Tracking in Wide Area Motion Imagery using Phase Vector Fields

Varun Santhaseelan and Vijayan K. Asari
University of Dayton
300 College Park, Dayton OH 45469
{santhaseelanv1, vasari1}@udayton.edu

Abstract

Tracking in wide area motion imagery (WAMI) is extremely complex because of low resolution of targets, low frame rate and a host of other detrimental environmental factors. In this paper, we propose a robust feature-based method to track objects in WAMI data by exploiting local phase and orientation information based on the monogenic signal representation. We present a detailed monogenic space based analysis to develop a robust method to track objects of low resolution in wide area aerial surveillance imagery. By exploiting local phase information at multiple scales, objects in shadows can be tracked. We also propose an efficient technique to make the feature representation of objects rotation invariant by utilizing local orientation of the object region. The proposed technique is shown to be robust in challenging situations that are characteristic of wide area motion imagery. Effectiveness of the proposed method is illustrated by comparing its performance with dense set of SIFT features and mean shift tracker.

Figure 1. A typical path of a car in CLIF data is marked in yellow on the image in left. Images (a), (b), (c) and (d) show the variation in views of the same car when it traverses the path marked in yellow.

1. Introduction

Huge strides made in the area of sensor technology have enabled data capture of large geographical areas from aerial platforms. One of the preliminary steps in extracting information from the images captured at such high altitudes is to track certain objects of interest. Although the field of object tracking is not new, the process is complicated when it comes to wide area motion imagery (WAMI) because of very low resolution of objects, low frame rate and the effect of other environmental factors. In this paper we tackle the problem of object tracking in WAMI data, specifically in the Columbus Large Image Format (CLIF) [1] dataset. Images in the CLIF dataset were captured using an array of monochromatic sensors at an altitude of 7000 feet and covers an expanse of 2 miles radius. The temporal sampling rate of the CLIF data is around 2 fps. While other datasets containing aerial imagery consist of objects of high resolution in the range of at least a couple of hundred pixels, the objects in CLIF database have an object region that is less than hundred pixels. The considerable distance traveled by an object of interest in between frames because of the low frame rate does not aid the deployment of existing tracking mechanisms.

In this paper, we consider the problem of feature representation of objects that can aid in tracking the object over successive frames. The major scenarios that we have considered in the development of such a detect and hence track mechanism are: (1) tracking in shadows, and (2) tracking when the object makes a turn. Most of the existing algorithms for tracking in WAMI consist of a step to detect moving objects in the scene and then a robust data association algorithm that can accurately predict where the object of interest is based on velocity patterns. The accuracy of such a framework for tracking depends on the accuracy of the algorithm to detect moving objects. Therefore, they tend to fail when the background is similar to the object region or when the object comes to a complete stop. When the ob-

978-0-7695-4990-3/13 $26.00 © 2013 IEEE
DOI 10.1109/CVPRW.2013.123

ject being tracked can be distinctively recognized in successive frames, the risk of losing track when the object is not moving can be eliminated. However, the task of modeling the appearance of objects of interest in WAMI is extremely challenging because of very low resolution.

1.1. Related Work

The most popular work on tracking in CLIF data was presented in [15]. The framework utilized the Hungarian algorithm to build a robust tracking technique based on graph theory. Velocity information was the primary constraint to achieve data association between detected moving objects in any particular scene. The assumption is that the object region stands out in comparison to the surroundings and constraints like shadow regions were not considered, thus enabling efficient extraction of moving objects in a scene. Another approach based on the use of tracklets was proposed in [11]. They used Bayesian networks to solve the data association problem. The method was extended in [14] to enable tracking multiple objects based on prior knowledge of motion patterns of the objects in the scene. These works are improvements made on standard approaches based on Kalman filter, particle filter, etc. to accommodate the challenges posed by the CLIF data. A comprehensive framework for tracking was presented in [10], where velocity and appearance models were taken into consideration for tracking objects in CLIF data. However, the appearance model used in [10] is the squared difference of the intensities of the object region. This measure would be distorted in regions where there is variation in illumination and when the object is rotated.

There has been very little success when it comes to design of feature based tracking methods for WAMI. The main hindrance is the lack of detail that can curtail the attempt to represent objects in a robust and efficient manner. A method proposed in [9] makes use of a histogram based representation of objects. Earth-movers distance was used as the metric to compare histogram of the object model and the test patches. However, the method is very sensitive to changes in illumination because of the use of intensity information to represent the object. In [16], contextual information (roads) was used to detect all moving objects in a frame and hence improve tracking. However, the initial step involved a tracking step which acts as the seed information for algorithm. In such a case, a robust feature representation would aid in generating the initial track.

