International Journal on Future Revolution in Computer Science & Communication Engineering
Volume: 4 Issue: 3
ISSN: 2454-4248
578 – 584

Study on Segmentation and Global Motion Estimation in Object Tracking Based on Compressed Domain

Kusuma T
Research Scholer, Computer Science and Engineering Department, Global Academy of Technlogy
Banagaluru, India

Dr Ashwini K
Research Supervisor, Computer Science and Engineering Department, Global Academy of Technlogy
Banagaluru, India

***Abstract*—** Object tracking is an interesting and needed procedure for many real time applications. But it is a challenging one, because of the presence of challenging sequences with abrupt motion occlusion, cluttered background and also the camera shake. In many video processing systems, the presence of moving objects limits the accuracy of Global Motion Estimation (GME). On the other hand, the inaccuracy of global motion parameter estimates affects the performance of motion segmentation. In the proposed method, we introduce a procedure for simultaneous object segmentation and GME from block-based motion vector (MV) field, motion vector is refined firstly by spatial and temporal correlation of motion and initial segmentation is produced by using the motion vector difference after global motion estimation.

Keywords- *Object tracking; Global Motion Estimation (GME); Segmentation, Motion Vector (MV) field.*

*****

I. INTRODUCTION

Global motion estimation (GME) and motion segmentation are two generally used techniques in video coding, computer vision, and content-based video analysis. GME estimates motion produced by camera movement in a video sequence and can be ready in either pixel domain or compressed domain. Compressed-domain approaches are computationally less demanding, since they utilize block based motion vectors (MVs) from the compressed bit stream. However, their accuracy may also suffer due to outliers in the MV field, caused either by imperfect motion estimation at the encoder, or by objects whose motion is different from the camera motion. Hence, identifying and removing outliers is one of the main challenges of compressed-domain GME. In the approach proposed here, motion segmentation is the cornerstone of outlier removal. In both these works, the flow is modeled as a Markov Random Field (MRF), and Bayesian segmentation is used to isolate regions that appear to move differently from each other. The method presented in this paper uses similar ideas about motion segmentation, specifically the MRF model and the Bayesian approach, but differs in a number of ways. First, our main goal is global motion estimation (GME), that is, the estimation of motion caused by camera movement. This motion is usually associated with the background, which is approximated as a flat surface far from the camera. Second, while the methods operate on the raw video in the pixel domain, our method is developed for compressed video to operate directly on MVs from the compressed bit stream, resulting in much lower complexity. Third, motion segmentation in our work is performed on the global motion-compensated MV field, that is, the field from which the estimated camera motion has been removed. Finally, the main purpose of motion segmentation in our work is to remove the MVs that belong to individual moving objects and thus improve the accuracy of GME, while motion estimation for individual objects is not explicitly considered.

Motion estimation and segmentation have been considered jointly in the context of optical flow estimation, as discussed above, in the literature on GME itself, motion estimation [3] and object segmentation are usually treated as two separate topics. A segmentation framework in compressed domain, where GM compensation is employed to obtain MV residuals prior to segmentation. In their approach, GME itself was conducted without moving object removal from the MV field. Pixel-domain GME basis for video object segmentation, where object information is obtained by performing GM compensation in the pixel domain, and used to predict outlier blocks for GME in the next frame.

The proposed approach couples object segmentation and GME on the block-based MV field, and introduce several contributions to the research in GME. First, motion segmentation offers a principled way to distinguish outlier MVs caused by moving objects or objects close to the camera, and leads to fast convergence of GM parameter estimates. Second, motion parameters are fed back to segmentation process to compensate for global motion, thus mitigating segmentation problems found in scenes with a moving camera. Third, the approach is applicable to any video bit stream compressed by a block-based standards-compliant encoder (e.g. H.264, etc.) since the MV field is the only information required. The proposed method has a higher degree of flexibility and portability compared to some compressed-domain approaches that rely on code specific information.

