

Dominant Flow Extraction and Analysis in Traffic Surveillance Videos

Srinivas S S Kruthiventi and R. Venkatesh Babu
Video Analytics Lab, SERC,
Indian Institute of Science, Bangalore, India

srinivask@ssl.serc.iisc.in, venky@serc.iisc.ernet.in

Abstract

Flow analysis of crowd and traffic videos is an important video surveillance task. In this work, we propose an algorithm for long-term flow segmentation and dominant flow extraction in traffic videos. Each flow segment is a temporal sequence of image segments indicating the motion of a vehicle in the camera view. This flow segmentation is done in the framework of Conditional Random Fields using motion and color features. We also propose a distance measure between any two flow segments based on Dynamic Time Warping and use this distance for clustering the flow segments into dominant flows. We then model each dominant flow by generating a representative flow segment, which is the mean of all the time-warped flow segments belonging to its cluster. Using these dominant flow models, we perform path prediction for the vehicles entering the view and detect anomalous motions. Experimental evaluation on a diverse set of challenging traffic videos demonstrates the effectiveness of the proposed method.

1. Introduction

Nowadays, we have millions of cameras, all around the world, in public places like offices, traffic junctions, railway stations etc., capturing video data round the clock. This has resulted in an increased need for research in automated analysis of surveillance videos, as it is impossible to monitor this humongous data manually. Research in surveillance automation has given rise to solutions for some of the classic problems in computer vision such as tracking, detection, recognition etc. In this work, we propose an algorithm for extracting long-term flow segments of vehicles and performing flow analysis on the traffic surveillance videos.

A flow in a video can be defined as a temporal sequence of image segments which are continuous and have a coherent motion along the length of the sequence. In the problems of flow analysis, motion becomes a key feature as it captures the essential temporal information of moving objects. Works such as Particle Video [22], Dense Point Tra-

jectories [23] etc. effectively characterize the long-term motion in videos. These are primarily particle-based approaches, where the particles are initialized in regions with rich texture and are propagated from frame to frame using the optical flow and appearance consistency criterion. In this work, we characterize the long-term motion of an object in the video as a sequence of image segments. Each image segment is a temporal instance of a homogeneously moving connected component. The union of these coherently moving image segments forms the long-term path for the connected component.

We refer to an entire sequence of the image segments corresponding to a single connected component as a long-term flow segment. Our work attempts to extract these flow segments for all the moving objects in the view frustum of the camera simultaneously rather than tracking/segmenting each independently. This task is different from the problem of object segmentation considering that we do not require the exact boundaries of moving objects but only smooth spatio-temporal volumes with homogeneous flow representing the motion present in the video.

This task of extracting multiple flow segments simultaneously is achieved in a frame-by-frame manner by propagating the flow segments from the current frame to the next frame and initializing any new flow segments on the new frame. Our algorithm achieves this using the framework of Conditional Random Fields (CRF) by defining costs for background, new flow segments and the existing flow segments. CRF [15] is a probabilistic model which finds an optimal labeling over structured data by minimizing a defined global energy under the conditions of local smoothness. CRF has been extensively used in vision research in the last two decades to deal with the problems of structured prediction [10, 19, 20, 27].

We also develop a method for clustering these extracted flow segments using a distance measure based on Dynamic Time Warping (DTW). DTW finds a mapping between two sequences such that the similarity between the mapped elements is maximized. Each of the obtained clusters represent a dominant flow in the video (Fig. 1). We create a model to

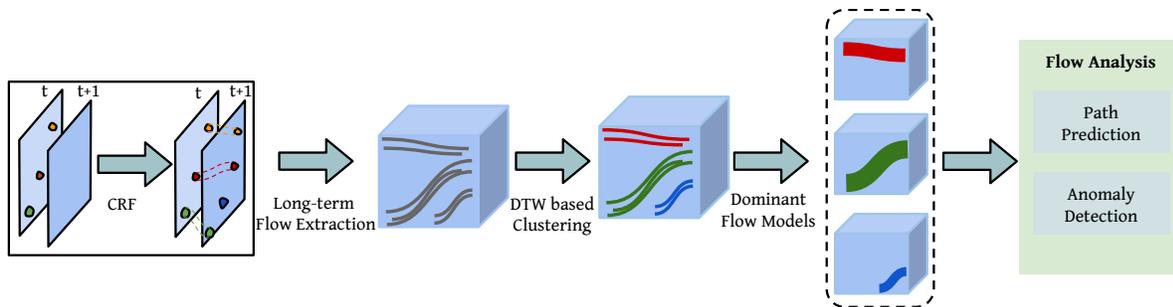


Figure 1: Overview of the proposed method

represent the flow segments of each dominant flow cluster by mean pooling the time warped samples. Further, we illustrate the applications of our model in predicting the path of a vehicle entering the camera field of view. Using the learnt dominant models, any anomalous movement i.e., a vehicle moving along an unusual path, can also be detected.

The rest of this paper is organized in the following manner: Section 2 discusses recent related work. Section 3 details the proposed method of extracting long-term flow segments from a video and clustering them to obtain dominant flows. We also discuss its applications: path prediction, anomaly detection. Section 4 discusses the experimental evaluation of the proposed method. We conclude with a short summary of the proposed method in Section 5.

