

University of Dayton eCommons

Electrical and Computer Engineering Faculty
Publications

Department of Electrical and Computer
Engineering

9-2015

Directional Ringlet Intensity Feature Transform for Tracking

Evan Krieger
University of Dayton

Paheding Sidike
University of Dayton

Theus H. Aspiras
University of Dayton

Vijayan K. Asari
University of Dayton, vasari1@udayton.edu

Follow this and additional works at: http://ecommons.udayton.edu/ece_fac_pub

 Part of the [Electromagnetics and Photonics Commons](#), and the [Optics Commons](#)

eCommons Citation

Krieger, Evan; Sidike, Paheding; Aspiras, Theus H.; and Asari, Vijayan K., "Directional Ringlet Intensity Feature Transform for Tracking" (2015). *Electrical and Computer Engineering Faculty Publications*. Paper 375.
http://ecommons.udayton.edu/ece_fac_pub/375

This Conference Paper is brought to you for free and open access by the Department of Electrical and Computer Engineering at eCommons. It has been accepted for inclusion in Electrical and Computer Engineering Faculty Publications by an authorized administrator of eCommons. For more information, please contact frice1@udayton.edu, mschlangen1@udayton.edu.

DIRECTIONAL RINGLET INTENSITY FEATURE TRANSFORM FOR TRACKING

Evan Krieger, Paheding Sidike, Theus Aspiras, Vijayan K. Asari

The University of Dayton, Dayton, OH 45469, USA

ABSTRACT

The challenges existing for current intensity-based histogram feature tracking methods in wide area motion imagery include object structural information distortions and background variations, such as different pavement or ground types. All of these challenges need to be met in order to have a robust object tracker, while attaining to be computed at an appropriate speed for real-time processing. To achieve this we propose a novel method, Directional Ringlet Intensity Feature Transform (DRIFT), that employs Kirsch kernel filtering and Gaussian ringlet feature mapping. We evaluated the DRIFT on two challenging datasets, namely Columbus Large Image Format (CLIF) and Large Area Image Recorder (LAIR), to evaluate its robustness and efficiency. Experimental results show that the proposed approach yields the highest accuracy compared to state-of-the-art object tracking methods.

Index Terms— object tracking, Kirsch mask, Gaussian ringlet, image enhancement, feature extraction

1. INTRODUCTION

Object tracking in low resolution wide area motion imagery (WAMI) datasets is a challenging task. The challenges for tracking in WAMI data include object occlusion, rotation, scaling, illumination changes, and background variations. Most object tracking algorithms intend to tackle these challenges using various feature extraction techniques. Several feature extraction methods use intensity-based histograms that perform only by using intensity information and creating a histogram of features for classification. The feature extraction methods that use intensity-based histograms differ in how they partition the image for constructing the histogram. Scale invariant Feature Transform (SIFT) and Speeded Up Robust Features (SURF) methods are scale invariant feature descriptors [1,2]. SIFT and SURF have been shown to be improper choices for the small object tracking in WAMI data. The Histogram of Oriented Gradients (HOG) descriptor uses simple gradient filters, such as Prewitt filters, to create a gradient image [3]. The gradient orientation of localized sections of the image is then used to compute the feature descriptor. The Local Binary Pattern (LBP) method is a texture descriptor that creates a descriptor for each pixel based on its neighboring pixels [4].

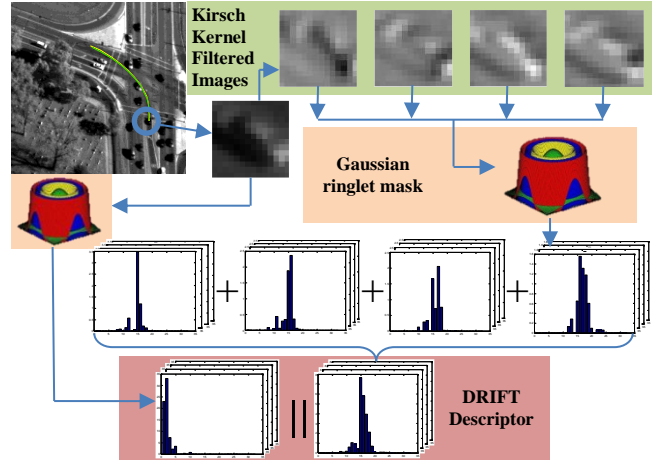


Figure 1. Proposed DRIFT scheme.

