MOTION ESTIMATION VIA BLOCK MATCHING ALGORITHMS

Peter RADOCZI
Department of Electronics and Multimedia Communication, Faculty of Electrical Engineering and Informatics,
Technical University of Košice, Letná 9, 042 00 Košice, tel. 095/602 2808, E-mail: Peter.Radoczi@tuke.sk

SUMMARY

Many motion estimation techniques have been proposed. For image sequence coding, two most commonly used approaches have been classified as the pel recursive algorithms and the block matching algorithms. The former estimates the motion on a pixel by pixel basis, whereas the latter predicts the motion on a block by block approach. Furthermore, the block matching approach is most appropriate for its applications to current standards such as MPEG or H.26X, which are based on discrete cosine transform coding (DCT).

This paper presents at first performance of full search algorithm in sense of difficulty implementation following criteria of distortion: normalised cross-correlation function, method of correlation coefficients, mean square error, mean absolute difference, minimized maximum error and pixel difference classification. Evaluation of full search algorithm has been evaluated for three different pair of frames in the raw-format, according to by peak-signal-to-noise ratio (PSNR) formula.

Block motion estimation using the full search is computationally exacting. Several efficient techniques have been recently proposed to reduce the computational complexity of block matching for motion estimation in video sequence coding. The goal is efficient motion estimation with minimal error in the motion compensated predicted image. This paper presents in the second part results of efficiency block matching motion estimation algorthms in sense to achieve the best PSNR values. The evaluation has been done for 21 different block matching algorithms. The main idea has been to provide a complex review of block motion estimation algorithms and its efficiency in image sequence coding.

Keywords: *motion estimation, interframe redundancy, cost function, block matching algorithm.*

1. INDRODUCTION

In image sequence coding, the statistical redundancies can be categorized into spatial and temporal domains. Transform coding is one of the most general ways to reduce the spatial redundancy, which is called intraframe coding, whereas the reduction of temporal redundancy is referred to as interframe techniques. For the purpose of reducing temporal redundancies, motion estimation techniques have been succesfully applied, which estimate the displacement of objects between successive frames. In image sequences coding the interframe coding methods with the prediction of the objects motion are used. For simplification of implementation, motion compensation with motion estimation is generally performed by stepwise translation of objects in image. In first stage of coding, the displacement of object is estimated by using of motion estimation methods. The main methods of motion estimation are pel recursive techniques and block matching techniques. The first one estimated the motion vector ob pel-by-pel basis, whereas the second one estimated the motion vector on block-by-block basis. The result of this step is identification of block in the current frame with the most similarity in the previous frame. The offset between both blocks is the displacement vector for motion compensated prediction. Consequently, the prediction error and motion vectors for each of block has to be transmitted. For prediction error coding the DCT transform is used in MPEG or H.26X standards. Transformation coefficient of DCT are quantizated and coded by VLC (Variable Lenght Code) method (Fig. 1).

2. MPEG STANDARDS

The first set of MPEG standards, commonly referred to as the MPEG-1 standards [1][2], was adopted in October 1992. MPEG-1 has the official name ISO/IEC-11172 and consists of four main parts: system, video, audio and the fourth part contains conformance tests stored on a CD ROM. The MPEG-1 standards are targeted for data rates of up to about 1.5 Mbit/s, which is roughly the data rate produced by a single-speed CD ROM player. The core parts of the second set of MPEG standards, commonly referred to as the MPEG-2 standards, were adopted in November 1994. A third MPEG standard, MPEG-3, was originally planned for handling bitrates between 20 and 40 Mbit/s. When the MPEG-2 standard was found to handle these rates sufficiently well, MPEG-3 was abandoned. A fourth MPEG standard, MPEG-4, is currently under development for encoding multimedia data at low bit rates (between 4,800 and 64,000 bit/s). This future standard is anticipated to be useful for such applications as mobile multimedia, video phones, video electronic mail, and sign language captioning.

In addition to the MPEG standards, there are other important standards currently being used to transfer multimedia data. One of these is the ITU (International Telecommunications Union, formerly CCITT) H.261 video codec for Audiovisual Services at multiples of 64 kbit/s data rates [6],[10],[12]. The H.261 standard was completed and approved in

December 1990, and is used mainly for video phone and teleconferencing applications.

