

# 3D-MRF BASED VIDEO TRACKING IN THE COMPRESSED DOMAIN

S. Vinci<sup>1</sup>, V. Josephine Sutha<sup>2</sup>

<sup>1</sup>PG Scholar, Department of Computer Science Engineering, SRCE, Tamil Nadu, India

<sup>2</sup>Assistant Professor, Department of Computer Science and Engineering, SRCE, Tamil Nadu, 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, drastic illumination change, large pose variation, occlusion, cluttered background and also the camera shake. This paper presents a novel method of object tracking by using the algorithms spatio-temporal Markov random field (ST-MRF) and online discriminative feature selection (ODFS), which overcome the above mentioned problems and provide a better tracking process. This method is also capable of tracking multiple objects in video sequence even in the presence of an object interactions and occlusions that achieves better results with real time performance.

**Keywords:** Video object tracking, spatio-temporal Markov random field (ST-MRF), online discriminative feature selection (ODFS).

\*\*\*

## 1. INTRODUCTION

In the past decades many algorithms were proposed for the object tracking process but it is still difficult to make a efficient method of tracking. Appearance changes created by abrupt motion, shape deformations, illumination variations and also the occlusion all these makes the tracking so tricky. Also the applications such as video indexing and editing, Automated surveillance, traffic monitoring and human-computer interaction are needed for better results.

Pixel domain and compressed domain are the few major groups of approaches to track moving objects in a video sequence. Pixel domain will provide high accuracy. It needs higher computational complexity. It also require decoding to get pixel domain information, because nowadays many of the video contents are available only in compressed form. So compressed domain approach is widely used. This approach uses the data from compressed video bit stream like block coding modes vectors, motion compressed prediction residual or its transform coefficients etc.

Based on their representation schemes tracking algorithms are categorized into two classes: 1) Generic 2) Discriminative models. Generative algorithms study on appearance model and use that for search the image regions thus provide tracking results with minimal reconstruction errors. Sparse representation methods are used to represent the object using trivial templates and a set of target to deal with pose variation and partial occlusion. But these generative models discard many of useful information, because it does not take surrounding visual context into account.

Discriminative models use a learned classifier to separate the target object from its surrounding background with in a local region. Collins et al.[1] demonstrate in an online manner to improve tracking performance. Grabner et al.[2] demonstrated on online boosting feature selection method. But these all used only one positive sample and multiple negative samples. If the detected object location is not precise. The positive samples for tracking with the MIL framework [3][4]. This paper demonstrates the object tracking with MRF algorithm combined with ODFS algorithm in the compresses domain.

## 2. OBJECT TRACKING

Object tracking is nothing but the process of extracting an object of interest from a video scene and continuous track of its motion, orientation and occlusion. It follows the segmenting step which is same as recognition step in image processing. First main step of information extraction is deletion of moving objects from video streams. Three approaches are there for object tracking. One is feature based approach, which extracts the characteristics like points, line segments. Differential methods of approach are mainly based on optical flow computation. The third approach uses the correlation to read the image displacements. Based on the domain of the problem any of these approaches its selected.

