A new algorithm for small object tracking based on super-resolution technique

Yabunayya Habibi, Dwi Ratna Sulistyaningrum, and Budi Setiyono

Citation: AIP Conference Proceedings **1867**, 020024 (2017); doi: 10.1063/1.4994427
View online: https://doi.org/10.1063/1.4994427
View Table of Contents: http://aip.scitation.org/toc/apc/1867/1
Published by the American Institute of Physics

A New Algorithm for Small Object Tracking Based on Super-Resolution Technique

Yabunayya Habibi[1, a)] , Dwi Ratna Sulistyaningrum[1, b)] and Budi Setiyono[1, c)]

[1]*Mathematics Departement, Institute of Technology Sepuluh Nopember, Jalan Raya ITS, keputih, Sukolilo, Surabaya, Jawa timur 60111, Indonesia*

.

[a)]Corresponding author: habibiexebppt@gmail.com
[b)]dwratna@gmail.com
[c)]masbudisetiyono@gmail.com

Abstract. Object tracking in a video is a problem of estimating the trajectory of an object in the image plane as it moves around a scene. In general, object tracking is a quite complicated problem. Difficulties in object tracking occur due to some constraints or conditions such as object motion, changing appearance patterns, non-rigid object structures, occlusions, and camera motion. Level of problems would be higher if the object tracking has relatively small. If it happens, an object will be difficult to identify and tracking becomes less precision because small object has little information. In order to overcome these problem, the tracking will be integrated with super-resolution where a high-resolution image will be built from several low-resolution image. In this research, tracking of moving object using adaptive particle filter which adaptive motion model is applied to get better proposal distribution approach. The simulation shows that tracking integration with super-resolution significantly increase the accuracy of small object tracking.

INTRODUCTION

Video surveillance is a topic of research in computer vision that tries to detect, recognize and track objects from a sequence of images. Object tracking is an important task in many computer vision applications such as surveillance, vehicle navigation, and autonomous robot navigation. Object tracking in a video is a problem of estimating the trajectory of an object in the image plane as it moves around a scene. One of the advantages in the military field is the utilizing on tracking as video surveillance for national patrol such as guard in the border area or around the base then detection and tracking of the moving enemy on the battlefield or automatic control system to track the target or the target position with a higher precision in firing missiles. In other filed such as sports, tracking of moving objects performed to determine the movement and position players on the field so that a sports analyst can determine the pattern of the game from players and utilization of Hawk-eye is used to track the position of the ball so it will greatly assist the task of linesmen. There are many things to do in tracking process but it depends on the level whom user need for retrieving information [1].

In general, tracking is a problem that is quite complicated. This problem can arise due to the movement of objects quickly, occlusion or the object invisible, noise in image sequences, the structure of non-rigid objects such as object rotating and changing the scale, and the camera moves to follow the movement of objects and more [2]. One of the more difficult problems will occur if object tracking is performed at considerable distance, so a objects look relatively small such as tracking of paragliding or plane movement, balls in sports like soccer video. If this case happen, the tracking process become less precision because small object isn't enough information to be identified and regarded as noise [3]. One way to overcome these problems and improve information of the small object is integration a super-

International Conference on Mathematics: Pure, Applied and Computation
AIP Conf. Proc. 1867, 020024-1–020024-10; doi: 10.1063/1.4994427
Published by AIP Publishing. 978-0-7354-1547-8/$30.00

resolution technique in tracking process. Super-resolution is a process of high-resolution images that are built from several low-resolution image [4][5]. Super-resolution will improve the visual quality of small objects so that objects has more information to be tracked.

Some earlier related research about tracking and super-resolution is a study from Sun [5]. Sun used template matching for tracking and super resolution to enhance image quality by reconstructing Projection Onto Convex Sets (POCs). However, final in his research utilized tracking result as enhance image on super-resolution to improvement visual quality. Integration between super-resolution and tracking has been established by Mise and Breckon [6] with a super-resolution imaging approach based on combination of the Sum of Absolut Diff erences (SAD) and gradient descent. The research used for improve target appearance that assists the overall tracking on high dynamic scene. Then, earlier research about small object tracking has been proposed by Davieshy [7]. This paper will present our approach that integrates a super-resolution technique into tracking in order to improve visual quality to get more information of the small object. The tracking method using adaptive particle filter proposed by Huang [8] whilst also can handles problem in tracking of small object. Improvement visual quality with super-resolution from small object will help the tracking process become more precision.