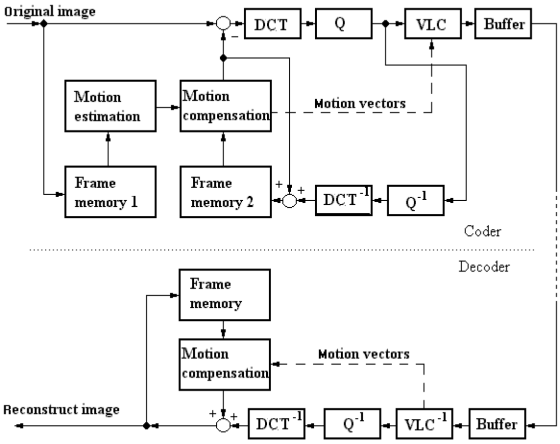

Fig. 1 MPEG coder-decoder

3. MOTION ESTIMATION

Block matching algorithm (BMA) estimates the motion vector in a block-by-block basis. In BMA, a current frame is divided into blocks of size $(M \times N)$ pixels. The block of pixels in the current frame is compared with the corresponding blocks within a search area of size $(M+2p)\times(N+2p)$ pixels in the previous frame, where p is the maximum displacemet allowed [5],[10]. The motion vector of the current block is found. We briefly describe the operation of BMA (Fig.2) between two consecutive frame.

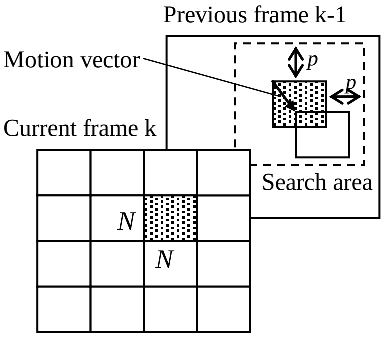

Fig. 2 Principle of block matching

Assume $x_k(m,n)$ to be location of the pixel of the current block in the current frame, and $x_{k-1}(m+i,n+j)$ to be the location of the pixel in the candidate block in the previous frame, shifted by the *i* pixels and *j* lines within the search area. For the best match, the motion vector (i,j) represents the estimate of displacement in horizontal and vertical direction, respectively. The accuracy of motion estimation depends on the matching criteria (cost function) applied in the search area. The most popular ones are briefly discribed as follows :

- *CrossCorrelation Function*

The CrossCorrelation Function [11] is derived from the correlation between two random variables. Correlation is the measure od dependence between the random variables. Correlation ranges form 1 to -1

- a correlation of 1 means the random variables are completely dependent
- a correlation of 0 means the random variables are completely independent
- a correlation of -1 means the random variables are inversely related.

The correlation ρ between two random variables U and V is defined as

$$\rho = \frac{Cov(U,V)}{Var(U)Var(V)} \qquad (1)$$

where

- $Cov(U,V)$ is the covariance between U,V
- $Var(U)$ and $Var(V)$ are the variances of U,V, respectively.

The CrossCorrelation Function defines the random variable U as pixel x_k values in the current block and V as pixel values x_{k-1} in the previous block. Thus a sample U_{mn} is defined as

$$U_{mn} = x_k(m,n) \qquad (2)$$

and the sample V_{mn} is defined as

$$V_{mn} = x_{k-1}(m,n) \qquad (3)$$

where

- m ranges from 1 to M
- n ranges from 1 to N.

The variance of the random variable U is defined as

$$Var(U) = E^2(U) - E(U^2) \qquad (4)$$

where

- $E^2(U)$ is the square of the expected value, or average, of U
- $E(U^2)$ is the expected value of the square of U.

The covariance between U and V is defined as

$$Cov(U,V) = E(U*V) - E(U)*E(V) \qquad (5)$$

where

- $E(U*V)$ is the expected value of the product of the random variables U and V
- $E(U)$, $E(V)$ are the expected values of random variables U,V, respectively.

