

# Content Adaptive True Motion Estimator for H.264 Video Compression

Jozef HUSKA, Peter KULLA

Dept. of Radio and Electronics, Slovak University of Technology, Ilkovičova 3, 812 19 Bratislava 1, Slovak Republic

jozef.huska@stuba.sk, ppkulla@elf.stuba.sk

**Abstract.** *Content adaptive true motion estimator for H.264 video coding is a fast block-based matching estimator with implemented multi-stage approach to estimate motion fields between two image frames. It considers the theory of 3D scene objects projection into 2D image plane for selection of motion vector candidates from the higher stages. The stages of the algorithm and its hierarchy are defined upon motion estimation reliability measurement (image blocks including two different directions of spatial gradient, blocks with one dominant spatial gradient and blocks including minimal spatial gradient). Parameters of the image classification into stages are set adaptively upon image structure. Due to search strategy are the estimated motion fields more corresponding to a true motion in an image sequence as in the case of conventional motion estimation algorithms that use fixed sets of motion vector candidates from tight neighborhood.*

## Keywords

Motion estimation, H.264, true motion field, multi stage, estimation reliability, objects projection.

## 1. Introduction

Motion estimation and compensation is an important part of the most of today used video compression standards, such as MPEG-1/2/4 and ITU-T H.261/263/264, since it can effectively reduce temporal redundancy that exists between frames of video sequences. Block based motion estimation (ME) techniques are the ones most widely used due to their relative implementation simplicity and efficiency. The idea is to partition the current frame in time  $t$  into small square blocks of pixels and by using a predefined distortion criterion locate the best match for these blocks inside a reference frame in time  $t-d_t$  or  $t+d_t$ . This best match is then used as a predictor for the block in the current frame, whereas the displacement between the two blocks defines a motion vector (MV) which is associated with the current block. Thus, it is only necessary to send the motion vector and additional residue block defined as the quantized difference between the current block

and the predictor. This method, in general, requires significantly fewer bits than the direct coding of the original.

The research in the field of ME has a long history but there are still things that can be improved or be done by another approach. Several algorithms have been proposed in an effort to reduce complexity by reducing the number of checking points examined. The most known techniques are for example the 3-step search [1] and the Diamond search (DS) [2], [3] algorithms. These techniques, even though could make the search faster, unfortunately balanced by considerably lower quality as by the exhaustive search also called Full search (FS). Then algorithms that not only reduce the number of checking points, but can also retain, and in many cases improve, video quality were presented. These techniques can accomplish this by initially considering several highly likely predictors (from the current block spatial/temporal neighborhood) and introducing early stopping criteria to terminate the search when some of the candidates met the threshold. This group of ME techniques represent for example the Predictive DS [4] and the Enhanced predictive zonal search (EPZS) [5]. The last named EPZS algorithm was accepted by Joint Video Team of ISO/IEC MPEG & ITU-T VCEG for implementation in H.264/MPEG-4 AVC Reference Software, [6].

There presented ME algorithm struggles to minimize negative influence of the ME ambiguity problems [7] on output quality that are still current despite of long year research in this field. Due to ME ambiguity the estimated motion fields are in some cases chaotic and have poor relation with the true motion of objects in the scene. Then the results are wrong predictors that cause larger bits demands to encode block residues. Used approach to accomplish minimization of ME ambiguity influence is completely new in its basis. Our adaptive ME reliability measurement is used [8] to distinguish the most problematic image blocks for ME. Not minor novelty consists in the selection of the most likely candidates for the current block MV which appears to be the most important feature and the key of algorithms performance. In the selection ME reliability measurement and the nature of the projection theory from 3-D scene to 2-D image plane is reflected [9], [10]. That is the largest difference against conventional “hard wired” tight spatial neighborhood selection in the

3DRS, PDS as well as the EPZS algorithms. Due to used approach in our ME algorithm it can be labeled as adaptive true motion estimator.

## 2. Prior Work

Our previous work in the field of ME involves two ME algorithms named “true region motion field estimator” (TRMF) [11], and “modified TRMF” [12], from which there presented algorithm outgoes. Both TRMF algorithms used the concepts of 3DRS and PDS algorithms considering predictive criteria based on recursive and multi stage approach. The difference between them was in the used approach in image blocks classification. In the TRMF search algorithm graph-based image segmentation in combination with feature detection algorithm [13], [14] was used, modified TRMF search involved only feature detection algorithm.

Both algorithms achieved better results as the DS and the 3DRS algorithms especially in the case of synthetic image sequences. The performance (in PSNR) of the both TRMF search algorithms against the PDS algorithm was balanced (+/- 0.1dB). The TRMF search achieved +0.05dB better PSNR value as the modified TRMF search but its execution time was 0.1 times longer. The largest drawback of the both TRMF search algorithms against the well known EPZS algorithm consists in missing zero candidate and early stop conditions.