In this paper, we explore the utility of phase vector fields derived from monogenic signal [5] representation to represent an object. The main contributions of this paper are: (i) development of framework for tracking objects in shadows using phase vector fields, and (ii) development of a rotation invariant representation of phase vector fields that can aid in tracking objects in WAMI.

2. Phase Vector Field from Monogenic Signal Representation

Local phase information of a signal gives ample information regarding the structure of the 2D signal even when illumination changes. Phase congruency [7] was one of the concepts using phase information in an image that was able to represent edges by finding locations where the Fourier components of the signal under consideration were in phase. However, phase congruency does not work for WAMI because of the lack of edge information in images. The analytical signal model in signal processing had enabled the extraction of phase information from signals for the one dimensional case. It was extended to two dimensions by using steerable filters or finding responses of filters in multiple orientations and combining them. The isotropic extension of the concept of analytical signal to two dimensions resulted in the monogenic signal representation. Monogenic signal is used to analyze intrinsic 1D (i1D) signals like edges and lines. While the Hilbert transform was used to find the local phase of a signal in one dimension, Riesz transform was used in the case of the monogenic signal.

The analytical signal model is as given in (1).

$$f_A(x) = f(x) - if_H(x) \tag{1}$$

$f_H(x)$ is the Hilbert transform of $f(x)$.In the frequency domain, $f_H(x)$ is represented as $F_H(u)$ and can be computed as in (2).

$$F_H(u) = H_1(u)F(u) \tag{2}$$

where $F(u)$ is the Fourier transform of $f(x)$ and $H_1(u) = i\,\mathrm{sign}(u)$ is the definition of the Hilbert transform. The phase for the 1D signal can be computed as in (3).

$$\varphi(x) = \arctan(f_H(x), f(x)) \tag{3}$$

The analytical signal model can be extended into 2D using the monogenic signal representation as shown in (4).

$$f_M(x_1, x_2) = (f, f_R)(x_1, x_2) \tag{4}$$

where $f_R(x_1, x_2) = (\mathbf{h} * f)(x_1, x_2)$ and $\mathbf{h} = (h_1, h_2)$ is the Riesz kernel. The spatial and frequency domain representation of the Riesz kernel is given in (5) and (6) respectively.

$$(h_1, h_2)(x_1, x_2) = \left(\frac{x_1}{2\pi|\mathbf{x}|^3}, \frac{x_2}{2\pi|\mathbf{x}|^3}\right), \mathbf{x} = (x_1, x_2) \in \mathbb{R}^2 \tag{5}$$

$$(H_1, H_2)(u_1, u_2) = \left(\frac{u_1}{2\pi|\mathbf{u}|}, \frac{u_2}{2\pi|\mathbf{u}|}\right), \mathbf{u} = (u_1, u_2) \in \mathbb{R}^2 \tag{6}$$

The local phase vector, $\Phi(\mathbf{x})$ from the monogenic signal model is as shown in (7).

$$\begin{aligned} \Phi(\mathbf{x}) &= \frac{f_R(\mathbf{x})}{|f_R(\mathbf{x})|} \arctan \frac{|f_R(\mathbf{x})|}{f(\mathbf{x})} \\ &= \varphi(\mathbf{x}) \exp(i\theta(\mathbf{x})) \end{aligned} \tag{7}$$

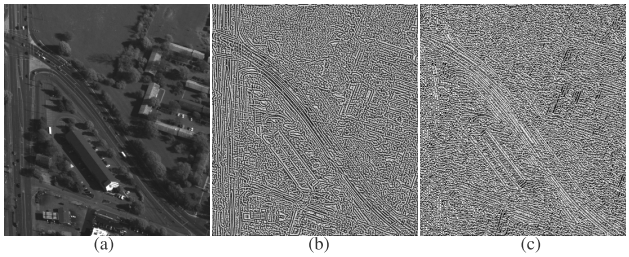

Figure 2. The decomposition of original image (a) into images corresponding to local phase (b) and local orientation (c).

where $\varphi(\mathbf{x})$ is the local phase and $\theta(\mathbf{x})$ is the local orientation and can be computed using the Riesz transform of the signal as shown in (8) and (9).

