Survey on Motion Vector Filtering and Object Segmentation Methods in Compressed Domain

R.Rajkumar[1], D.SaiKrishna[2], Jayanth.A.S[3]

[1,2] School of Computing Science and Engineering, VIT University, Vellore, Tamil Nadu

[3] Dept of E&C, Manipal Institute of Technology, Manipal, Karnataka

Abstract

In the compression domain video analysis, motion vectors and DCT coefficients are the two main components used to analyze the motion of objects. Compression domain video analysis has many advantages over pixel domain analysis, it is becoming popular very fast and thus there is a need to develop an efficient method to identify the objects in the compression domain. In due course, large number of compressed domain approaches have appeared over the years, including foreground segmentation and object tracking. As mentioned earlier motion vectors and DCT coefficients play a major role in object detection and there is a possibility of noise in both the components. An efficient algorithm is required to filter out these noisy components. Considering the issue, we discuss several approaches for both noise filtering and object segmentation in a compressed video.

Keywords : *Motion Vectors, DCT coefficients, noise filtering, object segmentation.*

1. Introduction

Video compression has gained significant importance in everyday life, as it finds its use in television broadcasting, video streaming and video conferencing to name a few. Various modern compression techniques like MPEG-1, MPEG-2, MPEG-4, H.261, H.263 and H.264/MPEG-4 have emerged over the past two decades. One major advantage of transferring the video in compressed format is that, the data rate is very low.

As video surveillance is rapidly taking new shape in this technology driven era, they are growing in size, complexity and capacity. Evolution of advanced coding standards have aided better compression of video data, but on other hand the decompression of video requires more resources and large volumes of memory to store the decompressed video. Higher resolution video frames delivered from the recent high definition [HD] cameras require more processing time to analyze the video. Also pixel domain analysis requires full length decoding of compressed video which is highly expensive. Hence there is a need for an alternate approach to analyze the compressed video. A better approach would be analyzing the data in compressed domain using motion vectors and DCT coefficients.

To avoid the complete decoding process and to reuse the work done during encoding, a constant endeavor is made at detecting moving objects directly from the compressed video stream. Many algorithms have been recommended to analyze video content in the MPEG and H.264/AVC compressed domain. H.264/AVC is a new and effective video coding standard introduced jointly by the Moving Picture Experts Group (MPEG) and the Video Coding Experts Group (VCEG). It contains several new features which make previous object detection techniques in MPEG compressed domain not directly reusable. Filtering out noisy motion vector and Object segmentation play a vital role in compressed domain video analysis. So in this discussion we present few methods to perform both the operations.

The rest of the paper is organized as follows. In Section 2, the basic architecture for compression domain analysis is discussed. Section 3 describes different noise filtering methods and section 4 provides a brief introduction to various object segmentation methods in compression domain. Finally we present the conclusion in section 5.

2. Basic Architecture Of Compression Domain Video Analysis

Generally to analyze the video in compression domain, we need to extract the basic components from encoded video without decoding it completely. As mentioned earlier the basic components are motion vectors (MVs) and DCT coefficients. Since the standard encoder encodes the video frames in units of 16 by 16 (16X16) macro blocks(MB), these MV's and DCT corresponds to MB's of different frames. With these extracted information we can define the motion, spatial frequency, edge and color contents of MB.
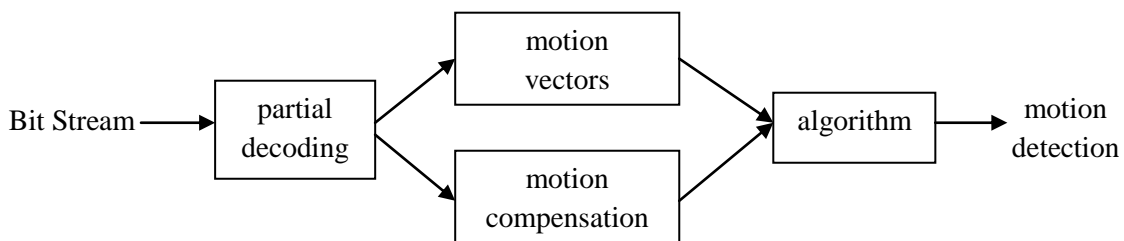

Fig. 1: Architecture for compression domain video analysis

So by utilizing these two components various algorithms have been proposed for different compression formats. Here we present different approaches taken by different authors to detect motion in compression domain by utilizing these MV's and DCT coefficients.

