A MULTISTAGE FAST MOTION ESTIMATION SCHEME FOR VIDEO COMPRESSION

Jiancong Luo, Ishfaq Ahmad, Yu Sun and Yongfang Liang

Department of Computer Science and Engineering
The University of Texas at Arlington, TX 76019
{jluo, iahmad, sunyu, yliang}@cse.uta.edu

ABSTRACT

This paper presents a novel multistage motion estimation (ME) scheme called *Content Adaptive Search Technique* (CAST). The proposed scheme consists of four stages: motion vector field (MVF) prediction, block-based segmentation, motion parameter extraction, and adaptive search strategy. Through pre-processing the MVF of the previous reference frame in the first three stages, CAST extracts the motion parameters for each region. The 4th stage is a combination of various techniques including MV prediction, search area decision and an adaptive fast search algorithm that is adjusted by a mathematical model for the block distortion surface (BDS). Experiment shows that the proposed scheme improves the visual quality, while yielding a faster speed, comparing with the other predictive ME algorithms.

1. INTRODUCTION

It is well known that motion estimation is the main computational bottleneck in block-based video coding. To avoid the intensive computation in full search algorithm, fast ME algorithms such as [2][3][8], have been proposed. In the last few years, a new class of ME adopting motion vector (MV) prediction techniques came forth [1][5][7]. Nevertheless, it has been shown that the performances of the above mentioned algorithms highly depend on the characteristics of the video contents. There is no single algorithm that can adapt to all kinds of video contents.

We intend to tackle this issue by introducing online analysis technique to the adaptive ME process. In this paper, we propose a multiple stage ME scheme for video compression, which is called *content adaptive search technique* (CAST). CAST can provide adaptability to the video contents to maximize the overall performance.

The innovative features of CAST include:

- Weighted Mean Inertia MVF prediction;
- Online analysis for motion parameter extraction;
- Mathematical model for block distortion surface;
- Adaptive search area decision and MV prediction;
- Adaptive fast search algorithm.

The remaining of the paper is organized as follows: Section 2 introduces the CAST scheme. Subsections are included to describe the various stages. Section 3 presents the performance evaluation results. Finally, the paper is concluded in Section 4.

2. CAST MULTISTAGE SCHEME

2.1 Overview

Various techniques are utilized in a ME algorithm, such as MV prediction, search range, search pattern, and search strategy. Most of these techniques are highly motion-dependent. CAST pre-processes the MVF of the previous reference frame to extract the motion parameters. Being aware of the characteristics of the video contents it is dealing with, the algorithm can choose the optimal parameters for these techniques. The block diagram of CAST is shown in Fig. 1.

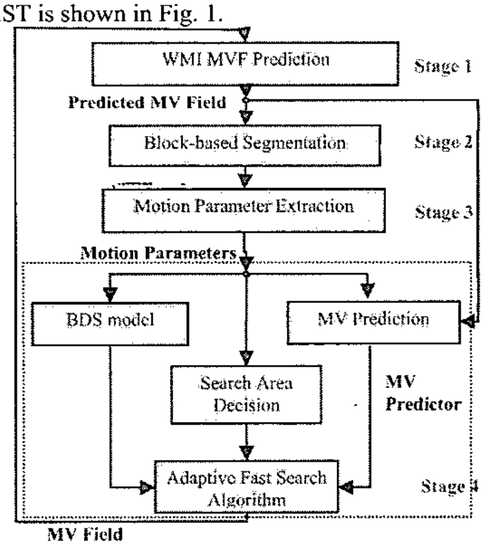

Fig. 1: Block diagram of CAST.

2.2 Motion Vector Field Prediction

Recently, new MVF prediction methods exploiting the motion inertia property are proposed [4] [7]. Here, we propose a novel *Weighted Mean Inertia* (WMI) MVF prediction method that not only increases the prediction accuracy, but also predicts the MB distortion. WMI is described as following.

