USING DISPERSION MEASURES FOR DETERMINING BLOCK-SIZE IN-MOTION ESTIMATION

USANDO MEDIDAS DE DISPERSIÓN PARA DETERMINAR EL TAMAÑO DEL BLOQUE EN LA ESTIMACIÓN DE MOVIMIENTO

CARLOS MERA
MSc., PhD. Student, Escuela de Sistemas, Facultad de Minas, Universidad Nacional de Colombia, campus Medellín, camerab@unal.edu.co

MARIA TRUJILLO
PhD., Professor, Escuela de Ingeniería de Sistemas y Computación, Universidad del Valle, Cali, maria.trujillo@correounivalle.edu.co

ABSTRACT: Video compression techniques remove temporal redundancy among frames and enable high compression efficiency in coding systems. Reduction of temporal redundancy is achieved by motion compensation. In turn, motion compensation requires motion estimation. Block matching is perhaps the most reliable and robust technique for motion estimation in video coding. However, block matching is computationally expensive. Different approaches have been proposed in order to improve block matching motion estimation accuracy and efficiency. In this paper a block-matching strategy for motion estimation is introduced. In the proposed approach the size of matching block is adapted according to the variability of the matching areas. That is, the block size is constrained by variations of the image intensity. The variability is assessed using two variability measures: the variance and the mean absolute deviation. Results of computer experiments aimed at validating the performance of the proposed approach are also reported.

KEYWORDS: Motion estimation, block matching, variance, mean absolute deviations, variable block size

1. INTRODUCTION

Motion estimation (ME) is the basis for motion compensation and removes temporal redundancies among frames in video coding systems [1,2]. Motion estimation usually counts for up to 70–80% of the computational complexity of a complete encoding system [3,4]. Although several approaches have been presented in the literature to determine the best choice of motion vectors, the block-matching algorithm (BMA) is the most popular. The BMA divides an image frame into blocks of size N x N and matches each block in the reference frame to the most similar block within a search area in the next frame assuming that all the pixels within a block have uniform translational motion. However, the use of fixed-size blocks has some drawbacks; when small block size is used, it may produce problems of ambiguity; and on the other hand, a large block size may cause inaccuracy problems. In addition, the boundaries of moving objects do not normally coincide with the boundaries of the blocks used for motion estimation [4,5].

Unlike the BMA, variable block-size (VBS) motion estimation is a method where the block size is varied according to the types of motion in the block. Accuracy in motion estimation can be improved by coding blocks of high motion detail, using a small block size. Similarly, blocks with less motion detail can be encoded using larger block sizes [6].

Received for review July 20th, 2011, accepted December 2nd, 2011, final version December 21st, 2011
Only a few approaches adaptively change the size of the matching block. A pioneering work in this direction was introduced by Levine et al. [7] in the early 1970s. In this work, the block size is adapted according to variations in intensity values. Vaisey and Gersho [8] discussed techniques in which the size of the block is varied according to the local detail of the image using quad-tree implementation. Puri et al. [9] presented a decision rule for determining the block size based on the comparison of motion compensated prediction error between two different block sizes. Oh and Lee [10] used a top-down partition within an image segment in order to establish a homogenous block. However, the homogeneity criterion is not included in the discussion. In [5] and [11], a VLSI architecture for VBS, and a VLSI processor with parallel architecture based on an adaptive block size are proposed, respectively. Veksler [12] proposed a variable block algorithm using integral images and defining a block cost, based on average error and variance, which explores a range of block shapes and sizes. In [4] the authors proposed an algorithm for adaptive selection of block sizes based on measuring texture areas as a criterion to stop the reduction of the block size for visually irrelevant areas.