TRACKING SYSTEM

Adaptive Particle Filter Particle Filter is a method with applying adaptive motion models to get a better distribution approach. To further refine the existing disorders, motion continuity and smoothness of the track combined with correlation template in the observation likelihood [9]. This section will be described literature review and the basic theory about adaptive particle filter that it proposed by [8].

Adaptive Particle Filter

An image can be represented in a two-dimensional matrix which each value in the matrix represents a value of illumination intensity. State vector is a variable state that describes the behavior of a system. If an object's image is defined as X with (x, y) as centroid of object, then object tracking can be modeled as

$$X_{t+1} = X_t + v_t + \mu_t \tag{1}$$

where v_t is motion vector for object obtained from motion estimation and a noise μ_t. The motion vectors can be represented by a translational or other models that can estimate the object motion from a sequential image of video.

In the observation model, the calculation of the weights particle measure based on intensity, motion and trajectory. It can use the likelihood function defined as

$$P(Z_t \mid X_t) = P(Z_t^{\text{int}} \mid X_t) P(Z_t^{mot} \mid X_t)^{O_{t-1}} P(Z_t^{trj} \mid X_t)^{t-O_{t-1}} \tag{2}$$

where $Z_t = \{Z_t^{\text{int}}, Z_t^{mot}, Z_t^{trj}\}$ is intensity measurement, motion measurement and trajectory measurement, respectively. Z_t is the observation vector that is aff ected by state vector X_t. From equation above, intensity measurement is assumed independent from either but not for motion measurement and trajectory measurement because both of them is foreign mutual (the equation is not processed in the same state). If an object isn't detected by motion estimation then $O = 0$ and 1 otherwise.

The intensity measurement is calculated with similarity between image blocks (template) and candidate particle. A uniformly distributed of Clutter with J candidates from correlation surface and then intensity likelihood is formulated as

$$P(Z_t \mid X_t) = q_0 U(.) + C_N \sum_{j=1}^{J} q_j N(r_t, \sigma_t) \tag{3}$$

where q_j is prior probability and a normalization factor C_N.

The motion object likelihood is calculated based on d_{mot} or the difference between the particle position change $(\Delta x, \Delta y)$ and average object position change in previous time as $\overline{\Delta x} = \sum_{s=t-k}^{t-1} \frac{|x_s - x_{s-1}|}{k}$, $\overline{\Delta y} = \sum_{s=t-k}^{t-1} \frac{|y_s - y_{s-1}|}{k}$ ([8] use $k = 10$). Then, the motion object likelihood calculation obtained as follows

$$P(Z_t^{mot} \mid X_t) = \frac{1}{\sqrt{2\pi}\sigma_{mot}} e^{-\frac{d_{mot}^2}{2\sigma_{mot}^2}} \tag{4}$$

where $(\Delta x, \Delta y)$ is particle position change, $(\overline{\Delta x}, \overline{\Delta y})$ is average object position change in previous time and is variance for motion object likelihood. While, the trajectory likelihood is calculated from particle closeness to a trajectory that is obtained from previous position object. Trajectory Likelihood can be written as follows

$$P(Z_t^{trj} \mid X_t) = \frac{1}{\sqrt{2\pi}\sigma_{trj}} e^{-\frac{d_{trj}^2 / F}{2\sigma_{trj}^2}} \tag{5}$$

where the closeness metric $d_{trj} = | y - \sum_{i=0}^{m} a_i x^i |$ with polynomial order m as trajectory function. [8] defined a forgotten factor $F = \lambda_f^{t_0}, (0 < \lambda_f) < 1$. In certain condition.