$$\varphi = \arctan\left(\sqrt{R_1^2\{f\} + R_2^2\{f\}}, f\right), \varphi \in [0, \pi) \quad (8)$$

$$\theta = \arctan\left(\frac{R_2\{f\}}{R_1\{f\}}\right), \theta \in [0, \pi) \quad (9)$$

where $R_1\{f\} = h_1 * f$ and $R_2\{f\} = h_2 * f$. The local magnitude A, which is a measure of local contrast, is defined as in (10).

$$A = \sqrt{f^2(x) + |f_R(x)|^2} \quad (10)$$

In this paper, we utilize the phase vector, for the pixels in the object region to represent the object. Thus, the field created by the object is represented in terms of the ordered pair containing local phase and local orientation, $\Phi = (\varphi, \theta)$. The assumption is that the object region is comprised of intrinsic 1D signals. Phase vector represents the phase and orientation of the intrinsic 1D signal in a local neighborhood. In terms of physical interpretation, the local phase captures the structural information of the object and the local orientation sheds light on the geometric information of the object. The contrast information in the image is given by the local amplitude which we have completely discarded in the phase vector field representation.

In practical scenarios, the signal being considered is of finite length. Therefore, the signal has to be convolved with a band pass filter before the application of Riesz transform. In this paper, we have used the log-Gabor filter [6] as the band-pass filter. The transfer function of a log-Gabor filter is given in (11).

$$G(\omega) = \exp\left(-\frac{(\log(\omega/\omega_0^2))^2}{2(\log(k/\omega_0^2))^2}\right) \quad (11)$$

where ω_0 is the center frequency of the filter, and k/ω_0 remains constant which is a measure of the bandwidth of the filter.

Robust orientation estimation: In scenarios similar to WAMI tracking, the estimation for local orientation could be affected by noise to a very large degree. Unser et al. [17] proposed to have a least square estimate of the orientation based on the local neighborhood. The robust estimate is obtained by maximizing the directional Hilbert transform of the function over a neighborhood as represented by the optimization function in (12).

$$\bar{\theta}(\mathbf{x}) = \underset{\theta \in [-\pi, \pi]}{\arg\max} \int_{\mathbb{R}^2} v_\sigma(\mathbf{x}' - \mathbf{x}) |\mathcal{H}_\theta\{f(\mathbf{x}')\}| d\mathbf{x}' \quad (12)$$

where v_σ is a Gaussian kernel and σ^2 is its variance, $\mathcal{H}_\theta(\cdot)$ is the directional Hilbert transform represented in the frequency domain as in (13).

$$\mathcal{H}_\theta(\omega) = \frac{\omega_x \cos(\theta) + \omega_y \sin(\theta)}{|\omega|} \quad (13)$$

where ω_x and ω_y are the components of angular frequency, ω. The aforementioned maximization problem is formulated as an eigen value problem. The solution is found using the eigen vector corresponding to the largest eigen value of the matrix represented as in (14) [2].

$$[\mathbf{T}(\mathbf{x})]_{mn} = \int_{\mathbb{R}^2} v_\sigma(\mathbf{x}' - \mathbf{x}) R_m\{f(\mathbf{x}')\} R_m\{f(\mathbf{x}')\} d\mathbf{x}' \quad (14)$$

where $m, n = [1, 2]$. The estimate of local orientation, $\bar{\theta}(\mathbf{x})$ can then be redefined as in (15).

$$\bar{\theta}(\mathbf{x}) = \frac{1}{2} \arctan\left(\frac{2[\mathbf{T}(\mathbf{x})]_{12}}{[\mathbf{T}(\mathbf{x})]_{22} - [\mathbf{T}(\mathbf{x})]_{11}}\right) \quad (15)$$

3. Tracking in shadows

One of the major challenges while tracking objects in WAMI is the presence of shadows. Shadows cast by trees or buildings can completely alter the feature representation of an object being tracked. This section gives an insight into how the phase vector field representation can be used to effectively track objects in shadows. The images are assumed to be registered. In our case, we have used Harris corners with SIFT [8] descriptors to match points, filter outliers using RANSAC and thus register images as done in [15].

For the purpose of tracking, we start by manually marking the object of interest to be tracked. The features of the object are extracted and matched to the features in the next frame based on nearest neighbor approach. The best match in the next frame is considered to be the model of the object of interest which will be matched in the subsequent frame.