The computer-vision community has also addressed the problem of block size. Kanade and Okutomi [13] proposed an iterative algorithm that controls the size and the shape of the block using a measure of the uncertainty in the motion/disparity estimation. Tzovaras et al. considered using two multiresolution block-matching methods where the size of the block varies with the resolution level of the frame [14]. Fusiello et al. presented a multi-block algorithm using nine asymmetric matching blocks, where the motion/disparity profile is used to select the appropriate size [15]. Ohm et al. [16] as well as Izquierdo [17] proposed a hierarchical block matching algorithm which uses large size blocks in the first level and small size blocks in the second level for refining the initial estimation. Although they proposed to use different block sizes, they are fixed heedless of the frame/image content. For continuous optimisation, Goulermas et al. [18] introduced a new regularisation to enforce the inter-block dependency. Odone, Trucco, and Verri [19] presented a method using matching blocks, based on dilation along the grey values and the spatial dimension, taking into account variations in grey levels and localization. In [20], a modified semivariogram function was introduced, which takes into account spatial variations for constraining the block size, in a feature-based matching strategy.

The H.264 standard [21,22] allows variable block sizes to be used in motion estimation. In H.264, a frame is first divided using macro-blocks (MB) of size 16x16. Each MB may then be segmented into micro-blocks (mB) of block sizes 8x16, 16x8, 8x8, 8x4, 4x8, and 4x4. These smaller blocks intend to describe complex motion and movements next to objects edges. Although this standard gives the ability to dynamically choose what block size will be used, the block size is predetermined and fixed.

In this paper, an approach for adaptively selecting the block size is proposed. The approach exploits intensity variation using relevant statistical measures. Intensity variations within macro and micro-blocks are compared in order to decide whether to perform a new partition. Experimental evaluations showed better estimations of motion vectors by using an appropriate block size.

This paper is organized as follows: Section 2 gives a brief discussion of the block-matching motion estimation and commonly used variability measures. The block-size estimation criterion is introduced in Section 3. Experimental evaluations and final remarks are presented in Sections 4 and 5, respectively.

2. BACKGROUND

Let \( D_1 \) and \( D_2 \) be adjacent video frames at \( t \) and \( t + 1 \). Block-matching is described as follows and illustrated.
in Fig. 1. Given a point \( u = (x_1, y_1) \), a reference block of size \( m \times n \) is positioned on \( D_1 \) with \( u \) at the left-up corner. A search area of size \( (2p + 1) \times (2p + 1) \) is defined on \( D_2 \) with \( u \) at its centre. A matching operation is performed between the reference block in \( D_1 \) and all matching blocks lying within the search area in \( D_2 \). This matching is based on a search for the most similar matching block within the search area in \( D_2 \) using a similarity measure, for instance mean absolute differences (MAD) or mean squared error (MSE) [23].

Estimated displacement vectors are calculated by the difference between the left-up point at the reference block in \( D_1 \), \( u = (x_1, y_1) \), and the left-up point at the most similar matched block within the search area in \( D_2 \), \( u' = (x_2, y_2) \). That is:

\[
d = (x_2 - x_1, y_2 - y_1)
\]

As mentioned before, the matching block size depends on the spatial variability in the surrounding areas. When there is large variability in the surrounding areas, a small matching block is required. On the other hand, when there is small variability in the surrounding areas, a large matching block is needed. Assessing variations in the surrounding areas can be done by means of a variability measure. A large number of variability measures have been explored in the literature [24]. The most frequently used are the range, the interquartile range, the variance, the coefficient of variation, and the standard deviation.

We consider most relevant statistical measures for establishing a homogeneous matching block based on deviations of intensity values from the mean and from the median within a block. The variance is calculated as the average of the squared deviations from the mean, using Norm-2. The absolute deviation is calculated as the average of the absolute deviations from the median, using Norm-1. Higher-order statistics are not considered since they are used for assessing shape rather than variations.

The following notation is used: Let \( I(x,y) \) be an intensity value of a pixel \( (x, y) \) lying within a block of size \( m \times n \). \( I \) and \( Me \) represent, respectively, the mean and median of the intensity in such a block.

