FAST MOTION ESTIMATION BASED ON NATURE OF ERROR SURFACES

1. OVERVIEW

This method is based on spatio temporal correlation and takes into account the nature of error surfaces that are encountered in the real world image sequences. A combination of spatial and temporal predictors has been used to find the initial search center. But there is always a possibility that if the prediction goes wrong the initial search point could be misleading. To avoid this we have used the concept of multiple predictors. For example at the first step instead of choosing only one initial search center we will chose multiple initial search centers to start search. The best one having the minimum error will be assumed to be closer to the global minimum. By accurately predicting the location of the best motion vector candidate we can search a relatively small area in the neighborhood of the predicted MV.

Motion estimation is a multi step process that involves a combination of techniques such as motion starting point, motion search patterns and adaptive control to curb the search, avoidance of search stationary regions and avoidance of local minimum. The collective efficiency of these techniques makes a motion estimation algorithm robust and efficient. The main objective of the proposed algorithm is to decrease the computational burden while keeping a good predicted image quality. Important aspects of the proposed algorithm are 1) Using spatio-temporal neighborhood information that leads to the prediction of initial search center, 2) Multiple initial starting point selection, 3) Adaptive search pattern, 4) Local minimum elimination criteria. All these points help in reducing the computational complexity and finding the true minimum error point. In the proposed algorithm the spatial and temporal correlation is utilized to adjust the size
of the rood shaped search pattern for matching different motion magnitudes and directions. This improves the search speed as well as accuracy.

2. MOTION VECTOR PREDICTION

The regions of support (ROS) for our proposed algorithm is shown in Figure 1. The blocks from the spatial domain are the left and top neighboring blocks (relative to the current block). The left neighboring block is not always correlated with the current block and is unavailable for left margin blocks. So we have chosen top block as well to compensate when the left block is not available. Only one temporal neighboring block i.e. block at the same location in the previous (reference) frame is used for prediction. Thus we have three initial MVs; two MVs are provided by spatial neighbors and one by temporal neighboring block.

The first step of our algorithm utilizes these neighboring MVs for predicting an initial search point which is closer to the global optimum. These motion vectors, $\text{MV}_{\text{SL}}(n)$ (Spatial Left), $\text{MV}_{\text{SA}}(n)$ (Spatial Above), and $\text{MV}_{\text{P}}(n-1)$ (Previous) perform as the candidates of the predicted motion vector $P(X_n)$ for the current block $X$ in frame $n$. If the predictor accuracy is high the optimal MV (inside a given search window) can be attained faster thus enabling computational
savings for fast searches. We will calculate the predicted MV by using the weighted mean method:

\[ P(X_n) = \frac{K}{\sum_{k=1}^{K} \alpha_k MV_k} \]  

(1)

Where \( \alpha \) is co-efficient of weighted mean and \( k \) is for the number of blocks. The \( x \) and \( y \) components of the weighted mean predicted MV are computed independently. Even with simple predictors like mean and median searching the 4x4 area around the initial search center would generally produce more than 90% of the motion vectors obtained by FS algorithm. This should not be surprising since many of the low resolution video sequences such as Claire and Miss America exhibit very small and slow motion and non motion related variations. Once the predicted motion vector is obtained the first step of the algorithm is to move the initial search center to the predicted motion vector location.

3. NATURE OF ERROR SURFACES

Most error surfaces encountered in real world video sequences are not truly unimodal. However the characteristics of the distortion surfaces can be considered unimodal within a small window in the neighboring region of the global optimal point (minimum error point). It has also been noted that localizing the search origin through appropriate predictors reduces the probability of getting trapped in local minimum as the predictors move the search center closer to global minimum. Therefore in our proposed algorithm we are using the idea of multiple predictors acting as multiple initial search points. Figure 2 shows a distortion surface in 1-D space. By considering multiple starting points we can clearly get more chance to reach the global minimum as compared to the case when only point no. 2 might be selected. In case of irregular motion, the chance of locating true motion vector increases by checking multiple points within the search window. This not only improves the local minimum problem but also speeds up the search process.
4. MULTIPLE INITIAL POINT PREDICTION

In our algorithm we have used two spatial neighboring blocks (left and above) and one temporal block (same block in the previous frame) for initial point prediction. The two initial point predictors can be obtained as follows:

From spatial frame (weighted mean of MVs of the two spatial neighbors)

\[ P_S(X_n) = \alpha \cdot MV_{SL} + \alpha_1 \cdot MV_{SA} \]  \hspace{1cm} (2)

From temporal frame (MV of the reference block)

\[ P_T(X_n) = MV_p \]  \hspace{1cm} (3)

Where, \( P_S(X_n) \) and \( P_T(X_n) \), are spatial and temporal predicted MVs, respectively. \( MV_{SL}, \ MV_{SA} \), and \( MV_p \) are the motion vectors of the spatial left, spatial above and temporal reference blocks respectively. \( \alpha \) and \( \alpha_1 \) are weighted mean coefficients.

For the case of starting corner block the spatial predictors are not available so instead we use zero MV for that, whereas for the left column the left block is not available and for the top row the above block is not available. The temporal reference blocks are not available for the first frame.