2. Related Work

Recently there has been a good amount of work in characterizing and analyzing motion patterns in surveillance videos. These approaches primarily fall into three categories —flow field/particle based [1], topic models based [7, 8, 14, 25] and the ones based on analysis of trajectories [16, 17, 18].

Ali et al., [1] proposed a flow field based approach for segmenting and analyzing motion in crowd videos. Their approach involves studying the Lagrangian dynamics of particles propagated on flow fields generated by a moving crowd. The evolution of these particles over time is tracked using a Flow Map and the eigenvalues of their deformation tensor is used to reveal the Lagrangian Coherent Structures (LCS) of the underlying flow. Using these structures, the flow field is divided into regions, each with different motion dynamics, and the final flow segments are identified. But this approach assumes that the flow in the video is static and doesn't change with time i.e., a pixel location in the frame can only have a single flow going through it.

Also, there have been approaches which analyze the recurrent patterns in surveillance videos to extract dominant flows and detect anomalies in the framework of topic models. Kwak et al., [14] used Latent Dirichlet Allocation (LDA) model for abnormal event detection. Varadarajan et

al., [25] proposed motif - a sequential topic model, which models not only the co-occurrence of words in a document but also their temporal order. Further, they proposed a method for recovery of sparse distributions to represent the starting times of various topics using regularization constraints.

Another popular approach in this area is based on the analysis of motion trajectories. Morris et al., [17, 18] demonstrated that trajectory patterns learned by clustering can demonstrate the various typical behaviors in a surveillance video. They have done extensive experiments with a wide variety of distance measures —DTW, PCA, Hausdroff etc., for clustering and demonstrated that warping based approaches generally work better.

Kim et al., [11] represented trajectories as continuous dense flow fields using Gaussian Process Regression. Upon learning different classes of motion using this representation and a novel random sampling strategy, they perform path prediction and anomalous event detection by comparing the trajectories of vehicles with the learned models.

Our approach for flow analysis deals with long-term flow segments, which can be thought of as an extension to the trajectories and capture the spatio-temporal information of the moving object more effectively.

3. Proposed Method

In this section, we present our method for long-term flow segmentation in videos followed by the algorithm for clustering them and modeling dominant flows. An overview of the proposed method is shown in Fig. 1.

3.1. Long-term Flow Segmentation

In this work, our goal is only to obtain smooth space-time flow segments with coherent motion and not to get a boundary accurate object segmentation. This means that two objects, in close proximity, moving together for their entire duration will be segmented out as a single flow. These extracted flow segments can be used for various flow analysis tasks in video surveillance, like dominant flow extraction, path prediction, anomalous flow detection etc.

For flow segmentation, our approach uses optical flow vectors and local color histograms as features in the framework of CRF which is a probabilistic graphical model for the joint labeling of structured data. In this model, the structure present in the data is captured by the edges of an undirected graph while the data itself is represented as features of the nodes in the graph. The model aims to find an optimal labeling for each node in the graph while maximizing the joint probability of the overall labeling. For this, the model assumes Markovian property - the probability of a node taking a label is conditionally independent of the non-neighboring nodes given the features of its neighboring nodes and their labels. With the Markovian assumption, the joint probability of a labeling when expressed as energy, Eq. (1), is a summation of potentials defined over cliques in the graph, Eq. (2).

$$P(\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \exp(-E(\mathbf{x})) \quad (1)$$

$$E(\mathbf{x}) = \sum_{c \in \mathcal{C}_{\mathcal{G}}} \phi_c(x_c) \quad (2)$$

Here, \mathbf{x} is a possible labeling over the nodes of an undirected graph \mathcal{G} and $\mathcal{C}_{\mathcal{G}}$ denotes the set of possible cliques in this graph. The optimal labeling is the one that minimizes the overall energy defined in Eq. (2).

The proposed flow segmentation method operates in a frame-by-frame manner by propagating the currently active flow segments on to the next frame and initializing new flow segments whenever a new moving object is found. Both the propagation and initialization are done using a CRF constructed on the image grid of the subsequent frame. Each pixel in the frame is considered to be a node and cliques of size 4 are formed by connecting each pixel with its top, left, bottom, and right pixels. The pixel nodes of this CRF are characterized by three features - location, optical flow and local color histogram.

Let there be K active flows at a time instant t . We propagate these flow segments to the frame at time instant $t + 1$ by constructing a CRF and solving it to obtain an optimum labeling for each pixel node into the following:

- Background
- New flow segment
- One of the existing K flow segments

Thus the optimization of the CRF is carried out over a set of $K + 2$ possible labels for each node. The unary and pair-wise terms for the CRF are described below.

3.1.1 Unary Cost

Background The nodes in the graph where the motion is negligibly low are encouraged to take the background label by defining the unary cost to be a decreasing function with

fall in the flow magnitude. Specifically, the unary cost for a node u at location x_u , with optical flow vector f_u , for taking the background label l_B is defined as

$$\varphi_u(l_B) = C_1 * \exp(-\alpha^2) \quad (3)$$

where C_1 is a scaling constant and α is defined as

$$\alpha = \frac{th_{mag}(x_u)}{\|f_u\|_2} \quad (4)$$

Here, $\|\cdot\|$ denotes the l_2 norm and th_{mag} is a location-dependent threshold on the magnitude of optical flow to determine noisy flow vectors. This is calculated using the median flow statistics at each location. This makes the threshold location dependent, enabling it to deal with the problem of near-by and far-from camera motion.