In our first case, we consider the amplitude of the phase vector field as the feature template to be matched in subsequent frames for detection. Let $\varphi_{\mathbf{x}}$ represent the template containing local phase information of a patch centered at $\mathbf{x} = (x_1, x_2)$ that is of the same size as the object region. The template containing the local phase information of object to be detected is represented as φ_{model}. The location of

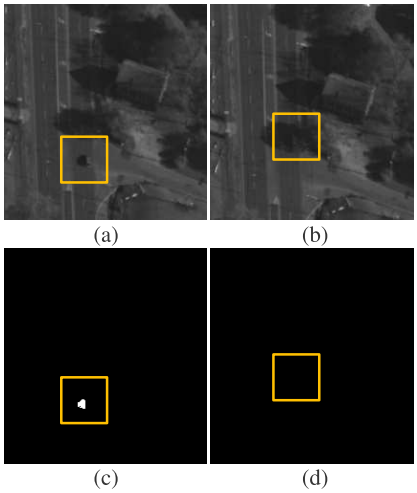

Figure 3. Illustration of why trackers based on moving object detection fail. (a) and (b) are consecutive frames with the yellow box showing the location of the object. (c) and (d) are the binary images of (a) and (b) respectively, showing moving objects in the scene. The object of interest is completely lost in (d) since it moved into the shadow region.

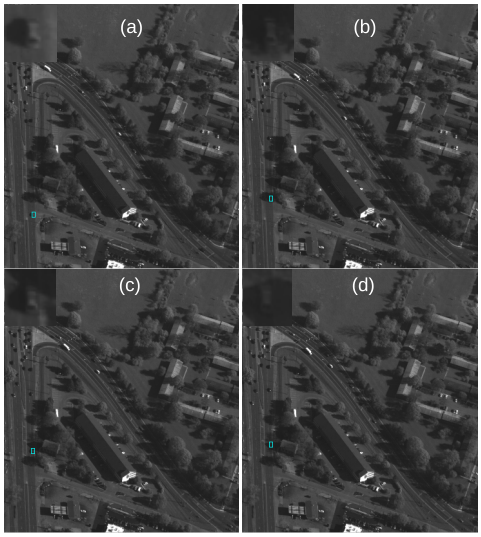

Figure 4. A vehicle (inside cyan box) being tracked through shadow regions in CLIF data. (a) Vehicle in a region with no shadow, (b) vehicle in partial shadow, (c) vehicle in between shadow regions and (d) vehicle in complete shadow.

the object in the next frame, $\mathbf{x}_{n+1}$ is derived from (16) as

$$\mathbf{x}_{n+1} = \arg\min_{\mathbf{x}} \sum_{D} \text{abs}(\varphi_{\mathbf{x}} - \varphi_{model}) \qquad (16)$$

where D represents the patch region. The result of this tracking procedure is shown in Fig. 4

When an object moves into a shadow region, the illumination is reduced in some neighborhood of the object region. This causes usual gradient or intensity based measures to fail. In the monogenic signal model, the effect of illumination change is reflected only in the local magnitude. If we consider the local phase information alone for tracking, the challenges due to change in illumination can be overcome to a very large extent. However, the design of parameters for the band-pass filter is critical to the proper functioning of the algorithm.

Filter design constraints: When objects move into a shadow region, humans are able to recognize the shape within the region of shadow because of the faintly visible edges of the object. This proves that the effective feature needs to contain the high frequency information of the object being tracked. This constraint is incorporated in the design of the band-pass filter, thus resulting in the filter with a center frequency in the high range. The center frequency can be moved towards the lower range depending on the high frequency content in the object being tracked. For instance, the center frequency may be lowered when there are objects of higher resolution (e.g. trucks) and the contrast information in the objects is relatively higher.As the object moves from an area of higher illumination to an area of lower illumination, the edge information allows the object to be tracked. When an object comes out of the shadow region, the most similar feature with respect to its previous appearance in the shadow region is the edges. Therefore, the features at frequencies corresponding to edge information have to be matched in order to have a perfect match. This suggests that the band pass filter that is applied before the application of Riesz kernel should have low bandwidth and high center frequency.

Multi-scale representation: When template of features is matched to patches in the next frame, it is possible that the feature set in a single scale may be inadequate for matching. Therefore, a multi-scale template is created to match with the next frame. In our experiments, we found that using two scales is better than using a single scale for feature matching.