We divide the search space into four quadrants and then see if both these vectors lie in same quadrant or not. The angles for the division of search space are defined as follows:

4.1 Case 1 (Same Quadrant)

When both spatial and temporal predicted MVs lie in the same quadrant we assume that the dominant motion is in this quadrant and we start our search from this quadrant. This case is shown in Figure 2 (a). This seems to be a simple case so we calculated \( P(X_n) \) by taking the
weighted mean of the two spatial and one temporal motion vector and start the search from this point (one point only). \( P(X_n) \) is calculated as follows [1]:

\[
P(X_n) = \alpha_2 \cdot P_T(X_n) + \alpha_3 \cdot P_S(X_n)
\]  

(4)

\( \alpha_2 \) and \( \alpha_3 \) are weighted mean coefficients.

\[ \text{(a)} \hspace{1cm} \text{(b)} \]

\[ \begin{array}{c}
\text{II} \\
\text{III} \\
\text{IV} \\
\text{I}
\end{array} \hspace{1cm} \begin{array}{c}
\text{II} \\
\text{III} \\
\text{IV} \\
\text{I}
\end{array} \]

\[ \begin{array}{c}
P_T \\
P_S
\end{array} \hspace{1cm} \begin{array}{c}
P_T \\
P_S
\end{array} \]

\[ \text{same} \hspace{1cm} \text{different} \]

**Figure 2.** Spatial and temporal predictors (a) lie in same quadrant, (b) lie in different quadrants.

4.2 Case 2 (Different Quadrants)

When spatial and temporal predicted MVs lie in different quadrants, then we use multiple predictors i.e. two initial predicted motion vectors and start our search from two separate initial points. This is shown in Figure 2 (b) and is explained as follows:

Spatial Predictors \( P_S(X_n) \), and Temporal predictors, \( P_T(X_n) \), as defined by separate equations and are taken as separate starting points, see [1] for detailed equations. This choice of multiple points decreases the risk of ignoring the actual motion and reduces the chance of being trapped in local minimum.
4.3 Local Minimum Elimination Criteria

From the characteristics of distortion surfaces it becomes quite clear that there are a number of local minimums in addition to the global minimum. So the beauty of the search algorithm is that it should get rid of local minimums while searching for the global minimum but keeping a low computational cost. The reason for selecting multiple initial points for prediction is that it can result in increasing a chance of selecting an initial point closer to the global minimum rather than the local minimum. This can be seen from Figure 3 which shows a distortion illustration in 1-D space. Two initial search points 1 and 2 are selected in the first step. Point 1 has lower distortion error so it is considered closer to the global minimum. In the later steps we will extend fine search around point 1 to reach the global minimum point. The local minimum elimination criteria (LMEC) to locate the global minimum point and stop the search in case of multiple initial starting points has been defined.

If LMEC has value higher than a predefined threshold, then we can safely assume that one of the two starting points is actually the global minimum point and stop the search at that point. Otherwise we will continue searching the minimum distortion point from the minimum of the two multiple points calculated.
4.4 Magnitude of Predicted Motion Vector and Motion Content

The magnitude of predicted motion vector is used to define the motion content of the blocks. The blocks are classified into three categories based on the motion content. These are stationary, small motion, and medium motion and large motion blocks.

4.5 Search Pattern

The distribution of the global minimum point in real world video sequences is centered at the position of zero motion, at the search window center as in TSS, FSS and NTSS etc. Most MVs are found to be enclosed in a circular support within a radius of 2–3 pels centered at the position of zero motion. Using these characteristic only 1 to 2 steps of the search pattern will give the final result. Since the refined search center is already closer to the global minimum point any local search using a small compact search pattern should be fairly efficient. Searching on a patterns first step search points is unavoidable as the minimum necessary computational cost of a search pattern is directly related to the number of their first step search points. In our proposed algorithm the search pattern is based on the motion content of the blocks, which is derived from the magnitude of the predicted motion vector. Search pattern also depends on single or multiple point prediction. Types of search pattern employed in the proposed algorithm are shown in Figure 4.

![Search patterns](image)

Figure 4. Search patterns employed in proposed algorithm (a) Large Rood Pattern, (b) Small Rood Pattern.
1. **Stationary Blocks**

For stationary blocks the initial search center is considered same as the actual search center. To capture any motion the algorithm takes the following steps.

Step 1: If SAD (search center) < threshold, then search only one point and the initial search point is taken as the final MV (which is the zero point) as shown in Figure 5 (a).

Step 2: If SAD (search center) ≥ threshold then we search five points, the search center and four neighboring points on horizontal and vertical axis at a step size of one, and then stop the search. Step Size is defined as the horizontal/vertical distance between two pixels. This is shown in Figure 5 (b).

![Figure 5. Search pattern for stationary blocks](a) if SAD < Threshold, (b) if SAD > Threshold.

2. **Motion Blocks (Single Point Prediction)**

Here we encounter two types of cases.

- For case of small motion blocks we use a small rood search pattern.

- For case of medium and large motion blocks again we observe the SAD of the search center. And follow rood pattern accordingly.

3. **Motion Blocks (Multiple Point Prediction)**

Multiple point prediction is further divided into three cases on the basis of distance between two starting points. And hence variable search pattern is choosen
4.6 EXPERIMENTAL RESULTS

The proposed algorithm is implemented in JM-12.2 [2] of H.264/AVC reference software. The simulation is carried out at 4 different quantization parameters (QP=28, 32, 36, 40) to test the algorithm at different bit rates. For encoding purposes JM-12.2 Main Encoder Profile has been used. The rate distortion curves are shown in Figure 6.

![Performance for News QCIF Sequence](chart1.png) ![Performance for Hall Monitor CIF Sequence](chart2.png)

**Figure 6.** Rate-distortion curves for, (a) News, (b) Hall Monitor.

REFERENCES