New Flow Segment This label denotes regions where a new vehicle has just entered the view or an existing vehicle has started to move. The unary cost for this label favors nodes with a high motion magnitude. The unary cost for a node u at location x_u , with optical flow vector f_u , to belong to the new segment label l_N is calculated as

$$\varphi_u(l_N) = C_1 * \exp\left(-\frac{1}{\alpha^2}\right) \quad (5)$$

where α is defined as in Eq. (4). Each connected component in this segment will be considered as an independent flow segment and will be assigned a new label in the subsequent frames.

Existing Flow Segments Now we define the unary costs for a node to belong to any of the K existing flow segments. Let the k^{th} flow segment be FS_k and the corresponding label be l_{FS_k} . To facilitate the computation of unary cost for a pixel node u to take this label l_{FS_k} , we create a likelihood map on the $t + 1^{th}$ frame. This likelihood map is computed using a non-parametric model built from the pixels of flow segment FS_k in the t^{th} frame.

Let these pixel nodes from frame t be $\{v_i\}_{i=1}^N$ and their location and optical flow respectively be $\{x_{v_i}\}_{i=1}^N$ and $\{f_{v_i}\}_{i=1}^N$. The likelihood for the node u in frame $t + 1$ to take the label l_{FS_k} is given as

$$\mathcal{L}(\text{label}(u) = l_{FS_k}) = \frac{1}{N} \sum_{i=1}^N \mathbf{K}_{\mathbf{x}}(x_{v_i} + f_{v_i} - x_u) \mathbf{K}_{\mathbf{f}}(f_{v_i} - f_u) \quad (6)$$

where $\mathbf{K}_{\mathbf{x}}$, $\mathbf{K}_{\mathbf{f}}$ are spatial and optical-flow kernels respectively. In this work, we have used multi-variate Gaussian kernels whose variances are determined from the statistics of magnitude of optical flow in the video. Since the Gaussian falls quickly, the likelihood is computed only for pixels within the vicinity of $\{x_{v_i} + f_{v_i}\}_{i=1}^N$ and is assumed a very low constant value elsewhere. This is done to avoid the overhead of computing likelihood over the entire frame.

The likelihood computed for a pixel node is converted to unary cost as

$$\varphi_u(l_{FS_k}) = -\log(\mathcal{L}(\text{label}(u) = l_{FS_k})) \quad (7)$$

This formulation makes the unary cost a decreasing function of the node’s spatial proximity and motion similarity to the pixels of the flow segment in the previous frame.

The regularization performed during the computation of optical flow can quite often result in over-smoothing. This results in blurring of the optical flow vectors near the object boundaries and can cause background pixel nodes to be labeled as flow segments. This problem is countered by scaling the defined non-background unary costs, i.e., $\varphi_u(l_N), \varphi_u(l_{FS_k})$, by a factor proportional to the Bhattacharya similarity between the local color histograms of pixel node in the frame $t + 1$ and that in the background model. This amplifies the non-background unary costs at background regions and helps in preserving the boundaries. Here, the local color histograms are computed using the linear-time implementation of Locality Sensitive Histograms [9].

3.1.2 Pair-wise Cost

In our CRF formulation, we have used the Potts model for the pair-wise term as there is no prior information regarding the neighborhood relationship between labels. The pair-wise cost for two neighboring nodes u and v taking different labels is given as

$$\psi_{u,v} = C_2 * \exp(-\|f_u - f_v\|^2) \quad (8)$$

where C_2 is a scaling constant, f_u and f_v are the optical flow vectors of nodes u and v respectively.

3.1.3 Inference in CRF

With the unary and pair-wise cost terms defined for the CRF, optimization is carried out to find a labeling which minimizes the following overall cost.

$$E(\mathbf{x}) = \sum_u \varphi_u(\text{label}(u)) + \sum_{\substack{u,v \\ u \neq v}} \psi_{u,v}(\text{label}(u), \text{label}(v)) \quad (9)$$

By assigning optimal labels for all the nodes, existing flow segments are propagated from the current frame t to the next frame $t + 1$ and simultaneously any possible new flow segments are initialized.

Since the problem of exact energy minimization for the formulated CRF is NP hard, we use an approximate solution. This is determined using the graph cuts based algorithm proposed by Boykov et al., [3, 4, 6, 12]. Their algorithm is known to converge quickly for CRFs formulated on

image grids by making very large moves towards the minima whenever possible. The moves made by the algorithm are primarily of two kinds: $\alpha - \beta$ swap and α expansion. In $\alpha - \beta$ swap, for an arbitrary set of pixels, the labels α, β are swapped. These swap moves are made until the global energy can not be further reduced by a swap. In α expansion, an arbitrary set of pixels are assigned the label α . These moves generate a labeling such that its energy can not be further brought down by any expansion move.