## 3. Ambiguity in Motion Estimation

The luminance variation in a video sequence is represented by  $I(x,y,t)$ . An image point  $(x,y)$  at time  $t$  is moved to  $(x+d_x,y+d_y)$  at time  $t+d_t$ . Under the constant intensity assumption (image point has the same intensity along the trajectory of movement) it can be written

$$I(x+d_x,y+d_y,t+d_t) = I(x,y,t). \quad (1)$$

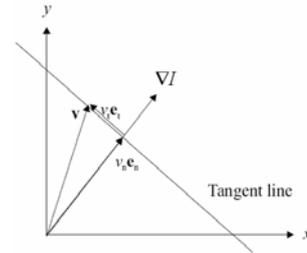
Using Taylor’s expansion, when  $d_x, d_y, d_t$  are small, and combining the result with (1) yields

$$\frac{\partial I}{\partial x}d_x + \frac{\partial I}{\partial y}d_y + \frac{\partial I}{\partial t}d_t = 0 \quad (2)$$

in terms of the motion vector  $(d_x,d_y)$ . Dividing both sides by  $d_t$  it results in

$$\frac{\partial I}{\partial x}v_x + \frac{\partial I}{\partial y}v_y + \frac{\partial I}{\partial t} = 0 \quad (3)$$

where  $(v_x,v_y)$  represent the components of velocity vector  $\mathbf{v}$ . The above equation is known as the optical flow equation [7].



**Fig. 1.** Decomposition of motion  $\mathbf{v}$  into normal  $(v_n e_n)$  and tangent  $(v_t e_t)$  components [7].

As shown in Fig. 1, the flow vector  $\mathbf{v}$  at any points  $(x, y)$  can be decomposed into two orthogonal components as

$$\mathbf{v} = v_n \mathbf{e}_n + v_t \mathbf{e}_t \quad (4)$$

where  $\mathbf{e}_n$  is the direction vector of the image gradient

$$\nabla I = \left[ \frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \right] \quad (5)$$

to be called the normal direction, and  $\mathbf{e}_t$  is orthogonal to  $\mathbf{e}_n$ , to be called the tangent direction. Three consequences as in [7] from (3) are:

- *Underdetermined component  $v_t$ :* there is only one equation for two unknowns ( $v_x$  and  $v_y$ , or  $v_n$  and  $v_t$ ). That can be dealt with the motion measurement applied to a block of pixels [15].
- *Aperture problem:* the projection of the motion vector along the normal direction is fixed, with

$$v_n = -\frac{\partial I}{\partial t} / \|\nabla I\| \quad (6)$$

where as the projection onto the tangent direction,  $v_t$ , is undetermined. Any value of  $v_t$  would satisfy the optical flow equation. The word “aperture” refers to the small window over which to apply the constant intensity assumption. The motion can be estimated uniquely only if the aperture contains at least two different gradient directions.

- *Indeterminate flow vector:* in regions with constant brightness so that  $\|\nabla I\| = 0$ , the flow vector is indeterminate. The estimation of motion is reliable only in regions with brightness variation, i.e., regions with edges or non-flat textures.

## 4. 3-D Motion Projection to the 2-D Image Plane

There are some few aspects to consider for ME used in the video coding systems. That are:

- the goal is to estimate the motion of image points, i.e., the 2-D motion or apparent motion,
- motion is estimated based on the variations of intensity and color in image points,
- in regard to a projection of 3-D scene objects onto the image plane, the motion of neighboring image points within the object's projection area is very similar [9].

The assumption that the motion of neighboring image points within the object's projection area (Fig. 2.) is very similar is in our algorithm included in a bigger manner than only estimating motion vectors for blocks of image points.

Let assume that the object from the scene is projected in 2-D image (Fig. 2.) in such a way, that the object's projection area (OPA) (Fig. 3.) is larger than one block of pixels. Than it could be assumed that the motion of all blocks within the same OPA is very similar. Taking into consideration simple rigid 3-D object that undergoes translation motion parallel to 2-D image scanning plane, it can be formulated these assumptions for the observed motion in the 2-D image plane:

- motion of neighboring image points/blocks within OPA is similar,
- motion of inner pixels/blocks of the OPA can be recovered from the motion of the OPA boundary pixels/objects (edges, corners),
- motion of edge pixels/blocks of the OPA can be recovered from the motion of corner pixels/blocks.

When size of the image blocks is small, in comparison with the size of OPA, the previous assumptions can be used for almost all types of motions. That showed useful [8], [16], [11], [12], for ME to recover missing motion components of image blocks where some of the ambiguity was present. In general way the corner elements of OPAs boundary are trouble free for ME, on contrary ME of edge components meets with the aperture problem and ME of OPAs inner is the most problematic due to obviously including small or zero brightness variability.