The Cross-Correlation Function may now be written

$$CCF(i,j)= \tag{6}$$
$$=\frac{\sum_{m=1}^{M}\sum_{n=1}^{N}x_k(m,n)-x_{k-1}(m+i,n+j)-\sum_{m=1}^{M}\sum_{n=1}^{N}x_k(m,n)\sum_{m=1}^{M}\sum_{n=1}^{N}x_{k-1}(m+i,n+j)}{\sqrt{\sum_{m=1}^{M}\sum_{n=1}^{N}x_k^2(m,n)-\left(\sum_{m=1}^{M}\sum_{n=1}^{N}x_k(m,n)\right)^2}\sqrt{\sum_{m=1}^{M}\sum_{n=1}^{N}x_{k-1}^2(m+i,n+j)-\left(\sum_{m=1}^{M}\sum_{n=1}^{N}x_{k-1}^2(m+i,n+j)\right)^2}}$$

If the assumption is made that the expectation of x_k and x_{k-1} are 0, we can obtain Normalized Cross-Correlation Function

$$NCCF(i,j)=\frac{\sum_{m=1}^{M}\sum_{n=1}^{N}x_k(m,n)x_{k-1}(m+i,n+j)}{\sqrt{\sum_{m=1}^{M}\sum_{n=1}^{N}x_k^2(m,n)}\sqrt{\sum_{m=1}^{M}\sum_{n=1}^{N}x_{k-1}^2(m+i,n+j)}} \tag{7}$$

In this measure, the highest *NCCF(i,j)* within the search area, represents the best match.

- *Method of Correlation Coefficients (CC)*

a) The mean value of block can be expressed in the form

$$\bar{X}_k=\frac{\sum_{m=1}^{M}\sum_{n=1}^{N}x_k(i,j)}{MN} \tag{8}$$

$$\bar{X}_{k-1}(i,j)=\frac{\sum_{m=1}^{M}\sum_{n=1}^{N}x_{k-1}(m+i,n+j)}{MN}, \tag{9}$$

b) The standard deviation has following form

$$S_k=\sqrt{\frac{\sum_{m=1}^{M}\sum_{n=1}^{N}\left[x_k(m,n)-\bar{X}_k\right]^2}{MN-1}} \tag{10}$$

$$S_{k-1}(i,j)=\sqrt{\frac{\sum_{m=1}^{M}\sum_{n=1}^{N}\left[x_{k-1}(m+i,n+j)-\bar{X}_{k-1}(i,j)\right]^2}{MN-1}} \tag{11}$$

c) The covariance between blocks can be calculated

$$Covx_kx_{k-1}(i,j)=\frac{\sum_{m=1}^{M}\sum_{n=1}^{N}x_k(m,n)x_{k-1}(m+i,n+j)}{MN}-\bar{X}_k\bar{X}_{k-1}(i,j) \tag{12}$$

d) At last correlation coefficient has following form

$$r'(i,j)=\frac{Covx_kx_{k-1}(i,j)}{S_kS_{k-1}(i,j)} \tag{13}$$

In this method, the maximum of absolute value is chosen for the best match.

- *Mean square error (MSE)*

$$MSE(i,j)=\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}\left[x_k(m,n)-x_{k-1}(m+i,n+j)\right]^2 \tag{14}$$

Reversal, for the smallest *MSE(i,j)* within the search area, (i,j) represents the motion vector of the block. MSE is simpler than NCCF in computational complexity [9].

- *Mean absolute difference (MAD)*

$$MAD(i,j)=\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}\left|x_k(m,n)-x_{k-1}(m+i,n+j)\right| \tag{15}$$

In this criterion, the motion vector is determined by the smallest *MAD(i,j)* for all possible displacement (i,j) within the search area. The MAD is well known applicated due to its lower computational complexity.

- *Pixel Difference Classification (PDC)*

In order to reduce the computational complexity of matching criteria presented above, Gharaviri and Mills [3] have proposed a simple block matching criterion. The Pixel Difference Classification (PDC) cost function is defines as:

$$PCD(i,j)=\sum_{m=1}^{M}\sum_{n=1}^{N}G(m,n,i,j), \tag{16}$$

where:

$$G(m,n,i,j)=\begin{cases}1\xrightarrow{if}\left|x_k(m,n)-x_{k-1}(m+i,n+j)\right|\le T\\ 0\longrightarrow otherwise\end{cases} \tag{17}$$

and T is a treshold.

- *Minimized Maximum Error (MiniMax)*

Minimized Maximum Error (MiniMax) [4] cost function is decribed as follows:

$$MiniMax(i,j)=\min\left\{\left|x_k(m,n)-x_{k-1}(m+i,n+j)\right|_{\max}\right\} \tag{18}$$

In each displacement of block matching is found the maximum of the absolute values of pixel difference among all pixels of the block. Then the minimum of the found maximum distorton values among candidate blocks is chosen as the best match.

