On Acquiring a Motion Field in the Compressed Domain

Cheng Quan Jia

Abstract

Video object segmentation in the Compressed Domain has gained research attention due to its reduced complexity. Without the need to decode a compressed bitstream to the pixel domain, this approach makes application in real-time feasible. However, the motion vectors obtained in the motion compensation step is not intended to capture real object motion, and a measure is needed to ensure the validity of the said vectors. Since the motion compensation step in the encoding process resembles that of obtaining the 2-D optical flow field, optical flow techniques could be applied to the Compressed Domain. This paper is an attempt to draw a connection between optical flow and the information available in the Compressed Domain.

1 Introduction

Image Change Detection is in essence a classification problem - to determine whether a change had occurred between images [11]. Video object segmentation is a subset of the problem - to distinguish different moving objects from the static background. For this end, motion information for each of the moving video objects must acquired before segmentation is performed. Traditional video object segmentation is performed in the pixel domain, in which pixel data are obtain from full decoding of the video bitstream. The motion flow is extracted by comparing consecutive frames and the basic data used is intensity value from each pixel.

However, the processing and storage overhead in decoding every frame from an encoded video bitstream prevents these methods from application in real-time applications. Video Object Segmentation in the Compressed Domain has gained interest because of its reduced computational and storage complexity compared to Pixel domain algorithms, making them applicable to real-time applications such as surveillance systems.

The term Compressed Domain in literature refers to video compression methods in which motion compensation and Discrete Cosine Transform (DCT) are used to reduce the number of bits required to represent a video, examples include MPEG-1/-2/-4, H.261 and H.263. All of these compression standards achieve compression by exploiting two observations. Firstly, it is unusual for intensity values to change frequently over a small area(spatial redundancy). Secondly, consecutive frames along time-ordered sequence of frames are similar(temporal redundancy). The Compressed Domain address the first observation with DCT and Quantization and the second with Motion Compensation. The products of the two processes an array of DCT coefficients and predicted motion vector(s) associated with each macroblock respectively. The DCT coefficients denoting the value of vertical and horizontal frequencies and the motion vectors the approximation of image motion, both of which could be easily obtained from a parsed bitstream [10].

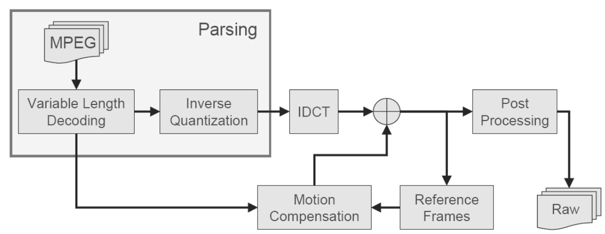

Figure 1. Parsing from MEPG bitstream[10]

For the following discussions, relatively small motion should exist in the input video bitstream, otherwise most of the macroblocks would be intracoded instead of intercoded because the encoders' inability to find an acceptable match in the search window.

2 Motion Accumulation and Median Filtering

The difficulty with the use of predicted motion vectors in video object segmentation is that the motion vectors are obtained to be the best-match of the reference frame rather than video object motion, and therefore not representative. A common approach to remove the outlying motion vectors is to accumulate the motion vectors from the bitstream over a few video frames then apply median filtering, used in [2],

[8], and [4]. However, the accumulation-and-filtering approach has two problems.

The first problem, addressed by Chen and Bajic [3], is that for repetitive motion (such as the bouncing motion of a ball) the motion vectors cancel out each other causing the accumulated motion to be of small magnitude and possibly undetected. This also leads to the discussion of the appropriate frame interval for accumulation. In [5], the accumulation interval is reduced to one (i.e. only using the motion vectors in the succeeding frame) if the average motion is estimated to be higher than a threshold. The second problem is that, while median filtering applied after motion accumulation would give a smooth motion field, we do not know if the accumulated motion field is contaminated by inaccurate motion vectors in the first place.

Porikli et al.'s investigation further suggests the of ineffectiveness of using motion accumulation alone for video segmentation. Porikli et al. [10] experimented with almost all of the information present in the Compressed Domain. The experimental results show that a slight over segmentation using DCT coefficients followed by aggregated motion based clustering produces more accurate boundaries than single stage joint segmentation. Also, using all of the DCT coefficients do not necessarily provide a stable segmentation in that the mean-shift algorithm becomes sensitive when AC components and spatial energy term are included. Ironically, the best combination stated above renders the system to segment video objects with similar average intensity value and texture, which in turn sensitive to intensity differences; in addition, the algorithm favours moderate motion since spatial-temporal volumes would be disjoint in the presence of motion larger than the area of segmented 2-D object.

