4-2015

Intensity and Resolution Enhancement of Local Regions for Object Detection and Tracking in Wide Area Surveillance

Evan Krieger
University of Dayton

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

Saibabu Arigela
University of Dayton

Theus H. Aspiras
University of Dayton

Intensity and resolution enhancement of local regions for object detection and tracking in wide area surveillance

Evan Krieger, Vijayan K. Asari, Saibabu Arigela, Theus Aspiras

University of Dayton, Department of Electrical and Computer Engineering, Dayton, Ohio 45469

E-mails: kriegere1@udayton.edu, vasari1@udayton.edu, arigelas1@udayton.edu, theus.aspiras@gmail.com

ABSTRACT

Object tracking in wide area motion imagery is a complex problem that consists of object detection and target tracking over time. This challenge can be solved by human analysts who naturally have the ability to keep track of an object in a scene. A computer vision solution for object tracking has the potential to be a much faster and efficient solution. However, a computer vision solution faces certain challenges that do not affect a human analyst. To overcome these challenges, a tracking process is proposed that is inspired by the known advantages of a human analyst. First, the focus of a human analyst is emulated by doing processing only the local object search area. Second, it is proposed that an intensity enhancement process should be done on the local area to allow features to be detected in poor lighting conditions. This simulates the ability of the human eye to discern objects in complex lighting conditions. Third, it is proposed that the spatial resolution of the local search area is increased to extract better features and provide more accurate feature matching. A quantitative evaluation is performed to show tracking improvement using the proposed method. The three databases, each grayscale sequences that were obtained from aircrafts, used for these evaluations include the Columbus Large Image Format database, the Large Area Image Recorder database, and the Sussex database.

Keywords: Nonlinear enhancement, super resolution, object detection, object tracking, dynamic range compression

1. INTRODUCTION

Object tracking in wide area motion imagery (WAMI) is a complex problem that includes object detection and target tracking over time. The object detection portion is well suited for human analyst. An analyst has the ability to focus on the object of interest and ignore unnecessary background data. The analyst can also recognize a severely distorted object from the context of the scene. In addition, the human eye has the ability to deal with high dynamic range lighting situations. The human eye can work effectively with variation in lighting [1-2]. Human have a duplex retina which allows for good visual discrimination in all lighting conditions. The eye also uses the pupil to adapt to local lighting conditions. These qualities of the human eye allow an analyst to detect an object that appears in complex lighting conditions. A human analyst can also use the aid of additional lenses or digital zoom to magnify the object of interest. The use of a magnification tool allows the analyst to see additional details of the object for more accurate object detection. However, a human analyst can likely only keep track of a single object in a scene at a reasonable frame rate.

A computer vision solution for object tracking has the potential to be a much faster and efficient solution. A computer vision solution can allow for multiple objects to be tracked in parallel. In addition, the computer does not suffer from the same fatigue that a human analyst will suffer from. However, a computer vision solution does face challenges that do not affect the human analyst. The imagery captured for computer analysis has a limited dynamic lighting range due to the globally set exposure of the camera. This results in overexposed and underexposed regions where there is a loss or suppression of data. In addition, wide area motion imagery is typically taken over huge areas where the objects of interest, like cars and people, have a very low spatial resolution. In WAMI data the object's spatial resolution chosen to be as minimal as possible for recognition in order for the total area of the data captured to be as large as possible. These very low resolution objects are very challenging to track for even the most advanced tracking algorithms.

The proposed solution for improving computer vision object tracking results is to mimic the abilities of a human analyst. The proposed solution has three aspects that are each based on an aspect of human object detection. First, the focus of a human analyst is emulated by only doing work within a local object search area. By focusing the processing efforts on

Mobile Multimedia/Image Processing, Security, and Applications 2015, edited by Sos S. Agaian, Sabah A. Jassim, Eliza Yingzi Du, Proc. of SPIE Vol. 9497, 949705 · © 2015 SPIE
CCC code: 0277-786X/15/$18 · doi: 10.1117/12.2177228

only the local search area, the computation requirements are decreased. Second, it is proposed that an intensity enhancement process should be done on the local area in order to improve the visual quality of image. An intensity enhancement process can improve the detection by improving the contrast in overexposed and underexposed regions. Third, it is proposed that the spatial resolution of the local search area is increased. By using a super resolved search area, better features can be extracted and matched to the object of interest. The concept of the proposed tracking method improvements are demonstrated in Figure 1.