4. BLOCK MATCHING ALGORITHMS

In this contribution some of the most commonly used block motion estimation algorithms are presented [7],[8],[13]. From the point of view PSNR has been evaluted these algorithms:

- Full Search algorithm (FS)
- Block-Based Gradient Descent Search (BBGDS)
- Binary Search algorithm (BINS)
- Boomerang Search algorithm (BOOMS)
- Conjugate Direction Search algorithm (CDS)
- Cross Search Algorithm (CSA)
- Diamond search algorithm (DIAMOND)
- Dynamic window adjustment search (DYNAM)
- Full Search of H a V Direction (FSHVD)
- Four Step Search Algorithm (FSSA)
- Modified Motion Estimation search (MME)
- New Prediction Search Algorithm (NPSA)
- New Three Step Search algorithm (NTSS)
- One-Dimens Full Search algorithm (ODFS)
- One-at-a-Time Search algorithm (OTS)
- Parallel Hierarch. One-Dimens. Search (PHODS)
- R search algorithm (R)
- Simple and Efficient Search algorithm (SES)
- Spiral search algorithm (SPIRAL)
- Two-Dimensional Logarithmic search (TDLOG)
- Three Step Search algorithm (TSS)

All presented fast search algorithm, excepting full search algorithm, eliminated the positions in search area by principles of quadrant monotonic function of distortion.

- *Definition of quadrant monotonic:*

Suppose $O=(x_O,y_O)$ is the optimum search point, and $A=(x_A,y_A)$ is any other point in search area. A function $D(x,y)$ is called *quadrant monotonic*, if $D(X)<D(A)$ for any $X=(x_X,y_X)$ from search area, that satisfies the following conditions:

a) X a A belong to the same quadrant with respect to O, that is, $x_X - x_O$ (and $y_X - y_O$) has the same sign as $x_A - x_O$ (and $y_A - y_O$)

b) $|x_X - x_O| < |x_A - x_O| and |y_X - y_O| \le |y_A - y_O|$, or
$|x_X - x_O| \le |x_A - x_O| and |y_X - y_O| < |y_A - y_O|$

The following properties are derived based on the quadrant monotonous model assuming that:

1. $O=(x_O,y_O)$ is the optimum (minimum) point in search area.

2. Two distinct search point $A=(x_A,y_A)$ and $B=(x_B,y_B)$ have been placed in search area and $D(A)>D(B)$.

Property 1:
If $y_A = y_B$ and $x_A > x_B$, then O cannot exist in the half plane defined by $\{(x,y)\in SA | x \ge x_A\}$.

Property 2:
If $x_A > x_B$ and $y_A > y_B$, then O cannot exist in the quadrant defined by $\{(x,y)\in SA | x > x_A and y > y_A\}$.

5. EXPERIMENTAL RESULTS

Block matching algorithms were tested on the two following frame by three different sequences (Fig. 3a,b,c) in the format **raw** (256x256, 8 bpp). Degree of reduction interframe redundancy was evaluated by Peak-signal-to-noise-ratio (PSNR).

$$PSNR = 10\log_{10}\left\{\frac{(255)^2}{\frac{1}{256*256}\sum_{i=1}^{256}\sum_{j=1}^{256}\left(x_{i,j}-x_{i,j}^{C}\right)^2}\right\}\ [dB] \qquad (19)$$

where
- $x_{i,j}$ are values of pixels into the current frame
- $x_{i,j}^{C}$ are values of pixels into the previous frame after motion compensation.

Fig. 3a Frame of sequence ***M.raw***

Fig. 3b Frame of sequence ***P.raw***

Fig. 3c Frame of sequence ***S.raw***

For evaluated presented cost function has been chosen to estimation the motion vector full search algorithm. The full search algorithm finding the motion vectoc by location all possible candidate blocks and do not reduce the number of calculated values of cost function, but given the best results in sense estimation of the motion vectors. For this purpose the full search algorithm is accounted as referent algorithm.

In the Tab. 1-3 are the evaluations of presented cost function. The values of Peak-Signal-to-Noise-Ratio (PSNR) has been obtained by range of parameter *p* from 2 to 10 and the size of block of pixel has been equal to 8x8.