The Rotation Invariant Local Binary Pattern (RILBP) creates a rotation invariant feature descriptor by using the magnitude of the discrete Fourier transform of the histogram [5]. The Rectangular Grid (RECT) method uses a rectangular grid pattern to partition the image for feature extraction [6]. The equal distance circular grid (CIRC-ED) method is created by partitioning the image into different equal distance circular rings [7]. This method is illumination and rotation invariant and can also handle partial occlusions. The equal area circular grid (CIRC-EA) method is similar but the image is partitioned by circular rings with equal areas [8]. The use of square ring histograms was proposed as an improvement of the RECT method in [9]. Improved features can be created using center weighted histograms as proposed for the rectangular grid in [10]. The circular grid methods can also weigh the different rings (WCIRC-ED and WCIRC-EA) with an emphasis on the center ring to obtain a stronger feature [11]. Recently, a feature tracking algorithm, named Gaussian ringlet intensity distribution (GRID) [11], is proposed which showed robustness for rotationally invariant tracking. Due to background variations, such as different pavement or ground types, and object structural information distortions, the abovementioned feature based tracking methods may encounter problems. The main reason is that the intensity changes around the target objects causes feature mismatching. These challenges necessitate a stronger feature descriptor to be constructed. Therefore, we propose a Gaussian ringlet masking strategy that utilizes rotational invariance of the Gaussian ringlet and directional edge information of the Kirsch kernel. The proposed technique,

Directional Ringlet Intensity Feature Transform (DRIFT) as seen in Fig. 1, weighs the intensity and edge information of the reference object with the ringlet features to be robust to object distortions and background variations.

2. PROPOSED METHOD

The proposed algorithm can be described from four aspects; feature extraction, intensity limitation, Kirsch masking, and nonlinear enhancement. These aspects robustly handle object tracking challenges including occlusion, rotation, orientation, and illumination.

2.1. Feature extraction: Occlusion and rotation handling

The proposed object tracking scheme is depicted in Fig. 2. In the first frame, the center point and the radius of the reference object is selected. The method uses a Kalman filter to estimate where the center of the object will be in the next frame. The estimation is determined using the previous information of the object to predict its trajectory. This estimation is then used to determine the search area for the object. The search area is determined through properties of the camera setup and the estimated properties of the object. These properties are used to determine the maximum distance, in pixels, the object can travel in one frame. This distance is used as the radius of the search area. The feature extraction method is then used to find the features in the search area. Classification is used to determine the most likely center point candidate from the extracted features. The classification used in this tracking method is determined by the minimum Earth Mover's Distance [12]. The most likely point is obtained along with a confidence factor. A threshold is used with the confidence factor to determine if the obtained point is valid. If the confidence is high the obtained point will be considered the new center pixel of the object. If the confidence is low, the estimated point from the Kalman filter will be used as the new center pixel of the object. If an occlusion occurs, the Kalman filter will be used as the location estimator until the object exits the occlusion. The final step of the tracking process is to use the new center point information to update the Kalman Tracker and the reference object features.