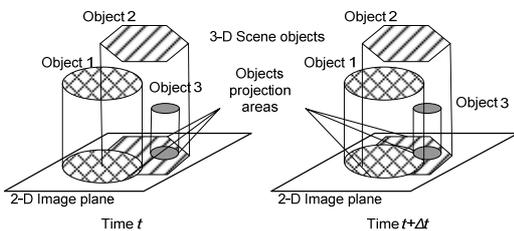


Fig. 2. 3-D Scene objects projection in 2-D image plane.

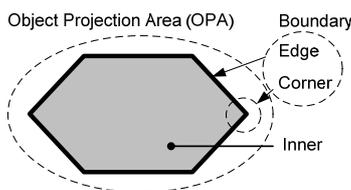


Fig. 3. Description of object projection area elements in 2-D image plane.

## 5. Description of the Content Adaptive True Motion Estimator

The Content Adaptive True Motion Estimator algorithm consists of three processing blocks as shown in Fig. 4.: adaptive image blocks evaluation of ME reliability, MV candidates selection and multi stage ME with use of EPZS framework including search and early stopping conditions. In comparison with the conventional ME algorithms that take into consideration MV candidates from spatial neighborhood [17], [4], [5] are instead of automatically taken candidates from the tight neighborhood (left, top, top-right, top-left) selected the candidates that are characterized with a higher ME reliability than current. The method of MV candidates selection upon ME reliability measure (presence of ME ambiguity type) coincides with the assumptions for 2-D motion from previous paragraph. Chosen approach requires multi stage ME scheme that starts with the highest and ends with the lowest reliability MVs.

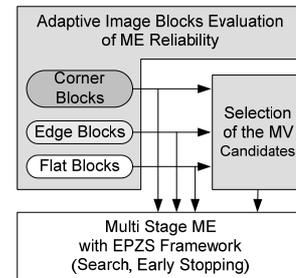


Fig. 4. Principal scheme of the Content Adaptive True Motion Estimator.

### 5.1 Adaptive Image Blocks Evaluation of ME Reliability

The reliability of ME for each image block is influenced with the occurrence of the type of ME ambiguity listed in the 3<sup>rd</sup> chapter. There are classified three types of image blocks in relation to ME from the highest to the lowest reliability estimation: trouble free estimation, aperture problem (undetermined  $v_t$  component) and indeterminate MV (undetermined both  $v_b$ ,  $v_n$  components). Covariance matrix of the image block which is a good indicator of the distribution of image spatial gradient over a rectangle block [10] is used to classify image blocks to three classes upon ME reliability. Measurement of image structures distribution is in the literature also called feature detection [13], [10], [18]. The algorithm is based on the relationship between eigenvalues  $\lambda_1, \lambda_2$  (8) of an image covariance matrix  $A$ :

$$A = \sum_{x \in R_x} \sum_{y \in R_y} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad (7)$$

where

$$a_{11} = \left( \frac{\partial I(x, y, t + \Delta t)}{\partial x} \right)^2, \quad a_{12} = \frac{\partial^2 I(x, y, t + \Delta t)}{\partial x \partial y},$$

$$a_{21} = \frac{\partial^2 I(x, y, t + \Delta t)}{\partial y \partial x}, \quad a_{22} = \left( \frac{\partial I(x, y, t + \Delta t)}{\partial y} \right)^2$$

$$\lambda_{1,2} = \frac{1}{2} \left[ (a_{11} + a_{22}) \pm \sqrt{4a_{12}a_{21} + (a_{11} - a_{22})^2} \right] \quad (8)$$

Summarizing the relationship between the matrix  $A$  and image block structure, small eigenvalues of  $A$  correspond to a relatively constant intensity ( $\|\nabla I\| = 0$ ) within an image block which is referred to as the block with the lowest reliability ME. It is usually valid for the inner of OPA. In the further text these blocks will be marked as flat blocks. A pair of large and small eigenvalues correspond to an uni-directional texture pattern (one dominant direction of the  $\nabla I$ ) and that block which satisfies this is referred to as the block with the middle reliability ME, in the next referred to as edge blocks (they commonly create edge element of OPAs boundary as shown in Fig. 3). Two large eigenvalues represent the blocks with the highest reliability ME (two different directions of the  $\nabla I$ ) and in the further text are denoted as the corner blocks (in general coincide with the corner elements of OPAs boundary).

Image blocks classification is influenced with two parameters: maximal permitted difference between eigenvalues  $DIF_{TH}$  (in percentage) for the corner blocks and the threshold for minimal eigenvalues  $MIN_{TH}$  for the edge and corner blocks (in ppm from the maximal theoretical eigenvalue for the selected image block size). Blocks classification depending on these parameters is described in Fig. 5.

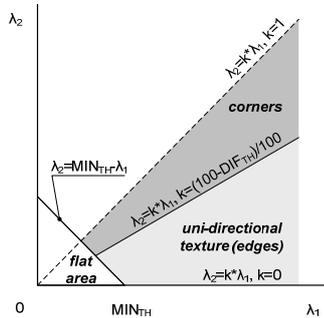


Fig. 5. Image blocks classification into three types: corner blocks, edge blocks and flat blocks [8].