## 2.1 Overview of the Proposed Method

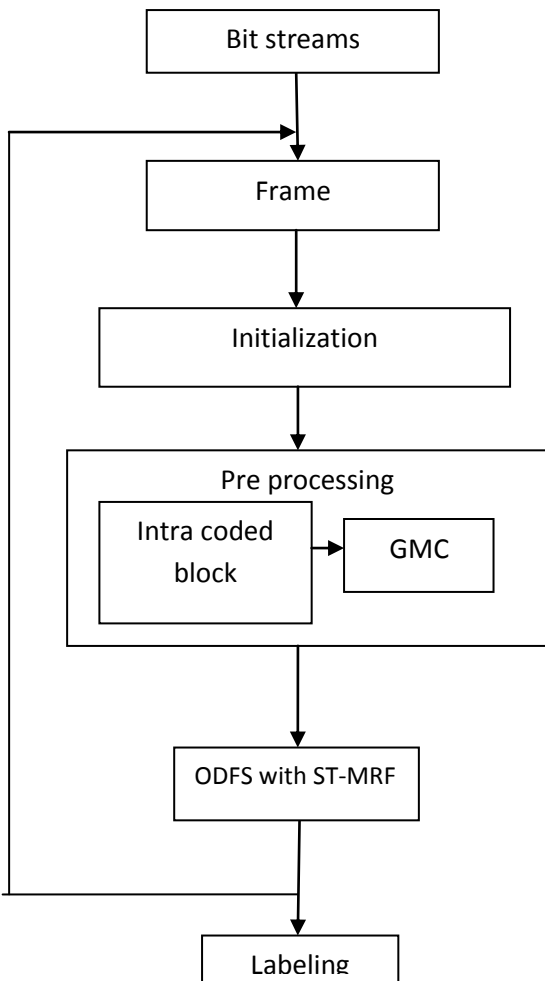


Fig -1: Flow diagram for tracking

## 2.2 Preprocessing

Tracking algorithm makes use of two types of information from the H.264/AVC compressed bit stream: MVs and block coding mode (partition) information. In this method texture data does not require to be decoded. Four basic MB modes are defined in the above mentioned bit stream:  $16 \times 16$ ,  $16 \times 8$ ,  $8 \times 16$ , and  $8 \times 8$ , here the  $8 \times 8$  mode can be again split into  $8 \times 4$ ,  $4 \times 8$ , and  $4 \times 4$  modes. Being  $4 \times 4$  is the smallest coding mode (partition) in H.264/AVC, so as to have a uniformly sampled MV field, this method map all MVs to  $4 \times 4$  blocks. This is simple in inter-coded blocks, as well as SKIP blocks where the MV is simply set to zero. But the motion interpretation in the intra-coded blocks is more involved.

## Global Motion Compensation

In a frame all the pixels will be affected by the global motion, which is caused because of camera movements. To get an accurate tracking, global motion should be removed because it adds to the object's native motion. Global Motion should be compensated from the motion vector field prior to further processing. 6-parameter affine model [8] and 8-parameter perspective model can be used to represent global motion. The affine model is less flexible than the 8-parameters perspective model.

Based on the single affine transform instruction, Global motion compensation describes the motion in a scene. The selected frame is rotated, panned and zoomed in accordance to GMC warp points to produce a prediction about the following frame. This operation can create an accurate prediction because it works on individual pixels. But the block based approaches cannot create such predictions. Every block in a frame can be compensated using global motion or local motion. A predicted frame can be called as sprite frame (s-frame) when it is created by using GMC. If it is encoded without GMC it can be either P-frame or B-frame. Frame predicted jointly by past and future frame is B-frame. Camera motions like tilt, roll, track, boom are basically reflected in the global motion compensation. Normally large blocks ( $8 \times 8$  and above) are more likely to be part of the background, whereas small blocks, are more likely to belong to the moving objects. In GME, only large blocks ( $16 \times 16$ ,  $16 \times 8$ ,  $8 \times 16$ ,  $8 \times 8$ ) are used, and so small blocks ( $8 \times 4$ ,  $4 \times 8$ ,  $4 \times 4$ ) and intra-coded blocks are discarded

## 2.3 ST-MRF Optimization

The characteristics of a moving rigid object are relative similarity of motion within the region occupied by the object, spatial compactness i.e. not dispersed in different parts of the frame, and a continuous motion trajectory. Though the motion of flexible objects is not so characteristic it is in principle that these objects can be treated in a divide and conquer manner as a group of smaller sufficiently rigid objects. So this ST-MRF model is based on rigid object motion characteristics. The frame should be divided into smaller blocks (i.e.  $4 \times 4$ ). Object blocks can be labeled 1 and non-object blocks can be labeled as 0. The block in  $t$ th frame labeled as  $\omega_t$  and the block in frame  $t-1$  is labeled as  $\omega_{t-1}$ . In past two algorithms Stochastic Relaxation (SR) and ICM were used. SR has some advantage in accuracy compared to ICM, but at a higher computational cost [11]. At first, the label of each block is initialized by projecting the previous frame labeling  $\omega_{t-1}$  into the current frame. Then each block is relabeled with the label (0 or 1) that leads to the largest reduction in the energy function. Relabeling procedure is done until no further energy reduction is achieved. Normally, six iterations are needed to reach the local minimum. This is worth mentioning that because results are dependent on the initial labeling.

## 2.4 ODFS

This method proposes an efficient and robust tracking algorithm which addresses all the issues faced by existing methods of tracking. It proposes a simple and effective online discriminative feature selection (ODFS) approach which directly couples the classifier score with the sample importance, and thus formulating a more robust and efficient tracker than state-of-the-art algorithms [11], [13], and 17 times faster than the MIL Track [14] method. It also shows that use of bag likelihood loss functions is unnecessary for feature selection which is proposed in the MIL Track method. As an alternative, this method directly selects features on the instance level by using a supervised learning method which is more efficient and robust than the MILTrack method. As the correct positive one and all the instances, can be labeled from the current classifier, they can be used for update via self-taught learning. Here, the correct positive instance can be effectively used as the tracking result of the current frame as similar to other discriminative models [9], [10], [11], [12].

## 3. RESULTS

### 3.1 Object Representation

In a tracking scenario, an object can be defined as anything that is of interest for further analysis. For instance, boats on the sea, fish inside an aquarium, vehicles on a road, planes in the air, people walking on a road, or bubbles in the water are a set of objects that may be important to track in a specific domain. Objects can be represented with use of their shapes and appearances.

