TRAJECTORY CLUSTERING FOR MOTION PATTERN EXTRACTION IN AERIAL VIDEOS

Tahir Nawaz[*†]   *Andrea Cavallaro*[*]   *Bernhard Rinner*[†]

[*]Centre for Intelligent Sensing, Queen Mary University of London, United Kingdom
[†]Institute of Networked and Embedded Systems, Alpen-Adria-Universität Klagenfurt, Austria
{tahir.nawaz,andrea.cavallaro}@eecs.qmul.ac.uk, bernhard.rinner@aau.at

ABSTRACT

We present an end-to-end approach for trajectory clustering from aerial videos that enables the extraction of motion patterns in urban scenes. Camera motion is first compensated by mapping object trajectories on a reference plane. Then clustering is performed based on statistics from the Discrete Wavelet Transform coefficients extracted from the trajectories. Finally, motion patterns are identified by distance minimization from the centroids of the trajectory clusters. The experimental validation on four datasets shows the effectiveness of the proposed approach in extracting trajectory clusters. We also make available two new real-world aerial video datasets together with the estimated object trajectories and ground-truth cluster labeling.

Index Terms— Aerial videos, trajectory clustering, motion patterns, trajectory features.

1. INTRODUCTION

The extraction of motion patterns corresponding to the movement of people and vehicles in a scene can support behavior prediction [1], abnormality detection [2,3] and tracking [4]. Motion patterns can be extracted by analyzing motion information between frame pairs [5–7] or by analyzing motion information across multiple frames (object trajectories) [8–12]. The former category of approaches is suitable for extracting short-range patterns, whereas the latter category helps extracting long-range patterns when the trajectory information is available [7,12]. Trajectory-based methods generally rely on clustering spatio-temporal features [8–11] or frequency-domain features such as DFT coefficients [12,13]. Hu *et al.* [8] presented a hierarchical trajectory clustering framework that separated the trajectories of vehicles and persons and then subclusters trajectories of each category to extract motion patterns. Anjum and Cavallaro [9] introduced a framework that performed trajectory clustering and then fused clusters obtained with different features to identify patterns. Wang *et al.* [14] proposed a method to learn motion patterns using the Dual Hierarchical Dirichlet Processes (Dual-HDP). The authors in [14] built on the Dual-HDP model introducing a Dynamic Dual-HDP model in [11] that enabled updating motion patterns dynamically. Zhang *et al.* [10] applied trajectory clustering in a block-based scene representation based on Gaussian Mixture Models (GMM) to learn motion patterns. Recently, Hu *et al.* [12] proposed an incremental trajectory clustering algorithm to learn motion patterns based on Dirichlet Process Mixture Model (DPMM).

As most frameworks assume stationary cameras [1, 10, 12, 14, 15], an important challenge is extracting motion patterns from aerial

videos as they require camera motion compensation. Methods exist that cope with camera motion but are not aimed at motion pattern extraction. These methods rely on the availability of Geo-spatial Information System (GIS) information and perform geo-registration of the aerial video to estimate depth cues for identifying buildings, trees and roads [16] or to segment areas using motion-based foreground segmentation without distinguishing motion patterns by registering the input frames with a generated background mosaic [17].

This paper presents an end-to-end approach for trajectory clustering for motion pattern extraction in aerial videos. The overall method involves compensating camera motion in the estimated trajectories and performing trajectory clustering to identify motion patterns (Fig. 1). To perform clustering, we use a feature that encapsulates trajectory information using its Discrete Wavelet Transform (DWT) coefficients. We demonstrate the effectiveness of the proposed approach compared to the state of the art on four real-world datasets. We also introduce two new real-world aerial datasets for parking lot and traffic junction scenes, which are made available online together with the estimated trajectories and ground-truth cluster labeling at http://uav.lakeside-labs.com/publications/test-data.

This paper is organized as follows. Sec. 2 describes the camera motion compensation in trajectories. Sec. 3 explains the feature extraction and trajectory clustering. The experimental setup is provided in Sec. 4 and results in Sec. 5. Sec. 6 concludes the paper.