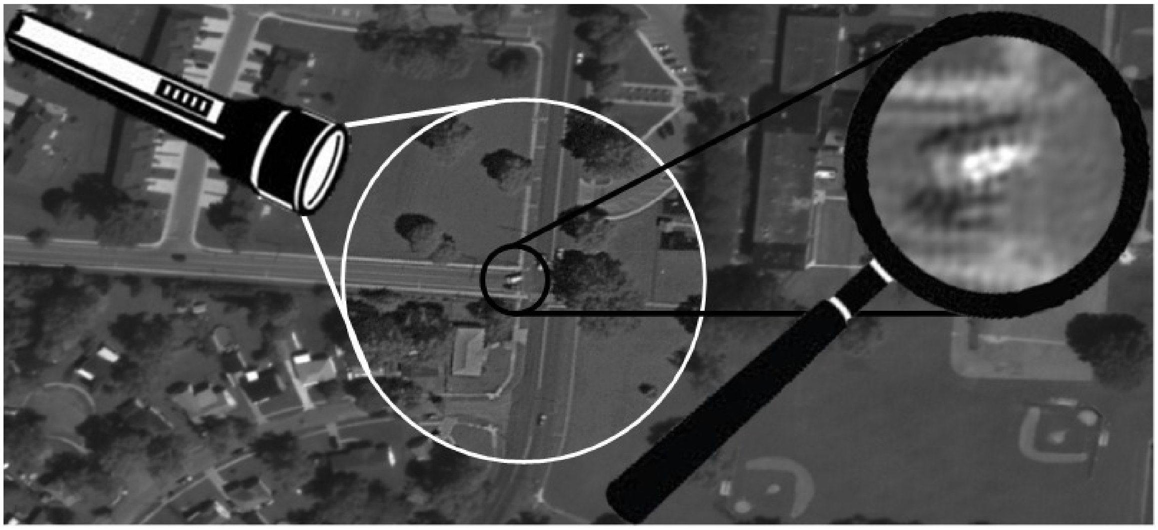

Figure 1. Concept of the proposed method. The flashlight represents the intensity enhancement method to improve lighting. The magnifying glass represents the super resolution method to increase spatial resolution.

In this paper, Sec. 2 presents background on the standard tracking methodology that the proposed method is built upon. Sec. 3 presents the proposed tacking methodology with the inclusion of local enhancement and super resolution. In Sec. 4, experimental results of the quantitative evaluation are shown and discussed. The paper is concluded in Sec. 5.

2. BACKGROUND

In order to implement the proposed improvements to a computer vision tracking method, an existing tracking process is selected. This tracking process uses Gaussian ringlet intensity distribution (GRID) feature extraction [3], Earth Mover's Distance for classification [4], and the Kalman filter for search area estimation. In this section, each component of this tracking method is explained. The object tracking process is shown in Figure 2.

2.1 Object selection

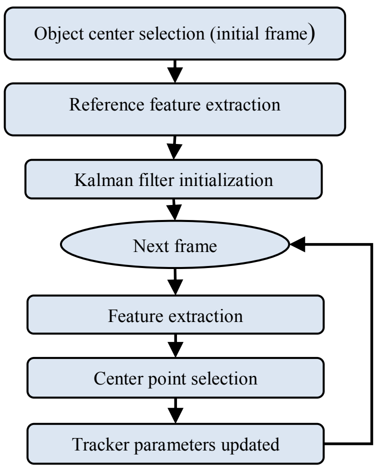

Figure 2. Kalman tracker with GRID detection tracking method.

In the first frame, the center point and the radius of the object of interest is selected by the user to be the reference object. For this method, the reference object is required to be a square window.

2.2 Reference feature extraction

Subsequently, the reference feature histogram is created using the GRID feature extraction method on the reference object. The GRID feature descriptor is described in Sec. 2.4.

2.3 Kalman filter initialization

Next, the Kalman filter is initialized using the estimated properties of the object of interest and the reference object location. The method uses a Kalman filter to estimate where the center of the object will be in the next frame. This estimation is determined using the previous information of the object to predict its trajectory. The Kalman filter tracking method can be written as

$$\boldsymbol{x}_k = \boldsymbol{A}\boldsymbol{x}_{k-1} + \boldsymbol{B}\boldsymbol{u}_k + \boldsymbol{w}_k \tag{1}$$

$$\text{and } \boldsymbol{z}_k = \boldsymbol{H}\boldsymbol{x}_k + \boldsymbol{v}_k \tag{2}$$

where k is the frame number, $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$are assumed to be zero mean Gaussian noise, $\boldsymbol{x}_k$ is the state model that includes the position and velocity information, $\boldsymbol{u}_k$is the control vector, $\boldsymbol{A}$ is the state transition matrix, $\boldsymbol{H}$ is the measurement matrix, $\boldsymbol{B}$ is the control-input matrix and $\boldsymbol{z}_k$is the measurement vector for the position.