Tab. 1 Cost functions - ***M.raw***

p	2	4	6	8	10
NCCF	37.538	37.572	37.490	37.469	37.358
KK	37.182	37.110	36.965	36.658	34.360
MSE	**37.581**	**37.620**	**37.627**	**37.631**	**37.648**
MAD	37.477	37.516	37.524	37.522	37.531
MiniMax	37.093	37.108	37.086	37.099	37.112
PDC	36.595	36.585	36.537	36.502	36.499

Tab. 2 Cost functions - ***P.raw***

p	2	4	6	8	10
NCCF	27.754	28.202	28.140	27.369	26.786
KK	26.967	26.673	26.353	25.628	25.066
MSE	**28.050**	**28.671**	**28.947**	**29.061**	**29.146**
MAD	27.907	28.515	28.803	28.897	28.968
MiniMax	27.306	27.795	28.019	28.115	28.136
PDC	27.607	28.076	28.244	28.362	28.413

Tab. 3 Cost functions - ***S.raw***

p	2	4	6	8	10
NCCF	34.229	34.136	34.023	33.444	32.938
KK	33.769	33.376	32.746	31.990	31.333
MSE	**34.384**	**34.449**	**34.465**	**34.485**	**34.498**
MAD	34.281	34.377	34.386	34.396	34.406
MiniMax	33.829	33.787	33.785	33.795	33.735
PDC	33.296	33.121	33.130	33.105	33.085

In Tab. 4 is the evaluation of the presented block matching algorithms together with comparison in sense of computational complexity. Mean square error has been applicated as a cost function by maximum displacement p=6.

Tab. 4 Comparison od presented block matching motion estimation algorithms

Sequences	**M**		**P**		**S**	
	PSNR	SP	PSNR	SP	PSNR	SP
FS	37.627	169.00	28.947	169.00	34.465	169.00
BBGDS	37.594	12.06	28.570	16.33	34.403	12.94
BINS	37.288	10.23	27.469	13.34	33.219	10.95
BOOMS	37.596	6.71	28.422	15.06	34.311	10.05
CDS	37.447	8.21	28.347	11.69	33.802	9.96
CSA	36.673	6.05	26.522	14.95	32.619	16.98
DIAMOND	36.954	27.09	28.193	23.76	33.217	24.39
DYNAM	36.491	17.86	27.660	20.25	31.920	19.77
FSHVD	37.328	25.00	28.660	25.00	33.691	25.00
FSSA	37.286	18.15	28.554	18.41	33.568	18.17
MME	36.655	7.11	28.004	15.05	32.702	17.10
NPSA	37.197	15.25	28.548	17.27	34.380	16.12
NTSS	37.591	18.09	28.729	22.72	34.384	20.31
ODFS	37.398	36.93	27.856	36.39	33.354	36.83
OTS	37.311	7.04	27.279	12.42	33.544	9.18
PHODS	36.566	13.00	26.298	13.00	30.614	13.00
R	37.593	9.38	28.355	11.80	34.385	10.46
SES	37.339	17.80	27.999	17.58	33.234	17.33
SPIRAL	34.776	23.52	26.787	23.10	30.205	23.54
TDLOG	37.311	14.94	28.556	19.45	33.724	16.53
TSS	36.783	25.00	27.072	25.00	32.241	25.00

where: SP is the average required number of search points

6. CONCLUSION

Full search algorithm is based on the matched all possible displaced candidate blocks within the search area in the previous frame, in order to find the block with the minimum distortion. The main disadvantage of this procedure is a very high computational complexity.

Computational complexity is higher by using some of the correlation methods. Using by some of the functions distortion are achieved better performances as in the case of using by some correlation functions. In sense of the computation complexity the application of MAD as distortion criterion is optimal, because MAD needs only the adding operations.

The best PSNR values given the FS procedure using by MSE. Main reason, why MSE achieved the best performances is that the denominator of PSNR formula has the same form as MSE. The denominator is the reason, why the performance of

NCCF and correlation coefficient is not monotonous function by parameter *p*, too.

For this purpose the MSE has been chossen as cost function in the second part of contribution to estimate the motion vector by fast search algorithms.

About full search algorithm we can say, that the algorithm always find the global minimum of cost function, but his computational complexity is very high.