2. MOTION COMPENSATION

We aim to identify motion patterns produced by people and vehicles in urban scenes using videos captured by UAVs equipped with a top-down looking camera. Let $\mathfrak{X} = \{\mathcal{X}_i\}_{i=1}^{I}$ be a set of trajectories $\mathcal{X}_i$ of moving objects on the image plane obtained using a video tracker, where I is the total number of trajectories. k_i^S and k_i^E denote the start and end frames for $\mathcal{X}_i$: $\mathcal{X}_i = [X_i^k]_{k=k_i^S}^{k_i^E}$, and K is the total number of frames in the video sequence, V. The position of an object at each frame k of V is defined as $X_i^k = [(x_i^k, y_i^k)]$, where (x_i^k, y_i^k) are the coordinates of an object on the image plane.

We compensate the camera motion in the trajectories, $\mathfrak{X}$, based on the homography computation under the assumption of planar object motion and minimal perspective distortions, described as follows. We map all trajectories on a common frame, $\mathcal{I}_r$, selected from the frames of V. The choice of $\mathcal{I}_r$ is made so as to ensure that it overlaps with the remaining frames. Given $\mathcal{X}_i$, X_i^k is to be mapped on $\mathcal{I}_r$ by computing a homography, $H_{k,r}$, between $\mathcal{I}_k$ and $\mathcal{I}_r$. To this end we use the standard feature-based alignment method [18] that involves extracting point features in $\mathcal{I}_k$ and $\mathcal{I}_r$, determining point correspondences and computing $H_{k,r}$ with the best correspondences obtained by applying RANSAC. We employ the widely-used SIFT point features [19]. RANSAC reduces the errors (caused due to the
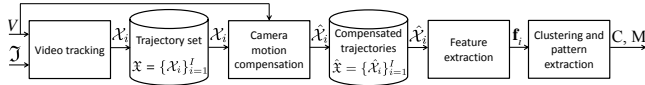
Fig. 1. Proposed pipeline - V: video sequence; $\mathfrak{I}$: tracker initialization set; $\mathfrak{X}$: trajectory set; $\mathcal{X}_i$: trajectory i; $\hat{\mathfrak{x}}$: compensated trajectory set; $\mathbf{f}_i$: feature vector; C, M: set of clusters and motion patterns.

presence of non-planar SIFT matches and mismatches) in the homography computation and hence in the compensated trajectories.

Although alternative approaches to homograhy-based camera motion compensation [20, 21] are suitable for segment-long [21] or sequence-long [20] optical flow-advected dense trajectories (belonging both to background and foreground), they are not directly applicable to object trajectories that can have variable lengths and different k_i^S and k_i^E.

After the computation of the homography matrix, (x_i^k, y_i^k) can be mapped onto $\mathcal{I}_r$ as: $\left[w\,\hat{x}_i^k, w\,\hat{y}_i^k, w\right]^T = H_{k,r}\left[x_i^k, y_i^k, 1\right]^T$, where $(\hat{x}_i^k, \hat{y}_i^k)$ are the corresponding coordinates of X_i^k in $\mathcal{I}_r$ obtained by dividing the left-hand-side of equation by w. Similarly, all the positions in $\mathcal{X}_i$ can be mapped onto $\mathcal{I}_r$ to get the corresponding compensated trajectory, $\hat{\mathcal{X}}_i$. In this way, all $\mathcal{X}_i$ are transformed onto $\mathcal{I}_r$ to obtain $\hat{\mathfrak{x}} = \{\hat{\mathcal{X}}_i\}_{i=1}^I$, the set of compensated trajectories on $\mathcal{I}_r$ (Fig. 2).

3. FEATURE EXTRACTION AND CLUSTERING

We first encode the time-varying information of the trajectories $\hat{\mathfrak{x}} = \{\hat{\mathcal{X}}_i\}_{i=1}^I$ for clustering. Then we apply a trajectory-clustering procedure to $\hat{\mathfrak{x}}$ thus yielding a set of clusters $C = \{C_n\}_{n=1}^{\bar{N}}$. Each cluster, C_n, is used to represent the corresponding motion pattern, M_n (i.e. each M_n is a representative spatio-temporal trend of object motion in the scene).

Feature extraction in the frequency domain is demonstrated to be appropriate for trajectory clustering [12, 13, 22] using a set of DFT or DWT coefficients. Due to the time localization, DWT can better capture the changing frequency information along trajectories and has a lower complexity ($O(N)$) than DFT ($O(N \log N)$) [23], where N is the number of points along the trajectory. DWT is also used for trajectory retrieval by Sahouria and Zakhor [24].