Why template matching? One of the common techniques to represent features is using a neighborhood based descriptor like histogram or SIFT. These techniques can be used for the phase vector field information as well. However, the aforementioned descriptors are constructed from information of the local neighborhood of a pixel. Therefore the effect of the background on the descriptors for the pixel in the object region is much more than desired. The method does not cause any problems for describing objects of high resolution. For WAMI, objects of interest pass through highly varying backgrounds and this could cause the feature matching technique to fail. For example, a car being tracked on the road may not be detected when it is within a zebra crossing. Therefore, the aim in our research has been to reduce the dependence of feature representation on the neighborhood of an object region. Since we manually marked the object region in our research experiments, it was made sure that not much of the background is enclosed

in the bounding box.

4. Rotation invariant representation for tracking

Amplitude of the phase vector is sufficient to detect and track an object from frame to frame as long as orientation of the object does not change. The template matching procedure is affected when the direction of movement of the object changes. However, it is to be noted that the whole object region is rotated and therefore the direction of the phase vector field is altered by the amount of rotation of the object. This implies that the local orientation component of the vector has a similar change throughout the region of the object. In our research we have developed the use of a relative orientation map of the object region that is being tracked to deal with variation in rotation. The key idea is that the local orientation relative to the local orientation of centroid of the object region remains the same throughout the object region even the object is rotated. Therefore, when the object is rotated, we match the relative orientation map of the object region to detect the object in subsequent frames. Let $\theta_{\mathbf{x}}$ represent the local orientation at a pixel location, $\mathbf{x} = (x_1, x_2)$. If D represents the object region, then the template containing local orientation of the object region can be transformed to a relative orientation map as shown in (17).

$$\theta'_{\mathbf{x}} = \theta_{\mathbf{x}} - \theta_{centroid} \tag{17}$$

where $\theta_{centroid}$ is the local orientation at the centroid of the pixel. (See Fig. 5 for illustration.) The relative orientation map of the object region can be matched to the next frame by finding the patch of the same size using the nearest neighbor approach as shown in (18).

$$\mathbf{x}_{n+1} = \arg\min_{\mathbf{x}} \sum_{D} \text{abs}(\theta'_{\mathbf{x}} - \theta'_{model}) \tag{18}$$

where $\mathbf{x}_{n+1}$ is the location in the next frame, D represents a patch of size equal to that of the object, $\theta'_{\mathbf{x}}$ is the relative orientation map of the test patch and θ'_{model} is the relative orientation map of the object region. Sample frames from video in which a turning car is tracked is shown in Fig. 6.

Filter design constraints: The critical part of computing the relative orientation map is the computation of the local orientation at the centroid, $\theta_{centroid}$. If $\theta_{centroid}$ is not related to the general orientation of the object of interest or if it is affected by the noise in the image, then the method will fail. In order to avoid such a failure, the local orientation is computed after the application of a band-pass filter with low center frequency. This would allow the dominant orientation of the signal in the neighborhood of the centroid to be captured and hence making the relative orientation map invariant to rotation.

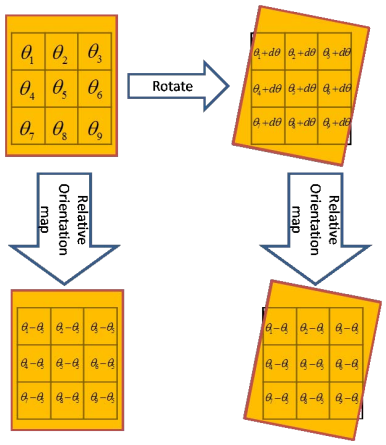

Figure 5. Illustration of how the relative orientation map is computed. The squares represent the pixels and contain the respective orientations. The yellow region represents the object of interest.

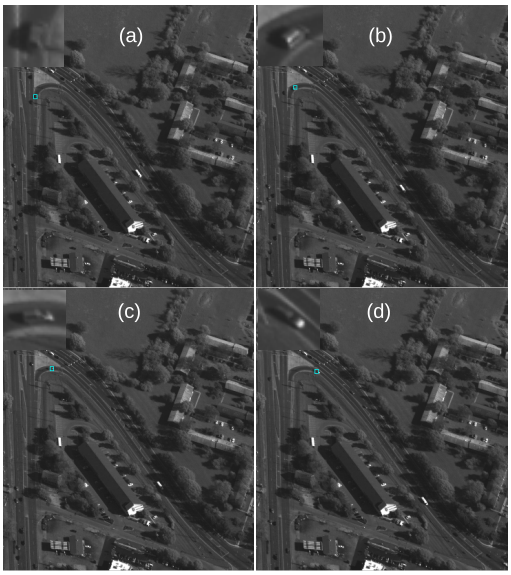

Figure 6. Tracking an object (in cyan box) while turning to demonstrate rotation invariance. Alternate frames in a sequence are shown in (a) to (d) with the magnified view of object in inset.