The classification of the image blocks into the named classes designates the blocks from that MV candidates for the ME of the current block will be taken. From that implies that image blocks classification parameters ( $DIF_{TH}$ ,  $MIN_{TH}$ ) have a large influence on the exactness of predicted MV which directly affects the ME performance as shown in [8]. Tests showed that the best performance of the ME algorithm was reached with the thresholds that are set adaptively upon current image content. We used statistic properties, in concrete the histogram, of image blocks covariance matrix eigenvalues distribution for the assumed standard image sequence [8].

### 5.1.1 Image Covariance Eigenvalues Statistic

Let  $\Lambda_1$ ,  $\Lambda_2$  be probability random variables with

values from the groups of discrete values

$$\Lambda_1 \in H_1 = \{ \lambda_1(0), \lambda_1(1), \dots, \lambda_1(i), \dots, \lambda_1(m) \}$$

$$\Lambda_2 \in H_2 = \{ \lambda_2(0), \lambda_2(1), \dots, \lambda_2(i), \dots, \lambda_2(m) \} \quad (9)$$

where  $\lambda_1(i)$ ,  $\lambda_2(i)$  are the eigenvalues of image covariance matrix  $A$  of the image block  $i=1, 2, \dots, m$ . The groups  $H_1$ ,  $H_2$  can be split into  $k$  class intervals of size step

$$I_{i,j} = (I(\lambda_{1,i}^*), I(\lambda_{2,j}^*)) \quad (10)$$

where

$$I(\lambda_{1,i}^*) = \left\langle \lambda_{1,i}^* - \frac{step}{2}, \lambda_{1,i}^* + \frac{step}{2} \right\rangle,$$

$$I(\lambda_{2,j}^*) = \left\langle \lambda_{2,j}^* - \frac{step}{2}, \lambda_{2,j}^* + \frac{step}{2} \right\rangle \quad (11)$$

and  $i=1, 2, \dots, k$ ,  $j=1, 2, \dots, k$ ,  $\lambda_{1,i}^*$ ,  $\lambda_{2,j}^*$  are middle values of interval  $I_{i,j}$ . When taking  $T(\Lambda_1, \Lambda_2)$  as discrete random vector with the probability function  $f(\lambda_1, \lambda_2)$ , it can be written for each class interval  $I_{i,j}$  the equation for occurrence probability of the discrete random vector  $T(\Lambda_1, \Lambda_2)$  inside the interval  $I_{i,j}$  (histogram's interval values):

$$P_{i,j}(\Lambda_1 \in I(\lambda_{1,i}^*), \Lambda_2 \in I(\lambda_{2,j}^*)) =$$

$$= \sum_{\lambda_1 \in I(\lambda_{1,i}^*)} \sum_{\lambda_2 \in I(\lambda_{2,j}^*)} f(\lambda_1, \lambda_2) \quad (12)$$

From the examination of the eigenvalues histograms of different common image sequences in [8] it can be written:

- the probability of image blocks occurrence decreases with the increase of  $\lambda_1$ ,  $\lambda_2$ ,
- in histograms there are notable gaps (lowest image blocks occurrence probability) within the regions with the high and low values of  $\lambda_1$ ,  $\lambda_2$  and as well as among regions with large and small difference between  $\lambda_1$ ,  $\lambda_2$ .

### 5.1.2 $MIN_{th}$ Threshold

It shows useful to compute the sums of histogram  $P_{i,j}$  values for the intervals  $I_{ij}$  those middle values fulfill condition

$$\lambda_{2,j}^* = MIN_{TH} - \lambda_{1,i}^* \quad (13)$$

for every  $MIN_{TH} \in \langle 0, MIN_{TH} \max \rangle$  (ppm). It is the equation of line on which the eigenvalues of the same sum are lying. The  $MIN_{TH}$  value that gives minimal sum of histogram  $P_{i,j}$  values after appearance of two largest local maxima, designates the proper  $MIN_{TH}$ . Minimal sum should not be searched in the whole  $MIN_{TH}$  range because from the global view the sum of  $P_{ij}$  for intervals  $I_{ij}$  that fulfill (13) is a decreasing function. We have empirically selected the  $MIN_{TH} \max$  to be 30000 ppm which corresponds to a standard deviation of brightness values inside a block equal to 15.61. It is searched for the minimum after two largest

local maxima due to expectation that each ordinary image contains two most numerous types of image blocks that should be classified as flat: blocks with almost zero spatial gradient (constant brightness regions) and blocks with “salt and pepper” texture with small spatial gradient (noise and soft textured areas) [8].