We use Haar wavelets to capture local variations in trajectories using the single-level implementation of Mallat's algorithm [25] for computing the DWT coefficients. Haar wavelets are also used in the existing works [23, 24] and are reported to perform better than Daubechies and Coiflet wavelets [23]. For a given trajectory $\hat{\mathcal{X}}_i$, we therefore compute the DWT (Haar wavelets) of the 1-D data, $\hat{x}_i(k) = \{\hat{x}_i^k\}_{k=k_i^S}^{k_i^E}$ and $\hat{y}_i(k) = \{\hat{y}_i^k\}_{k=k_i^S}^{k_i^E}$.

To build a feature vector, we use the computed approximation DWT coefficients of $\hat{x}_i(k)$, $\mathcal{C}^{\hat{x}_i}$, and $\hat{y}_i(k)$, $\mathcal{C}^{\hat{y}_i}$, to formulate the DWT-based feature for $\hat{\mathcal{X}}_i$ as follows: $\mathbf{f}_i = (\mathbf{f}^{\hat{x}_i}, \mathbf{f}^{\hat{y}_i})$, where $\mathbf{f}^{\hat{x}_i} = (\min(\mathcal{C}^{\hat{x}_i}), Q_{25}^{\hat{x}_i}, Q_{50}^{\hat{x}_i}, Q_{75}^{\hat{x}_i}, \max(\mathcal{C}^{\hat{x}_i}))$ encapsulates the non-parametric statistics for the coefficients including the minimum coefficient value, the first quartile or 25th percentile ($Q_{25}^{\hat{x}_i}$), the second quartile or 50th percentile ($Q_{50}^{\hat{x}_i}$), the third quartile or 75th percentile ($Q_{75}^{\hat{x}_i}$), and the maximum coefficient value in $\mathcal{C}^{\hat{x}_i}$. Likewise, $\mathbf{f}^{\hat{y}_i} = (\min(\mathcal{C}^{\hat{y}_i}), Q_{25}^{\hat{y}_i}, Q_{50}^{\hat{y}_i}, Q_{75}^{\hat{y}_i}, \max(\mathcal{C}^{\hat{y}_i}))$. Instead of using the first few coefficients [23, 24], $\mathbf{f}_i$ captures the overall distribution of coefficients non parametrically in terms of $\mathbf{f}^{\hat{x}_i}$ and $\mathbf{f}^{\hat{y}_i}$, thus providing a more comprehensive trajectory description.

Fig. 2. Motion-compensated trajectories overlaid on the image created by registering frame 1 ($\mathcal{I}_r$) and frame 7729 of the Parking Lot dataset (left) and by registering frame 13468 ($\mathcal{I}_r$) and frame 4785 of the Traffic Junction dataset (right).

We perform clustering using the feature $\mathbf{f}_i$ computed for each $\hat{\mathcal{X}}_i$ to extract a set $C = \{C_n\}_{n=1}^{\bar{N}}$ of $\bar{N}$ trajectory clusters, where C_n denotes cluster n. We use the trajectory clustering algorithm proposed in [9], which uses an incremental procedure to select the bandwidth parameter in the Mean-Shift procedure and does not require the prior knowledge of $\bar{N}$, the number of clusters. The bandwidth parameter is initialized with 20% of the range of the feature space of $\mathbf{f}_i$. As done in [9], trajectories in sparse clusters (i.e. whose cardinality is smaller than 10% of the median cardinality of all the clusters) and those with a normalized absolute distance from the centroid of the corresponding dense clusters greater than $\tau_1 = 0.95$ are defined as outlier trajectories. Unlike the method in [9], the proposed framework uses a DWT-based feature space and addresses the challenge of camera motion compensation in the trajectories.

Finally, a motion pattern M_n is defined by the trajectory that minimizes the distance from the centroid of cluster C_n without considering the direction of motion. The minimization uses the trajectory mean point and length ($k_i^E - k_i^S$) to capture the spatial location and the elongation of the patterns. Examples of C_n and M_n are shown in Fig. 3.

4. EXPERIMENTAL SETUP

We perform the experimental validation of the proposed framework on four datasets (Tab. 1). The first two are a Parking Lot and a Traffic Junction scene containing persons and vehicles. These sequences are captured using an octocopter UAV (AscTec Falcon 8) at low altitudes ($\approx 20 - 40$ m). We extracted real trajectories, $\mathfrak{X}$, using the Mean-Shift tracker [26] with manual initializations to track the moving objects until they leave the scene. $\mathfrak{I}$ denotes the set of initializations for all targets.