Effect of background: The orientation of the edges (i1D signals) that significantly contribute to the object shape is to be considered during the feature matching process. This criterion is taken care of during the design of the filter. During our experiments, it has been observed that the background contributes towards perfect definition of the feature template in terms of the relative orientation map. This is due to the fact that the orientation of the object is relevant with respect to its background as reference.

5. Framework for Tracking

In the previous two sections, we demonstrated how the local phase and local orientation can separately help in detecting an object in subsequent frames. In this section, we

present a framework to integrate both local phase and local orientation features to have a robust feature tracking mechanism.

If motion in video is segmented using local phase information, then almost all changes that happen in a scene can be detected. In order to model the background, we use the technique from [15]. The difference in phase information of the background and the actual scene represents the change that has occurred. Therefore the search for the object of interest can be narrowed down to the region of changes along with the prior location of the object. In this scenario, we take the liberty of representing the local phase information of the object as a histogram. This representation would negate the effect of rotation. Thus the new feature representation for an object is the histogram of phase concatenated with the relative local orientation.

The architecture of designed framework is shown in Fig. 7 and an illustration of change detection in phase domain is shown in Fig. 8. It is to be noted that the previous location of the object of interest is always a candidate location to be searched irrespective of change detection.

When there is variation in illumination, it is reflected in the amplitude component of the monogenic signal model. This measure would dictate the dependence of the final correlation score on the phase component or the orientation component. In our case, we calculated the mean magnitude for the selected patch, D. If the value crosses a threshold, then the weights would be the same. A reduced value for the mean of magnitude would indicate a greater weight for the correlation score based on phase component.

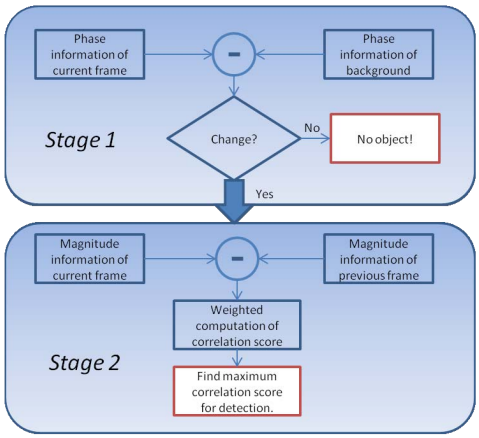

Figure 7. Block diagram illustration of the framework for feature based detect and track mechanism.

6. Experiments

We conducted experiments in multiple scenarios that are characteristic of WAMI. For the experiments, we have considered the images from a single sensor only. The images can be stitched using existing algorithms like [13] as well.

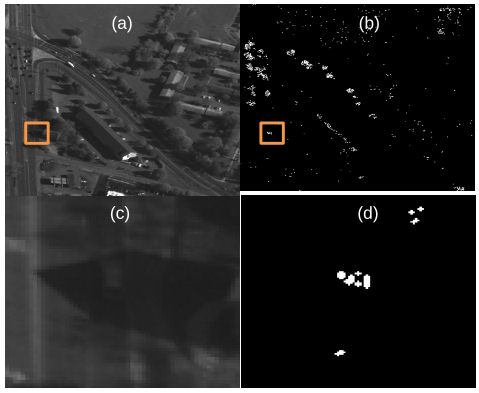

Figure 8. Improved change detection based on difference of phase component. (a) original image, (b) change detected in terms of local phase for object in shadow, (c) the object of interest in shadow and (d) binary image of change detected for the object region within the shadow.

In previous research related to WAMI, the emphasis has been on tracking multiple objects in situations where the object is completely visible or occluded for a short time. In this section, we explore some challenging situations where the feature based tracking process will be a powerful aid to the development of a perfect tracking framework as well as a quantitative comparison with the mean shift tracker.

6.1. Quantitative evaluation

We compared the performance of the proposed method with that of the mean shift tracker [4]. Object tracking error (OTE) from [3] was the metric used for comparison and is computed using (19).

$$OTE = \frac{1}{N_{rg}} \sum \sqrt{(xg_i - xr_i)^2 + (yg_i - yr_i)^2} \quad (19)$$

where N_rg is the number of frames for which the ground truth and the detections are available, (xg_i, yg_i) is the ground truth and (xr_i, yr_i) is the location of detection during tracking. The locations are computed as the centroid of the bounding box of the tracked objects.