The variance can be seen as a measurement of global variations around the mean. These variations can be measured in a region regardless of the spatial location of the intensity value in the frame. The use of the variance assumes outcomes to be independent.

\[
Var = \frac{1}{mn} \left[ \sum_{x=1}^{m} \sum_{y=1}^{n} (I(x,y) - \bar{I})^2 \right] - \bar{I}^2
\]

The coefficient of variation (CV) is a normalized measure of dispersion [25] useful to compare dispersions at different scales.

\[
CV = \frac{\sigma}{\bar{I}} = \sqrt{Var} / \bar{I}
\]

The mean absolute deviation can be seen as a measurement of global variations around the median. These variations can be measured in a region regardless of the location of the frame intensity. The use of the mean absolute deviation assumes outcomes to be independent and it assures robustness against noise.

\[
MAD = \frac{1}{mn} \sum_{x=1}^{m} \sum_{y=1}^{n} |I(x,y) - Me|
\]

Table 1. Values of Var, CV, and MAD calculated using the blocks of Fig. 2

<table>
<thead>
<tr>
<th>Block/Measure</th>
<th>Blue Block</th>
<th>Green Block</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var</td>
<td>3148.01</td>
<td>54.33</td>
</tr>
<tr>
<td>CV</td>
<td>0.44</td>
<td>0.11</td>
</tr>
<tr>
<td>MAD</td>
<td>43.82</td>
<td>5.88</td>
</tr>
</tbody>
</table>
Selected blocks and their intensity values, in Fig. 2, are used to illustrate how the measures can be used to assess homogeneity in a block. The green block has homogeneous intensity values, and the blue one does not have them. Table 1 shows the variance, coefficient of variation and mean absolute deviation, computed for the blocks of Fig. 2. It can be observed that small values of these measures indicate small variations in intensity values. Therefore, we can detect homogeneous blocks by comparing the values obtained in the two blocks.

3. THE BLOCK-SIZE ESTIMATION

Homogeneous blocks can be detected by comparing the intensity variation in a large area with the intensity variation in small areas lying within the large area. Large areas are called macro-blocks (MB) and small areas are called micro-blocks (mB). The homogeneity of an area is assessed using a variability measure such as: the variance, the coefficient of variation, and the mean absolute deviation.

At each node a test is performed to evaluate if the mB represented by the node is homogeneous by comparing the intensity variation of that mB and its parent. If the test is positive the node becomes a leaf, otherwise the node becomes a MB and it is divided. The partition strategy is described in detailed as follows using the subsequent notation:

- \( \text{Var}(MB_k) \) is the value of the variance in the \( k \)-th MB.
- \( \text{Var}(mB_{ki}) \) is the value of the variance in the \( i \)-th mB within the \( k \)-th MB.
- \( \text{CV}(MB_k) \) is the value of the coefficient of variation in the \( k \)-th MB.
- \( \text{CV}(mB_{ki}) \) is the value of the coefficient of variation in the \( i \)-th mB within the \( k \)-th MB.
- \( \text{MAD}(MB_k) \) is the value of the mean absolute deviation in the \( k \)-th MB.
- \( \text{MAD}(mB_{ki}) \) is the value of the mean absolute deviation in the \( i \)-th mB within the \( k \)-th MB.

Initially, a median filter is used to homogenise the content in the frame and reduce the effect of noise and textures. Then, as it is illustrated in Fig. 4, the whole frame content is divided into four MBs (MB\(_k\)). Then, each MB is divided into four mBs (mB\(_{ki}\)). Homogeneity in an mB is assessed by variability measures in order to decide whether or not to subdivide each mB (mB\(_{ki}\)). In this way, we use \( \omega \), whose value is equal to one (\( \omega = 1 \)) when the \( i \)-th mB, within the \( k \)-th MB (mB\(_{ki}\)), is not homogeneous and it must be subdivided into new four mBs.

