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 algorithms 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. INTRODUCTION

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 successfully 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 on 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 quantized 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.

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 displacement 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.

Assume \(x_i(m,n)\) to be location of the pixel of the current block in the current frame, and \(x_{i+j}(m+i,n+j)\) 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 described as follows:

- **CrossCorrelation Function**

  The CrossCorrelation Function [11] is derived from the correlation between two random variables. Correlation is the measure of 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 \(\rho\) between two random variables \(U\) and \(V\) is defined as

  \[
  \rho = \frac{\text{Cov}(U,V)}{\text{Var}(U)\text{Var}(V)}
  \tag{1}
  \]

  where
  - \(\text{Cov}(U,V)\) is the covariance between \(U,V\)
  - \(\text{Var}(U)\) and \(\text{Var}(V)\) are the variances of \(U,V\), respectively.

  The CrossCorrelation Function defines the random variable \(U\) as pixel \(x_i\) values in the current block and \(V\) as pixel values \(x_{i+j}\) in the previous block. Thus a sample \(U_{mn}\) is defined as

  \[
  U_{mn} = x_i(m,n)
  \tag{2}
  \]

  and the sample \(V_{mn}\) is defined as

  \[
  V_{mn} = x_{i+j}(m,n)
  \tag{3}
  \]

  where
  - \(m\) ranges from 1 to \(M\)
  - \(n\) ranges from 1 to \(N\).

  The variance of the random variable \(U\) is defined as

  \[
  \text{Var}(U) = E(U^2) - E(U)^2
  \tag{4}
  \]

  where
  - \(E(U)\) 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

  \[
  \text{Cov}(U,V) = E(U \cdot V) - E(U) \cdot E(V)
  \tag{5}
  \]
where

- $E(U \cdot V)$ 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

$$CCF(i,j) = \frac{\sum_{m,n} x(m,n) \cdot x_{m+n}(m+i,n+j)}{\sqrt{\sum_{m,n} x(m,n)^2 \cdot \sum_{m,n} x_{m+n}(m+i,n+j)^2}}$$

If the assumption is made that the expectation of $x_i$ and $x_{m+n}$ are 0, we can obtain Normalized Cross-Correlation Function

$$NCCF(i,j) = \frac{\sum_{m,n} x_i(m,n) \cdot x_{m+n}(m+i,n+j)}{\sqrt{\sum_{m,n} x_i(m,n)^2} \cdot \sqrt{\sum_{m,n} x_{m+n}(m+i,n+j)^2}}$$

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

$$X = \frac{\sum_{i,j} x(i,j)}{MN}$$

$$X_{m+n}(i,j) = \frac{\sum_{i,j} x_{m+n}(m+i,n+j)}{MN}$$

b) The standard deviation has following form

$$S_i = \sqrt{\frac{\sum_{i,j} x_i(m,n)^2 - X_i^2}{MN-1}}$$

$$S_{m+n}(i,j) = \sqrt{\frac{\sum_{i,j} x_{m+n}(m+i,n+j)^2 - X_{m+n}(i,j)^2}{MN-1}}$$

c) The covariance between blocks can be calculated

$$Cov_{x_i, x_{m+n}}(i,j) = \frac{\sum_{m,n} x_i(m,n) \cdot x_{m+n}(m+i,n+j)}{MN} - X_i \cdot X_{m+n}(i,j)$$

d) At last correlation coefficient has following form

$$r(i,j) = \frac{Cov_{x_i, x_{m+n}}(i,j)}{S_i S_{m+n}(i,j)}$$

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

- **Mean square error (MSE)**

$$MSE(i,j) = \frac{1}{MN} \sum_{m,n} [x_i(m,n) - x_{m+n}(m+i,n+j)]^2$$

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)**

$$MAD(i,j) = \frac{1}{MN} \sum_{m,n} |x_i(m,n) - x_{m+n}(m+i,n+j)|$$

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 applied 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 defined as:

$$PCD(i,j) = \sum_{m,n} G(m,n,i,j)$$

where:

$$G(m,n,i,j) = \begin{cases} 1 & \text{if} \quad |x_i(m,n) - x_{m+n}(m+i,n+j)| \leq T \\ 0 & \text{otherwise} \end{cases}$$

and $T$ is a threshold.