 ISSN: 2454-4248
578 – 584

II. PRIOR WORK ON COMPRESSED-DOMAIN SEGMENTATION AND TRACKING

Object Detection and Tracking in H.264/AVC Bitstreams: Many methods have been developed for moving object detection and tracking in H.264/AVC bitstream domain. A method based on partial decoding and initial object position determination is proposed in. Although additional information such as colors and initial position of the objects provide consistent detection results, it cannot be applied for automatic monitoring systems because this is a semi-automatic approach which obliges initial object position information by human interference. It performs automatic object detection by observing the bitstream in MB level.

As a result, this method might fail to detect small objects, specifically those of size smaller than an MB size of 16 pixels. Another method using partial decoding is proposed by [8] to detect moving vehicles in traffic surveillance recording. Since the method assumes that the objects always move in two opposite directions in two different lanes of roads, it may not be used for more general applications. Moreover, the use of partial decoding and background segmentation may require high computational complexity.

Compressed-Domain Segmentation and Tracking: An iterative pattern that combines Global Motion Estimation (GME) and Macroblock (MB) rejection is exploited. To identify moving object blocks, which are then tracked via MB-level tracking. This scheme, however, is not able to segment and track the moving objects whose motion is not sufficiently distinct from the background motion and estimate the trajectories of moving objects from H.264-AVC/SVC MVs. Foreground objects are identified by applying the background subtraction technique monitored by temporal filtering to remove the noise. Afterwards, motion segmentation is performed by Timed Motion History Images approach, and finally, the trajectory is estimated by object correspondence processing.

Mean shift clustering is used to segment moving objects from MVs and partition size in H.264/AVC bitstream. After obtaining salient MVs by applying spatial-temporal median filter and Global Motion Compensation (GMC), this method applies spatial-range mean shift to find motion-homogenous regions, and then smoothers the regions by temporal-range mean shift and presented an algorithm to track multiple moving objects in H.264/AVC compressed video based on probabilistic spatiotemporal MB filtering and partial decoding. Their work assumes stationary background and relatively slow-moving objects. We perform iterative backward projection for accumulating over time in order to obtain the salient MVs. The spatial homogenous moving regions are formed by statistical regions. In [11], the segmented regions are further classified temporally using the block residuals of GMC.

Another line of research addresses segmentation and tracking problems using Markov Random Field (MRF) models, which provide a meaningful framework for imposing spatial constraints. Treetasanatavorn *et al.* [11] suggested an algorithm for motion segmentation and tracking from MV fields through OC.

III. OVERVIEW OF THE PROPOSED METHOD

Global motion estimation (GME) and motion segmentation are two commonly used techniques in video coding [4], computer vision and content-based video analysis [7-9]. GME estimates motion caused by camera movement in a video sequence [1], and can be done in either pixel domain [2] or compressed domain [4]. Identifying and removing outliers is one of the main challenges of H.264/AVC compressed-domain GME. In the approach proposed here, motion segmentation is the cornerstone of outlier removal. In both these works, the flow is modeled as a Markov Random Field (MRF), and Bayesian segmentation is used to isolate regions that appear to move differently from each other. The method presented in this paper uses similar ideas about motion segmentation, specifically the MRF model and the Bayesian approach, but differs in a number of ways. First, our main goal is global motion estimation (GME), that is, the estimation of motion caused by camera movement. This motion is usually associated with the background, which is approximated as a flat surface far from the camera. Hence, for this purpose, an 8-parameter perspective motion model [6], which we use here, is more appropriate than the 6-parameter affine model used in. Second, while the methods in [8] and [9] operate on the raw video in the pixel domain, our method is developed for compressed video to operate directly on MVs from the compressed bit stream, resulting in much lower complexity. Third, motion segmentation in our work is performed on the global motion-compensated MV field, that is, the field from which the estimated camera motion has been removed.