The \( \omega \) criterion is computed as follows:

\[
\omega = \begin{cases} 
1 & \text{if } CV(MB_k) \geq \tau \text{ and } \Omega = 1 \\
0 & \text{in other case}
\end{cases}
\]  
\( (5) \)

where

\[
\Omega = \begin{cases} 
1 & \text{if } A = 1 \text{ and } B = 1 \text{ and } C = 1 \\
0 & \text{in other case}
\end{cases}
\]  
\( (6) \)

\[
A = \begin{cases} 
1 & \text{if } CV(mB_{ki}) \geq \varphi \\
0 & \text{in other case}
\end{cases}
\]  
\( (7) \)

\[
B = \begin{cases} 
1 & \text{if } \frac{\text{Var}(mB_{ki})}{\text{Var}(MB_k)} \geq \kappa \\
0 & \text{in other case}
\end{cases}
\]  
\( (8) \)

\[
C = \begin{cases} 
1 & \text{if } \frac{\text{MAD}(mB_{ki})}{\text{MAD}(MB_k)} \geq \delta \\
0 & \text{in other case}
\end{cases}
\]  
\( (9) \)
Figure 4 illustrates the blocks obtained in the first frame of table tennis video sequence, when the proposed criterion is used. Figure 5 depicts the proposed algorithm.

Table 2 shows the values obtained using the selected measures in MBs and mBs of the first frame of the table tennis video sequence (step 1). It can be observed that the variability in mB\textsubscript{14} is greater than the variability in MB\textsubscript{1}, which suggests that mB\textsubscript{14} should be subdivided (here Eq. 7 is used). Human visual perception, in Fig. 4, shows that mB\textsubscript{12} must be subdivided too. If the values of mB\textsubscript{14} are compared with the values of MB\textsubscript{1}, it can be seen that these values are close.

The values in Table 2 suggest that when the CV value of the MB is greater than 0.1, mBs may be subdivided (Eq. 5). A similar situation happens with the CV values of the mBs (Eq. 7). Human visual perception indicates that an mB must be subdivided when its variance is greater than or close to the variance of MB or when the MAD for an mB is greater than the MAD of an MB (Eqs. 8 and 9, respectively).

Table 2. Numerical Information about MB and mB in the first step using the first frame of table tennis video sequence