In the first step of feature extraction we use intensity-based histograms on an image that is partitioned using Gaussian ringlets [11]. A ring method like in CIRC [7] is used except that each ring is not a discrete filter but a Gaussian ring function. The center Gaussian ring is the Gaussian function

$$G_1(x, y) = ce^{-\frac{(x-x_o)^2+(y-y_o)^2}{\frac{1}{2}(R_i-R_{i-1})^2}} \quad (1)$$

where c is a constant, x_o and y_o are the coordinates of the center points, and R_i is the radius of ring i . The other Gaussian rings are computed as

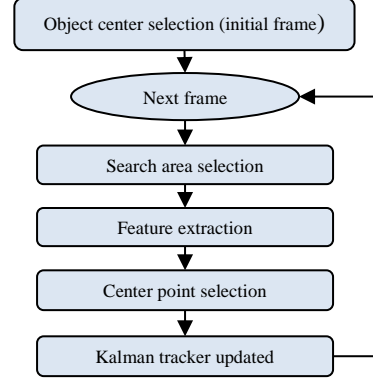


Figure 2. Object tracking method.

$$G_i(x, y) = ce^{-\frac{(\sqrt{(x-x_o)^2+(y-y_o)^2}-\frac{1}{2}(R_i-R_{i-1}))^2}{\frac{1}{2}(R_i-R_{i-1})^2}} \quad (2)$$

A histogram is calculated for each ring with respect to the Gaussian ring as a mask. These histogram are normalized using

$$N_i = \sum_{v=0}^V H_i(v) \quad (3)$$

where N_i is the normalization factor for ring i , H_i is the histogram for ring i , and V is the number of bins in the histogram. Each value of the histogram is divided by this normalization factor to equally weight each ring of the histogram. Additional weighting can also be applied to the histograms to emphasize the central rings. For this evaluation, a total of four rings each having 32 bin histograms were used to reduce the feature size without a loss of accuracy.

2.2. Intensity limitation: Speed improvement

The second proposed aspect of the feature extraction method is to limit the search area of the image. In the GRID algorithm, each pixel of the search area is computed for feature matching. This can be computationally intensive, especially as the search area increases. By limiting the search area, the computation time can be decreased significantly with little effect on the accuracy of the tracker. Therefore, we use an intensity based pixel selection method to remove areas that have a large illumination difference compared to the reference object. This is achieved by first finding the average value of the test area of each point in the search area expressed by

$$I_{avg}(x, y) = \frac{1}{d_o^2} \sum_{n=x-d_o}^{x+d_o} \sum_{m=y-d_o}^{y+d_o} I(n, m) \quad (4)$$

where $I_{avg}(x, y)$ is the image of averaged test objects, $I(x, y)$ is the search area image, and d_o is the object diameter for a square object. The search area is then computed by comparing $I_{avg}(x, y)$ to a set of low and high limits. These

limits are based around the average of the reference image computed as

$$limit_{high} = \frac{1}{d_o^2} \sum_{n=0}^{d_o} \sum_{m=0}^{d_o} I_{ref}(n, m) + l \quad (5)$$

$$limit_{low} = \frac{1}{d_o^2} \sum_{n=0}^{d_o} \sum_{m=0}^{d_o} I_{ref}(n, m) - l \quad (6)$$

where l is the limiting intensity difference factor. The value of l will determine the amount of undesired search area points that will be removed. A higher l will decrease the speedup potential while reducing the likelihood of removing the best matched object.

It was found through experimentation that the accuracy for most objects will not be affected until the limiting factor l is small (~ 15 for an 8 bit image). This is due to the properties of the Gaussian ringlet algorithms and the Kalman based tracking method. The GRID algorithm is affected by large illumination changes and will be unable to match objects that have gone under global illumination changes. In addition, the tracker will use the Kalman filter results for insufficiently matched objects. A large global illumination change will result in a mismatched object with or without the limiting factor. The only real risk is for partially occluded objects. A large enough limiting factor is needed to retain the partially occluded objects as valid search points.

2.3. Kirsch masking: Orientation handling

The third part of the proposed algorithm is the inclusion of features from Kirsch operator filtered images into the feature descriptor. It is observed from the results of the tracked cars using GRID method that the vehicles are lost when there is a change in intensity in the background around the target object. This occurs because the reference object is a square around the vehicle center which results in pavement and nearby objects being included as part of the reference object. As the vehicle drives, the area around the car can drastically change in intensity. In order to reduce this effect, some edge information of the object can be utilized as part of the feature descriptor. This will be accomplished using Kirsch compass kernels to filter the input frame.