Finally, the main purpose of motion segmentation in our work is to remove the MVs that belong to individual moving objects (i.e., outliers) and thus improve the accuracy of GME, while motion estimation for individual objects is not explicitly considered. Although motion estimation and segmentation have been considered jointly in the context of optical flow estimation, as discussed above, in the literature on GME itself, motion estimation [2] and object segmentation are usually treated as two separate topics. Notable exceptions, where some connection between GME and segmentation is made include [11] and [10]. In [10], Liu *et al*. proposed a segmentation framework in compressed domain, where GM compensation is employed to obtain MV residuals prior to segmentation. In their approach, GME itself was conducted without moving object removal from the MV field. GME and motion

 ISSN: 2454-4248
578 – 584

segmentation are securely coupled in the proposed approach, Motion segmentation helps remove MV outliers from the input MV field, leading to more accurate GME. Global motion is then used to perform GM compensation, which leads to more accurate motion segmentation on the GM-compensated MV field. The proposed method evaluates the labeling dependence of consecutive frames through MVs as well as the similarity of context and neighboring MVs within the frame, and then assigns MVs into the most probable class (object vs. non-object), as measured by the posterior probability. A two-dimensional vector used for inter prediction that provides an offset from the coordinates in the decoded picture to the coordinates in a reference picture measured by Posterior probability is a conditional probability conditioned on randomly observed data. Hence it is a random variable. The aim of this paper is to provide a framework for tracking moving objects in compressed domain based on MVs and associated block coding modes alone. The object of interest is selected by the user in the first frame, and then tracked through subsequent frames as shown in figure 1

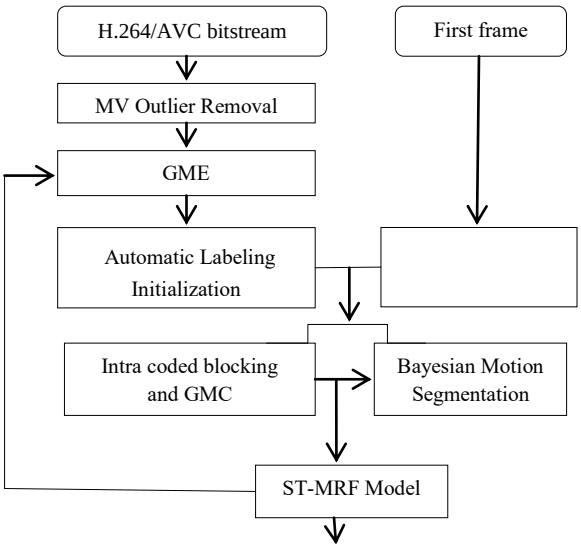


Figure 1: Flow chart of the proposed method

IV. GLOBAL MOTION AND GLOBAL ESTIMATION

Information is very important for video content analysis. In surveillance video, usually the camera is stationary, and the motions of the video frames are often caused by local motion objects. Thus detecting motions in the video sequences can be utilized in anomalous events detection, in sports video, the heavy motions are also related to highlights. Motion estimation and compensation is the fundamental of video coding. Coding the residual component after motion compensated can save bit-rates significantly. In video sequences, the motion configuration can be classified into two types: Local motion and Global motion.

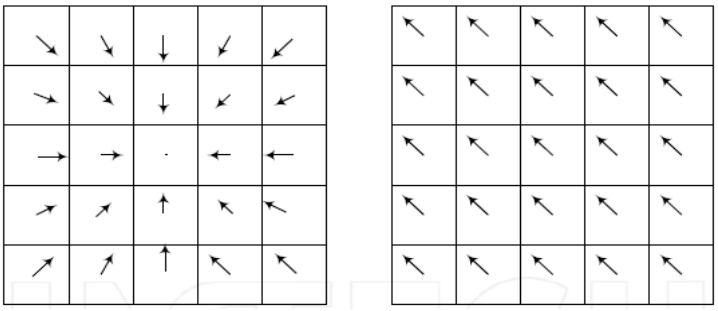
A. Global Motion

Global motions in a video sequence are caused by camera motion, which can be demonstrated by parametric transforms. The process estimating the transform parameters is called global motion estimation [4]. The global motions have certain consistence for the complete frame as shown in below figure. The global motion in figure (a) is a zoom out and that in figure (b) is a translation respectively.

We find that the motion track is from outer to inner regions, which means that the coordinates of a current frame t can be generated in the inner regions of the reference frame V(t>v) . In figure 2, the motion vectors in the motion field correspond to the global motion vectors at the coordinates.

