3^{rd}Generation Approach to Video Compression for Multimedia

Pavel Hanzlík, Petr Páta

Dept. of Radioelectronics, Czech Technical University in Prague, Technická 2, 166 27, Praha 6, Czech Republic

Hanzlip@feld.cvut.cz, Pata@feld.cvut.cz

Abstract. *This paper introduces a video compression method based on methods and algorithms of the digital third generation. Standards designed for video coding (MPEG-4 and MPEG-7), being not strictly defined, present a large field for further experiments and improvements. Our method was developed on fundamentals of Vector Quantization (VQ) and the most advantageous qualities of Karhunen-Loève Transform (KLT) (known also as Principal Component Analysis – PCA). We expect the new method to be used in high-speed video transfer channels in present multimedia systems. Among other compression techniques, object coding, quad-tree decomposition and optimalization of existing algorithms is discussed as a purpose for further improvement.*

1. Introduction

The main objective of this paper is to present a new alternative approach to video compression while there are many standardized methods used worldwide in various technical applications. In time of MPEG-4, MPEG-7 and on-coming MPEG-21 we can ask where do these standards allow us to improve more in the meaning of video compression. Whether it is better either to build a brand-new video coder based on the latest break-throughs in computer science or to concentrate on mathematic and physiological background of the original compression standards and techniques, being currently already in place.

2. Vector quantization and Karhunen–Loève transform

Vector quantization (VQ), which Gray [1] or Abut [2] presented in the early 80's, is one of the powerful compression methods, which became a fundamental of present MPEG standards. The idea of an optimized codebook, adaptive codebook forming or training set iterative improvement is not new; it has been discussed and adjusted to many different variations according to the specialized applications in audio, video, robotics or neural networks already.

With advance, we used the profit of VQ qualities to build a video coder based only on VQ [3], what made us to concentrate on the best/appropriate design of a video coder [4]. The experiments confirmed that although there were excellent results in vector coding of the image base characteristics (luminance, color), it would also be an interesting challenge, to vector quantize spectrum of any of the well-known integral transforms e.g. *Discrete Cosine Transform* (DCT), *Fast Fourier Transform* (FFT), *Walsh-Hadamard Transform* (WHT), *Wavelet Transform* (WT), Slant or *Karhunen-Loève Transform* (KLT). Vector quantization allows us to eliminate the need of re-forming the base functions according to each transform usage in the process simply by vector quantizing the necessary base vector function into a complex set.

From the above-mentioned integral transforms, we chose KLT for our application. We compared the characteristics of various transforms with special attention to optimality in image reconstruction and compression. Although there are DCT or WT often used in image compression nowadays, we chose KLT (mainly for it‘s mathematical and statistical qualities), also often used in computer science terminology as *Principal Component Analysis* (PCA).

Accordingly with Jollife [5] we can prove that KLT is an optimal transform, in manner of the steepest reduction in the least possible number of spectral coefficients, to signals with

Gaussian distribution probability density. While an image is meant to be Gaussian distributed signal, we can only transfer 80-95% of the KLT spectrum without this having any impact on the quality of the reconstructed image [6]. With help of little mathematics we also can show that the reduction in spectrum results in redundancy in base vectors (KLT base coefficients), which are (in fact) elements of the image covariance matrix. The percentage is dependent to image or scene character and of course, on the elementary transform block size. The more the coefficients in KLT base vector are decorrelated, the more KLT spectral coefficients can be eliminated being so-called *near zero coefficients*.

2.1. Elementary Karhunen-Loève Theory

Given a general (n x n) covariance matrix:

$$K_Y = \begin{bmatrix} \mathrm{cov}(1,1) & \mathrm{cov}(1,2 & \dots & \dots & \mathrm{cov}(1,n) \\ \mathrm{cov}(2,1) & \mathrm{cov}(2,2) & \dots & \dots & \mathrm{cov}(2,n) \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \mathrm{cov}(n,1) & \mathrm{cov}(n,1) & \dots & \dots & \mathrm{cov}(n,n) \end{bmatrix},$$

a transformation matrix A is to be constructed such that the n uncorrelated random variables X_j are to be transformed into Y_j random variables which possess the properties of having zero means and the given covariance matrix K. The only constraints on the X_j are: mutual independence, zero mean, and unit variance.