2.4 Feature extraction

In each subsequent frame, Kalman filter is used to determine the center of the search area. The GRID feature extraction method is then used to find the features in the search area. Within the search area a sliding window approach is used to obtain the features of each testing area centered at points in the search area. The window size used is the same size as the reference object. The window starting point is selected using the equation

$$I_w(x, y) = I(x + ir - r, y + jr - r) \tag{3}$$

where i and j are the iteration of the window and r is the distance in pixels the window is moved in each iteration where the search area dimensions are equal to $[ir, jr]$. When r is set to 1, every possible window within the search area has its features extracted. Setting r to a larger value will reduce computation times but may result in less accurate results.

The GRID detection method is accomplished by partitioning the window into rings (circular areas) [3]. The use of rings gives results that are rotation invariant. In addition, more weight can be given to the inner rings as more important features occur near the object center. A mask to partition a window into three ring partitions with equal weighting can be seen in Figure 3. Each partition is then used to calculate a Gaussian ring histogram that is used as the feature descriptor.

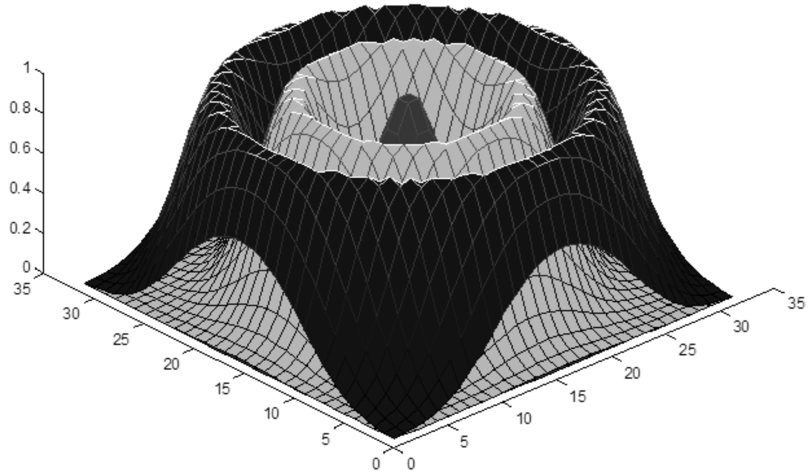

Figure 3. Circular Gaussian ring histogram kernel.

The Gaussian ring partitions are constructed to have equally spaced rings based on the object window size. The center Gaussian ring is created using the Gaussian function

$$G_1(x,y) = ce^{-\frac{(x-x_O)^2+(y-y_O)^2}{\frac{1}{2}(R_i-R_{i-1})^2}} \tag{4}$$

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 remaining Gaussian rings are computed as

$$G_i(x,y) = ce^{-\frac{(\sqrt{(x-x_O)^2+(y-y_O)^2}-\frac{1}{2}(R_i-R_{i-1}))^2}{\frac{1}{2}(R_i-R_{i-1})^2}} \tag{5}$$

A histogram is calculated for each ring with respect to the Gaussian ring as a mask. The histogram can be created with various bin sizes to improve misidentification and computation time. For this implementation, a bin size of 256 was used. These histograms are normalized using

$$N_i = \sum_{v=0}^{V} H_i(v) \tag{6}$$

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 then be applied to the histograms to emphasize the central rings. A total of three rings were used for this evaluation. The central ring was given a weight of 3, the middle ring a weight of 2, and the outside ring a weight of 1.

2.5 Center point selection

A classification strategy 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 [4]. 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 higher than the threshold, the obtained point will be considered the new center pixel of the object. If the confidence is lower than the threshold, the estimated point from the Kalman filter will be used as the new center pixel of the object. If an occlusion occurs, it is expected that confidence factor will be lower than the threshold. When the object of interest is occluded for multiple frames, the Kalman filter will keep track of the object trajectory and be used as the location estimator until the object exits the occlusion.

2.6 Tracking parameters update

The final step of the tracking process is to use the new center point information to update the Kalman Tracker and the reference object feature histogram. The reference object feature histogram is updated to keep track of global lighting changes and the gradual view point changes of the object. The update to the reference histogram is done using the equation

$$Hist_{ref} = (1-n) * Hist_{ref} + n * Hist \tag{7}$$

where n is the update factor that is set to 0.6 for the evaluation.

3. PROPOSED TRACKING METHOD

The proposed object tracking process is shown in Figure 4 with the proposed components in bold. An example of a frame that will be used for feature extraction in the proposed method is seen in Figure 5. This shows the relationship between the search, enhancement, and super resolution areas.