### 5.1.3 DIF<sub>th</sub> Threshold

As in the case of MIN<sub>TH</sub>, the DIF<sub>TH</sub> is assigned following the eigenvalues histogram properties. In contrast to finding MIN<sub>TH</sub>, in the case of DIF<sub>TH</sub> determination is searched for the first local maximum of the sums of histogram  $P_{i,j}$  values for the intervals  $I_{ij}$  those middle values fulfill condition:

$$\lambda_{2,j}^* = \frac{100 - DIF_{TH}}{100} \lambda_{1,i}^* \quad (14)$$

for every  $DIF_{TH} \in \langle 0, 100 \rangle$  (%). It is the equation of line on which the eigenvalues with the same percentage difference between them are lying. The sum of  $P_{i,j}$  values increases with increasing percentage difference between eigenvalues. This property results from the ordinary image in which the probability occurrence of sharp edges is higher than of sharp corners. The search for the first maximal increase of the occurrence probability sum starting at DIF<sub>TH</sub>=0% is done due to finding most numerous group of image blocks with the lowest difference between eigenvalues. That image blocks are the highest quality blocks from the view of the largest ME reliability due to including minimal two different directions of spatial gradient [8].

## 5.2 Selection of the MV Candidates

The MV candidates are defined in two sets A and B. In the selection ME reliability measurement and the nature of the projection theory from 3-D scene to 2-D image plane is reflected. It means that the candidates are selected among previously calculated higher reliability MV than current. Each image blocks class has its own definition of MV candidates except for the corner blocks that do not have any MV candidate (A, B sets include only zero candidates).

The second B candidate set for edge and flat blocks  $MV=(d_x, d_y)$  with coordinate  $[x,y]$  consists of zero predictor  $B_0=(0,0)$  and four predictors  $B_c=(db_{xc}, db_{yc})$ ,  $c=1,2,3,4$  with positions  $[x_c, y_c]$  found in four directions in maximal TH distance from the current block:

- left-up  $B_1[\langle x-TH, x \rangle, \langle y, y-TH \rangle]$ ,
- right-up  $B_2[\langle x, x+TH \rangle, \langle y, y-TH \rangle]$ ,
- left-down  $B_3[\langle x-TH, x \rangle, \langle y, y+TH \rangle]$ ,
- right-down  $B_4[\langle x, x+TH \rangle, \langle y, y+TH \rangle]$ .

As predictors MVs of the nearest blocks that belong to a higher ME reliability level than current are taken. It means that for the edge blocks the nearest MV predictors within corner blocks are selected and for the flat blocks the near-

est MV predictors within corner and edge blocks are selected.

The A candidate set includes only one predictor  $A=(da_x, da_y)$  that is considered to be the most likely candidate (optimal predictor). It is calculated from the B set predictors (except zero predictor  $B_0$ ) as the weighted mean value [19]:

$$da_x = \frac{\sum_{c=1}^4 w_c db_{xc}}{\sum_{c=1}^4 w_c}, \quad da_y = \frac{\sum_{c=1}^4 w_c db_{yc}}{\sum_{c=1}^4 w_c} \quad (15)$$

where weights are used

$$\{w_c\}_{c=1}^4 = \left\{ \frac{\min\{|CD_c|\}_{c=1}^4}{|CD_c|} \right\}_{c=1}^4 \quad (16)$$

involving the distance  $|CD_c|$  between the current MV and the candidate predictor  $B_c$ . This proceeding guaranties continuous tracking of the motion speed and direction variations in the case of rotation, zoom or translation not parallel with 2-D image scanning plane.

## 5.3 Multi Stage ME with EPZS Scheme

An algorithm for the adaptive true motion estimator is in compact form represented with the scheme in Fig. 6. In the scheme is in addition to findings described in previous sections and subsections annexed EPZS scheme [5] for the own search and early stopping conditions. The ME algorithm was implemented in the H.264 reference software version Jm12.2 from the Joint Video Team of ISO/IEC MPEG & ITU-T VCEG [6]. The existing EPZS implementation was used to which additional processing steps for adaptive image blocks evaluation of ME reliability and the selection of the MV candidates were added. The MV candidates in original EPZS were replaced by there described.

## 6. Results

The proposed motion estimation algorithm is applied to the JVT encoder [6] with the variable-block ME/MC delimited to only 16:16 blocks,  $\pm 16$  pixels motion search range, full pixel precision, disabled chroma ME, using one reference frame, and rate-distortion optimization. Several image sequences from [21], each of them with 100 frames, were used for this experiment. Each sequence is compressed with the scheme of I,P,P,..., i.e., only the 1<sup>st</sup> frame is INTRA, while the others are INTER frames without the B frame. The proposed method is compared to the fast ME algorithms implemented in JM12.2 source code: UMHexagon search [20] and EPZS [5] algorithm. The EPZS algorithm had disabled temporal and spatial memory predictors to discover the influence of the spatial predictors selection. There presented algorithm used for the

image blocks classification the size of image block identical to the size of blocks for ME/MC, i.e., 16:16 pixels.