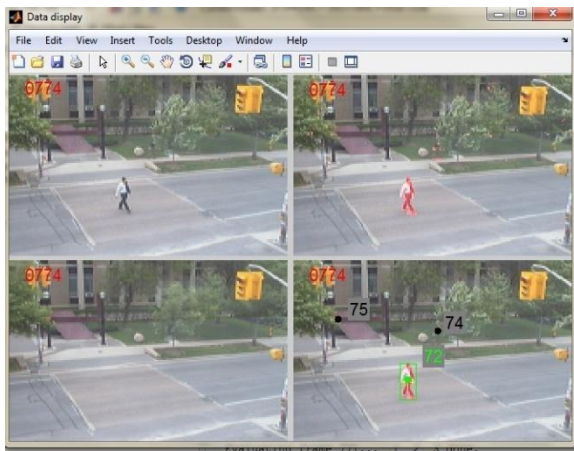


Fig -2: Proposed Method Of Tracking

### 3.2 Feature Selection for Tracking

Selecting the right features plays a critical role in tracking. In general, the most desirable property of a visual feature is its uniqueness so that the objects can be easily distinguished in the feature space. Feature selection is most related to the object representation. For example, in histogram based

appearance representations color is used as a feature, where as for contour-based representation, object edges are usually used as features. Many tracking algorithms use a combination of these features.

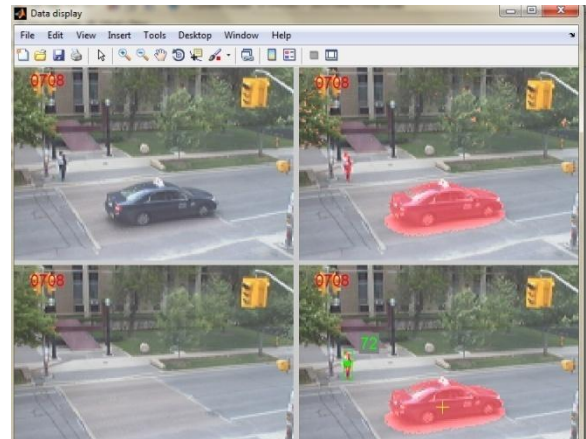


Fig -3: Feature Selection for Tracking

## 4. CONCLUSIONS

This algorithm provides a novel approach to track a moving object in -compressed video. The only data from the compressed stream used in the proposed method are the motion vectors and block coding modes. As a result, the proposed method has a fairly low processing time, yet still provides high accuracy. After the preprocessing stage, which consists of intra-coded block motion approximation and global motion compensation, an estimate of the labeling of the current frame is formed based on the previous frame labeling and current motion information. The results of experimental evaluations on ground truth video demonstrate superior functionality and accuracy of our approach against other state-of-the-art compressed-domain tracking Approaches. Although our algorithm works well even with fixed parameter values, possibly better performance may be obtained by adaptive tuning, although this would in general increase the complexity.

## 5. REFERENCES

- [1]. T. Drummond and R. Cipolla, "Real time visual tracking of complex structures," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol., 24, no. 7, pp. 932-946, Jul. 2002.
- [2]. 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.
- [3]. W. Fei and S. Zhu, "Mean shift clustering-based moving object segmen-tation in the H.264 compressed domain," *IET Image Process.*, vol. 4, no. 1, pp. 11-18, Feb. 2010.
- [4]. Y.-M. Chen, I. V. Bajic, and P. Saeedi, "Moving region segmentation from compressed video using global motion

estimation and Markov random fields,” IEEE Trans. Multimedia, vol. 13, no. 3, pp. 421-431, Jun. 2011.

[5]. 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.

[6]. 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.

[7]. R. Szeliski, R. Zabih ,D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, "A comparative study of energy minimization methods for Markov random fields," in Proc. Eur. Conf. Comput. Vis., LNCS 3952. 2006, pp. 16-29.

[8]. R. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2004, pp. 39-44.

[9]. S. Avidan, "Support vector tracking," IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 8, pp. 1064-1072, 2004. 1 ,2 ,5

[10]. R. Collins, Y. Liu, and M. Leordeanu, "Online selection of discriminative tracking features," IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 10, pp. 1631-1643, 2005. 1 ,2 ,3 ,5 ,6.

[11]. H. Grabner, M. Grabner, and H. Bischof, "Real-time tracking via online boosting," In British Machine Vision Conference, pp. 47-56 , 2006. 1, 2,5 ,6,7.

[12]. S. Avidan, "Ensembletracking," IEEE Trans. Pattern Anal. Mach. Intell., vol. 29,no. 2,pp. 261-271, 2007. 1 ,2 ,5 ,6.

[13]. S. Hare, A. Saffari, and P.Torr, "Struck: structured output tracking with kernels," In Proc. Int. Conf. Comput. Vis., 2011. 1 ,2 ,6 ,7.

[14]. B. Babenko, M.-H. Yang, and S. Belongie, "Robust object tracking with Online multiple instance learning," IEEE Trans. Pattern. Anal. Mach. Intell., vol.33, no.8, pp. 1619-1632, 2011. 1 ,2 ,3 ,6 ,7 ,9 ,11.