The other two datasets are Students003 [27] and Train Station [28] captured from a high viewpoint static camera (i.e. no need for camera motion compensation). We use the provided ground-truth trajectories for Students003 and the provided real trajectories extracted using KLT tracker [29] for Train Station. Most of the trajectories in Train Station are short-duration tracklets obtained by repeated tracker initializiations and dealing with tracklets is out of the scope of the proposed framework. We therefore use only the longer trajectories (length, $k_i^E - k_i^S > 600$) in our experiments with Train Station.

We compare the proposed methods with three alternative approaches, namely $\mathcal{M}1$, $\mathcal{M}2$ and $\mathcal{M}3$. $\mathcal{M}1$ uses a DFT-based feature ($\mathbf{f}_i^D$) [12,13] that represents a trajectory using the first five DFT coefficients of x- and y- coordinates of $\hat{\mathcal{X}}_i$. $\mathcal{M}2$ uses as features the start and end points of each trajectory, $\mathbf{f}_i^{SE}$. To compare with $\mathbf{f}_i^D(\mathbf{f}_i^{SE})$, $\mathbf{f}_i$ is replaced with $\mathbf{f}_i^D(\mathbf{f}_i^{SE})$ in the proposed framework. $\mathcal{M}3$ is the method performing mean-shift trajectory clustering using multiple
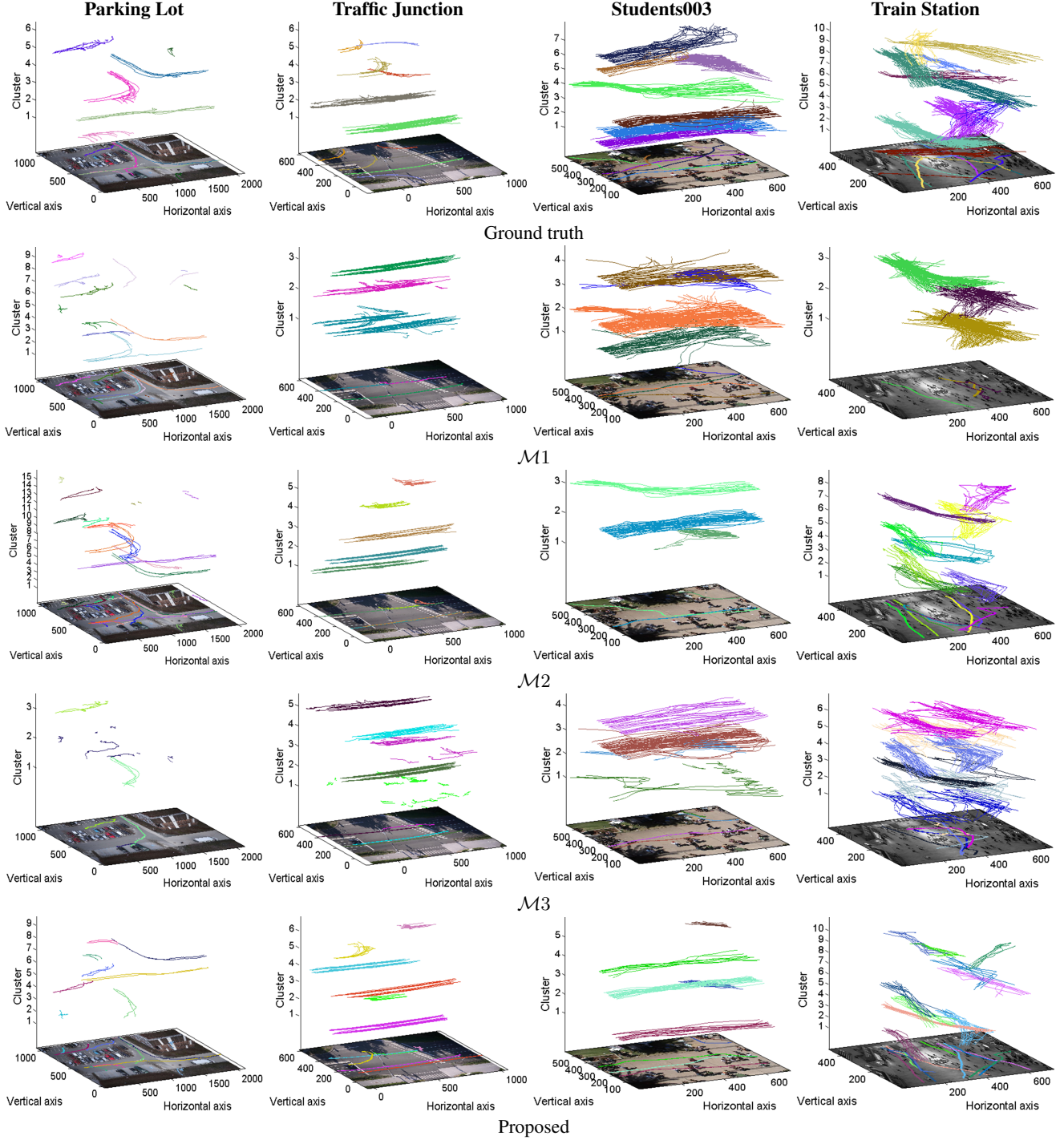
Fig. 3. Visualization of the results for the extracted clusters and motion patterns (color coded) for $\mathcal{M}1$ (second row), $\mathcal{M}2$ (third row), $\mathcal{M}3$ (fourth row) and the proposed method (fifth row) to be compared with the ground truth (first row). The clusters are shown on planes along z-axis in each plot. Motion patterns are superimposed on the original frame and shown in the lowest plane in each plot.

