Object Oriented Motion Estimation by Sliced-Block Matching Algorithm

Cataldo Guaragnella and Eugenio Di Sciascio
Dipartimento di Elettrotecnica e Elettronica - Politecnico di Bari
{guaragnella, disciascio}@poliba.it

Abstract
In this paper a novel algorithm for object-oriented motion estimation is presented. The algorithm initially determines a macro-block partition on the basis of the computed current frame difference, using the hidden information that a foreground moving object produces high absolute frame difference values in the neighborhood of the object boundaries. An inter-frame coding algorithm, adopting a modified version of classical block matching is then applied on separate slices of each macro-block. Resulting data are used to obtain a preliminary object segmentation. The algorithm further splits each macro-block if characterized by the presence of more than one motion vector into sub-areas. The approach allows to obtain a global object segmentation - with the possibility of tracking - and lower prediction errors with respect to classical Block Matching, without increasing the computational complexity.

1. Introduction
The interest in image sequence analysis for MPEG-based video coding ([1-2]) aims to provide standardized core technologies allowing efficient storage, transmission and manipulation of video data in multimedia environments. The newly defined MPEG-4 [4, 5, 6] aims at providing a standard to cope with the requirements of current and future multimedia applications. Diffrently from previous techniques [1,2,7,8], MPEG-4 considers a scene to be a composition of Video Objects (VO) with intrinsic properties such as shape, motion, and texture. In fact, the video Object Oriented MPEG standard [3], which is based on motion estimation and compensation on arbitrary shaped macro-blocks, depends on the actual motion field. Such a content based representation is key for interactivity with objects for a variety of multimedia applications.

In this contribution a modified block matching (BM) algorithm is proposed, which obtains the subdivision of each macro-block into non-overlapping areas with the aid of the available information on difference between couple of frames, allowing a better object boundary definition and lower prediction errors. The rationale of the algorithm is to detect zones of the current image where the frame difference (FD) is high. In such areas the presence of a sharp moving contour is assumed, so that each block can be further divided into subsets defined by thresholding the FD. Each of these subsets is considered candidate for the sub-block segmentation and its motion is estimated. In doing that, the sign of the FD must be kept into account. Considering the sign of the FD allows to properly take care of the different contributions to the motion estimate due to the uncovered background appeared as a consequence of the foreground object motion.

The thresholding procedure cannot be simply applied on the absolute frame difference, because this segmentation can be misleading in determining the prediction: the FD sign depends both on the luminance level of the object and the background, and on the object motion. Where the frame difference assumes different signs, different pixel areas must be considered in order to take care properly of uncovered background areas appeared as a consequence of the foreground object motion.

An arbitrarily shaped macro-block segmentation results, allowing the definition of moving objects borders, thus reducing the Displaces FD (DFD) of standard BM algorithms, depending on the motion present in the scene, without increasing the computational burden.

The remaining of this contribution is so structured: in section 2 a few considerations on the behavior the frame difference of a moving object with sharp contours on a fixed background are proposed to highlight the possibility of extracting object borders and their motion in a scene sequence. In section 3 the proposed sliced BM algorithm for motion estimation is presented, showing the capability of the algorithm to cope with object boundaries. Numerical results of tests carried out on standard sequences, discussion and future work are presented in section 4.

2. Frame difference and moving object boundaries
The motion field in video sequences is usually computed by means of BM based techniques. In such techniques the image to be predicted - at the coder end - is divided into blocks of fixed size, usually squares of 16 x 16 pixels. Each block of this image is compared with a displaced block on the preceding frame, and a measure of the approximation of the block is calculated, usually as Mean Absolute Error (MAE) or Mean Square Error (MSE). The displaced block on the previous image having the least prediction error is evaluated, and its displacement from the actual block size is chosen as the motion vector. The
simplicity and good performances of BM based motion estimation algorithms have favored its use in most coding standards. This simple algorithm, however, suffers from several drawbacks, such as blocking artifacts appearing on predicted images at high compression ratios. In order to introduce the proposed Sliced-Block-Matching algorithm, let us consider a main object moving on a fixed background, in a given video sequence, and suppose the presence of high color contrast between object and background. Computation of the FD between two successive frames will show two peaks in the neighborhood of the moving object borders. Blocks of the frame to be predicted containing pixels of the background appeared after the motion has occurred cannot be well predicted, normally, by any of the displaced blocks of the preceding frame using standard BM.