3. Filtering Methods

Motion vector filtering is a very important task in the compression domain motion analysis. Since motion of an object is to be detected from MV fields, the reliability of result depends on the filtered motion-vectors. So we present some of the motion vector filtering approaches in the compression domain.

Fig. 2: Example of Compressed Domain MV Noise Filtering

3.1 Mean Accumulated Thresholded (MAT) Filter

Golam Sorwar et al. in [3] have proposed a mean accumulated thresholded method for removing noisy motion vector. This method involves following steps:
- Involves an iterative "in-place" application of mean filter [3].
- Condition : With the mean and median filters, after many iterations of "in-place", the length of new motion vectors will not cross the maximum length of original vectors in the neighborhood.

From satisfying the above condition only true motion vectors are extracted eliminating the noisy one.

3.2 Cascade Filter

The method proposed in [4], is a novel approach for object detection in MPEG-2. In order to detect an object in a video, motion vectors from MPEG-2 stream is used. Since robust object detection is not possible through directly extracted MV fields, motion vector smoothing and noise reduction are carried out here.

To avoid independent noise, which is present in the content of motion vectors field and distributed over the entire frequency spectrum, either a frequency filter or a spatial filter can be used. As spatial filter is computationally less expensive than a frequency filter, a cascade filter which is composed of two spatial filters namely Gaussian filter and Median filter is used in this method to reduce the noise from the raw motion vectors.

Cascade filter = Gaussian filter + Median filter

3.3 Spatio-Temporal Filter (STF)

The method in paper [12], is based on Spatio-temporal consistency of motion vectors. In order to filter the noisy motion vectors the method implements the following steps.

- MV Normalization: To make the motion vectors independent of frame type (P , B) they are normalized in this step [12].

- Temporal Consistency: Vector matching ratio is used to analyze the MVs related to blocks and its position in the previous frame [12].Temporal consistency between the consecutive frames is calculated here [12].

- Spatial Consistency: Spatial consistency is calculated between the reference MVs and surrounding MVs based on the distance.

Finally a motion vector can be stated spatio-temporally consistent if its Temporal consistency Index or Neighbor Consistency Index (TCI or NCI) value is above the minimum threshold [12].

3.4 Median Filter With Confidence Measure

The paper [1] in entirety deals with accuracy estimation and smoothening of motion field using confident measures which depends on DCT coefficients and spatial/temporal consistency of motion. This method determines spatial, temporal and directional confidence of the input stream, using three adjacent frames. Considering this combined confidence score, low confident macro block with motion vectors are removed. Sequential approach of this method involves the following steps.

Confidence measure = (Spatial + Temporal + Texture) Confidence

The spatial filtering is ensured with in a box [2].This is done using a median filter or morphological filter. The temporal filtering is performed based on the confidence score of three frames and frame types. It is simply a one dimensional weighted sum of MV's from the previous, current and next frame's corresponding macroblocks[2].

4. Object Segmentation Methods

Compression domain video object segmentation (VOS) plays a major role in real time application such as video surveillance, video transcoding and indexing. The purpose of video object segmentation is to partition the image sequence into meaningful regions. As VOS in compression domain is new and challenging task we briefly discuss some of the object segmentation methods in compression domain.

Fig. 3: Example of Compressed Domain Object Segmentation

4.1 Expectation Maximization (EM) Algorithm

Babu in [2] has proposed an EM algorithm for object segmentation on compressed MPEG video. As an initial step of object segmentation, motion vectors and DCT coefficients are extracted from the compressed MPEG video. Based on the two dimensional median filter and DCT error energy of macroblock, reliable motion vectors are separated from the noisy MVs. Finally by using Delaunay triangle based surface interpolation and Gaussian filter, smoothed dense motion field for current frame is obtained.

By using the obtained dense motion field, the object segmentation is carried out in the following way [2] .

- **E-Step** — Calculates the probabilities associated with classification of each pixel.
- **M-Step** — Filters the motion model estimates.
- **Motion Model Estimation** — determines the number of motion models, which are important in the final stage of segmentation. K-means clustering is used to determine the motion models.

After determining the number of motion models Edge refinement is done to overcome poor edge localization[2]. Finally the object is segmented.

4.2 Block Based MRF Model

The algorithm proposed in this paper [7] is based on block-based Markov Random Field (MRF) model for moving object segmentation — from motion vectors and DCT coefficients obtained from the bit stream. Block based MRF model considers H.264/AVC which is an advanced video coding standard for object segmentation.

