressed file of size 250 KB, the mean PSNR value for the restored frames for MPEG was 35,9 dB, i.e. the PSNR is a little higher, but as it is known, this is not the best way to evaluate the quality of a moving object. In fact, the visual quality of the fast moving object (the hand) much better for the new method, while the background was reproduced a little better by the MPEG-2 standard. Additional advantage of the presented method it its’ lower computational complexity. Further quality enhancement of the restored images in accordance with the new method is possible applying post-processing of the restored TV frames. For this could be used the adaptive fuzzy filter for blocking artifacts reduction, developed by the authors [9]. This filter suits very well the special features of the Hierarchical decomposition and the parameters of the decoded image and ensures mean quality enhancement of 0,4 dB.

V. CONCLUSION

The experiments performed with the new method for intra-frame compression of video sequences, presented in this paper, proved its efficiency. The HD offers some advantages, which still have to be investigated in detail. For example, the relations, existing between some of the 2D orthogonal transform functions permit reduction of the used coefficients number in the decomposition layers [7]. The felicitous use of transform coefficients can ensure quality enhancement with negligible augmentation of the coded data. Instead of evaluating the values of the calculated coefficients, it is possible to prepare some models for certain textures, etc., which will decrease the computational complexity of the method even more.

The influence of the decomposition layers number on the compression ratio and the quality of the restored sequence is closely related to the image content and movements speed. Additional investigations will be aimed at these problems.

The HD could be implemented using any kind of orthogonal transform (DCT, Haar, wavelets, etc.), but the Walsh-Hadamard transform has lower computational complexity and together with this ensures the needed quality of the restored images.

The new method is suitable for applications with fixed camera where the motion compensation is not necessary. The new point in this work is that the dynamic signature is saved in the process of signing and after that – coded with inter/intra-frame compression. This method permits together with the traditional similarity comparison to analyze and match a great number of additional parameters and characteristics, which else could not be detected or described. The traditional signature comparison could be done using the same TV sequence of the dynamic signature, because in the last frames the signature is already ready and could be successfully extracted. Example signature extracted from the TV sequence is shown in Fig. 3.a,b. The high compression ratio, achieved for such kind of video information with the new method for inter/intra-frame coding, permits all the information to be stored as a relatively small file.

The presented method for efficient coding could be used for comparison of video signatures by using content-based information retrieval methods (the compressed document saved in the database is compared with the new signature taken online). Additional possible application areas are the medical imaging, videoconferencing, surveillance, etc.

a) A signature in the TV frame b) the extracted signature

Fig. 3. Part of the TV frame with the finished signature

ACKNOWLEDGEMENT

This paper was supported by the National Fund for Scientific Research of the Bulgarian Ministry of Education and Science (Contract VU-I 305/2007).

REFERENCES

[1] D. Dessimoz, J. Richiardi, C. Champod, A. Drygajlo. Multimodal biometrics for identity documents: state of the art. Research report, Universite de Lausanne, 2005, pp. 87-94.

[2] B. Haskell, A. Puri and A. Netravali. Digital video: Introduction to MPEG-2. Chapman Hall, NY, 1997

[3] I. Richardson. H.264 and MPEG-4 video compression. J. Wiley & Sons, NY, 2003.

[4] L. Hanzo, P. Cherriman and J. Streit. Video compression and communications: H.261, H.263, H.264, MPEG-4 and HSDPA-style adaptive turbo-transceivers. J. Wiley and IEEE press, 2007.

[5] T. Acharya and P. Tsai. JPEG 2000 standard for image compression: concepts, algorithms and VLSI architectures. J. Wiley & Sons, INC Publication, NJ, 2005.

[6] R. Kountchev, S. Rubin, M. Milanova, Vl. Todorov, R. Kountcheva. Cognitive Image Representation Based on Spectrum Pyramid Decomposition. Proceedings of the WSEAS Intern. Conf. on Mathematical Methods and Computational Techniques in Electrical Engineering (MMACTEE), 2008, pp. 230-235.

[7] R. Kountchev, R. Kountcheva. Image Representation with Reduced Spectrum Pyramid. Book chapter in: New Directions in Intelligent Interactive Multimedia, Eds. G. Tsihrintzis, M. Virvou, R. Howlett, L. Jain, Springer-Verlag, Berlin, Heidelberg, 2008, pp. 275-284.

[8] W. Pratt (Ed.). Image Transmission Techniques. Academic Press, NY, 1979.

[9] R. Kountchev, Vl. Todorov, R. Kountcheva, M. Milanova. Adaptive Fuzzy Filter for Reduction of Blocking Artifacts in Images Compressed with IDP Decomposition. 10th WSEAS Int. Conf. on Computers (CSCC'06), Athens, Greece, July 13-15, 2006, pp. 502-507.

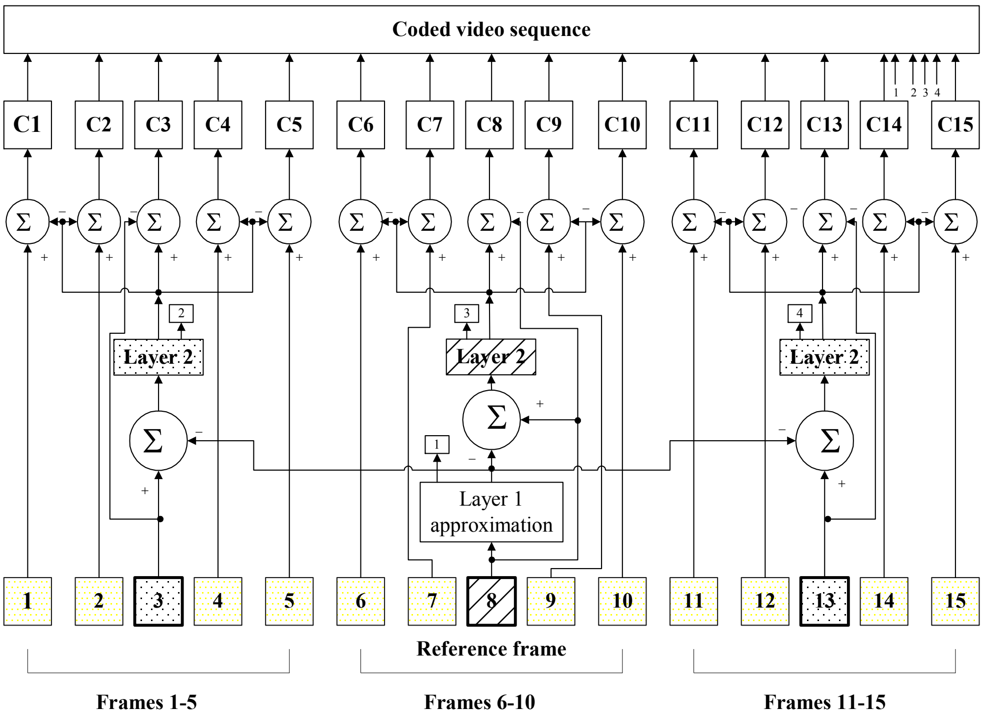

Fig. 1. Block diagram of the inter/intra-frame coding of TV sequences

Abbreviations: 1-15 – original TV frames; C1 – C15 – the coded data for the second-layer approximations of the corresponding TV frames; Σ - block in which is calculated the difference image; Frame 8 – used for the low- layer approximation; Frames 3, 8 and 13 – used for the second layer approximation.