![Figure 1: 1D frame difference due to the motion of a foreground object on a fixed background.](image)

In the successive frame, the target of the motion estimation procedure, the part of the image corresponding to the negative (in this case) frame difference, will not be predictable by any motion vector, as pixels of that area belong to the background region uncovered by the motion object, as sketched in Figure 1. Based on this observation, a new strategy of block motion estimation can be adopted, to reduce the unwanted block motion estimation behavior.

If, for example, a rectangular object is moving on an arbitrary background with a given motion vector, as it is exemplified in Figure 2, the frame difference will show two high-differences frame regions around moving object borders.

![Figure 2: Moving object on a fixed background with highlighting of the regions of strong frame difference.](image)

In the leftmost region contains, in the successive frame $\text{IM}_{t+1}(i,j)$, a portion of the object whose prediction will not be possible with forward motion estimation BM procedure, while the rightmost area is predictable by BM, though compensation results may vary with the relative positions of the block to be predicted and the object border position. The distinction between positive and negative frame difference areas can allow to separate motion prediction of objects boundaries.

The presence of high frame difference inside a given block can allow the segmentation of block into disjoint areas, based on the sign of the FD. Each one can be predicted regardless of the other inside a block, thus avoiding background pixel misleading contribution to the estimation of the motion for a boundary block of the moving object; this feature will increase the performances of the motion estimation algorithm, allowing lower prediction errors.

The sub-block splitting, together with lower errors, can allow a first contour-based macro-block segmentation, and can represent a first step toward an object oriented MPEG-4 compliant motion estimation technique. If proper image segmentation takes place, uncovered background areas can be detected and coded separately. In some applications such as video-conference, the background is known and such areas can be left not coded at all.

3. Sliced BM algorithm

The problem here addressed is the detection of blocks where one motion vector is not enough to model the local motion field and to predict a subsequent block by conventional motion compensation algorithm. The presence of uncovered background due to a foreground moving object asks for a contour based macro-block segmentation into non overlapping areas to be coded separately. For such blocks we determine a sub-block partition, using the knowledge of the frame difference values between contiguous frames.

Frame difference is normally computed as:

$$FD(i,j) = \text{IM}_{t+1}(i,j) - \text{IM}(i,j)$$

where $(i,j)$ represents the pixel location on the image, and the time distance between frames is assumed normalized to 1 s.

Equation (1) represents a discrete time-time domain approximation of a derivative; regions showing peaks in $FD(i,j)$ should be linked to the time contour positions, so that such regions can be considered as evidence of the presence of moving objects.

Selection of areas can be obtained with the aid of a threshold, separating positive FD high energy zones from negative ones and isolating regions with low FD. Assuming $Th$ as the selected threshold value, three types of regions can be defined, as:

$$P = \{ (i,j) | FD(i,j) > Th, 0 \leq i \leq N, 0 \leq j \leq M \}$$

$$N = \{ (i,j) | FD(i,j) < -Th, 0 \leq i \leq N, 0 \leq j \leq M \}$$

$$Z = \{ (i,j) | FD(i,j) < Th, 0 \leq i \leq N, 0 \leq j \leq M \}$$

with $N$ and $M$ the horizontal and vertical size of the frame, respectively.
BM algorithm can then be separately applied on slices belonging to the three partitions, allowing to evaluate separately the motion vectors of each defined area inside macro-blocks. This implies the same total computational load, as the three (at most) areas in each macro-block do not overlap.

As long as the motion is not negligible, the gain in prediction error increases.
If such conditions are met, high values of the frame difference are a clear sign that a contour has appeared/disappeared in a block, due to an object in motion through the sequence.