One of the problems that could be encountered during flow segmentation is flow bifurcation. This can happen when two objects, which are very close by, move together for a certain distance and then take different paths. To handle such scenarios, we assign different labels to each of the component flows, in the frame where they begin to diverge. The segmentation procedure is then reapplied in the reverse temporal order, over the original combined flow pipe, with the two new flows as possible labels. This splitting up of the flow is done within the same CRF framework using the above formulated unary and pair-wise costs.

3.2. Flow Clustering

Each flow segment can be thought of as a sequence of binary masks. Each mask contains exactly a single connected component indicating the position of an object at a time instant. Let $FS_1 = \{B_1^1, B_1^2, B_1^3, \dots, B_1^M\}$, $FS_2 = \{B_2^1, B_2^2, B_2^3, \dots, B_2^N\}$ be two flow segments and B_k^i be the i^{th} binary mask of the k^{th} flow segment. In order to perform clustering of these flow segments, we first define a measure of distance between any two binary masks followed by a distance measure for flow segments.

The distance between two binary masks B_1^i and B_2^j is defined as

$$\text{dist}(B_1^i, B_2^j) = \|c(B_1^i) - c(B_2^j)\|_2 + |r(B_1^i) - r(B_2^j)| \quad (10)$$

where $c(B_1^i)$, $r(B_1^i)$ denote the centroid location and mean radius of the connected component in binary mask B_1^i . This is illustrated in Fig. 2 by placing the connected components of both the binary masks on the same frame. Also, we define the mean of two binary masks B_1^i and B_2^j to be a binary mask with a single connected component with centroid $\frac{1}{2}(c(B_1^i) + c(B_2^j))$ and radius $\frac{1}{2}(r(B_1^i) + r(B_2^j))$.

With the distance measure defined by Eq. (10), we find a frame-to-frame mapping between FS_1 and FS_2 using the technique of Dynamic Time Warping (DTW) [21, 24]. The process of time warping results in a frame-to-frame mapping between the two flow segments with minimum cumulative cost over the mapped frames while ensuring that the mapping is both continuous and monotonic. Finally, we define the distance between two flow segments to be the overall cost of warping them with respect to each other.

Similar to Kim et al. [11], we have considered only complete flow segments for the clustering. Let there be N flow

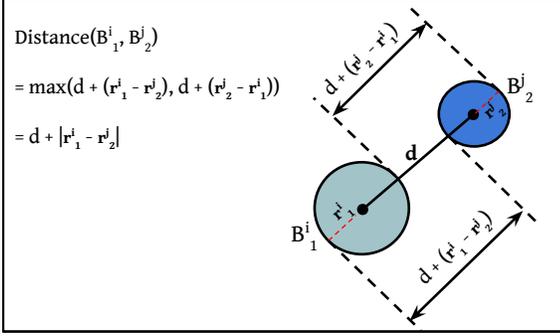


Figure 2: Distance between binary masks

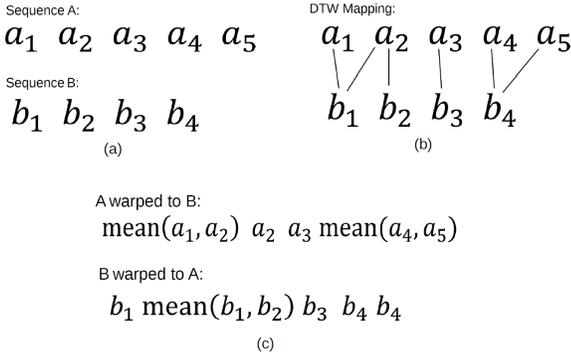


Figure 3: (a) Sequences A and B (b) Dynamic Time Warping of A and B (c) Sequence A warped to B and sequence B warped to A

segments obtained from the flow extraction step after pruning out noisy and incomplete flow segments. We construct a distance matrix D of size $N \times N$ where $D(i, j)$ is the overall warping cost between the flow segments FS_i and FS_j . Using this distance matrix, we cluster the flow segments using the Spectral Clustering algorithm [5, 26]. This will result in semantic clusters i.e., flow segments in a cluster will closely resemble each other in terms of their path and correspond to a dominant flow of the video.

3.3. Modeling the Dominant Flows

Now, we create model flow segments to effectively represent each of the clusters. This model flow segment is also a sequence of binary masks.

Let there be M flow segments FS_1, FS_2, \dots, FS_M , belonging to a cluster k . The representative model, FS_{model_k} , for this cluster can be defined to be the centroid of this cluster. In the sense of the proposed distance measure for flow segments, this centroid can be obtained as

$$FS_{model_k} = \arg \min_{FS} \sum_{i=1}^M d_{warp}(FS, FS_i) \quad (11)$$

Algorithm 1 : Modeling a Dominant Flow

Require: M Flow Segments of a cluster k : $\{FS_i\}_{i=1}^M$

Ensure: Model Flow Segment: FS_{model_k}

```

%% Assume the longest flow segment as initial model
FS_model_k = FS^j s.t. j ∈ [1 M]

while !convergence(FS_model_k) do
  for i = 1 → M do
    %% Time align the sequence to model using DTW
    as shown in Fig. 3
    FS_i^warp = warp(FS_i, FS_model_k)
  end for

  %% Update the model to be mean of all warped sequences
  FS_model_k = mean({FS_i^warp}_{i=1}^M)
end while

```

where $d_{warp}(FS, FS_i)$ denotes the warping cost between FS and FS_i .