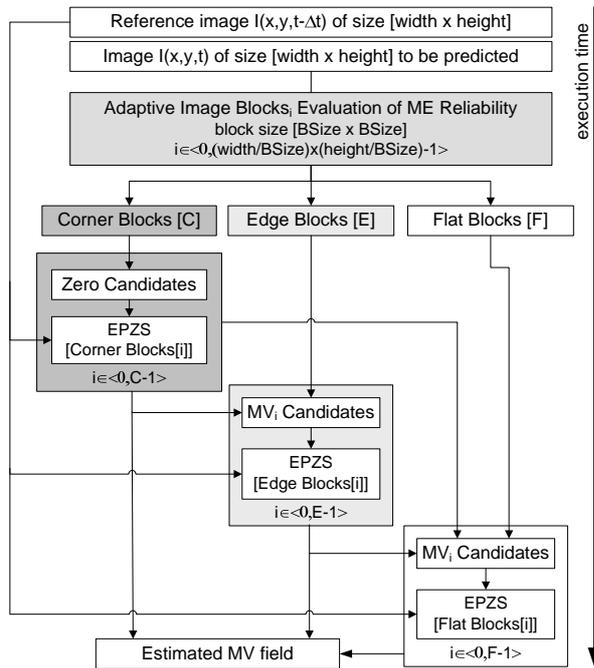


Fig. 6. Algorithm of the adaptive true motion estimator (C, E,F-counts of detected corner, edge and flat blocks).

Results of algorithms comparison with the test sequences are reported with the summarization table Tab. 1. where there are the Y component SNR values at the 110kbps. Dependence graphs of SNR from the bit-rate are shown in Fig. 7.

| test sequence<br>100 frames<br>QCIF | ME algorithm   |                    |                |                    |                |                    |
|-------------------------------------|----------------|--------------------|----------------|--------------------|----------------|--------------------|
|                                     | EPZS           |                    | UMHex          |                    | CATME          |                    |
|                                     | SNR(Y)<br>[dB] | Bit-Rate<br>[kbps] | SNR(Y)<br>[dB] | Bit-Rate<br>[kbps] | SNR(Y)<br>[dB] | Bit-Rate<br>[kbps] |
| salesman                            | 38.49          | 110.2              | 38.52          | 110.0              | 38.52          | 110.1              |
| news                                | 38.16          | 110.2              | 38.28          | 110.2              | 38.49          | 110.2              |
| highway                             | 38.08          | 110.1              | 38.19          | 109.8              | 38.04          | 110.4              |
| claire                              | 44.06          | 108.9              | 44.17          | 109.1              | 43.85          | 109.5              |
| carphone                            | 35.00          | 110.2              | 35.06          | 110.1              | 34.96          | 110.1              |
| bridgeclose                         | 36.45          | 110.7              | 36.48          | 110.7              | 36.48          | 110.7              |

Tab. 1. Comparison of the conventional algorithms EPZS and UMHex with there presented Content Adaptive True Motion Estimator (CATME).

The SNR performance of all three compared algorithms is very similar (average deviation 0.06 dB) within all video

test sequences despite another used approach. There presented ME algorithm (CATME) does not overcome another two algorithms throughout all the test sequences, but the improvement is observable in some of them. In comparison with EPZS CATME is better in the case of “bridge-close“ (in average +0.023 dB), “claire“ (in average +0.03 dB) and especially in “news“ sequence (in average +0.17 dB). In average CATME achieved higher SNR as UMHex in the case of “news“ (+0.05 dB). The video sequences in which CATME achieved higher SNR include large areas with low brightness change and the scene composition is easy to recognize. From the graphs of SNR=f(Bit-Rate) the increase of CATME performance can be seen especially for the Bit-Rate range from 80 kbps to 140 kbps.