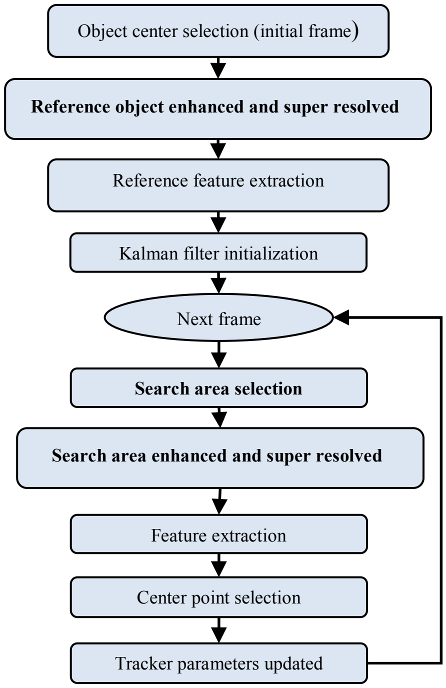

Figure 4. Proposed tracking method workflow.

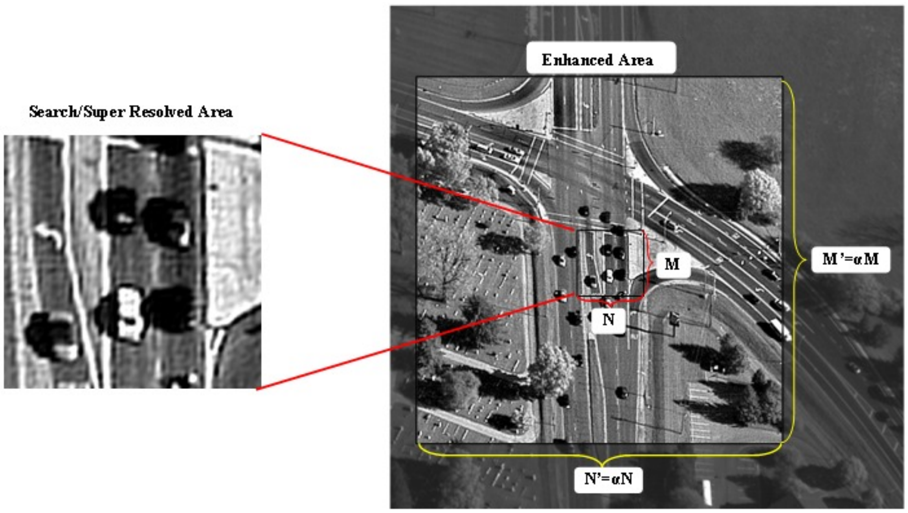

Figure 5. Proposed frame layout

3.1 Search area selection

For the proposed tracking method, the search area boundaries are selected to include only the local area around the target object. For each frame, this is accomplished by using the Kalman filter estimated center as the center for the search area. The search area boundaries are 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 and is found using the equation

$$D_{max} = \frac{s * gsd}{3600 fps}, \tag{8}$$

where D_{max} is the maximum distance the object can travel in pixels, s is the expected max speed of the object in miles per hour, gsd is the ground spatial distance in pixel per mile, and fps is the frame rate in frames per second. The search area is also used to weight the feature descriptors. The distance of the testing area from the center of the search area determines a weight that increases as the distance increases. The weight is used with the Earth Mover's Distance to determine the best match that is the closest to the center of the search area.

3.2 Intensity enhancement

Next, the intensity enhancement process is performed on the search area. The proposed solutions for enhancing the images include contrast limited adaptive histogram equalization (CLAHE), locally tuned sine nonlinearity (LTSN), and self-tunable transformation function (STTF). In Sec. 4.1 these methods are tested on various databases to compare their effectiveness in improving the tracking.

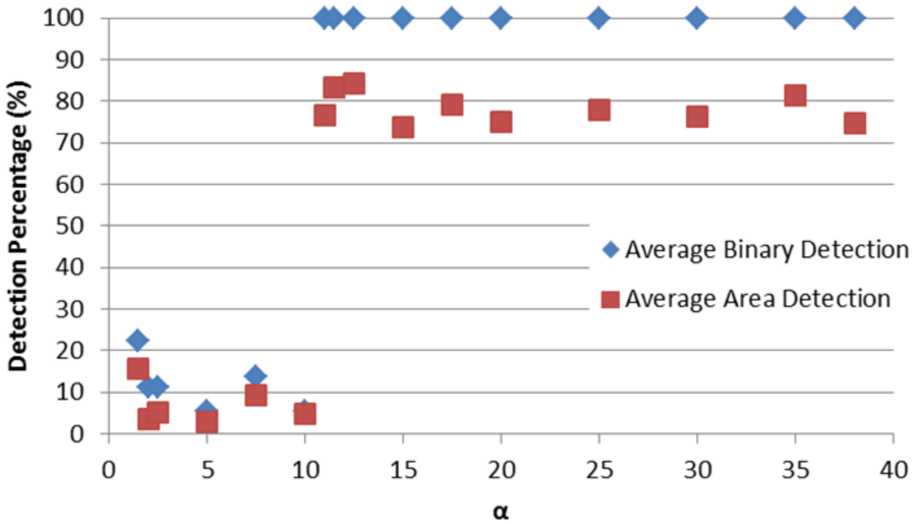

Figure 6. Detection percentage for different enhancement window sizes. The multiplier amount is multiplied to the size of the search area to determine the enhancement window size.