The Kirsch compass kernels are non-linear edge detectors [13]. The Kirsch operator includes a single kernel mask that is rotated in eight directions. This rotation finds the maximum edge strength in 45 degree increments. The kernels are seen in Fig. 3. It is found that not all eight directions were needed to have a positive impact of the accuracy of the tracker. It is determined that by only using four kernels the accuracy is increased significantly while the computation time is kept low.

The implementation of the Kirsch compass kernels into the tracking algorithm is to first filter the frame with each of the four kernels. The four filtered images are used with the Gaussian ring masks to create the feature histograms. The

four histograms are then added together and vertically concatenated with the Gaussian ringlet histograms of the unfiltered image to create the final 8 by 32 feature descriptor as shown in Fig. 1.

$\begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{bmatrix}$	$\begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix}$	$\begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix}$	$\begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix}$
East	West	Northeast	Southwest
$\begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{bmatrix}$	$\begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{bmatrix}$	$\begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}$	$\begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}$
North	Northwest	Southeast	South

Figure 3. Kirsch compass kernels

To create the feature histogram with the Kirsch filtered image, the image should be normalized first to be between [0 255]. The normalized filtered image can be expressed as

$$I_{Knorm} = \begin{cases} 0 & I_{Kirsch} \leq -127 \\ I_{Kirsch} + 127, & 127 < I_{Kirsch} < 127 \\ 255, & I_{Kirsch} \geq 127 \end{cases} \quad (7)$$

This normalization will allow the filtered images to be used like the 8-bit intensity images as inputs for the feature extraction. To reduce the amount of information that is cut-off in this normalization, the filtered image may be expressed as

$$I_{Kirsch}(x, y) = \sum_{m=-1}^1 \sum_{n=-1}^1 Kirsch_i(m, n) \frac{1}{b} I(m+x, n+y) \quad (8)$$

where $Kirsch_i$ is the Kirsch kernel in the i th direction and b is a constant that reduces that maximum and minimum value of $I_{Kirsch}(x, y)$. The value of b should be selected to reduce the amount of information cutoff in the normalization in Eq. 7. The value for b selected for evaluations is $b = 10$ which gives a better feature descriptor by increasing the range of the typical vehicle edge. The four Kirsch kernel filtered images of a reference object can be seen in Fig. 1.

2.4. Nonlinear enhancement: Illumination handling

The fourth part of the proposed algorithm is to use a nonlinear image enhancement function to improve the illumination in the imagery. The image enhancement will allow object in underexposed or overexposed regions to be more visible. It also allows better features to be obtained by the feature extraction. The self-tunable transformation function (STTF) enhancement algorithm was selected as the preprocessing algorithm [14-15]. The algorithm using image enhancement will be referred to as DRIFT-STTF.

3. EXPERIMENTAL RESULTS AND DISCUSSION

Database: The DRIFT algorithm introduced in this paper is tested by finding its tracking ability in an object tracking scenario in WAMI data. The datasets used in this experimentation are the Columbus Large Image Format (CLIF)

Table 1. Object Tracking Frame Detection Accuracy (%)