spatio-temporal feature spaces [9] applied on the compensated trajectories.

The quantitative evaluation is done by computing the accuracy of the learned clusters, A, and the precision (P) and recall (R) of the extracted motion patterns. The accuracy is computed as follows [12]: $A = \frac{1}{N} \sum_{n=1}^{N} \frac{b_n}{B_n}$, where b_n is the number of trajectories with the same ground-truth cluster label and the highest proportion in the learned cluster C_n, and B_n is the number of trajectories in C_n

Table 1. Characteristics of the dataset. Key - FS: frame size as height × width (pixels); NF: number of frames; NT: no. of trajectories; CLT: combined length of all trajectories (frames); FPS: frames per second.

Dataset	FS	NF	NT	CLT	FPS
Parking Lot	1080×1920	9517	54	29483	30
Traffic Junction	540×960	16154	236	42311	30
Students003	576×720	5405	417	207304	25
Train Station	480×720	46009	762	557345	23

with $\bar{N}$ denoting the number of learned clusters. The ground-truth cluster labeling was done manually by multiple annotators for Parking Lot and Traffic Junction, and by one annotator for Students003 and Train Station. P and R are computed using correct (true positive), incorrect (false positive) and missed (false negative) patterns. A motion pattern is considered correct if it lies within a ground-truth cluster. For a complete evaluation, P and R should be analyzed with A since a correct pattern may have originated from an inaccurate cluster. The average A, P and R are computed for five runs on each dataset.

5. ANALYSIS OF THE RESULTS

In this section we present the evaluation of the proposed framework in extracting trajectory clusters and patterns, its robustness and discuss its computational complexity.

We performed the evaluation and comparison qualitatively (Fig. 3) and quantitatively (Tab. 2). The clusters generated using the proposed feature, $\mathbf{f}_i$, are more accurate (highest A and highest R). The highest A and R are however associated with a small P on Traffic Junction, Students003 and Train Station due to false positives. Except for Parking Lot, $\mathcal{M}1$ has the best P. $\mathcal{M}2$ is the second best in terms of A and R (its R is the same as for the proposed method on Parking Lot).

As the performance of the pipeline can be affected by the presence of tracking failure-ridden trajectories, we induce tracking failures by selecting the first half of randomly selected $p\%$ trajectories in each dataset, $p = 0, 10, \ldots, 50$, and analyze its effect on A for $\mathcal{M}1$, $\mathcal{M}2$, $\mathcal{M}3$ and the proposed method (Tab. 3). In the evaluation of A the tracking failure-ridden trajectories are removed from the ground truth clusters as outliers. The results show that the proposed method has the best mean A on all datasets except Train Station where it is the second best to $\mathcal{M}2$. In terms of variation of A, $\mathcal{M}1$ is better due to its smallest σ on all datasets except Parking Lot. From the viewpoint of UAV operations, the tracking failures may be caused as a result of abrupt UAV movements leading to larger inter-frame target displacement on the image plane (Fig. 4), which could be accounted for in the tracking algorithm [30].

Finally, we consider the computational cost for the whole pipeline. The computational effort used by the camera motion compensation in trajectories is significantly greater than that for the

Table 2. Evaluation of the clustering and motion pattern extraction for $\mathcal{M}1$, $\mathcal{M}2$, $\mathcal{M}3$ and the proposed method in terms of A, P and R.