In order for the intensity enhancement process to successfully enhance the search area, it may be necessary to enhance an area larger than the search area. The purpose of enhancing an area that is larger than the search area itself is due to the global nature of the proposed enhancement methods. In order to have tracking improved, the object and the area around it need to have a consistent intensity across frames. The enhancement area has to be large enough that the intensity range of the object will be the same as the scene changes. If the enhanced area is too small, problems will occur as new elements are introduced into the scene which have a large intensity difference from the previous scene. This will cause the intensity of the object of interest to change too greatly as compensation for the new elements. Conversely, enhancing the entire area captured by the camera is unnecessary and time wasting. A balance must be found to obtain the smallest enhancement area that will have no adverse effects to the tracking. It is proposed that the size of the area to be enhanced is selected based on the object size and scale of database being used. Various enhancement size evaluations are evaluated to find the best size to use for the evaluation seen in Sec. 4. The evaluation was done for one object tracking sequence that has a window size of 10 pixels with a search area of 26 pixels. The total size of the frame is 1000x1000 pixels. This sequence was tested using the STTF enhancement process at various enhancement area sizes. The enhancement area size is determined using

$$N' = \alpha N, \tag{9}$$

$$M' = \alpha M, \tag{10}$$

where $[N\ M]$ are the dimensions of the search area, $[N'\ M']$ are the dimensions of the enhancement area, and α is a constant for the size increase. This is shown in Figure 5. The values for α that were used for testing varied between 1 and 38.5. This results in an enhancement area size that ranges from 26 pixels to 1001 pixels. As seen in Figure 6, the results show that the when the enhancement area is the between 1 times and 11 times the search area the enhancement process is very detrimental to the ability to track the object. When the enhancement area is between 11 pixels and the maximum 38.5 times the search area, the results show that the object can be tracked for each of the 36 frames. The results for the average area for these enhancement areas do fluctuate, but are much higher than if the enhancement area is too small. For the purpose of the evaluation in Sec. 4, the enhancement window size was chosen as 20 times the size of the search window to ensure it is acceptable for all of the tracking sequences.

The contrast limited adaptive histogram equalization enhancement method is an improved version of adaptive histogram equalization (AHE) [2, 5-8]. For AHE, each pixel is changed based on the histogram of its neighborhood. The new value is selected by redistributing the histogram so it is more flat and the peaks are more spread out. For CLAHE, each neighborhood is contrast limiting in order to limit the amplification available for the pixel.

The locally tuned sine nonlinearity is an enhancement method that relies on a nonlinear sine squared function to reassign each pixel [9]. The function is constructed so it can simultaneously improve dark and bright regions. The sine function used in LTSN is

$$I_{enh}(x, y) = \sin\left(\frac{\pi}{2} I_n(x, y)^q\right)^2, \tag{11}$$

where $I_n(x, y)$ is the intensity image normalized to [0 1] and q is the control parameter. The control parameter is based on the Gaussian mean of the neighborhood of the pixel. The LTSN does have trouble with over enhancement of under exposed regions.

The self-tunable transformation function is an improved version of the LTSN that uses an inverse sine function as it nonlinear function [10-11]. The function that is used for the STTF is

$$I_{enh}(x, y) = \frac{2}{\pi} \sin^{-1}\left(I_n(x, y)^{\frac{q}{2}}\right). \tag{12}$$

This function gives improved nonlinear curves that better enhance the overexposed and underexposed regions of an image. The STTF also has a Laplacian sharpening and a contrast enhancement steps that further improve the image quality.

