Pattern Recognition Letters 33 (2012) 52–61

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Content adaptive fast motion estimation based on spatio-temporal homogeneity analysis and motion classification

Humaira Nisar [a,*], Aamir Saeed Malik [b], Tae-Sun Choi [c]

[a] Universiti Tunku Abdul Rahman, Kampar, Perak, Malaysia
[b] Universiti Teknologi Petronas, Tronoh, Malaysia
[c] Gwangju Institute of Science and Technology, Gwangju, Republic of Korea

A R T I C L E   I N F O

Article history:
Received 9 November 2009
Available online 28 September 2011
Communicated by A. Shokoufandeh

Keywords:
Motion estimation
Block matching
Full Search
Motion classification
Video coding
Spatial correlation

A B S T R A C T

In video coding, research is focused on the development of fast motion estimation (ME) algorithms while keeping the coding distortion as small as possible. It has been observed that the real world video sequences exhibit a wide range of motion content, from uniform to random, therefore if the motion characteristics of video sequences are taken into account before hand, it is possible to develop a robust motion estimation algorithm that is suitable for all kinds of video sequences. This is the basis of the proposed algorithm. The proposed algorithm involves a multistage approach that includes motion vector prediction and motion classification using the characteristics of video sequences. In the first step, spatio-temporal correlation has been used for initial search centre prediction. This strategy decreases the effect of unimodal error surface assumption and it also moves the search closer to the global minimum hence increasing the computation speed. Secondly, the homogeneity analysis helps to identify smooth and random motion. Thirdly, global minimum prediction based on unimodal error surface assumption helps to identify the proximity of global minimum. Fourthly, adaptive search pattern selection takes into account various types of motion content by dynamically switching between stationary, center biased and, uniform search patterns. Finally, the early termination of the search process is adaptive and is based on the homogeneity between the neighboring blocks.

Extensive simulation results for several video sequences affirm the effectiveness of the proposed algorithm. The self-tuning property enables the algorithm to perform well for several types of benchmark sequences, yielding better video quality and less complexity as compared to other ME algorithms. Implementation of proposed algorithm in JM12.2 of H.264/AVC shows reduction in computational complexity measured in terms of encoding time while maintaining almost same bit rate and PSNR as compared to Full Search algorithm.

1. Introduction

H.264 AVC (audio video coding) standard shows higher compression ratio with better visual quality than other video coding standards (LeGall, 1991; ISO/IEC JTC// SC29/WG11 Moving Picture Experts Group, 1993; CCITT SG XV, 1990; Wiegand et al., 1449). The major coding efficiency of H.264 is because of spatial intra prediction, 4×4 block based transformation, variable block size motion estimation etc. Because of these additions, the complexity of the H.264 standard is very high; in particular motion estimation (ME) occupies 60–90% of the total encoding time. Therefore, low complexity fast ME algorithms are required to support real time video services. Full Search (FS) is the most straightforward and optimal block matching motion estimation algorithm but is computationally expensive as it covers every pixel in every frame. Also, it may fail to obtain reliable motion vectors (MVs) under the presence of noise. In recent years, several fast ME algorithms based on a variety of techniques have been proposed to reduce the encoding time of ME while keeping a good predicted image quality.

Fast block ME algorithms can generally be classified into two categories; (1) reducing the number of checking (search) points, and, (2) lowering the computational complexity in calculating the block matching criterion for each checking point. Our focus is on the algorithms in the first category. Some well-known ME algorithms like three-step search (TSS) (Koga et al., 1981), new three-step search (NTSS) (Li et al., 1994), four-step search (FSS) (Po and Ma, 1996), diamond search (DS) (Zhu and Ma, 2000), and hexagonal search (HEXBS) (Zhu et al., 2002) are based on unimodal error surface assumption (UESA). The UESA states that the block matching error increases monotonically as we move away from

* Corresponding author. Address: Department of Electronic Engineering, Faculty of Engineering and Green Technology, Universiti Tunku Abdul Rahman, Jalan Universiti, Bandar Barat, 31900 Kampar, Perak, Malaysia. Tel.: +605 468 8888; fax: +605 466 7449.

E-mail address: humaira@utar.edu.my (H. Nisar).

0167-8655/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2011.09.015

the global minimum. But this assumption does not always hold true in real-world video sequences, and the algorithm may fall into a local minimum. This results in degradation of the reconstructed image quality. Experimental results show that these fixed search pattern algorithms reduce the computational requirements significantly by checking only some points inside the search window, while maintaining good error performance compared with FS. DS (Zhu and Ma, 2000) is an outstanding algorithm adopted by MPEG-4 verification model (VM) due to its superiority to other methods in the class of fixed search pattern algorithms. It is a center biased search algorithm. However, it is also observed that fixed search patterns are unable to constantly match the dynamic motion content, thus not only wasting the computational power but also leading to local minimum matching error trapping and large prediction errors.