If an object in previous frame $(cur - 1)$ isn't detected by motion estimation $O_{cur-1} = 0$, a projectiion one prediction results on the estimated trajectory will be done if object is not detected by motion estimation. Given two positions when the object is detected in previous frame, namely X_j on frame j th and X_i on frame i th with $(i > j)$, the estimated trajectory calculated as

$$X_{cur} = (1 - \lambda_f^{t_0})\hat{X}_{cur} + \tilde{X}_{cur}\lambda_f^{t_0} \tag{6}$$

with a prediction

$$\hat{X}_{cur} = X_i + (X_i - X_j)\frac{cur - i}{i - j}, i > j \tag{7}$$

where a projection of $\hat{X}_{cur}$ namely $\tilde{X}_{cur}$ is defined as the point on the closest and t_0 is the number of previous frame that isn't detected by motion estimation.

SUPER-RESOLUTION

Super-resolution is a process of high-resolution images that are built from several low-resolution image [5][4]. Low resolution images can be used form a single image or images series from the same scene. Because of the same scene, a process will obtain a information that can be used to reconstruct the high-resolution image [6] [10]. This paper focuses on a multi-frame super-resolution reconstruction technique and then a simplified ilustrated from super-resolution process from a images series such low quality video frame .

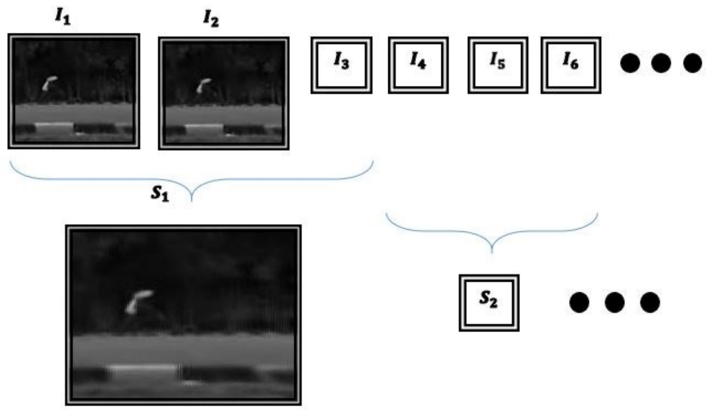


FIGURE 1. The process of Super-resolution image from a sequential low resolution image

In general, super-resolution consists of two process, namely the registration and reconstruction image. Registration process including an important process in super-resolution moreover images are registration with sub-pixel accuracy because it very important aspect in there construction process [4]. For implementation, Phased Based Image Matching (PBIM) method will be used for the registration process super-resolution.

Phased Based Image Matching

This method is actively developed in recent times because PBIM basically use a discrete Fourier transform and is often used in image registration process because of its reability and the computong time required is quite simple [4]. In PBIM, The calculation sample cross correlation between the reference and target image using the Fast Fourier Transform (FFT) and find the location of the peak which is a great translation between the two images. Suppose two images $f(n_1, n_2)$ and $g(n_1, n_2)$ with $N_1 \times N_2$ dimension and Cross phase spectrum (normalized cross spectrum) defined as follows

$$\hat{R}(k_1, k_2) = \frac{F(k_1, k_2)\overline{G(k_1, k_2)}}{|F(k_1, k_2)G(k_1, k_2)|} = e^{J\theta(k_1, k_2)} \tag{8}$$

And its invers

$$\hat{r}(n_1, n_2) = \frac{1}{N_1 N_2} \sum_{n_1=0}^{N_1} \sum_{n_2}^{N_2} \hat{R}(k_1, k_2) e^{j2\pi\left(\frac{k_1 n_1}{N_1} + \frac{k_2 n_2}{N_2}\right)} \tag{9}$$

where $F(k_1, k_2)$ and $G(k_1, k_2)$ a discrete Fourier transform of the spatial domain image. The results of this method is pixels translation that will be used in the reconstruction process with POCs. Super-resolution image reconstruction on stated redevelopment or rearrangements with projections to a high-resolution grid after value from the movement of the registration process have been obtained [4]. The reconstruction process in the super-resolution image based on image registration is illustrated as follows.