Let $MB_{i,t-1}$ denote a macroblock (MB) i in frame $t-1$. The MV and distortion of $MB_{i,t-1}$ are $MV_{i,t-1}$ and $D_{i,t-1}$. Due to the motion inertia property, $MB_{i,t-1}$ intends to maintain its motion and has a displacement $-MV_{i,t-1}$ in the next frame. Let $P_{i,t}$ denote the displaced $MB_{i,t-1}$ in frame t. $P_{i,t}$ overlaps one or more MBs in frame t. Let $S_{i,j}$ denote the overlap of $P_{i,t}$ and $MB_{j,t}$, $PMV_{j,t}$ and $PD_{j,t}$ denote the predicted MV and distortion of $MB_{j,t}$. $PMV_{j,t}$ and $PD_{j,t}$ are given below. An illustration of WMI is shown in Fig. 6.

$$PMV_{j,t} = \frac{\sum_i MV_{i,t-1} S_{i,j}}{\sum_i S_{i,j}}, \quad PD_{j,t} = \frac{\sum_i D_{i,t-1} S_{i,j}}{\sum_i S_{i,j}} \tag{1}$$

The correlation coefficients between PMVF and the true MVF reflect the prediction accuracy. Table 2 shows that the WMI MVF prediction is more accurate than the inertia MVF prediction proposed in [7].

2.3 Block-based Segmentation

The segmentation is based on the WMI PMVF obtained in stage 1. A frame is segmented into three regions, i.e. foreground, background and uncovered background. Uncovered background is the region from which the foreground object moves out.

The block-based segmentation consists of the following steps:

Step 1: Estimate the affine transform parameters from the WMI PMVF using the approach in [6].

Step 2: Reconstruct the background MVF with the estimated affine transform parameters.

Step 3: Subtract the reconstructed MVF from the WMI PMVF. Compare the reconstruction errors with a predefined threshold T.

Step 4: Cluster the MBs with the following rules:
Let F, B and U denote the sets of MBs in foreground, background and uncovered background. For each $MB_{j,t}$:

1. If $\sum_i S_{i,j} < \frac{\text{MB size}}{2}$, $MB_{j,t} \in U$.
2. Else, if the reconstruction error is larger than T, $MB_{j,t} \in F$, otherwise $MB_{j,t} \in B$.

Fig. 7 is an example of background and foreground MVFs.

2.4 Motion Parameter Extraction

Motion parameters representing the motion characteristics of each region are extracted. M_{Fvel} (M_{Bvel}) indicates the motion velocity in foreground (background). M_{Fcomp} (M_{Bcomp}) indicates the motion complexity in foreground (background). M_{Fvel} (M_{Bvel}) is the average length of MVs in foreground (background), while M_{Fcomp} (M_{Bcomp}) is the standard derivation of MVs in the foreground (background). In the context below, M_{vel} represents either M_{Fvel} or M_{Bvel}, and M_{comp} represents either M_{Fcomp} or M_{Bcomp}, depending on the region of the current MB.

2.5 Adaptive Search Strategy

The adaptive search strategy consists of search area decision, MV prediction and an adaptive fast search algorithm, which is tuned by the motion parameters and the block distortion surface model.

2.5.1. Adaptive Search Area

We observe that the prediction errors are concentrated to zero in regions with low M_{comp}. On the contrary, in regions with high M_{comp}, the prediction errors distributes in a wide range. We exploit this property by employing a small search area in regions with low M_{comp}, while applying a large search area to regions with high M_{comp}.

We categorize M_{comp} into three levels: low, medium and high. The search area for each level covers a diamond shape with a radius defined in Table 1. The radius is obtained through experiments using the following method. Let p_i be the probability density function of MV prediction error i. The radius is equal to the minimum l satisfying $\sum_{i=0}^{l} p_i \geq P_{target}$. P_{target} is the accumulated probability, which is set to 99%.

Table 1: Search area decision

	Motion complexity level	Search area radius
M_{comp}<0.1	Low	3
0.1≤ M_{comp} <0.6	Medium	7
M_{comp}≥0.6	High	Unrestricted