	CLIF 1	CLIF 2	CLIF 3	CLIF 4	LAIR 1	LAIR 2	LAIR 3	LAIR 4	Average
HOG [1]	40.10	8.05	16.11	11.35	9.94	13.98	5.26	22.99	15.97
LBP [2]	46.45	8.50	5.00	47.30	5.26	17.43	10.09	16.11	19.52
RILBP [3]	58.00	7.55	9.17	8.70	9.36	24.26	5.26	14.37	17.08
RECT [4]	82.40	75.40	61.39	6.60	47.69	73.85	81.08	24.65	56.63
CIRC-EA [6]	65.90	75.25	83.06	68.30	35.61	81.58	77.69	55.90	67.91
CIRC-ED [5]	69.90	69.25	83.06	68.50	39.38	49.75	79.12	68.06	65.88
GRID-EA [7]	18.00	69.35	84.72	60.50	37.23	76.23	76.69	42.57	58.16
GRID-ED [7]	18.00	69.10	81.53	69.70	38.47	73.11	76.77	26.04	56.59
WCIRC-EA [7]	75.35	72.75	81.53	67.60	36.19	75.49	77.51	40.69	65.89
WCIRC-ED [7]	70.00	67.85	81.53	70.65	38.73	79.69	84.43	67.50	70.05
WGRID-EA[7]	73.25	68.80	81.67	69.90	37.75	76.23	83.91	37.78	66.16
WGRID-ED[7]	80.30	68.35	81.53	73.40	38.21	78.37	82.12	85.21	73.44
DRIFT	85.70	68.25	88.06	79.30	61.40	86.51	83.86	85.28	79.80
DRIFT -STTF	82.60	58.10	94.31	71.90	68.81	90.21	88.86	86.32	80.14

[16] dataset and the Large Area Image Recorder (LAIR) public released dataset.

Test Setup: The sequences used in this experimentation are registered to remove shifts caused by the moving sensors [17]. The data is captured at approximately two frames per second. The eight object sequences that are used in the experimentation are selected for their challenging aspects. These challenges include object turning, similar objects in scene, small object (8-10 pixels across), changes in pavement, and changes in lighting. These sequences can be seen in the yellow line in Fig. 4. The sequences used were between 10 and 20 frames. The initial search area around the object has radius of 15 pixels for each method. This search area was determined using object and dataset information to make sure the object will be within the search area. The evaluation is performed by comparing the truth data to the testing results. The truth data is a collection of center points of the object in the sequence. The evaluation performed using this information is the Frame Detection Accuracy

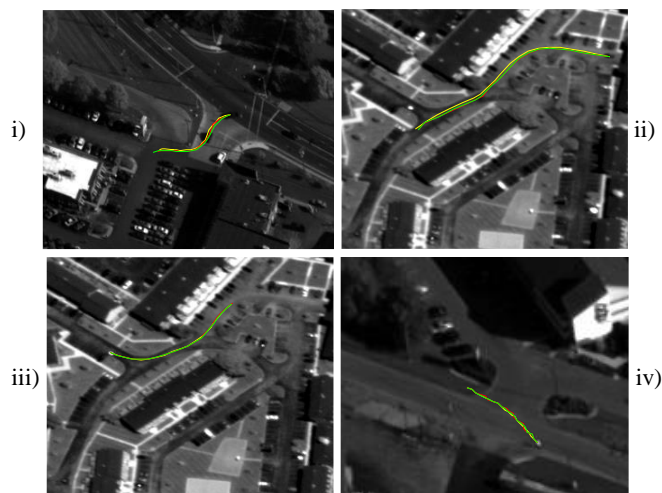


Figure 4. Object tracking: i) Car 1 in set CLIF 4; ii) Car 2 in set LAIR 1; iii) Car 3 in set LAIR 2; iv) Pedestrian in set CLIF 3. Yellow-ground truth, red-DRIFT, and green-DRIFT-STTF.

(FDA) [18]. The FDA in each frame is the overlap of the truth object with the testing result object. These results are averaged over all frames of the sequence to obtain the average area detection result.

Results and Comparison: The evaluation is performed using many different tracking methods. We also investigated the performance of the proposed method without image enhancement, i.e., DRIFT, and with image enhancement, i.e., DRIFT -STTF, as shown in Table 1. The timing results for some of the methods are summarized in Table 2. From Table 1, it is evident that the proposed DRIFT and DRIFT-STTF methods give overall better results compared to the alternate tracking methods. The tracking paths for sequences in Table 1 using DRIFT and DRIFT-STTF are shown in Fig. 4 with the red and green lines, respectively. It can be seen from these results that the proposed method tracks the objects more accurately. The timing results from Table 2 indicate that the processing speed of DRIFT is faster than WGRID-ED algorithm. However, the DRIFT-STTF has an increased computation time because of the image enhancement preprocessing. The computing platform is an Intel Core i7-3630QM CPU – 2.4GHZ dual core processor and 8GB RAM. The operating system is Windows 7 Professional Edition. All evaluations were implemented in MATLAB.