The comparison of metrics for the proposed tracker and the mean shift tracker for the scenario displayed in Fig. 1 is shown in Fig. 9. Another comparison is illustrated in Fig. 10 where illumination of scene is very low. From the results, it was evident that the proposed tracker worked much better than the mean shift tracker.

6.2. Qualitative evaluation

In the first part of evaluation, we tested the algorithm on various cars that pass through shadows. One of the results is shown in Fig. 11. The car goes in and out of shadow regions thus making it a tough problem to tackle. The proposed framework was able to consistently track the vehicle as illustrated.

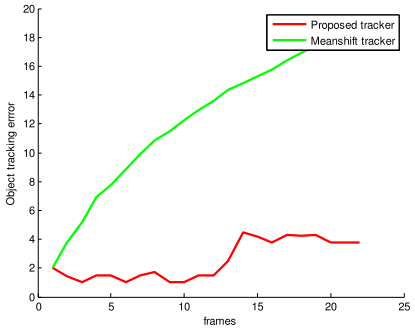

Figure 9. The comparison of object tracking error for proposed tracker and the mean shift tracker for track in Fig.1.

As the second part of qualitative evaluation, the proposed feature representation is compared with the performance of SIFT features. A dense set of SIFT features has been proven to be one of the most effective representation for regions. Since only regions can be matched from frame to frame for WAMI, we analyzed the performance of SIFT features as well. The key strength of the SIFT feature is its descriptor. The descriptor is computed as the magnitude weighted histogram of gradients of the pixels in the local neighborhood of a point. When every pixel in the object region is represented using corresponding SIFT descriptor, the effect of background in the overall description of an object becomes significant. Therefore, slight change in background coupled with the presence of similar objects in the vicinity causes SIFT feature set to fail. This is illustrated in Fig. 12.

7. Conclusion

We have proposed an effective framework for tracking objects in WAMI using phase vector fields. The technique is robust to variations in illumination thus enabling tracking of objects in regions containing shadows. The rotation invariance property of phase vector fields aids in persistent surveillance of objects in WAMI. The method can be used in conjunction with other robust multi-object tracking mechanisms as well, thus paving the way towards perfection in the realm of object tracking for wide area surveillance. The other major challenge in tracking in WAMI is the presence of large occlusions like buildings. If the presence of buildings and their effect on the path of the object can be modeled as in [12], the proposed feature matching technique can be employed to achieve a robust data association method after the object comes out from the occlusion.

References

[1] Columbus large image format dataset, https://www.sdms.afrl.af.mil/index.php?collection=clif2007, 2007.

[2] M. Alessandrini, A. Basarab, H. Liebgott, and O. Bernard. Myocardial motion estimation from medical images using the monogenic signal. *IEEE Transactions on Image Processing*, 22(3):1084–1095, 2013.

[3] J. Black, T. Ellis, and P. Rosin. A novel method for video tracking performance evaluation. In *In Joint IEEE Int. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS)*, pages 125–132, 2003.

[4] D. Comaniciu and V. Ramesh. Mean shift and optimal prediction for efficient object tracking. In *Proceedings of International Conference on Image Processing*, volume 3, pages 70–73, 2000.

[5] M. Felsberg and G. Sommer. The monogenic signal. *IEEE Transactions on Signal Processing*, 49(12):3136–3144, 2001.

[6] D. J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A*, 4(12):2379–2394, 1987.

[7] P. Kovesi. Image features from phase congruency. *Videre: Journal of computer vision research*, 1999.

[8] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, Nov. 2004.

[9] A. Mathew and V. Asari. Local region statistical distance measure for tracking in wide area motion imagery. In *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 248–253, 2012.

[10] T. Pollard and M. Antone. Detecting and tracking all moving objects in wide-area aerial video. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 15–22, 2012.

[11] J. Prokaj, M. Duchaineau, and G. Medioni. Inferring tracklets for multi-object tracking. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 37–44, 2011.

[12] J. Prokaj and G. Medioni. Using 3d scene structure to improve tracking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1337–1344. IEEE Computer Society, 2011.

[13] J. Prokaj and G. Medioni. Accurate efficient mosaicking for wide area aerial surveillance. In *IEEE Workshop on Applications of Computer Vision (WACV)*, pages 273–280, 2012.

[14] J. Prokaj, X. Zhao, and G. Medioni. Tracking many vehicles in wide area aerial surveillance. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 37–43, 2012.

[15] V. Reilly, H. Idrees, and M. Shah. Detection and tracking of large number of targets in wide area surveillance. In *Proceedings of the 11th European conference on computer vision conference on Computer vision: Part III*, pages 186–199, 2010.