3.3 Spatial resolution enhancement

In the proposed method, after the image goes through intensity enhancement, the search area is spatially enhanced. The goal of super resolution is to increase the spatial resolution of an image by a predetermined factor using information from the image or images in the same scene. Super resolution can be obtained using single frame or multiple frame techniques. The multi-frame technique is applied to the temporal domain of the images. This technique uses a set of frames that are related by sub-pixel displacements [12]. For static scenes, the sub-pixel displacements are global due to camera motion. For dynamic scenes, the sub-pixel displacements are local due to object movement. The low resolution frames are used to reconstruct a high resolution image by employing motion estimation. However, this type of technique is only useful for a small increase in resolution [13]. In addition, a high frame rate or low movement in the scene is needed to increase spatial resolution. For the desired application of object tracking in WAMI data, the frame rate is too low to used multi-frame techniques successfully. Instead, a single frame super resolution technique is used that achieves the resolution increase using only the information from one image. The super resolution process that is used in the proposed technique is as seen in [11, 14] using a machine learning approach with kernel regression. The super resolution method is a single frame super resolution technique that does not rely on temporal information. As a machine learning approach, training and association of trained parameters with the low resolution image is used to obtain the high resolution image. The super resolution process consists of finding the multi-level covariance estimation and using it to perform Kernel regression. The image is then sharpened using the natural image prior.

In the proposed technique, the single frame super resolution method used to increase the spatial resolution of the whole search area. An increase in resolution results in an increase of discerning features for small low quality target objects. The resolution increase also results in additional testing features to be used for classification. For the evaluation in Sec. 4, the spatial resolution in each dimension was increased by a factor of 4.

4. EXPERIMENTAL RESULTS AND DISCUSSION

The use of local enhancement and super resolution on tracking applications is evaluated in this section. The testing is done first to find which enhancement method has the most benefit for the tracking process. The local enhancement and super resolution process is then evaluated using the best enhancement method. The process to evaluate the tracking is done by comparing the true center of the object in the sequence to the object center found by the tracker. The databases that are used for these evaluations include the Columbus Large Image Format (CLIF) database [15], the Large Area Image Recorder (LAIR) database, and the Sussex database. All three databases are grayscale sequences that were taken from an aircraft.

4.1 Evaluation results for different enhancement methods

The evaluation for finding the most benefit of an enhancement method was performed by testing the standard tracking process seen in Sec. 2 with the inclusion of various enhancement processes. The tracking is performed by first enhancing all of the whole frames. Next, object is selected and tracked as seen in Sec. 2. The enhancement methods that are tested with are CLAHE, LTSN, and STTF. The tracking without enhancement is also evaluated.

The evaluation to find the most useful enhancement process is done on 20 different objects in the three databases. The objects are of various sizes between 4x4 and 14x14 pixels. The objects also have various sequence sizes between 19 and 36 frames. Each object is selected from the first frame and the tracking method is used to detect the object in the subsequent frames. The true centers were found for each object sequence and are used to determine the accuracy of the detection. The performance measurements of Frame Detection Accuracy (FDA) and Binary Detection Accuracy (BDA) will be used for this evaluation [16]. The FDA is the average object overlap of the truth and testing windows for all frames. The BDA is the average binary overlap for all frames.

Table 1. BDA evaluation results for object tracking using differing enhancement techniques. The objects tracked include cars, trucks, bikes, and pedestrians. The databases used were CLIF, LAIR, and Sussex (SUS).

Object ID	Database	Object Window Size	Frames	Detection Rate (%)			
				Enhancement Method			
				None	CLAHE	LTSN	STTF
1	CLIF	10	36	**100.0**	**100.0**	**100.0**	**100.0**
2	CLIF	10	29	**93.1**	**93.1**	**93.1**	**93.1**
3	CLIF	10	25	**100.0**	**100.0**	**100.0**	**100.0**
4	CLIF	14	20	**100.0**	**100.0**	**100.0**	**100.0**
5	CLIF	14	26	**100.0**	**100.0**	**100.0**	96.2
6	CLIF	10	36	**100.0**	**100.0**	**100.0**	**100.0**
7	CLIF	10	36	**100.0**	**100.0**	**100.0**	**100.0**
8	CLIF	10	33	60.6	87.9	66.7	**93.9**
9	CLIF	10	36	**100.0**	**100.0**	**100.0**	**100.0**
10	CLIF	6	36	**100.0**	97.2	**100.0**	58.3
11	CLIF	4	36	52.8	88.9	25.0	**100.0**
12	CLIF	8	20	**30.0**	**30.0**	25.0	**30.0**
13	LAIR	12	20	60.0	**100.0**	**100.0**	**100.0**
14	LAIR	14	20	15.0	**100.0**	**100.0**	**100.0**
15	LAIR	12	20	**100.0**	**100.0**	**100.0**	**100.0**
16	LAIR	12	20	15.0	**100.0**	55.0	85.0
17	LAIR	12	20	15.0	30.0	**100.0**	**100.0**
18	LAIR	12	19	**100.0**	**100.0**	**100.0**	**100.0**
19	LAIR	12	20	90.0	**100.0**	90.0	95.0
20	SUS	10	30	**100.0**	46.7	50.0	93.3
			Average	76.58	88.69	85.24	**92.24**

The first results of the evaluation are seen in Table 1. These results show the BDA of each object sequence. The BDA best indicates how often the tracker loses the object of interest. It can be seen from these results that the average of the average binary detection is highest for the process using STTF enhancement at 92.2 %. It is also seen that the method with no enhancement has significantly worse performance compared to any of the three enhancement techniques.