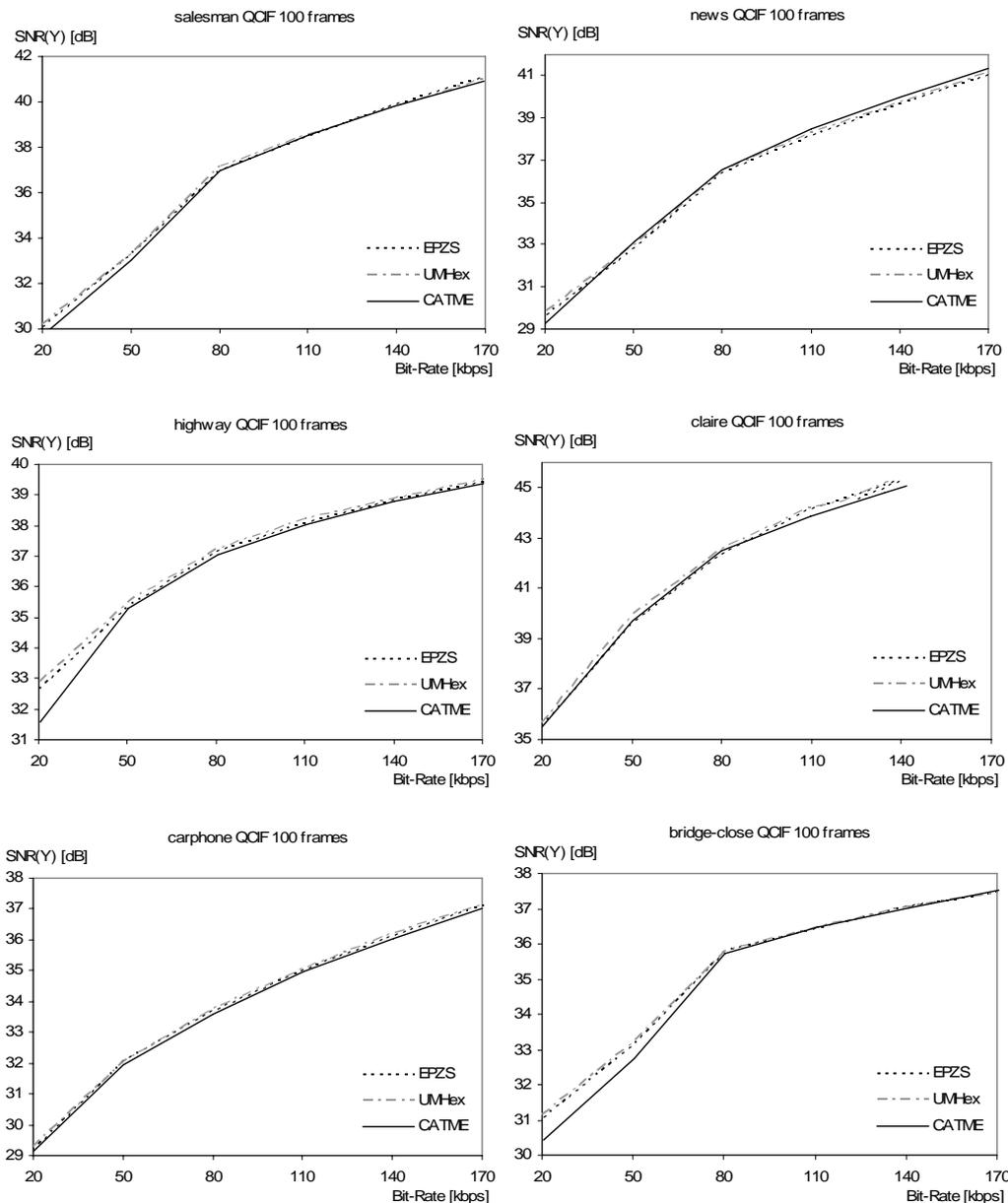
### 7. Conclusions

We presented the original ME algorithm that adapts itself to the input data to overcome ME ambiguity and tries to estimate motion fields more corresponding to the true motion of objects in a scene. Our content adaptive true motion estimator incorporates adaptive ME reliability measurement in conjunction with the nature of the projection theory from the 3-D scene to the 2-D image plane. This approach helps to select more corresponding candidates for the MV estimation that lead to higher performance especially in video sequences where large low brightness variability areas are present as well as consisting from easy separable objects. In the future CATME algorithm would be applied in the real-time frame-rate up conversion and examined prediction quality metric taking into consideration the human visual system, for example [22].

Performance within the “bridge-close, “claire“, “news“ test sequences as well as results of our previous works [8],[11],[12],[16] justify the theory we used for there presented ME algorithm. Essential attribute of the algorithm is not serious decrease of its performance in the case of inputs that do not meet inquiry of easy separable objects.

### Acknowledgements

The research described in the paper was financially supported by the Slovak Ministry of Education under VEGA Grant No.G-1/3107/06.



**Fig. 7.** Graphs of SNR dependent from the Bit-Rate for "salesman", "news", "highway", "claire", "carphone" and "bridge-close" video QCIF sequences.

## References

- [1] VIRK, K., KHAN, N., MASUD, S. Low complexity recursive search based motion estimation algorithm for video coding applications. In *Proc. of the 13<sup>th</sup> European Signal Processing Conference (EUSIPCO 2005)*, Sept. 2005.
- [2] ZHU, S., MA, K. K. A new diamond search algorithm for fast block matching motion estimation. In *Proc. of Int. Conf. Information, Communications and Signal Processing*, 1997, vol.1, pp. 292-6.
- [3] THAM, J. Y., RANGANATH, S., RANGANATH, M., KASSIM, A. A. A novel unrestricted center-biased diamond search algorithm for block motion estimation. *IEEE Trans. on Circuits & Systems for Video Technology*, 1998, vol.8, pp.369-77.
- [4] TOURAPIS, A. M., SHEN, G., LIU, M.L. A new predictive diamond search algorithm for block based motion estimation. In *Proc. of Visual Communications and Image Processing 2000 (VCIP-2000)*, June 2000.
- [5] TOURAPIS, A. M. Fast motion estimation within the H.264 codec. In *Proc. 2003 International Conference on Multimedia and Expo*, vol. 3, pp. 517-20, July 2003.
- [6] TOURAPIS, A. M., H.264/MPEG-4 AVC Reference Software Manual JVT-X072. In *Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG 24th Meeting: Geneva*, pp. 75, June 2007.
- [7] WANG, Y., OSTERMANN, J., ZHANG, Y. Q. *Video Processing and Communications*. Prentice-Hall, 2002, pp. 609.