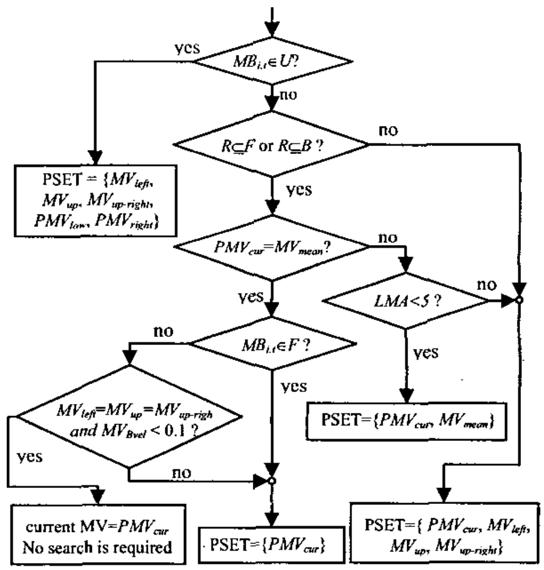

Fig. 2: PSET construction process

2.5.2. MV Prediction

CAST maintains a set of MV predictor candidates (PSET). Let MV_{left}, MV_{up} and $MV_{up\text{-}right}$ denote the MVs of the left, upper and upper-right coded MB in the current frame. PMV_{cur}, PMV_{low}, PMV_{right} denote the WMI predicted MVs of the current, lower and right MB. MV_{mean} denotes the mean of MV_{left}, MV_{up}, and $MV_{up\text{-}right}$. The members of PSET are selected from the above MVs.

We define R as the set of the left, upper and upper-right MBs of the current MB. Local motion activity

(LMA) is defined as the maximum Manhattan length of MV_{left}, MV_{up} and $MV_{up\text{-}right}$, representing the local motion consistence. Fig. 2 illustrates the PSET construction.

After the construction, members of PSET are evaluated first. The candidate associated with the minimum distortion is selected to be the MV predictor.

2.5.3. Block Distortion Surface (BDS) Model

The BDS consists of distortion values of all check points. The center of BDS is the global minimum position (GMP), as illustrated in Fig. 3. We propose a mathematical model for BDS to estimate the distance from the current check point to the GMP.

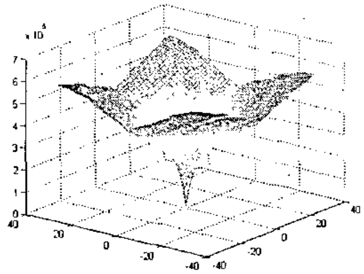

Fig. 3: Average block distortion surface of *Stefan*.

Let r be the distance from a check point to the GMP. $D(r)$ is the distortion function. Through extensive experiments, we observe that $D(r)$ is related to the global minimum distortion $D(0)$. The relation can be described by $(D(r)-D(0))^2/D(0)^2 = g \cdot r$, as shown in Fig. 4.

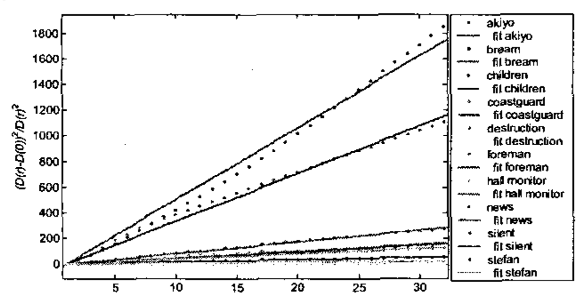

Fig. 4: $(D(r)\text{-}D(0))^2/D(0)^2$ is closely linear to r

Here, we model $g(M_{vel})$ with a rational quadratic polynomial: $g(M_{vel})=1/(a+bM_{vel}+cM_{vel}^2)$, ($a = 0.013$, $b = 0.1$, $c = 0.081$). Finally, the equation can be reduced to:

$$r = \left(\frac{D(r)-D(0)}{D(0)}\right)^2 (a + bM_{vel} + cM_{vel}^2) \qquad (2)$$

The equation allows to estimate the distance r, given $D(r)$, M_{vel} and $D(0)$. We approximate the value of $D(0)$ by the WMI predicted distortion in (1).