3.1. Algorithm implementation steps
Though the target of the whole work is the joint definition of a suitable segmentation of the moving object together with a compact description of its motion field, at this implementation stage, the algorithm performs only motion estimation on arbitrary shaped macro-blocks.
 Its main steps are, synthetically:
- computation of frame difference;
- threshold application to FD to select the high energy frame difference zones of the image and block labeling;
- Sliced-BM motion estimation.

4. Experimental results and future work
The method proposed here has been applied to several standard image sequences. Here results are reported for the “Akiyo”, “Carphone”, “Foreman”, “Mother and Daughter” and “Salesman” test sequences, represented in the QCIF format of 144x176 pixels, luminance only.
In order to assess the capability of the method to deal with sub-block area motions, obtaining better results in terms of displaced frame difference (DFD) compared with BMA, couple of successive frames have been used.
Block size is 16x16 pixels. A fixed threshold value (Th=5) has been used for all frames of each processed sequence. The value has been determined, after experiments on several sequences, as the one that on average performs better, see for an example fig. 5.
It should be noticed that in these experiments the unpredictable areas due to uncovered background regions were not considered separately, but predicted from the actual frame, so that their contribute worsens the possible algorithm performance. For visual inspection purposes figure 3 reports the raw segmentation obtained by thresholding the frame difference for four of the used test sequences.

Table 1: Pixel Mean Absolute Error gain (dB)
Used frame rates are 10 frames/s for Akiyo and Foreman, 30 frames/s for the other sequences.

<table>
<thead>
<tr>
<th>Fr no.</th>
<th>Akiyo</th>
<th>Carphone</th>
<th>Foreman</th>
<th>Mother</th>
<th>Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2.04</td>
<td>1.79</td>
<td>1.41</td>
<td>0.01</td>
<td>0.10</td>
</tr>
<tr>
<td>3</td>
<td>1.51</td>
<td>1.63</td>
<td>0.77</td>
<td>0.07</td>
<td>0.31</td>
</tr>
<tr>
<td>4</td>
<td>1.83</td>
<td>1.78</td>
<td>0.67</td>
<td>0.06</td>
<td>0.47</td>
</tr>
<tr>
<td>5</td>
<td>0.31</td>
<td>1.65</td>
<td>0.78</td>
<td>0.06</td>
<td>0.89</td>
</tr>
<tr>
<td>6</td>
<td>1.04</td>
<td>1.43</td>
<td>0.59</td>
<td>0.08</td>
<td>0.65</td>
</tr>
<tr>
<td>7</td>
<td>0.99</td>
<td>1.50</td>
<td>0.67</td>
<td>0.11</td>
<td>1.26</td>
</tr>
<tr>
<td>8</td>
<td>0.52</td>
<td>1.20</td>
<td>0.72</td>
<td>0.24</td>
<td>1.89</td>
</tr>
<tr>
<td>9</td>
<td>0.74</td>
<td>0.97</td>
<td>0.73</td>
<td>0.25</td>
<td>2.04</td>
</tr>
</tbody>
</table>

859
With reference to numerical results, table 1 shows comparative results for BMA and the proposed Sliced-BMA for the first nine frames of the test sequences in terms of prediction error gain, defined as the ratio between the BMA prediction error and the proposed algorithm one, expressed in dBs. Figure 4 diagrammatically shows the gain obtained with the proposed method applied on 30 frames of the sequences. As a general consideration on all experiments carried out, results indicate that a fair improvement in prediction error can be obtained with respect to standard Block Matching, without increasing the computational burden. The greater the motion occurring, the larger the proposed method gain appears on every tested sequence, allowing a lower frame rate for very low bit rate coding purposes.

The main advantage of the proposed motion estimation procedure, at this stage, lies in its intrinsic segmentation oriented structure, able to cope with moving objects boundaries, as experimental results show. Future work will focus on the possibility of grouping macro-blocks belonging to the same moving object thus obtaining the moving object contour together with a compact parametric description of the motion fields it is experiencing [10]. The possibility of isolating background regions uncovered by the foreground object in motion is a topic that will be stressed, too.

References

![Sliced BMA](image)

**Figure 5:** Pixel MAE for Foreman sequence as a function of the chosen threshold level. The curves refer to frames 2 to 7.