<table>
<thead>
<tr>
<th>Block</th>
<th>Var</th>
<th>CV</th>
<th>MAD</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>MB\textsubscript{1}</td>
<td>146.17</td>
<td>0.09146</td>
<td>6.4789</td>
<td>0.194</td>
<td>0.423</td>
</tr>
<tr>
<td>mB\textsubscript{11}</td>
<td>32.447</td>
<td>0.04379</td>
<td>4.461</td>
<td>0.223</td>
<td>0.689</td>
</tr>
<tr>
<td>mB\textsubscript{12}</td>
<td>126.27</td>
<td>0.08222</td>
<td>5.7985</td>
<td>0.863</td>
<td>0.894</td>
</tr>
<tr>
<td>mB\textsubscript{13}</td>
<td>37.212</td>
<td>0.04808</td>
<td>4.684</td>
<td>0.254</td>
<td>0.722</td>
</tr>
<tr>
<td>mB\textsubscript{14}</td>
<td>327.23</td>
<td>0.13385</td>
<td>8.1936</td>
<td>2.228</td>
<td>1.264</td>
</tr>
<tr>
<td>MB\textsubscript{1}</td>
<td>840.40</td>
<td>0.2451</td>
<td>18.899</td>
<td>1.112</td>
<td>1.233</td>
</tr>
<tr>
<td>mB\textsubscript{21}</td>
<td>35.480</td>
<td>0.04428</td>
<td>4.6735</td>
<td>0.042</td>
<td>0.247</td>
</tr>
<tr>
<td>mB\textsubscript{22}</td>
<td>846.35</td>
<td>0.25829</td>
<td>21.400</td>
<td>1.007</td>
<td>1.132</td>
</tr>
<tr>
<td>mB\textsubscript{23}</td>
<td>97.89</td>
<td>0.07376</td>
<td>5.6581</td>
<td>0.116</td>
<td>0.299</td>
</tr>
<tr>
<td>mB\textsubscript{24}</td>
<td>1136.5</td>
<td>0.36699</td>
<td>29.214</td>
<td>1.352</td>
<td>1.545</td>
</tr>
<tr>
<td>MB\textsubscript{1}</td>
<td>388.90</td>
<td>0.15807</td>
<td>8.9694</td>
<td>0.514</td>
<td>0.585</td>
</tr>
<tr>
<td>mB\textsubscript{31}</td>
<td>32.832</td>
<td>0.04639</td>
<td>10.458</td>
<td>0.084</td>
<td>0.488</td>
</tr>
<tr>
<td>mB\textsubscript{32}</td>
<td>303.59</td>
<td>0.14371</td>
<td>9.9140</td>
<td>0.780</td>
<td>1.165</td>
</tr>
<tr>
<td>mB\textsubscript{33}</td>
<td>599.81</td>
<td>0.19484</td>
<td>10.398</td>
<td>1.542</td>
<td>1.105</td>
</tr>
<tr>
<td>mB\textsubscript{34}</td>
<td>590.05</td>
<td>0.18893</td>
<td>24.628</td>
<td>1.517</td>
<td>1.159</td>
</tr>
<tr>
<td>MB\textsubscript{1}</td>
<td>1548.1</td>
<td>0.31672</td>
<td>24.628</td>
<td>2.049</td>
<td>1.607</td>
</tr>
<tr>
<td>mB\textsubscript{41}</td>
<td>2008.6</td>
<td>0.36655</td>
<td>31.692</td>
<td>1.297</td>
<td>1.286</td>
</tr>
<tr>
<td>mB\textsubscript{42}</td>
<td>1517.85</td>
<td>0.3720</td>
<td>30.11</td>
<td>0.980</td>
<td>1.222</td>
</tr>
<tr>
<td>mB\textsubscript{43}</td>
<td>613.052</td>
<td>0.1870</td>
<td>13.019</td>
<td>0.396</td>
<td>0.528</td>
</tr>
<tr>
<td>mB\textsubscript{44}</td>
<td>1425.41</td>
<td>0.2746</td>
<td>22.505</td>
<td>0.920</td>
<td>0.913</td>
</tr>
</tbody>
</table>

The $\tau$, $\varphi$, $\kappa$, and $\delta$ values were tuned as follows, according to the selected variability measure:

$\tau = 0.055$ \hspace{1cm} $\kappa = 0.70$

$\varphi = 0.11$ \hspace{1cm} $\delta = 0.70$

4. EXPERIMENTAL VALIDATION

The performance of the proposed method has been verified by processing various video sequences using the full search method in which the motion vector is searched at all the locations within the search window. These video sequences are available at http://media.xiph.org/video/derf/ and have different motion activity and different content. Table 3 describe the video sequences tests.
The blocks are obtained from the first frame and these are used for estimating motion in the shot. Results are reported using the PNSR values, time in milliseconds, and the number of blocks obtained.

The proposed algorithm efficiently partitioned the frames. The block size is small both in the surroundings of object’s edges and in highly textured areas. In contrast, large block sizes are obtained in remaining areas. (This can be observed on the images in Table 4.)
Table 6 shows the number of blocks when fixed block sizes (4x4, 8x8, and 16x16) and variable block-size algorithms are used. The number of blocks is reduced compared to when 4x4 and 8x8 blocks are used. However, the number of blocks for the proposed technique is larger than the number of blocks when 16x16 blocks are used. This is because the proposed algorithm reduces the block size in frames with high textures, like Coastguard, Garden, Mobile, Silent, and Stefan.