With these constraints, it can be shown that $K = AA^T$ and any A matrix with this property is such a transformation matrix. The symmetry of K implies K can be decomposed into the product of matrices of *eigenvectors* and *eigenvalues*, namely the orthogonal decomposition $K = P\Lambda P^T$. Letting X be the *(n x 1)* matrix consisting of the X_j and Y be the *(n x 1)* matrix consisting of the Y_j then $AX = Y$ and $P^T Y = \overline{X}$, where the components of $\overline{X}$ are $\overline{X}_i$, uncorrelated random variables with variance, $\sigma^2_{Y_i}$.

To summarize, K-L relation in matrix form can be defined as follows:

$$\mathbf{C}_x \boldsymbol{\tau}^* = \boldsymbol{\tau}^* \Lambda; \qquad \text{where} \quad \Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & 0 & \\ & 0 & \ddots & \\ & & & \lambda_M \end{bmatrix} \longleftarrow \text{Diagonal matrix}$$

These are the fundamental relationships of elementary K-L theory, for a finite number of random variables (a discrete process), which gives a procedure for generating correlated random variables. We studied the problematics of optimality and sub-optimality when infracting some of the mathematical prerequisitions in KLT definition and this branch still remains open for future studies. Nevertheless, several tests proved that KLT is optimal only in the meaning of zero mean and when the distribution of the covariance coefficients matrix of each KLT elementary sub block has normal distribution (See Fig.1).

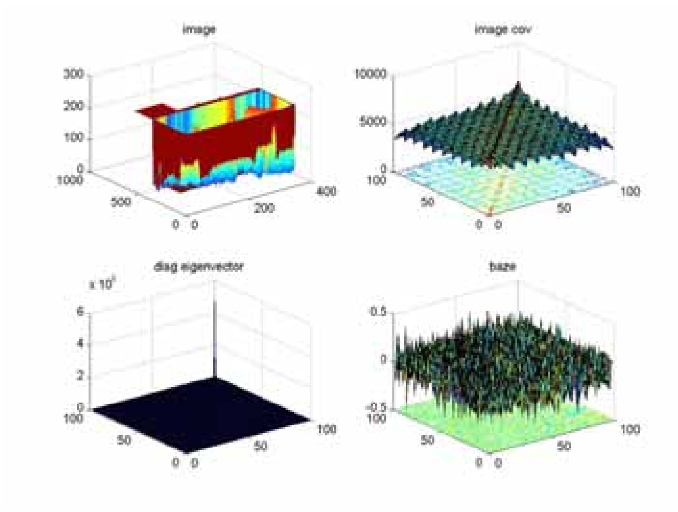

Fig.1. Example of calculation of KLT eigenvectors, coefficients and KLT base

2.2. Karhunen-Loève Transform characteristics

The first tests with black-and-white static images showed that the redundancy thanks to KLT could rise up to 80-98% without visible image quality loss (See Fig. 2). This percentage is highly dependent to image or scene character and of course on the elementary block size of the transform.