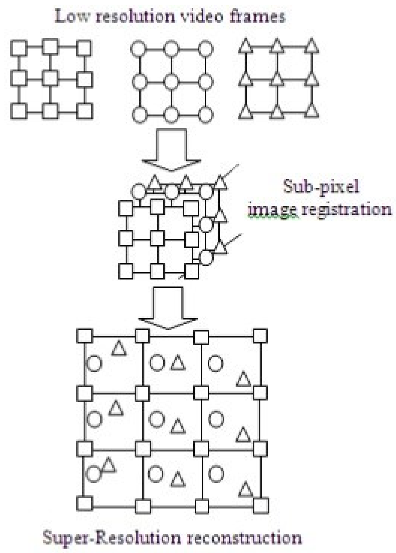


FIGURE 2. Super-resolution technique using sub-pixel image registration

From the illustration above, the super-resolution will be reconstructed or a redevelopment from a low-resolution image sequence after through the image registration process.

Projection Onto Convex Sets

Projection Onto Convex Sets (POCS) algorithm in image reconstruction is an algorithm that quite simple but it also can provide information more detail in the images reconstructed super-resolution from several low-resolution image with motion blur and noise [4]. The first, POCS as images reconstructed super-resolution was suggested by Stark and Oskoui [11]. The main concept is to estimate the super-resolution image is constrained in a closed convex set and get results by iterating [12]. Let a Low-resolution image $g(x, y)$ which a high resolution image $f(x, y)$ who experienced a shift (s_x, s_y), a process of degradation or blurring by a point spread function $h(x, y)$ and the addition of noise $N(x, y)$ can be modeled as

$$g(x, y) = h(x, y) f(x + x_s, y + s_y) + N(x, y) \tag{10}$$

So from the equation obtained convex set

$$C_i = \{f :| g(x, y) - h(x, y) f(x, y) | \leq N(x, y)\} \tag{11}$$

Suppose given projection operator who project an image to the set of convex closed Pc, f_k the magnification of the image by interpolating algorithm after k iteration, and f_0 first iteration from high resolution so a projection recontruction equation to convex set as follow

$$f_{k+1} = Tc_m Tc_{m-1} .....Tc_2 Tc_1 f_k \tag{12}$$

where $Tc_i = I + \lambda_i (Pc_i - I)$ with $0 < \lambda_i < 2$ [4][12]. The equation above can be solved with obtain completion iteratively on orthogonal projection to convex set by the constraints from noise level of low-resolution image. If projection operator is substituted into equation above to obtain following equation

$$f_{k+1} = f_k + \lambda_i \frac{g_i - h_i^{'} f_k}{\|h_i\|^2} h_i^2 \tag{13}$$

where g_i is i th element of vector $g(x, y)$ and $h_i^{'}$ is a row of i from matrix $h(x, y)$. The iteration process will continue until the termination criteria is obtained and the iteration will stop[4].

TRACKING INTEGRATION WITH SUPER-RESOLUTION

This section will discuss our proposed tracking integration with super-resolution technique. Super-resolution imaging applied into frame of video in order to improve information of the tracked target (small object). Figure 3 shows block diagram of the tracking integration with super-resolution