### Table 6. Total number of blocks for BM

<table>
<thead>
<tr>
<th></th>
<th>4x4</th>
<th>8x8</th>
<th>16x16</th>
<th>VBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akiyo</td>
<td>6336</td>
<td>1584</td>
<td>396</td>
<td>589</td>
</tr>
<tr>
<td>Coastguard</td>
<td>6336</td>
<td>1584</td>
<td>396</td>
<td>580</td>
</tr>
<tr>
<td>Foreman</td>
<td>6336</td>
<td>1584</td>
<td>396</td>
<td>679</td>
</tr>
<tr>
<td>Garden</td>
<td>5280</td>
<td>1320</td>
<td>330</td>
<td>748</td>
</tr>
<tr>
<td>Mobile</td>
<td>5280</td>
<td>1320</td>
<td>330</td>
<td>877</td>
</tr>
<tr>
<td>Mother</td>
<td>5280</td>
<td>1320</td>
<td>330</td>
<td>562</td>
</tr>
<tr>
<td>Silent</td>
<td>6336</td>
<td>1584</td>
<td>396</td>
<td>676</td>
</tr>
<tr>
<td>Stefan</td>
<td>5280</td>
<td>1320</td>
<td>330</td>
<td>736</td>
</tr>
<tr>
<td>Tennis</td>
<td>5280</td>
<td>1320</td>
<td>330</td>
<td>325</td>
</tr>
</tbody>
</table>

Table 7 shows average computation time—in milliseconds—obtained for the test sequences using the full search algorithm with fixed block sizes (4x4, 8x8, and 16x16) and the proposed technique include the time used for dividing the frame. Results show that in most of the test sequences, the computation of motion vectors is faster.

### Table 7. Average time in milliseconds

<table>
<thead>
<tr>
<th></th>
<th>4x4</th>
<th>8x8</th>
<th>16x16</th>
<th>VBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akiyo</td>
<td>2637</td>
<td>2558</td>
<td>2543</td>
<td>2513</td>
</tr>
<tr>
<td>Coastguard</td>
<td>2675</td>
<td>2599</td>
<td>2551</td>
<td>2551</td>
</tr>
<tr>
<td>Foreman</td>
<td>2614</td>
<td>2542</td>
<td>2511</td>
<td>2521</td>
</tr>
<tr>
<td>Garden</td>
<td>2214</td>
<td>2142</td>
<td>2116</td>
<td>2556</td>
</tr>
<tr>
<td>Mobile</td>
<td>2682</td>
<td>2586</td>
<td>2549</td>
<td>2564</td>
</tr>
<tr>
<td>Mother</td>
<td>2673</td>
<td>2569</td>
<td>2533</td>
<td>2521</td>
</tr>
<tr>
<td>Silent</td>
<td>2643</td>
<td>2563</td>
<td>2525</td>
<td>2523</td>
</tr>
<tr>
<td>Stefan</td>
<td>2214</td>
<td>2157</td>
<td>2129</td>
<td>2148</td>
</tr>
<tr>
<td>Tennis</td>
<td>2243</td>
<td>2164</td>
<td>2135</td>
<td>2117</td>
</tr>
</tbody>
</table>

5. CONCLUSION

We consider statistic measures for establishing homogeneous matching blocks. Intensity variations were assessed by the coefficient of variation, the variance, and the mean absolute deviation. These measures rely on the assumption that intensity values are independent.

The results show that the proposed method avoids unnecessary reduction in the block size in homogeneous areas and reduces the block size in areas that contain complex motions. Subsequently, the number of blocks is reduced compared with the use of 4x4 and 8x8 blocks, and hence the number of bits for encoding all kinds of video sequences is decreased.

The average PSNR does not show significant differences between the reconstruction obtained using prefixed block size and the reconstruction obtained using variable block size. However, a reduction of the computational cost of block matching algorithm using larger block size may justify implementing the proposed approach.

ACKNOWLEDGMENTS

The first author gratefully acknowledges the support of the Universidad de San Buenaventura Seccional Cali. Most of the writing of this paper was done while he was professor at Programa de Ingeniería de Sistemas in that university.

REFERENCES