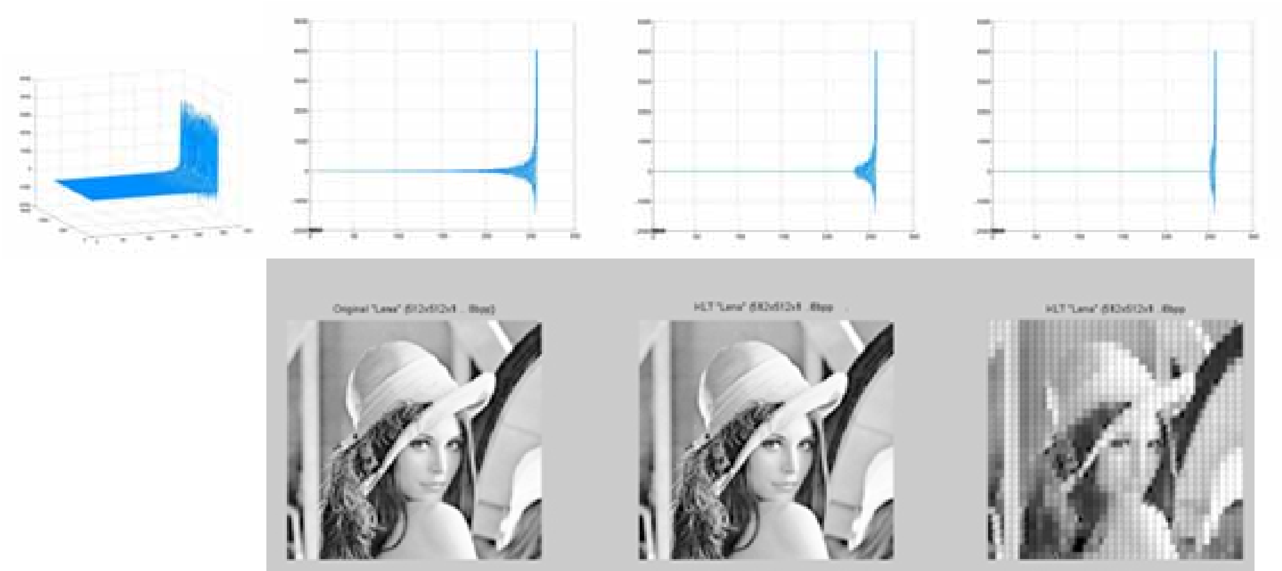

Fig. 2. Image „Lena“ coded with KLT block size 16x16, a/ KLT spectrum matrix, b/ Spectrum matrix – side view, c/ 80% of near zero coefficients zeroed, d/ 99,96% of near zero coefficients zeroed with relation to reconstructed image

Another approach can be in calculating so-called *decorrelation efficiency* of KLT bases. The more coefficients are decorrelated, the more spectral components can be eliminated without image quality loss. We can see that compared to DCT or Wavelet Transform (WT) the elementary base stones are highly optimized to the current image/scene. (See Fig. 3).

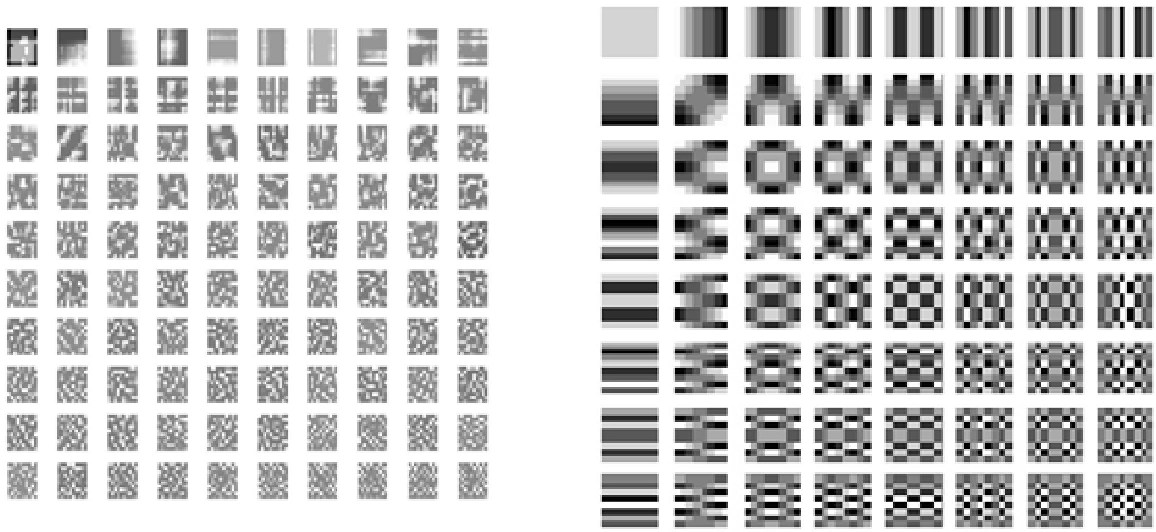

Fig. 3. Base elementary stone vector formed from the KLT base or DCT base

2.3. Elementary Vector Quantization Theory

While we know the statistic dissipation of any data vectors in the Euclidian space (*m*-dimensional vector) and the starting number of vectors, we can form a Linde-Buzo-Gray iterative algorithm for VQ by repeatedly reassigning the reference vector set according to the center of particular *Voronoi cell*. The VQ is a relation *Q* from *m*-dimensional Euclidian space A^{rk} into finite set Y from Ark. We have:

$$Q: R^k \to Y,$$

where

$$Y = (\hat{x}_i ; i = 1,2,.., N)$$

is the set of predefined vectors and N is the number of the vectors in the set.

