Super-resolution imaging applied to moving object tracking

Galandaru Swalaganata, Dwi Ratna Sulistyaningrum, and Budi Setiyono
Mathematics Department, Institute of Technology Sepuluh Nopember, Surabaya, Indonesia
E-mail: galandaru.swalaganata@gmail.com

Abstract. Moving object tracking in a video is a method used to detect and analyze changes that occur in an object that is being observed. Visual quality and the precision of the tracked target are highly wished in modern tracking system. The fact that the tracked object does not always seem clear causes the tracking result less precise. The reasons are low quality video, system noise, small object, and other factors. In order to improve the precision of the tracked object especially for small object, we propose a two step solution that integrates a super-resolution technique into tracking approach. First step is super-resolution imaging applied into frame sequences. Second step is tracking the result of super-resolution images. Super-resolution image is a technique to obtain high-resolution images from low-resolution images. In this research single frame super-resolution technique is proposed for tracking approach. Single frame super-resolution was a kind of super-resolution that it has the advantage of fast computation time. The method used for tracking is Camshift. The advantages of Camshift was simple calculation based on HSV color that use its histogram for some condition and color of the object varies. The computational complexity and large memory requirements required for the implementation of super-resolution and tracking were reduced and the precision of the tracked target was good. Experiment showed that integrate a super-resolution imaging into tracking technique can track the object precisely with various background, shape changes of the object, and in a good light conditions.

1. Introduction

Tracking of moving objects is a way used to detect and analyze changes that occur in an object from one frame to the next frame in the video [1]. The purpose of tracking a moving object is estimated trajectory of objects in the image plane is moved in a scene. Tracking moving objects can help to detect, identify, and analyze such objects. There are three basic steps in tracking a moving object. The first step is the detection or select Region of Interest (ROI) of a moving object in a video. The second step is tracking an object that has been selected from the first step of several parts of the frame. The third step is the analysis of movement of the object to identify the object’s behavior [2].

The ability to obtain visual quality and the precision of the tracked target is highly wished in modern tracking system. The fact that the tracked object does not always seem clear causes the tracking result less precise. The reasons are low quality video, system noise, small object, and other factors. By utilizing the super-resolution images in tracking process are expected to get results tracking moving objects in video more precision. Super-resolution is a technique of constructing the high resolution images from low-resolution images. There are two types of
super-resolution based on the input frame, namely single frame super-resolution and multi-frame super-resolution [3]. Super-resolution single frame has advantages in terms of computation time faster than multi-frame super-resolution [4]. The process of getting high-resolution image on a single frame super-resolution is raising the image of the target so that the process does not eliminate and still contains information prior to the super-resolution [5].

Earlier research on tracking of moving objects with the Kalman filter has been done under the title Moving Object Tracking Using Kalman Filter [6]. The study was concern on objects with varied backgrounds with a target single object. Another new study with title Moving Target Tracking Based on Camshift Approach and Kalman Filter [7] successfully combines Camshift and Kalman filter. From both these studies did not use a small object. Other studies with title Detection and Tracking of Very Small Objects Low Contrast [8] succeed to detect small objects using Wavelets for a single object and Kalman filter for multiple objects. Research by Olegs Mise and Toby P. Breckonb entitled Super-Resolution Imaging Applied to Moving Targets in High Dynamic Scenes [9] successfully applying super-resolution images to obtain the results of tracking moving objects. In [9] used multi-frame super-resolution and block matching method correlation for tracking.

Studies about a small object tracking is still slightly caused by very few of the features available. Therefore, by integrated super-resolution with tracking can be obtained more features that enable tracking a small object. There is a difficulty to detect relatively small objects that will be track. The problem would be too much noise appears and objects are relatively small because it would be difficult to distinguish between the object and the noise [8]. Another difficulty was the small object is represented by a set of pixels (between 10 and 100) on a high resolution image, so it is not enough information obtained or modeled correctly [10].

In this paper discuss the result of integrated super-resolution technique into tracking approach. First step is super-resolution imaging applied into frame sequences to improve visual quality of the tracked target. This step is done by cropping the frame in several frame or all of frame. Second step is track the result of super-resolution images. The overview of super-resolution method is described in Section 2. Our main concept is introduced in Section 3. Furthermore, in Section 4 we highlight the main concepts with results shown over a variety video.