We propose an approximate solution to the optimization problem in Eq. (11) as follows: we start by assuming the longest flow segment from the cluster as our initial model. All the flow segments in that cluster are warped to this model as illustrated in Fig. 3 (c). After warping, each of the warped flow segments will have the same number of frames as that of the model. These are then average pooled across each frame to obtain the new model. This process is then repeated until the model converges as described in Algorithm 1.

3.4. Path Prediction

Since the dominant flow segments indicate the paths commonly traversed in a video, we can predict the path of a vehicle entering the camera field of view by comparing its incomplete flow segment with the representative models learned for each dominant flow. Let the partial flow segment of a vehicle be denoted by FS_{test} . We warp this flow segment with each of the dominant flow models and obtain a warping cost as described in the Section 3.2. The likelihood for a vehicle with the partial flow segment FS_{test} to move along the k^{th} dominant flow can be related to their warping cost as

$$\mathcal{L}_k(FS_{test} \in FS_{model_k}) = 1 - \frac{d_{warp}(FS_{test}, FS_{model_k})}{\sum_{i=1}^K d_{warp}(FS_{test}, FS_{model_i})} \quad (12)$$

where $d_{warp}(FS_{test}, FS_{model_k})$ denotes the proposed warping based distance between the flow segments FS_{test}

and FS_{model_k} .

3.5. Anomaly Detection

The proposed dominant flow models can also be used to detect anomalies due to wrong turn/path in traffic. The flow segment of an object moving in an anomalous path does not warp well to any of the learned dominant flow models resulting in a high warping cost. Further, the anomaly score for a test flow segment can be defined as

$$A(FS_{test}) = \min_k(d_{warp}(FS_{test}, FS_{model_k})) \quad (13)$$

4. Experiments

We have evaluated the proposed method of dominant flow extraction on the following traffic videos: UCF video [2], Adam dataset [11], Junction video.

The UCF video is a short video of 1 minute 11 seconds duration with 5 dominant flows. Among these dominant flows, a few flows have paths with a common origin/destination. Also some of the flows have paths intersecting with each other. However these intersecting flows do not occur at the same time. The proposed method for flow extraction has obtained 45 complete flow segments from this video. Since the clustering is unsupervised, quantitative evaluation of the clustering would require obtaining a one-to-one correspondence between clusters formed and the ground-truth labels. We have used the Hungarian algorithm [13] to find the correspondence which minimizes the number of misclassified flow segments. Using this, we have obtained a clustering accuracy of 100% on this video.

In the Adam dataset, we have used three videos which do not contain any anomalous motions to learn the dominant flows. These videos together are about 7 minutes 23 seconds duration and they have 4 dominant flows. This video has the additional complexity of amplification of vehicle movements near-by camera as the view plane is not parallel with the camera (Fig. 4). We have performed clustering on the extracted complete flow segments using the proposed warping based distance and obtained a clustering accuracy of 100%.

The 4 dominant flow models built using these clustered flow segments are shown in Fig. 4. The flow models are represented as points scattered along the length of the flow. The color of the points denote the temporal evolution of the flow where green indicates the starting point and red indicates the ending point of the flow.

The Junction video is captured at a traffic intersection for a duration of 13 minutes 47 seconds. It contains 12 different dominant flows of all possible paths in a 4-road junction. This video also has high intra-class variance for each dominant flow. 267 complete flow segments are extracted from this video. These flow segments, upon clustering, are clas-



Figure 4: 4 dominant flows captured for the Adam video

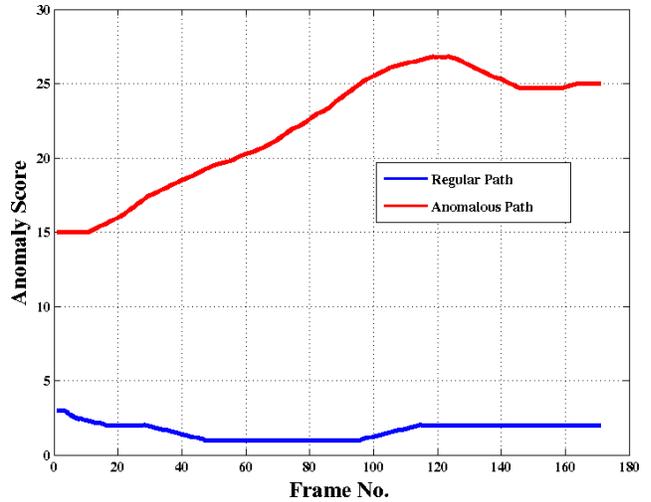


Figure 5: Anomaly Scores for a vehicle moving along a regular path and an anomalous path.

sified into the 12 dominant flows and the proposed method has obtained a classification accuracy of 96.25%.

The 12 dominant flow models built using these clustered flow segments are shown in Fig. 6. Out of the 12 possible paths in the junction, the flow which starts from the left-center and ends in the top-center of the image frame has not been captured by the algorithm. This is because of the very few complete flow segments extracted which represent this path.