Global motion vector is the motion vector calculated from the estimated global motion parameters. Global motion vector ($GMVx_t$, $GMVy_t$ **)** for the current pixel with its coordinates (x_t, y_t) is determined as

$$GMVx_t = x^1 - x_t \quad (1)$$
$$GMVy_t = y^1 - y_t$$



. (a) Zoom-out (b) Translation

Figure 2: Global motion fields. (a)Zoom out and (b) Translation

B. Global Motion Models

Global motion can be represented by global motion models with several parameters. The simplest global motion model is conversion with only two parameters. The complex global motion model is quadric model with 12 parameters, which is expressed as follows.

$$x^1 = \frac{m_3 x + m_4 x + m_5}{m_6 x + m_7 y + 1} \quad (2)$$

$$y^1 = \frac{m_0 x + m_1 x + m_2}{m_6 x + m_7 y + 1}$$

 ISSN: 2454-4248
578 – 584

C. *Global Motion Estimation (GME) approaches*

Global motion estimation can be carried out in pixel domain. In the pixel domain based approaches, all the pixels are involved in the estimation of global motion parameters. There are two short comings in pixel domain based approach: 1) it is very computational intensive; 2) it is often sensitive to noises (local object motions). In order to improve the convergence and speed up the calculation, coarse to fine searching approach is often adopted. Moreover, the subset of pixels having the largest gradient magnitude is adopted to estimate the global motion parameters [6]. Sub-point based global motion estimation approaches are very effective in reducing computational costs. To guarantee the accuracy of global motion estimation, how to determine the optimal sub-sets are the key steps. Except the pixel domain based global motion estimation, compressed domain based global motion estimation approaches are also very popular.

D. *Pixel domain based GME*

In GME involving two image frames Ik and Iv (with k<v), one seeks to minimize the following sum of squared differences between *Iv* and its predicted image Ik (x(i, j), y(i, j)) which is obtained after transforming all the pixels in Ik.

$$E = \sum_j e(i,j)^2 \quad (3)$$

where e(i, j) denotes the error of predicting a pixel located at (i, j) of frame Iv, by using a pixel at location [x(i, j), y(i, j)] of previous frame Ik.

$$e(i,j) = I_v(i,j) - I_k(x(i,j), y(i,j)) \quad (4)$$

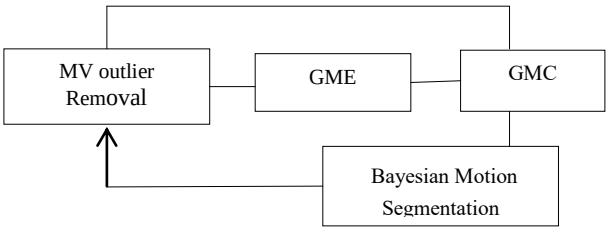


Figure 3: The system diagram of joint approach to compressed domain global motion estimation and motion segmentation. The framework offers GM parameters and video moving object information

F. *Compressed domain based GME*

In video coding standards, the motion estimation algorithms calculate the motions between successive video frames and guess the current frame from previously transmitted frames using the motion information. Hence, the motion vectors have some relationship with the global motion. A global motion estimation method is proposed based on randomly selected MV groups from motion vector field with adaptive parametric model determination. A non-iterative GME approach is proposed by Su *et al.* by solving a set of Exactly-determined matrix equations corresponding to a set of motion vector groups [4]. Each MV group consists of four MVs selected from the MV field by a fixed spatial pattern.

The global motion parameters for each of the MV group are obtained by solving the exactly determined matrix equation using singular value decomposition (SVD) based pseudo inverse technique. The final global motion parameters are obtained by a weighted histogram-based method. Moreover, a least-square based GME method by coarsely sampled MVs from the input motion vector field is proposed for compressed video sequences. The global motion parameters are optimized by minimizing the fitting error between the input motion vectors and the wrapped ones from estimated global motion parameters. In order to estimate global motions robustly, motion vectors in local motion region, homogeneous region with zero or near-zero amplitude and regions with larger matching errors are rejected.