2. Single Frame Super-Resolution
Super-resolution methods generally construct a high-resolution image from single or multiple low-resolution images. There are two types of super-resolution, namely single frame super-resolution and multi-frame super-resolution [3]. In general, super-resolution consists of two phases, namely the registration of the image and image reconstruction. In process of getting high-resolution image on a single frame technique is raising the size of the image \( n \times n \), where \( n \) is natural number, so that in the process of super-resolution does not eliminate and still contains information prior to the super-resolution. The method which is quite popular to perform single frame super-resolution is Nearest neighbor, Bilinear, and Bicubic [5].

In this study focused on single frame super-resolution. Bicubic Interpolation [5] and Directional Bicubic Interpolation [4] were proposed for single frame super-resolution. Fig. 1 shows \( 2 \times \) zoom process of \( 4 \times 4 \) LR (low resolution) image by Directional Bicubic Interpolation.

3. Tracking System Enhanced by Super-Resolution
In this section discuss integrated super-resolution framework into tracking system. First step is super-resolution imaging applied into frame sequences. This step was done by cropping the frame in several frame or all of frame in order to improve visual quality of the tracked target. Second step is tracking the result of super-resolution images. Bicubic Interpolation [5] and Directional Bicubic Interpolation [4] were proposed for super-resolution step. Every
interpolation has their own advantages, so it will be compared for consequent after tracked. For some reason, the object changes the shape (zoom, rotate) and various color. In order to handle them, Camshift method was proposed for tracking approach. Fig. 2 shows block diagram of the tracking system integrated with super-resolution framework used in this work.

There were two options to enhance tracking system with super-resolution, one of them was super-resolution first and then tracking the result as shown in Fig. 2. Here the chosen ROI (initialized by user) was as the first super-resolution image that being tracked. Single frame super-resolution was done by cropping the frame in several frames or all of frames. The purpose of cropping was create the search window area smaller. Therefore, it needs less computation time in searching the tracked target. An optional step was given before tracking, called histogram equalization. The purpose of equalize the histogram make a differences between forground (object) and background. It works by normalized the histogram distribution.
3.1. Bicubic and Directional Bicubic Interpolation

Bicubic interpolation interpolates indiscriminately the missing pixels in the same direction (horizontal or vertical) and typically results in blurring, blocking, ringing or other artifacts in interpolated images. Directional Bicubic first explicitly decides the local edge direction in terms of the ratio of the two orthogonal directional gradients for an interpolation pixel. If the pixel is on a strong edge, it will be interpolated using Bicubic along the edge direction; if the pixel is in a weak edge or texture region, it will be interpolated by combining the two orthogonal directional Bicubic interpolation values in direct proportion to the two orthogonal directional inverse gradients [4]. Computation time of the two interpolation is another difference. Bicubic interpolation indiscriminately the missing pixels in the same direction (horizontal or vertical) and typically results in blurring, blocking, ringing or other artifacts in interpolated images. Directional Bicubic first explicitly decides the local edge direction in terms of the ratio of the two orthogonal directional gradients for an interpolation pixel.

The advantages of Camshift based on simple calculation based on HSV color that use its histogram for some condition and color of the object varies. Camshift method improved the

---

**Step 1: Initialization.** This step expands the known \( n \times n \) LR image \( I_l \) onto the grid of the \((2n - 1) \times (2n - 1)\) HR image \( I_h \) as shown in Fig. 1(a). The mapping relation is \( I_h(2x - 1, 2y - 1) = I_l(x, y), x, y = 1, 2, ..., n \) where \((x, y)\) is the pixel location.

**Step 2: Interpolation in the first pass.** The whole gray pixels in Fig. 1(a) are interpolated. The filled pixels are \( I_h(2x, 2y), x, y = 1, 2, 3, ..., n \) (including border pixels). For every gray pixel, compute the \( 45^\circ(G_1) \) and \( 135^\circ(G_2) \) diagonal directional Bicubic interpolation values \( p_1 \) and \( p_2 \), the two diagonal directional gradients \( G_1 \) and \( G_2 \) using

\[
G_1 = \sum_{m=3,\pm1} \sum_{n=3,\pm1} |I(i + m, j - n) - I(i + m - 2, j - n + 2)| \quad (1)
\]

\[
G_2 = \sum_{m=3,\pm1} \sum_{n=3,\pm1} |I(i + m, j + n) - I(i + m - 2, j + n - 2)| \quad (2)
\]