The application of dominant flows in predicting the path of a vehicle is illustrated for the Junction Video in Fig. 7. In each of Fig. 7 (a), (b) the red bounding box indicates the vehicle for which the path is being predicted. At a time instant, the predicted path for the vehicle is indicated by color markers - green indicating the starting position and red indicates the end of the path. The size of the markers indicate the likelihood of the vehicle taking a particular path. As

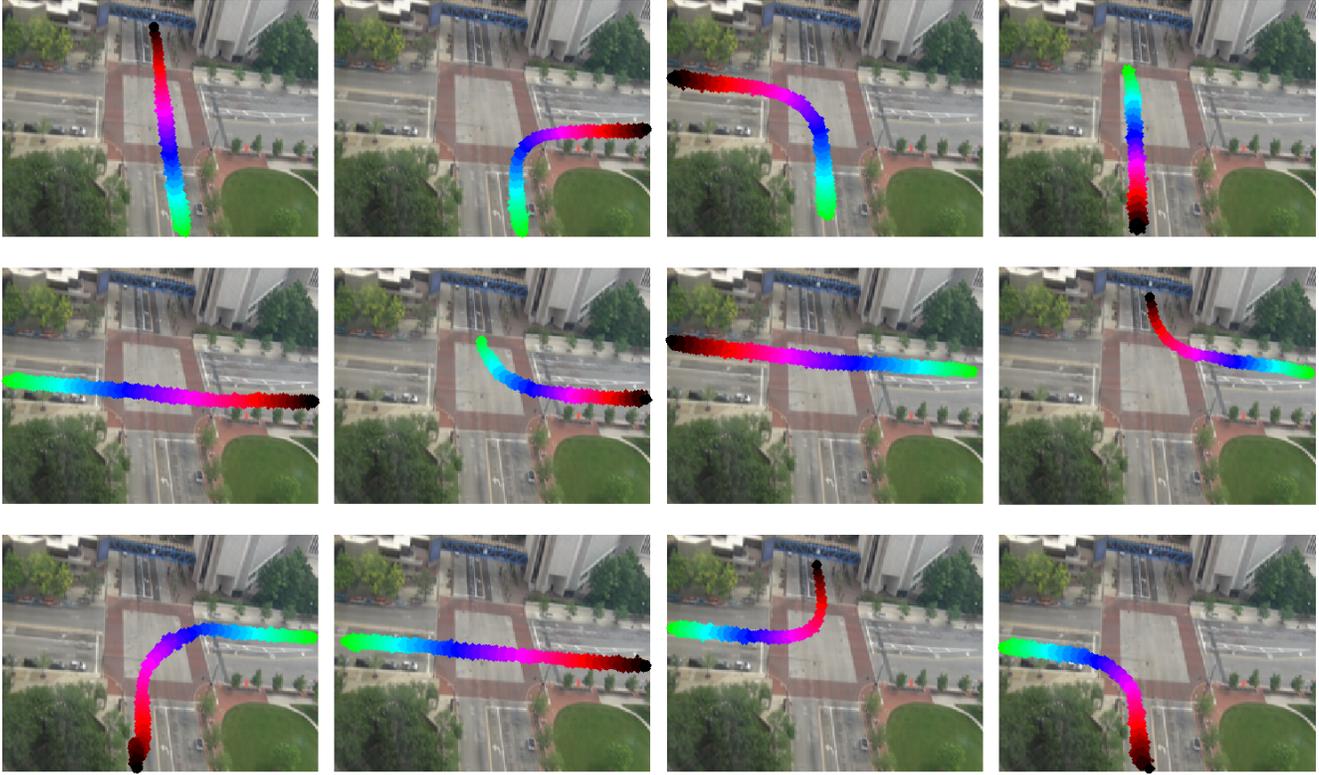


Figure 6: 12 dominant flows captured for the Junction video

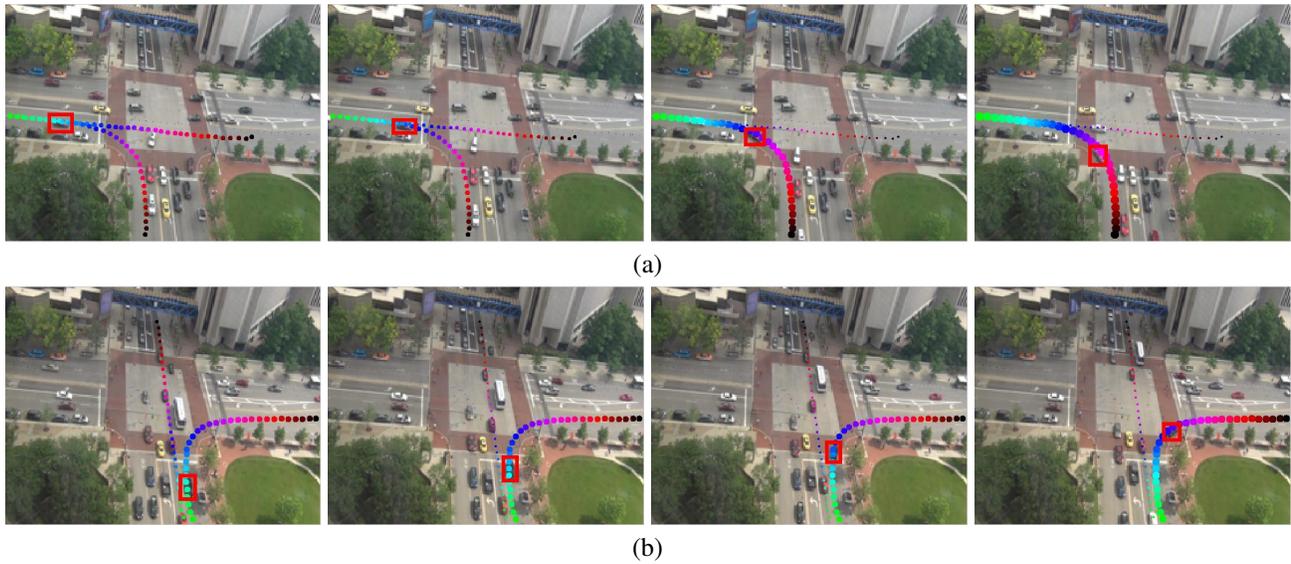


Figure 7: Path prediction for vehicles entering the camera view. Colored markers indicate predicted paths for vehicle in the red bounding box and the size of the markers indicated the likelihood of each path.

it is evident from the Fig. 7 the likelihoods are updated as the time progresses and they become less degenerate. Also the path with the highest likelihood matches with the actual path taken by the vehicle.

In order to demonstrate the performance of the proposed method in detecting vehicles taking unusual paths, we have introduced synthetic anomalies in the Junction video. These are generated by time reversing the motion of certain vehi-