MRF algorithm has two main stages
- Classification of Motion vectors
- Classification of MRF

In order to filter the noisy MV's and figure out the original MV's related to object motion, MV's are classified into four types [3] here. Based on the classified motion vectors Markovian labeling procedure is used for object segmentation.

MRF model makes use of
- **MV similarity** — used to merge blocks.
- **Temporal consistency** — track the object labels from many frames.
- **Block size** — to remove noise in the final segmentation.

4.3 Using Dynamic Design Of Fuzzy Sets

This algorithm is based on Motion vectors and decision modes of H.264/AVC compressed video. It makes use of fuzzy logic, which helps in defining position, velocity and size of the detected regions and it does not use luminance and chrominance values for segmentation. Fuzzy logic allows to represent the motion information in linguistic way [10], which gives us the linguistic description of motion of macroblocks among adjacent frames[10]. Here the motion vectors are defined as linguistic MVs if it contains the related information about velocity and direction of the object. A linguistic blob is a 7-tuple, composed of similar MVs that represent the region in the frame.

In the segmentation work carried out here, the motion between frames is calculated using fuzzification on motion vectors and it neither uses DCT information nor statistical filtering. After fuzzification, valid linguistic MVs are converted to linguistic blob using clustering algorithm. Following this dynamic design of linguistic variable is done as a modification and adaption of values of fuzzy sets[10]. Dynamic design of fuzzy sets depends on common features of fuzzy systems like supper set, height of fuzzy set and alpha cut threshold[10].

OBJECT SEGMENTATION = LINGUISTIC BLOBS + DYNAMIC DESIGN OF FUZZY SETS

4.4 Ant Colony Algorithm

From experimental studies it is shown that, a group of ants can discover food source easily, while a single ant cannot. So this method is a positive feedback method. Initially the motion vector fields are extracted from the H.264/AVC compressed video and by making use of magnitude and orientation, MV fields are classified into background, foreground and noisy. Here each MV element mv_i is considered as an ant and j cluster centers C_j food sources. As group of ants find the food, motion vectors gather into j categories and before object segmentation, MVs are normalized both spatially and temporally and at the same time Euclidean distance is calculated between motion vector and clustered center[14]. A heuristic function defines the degree of similarity of motion and clustered center [14]. Based on the result of clustering and orientation histograms moving object area is determined.

4.5 Volume Growth Method

The methods proposed by Fatih Porikli et al.[13] is based on MPEG video. The following steps are carried out prior to object segmentation process.

- **MPEG Video Parsing** — MPEG video is partially decoded by extracting the DCT and MV.
- **Frequency-Temporal Data Structure** — Frequency Temporal data is formed into a single layer by assembling the DCT values of I-frames and MV of P-frames.
- **Motion Aggregation** — Motion vectors for each intra-coded block within its immediate local neighborhood are interpolated in order to find the regular motion field. 3×3 Gaussian-shaped kernel is used to remove the extremities and keeping original boundaries.

Volume growth method is a iterative method, which expands the volume of Frequency Temporal (FT) data one after the other starting from seed point.

- Selection of a seed from a GOP.
- Initializing a volume descriptor **'v'** using the feature vector **'f'***(pseed)* of the seed point *P*seed.
- Define set of active boundaries.
- Comparison between current volume descriptor **v** and adjacent points of each active boundary points.
- Calculate the distance of a feature vector **f** of the neighboring point to the volume descriptor **v**.
- Select a threshold value that prevents under segmentation.
- Volume growth and seed selection process is done iteratively until no more point remains in the Frequency Temporal data structure.

4.6 Multikernel Meanshift Segmentation

Prior to object segmentation process, the steps defined in the previous method are carried out here as well.
Mean shift is an iterative method and it involves the following steps in object segmentation.

- Iteratively calculate the local gradient direction within a kernel and kernel's mean is shifted in the space.
- Base on the point where this shift operation moves and converges to the kernel is called sink point.
- Identify the clusters of sink points by connecting together all $\mathbf{f}*j$ (spatial sink points) which are closer than a preset value from each other in the joint domain.
- Finally, points that are assigned to same cluster of sink points are grouped together.

5. Conclusion

In our brief discussion here we have considered different motion vector filtering and object segmentation methods, of which some of them are applicable for MPEG and some for H.264/AVC video standards. Each of them have an unique approach to give best results. From the above mentioned filtering algorithms, it can be stated that spatio-temporal motion vector filtering would be the best option for filtering MVs in either of the video coding standards, as it filters the motion vectors both spatially and temporally. From the various object segmentation methods discussed we observe that MRF model which works on H.264/AVC would be the best one, as it uses the MV classification method to avoid the noisy motion vectors. Finally we conclude saying, if spatio temporal filtering of motion vectors is combined with MRF model it would yield results better than any previously obtained results.