Table 2. Processing speed comparison

	WGRID-ED	DRIFT	DRIFT-STTF
Average Time (sec./ frame)	1.827	1.186	2.227

4. CONCLUSIONS

We presented a novel object tracking algorithm that can handle the challenges of background variations and object distortions. The proposed method, DRIFT, constructs a stronger feature descriptor by utilizing methods including Gaussian ringlet masking and Kirsch kernel filtering. Test results show that that the proposed method improved the accuracy of the tracking process while having a positive effect on the computation time.

5. REFERENCES

- [1] H. Bay, A. Ess, T. Tuytelaars, and L. Gool, "SURF: Speeded Up Robust Features," *Computer Vision and Image Understanding (CVIU)*, vol. 110(3), pp. 346–359, 2008.
- [2] D. G. Lowe, "Distinctive Image Features from Scale-invariant Keypoints," *International Journal Computer Vision*, vol. 60(2), pp. 91–110, 2004.
- [3] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 886–893, 2005.
- [4] T. Ojala, M. Pietikäinen, and M. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 971–987, 2002.
- [5] T. Ahonen, J. Matas, C. He, and M. Pietikäinen "Rotation invariant image description with local binary pattern histogram Fourier features," *In Image Analysis*, pp. 61–70, 2009.
- [6] A. Mathew and V. K. Asari, "Local histogram based descriptor for tracking in wide area imagery," *Lecture Notes in Computer Science, Published by Springer-Verlag Berlin Heidelberg, Wireless Networks and Computational Intelligence (Communications in Computer and Information Science series)*, vol. 292, pp. 119–128, 2012.
- [7] A. Soetedjo and K. Yamada, "Traffic sign classification using ring partitioned method," *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, E88-A(9), pp. 2419–2426, 2005.
- [8] Z. Tang, X. Zhang, and S. Zhang, "Robust perceptual image hashing based on ring partition and nmf," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, pp. 711–724, 2014.
- [9] W. Xiaoling and M. Hongyan, "Enhancing Color histogram for Image Retrieval," *In Int. Workshop on Information Security and Application*, pp. 622-625, 2009.
- [10] S. He, Q. Yang, R. W. Lau, J. Wang, and M.-H. Yang, "Visual tracking via locality sensitive histograms," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2427–2434, 2013.
- [11] T. Aspiras, V. K. Asari, and J. Vasquez, "Gaussian ringlet intensity distribution (GRID) features for rotation-invariant object detection in wide area motion imagery," *IEEE International conference on Image Processing*, 2014.
- [12] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *International Journal of Computer Vision*, vol. 40, pp. 99-121, 2000.
- [13] B. H. Shekar, K. Raghurama Holla, and M. Sharmila Kumari, "KID: Kirsch Directional Features Based Image Descriptor," *Pattern Recog. and Machine Int. Lecture Notes in Comp. Sci.*, vol. 8251, pp. 327-334, 2013.
- [14] S. Arigela and V. K. Asari, "Self-tunable transformation function for enhancement of high contrast color images," *J. Electron. Imaging*, vol. 22(2), pp. 023010, 2013.
- [15] E. Krieger, V. K. Asari and S. Arigela, "Color image enhancement of low-resolution images capture in extreme lighting," *Proc. SPIE 9120*, pp. 91200Q, 2014.
- [16] CLIF data from AFRL, ID: HAAA08E09D, <https://www.sdms.afrl.af.mil/index.php?collection=clif2007>.
- [17] B. Zitova and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, pp. 977–1000, 2003.
- [18] V. Manohar, P. Soundararajan, D. Goldgof, R. Kasturi and J. Garofolo, "Performance Evaluation of Object Detection and Tracking in Video," *In Proc. of the seventh Asian Conference on Computer Vision*, pp. 151-161, 2006.