[16] X. Shi, H. Ling, E. Blasch, and W. Hu. Context-driven moving vehicle detection in wide area motion imagery. In *21st International Conference on Pattern Recognition (ICPR)*, pages 2512–2515, 2012.

[17] M. Unser, D. Sage, and D. Van De Ville. Multiresolution monogenic signal analysis using the riesz-laplace wavelet transform. *IEEE Transactions on Image Processing*, 18(11):2402–2418, 2009.

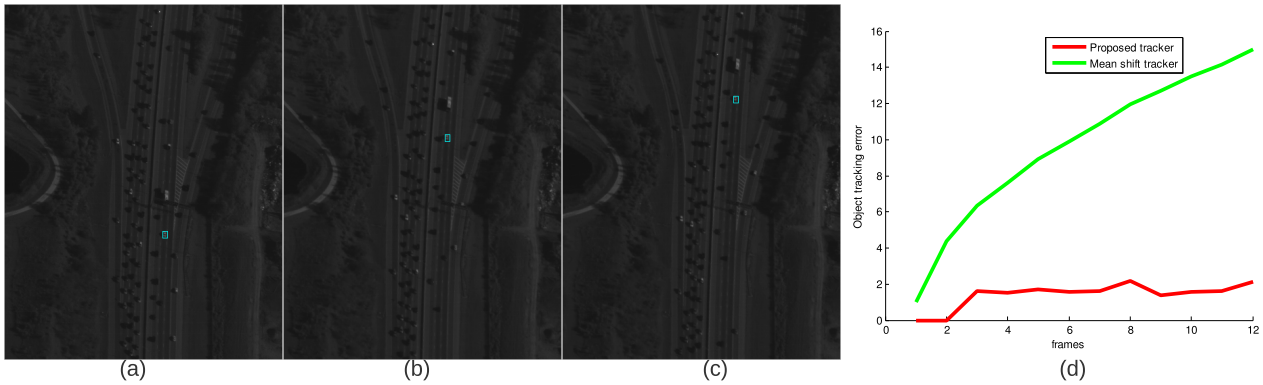

Figure 10. Performance of the tracker in a scene with low illumination. (a), (b) and (c) illustrate the performance of proposed tracker and (d) is the comparison with mean-shift tracker for the same object.

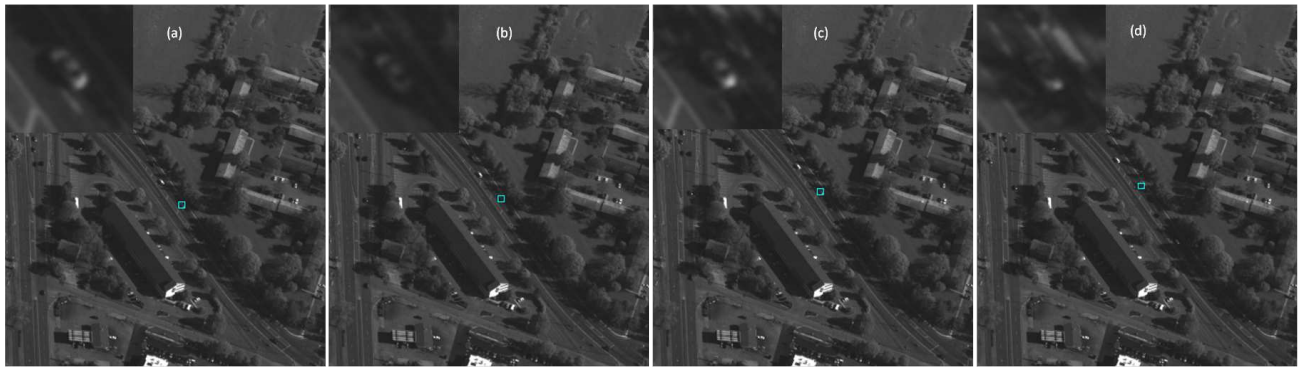

Figure 11. Car in the inset is tracked through regions containing shadow. The appearance of car in each image is shown in inset.

Figure 12. Comparison with dense SIFT features. (a) Object of interest marked in frame 1. (b) SIFT features matched to wrong car in frame 2. (c) Incorrect matching resulted in incorrect association. (d) Same object marked in cyan box. (e) Magnified view of object and neighborhood. (f) Search and match to same object in frame 2. (g) The overall search region to find the object in frame 2.