We define a distortion measure:

$$D(x, q(x)) = \frac{1}{N} \sum_{i=0}^{N-1} (x_i - \hat{x}_i)$$

that represents the error of representing vector *x* by it's model *x*‘. Changing the vectors in the so-called training set can minimize iteratively this error. The *centroid* (part of the training set) is generated from *n* Voronoi cells, corresponding to vectors dissipation. These two steps converge after several iterations to the minimized (square) error (MSE criteria) (See Fig. 4).

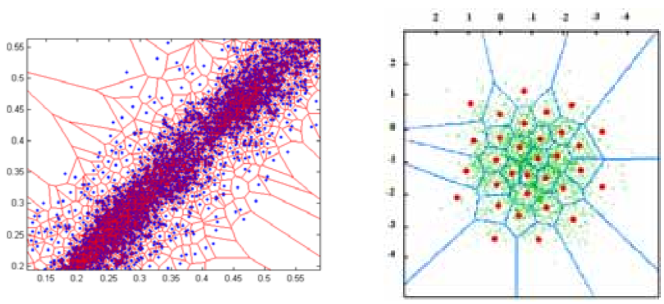

Fig.4. Voronoi diagram with assigned centroids and Voronoi diagram with finished LBG algorithm iterations

Every iteration though has two parts – assigning the codebook and adjusting it according to the image character. The adjusting is done iteratively by reassigning the particular training set with a new one. When achieving better results in MSE, the new one becomes reference. This process converges and minimizes the quantization error.

3. Experiment

Now the next logic step is to integrate KLT and VQ into one functional system. Based on the experience from the precedent experiments this idea is very effective. This is the stage where we could impact on a problem with time-consuming feedback used in Generalized Lloyd or Linde-Buzo-Gray iterative algorithms for VQ [1]. Instead we could use the uniform codebook reduction algorithm for VQ [4] which unifies the advantageous characteristics of (slow) iterative algorithms generating optimized codebooks and (fast) algorithms generating universal sub optimal codebooks characterized by small mean square error distortion. We have performed an experiment with a short color video-sequence which was divided into frames and vectors, formed from the three R, G, B planes in frame area sub block of size 30 x 12 x 3 pixels. Each vector has been then 1080 pixels long (dimension is 1080).

Karhunen-Loève Transform was applied to each frame and we achieved to determine an orthonormal highly decorrelated KLT base functions matrix and KLT spectrum (See Fig. 5).

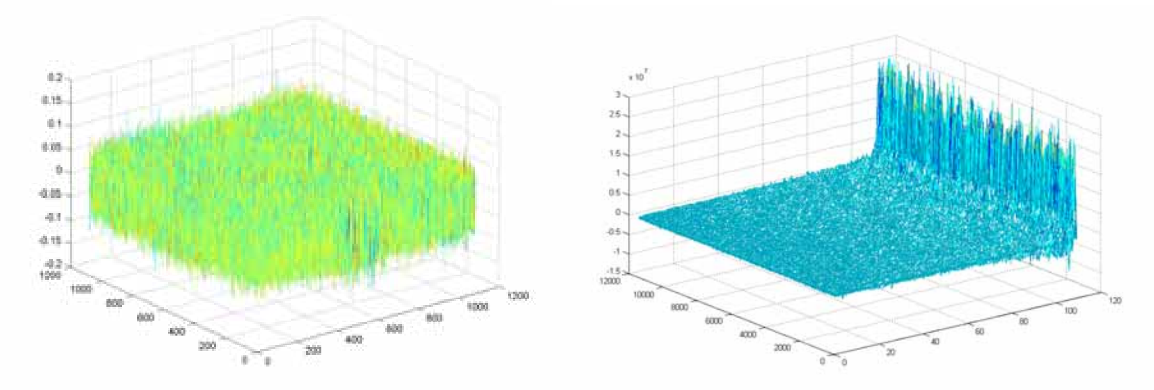

Fig. 5. Decorrelated KLT base function matrix and KLT spectrum matrix (10 frames)

3.1. KLT Spectrum

It is obvious that approximately 80% of all spectral coefficients are near zero and then there is no need to use them in the communication channel. When minimizing the redundant coefficients, new vectors are formed, with smaller dimension (in our case) 110 and represent the input to VQ algorithm.