E. *GMC and LMC based video coding*

The objective of this part is to illustrate how video compression performances can be improved by utilizing adaptive GMC/LMC mode determination. GMC/LMC based motion compensation mode selection approach in H.264/AVC is given [1], [2]. Global motion estimation and compensation is used in H.264/AVC advanced simple profiles (ASP) to remove the residual information of global motion. Global motion compensation (GMC) is a new coding technology for video compression in H.264/AVC standard. By extracting camera motion, H.264/AVC coder can remove the global motion redundancy from the video. In H.264/AVC 4 each macro block (MB) can be selected to be coded use GMC or local motion compensation (LMC) adaptively during mode determination. Intuitively, some types of motion, e.g., panning, zooming or rotation; could be described using one set of motion parameters for the entire VOP (video object plane). For example, each MB could potentially have the exact same MV for the panning. GMC allows the encoder to pass one set of motion parameters in the VOP header to describe the motion of all MBs. Additionally H.264/AVC allows each MB to specify its own MV to be used in place of the global MV. In H.264/AVC Advanced simple contour, the main target of Global Motion Compensation (GMC) is to encode the global motion in a VOP (video object plane) using a small number of parameters.

Each MB can be predicted either from the previous VOP by global motion compensation (GMC) using warping parameters or from the previous VOP by local motion compensation (LMC) using local motion vectors as in the classical scheme. The selection is made based on which predictor leads to the lower prediction error. In this Section we only expressed the GMC mode selection approach.

 ISSN: 2454-4248
578 – 584

V. MOTION SEGMENTATION

In this proposed work, the coder may perform a locally defined pre-processing, aimed at the identification of the objects appearing in the sequence. Hence, segmentation aiming at the generation of video objects is a key issue in efficiently applying the H.264/AVC coding scheme, however not affecting at all the bit-stream syntax and thus not being a normative part of the standard. From the very beginning there have been activities within a so-called fundamental experiment, aiming at the development of segmentation algorithms which are able to usually extract video objects from a captured scene. The examined approaches perform a classification of the pels in a video sequence into two classes, namely moving objects (foreground) and background.

5.1. Description of a combined temporal and spatial segmentation framework

Throughout the work on spontaneous segmentation of moving objects, different proposals for temporal and spatial segmentation algorithms have been proposed and investigated. This resulted at the end in a combined temporal and spatial segmentation framework which is shown in a high level block diagram in Figure 4.

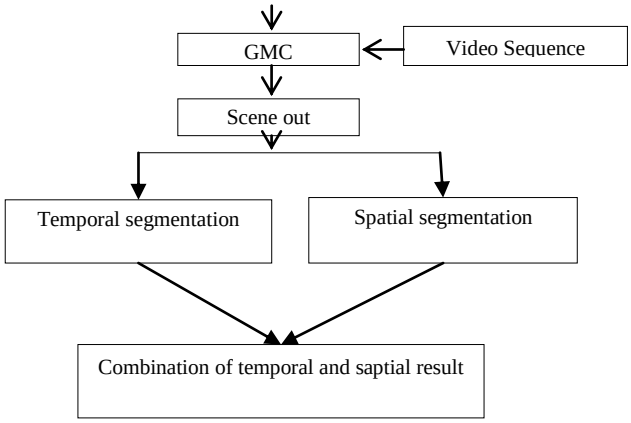


Figure 4: Combined temporal and spatial segmentation framework

5.2. Temporal segmentation based on change detection

The algorithm is mainly based on change detection. It can be subdivided into two main steps, assuming that a possible camera motion has already been compensated:

- Change detection mask between two successive images is estimated. Therefore, first an initial change detection mask between the two successive images is generated by global thresholding.
- Boundaries of changed image areas are smoothed by a relaxation technique using local adaptive thresholds. Thereby, the algorithm adapts frame-wise automatically to camera noise. In order to finally get temporal stable object regions, an object mask memory with scene adaptive memory length is applied.
- The mask is simplified and small regions are eliminated, resulting in the final change detection mask.
- Finally, an object mask is calculated by eliminating the uncovered background areas from the change detection mask.

5.3. Temporal segmentation using higher order moments and motion tracking

The algorithm produces the segmentation map of each image of the sequence by processing a group of images. The number of images n varies on the basis of the estimated object speed. For each image, the algorithm consists of three steps:

First, the differences of each image of the group with respect to the first image are evaluated in order to detect the changed areas, due to object motion, uncovered background and noise. In order to reject the luminance variations due to noise, a Higher Order Statistic (HOS) test is performed.