Another popular group of block ME algorithms employ spatio-temporal correlation (Chen et al., 2002; Nie and Ma, 2002; Ma et al., 2003; Kim et al., 2006; Ahmad et al., 2006; Tai et al., 2008; Nisar and Choi, 2008; Nisar and Choi, 2006; Nisar, 2008; Nisar and Choi, 2009; Luo et al., 2008; Zhang et al., 2010; C et al., 2004; Tsai and Hang, 2009; Cai et al., 2009; Liaw et al., 2009; Joint Video Team Reference Software, 2007), using the neighboring blocks in spatial and temporal domain. The main advantage of prediction based algorithms (algorithms using spatial or temporal neighboring information) is that they alleviate the local minimum problem to some extent. Since the new initial or predicted search center is usually closer to the global minimum, the chance of getting trapped in a local minimum decreases. This idea has been incorporated by many fast block motion estimation algorithms. It is possible that fast prediction based initial search may generate more sensible initial MVs that work better for the next stage of refinement. Also, MVs estimated in a prediction algorithm are more realistic in that they reflect a real physical phenomenon, whereas FS just amounts to finding the minimum sum of absolute difference (SAD), which does not necessarily give realistic MVs and is sometimes a bad choice because of the complexity and randomness of real-world imagery. UMHexagonS (Chen et al., 2002) is a very successful ME algorithm; it has reduced more than 90% ME time as compared to FS while maintaining the rate-distortion performance to a very good level. UMHexagonS uses a number of predictors, a hexagon based search pattern and some threshold based early termination strategies to quit the current search. Another important algorithm based on prediction is adaptive rood pattern search (ARPS) Nie and Ma, 2002. It uses neighboring block MVs for prediction; the adaptive rood pattern size is dynamically determined at the initial search stage. Then a unit-size rood pattern is exploited repeatedly and unrestrictedly to find the best MV in the refined local search stage. ARPS-2 (Ma et al., 2003) is an improved version of ARPS. Enhanced block motion estimation (EBME) based on distortion-directional search patterns (Kim et al., 2006) is another good addition to the store of prediction-based algorithms, which can be taken as an advanced version of ARPS and ARPS-2. The initial search is performed on a specially defined rood pattern with the arm length determined adaptively. In the fine search stage, variable-size directional rood patterns are used to increase the speed of the search process. Some famous prediction based adaptive search algorithms are fast adaptive motion estimation (FAME) Ahmad et al., 2006 and the fast motion estimation algorithm using motion adaptive search (MAS) Tai et al., 2008. Multiple initial point prediction based search pattern (Nisar and Choi, 2009) selection is another fast motion estimation algorithm that uses spatio-temporal correlation to adaptively select multiple initial starting points and patterns. Another fast motion estimation algorithm is content adaptive search technique (CAST) Luo et al., 2008 that performs motion analysis to assist in motion vector search. In this algorithm, the search is performed separately for foreground and background. Content adaptive Lagrange multiplier (CALM) Zhang et al., 2010 dynamically adapts the Lagrange multipliers for each macro block based on the content of neighboring or upper layer blocks to improve rate distortion performance.

Motion estimation based on two stage predictive search algorithms based on joint spatio-temporal correlation information (Hsieh et al., 2011) is a recently proposed fast motion estimation algorithm. Its two stage predictive search helps to reduce the computational complexity. In the first stage a rough search is conducted using six spatially and temporally correlated blocks to find a starting point that is closer to the global minimum. In the second stage BBGDS (Liu and Feig, 1996) algorithm and proposed predictive partial search algorithm are used for fine search. Another recently proposed fast motion estimation algorithm for H.264 (Cai et al., 2009) uses a mode discriminant approach so that encoder does not need to check small block size modes in homogeneous regions. Secondly a condensed hierarchal block matching method and a spatial neighbor searching scheme are employed to find the best full pixel motion vector. In the final step direction based selection rule is utilized to reduce the search range in sub pixel ME process. Fast block matching using prediction and rejection criteria is another fast ME algorithm that uses correlation between layers of the sum pyramid for a block. Initial motion vector prediction for a template block is also used (Liaw et al., 2009).

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 of motion and content (slow, fast, smooth, irregular) can help to define the search approach. Here we present a predictive motion estimation technique employing spatio-temporal correlation, homogeneity analysis and unimodal error surface assumption. The proposed algorithm follows a multistage approach that involves motion vector prediction, motion analysis and classification. The analysis stage helps the search technique to adapt to motion characteristics and classifies it into various categories that control the search process by avoiding search stationary regions and local minimum, and keeping track of motion content (slow, medium, fast, smooth and complex motions), by using switchable search patterns. The search patterns play a key role in deciding the performance of a search algorithm especially when the data correlation is low. A single static search pattern cannot handle all the varying real world sequences simultaneously. Each of the techniques mentioned above have several diversions that may suit a particular set of video characteristics. It would be hard to conceive an algorithm that can perform well for all kinds of video contents. However, if important characteristics of a video sequence can be identified and utilized for adjusting various steps of motion estimation, one can design an adaptive algorithm that can tune its parameters to suit the video at hand. A multistage motion estimation algorithm that includes a pre-stage for analyzing the motion characteristics of a video sequence in real time is required and hence proposed in this work. Prior information of the video sequences can help a lot in deriving a dynamic search pattern that can suit different video sequences encountered. Adaptive early termination criteria based on motion content of video sequence further accelerates the search process. We have evaluated the proposed algorithm through a comprehensive performance study that shows that the proposed algorithm achieves substantial speedup without quality loss for a wide range of video sequences.

The rest of the paper is organized as follows: The proposed algorithm is discussed in Section 2. In Section 2.1, motion vector prediction scheme is introduced. In Section 2.2, the criteria for homogeneity between the neighboring blocks, is defined. Stationary search regions have been identified in Section 2.3 that is followed by the classification of different types of motion encountered in video sequences in Section 2.4. Section 2.5 throws