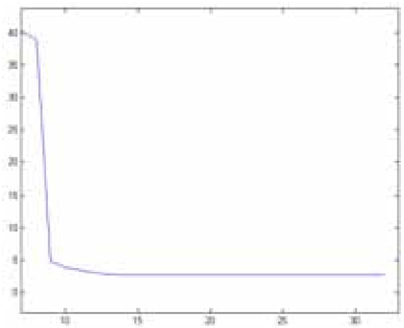

Fig. 6. Study of impact of bit resolution reduction on image quality (MSE)

The spectral coefficients need to be scalar quantize first, while the original resolution of the full dynamic range would be 28 bits and though too high. We suppressed the resolution to 8 bits with still not recognizable impact on image quality. The study of bit resolution impact is presented in Fig 6.

3.2. KLT base functions

With help of little mathematics it can be proved that if some of the spectrum coefficients can be eliminated, also some of the KLT base vectors are surplus. In Fig. 7. we demonstrate that only the most important vectors are preserved and that we can still demonstrate the multi-dimensionality in graphical representation by color (three planes) interpretation

Fig. 7aa – Image "Lena" – all calculated KLT base function stones, 7ab – only the most important functions are included in the compressed signal

Fig. 7ba – Video sequence "Harbor" – all calculated KLT base function stones – multi-dimensionality demonstration (color) 7bb – only the most important functions are included in the compressed signal

3.3 Vector Quantization purpose

The VQ is applied on remnants of KLT spectral coefficients formed to vectors. When we sort all the spectral vectors (See Fig 8) in the video sequence from the most to the least important spectral vector, we can see that some of the spectral vectors repeat often, or form a cluster in the Euclidian space (minimal MSE distortion). This could be predicted, because the images, as well as their spectral representation, will correlate a lot in the video sequence. With profit, we can use these findings to form an adaptive codebook vector quantization algorithm that will only include the most relevant parts of the signal. The MSE calculation will determine the cluster position differences in the Euclidian space, without need of construction (in VQ often used) Voronoi cell diagram.

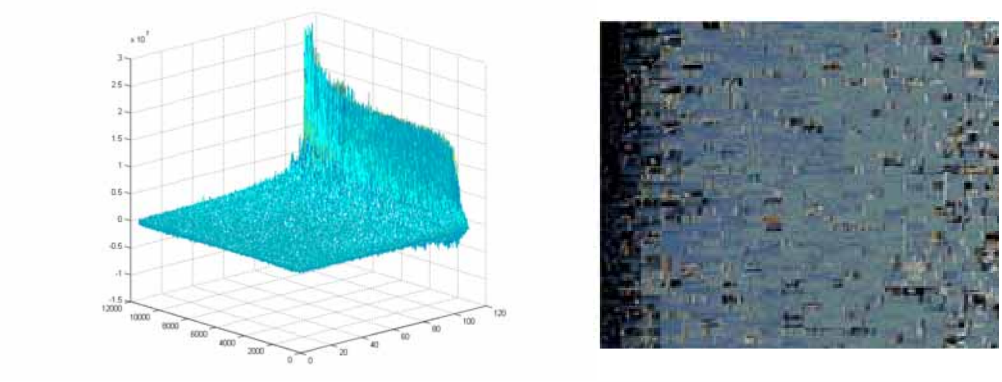

Fig. 8. All partial KLT spectrum vectors sorted and their reconstruction into image space (similarity)

Then, using the algorithms of *nearest neighbor search* in Euclidian space, only predetermined quantity of spectral vectors is chosen for the initial codebook to allocate the VQ codebook size for future lossless *Huffmann coding* or *Run-Length Coding* (RLC). Without having an iterative optimizing algorithm, it is not easy, to find out the best sub-optimal codebook. We used previously developed algorithm of uniform reduction [3] and similarly to the previous experiments, with a robust reduction approach, we obtained effective results (See Fig. 9).