Then the missing pixel \( p \) is estimated as

- if \((1 + G_1)/(1 + G_2) > T \)
  
  \[ p = p_2; \]
- else if \((1 + G_2)/(1 + G_1) > T \)
  
  \[ p = p_1; \]
- else
  
  \[ p = (w_1p_1 + w_2p_2)/(w_1 + w_2) \]

end

**Step 3: Interpolation of the case 1 of the second pass.** The whole gray pixels in Fig. 1(b) are interpolated. The filled pixels are \( I_h(2x - 1, 2y), x, y = 1, 2, 3, ..., n \) (including border pixels). For every gray pixel, compute the horizontal and vertical bicubic interpolation values \( p_1 \) and \( p_2 \), the horizontal and vertical gradients \( G_1 \) and \( G_2 \) using (1) and (2). Then the missing pixel \( p \) is estimated like Step 2.

**Step 4: Interpolation of the case 2 of the second pass.** The whole gray pixels in Fig. 1(c) are interpolated. The filled pixels are \( I_h(2x, 2y - 1), x, y = 1, 2, 3, ..., n \) (including border pixels). For every gray pixel, compute the horizontal and vertical bicubic interpolation values \( p_1 \) and \( p_2 \), the two horizontal and vertical gradients \( G_1 \) and \( G_2 \) using (1) and (2). Then the missing pixel \( p \) is estimated like Step 2.

3.2. Camshift (Continuously Adaptive Mean Shift)

The advantages of Camshift based on simple calculation based on HSV color that use its histogram for some condition and color of the object varies. Camshift method improved the
searching efficiency of mean shift tracking algorithm. The coordinate of the best matching block to be the initial searching position at next continuous image is recorded [1] [7]. The procedure of Camshift algorithm is illustrated as in Fig. 3.

The centroid of color probability distribution image in the search window can be searched by this equation [1].

The $0^{th}$ moment:

$$M_{00} = \sum_x \sum_y I(x, y) \quad (3)$$

The $1^{st}$ moments for $x$ and $y$:

$$M_{10} = \sum_x \sum_y xI(x, y) \quad (4)$$

$$M_{01} = \sum_x \sum_y yI(x, y) \quad (5)$$

The centroid of search window:

$$x_c = \frac{M_{10}}{M_{00}}; \quad y_c = \frac{M_{01}}{M_{00}} \quad (6)$$

where $I(x, y)$ is a pixel color value at position $(x, y)$ in the image and $(x, y)$ are in the search window. $M_{10}$ is the first moment for $x$, while $M_{01}$ is the first moment for $y$. 

**Figure 3.** Camshift Method [7]
Table 1. Dataset Video

| Name          | Screenshot | Frame(s) |
|---------------|------------|----------|
| 1. Paralayang.avi | ![Paralayang.png](image) | 168      |
| 2. bike.avi   | ![bike.png](image)       | 151      |
| 3. soccer.avi | ![soccer.png](image)     | 76       |

4. Result and Discussion
In this section used Matlab application for implementation and 3 video for experiment (see Table 1). All of the video contain single moving object, small object, and various size. Experiment on this paper was conducted with extension ”.avi” video which will be converted into frames sequences.

Every video in Table 1 has their own characteristics. Paralayang.avi video has size of 640×360 pixels and contains a small moving object and colour difference (between the object and background). Bike.avi video has size of 320×240 pixels and contains a small moving object and the moving camera. The last video, soccer.avi, has size of 856×480 pixels and contains more than one moving object.

4.1. Native Tracking without Super-resolution
Compares the racking results before and after the super-resolution is important to find out whether the proposed method will produce significant results. Thus, each video will tracked by Camshift method without super-resolution first.