The FDA results for the sequences are seen in Table 2. This percentage is calculated by computing the average overlap of the detected object window and the true object window. The average area detection reinforces the trends seen in Table 1. The average of the average area detection is highest for the process using STTF enhancement at 71.5 %. Again, it is observed that the method with no enhancement has the lowest detection rate.

The results seen in Table 3 are the average pixel distance for all object tracked with 100% binary detection as seen in Table 1. The average pixel distance is calculated by averaging the pixel distance between the detected object center and the true object center in each frame. These results show that, for these 8 objects, the STTF enhancement has the lowest average pixel distance and the method with no enhancement has the highest average pixel distance.

Table 2. FDA evaluation results for object tracking using differing enhancement techniques. The objects tracked include cars, trucks, bikes, and pedestrians. The databases used were CLIF, LAIR, and Sussex (SUS).

Object ID	Database	Object Window Size	Frames	Detection Rate (%)			
				Enhancement Method			
				None	CLAHE	LTSN	STTF
1	CLIF	10	36	80.2	**81.7**	71.8	77.5
2	CLIF	10	29	73.5	**77.7**	72.4	75.5
3	CLIF	10	25	83.5	82.2	79.9	**84.0**
4	CLIF	14	20	77.6	75.3	**81.7**	81.2
5	CLIF	14	26	83.3	**85.7**	81.1	80.1
6	CLIF	10	36	**82.1**	79.5	78.3	76.4
7	CLIF	10	36	67.3	72.8	78.2	**78.9**
8	CLIF	10	33	48.0	68.2	51.7	**80.1**
9	CLIF	10	36	74.6	80.5	**85.4**	70.8
10	CLIF	6	36	**82.5**	78.6	69.8	43.1
11	CLIF	4	36	29.5	46.0	13.7	**67.4**
12	CLIF	8	20	16.6	**27.4**	16.9	15.4
13	LAIR	12	20	32.6	57.1	72.9	**75.7**
14	LAIR	14	20	11.4	74.6	**89.5**	86.0
15	LAIR	12	20	70.9	61.9	66.4	**73.7**
16	LAIR	12	20	13.7	**74.8**	41.7	64.0
17	LAIR	12	20	13.9	24.0	**73.3**	70.8
18	LAIR	12	19	71.5	**82.1**	78.1	77.9
19	LAIR	12	20	62.4	**79.8**	65.6	76.4
20	SUS	10	30	**88.0**	41.8	43.2	75.2
			Average	58.16	67.59	65.58	**71.51**

Table 3. Average pixel distance results for objects track with 100% average binary detection. The objects tracked include cars, trucks, bikes, and pedestrians.

Object ID	Database	Object Window Size	Frames	Pixel Distance (pixels)			
				Enhancement Method			
				None	CLAHE	LTSN	STTF
1	CLIF	10	36	1.663	**1.543**	2.329	1.896
3	CLIF	10	25	1.421	1.497	1.693	**1.362**
4	CLIF	14	20	2.555	2.937	**2.117**	2.234
6	CLIF	10	36	**1.494**	1.772	1.825	1.991
7	CLIF	10	36	2.754	2.465	1.727	**1.693**
9	CLIF	10	36	2.135	1.574	**1.161**	2.391
15	LAIR	12	20	3.154	4.294	3.534	**2.615**
18	LAIR	12	19	2.965	**1.739**	2.218	2.253
			Average	2.268	2.228	2.076	**2.054**

From these results it was determined that the STTF enhancement method was the best enhancement method for this tracking method. In the next subsection the STTF enhancement was selected as the enhancement technique for the proposed object tracking method.

4.2 Proposed object tracking method results

The evaluation for determining the effectiveness of the proposed method seen in Sec. 3 is accomplished in this section. The evaluation is done on the same 20 objects used in Sec 4.1. Each object is selected from the first frame and the tracking method is used to detect the object in the subsequent frames. The same true centers are used to determine the accuracy of the detection.

Table 4. BDA and FDA evaluation results for the proposed object tracking method. The objects tracked include cars, trucks, bikes, and pedestrians. The databases used were CLIF, LAIR, and Sussex (SUS).