Reason, why the fast block matching algorithms given lower performances as FS procedure is, that in the search area does not exist always only one global minimum, but often in search area exist several local minimums of cost function. Currently location of these local minimums is ultimated for finding optimal point.

From the tab. 4 we can see, that the best performances in sense to achieve the highest PSNR values are obtained by the Block-based gradient descent search, Boomerang search algorithm, New prediction search algorithm, New three step search algorithm and , R search algorithm, in generally. By the block matching technique the second criterion is the average required number of seach points. The Boomerang search algorithm, Conjugate direction search algorithm, One-at-a-time search algorithm, Parallel hierarchical 1-dimens. search, R search algorithm finding the motion vector with the minimum search points. At least, the compromise between two criteria are the Boomerang search algorithm and R seach algorithm.

Correctly estimation of the motion vector can be executed by the assumption, that all pixels in block do the same move. Reliability of the estimation displacement of block depends on sizes of blocks and on largeness of movement. Using by less sizes of blocks increasing probability, that in the SA will exit more identical blocks.

Block matching algorithms come under algorithms, which are often implemented to the videoconference and videotelephone systems. Their application is based on the high efficiency by reduction interframe redundancy with low computation complexity.

REFERENCES

[1] Dufaux, F., Moscheni, F.: Motion Estimation Techniques for Digital TV : A Review and a New Contribution, Proceedings of the IEEE, Vol. 83, No. 6. June 1995, pp. 858-876.

[2] Furht, B., Greenberg, J., Westwater, R.: Motion Estimation Algorithms for Video Compression, Kluwer Academic Publishers, Massachusetts 1997.

[3] Gharavi, H., Mills, M. : Blockmatching Motion Estimation Algorithms - New Results, IEEE Transactions on Circuits and Systems, Vol. 37, No. 5, May 1990, pp. 649-651.

[4] Chen, M. J., Chen, L. G., Chiuech, T., D., Lee, Y., P. : A New Block Matching Criterion for Motion Estimation and its Implementation, IEEE Transactions on Circuits and Systems for video technology, Vol. 5, No. 3, June 1995, pp. 231-236.

[5] Chu, Ch. T., Anastassiou, D., Chang, S. F. : Hybrid Object-based/block-based Coding in Video Compression at Very Low bit-rate, Signal Processing : Image Communication , Vol. 10, 1997, pp. 159-171

[6] Karpf, M.: Kódovanie videa, metódy druhej generácie, Telekomunikácie a podnikanie, 2/2000, s. 16-18.

[7] Liu, L. K., Feig, E. : A Block-Based Gradient Descent Search Algorithm for Block Motion Estimation in Video Coding, IEEE Transactions on Circuits and System for Video Technology, Vol. 6, No. 4, August 1996, pp. 419-422.

[8] Po, L. M., Ma, W. Ch. : A Novel Four-Step Search Algorithm for Fast Block Motion Estimation, IEEE Transactions on Circuits and System for Video Technology, Vol. 6, No. 3, June 1996, pp. 313-317.

[9] Radoczi, P. : Block matching motion estimation algorithm with least square approximation, In Zborník I. Doktorandskej konferencie FEI TU Košice, Košice 2001, s. 75-76.

[10] Lukáč, R. : Impulse Detectors for Noised Sequences, Radioengineering, Vol. 10, No. 2, July 2001, pp. 24-31.

[11] Radoczi, P., Levický, D. : Comparison Function of Distortion for Block Matching Algorithms, In Zborník Nové smery v spracovaní signálov V, Liptovský Mikuláš 2000, s. 147-152.

[12] Grgic,M., Grgic.S., Zovko-Cihlar.B. : Image Retrieval Based on Texture Features Extraction, Proc. ECMCS 2001 EURASIP Conference, Budapest, Hungary, pp. 152-155.

[13] Radoczi, P., Levický, D. : Evaluation of Efficiency Some New Algorithms for Block Matching, In 10th International Czech-Slovak Scientific Conference, Proceedings Radioelektronika 2000, Bratislava, pp. III52-III55.

BIOGRAPHY

Peter Radoczi was born on 1976. In 1999 he graduated (MSc.) at the department of Electronics and Multimedia Communication of the Faculty of Electrical Engineering and Informatics at Technical University of Košice. In this time he is PhD. student. His research interest includes motion estimation, digital image processing, and multimedia communications.