Method	Parking Lot			Traffic Junction			Students003			Train Station		
	A	P	R	A	P	R	A	P	R	A	P	R
$\mathcal{M}1$	.64	.48	.53	.67	.67	.27	.41	.90	.40	.32	.60	.18
$\mathcal{M}2$	.72	.41	1	.82	.46	.40	.78	.74	.43	.80	.46	.36
$\mathcal{M}3$	.56	.63	.33	.70	.40	.33	.51	.60	.28	.33	.35	.18
Proposed	.89	.65	1	.88	.52	.50	.90	.58	.51	.82	.45	.50

Table 3. Effect of inducing tracking failures to $p\%$ randomly selected trajectories on the clustering accuracy in terms of the mean (μ) and standard deviation (σ) of A for $p = 0, 10, \ldots, 50$.

Method	Parking Lot $\mu\,(\sigma)$	Traffic Junction $\mu\,(\sigma)$	Students003 $\mu\,(\sigma)$	Train Station $\mu\,(\sigma)$
$\mathcal{M}1$	.45 (.18)	.53 (.13)	.29 (.10)	.23 (.07)
$\mathcal{M}2$	.57 (.13)	.66 (.19)	.66 (.21)	.60 (.19)
$\mathcal{M}3$	.41 (.20)	.54 (.15)	.37 (.13)	.20 (.09)
Proposed	.64 (.27)	.67 (.19)	.69 (.22)	.52 (.22)

Fig. 4. Tracking results for inter-frame target displacements simulated by regularly dropping $(m - 1)$ frames from the sequence: $m = 0, 2, 4, 6, 8$ (blue, green, magenta, black, yellow) [32]. A tracking failure occurs for $m = 8$ (rightmost image).

other stages due to the need of calculating homography for each frame where trajectories exist. The motion-compensation block is seven orders of magnitude larger (for Parking Lot) and six orders of magnitude larger (for Traffic Junction) than each of the remaining stages. The higher computational effort for the motion compensation stage for the Parking Lot is due to its larger frame size (Tab. 1). The major contributor in the computational effort of the motion compensation stage is the computation of SIFT features due to the combined computational complexity of its multiple steps [31].

6. CONCLUSIONS

We presented a pipeline for extracting trajectory clusters and motion patterns from aerial videos of urban scenes. The pipeline involves applying camera motion compensation to trajectories extracted in the image plane and performing clustering using a feature that encapsulates trajectory information non-parametrically using DWT coefficients. We performed the experimental validation and comparison of the framework on four datasets. The results showed the effectiveness of the proposed method in identifying trajectory clusters and motion patterns. Moreover, considering the scarcity of aerial datasets, we also introduced two new real-world aerial datasets of urban scenes and made them available online together with the estimated trajectories and ground-truth cluster labeling. Future work could involve reducing the computational effort of the compensation stage.

7. REFERENCES

[1] I. Saleemi, K. Shafique, and M. Shah, "Probabilistic modeling of scene dynamics for applications in visual surveillance," *IEEE Trans. on PAMI*, vol. 31, no. 8, pp. 1472–1485, August 2009.

[2] S. Calderara, A. Prati, and R. Cucchiara, "Mixtures of von mises distributions for people trajectory shape analysis," *IEEE Trans. on CSVT*, vol. 21, no. 4, pp. 457–471, 2011.

[3] Y. Zhou, S. Yan, and T. S. Huang, "Detecting anomaly in videos from trajectory similarity analysis," in *Proc. of IEEE ICME*, Beijing, 2007.

[4] B. Yang and R. Nevatia, "Multi-target tracking by online learning of non-linear motion patterns and robust appearance models," in *Proc. of IEEE CVPR*, Providence, 2012.

[5] S. Ali and M. Shah, "A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis," in *Proc. of IEEE CVPR*, Minneapolis, 2007.

[6] S. Wu, B. Moore, and M. Shah, "Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scnes," in *Proc. of IEEE CVPR*, San Francisco, 2010.

[7] P.-M. Jodoin, Y. Benezeth, and Y. Wang, "Meta-tracking for video scene understanding," in *Proc. of IEEE AVSS*, Krakow, 2013.