- [8] HUSKA, J., KULLA, P. Adaptive image regions segmentation based on motion estimation confidence. In *ELITECH 2007 9th Scientific Conference for PhD. Students*. Bratislava: FEI STU v Bratislave, 2007, ISBN 978-80-227-2655-9, CD-ROM.
- [9] CHEN, Y. K., LIN, Y. T., KUNG, S. Y. A feature tracking algorithm using neighborhood relaxation with multi-candidate pre-screening. In *Proc. of International Conference on Image Processing*, Sept. 1996, vol. II., pp. 513-516.
- [10] NEUMANN, U., YOU, S. Adaptive multi-stage 2D image motion field estimation. In *Proc. of SPIE Conf. on Applications of Digital Image Processing XXI*, July 1998, vol. 3460, pp. 116-123.
- [11] HUSKA, J., KULLA, P. A new block based motion estimation with true region motion field. In *IEEE Region 8 EUROCON 2007 The International Conference on Computer as a Tool*. Warsaw, 2007, ISBN 1-4244-0813-X, pp. 182-188.
- [12] HUSKA, J., KULLA, P. A new recursive search with multi stage approach for fast block based true motion estimation. In *Proceedings of 17th International Conference RADIOELEKTRONIKA 2007*. Brno: MJ servis, 2007, ISBN 978-80-214-3390-8, pp. 227-232.
- [13] WITTEBROOD, R., DE HAAN, G. Feature point selection for object-based motion estimation on a programmable device. In *Proc. of Visual Communications and Image Processing*, 2002, vol. 4671, pp. 687-697.
- [14] FELZENSZWALB, P. F., HUTTENLOCHER, D. P. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 2004, vol. 59, no. 2.
- [15] BOVIK, A. *Handbook of Image and Video Processing*. Academic Press Series in Communications, Networking, and Multimedia, 2000, pp. 891.
- [16] HUSKA, J., KULLA, P. Possibilities of motion vectors interpolation in block based motion estimation for H.264 video compression. In *TRANSCOM 2007 7th European Conference of Young Research and Scientific Workers*. Žilina: University of Žilina, 2007, ISBN 978-80-8070-693-7, pp. 93-98.
- [17] DE HAAN, G., LBANIY, L., OLIVIERIY, S. Noise robust recursive motion estimation for H.263 based video conferencing systems. In *Proc. of International Workshop on Multimedia Signal Processing*, Sep. 1999, pp. 345-350.
- [18] HARRIS, C., STEPHENS, M. A combined corner and edge detection. In *Proc. of the 4th Alvey Vision Conf.* 1988, pp. 147-151.
- [19] Weighted average. In *Wikipedia The Free Encyclopedia*, [http://en.wikipedia.org/wiki/Weighted\\_average](http://en.wikipedia.org/wiki/Weighted_average), online, 27<sup>th</sup> February 2007.
- [20] CHEN, Z., ZHOU, P., HE, Y. Fast Motion Estimation for JVT, *JVT of ISO/IEC MPEG & ITU-T VCEG*, JVT-G016, March 2003.
- [21] Video Traces Research Group, YUV video sequences, <http://trace.eas.asu.edu/yuv/index.html>, online 27<sup>th</sup> February 2007.
- [22] RIES, M., CRESPI DE ARRIBA, C., NEMETHOVA, O., RUPP, M. Content based video quality estimation for H.264/AVC video streaming. In *Proceedings of IEEE Wireless and Communications & Networking Conference*, Hong Kong, March 2007.

## About Authors...

**Jozef HUSKA** was born in 1981 in Brezno, Czech and Slovak Republic. In 2005 he graduated (MSc.) from the Dept. of Radio and Electronics, Faculty of Electrical Engineering and Information Technology, Slovak University of Technology, Bratislava. Since then he is a PhD. student at the Department of Radio and Electronics. His scientific research is focused on digital television, image and video processing and compression.

**Peter Kulla** (Fellow of IET) was born in 1949 in Šumiac, Czech and Slovak Republic. He received the MSc. degree in Electrical Engineering from the FEE of Czech TU Prague in 1973 and the Ph.D. degree in Radio and Electronics from FEE of Slovak TU in Bratislava, 1980. From 1984 he is the Associate professor at the Department of Radio and Electronics, Faculty of Electrical Engineering and Information Technology, Slovak University of Technology in Bratislava. His lifelong scientific research and education interests are focused on Television engineering, Applications of image sensors CCD, Digital Circuits, Digital image processing and encoding, Analogue and Digital Television.



<http://www.eucap2009.org>