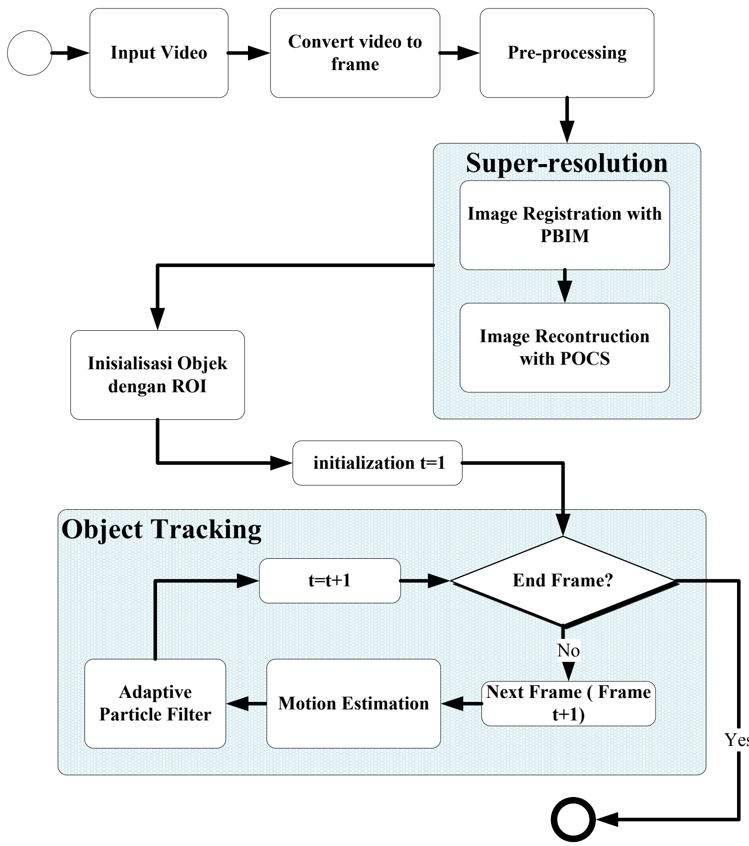


FIGURE 3. Block diagram of tracking based on super-resolution

Firstly, the process begins with input video and converting video to the frame to get all of the video frame. From of the frames, the result frames will be processed further called pre-processing. Pre-processing process is done to get the histogram that is evenly distributed with histogram equalization to all video frames and the result will be used as new frame. Then, the result of pre-processing will be used as input to the super-resolution process. Super-resolution process using super-resolution technique of multi-frame that consists of two stages, registration and recontruction. In a frame sequence, the object tracking process is processed on each frame in accordance with the specified object. ROI (Region of Interest) is a part of the selected image as an area for separating between the foreground and background (segmentation) as a reference for object tracking. Initialize of target is accepted as reference image initialized by user with ROI (Region Of Interest). In the next stage, super-resolution results or new frame are used for tracking of moving object with Adaptive Particle Filter method. A motion estimation is used to determine the motion vector based on the model in equation (1). Object tracking process start from frame 2^{th} until the number of video frames and according to the selected object with ROI then mark it with a bounding box.

RESULT AND DISCUSSION

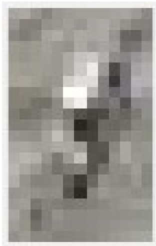
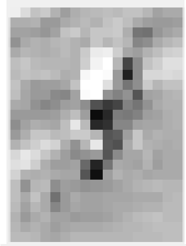
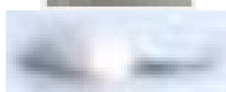
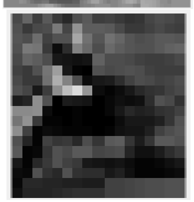
We are using MATLAB software to implements this concept with four video dataset in ".avi" and min 25 fps for simulation in order to determine its performance over a series simulation. Every video have various characteristic especially for object on every video that is presented in Table 1. In Tabel 1, Helicopter video has 60 frames with a large resolution (1920×1080) pixels for avoid a mistake of object detection as noise if the video is displayed in a lower resolution. Bicycle video and Paragliding video respectively has 50 frames and 60 frames with (640×360) pixels which contain a very low resolution object. Then, Motorcycle video has 63 frame with (640×360) pixes which has a object with a tendency of the similarity color to background and small part with intensity of diff erent color inside it.

"**TABLE 1,**" Video Dataset

Video	**Screenshoot Video**	**Number of Frame**
Motorcycle		63
Bicycle		50
Helicopter		59
Paragliding		60

From every video above, each frame of video will be processed become high resolution image, super-resolution, with object size according Lefevre and Vincent [3] typically between 10 and 100 pixels. The super-resolution result is shown in Table 2 that illustrates selected object by ROI.

TABLE 2. Super-resolution Image results

Video	Object Size (pixel)	Object Image	Grayscale Low Resolution Image with Histogram Equalization	Super-resolution Image
Motorcycle	(44×47)			
Bicycle	(23×15)			
Helicopter	(12×37)			
Paragliding	(17×16)			