As the accumulation-and-filtering approach is ineffective as it is unable to identify the validity of motion vector, some measure is required to ensure that the motion vector from the Compressed Domain is reliable to be used in video object segmentation. The motivation of carrying out this investigation comes from the publication by Coimbra and Davies [4] that draws the connection between information from the Compressed Domain and Lucas and Kanade optical flow method. The result is an accurate motion estimation scheme that is independent of GOP structure and approximates closely to a Lucuas-Kanade optical flow method.

Coimbra and Davies' discovery triggered the interest to find the connection between optical flow and the Compressed Domain, in particular the use of confidence measure.

3 Optical Flow

The first formal definition of optical flow is found in the publication by Horn and Schunck [6], in that "optical

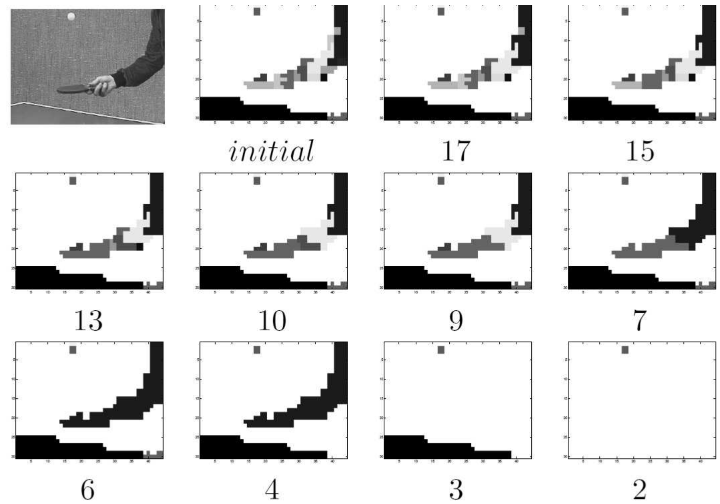

Figure 2. Porikli et al.'s segmentation results at the corresponding clustering levels. Note the volume growing process could not blend the lower part of the arm into other regions since its DCT coefficients were also significantly different.[10]

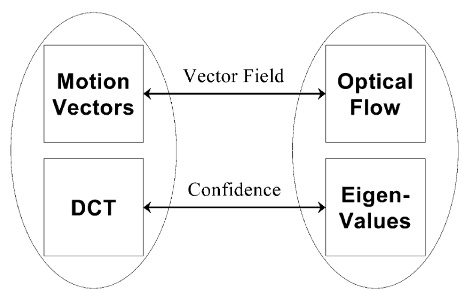

Figure 3. Comparison of the LK and MPEG-2 system[4]

flow is the distribution of apparent velocities of movement of brightness patterns in an image". According to [6], the optical flow problem is formulated as follows.

Let a brightness value at point (x,y) at time t be E(x,y,t), and the x- and y-component of optical flow be u and v respectively. For small motion (such that a point in the moving brightness pattern remains constant),

$$\frac{dE}{dx}u + \frac{dE}{dy}v + \frac{dE}{dt} = 0 \tag{1}$$

or,

$$\nabla I^T \cdot (u, v) = -\frac{dE}{dt} \tag{2}$$

where ∇ I is the gradient of image intensity.

Since the constraint that a point in the moving brightness pattern is constant is not enough to derive the value of (u,v) (often referred to as the Aperture Problem in literature), additional constraints has to be applied in addition to the above equation. For example, Horn and Schunck [6] introduced a smoothness constraint in that the velocity field of the brightness patterns in the image varies smoothly.

Following the classification in Barron et al.[7], methods of approximating optical flow is divided into four approaches:

1. Differential Techniques
2. Region-based Matching
3. Energy-based Methods
4. Phase-based Methods

In comparing the performance of optical flow techniques [7] emphasized on the accuracy of the optical flow measurements. They found out that in general the local differential approaches gives the most accurate results, with the method proposed by Lucas and Kanade [9] being the most accurate and least expensive. In addition, Barron et al.'s assessment [7] point out the importance of confidence measures and thresholds in their publication, stating that the use of confidence ensures the accuracy of the approximated optical flow fields.