Table 2 describes Paralayang.avi video when ROI has been initialized by user, the result of tracking a man was successfully tracked until frame 48 (28%). The object in frame 48 until end was not detected, the error was the color distribution did not read correctly. Bike.avi video succeed tracked a man in all of frames, but the rectangle for some frames was missed (55.6%). It happened because the moving camera and the object moving forward (zoom in)(see Fig.4). Bola.avi video successfully too in tracking a player in all of frames, but like previous video, the
Table 2. Tracking Result of Single Object

| Video          | Native Tracking | SR* Tracking |
|----------------|-----------------|--------------|
| 1. Paralayang.avi | 47 frames       | 168 frames   |
| 2. bike.avi     | Missed ±67 frames | Missed ±12 frames |
| 3. soccer.avi   | Missed ±23 frames | 76 frames    |

*Bicubic and Directional Bicubic Interpolation gave same result

Figure 4. Missed rectangle in result of bike.avi tracking (frame 10, frame 112, and frame 149)

Figure 5. Missed rectangle in result of soccer.avi tracking (frame 57, frame 69, and frame 76)

rectangle for some frames was missed (70%). It was caused by a small object and the object moving too fast (see Fig.5).

4.2. Tracking Integrated with Super-resolution

Table 2 shows the tracking result of each video. As we can see, the result does not precise because of many reason depend on each video. Our proposed method proposed a solution to track the target object with integrated single frame super-resolution into tracking system. Directional Bicubic Interpolation is better than Bicubic Interpolation in high resolution image result. However, computation time Bicubic was better than Directional Bicubic. Table 2 also shows the result of tracking the target after single frame super-resolution.

Paralayang.avi and soccer.avi video can tracked successfully a single object in all frames without histogram equalization. While bike.avi video should be use histogram equalization, because the moving camera and the object moving forward (zoom in). Thus, histogram equalization is for make a differences between forground (object) and background. The rectangle can place in almost frames of bike.avi video. It missed ±12 frames of 151 frames after super-resolution.

Both interpolation, $2 \times$ Bicubic Interpolation and $2 \times$ Directional Bicubic Interpolation gave the same result for tracking approach. Our proposed method was resulting great impact with 100% successfully track first and third video, and 92.1% for second video. It means that integrated single frame super-resolution into tracking system can track the object precisely with various background, shape changes of the object, and good light conditions.

5. Conclusion

In this paper we have presented an integration method between single frame super-resolution technique and tracking system. This approach significantly improves the overall tracking system. The results show significant improvement in visual target object and the precision of object
tracking. The proposed method also works well in the various video size. Bicubic Interpolation and Directional Bicubic Interpolation gave the same result for tracking approach. The average of tracked target successfully was 51.2% for native tracking and 97.4% for our proposed method.

**Acknowledgment**

This research was supported by Ministry of Research, Technology and Higher Education of the Republic of Indonesia through the funding from Block Grant research. We thank our colleagues for encouraging us to present this paper on the International Conference.

**References**

[1] Chen X, Wu H, Li X, Luo X, and Qiu T 2012 Real-time Visual Object Tracking via Camshift-Based Robust Framework *International Journal of Fuzzy Systems* 14 p 262

[2] Yilmaz A, Javed O, and Shah M 2006 Object tracking: a survey *Acm computing surveys (CSUR)* 38 p 13

[3] Setiyono B, Hariadi M, and Purnomo M H 2012 Survey of Superresolution Using Phased Based Image Matching *Journal of Theoretical and Applied Information Technology* 43

[4] Jing L, Zongliang G, and Xiuchang Z 2013 Directional Bicubic Interpolation-A New Method of Image Super-Resolution *Proceedings of International Conference on Multimedia Technology* (Atlantis Press) pp 470–477

[5] Raval H and Kothari A 2014 Analysis of Single Frame Super Resolution Methods *International Journal Of Engineering And Computer Science* 3 pp 9073–9076

[6] Patel A H and Thakore G D 2013 Moving Object Tracking Using Kalman Filter *International Journal of Computer Science and Mobile Computing* 2 pp 326–332

[7] Lien S, Hsia K, and Su J 2015 Moving Target Tracking Based on Camshift Approach and Kalman Filter *An International Journal Applied Mathematics and Information Sciences* 9 pp 395–401

[8] Daviesy D, Palmery P and Mirmehdiz M 1998 Detection and Tracking of Very Small Low Contrast Objects *Proceedings of the British Machine Vision Conference* (BMVA Press) pp 60.1–60.10

[9] Mise O and Breckonb T P 2013 Super-Resolution Imaging Applied to Moving Targets in High Dynamic Scenes *Emerging Technologies in Security and Defence; and Quantum Security II; and Unmanned Sensor Systems* X 8899

[10] Lefevre s, Foissotte T and Vincent N 2003 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