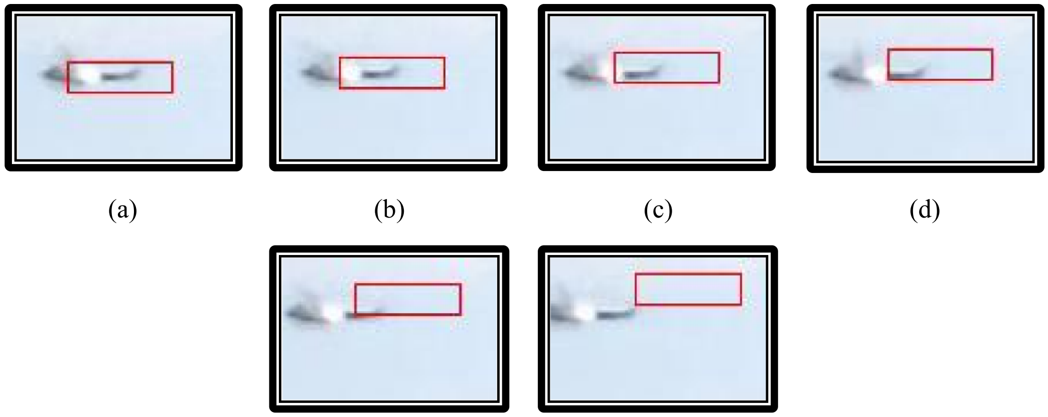
This paper gives three level of accuracy in object tracking where the assessment is done visually with the details shown in Figure 4.

(a) (b) (c)

FIGURE 4. Three level of accuracy. (a) Precision, (b) Less-precision, (c) Not precision

From the picture above, the object tracking is categorized into precision if most of the objects are in the area bounding box. Object tracking is categorized into less precision tracking if there are only a small part of the tracked object is in the bounding box area and not precision if no part of the object is in the bounding box area.

The same tracking method without super-resolution process will be used to track video above to make a comparison and find out significant results from our proposed.



(a) (b) (c) (d)

(e) (f)

FIGURE 5. Illustration tracking results

In Figure 5, section (a), (b) and (c) show the tracking result with category precision based on the level of accuracy as in Figure 5. Object tracking with less precision is shown in Figure 5 section (d), (e) and an imprecision tracking shown in Figure 5 section (f). For further calculation of the overall video, object tracking results without super-resolution process are presented in Table below. The comparison among both of them based on the number of frame from tracked object corresponding to the level of accuracy above then the result is shown in Table 2

TABLE 2. The Comparison of the accuracy result (frames)

Number	Video	Tracking without super-resolution			Tracking based on super-resolution		
		Not precision	Less-precision	Precision	Not precision	Less-precision	Precision
1	Motorcycle.avi	38	4	21	3	6	54
2	Bicycle.avi	17	13	20	0	0	50
3	Helikopter.avi	18	10	31	22	1	36
4	Paragliding.avi	0	12	48	0	0	60

Motorcycle video is failed to track moving motorcycle further because the acquired video is done in the area of the shadow of a building so object has a tendency of the same intensity with background. It is caused by motion estimation failure to detect object although a small part of the object in motorcycle video has difference in color intensity with background. A Tracking without super-resolution process on motorcycle video shows that object can only be tracked until frame 21th and the next frame tracking becomes less-precision or not precision. Bicycle video is also failed for further tracking process because object is smaller and has similarities with the background (in this case the background is the street where people cycling). Moreover, Helikopter and Paragliding video with tracking without super-resolution process show a tracking with accuracy quite good but something to know further about effect of the super-resolution process in tracking.

Table 2 also describes the tracking results with our proposed method such as a tracking on the helicopter video where the object can be tracked correctly up to 36 frames. However, object from frame 38th until the last has missed in tracking process because of the vibration effect during video acquisition so that object becomes blurred for a moment. It is likewise on the video motorcycle, a motorcycle movement can be traced correctly until frame 54th. Paragliding video and Bicycle video successfully tracked until last frame for track a man. The tracking result in Table 2 show an increase of tracking accuracy with our proposed method that utilize super-resolution result into tracking process. An increase of tracking accuracy significantly occurred on the motorcycle video. The difference of visualization in tracking result from these video is shown in Figure 6.