2.5.4. Adaptive Fast Search Algorithm

In the final stage, an adaptive fast search algorithm is performed to find the MV. The search starts from the *Initial Check Point (ICP)* pointed by the MV predictor. First, a cross pattern centered at the *ICP* is evaluated, which consists of 4 check points: (s,0), (0,-s), (-s,0), (0,s). s is the size of the cross pattern. Initially, s =1 if $r < 4$, otherwise $s = 2$. If one of these 4 check points has the minimum distortion, the direction indicated by this check point is the distortion decreasing direction. Then, a one-dimensional search will be performed along this direction. The step size of the one-dimensional search increases exponentially. E.g. if (s,0) has the minimum distortion, (2s,0), (4s,0), (8s,0)…, will be checked sequentially. It will continue until the distortion increases. In the above example, if the distortion begins to increase at (8s,0), the one-dimensional search is stopped and (4s,0) will be the new search center. Then, a cross pattern on the new search center is employed again, to adjust the search direction. The process repeats until the minimum distortion occurs at the center of the cross pattern. Then, a cross pattern with smaller size is used to verify the GMP. Fig. 5 illustrates the adaptive fast search algorithm.

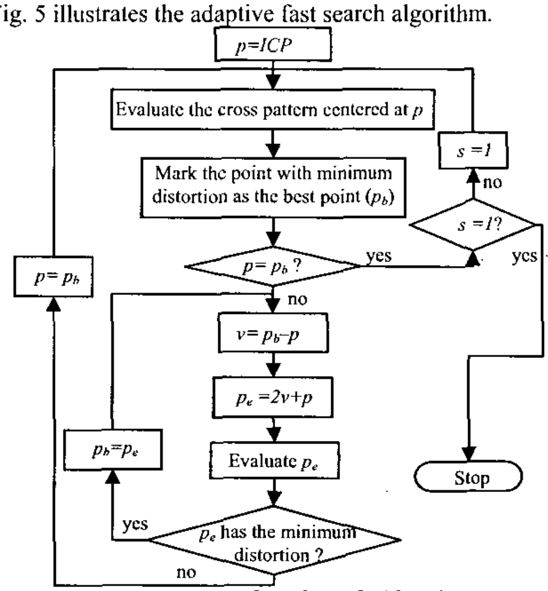

Fig. 5: Adaptive Fast Search Algorithm

3. PERFORMANCE EVALUATION

This section presents the performance evaluation. We compare CAST with two ME algorithms, MVFAST [1] and AMSED [7], which are based on MV prediction technique. MPEG-4 recommended MVFAST is used as reference. Several bench mark sequences, from slow to fast motion, are encoded in the simulation. PSNR and speedup are used to evaluate the quality and speed performance.

Table 3 shows that CAST substantially improves the search speed and achieves higher visual quality than the other two ME algorithms, for all kinds of video contents. In particular, CAST is much faster to encode the slow motion sequence *container*, without any quality degradation. This is due to the fact that CAST directly adopts the MV predictor as the MV in most slow motion background MBs, therefore no search is required.

4. CONCLUSION

A content adaptive search scheme is proposed. CAST provides adaptability to various types of video contents. Experiment shows that CAST can estimate the MVF accurately. It outscores MVFAST and AMSED in both visual quality and computational cost. The proposed scheme has the best overall performance among the compared algorithms after considering the overhead introduced by the online video analysis process.

5. REFERENCES

[1] P.I. Hosur and K.K. Ma, "Motion Vector Field Adaptive Fast Motion Estimation," *Second International Conference on Information, Communications and Signal Processing (ICICS'99)*, Singapore, pp.7-10, 1999.

[2] J.R. Jain and A.K. Jain, "Displacement Measurement and Its Application in Interframe Image Coding," *IEEE Trans. on Communication*, Vol. COM-29, pp.1799-1808, 1981.

[3] R. Li, B. Zeng, and M. Liou, "A New Three-Step Search Algorithm for Block Motion Estimation," *IEEE Trans. Circuit and Systems Video Technology*, Vol. 4, No.4, pp.438-442, 1994.

[4] T. Liu, K.T. Lo, J. Feng and X. Zhang, "Frame Interpolation Scheme using Inertia Motion Prediction," *Signal Processing: Image Communication*, Vol. 18, 2003.

[5] A.M. Tourapis, O.C. Au, and M.L. Liou, "Highly Efficient Predictive Zonal Algorithms for Fast Block-matching Motion Estimation," *IEEE Trans. on Circuits and Systems Video Technology*, Vol. 12, No.10, pp.934 – 947, 2002.