The optical flow method proposed by Lucas and Kanade [9] aims to find the disparity vector h that minimizes the difference between the original image F(x) and the translated image G(x). Their generalized algorithm, which can register translation as well as rotation, scaling and shearing, can be expressed as

$$G(x) = F(xA + h) \tag{3}$$

where A is a matrix of linear transformations for each pixel inside the region of interest R, in order to find the disparity which minimizes the sum of squared differences, i.e.

$$\sum_x [F(xA + h) - G(x)]^2 \tag{4}$$

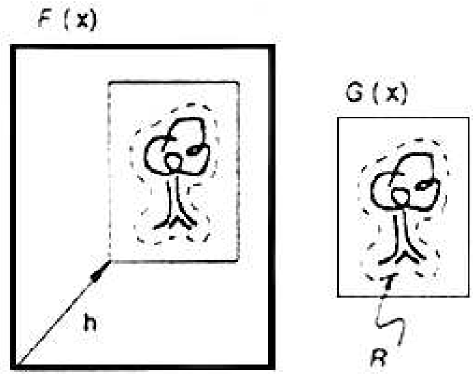

Figure 4. Lucuas and Kanade formulate the problem as the search for the disparity vector h which minimizes the difference between F(x + h) and G(x), for x in some region of interest R.[9]

4 Compressed Domain and Optical Flow

The process of Motion Compensation in video compression approximates optical flow calculation. The block-matching step, in that the encoder looks for the motion vector that gives the least difference between the reference and predicted macroblocks, resembles the optical flow approximation by Anandan [1] and Lucas and Kanade [9], in which both methods searches for the displacement with least error in the search window.

As a consequence, any confidence measure used in optical flow techniques could apply to the Compressed Domain. Coimbra and Davies [4] associates MPEG motion vectors and horizontal- and vertical-frequency DCT coefficients with optical flow and eigenvalues for confidence measure.

Unfortunately, this implies that the limitations in the optical flow methods also applies to Video Object Segmentation in the Compressed Domain. Aperture problem is present in all optical flow algorithms, and optical flow from less-textured image ares tends to be inaccurate. More importantly, sharp changes in intensities, such as occlusions and opening/closing of background lights. Such problems could be perhaps addressed separately, as in the Wallflower algorithm [12].

5 Conclusion

This paper presents the findings that are related to the acquisition of accurate Video Object Motion from the Compressed Domain. Many Compressed Domain video object segmentation algorithms involves the use of Motion Vectors in the input bit stream, which resembles the approximation of optical flow methods. While using these Motion Vectors saves the work of finding the optical flow for segmentation, it introduces the difficulties introduced by the fact that the motion vectors in the video bitstream does not necessarily reflect true motion.

Remedies have been introduced to deal with inaccurate Motion Vectors. A common approach is to accumulate Motion Vectors over a few picture frames then perform filtering to remove outliers. This approach, however, introduces problems such as motion cancelation and inclusion and inaccurate motion vectors. A better approach is to introduce confidence measures to remove potentially inaccurate Motion Vectors.

The application of optical flow methods in video object segmentation in the Compressed Domain, in particular the use of confidence measure, is an interesting topic and deserves further investigation.

References

[1] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2(3):283–310, January 1989.

[2] R. V. Babu, K. R. Ramakrishnan, and S. H. Srinivasan. Video Object Segmentation: A Compressed Domain Approach. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(4):462–473, April 2004.

[3] Y.-M. Chen and I. V. Bajic. Compressed-Domain Moving Region Segmentation with Pixel Precision using Motion Integration. In *IEEE Pacific Rim Conference on Computers and Signal Processing, 2009*, pages 442 – 447, August 2009.

[4] M. T. Coimbra and M. Davies. Approximating Optical Flow Within the MPEG-2 Compressed Domain. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(1):103–107, January 2005.

[5] K. Hariharakrishnan and D. Schonfeld. Fast Object Tracking Using Adaptive Block Matching. *IEEE Transactions on Multimedia*, 7(5):853–859, October 2005.

[6] B. K. P. Horn and B. G. Schunck. Determining Optical Flow. *ARTIFICAL INTELLIGENCE*, 17(1-3):185–203, August 1981.

[7] B. J.L., F. D.J., B. S.S., and T. Burkitt. Performance of optical flow techniques. pages 236–242, June 1992.

[8] L. Long, F. Xingle, J. Ruirui, and D. Yi. A Moving Object Segmentation in MPEG Compressed Domain Based on Motion Vectors and DCT Coefficients. In *Congress on Image and Signal Processing, 2008*, volume 3, pages 605–609, May 2008.

[9] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. pages 674–679, 1981.

[10] F. Porikli, F. Bashir, and H. Sun. Compressed Domain Video Object Segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*, 20(1):2–14, January 2010.

[11] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam. Image Change Detection Algorithms: A Systematic Survey. *IEEE Transactions on Image Processing*, 14(3):294–307, March 2005.

[12] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers. Wallflower: Principles and practice of background maintenance, 1999.