FAST MOTION ESTIMATION BASED ON MOTION CLASSIFICATION AND DIRECTIONAL SEARCH PATTERNS

1. OVERVIEW

In real world video sequences the motion field is very structured and there exists a strong correlation among neighboring blocks in spatial and temporal domain. This suggests that significant computational gain can be achieved by taking advantage of strong spatio-temporal dependencies that exist between the motion vectors. The main advantage of prediction based algorithms is that it can help alleviate the local minimum problem to some extent. As the new initial/predicted search center is usually closer to the global minimum candidate, the chances of getting trapped in local minimum decreases. It is possible that fast prediction based initial search may generate more sensible initial motion vectors that work better for the next stage of refinement search. Also motion vectors estimated in prediction algorithm are more realistic since it reflects the real physical phenomenon i.e. real motion, where as FS just finds the minimum SAD which does not necessarily give the realistic motion vectors (and is sometimes a bad choice) because of the complexity and randomness of the real world imagery.

Since moving objects and image features are different in different video sequences, degradation is always accompanied with a reduction in computational complexity if prior information about the video sequence is not taken into account. Some valuable information such as direction and content (slow, fast, smooth, irregular) can help to define the search approach. Here we present a predictive motion estimation technique employing both spatial and temporal correlation. The proposed algorithm is a multistage approach that predicts the initial motion vector, analyses motion and classifies it. The analysis stage helps the search technique to adapt to the motion
characteristics and classifies it into various categories that help to control the search process by avoiding search stationary regions and keeping track of slow, medium, fast, smooth and complex motions, by using switchable search patterns. Early termination criteria further accelerates the search process. We have evaluated the algorithms through a comprehensive performance study that shows that the proposed algorithms achieve substantial speedup without quality loss for a wide range of video sequences.

In this method [1] the spatio-temporal information available between the video frames is effectively exploited using median prediction method. By accurately predicting the location of the best MV candidate the final predicted MV can be located in a relatively small area. The reduced motion search area provides additional compression since the overhead information of MV is less. This is followed by identifying the direction and content of motion based on the direction and magnitude of the predicted motion vector. Final search pattern is adaptively chosen based on the motion classification. Experimental results show a good speedup with PSNR values close to FS algorithm.

2. MOTION VECTOR PREDICTION

The MV prediction is based on spatio temporal correlation information. Only horizontal and vertical neighboring blocks are used for prediction, as shown in Figure 1. The first step of our algorithm roughly searches five neighboring MVs for an initial search point [2]. The motion vectors, \( \text{MV}_{SL}(n) \) (Spatial Left), \( \text{MV}_{SA}(n) \) (Spatial Above), \( \text{MV}_{TR}(n-1) \) (Temporal Right), \( \text{MV}_{TB}(n-1) \) (Temporal Below) and \( \text{MV}_{P}(n-1) \) (Previous) perform as candidates of the predicted motion vector \( P(X_{n}) \). Median predictor has the ability to make a non-center biased distribution center biased. Median predictor is also a good choice when taking into consideration complexity, performance and robustness. Therefore we will calculate the predicted MV by median prediction
method. The x and y components of the median predicted MV are computed independently. Once the predicted MV is obtained the first step of the algorithm is to move the initial search center to the predicted location.

![Diagram of search pattern](image)

**Figure 1.** **Neighboring blocks for spatio-temporal correlation information.**

In fast ME algorithms, the search strategy, shape and size of search pattern plays an important role in search speed and performance of the algorithm. Block distortion forms an error surface over the search window and global minimum point corresponds to the point where the best matching occurs i.e. the minimum distortion point. Based on the MV distribution probabilities [3] of real world video sequences using FS algorithm, it is observed that about 52.76% to 98.70% of MVs are enclosed in a circular support within a radius of 2 pels centered on the zero motion position. And above 93%, 95.99 % (for Table Tennis, that is considered a fast moving sequence) of MVs are enclosed in a circular support with a radius of 7 pels centered on the zero motion position.
position. Based on this information we can assume that for small and medium motion blocks only few steps of search pattern can end up in finding the true MV. Same is also true for large motion however it is possible in some cases that the algorithm is unable to locate the true MV or the minimum distortion point when it lies outside the boundary of the search window. If the minimum distortion point lies outside the boundary then the algorithm will converge to a point on the boundary as close as possible to the minimum distortion point.