For each picture element, an estimated displacement is evaluated on a 3x3 window; if the displacement is not null the pixel is classified as moving. Finally, a regularization algorithm re-assigns still regions, internal to moving regions, to foreground and refines the segmentation results imposing a priori topological constraints on the size of objects irregularities such as holes, is the mini, gulfs and isles by morphological filtering. A post-processing operation refines the results on the basis of spatial edges.

RESULTS

A number of standard test sequences were used to estimate the performance of our proposed approach. The GOP structure IPPP, i.e., the first frame is coded as intra (I), and the subsequent frames are coded predictively (P). Motion and partition information were extracted from the compressed bitstream, and MVs were remapped to 4 × 4 blocks, as explained. Some of the features of the proposed approach are its robustness and stability. To show this, we use the *same* parameters throughout all the experiments and found that the average performance does not change much if some of these parameter values are changed, especially parameters represent the effect of camera motion on energy function, and only affect a few frames. Figure5 illustrates a few intermediate

 ISSN: 2454-4248
578 – 584

results from the tracking process for a sample frame from *Coastguard*. As s combined temporal and spatial segmentation framework from figure 5(a), the MV field around the target (small boat) is somewhat erratic, due to fast camera motion in this part of the sequence. The proposed ST-MRF-based tracking algorithm computes the energy function for the chosen MRF model, which is shown in figure 5(b). Here, the darker the value, the smaller the energy, hence the higher the posterior probability. Therefore, despite the erratic MV field, the target seems to be localized reasonably well, figure 5(c) shows the detected target region after the tracking process has been completed

A number of standard test sequences were used to estimate the performance of our proposed approach. The GOP structure IPPP, i.e., the first frame is coded as intra (I), and the subsequent frames are coded predictively (P). Motion and partition information were extracted from the compressed bitstream, and MVs were remapped to 4 × 4 blocks, as explained. Some of the features of the proposed approach are its robustness and stability..

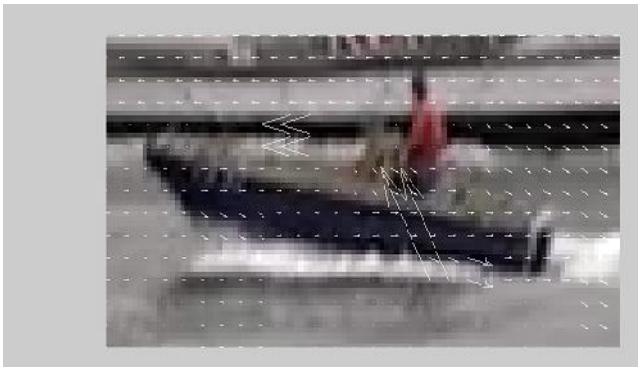


Figure 5(a): Object detection during ST-MRF-based tracking, target at frame #70 of *Coastguard* superimposed by scaled MV field after GMC.

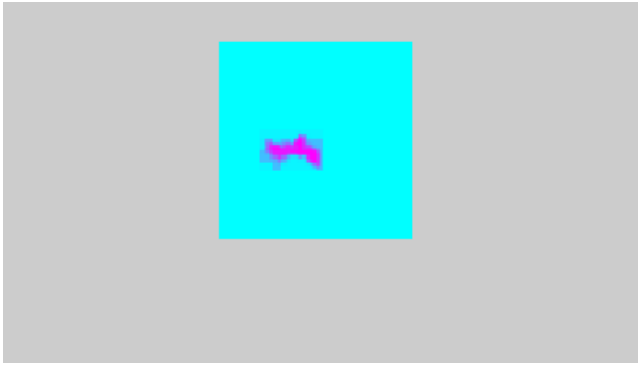


Figure 5(b): MRF energy value—the darker the color, the higher the energy

Figure 5(c): Tracking results by the proposed method.