Object ID	Database	Object Window Size	Frames	BDA	FDA
1	CLIF	10	36	100.0	75.0
2	CLIF	10	29	100.0	78.2
3	CLIF	10	25	100.0	81.0
4	CLIF	14	20	100.0	73.6
5	CLIF	14	26	100.0	77.4
6	CLIF	10	36	100.0	72.1
7	CLIF	10	36	100.0	76.9
8	CLIF	10	33	100.0	81.3
9	CLIF	10	36	100.0	71.7
10	CLIF	6	36	100.0	81.5
11	CLIF	4	36	100.0	70.3
12	CLIF	8	20	100.0	44.9
13	LAIR	12	20	100.0	82.0
14	LAIR	14	20	100.0	73.0
15	LAIR	12	20	100.0	77.3
16	LAIR	12	20	100.0	76.0
17	LAIR	12	20	100.0	73.5
18	LAIR	12	19	100.0	73.1
19	LAIR	12	20	95.0	77.3
20	SUS	10	30	96.7	84.1
			Average	**99.59**	**75.01**

The results for the proposed method are seen in Table 4. This table shows the BDA and FDA results for the proposed method. Compared to the results for the original tracking method seen in Tables 1 and 2, the proposed method has a higher detection rate in both BDA and FDA.

4.3 Computational time

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. The processes were implemented in MATLAB and C++. The computation time of the proposed method is compared to the no enhancement and STTF enhancement tracking methods. The results are shown in Table 5.

Table 5. Computation times in MATLAB and C++ implementations.

	Time per frame (s)		
	Enhancement Method		
	None	STTF	Proposed method
MATLAB	1.20	1.99	10.30
C++	0.10	0.34	1.53

5. CONCLUSIONS

In this paper, a novel tracking process is proposed to improve the accuracy of an object tracking application. The proposed techniques are introduced based on advantages seen in the human tracking ability. The inclusion of local intensity enhancement and spatial resolution increase within a tracking technique was proposed to simulate the focus and adaptive qualities of human object detection. The proposed technique's use of super resolution allows for better features to be collected and used for detection. The use of enhancement allows for the objects to be detected in challenging lighting conditions. In addition, the locality of the technique allows for a better performance without being detrimental to the accuracy of the tracker. Through evaluations on multiple sequences and databases, the proposed tracking technique was shown to more accurately track an object in comparison to the basic tracking method.

REFERENCES

[1] Kolb, H., "How the retina works," American Scientist, 91(1), (2003).

[2] Sherrier, R. H. and Johnson G. A., "Regionally adaptive histogram equalization of the chest," IEEE Trans. Med. Imag., 6(1), 1-7 (1987).

[3] 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).

[4] 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).

[5] Pizer, S. M., et al., "Adaptive histogram equalization and its variations," Proc. SPIE 1092, 290-300 (1989).

[6] Kim, T. K., Paik, J. K. and Kang, B. S., "Contrast enhancement system using spatially adaptive histogram equalization with temporal filtering," IEEE Trans. Consumer Electron. 44(1), 82–86 (1998).

[7] Kim, J. Y., Kim, L. S. and Hwang, S. H., "An advanced contrast enhancement using partially overlapped sub-block histogram equalization," IEEE Trans. Circuits Syst. Video Technol. 11(4), 475–484 (2001).

[8] Duan, J. and Qiu, G. "Novel histogram processing for colour image enhancement," Proc. 3rd Int. Conf. Imag. and Graphics, 55–58 (2004).

[9] Arigela, S. and Asari, V. K., "A locally tuned nonlinear technique for color image enhancement," WSEAS Trans. Signal Proc., 4(8), 524-519 (2008).

[10] Arigela, S. and Asari, V. K., "Self-tunable transformation function for enhancement of high contrast color images," J. Electron. Imaging, 22(2), 023010 (2013).

[11] Krieger, E., Asari, V. K. and Arigela, S., "Color image enhancement of low-resolution images capture in extreme lighting," Proc. SPIE 9120, 91200Q (2014).

[12] Prendergast, R. and Nguyen, T., "A block-based super-resolution for video sequences," Proc. of Int. Conf. on Imag. Proc., 1240-1243 (2008).

[13] Baker, S. and Kanade, T., "Limits on super-resolution and how to break them," IEEE Trans. on Pattern Analysis and Machine Intelligence, 9, 1167-1183 (2002).

[14] Arigela, S., Asari, V. K., and Qumsiyeh, M. B., "Learning Multi-level Local Phase Relationship for Single Image Resolution Enhancement," Proc. 2013 IEEE Int. Conf. on SMC, 2671-2676 (2013).

[15] CLIF data from AFRL, ID: HAAA08E09D, https://www.sdms.afrl.af.mil/index.php?collection=clif2007 (4 December 2014).

[16] 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).