Table 1 : Comparison Of Different Object Segmentation Methods In Compressed Domain

ALGORITHM	VIDEO FORMAT	COMPONENTS	FILTERING METHOD	BLOCKSIZE	RESOLUTION
EM*	MPEG	MVs & DCT	Median Filter	8 X 8	-
VOLUME GROWTH	MPEG	MVs & DCT	Gaussian Filter	8 X 8	352 X 288
MEAN SHIFT	MPEG	MVs & DCT	Gaussian Filter	8 X 8	352 X 288
MRF*	H.264/AVC	MVs & DCT	MV classification	Variable	352 X 288, 352 X 240
FUZZY SETS	H.264/AVC	MV	Decision mode selection	Variable	320 X 240, 640 X 240, 640 X 480
ANT COLONY	H.264/AVC	MV	Magnitude & Orientation	4 X 4	-

*   EM- Expectation Maximization, MRF- Markovian Random Field

References

[1]   Roy Wang, Hong-Jiang Zhang2,Ya-Qin Zhang ,” A confidence measure based moving object extraction system built for compressed domain”, The work was performed in MSR Beijing, where the 1st author interned in summer 1999.

[2]   R. Venkatesh Babu and K. R. Ramakrishnan, “Compression domain motion segmentation for video object extraction “, IEEE International Conference on Electromagnetic Interference and Compatibility, 2002.

[3]   Golam Sorwar, Manzur Murshed, and Laurence Dooley, “A Novel Filter for Block-Based Object Motion Estimation”, Digital Image Computing Techniques and Applications, Melbourne, Australia ,2002.

[4]   Ashraf M.A. Ahmad; Duan-Yu Chen and Suh-Yin Lee , “Robust Object Detection Using Cascade Filter in MPEG Videos “,Proceedings of the IEEE Fifth International Symposium on Multimedia Software Engineering ,2003.

[5]   Thomas Wiegand, Gary J. Sullivan, Gisle Bjontegaard, and Ajay Luthra ,” Overview of the H.264 / AVC Video Coding Standard “ , IEEE Transactions on circuits and Systems for Video Technology, july 2003.

[6]   Ashraf M. A. Ahmad; Bashar M. A. Ahmad and Suh-Yin Lee , “Fast and Robust Object Detection Framework in Compressed Domain”, Proceedings of the IEEE Sixth International Symposium on Multimedia Software Engineering ,2004 .

[7]   Wei Zeng, Jun Du, Wen Gao, Qingming Huang “Robust moving object segmentation on H.264/AVC compressedvideo using the block-based MRF model “ , 2005 Elsevier Ltd.

[8]   Ibrahim, M.M, Supriya Rao, “Motion Analysis In Compressed Video - An Hybrid Approach”, IEEE Workshop on Motion and Video Computing ,2007.

[9]   Takanori Yokoyama, Toshiki Iwasaki, and Toshinori Watanabe, “Motion Vector Based Moving Object Detection and Tracking in the MPEG Compressed Domain”, Seventh International Workshop on Content-Based Multimedia Indexing ,2009.

[10]   C. Solana-Cipres , L. Rodriguez-Benitez, J. Moreno-Garcia L. Jimenez G. Fernandez-Escribano, “Real-time segmentation of moving objects in H.264 compressed domain with dynamic design of fuzzy sets “,IFSA-EUSFLAT 2009.

[11]   Chris Poppe , Sarah De Bruyne, Tom Paridaens, Peter Lambert, Rik Van de Walle ,” Moving object detection in the H.264/AVC compressed domain for video surveillance applications “, Elsevier 2009.

[12]   Ronaldo C. Moura ,Monity, Elder Moreira Hemerly , “A Spatiotemporal Motion-Vector Filter for Object Tracking on Compressed Video “,Seventh IEEE International Conference on Advanced Video and Signal Based Surveillance,2010.

[13]   Fatih Porikli, Faisal Bashir, Huifang Sun, “Compressed Domain Video Object Segmentation”, IEEE Transactions on Circuits and system for Video Technology, VOL. 20, NO. 1, JANUARY 2010 .

[14]   Wang Pei , Wu Zhixia , “Moving Object Segmentation in H.264/AVC Compressed Domain Using Ant Colony Algorithm” , 2nd International Conference on Signal Processing Systems ,2010.