3. DIRECTION OF MOTION

The prediction information is used to predict the direction of motion of the current block. This is based on the predicted (initial) motion vector \( P(X_n) \) and the MV of the previous block (i.e. the MV of the same block in the previous frame) \( MV_{n-1}(P) \). The direction information helps to check only some points along that direction. The search space is partitioned into eight sectors, as shown in Figure 2.

![Figure 2. Configuration of sectors w.r.t. different possible directions of \( P(X_n) \).](image)

4. MOTION CONTENT

The magnitude of predicted motion vector is used to define the motion content of the blocks. The search range for our experiments is chosen to be \( \pm 7 \) and \( \pm 15 \), so the magnitude of \( x \) or \( y \) component of the MV varies from 0 to 7 or 0 to 15 respectively. The blocks are classified into
four categories based on the motion content. These are stationary blocks, small motion, medium motion, and large motion blocks.

![Diagram](image)

**Figure 3.** Stationary and small motion blocks, dark points represent the 1st search step whereas light points represent 2nd search step.

5. **SEARCH PATTERN**

The search pattern is based on the motion content of the blocks, which is derived from the magnitude and direction of the predicted MV. Hence the search pattern is highly dynamic and is divided into four cases and eight sectors respectively.

5.1 **Stationary and Small Motion Blocks**

To capture stationary blocks or any motion in small motion blocks the algorithm takes the following steps, as shown in Figure 3.

Step 1: Initially the algorithm checks four points around the search center. If the center point is found to be the minimum distortion point then we will stop the search, otherwise go to step 2.

Step 2: In the subsequent steps, three new points are checked and search is stopped when either the minimum point falls on the search center or its value is less than a fixed threshold T or the end of search window is reached.
The advantage of early search termination (before the end of search window is reached) is that it can substantially reduce the cost of block matching without degrading much video quality. A constant threshold (T=512) has been used in our experiments. It must also be noted that this threshold is certainly adjustable depending on the application requirement. For example if the video quality is not so demanding then T can be increased to a larger value for more speedup.

5.2 Medium and Large Motion Blocks

For medium and large motion blocks while defining the sector boundaries it has been considered that the motion normally occurs in the horizontal and vertical direction, so instead of defining a sector from 0-45 degrees we have defined it from -22.5 to 22.5 degrees. The directional search pattern following this approach is shown in Figure 4. Hence in the first step we follow the directional search pattern, and then for each direction selected (e.g. direction I or direction IV) we will need only one additional point, in addition to the search center. This additional point is selected at a step size of two pixels. This step is followed by two more steps. The whole process is as follows:

Step 1: 1st step directional search pattern is shown in Figure 4.

Step 2: In the 2nd step horizontal and vertical points around the search center are checked at a step size of two, and this is continued till the minimum point lies on the search center or its value is less than a fixed threshold T or the end of search window is reached.

Step 3: In the last step same pattern is followed with a step size of one.

6. EXPERIMENTAL RESULTS

The proposed algorithm is simulated using a number of video sequences that cover a wide range of motion contents and various formats including QCIF, CIF and SIF. For example, Miss America is a typical videoconferencing sequence, with movement restricted to the face area of the
speaker with a fixed background. Hall Monitor represents a ‘security camera’ case. The camera shows a corridor with two people entering, walking through it and eventually leaving it again. In Carphone (a more complex sequence) you can observe moderate foreground movement and fast movement in the background. Football and Table Tennis are fast moving sequences.

Each pixel in the sequences is uniformly quantized to 8 bits. The block size selected is 16x16 pixels. Sum of absolute difference (SAD) distortion function is used as the block distortion measure (BDM). The experimental results are shown in Figure 5 and Figure 6.

![Figure 4](image-url)  

**Figure 4.** 1\textsuperscript{st} step search pattern corresponding to eight sectors, arrow shows the direction of P(X\textsubscript{n}).
Figure 5. Average PSNR versus frame number for (a) frame no. 70 to 170 of Table Tennis sequence, (b) closer view from frame no. 90 to 140.
Figure 6. Average NSPs versus frame no. for 100 frames of Table Tennis sequence.

REFERENCES