[6] D. Wong and L. Wang, "Global Motion Parameters Estimation using a Fast and Robust Algorithm," *IEEE Trans. on Circuits and Systems Video Technology*, Vol. 7, No.5, 1997.

[7] W. Zheng and I. Ahmad, and M.L. Liou, "Adaptive Motion Search with Elastic Diamond for MPEG-4 Video Coding," in *Proc. Image Processing (ICIP2001)*, Vol. 1, pp.377-380, 2001.

[8] S. Zhu and K.K. Ma, "A New Diamond Search Algorithm for Fast Block-Matching Motion Estimation," *IEEE Trans. Image Processing*, Vol. 9, No.2, pp.287-290, 2000.

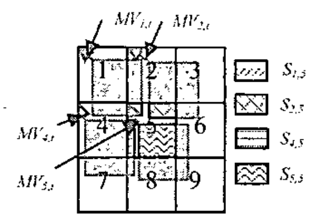

Fig. 6: WMI MVF prediction.

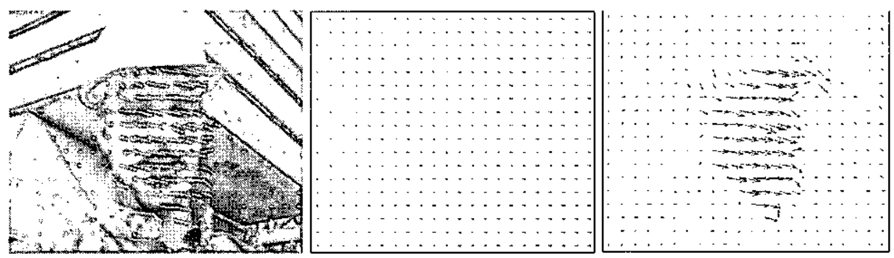

Fig. 7: Original MVF (left), background MVF (middle), foreground MVF (right).

Table 2: Correlation coefficients between PMVF and the true MVF

a) Predicted from frame *t-1*

	WMI		Inertia[7]	
	X	Y	X	Y
Coastguard	0.825	0.028	0.829	0.036
Foreman	0.569	0.366	0.548	0.310
Table tennis	0.586	0.547	0.524	0.500
Stefan	0.694	0.317	0.653	0.311
Average	0.668	0.314	0.638	0.289

b) Predicted from frame *t-2*

	WMI		Inertia[7]	
	X	Y	X	Y
Coastguard	0.766	0.038	0.754	0.051
Foreman	0.570	0.330	0.541	0.284
Table tennis	0.545	0.503	0.493	0.455
Stefan	0.657	0.410	0.617	0.357
Average	0.634	0.320	0.601	0.287

c) Predicted from frame *t-3*

	WMI		Inertia[7]	
	X	Y	X	Y
Coastguard	0.439	0.175	0.462	0.100
Foreman	0.448	0.260	0.410	0.213
Table tennis	0.589	0.433	0.554	0.419
Stefan	0.318	0.179	0.288	0.159
Average	0.448	0.262	0.429	0.223

Table 3: PSNR and search speed comparison

		CIF				QCIF			
Sequences	ME	PSNR	Check Points	speedup	PSNR gain	PSNR	Check Points	speedup	PSNR gain
Container	MVFAST	39.065	52816	1	0	48.951	10639	1	0
	AMSED	39.065	43228	1.22	0	48.951	10235	1.04	0
	CAST	39.065	4164	12.68	0	48.951	909	11.7	0
Foreman	MVFAST	38.341	281823	1	0	43.715	53120	1	0
	AMSED	38.305	158043	1.78	-0.036	43.729	33356	1.59	0.014
	CAST	38.368	146070	1.93	0.027	43.73	31415	1.69	0.015
Stefan	MVFAST	32.451	219872	1	0	38.252	66857	1	0
	AMSED	32.404	169307	1.3	-0.047	38.203	37884	1.76	-0.049
	CAST	32.597	127126	1.73	0.146	38.316	24377	2.74	0.064
Table Tennis	MVFAST	35.998	265012	1	0	42.546	78212	1	0
	AMSED	35.882	165927	1.6	-0.116	42.506	36549	2.14	-0.04
	CAST	36.089	147297	1.80	+0.091	42.586	32153	2.43	+0.040