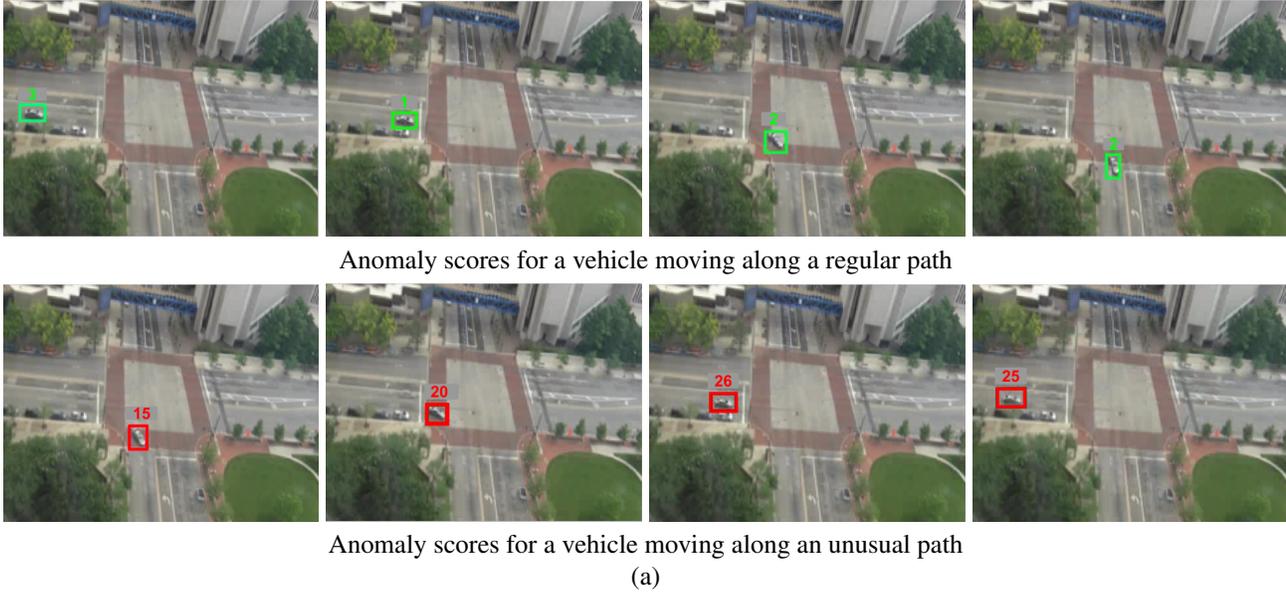


Figure 8: Detection of vehicles moving in unusual paths using the extracted dominant flows. Score above each bounding box indicates the anomaly score for the vehicle.

cles. The frames in the top row of the Fig. 8 (a) are taken from a sequence depicting a vehicle moving along a regular path. This sequence is temporally reversed to introduce anomalous motion and the corresponding frames are shown in the bottom row.

We have performed long-term flow extraction on both of these sequences and calculated the anomaly score at each frame. The anomaly score at a time instant is shown a the bounding box of the vehicle in each frame. The color of the bounding box also indicates if the vehicle movement is an anomaly - green for vehicles moving along dominant paths and red for the anomalies. As evident from the figure, the anomaly scores for the bottom row, in Fig. 8 (a), are higher than that of the top row indicating that the vehicle is moving along an unusual path. For the case of Fig. 8 (a), the anomaly scores at each frame in the sequence are plotted in Fig. 5.