Fig. 9. One of the video sequence frames – original image and reconstructed image.

4. Discussion

Especially the stage of defining the codebook and assigning the vectors to the compressed KLT spectrum is critical to the video quality. With scalable codebook size we can realize different modes of quality in broadcasting in Video-On-Demand, for example. A solution for keeping the quality on acceptable level is not easy, while an iterative optimalization algorithm would need much more time and could not be realized online by coding process with buffer then. We can improve the algorithms of sub-optimal codebook generation either by emphasizing the hierarchy importance in spectrum coefficients and eliminating similar spectral vectors, or – and what is more challenging – by decomposing the related frame in more effective way. There opens a possibility to use quad-tree decomposition or simply define objects or video-objects (MPEG-4, MPEG-7) and compress them with PCA approach. There is a novel concept of coding available in computer science where complete video-objects are defined and the compression remains in “what is not visible is not transferred” and synthetisation from the probability distribution functions approach.

5. Conclusion

The experiment came into stage where a novel approach is required, not to waste all the advantageous features of preceding transforms. The concept is mathematically one of the most effective in theory, with using one of the most optimal reductions of image information. The technical possibilities are limited and always bring constraints and distortion. This is what we are trying to solve too. The new compression method is planned to reach with ease compression ratio 1:120 – 1:150, depending on the demanded video quality and buffer block size. We do not expect this method to compete to DivX, DV or MPEG standards, but to take part in science applications like trans-satellite communication, down-links for satellites and probes with video information or interfered video communication.

6. Acknowledgement

This research and publication has been created on dept. of Radioelectronics, ČVUT – FEL in Prague (CTU) under the auspices of research project **"Quality aspects of image compression methods in multimedia systems"** and was supported by grant No. 102/02/0133 of the Czech Republic Grant Agency

The parts that deal with design of compression methods for high-speed video transfer were supported by grant within research intention of the pgs. project GAČR No. 102/03/H109 **„Methods, structures and components of electronic wireless communication "**.

7. References

[1] Gersho,A., Gray,R.M. – **VQ and signal compression** C1992, Kluwer Academic Press
[2] Abut, H. – **Vector Quantization** C1990.
[3] Hanzlík,P., Páta, P.: **Behind the Structure of Video VQ-Coder**, *Radioengineering September 2004, Volume 13, Number 3.*
[4] Hanzlík, P. - Páta, P.: **Video Coder based on Vector Quantization**. *Proceedings of Workshop 2004*. Prague: CTU, 2004, vol. 2, s. 288-289. ISBN 80-01-02945-X.
[5] Jolliffe I.T.: **Principal Component Analysis**. New York: Springer, 1986.
[6] Hanzlík, P. Páta, P.: Video **VQ-kodér – implementace v Matlabu** – *Matlab 2003 - proceedings of conference, vol 1*, November 2003, p. 173-179.
[7] **Encyclopedia of computing science and technology 29** (supplement 14).
[8] Klíma M., Bernas M., Hozman J.,Dvořák P. - **Zpracování obrazové informace** C1996.
[9] ITU-T, **"Methods for the Subjective Assessment of the Quality of Television Pictures"**, ITU-T Recommendation 500-2, 1982.
[10] ITU-T **"Subjective Video Quality Assessment Methods for Multimedia *Applications*"**, ITU-T Recommendation P 910, August 1996.
[11] ITU-T, "**Video coding for low bitrate communication"**, ITU-T Recommendation

8. About author…

Ing. Pavel Hanzlík was born in 1979 (Děčín, Czech Rep.), in 2003 he graduated on CTU, Prague. He continues in the studies on the PhD. program and his main interests are image and video compression, theory and comparison of used methods, and practical implementation of the new ones developed in his research work.
Contact: Dept. Of Radioelectronics K13137, CTU – FEE Prague, Technická 2, Prague 6, 16628.