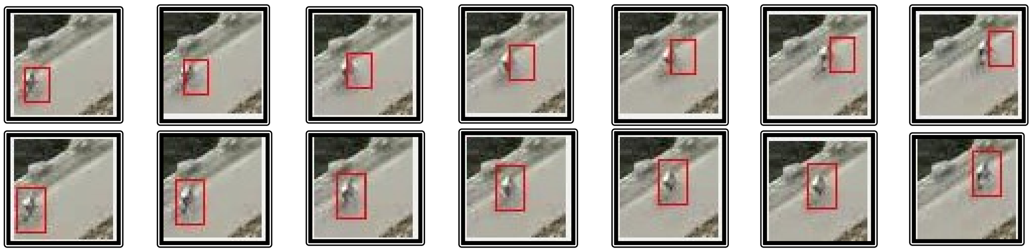


FIGURE 6. A scene tracking results from Bicycle video. Top image is tracking results without super-resolution. The bottom image is tracking result based on super-resolution

An increase in accuracy of tracking occurs on the Motorcycle video and Paragliding video. In addition, a small increase of accuracy on Helicopter video due to obstacles in the process of tracking such as some difficulty of tracking described in the Yilmaz et al [2].

CONCLUSIONS

This paper presents an approach that integrates a super-resolution technique into tracking process. The tracking method using adaptive particle filter proposed by Huang [8] which an adaptive motion model is applied to get good proposal distribution with varied diversity of particles. Super-resolution succesfully improves visual quality to get more information of the tracked target (small object). In order to make better result, multi-frame super-resolution is processing into tracking where phased based image matching (PBIM) is used in the registration process to get pixels translation and Projection Onto Convex Sets (POCs) as redevelopment of super-resolution image in recontruction process. The simulation results on the four video dataset with different characteristic and condition show that our approach significantly increase of tracking accuracy.

REFERENCES

1. K. R. Rupesh, A Survey on Object Detection and Tracking Algorithms, Tesis, National Institute of Technology Rourkela, Rourkela, 2013.
2. A. Yilmaz, O. Javed, and M. Shah ," Object tracking: A survey", ACM Computing Survey, Vol 38, No 4, Article 13, December, 2006.
3. S. Lefevre, T. Foissotte, and N. Vincent, " Detection and Tracking of Small Objects by Details Removal", Proceedings of 4th European Workshop on Image Analysis for Multimedia Interactive Services, University of London, Queen Mary, pp 137-140, April 2003.
4. B. Setiyono , M. Hariadi, H. M. Purnomo, Survei of Superresolution using Phased Based Image Matching, Jurnal Of Theoritical And Applied Information technology, Vol.43, No.2, September 2012.
5. S. Hao, L. Lin, Z. Weiping, and L. Limin, " Location and Super-resolution Enhancement of License Plates Based on Video Sequences", First International Conference on Information Science and Engineering, pp.1319-1322, December 2009.
6. O. Mise and T.P. Breckon , Super-Resolution Imaging Applied to Moving Targets in High Dynamic Scenes , Proceedings of SPIE - The International Society for Optical Engineering, Vol.889, pp.889917-889917-12, October 2013.
7. D. Davies, P. Palmer, and M. Mirmehdi, " Detection and tracking of very small low contrast object", Proceedings of the British Machine Conference , BMVA press, pp.60.1-60.10, 1998.
8. Y. Huang and J. Llach, Tracking the small object through Clutter with Adaptive Particle Filter, Audio, Language and Image Processing, pp.357 362, July 2008.
9. C. Stauff er and W. E. L. " Grimson, Adaptive background mixture models for real-time tracking", Computer Vision and Pattern Recognition, IEEE Computer Society Conference, Vol.2 , pp.252, 1999.
10. N. Panchal, B. Limbasiya, and A. Prajapati, Survey On Multi-Frame Image Super-Resolution, International Journal Of Scientific and Technology Research , Vol 2, pp.11, 2013.
11. H. Stark and P. Oskoui," High-resolution image recovery from image-plane arrays, using convex projections", JOSA A, Optical Society of America, Vol.6, No.11, pp.1715-1726, 1989.
12. C. Fan, J. Zhu, J. Gong, and C. Kuang, " POCS Super-resolution Sequence Image Reconstruction Based on Improvement Approach of Keren Registration Method",Sixth Internasional Conference on Intelligent Systems Design and Applications , Vol 2, pp.333-337, October 2006.