The dominant flows obtained from these three datasets are evaluated quantitatively by calculating the Jaccard similarity measure of the obtained outputs with the ground truth dominant flows. These results are presented in Table. 1

5. Conclusion

The main contribution of this work is a novel method for the extraction of long-term flow segments which characterize the movements of vehicles in traffic surveillance videos. This long-term flow segmentation is done using the Conditional Random Fields using the motion and color information. These flow segments are more informative than the usual trajectories as they capture the spatial extent of a

Video Sequences	Jaccard Similarity
UCF	0.71
Adams	0.80
Junction	0.63

Table 1: Jaccard Similarity Measure with Ground-truth Dominant Flows

moving body along with its path.

We have defined a time warping based distance measure for these flow segments and used this distance measure for clustering the flow segments into dominant flow clusters. We have also created a model for representing each dominant flow by mean pooling over the warped flow segments belonging to its cluster. Further, we have illustrated its applications in predicting the path of vehicles entering the camera view and detecting vehicles moving along unusual paths.

6. Acknowledgement

This work was supported by Defence Research Development Laboratory (DRDO), project No. DRDO0672.

References

- [1] S. Ali and M. Shah. A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–6, June 2007. 2

- [2] S. Ali and M. Shah. Floor fields for tracking in high density crowd scenes. In D. Forsyth, P. Torr, and A. Zisserman, editors, *Computer Vision ECCV*, pages 1–14. Springer, 2008. 6
- [3] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1124–1137, Sept 2004. 4
- [4] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, Nov 2001. 4
- [5] W.-Y. Chen, Y. Song, H. Bai, C.-J. Lin, and E. Y. Chang. Parallel spectral clustering in distributed systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(3):568–586, 2011. 5
- [6] A. DeLong, A. Osokin, H. Isack, and Y. Boykov. Fast approximate energy minimization with label costs. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2173–2180, June 2010. 4
- [7] R. Emonet, J. Varadarajan, and J. Odobez. Extracting and locating temporal motifs in video scenes using a hierarchical non parametric bayesian model. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3233–3240, June 2011. 2
- [8] R. Emonet, J. Varadarajan, and J.-M. Odobez. Temporal analysis of motif mixtures using dirichlet processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(1):140–156, Jan 2014. 2
- [9] S. He, Q. Yang, R. W. H. Lau, J. Wang, and M.-H. Yang. Visual tracking via locality sensitive histograms. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2427–2434. IEEE Computer Society, 2013. 4
- [10] X. He, R. Zemel, and M. Carreira-Perpindn. Multiscale conditional random fields for image labeling. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages II–695–II–702 Vol.2, June 2004. 1
- [11] K. Kim, D. Lee, and I. Essa. Gaussian process regression flow for analysis of motion trajectories. In *In Proceedings of IEEE ICCV*, 2011. 2, 4, 6
- [12] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):147–159, Feb 2004. 4
- [13] H. W. Kuhn. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2):83–97, 1955. 6
- [14] S. Kwak and H. Byun. Detection of dominant flow and abnormal events in surveillance video. *Optical Engineering*, 50(2):027202–027202–8, 2011. 2
- [15] J. D. Lafferty, A. McCallum, and F. C. N. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *International Conference on Machine Learning*, pages 282–289, 2001. 1
- [16] B. Morris and M. Trivedi. Learning and classification of trajectories in dynamic scenes: A general framework for live video analysis. In *IEEE International Conference on Advanced Video and Signal Based Surveillance*, pages 154–161, Sept 2008. 2
- [17] B. Morris and M. Trivedi. Learning trajectory patterns by clustering: Experimental studies and comparative evaluation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 312–319, June 2009. 2
- [18] B. T. Morris and M. M. Trivedi. A survey of vision-based trajectory learning and analysis for surveillance. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(8):1114–1127, 2008. 2
- [19] A. Quattoni, M. Collins, and T. Darrell. Conditional random fields for object recognition. In *NIPS*, pages 1097–1104. MIT Press, 2004. 1
- [20] A. Quattoni, S. Wang, L. Morency, M. Collins, and T. Darrell. Hidden conditional random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(10):1848–1852, Oct 2007. 1
- [21] H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 26(1):43–49, Feb 1978. 4
- [22] P. Sand and S. Teller. Particle video: Long-range motion estimation using point trajectories. *International Journal of Computer Vision*, 80(1):72–91, 2008. 1
- [23] N. Sundaram, T. Brox, and K. Keutzer. Dense point trajectories by gpu-accelerated large displacement optical flow. In *Computer Vision ECCV 2010*, pages 438–451. Springer, 2010. 1
- [24] R. J. Turetsky and D. P. Ellis. Ground-truth transcriptions of real music from force-aligned midi syntheses. *ISMIR 2003*, pages 135–141, 2003. 4
- [25] J. Varadarajan, R. Emonet, and J.-M. Odobez. A sequential topic model for mining recurrent activities from long term video logs. *International Journal of Computer Vision*, 103(1):100–126, 2013. 2
- [26] U. von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007. 5
- [27] S. B. Wang, A. Quattoni, L. Morency, D. Demirdjian, and T. Darrell. Hidden conditional random fields for gesture recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 1521–1527, 2006. 1