Figure 5(d): Moving region segmentation from compressed video using global motion estimation

To show this, we use the *same* parameters throughout all the experiments and found that the average performance does not change much if some of these parameter values are changed, especially parameters represent the effect of camera motion on energy function, and only affect a few frames. Figure 5 illustrates a few intermediate results from the tracking process for a sample frame from *Coastguard*. As seen from figure 5(a), the MV field around the target (small boat) is somewhat erratic, due to fast camera motion in this part of the sequence. The proposed ST-MRF-based tracking algorithm computes the energy function for the chosen MRF model, which is shown in figure 5(b). Here, the darker the value, the smaller the energy, hence the higher the posterior probability. Therefore, despite the erratic MV field, the target seems to be localized reasonably well, figure 5(c) shows the detected target region after the tracking process has been completed. For comparison purposes, the segmentation result from two other methods is illustrated in figure 5(d) and figure 5(e), respectively

 ISSN: 2454-4248
578 – 584

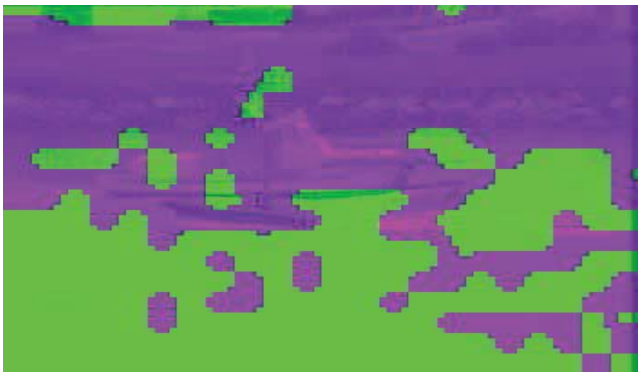


Figure 5 (e): "Robust moving object segmentation on H.264/AVC compressed video

CONCLUSION

Global motion estimation (GME) and motion segmentation are two generally used techniques in video coding, computer vision, and content-based video analysis. The proposed method has a higher degree of flexibility and portability compared to some compressed-domain approaches that rely on code specific information. The proposed approach is demonstrated to reliably segment moving objects with good quality from compressed video.

REFERENCES

[1] C. Aeschliman, J. Park, and A. C. Kak, "A probabilistic framework for joint segmentation and tracking," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, San Francisco, CA, Jun. 2010, pp. 1371–1378

[2] [*Su, M. T. Sun, "A Non-iterative motion vector based global motion estimation algorithm", Proc. Int. Conf. Multimedia and Expo, Taipei, Taiwan, June 27-30, 2004,pp.703-706.*

[3] Z. Kato, T.-C. Pong, and J. C.-M. Lee, "Color image segmentation and parameter estimation in a Markovian framework," *Pattern Recognit. Lett.*, vol. 22, nos. 3–4, pp. 309–321, 2001.

[4] Y. P. Su, M. T. Sun, "A Non-iterative motion vector based global motion estimation algorithm", Proc. Int. Conf. Multimedia and Expo, Taipei, Taiwan, June 27-30, 2004, pp.703-706.

[5] H. Wang, J. Wang, Q. Liu, H. Lu, Fast progressive model refinement global motion estimation algorithm with prediction, in proc. ICME 2006.

[6] H. Alzoubi, W. Pan, "very fast global motion estimation using partial data", in Proc. ICASSP 2006.

[7] H. Alzoubi, W. Pan, "efficient global motion estimation using fixed and random subsampling patterns".

[8] Y. Hu and T. J. Dennis, "Simulated annealing and iterated conditional modes with selective and confidence enhanced update schemes," in *Proc. 5th Annu. IEEE Symp. Comput.-Based Med. Syst.*, Jun. 1992, pp. 257– 264.

[9] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. 6, pp. 721–741, Nov. 1984.

[10] Y.-M. Chen and I. V. Bajic, "A joint approach to global motion estimation and motion segmentation from a coarsely sampled motion vector field," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 9, pp. 1316–1328, Sep. 2011.

[11] R. Huang, V. Pavlovic, and D. N. Metaxas, "A new spatio-temporal MRF framework for video-based object segmentation," in *Proc. MLVMA Conjunct. Eur. Conf. Comput. Vis.*, 2008.