[8] W. Hu, X. Xiao, Z. Fu, D. Xie, T. Tan, and S. Maybank, "A system for learning statistical motion patterns," *IEEE Trans. on PAMI*, vol. 28, no. 9, pp. 1450–1464, September 2006.

[9] N. Anjum and A. Cavallaro, "Multi-feature object trajectory clustering for video analysis," *IEEE Trans. on CSVT*, vol. 18, no. 11, pp. 1555 – 1564, 2008.

[10] T. Zhang, H. Lu, and S. Z. Li, "Learning semantic scene models by object classification and trajectory clustering," in *Proc. of IEEE CVPR*, Miami, 2009.

[11] X. Wang, K. T. Ma, G.-W. Ng, and W. E. L. Grimson, "Trajectory analysis and semantic region modeling using nonparametric hierarchical bayesian models," *IJCV*, vol. 95, no. 3, pp. 287–312, December 2011.

[12] W. Hu, X. Li, G. Tian, S. Maybank, and Z. Zhang, "An incremental DPMM-based method for trajectory clustering, modeling, and retrieval," *IEEE Trans. on PAMI*, vol. 35, no. 5, pp. 1051–1065, May 2013.

[13] A. Naftel and S. Khalid, "Motion trajectory learning in the DFT-coefficient feature space," in *Proc. of IEEE ICVS*, New York, 2006.

[14] X. Wang, K. T. Ma, G.-W. Ng, and W. E. L. Grimson, "Trajectory analysis and semantic region modeling using a nonparametric bayesian model," in *Proc. of IEEE CVPR*, Anchorage, 2008.

[15] N. Piotto, N. Conci, and F. G. B. De Natale, "Syntactic matching of trajectories for ambient intelligence applications," *IEEE Trans. on MM*, vol. 11, no. 7, pp. 1266–1275, 2009.

[16] J. Xiao, H. Cheng, F. Han, and H. Sawhney, "Geo-spatial aerial video processing for scene understanding and object tracking," in *Proc. of IEEE CVPR*, Anchorage, 2008.

[17] A. Mittal and D. Huttenlocher, "Scene modeling for wide area surveillance and image synthesis," in *Proc. of IEEE CVPR*, Hilton Head Island, June 2000.

[18] R. Szeliski, *Computer Vision: Algorithms and Applications*, Springer, 2011.

[19] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, no. 2, pp. 91–110, November 2004.

[20] S. Wu, O. Oreifej, and M. Shah, "Action recognition in videos acquired by a moving camera using motion decomposition of lagrangian particle trajectories," in *Proc. of ICCV*, Barcelona, 2011.

[21] S. Dey, V. Reilly, I. Saleemi, and M. Shah, "Detection of independently moving objects in non-planar scenes via multi-frame monocular epipolar constraint," in *Proc. of ECCV*, Florence, 2012.

[22] F. Moerchen, "Time series feature extraction for data mining using DWT and DFT," Tech. Rep., Philipps-University Marburg, Germany, 2003.

[23] K.-P. Chan and A. W.-C. Fu, "Efficient time series matching by wavelets," in *Proc. of IEEE ICDE*, Sydney, 1999.

[24] E. Sahouria and A. Zakhor, "Motion indexing of video," in *Proc. of IEEE ICIP*, Santa Barbara, 1997.

[25] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. on PAMI*, vol. 11, no. 7, pp. 674–693, July 1989.

[26] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Trans. on PAMI*, vol. 25, no. 5, pp. 564–577, 2003.

[27] J. Sochman and D. C. Hogg, "Who knows who - inverting the social force model for finding groups," in *Proc. of ICCV Workshops*, Barcelona, 2011.

[28] B. Zhou, X. Wang, and X. Tang, "Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents," in *Proc. of IEEE CVPR*, Providence 2012.

[29] C. Tomasi and T. Kanade, "Detection and tracking of point features," Tech. Rep. CMU-CS-91-132, Carnegie Mellon University, 1991.

[30] C. Teuliere, L. Eck, and E. Marchand, "Chasing a moving target from a flying UAV," in *Proc. of IEEE/RSJ IROS*, San Francisco, 2011.

[31] P. Vinukonda, "A study of the scale-invariant feature transform on a parallel pipeline," *MSc thesis. Louisiana State University and Agricultural and Mechanical College*, 2011.

[32] T. Nawaz and A. Cavallaro, "A protocol for evaluating video trackers under real-world conditions," *IEEE Trans. on IP*, vol. 22, no. 4, pp. 1354–1361, April 2013.