- **Minimized Maximum Error (MiniMax)**

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

$$MiniMax(i,j) = \min \{ |x_i(m,n) - x_{m+n}(m+i,n+j)| \}_{m,n}$$

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 distortion 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 evaluated 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 (MPSA)
- New Prediction Search Algorithm (NPSSA)
- 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) \([x_X-x_O]<[x_A-x_O] \text{ and } [y_X-y_O]<[y_A-y_O]\), or

  \([x_X-x_O]=[x_A-x_O] \text{ and } [y_X-y_O]=[y_A-y_O]\)

  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 \(\{x,y\in S \mid x > x_A\}\).

**Property 2:**

If \(x_A > x_B\) and \(y_A > y_B\), then \(O\) cannot exist in the quadrant defined by \(\{x,y\in S \mid x > x_A \text{ and } y > y_A\}\).

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).

\[
PSNR = 10 \log_{10} \left( \frac{255^2}{\sum_{i,j} (P_{raw}(i,j) - C_{ij}(i,j))^2} \right) [dB]
\]

where

- \(x_{ij}\) are values of pixels into the current frame
- \(x_{'ij}\) 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 vector 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.

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 applied as a cost function by maximum displacement \( p=6 \).

### Tab. 4 Comparison od presented block matching motion estimation algorithms

<table>
<thead>
<tr>
<th>Sequences</th>
<th>M</th>
<th>P</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR</td>
<td>SP</td>
<td>PSNR</td>
<td>SP</td>
</tr>
<tr>
<td>FS</td>
<td>37.627</td>
<td>169.00</td>
<td>28.947</td>
</tr>
<tr>
<td>BGDS</td>
<td>37.594</td>
<td>12.06</td>
<td>28.570</td>
</tr>
<tr>
<td>RNS</td>
<td>37.288</td>
<td>10.23</td>
<td>27.469</td>
</tr>
<tr>
<td>BOOMS</td>
<td>37.596</td>
<td>6.71</td>
<td>28.422</td>
</tr>
<tr>
<td>CDS</td>
<td>37.447</td>
<td>8.21</td>
<td>28.347</td>
</tr>
<tr>
<td>CSA</td>
<td>36.673</td>
<td>6.05</td>
<td>26.522</td>
</tr>
<tr>
<td>DIAMOND</td>
<td>36.954</td>
<td>27.09</td>
<td>28.193</td>
</tr>
<tr>
<td>DYNAM</td>
<td>36.491</td>
<td>17.86</td>
<td>27.660</td>
</tr>
<tr>
<td>FSHV</td>
<td>37.328</td>
<td>25.00</td>
<td>28.660</td>
</tr>
<tr>
<td>FSSA</td>
<td>37.286</td>
<td>18.15</td>
<td>28.554</td>
</tr>
<tr>
<td>MME</td>
<td>36.655</td>
<td>7.11</td>
<td>28.004</td>
</tr>
<tr>
<td>NPSA</td>
<td>37.197</td>
<td>15.25</td>
<td>28.548</td>
</tr>
<tr>
<td>NTSS</td>
<td>37.591</td>
<td>18.09</td>
<td>28.729</td>
</tr>
<tr>
<td>ODFS</td>
<td>37.398</td>
<td>36.93</td>
<td>27.856</td>
</tr>
<tr>
<td>OTS</td>
<td>37.311</td>
<td>7.04</td>
<td>27.279</td>
</tr>
<tr>
<td>PHODS</td>
<td>36.566</td>
<td>13.00</td>
<td>26.298</td>
</tr>
<tr>
<td>R</td>
<td>37.593</td>
<td>9.38</td>
<td>28.355</td>
</tr>
<tr>
<td>SES</td>
<td>37.339</td>
<td>17.80</td>
<td>27.999</td>
</tr>
<tr>
<td>SPIRAL</td>
<td>34.776</td>
<td>23.52</td>
<td>26.787</td>
</tr>
<tr>
<td>TDLOG</td>
<td>37.311</td>
<td>14.94</td>
<td>28.556</td>
</tr>
<tr>
<td>TSS</td>
<td>36.783</td>
<td>25.00</td>
<td>27.072</td>
</tr>
</tbody>
</table>

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 chosen 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 minimaums of cost function. Currently location of these local minimaums 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 criterior is the average required number of search 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 search 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


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.