Investigation of a new method for improving image resolution for camera tracking applications

Hamza Alzarok[1], Simon Fletcher[1], Andrew Longstaff[1], Alan Myers[1]
[1]*Centre for Precision Technologies, School of Computing and Engineering, University of Huddersfield, UK*
Email: u0950134@hud.ac.uk

Abstract

Camera based systems have been a preferred choice in many motion tracking applications due to the ease of installation and the ability to work in unprepared environments. The concept of these systems is based on extracting image information (colour and shape properties) to detect the object location. However, the resolution of the image and the camera field-of- view (FOV) are two main factors that can restrict the tracking applications for which these systems can be used. Resolution can be addressed partially by using higher resolution cameras but this may not always be possible or cost effective.
This research paper investigates a new method utilising averaging of offset images to improve the effective resolution using a standard camera. The initial results show that the minimum detectable position change of a tracked object could be improved by up to 4 times.

1 Introduction

In the last decade, with the rapid development in the applications of image processing, obtaining high quality images has become increasingly important. However, image enhancement sometimes exceeds the abilities of available cameras due to various limitations. Extensive work has been proposed by researchers to make this demand applicable with image resolution being one such important area. The term “image resolution” is often misunderstood in describing the properties of visual images since it has a large number of definitions. The resolution was defined by some researchers in the field of optics in terms of the modulation transfer function (MTF). However, MTF was also used for characterizing the response of the vision system to an arbitrary input [1]. On the other hand, in the field of image processing and computer vision, the term resolution can be described in three different ways; spatial resolution, brightness resolution and temporal resolution. In this paper, the term resolution only refers to the spatial resolution and therefore simply defined as the smallest measurable detail in a visual presentation [2].

1.1 Challenges in imaging enhancement

The spatial resolution of the image is restricted by the imaging sensors or the imaging acquisition device. Modern image sensors such as a charge-coupled device (CCD) and a complementary metal-oxide-semiconductor (CMOS) active-pixel sensor are basically arranged in a two-dimensional array to obtain two-dimensional image signals. The size of the pixel in the first place defines the spatial resolution of the captured image. The higher spatial resolution of the imaging system can be obtained if a higher density of the sensors is used. When an imaging system with inadequate detectors is used to generate images, the aliasing from low spatial sampling frequency produces low resolution images with “blocky” effects.

In order to enhance the spatial resolution of an imaging system, one of the straightforward ways is to increase the density of the sensor by minimizing the size of the pixel [3]. However, as the pixel size of the sensor decreases, the amount of light incident on each sensor also reduces, and leads to an increase in the shot noise [4]. Moreover, the hardware cost of the sensor rises with the increase of the density of the sensor or correspondingly pixel density of the image. Therefore, the limitation of the hardware on the sensor size restricts the spatial image resolution that can be obtained [5].

While the spatial image resolution is limited by the image sensor, the details of the image (high frequency bands) are also restricted by the optics, because of lens blur (associated with point spread function of the image sensor (PSF)), effects of the lens aberration, diffractions of the aperture and optical blurring due to motion [3].

Designing and building imaging chips and optical components to generate very high-resolution images is often not a feasible or practical solution in most real applications, such as tracking cameras, due to the increased cost. Furthermore, increasing the size of the chip in order to involve a large number of pixels requires an increase in the capacitance, and that leads to a reduction in the frame and/or data transfer rate [6].

In target tracking applications, a wide FOV camera should be used if a large area needs to be under view. However, the downside will be low resolution which can be partially addressed by using higher resolution cameras but as stated this may not be a practical solution. In order to use multiple cameras, the relationship between camera views needs to be manually or automatically computed [7], hardware and high computational cost again is another concern. If the multiple cameras are not stationary, then the speed and path of the object has to be considered in order to obtain the correspondence between cameras.

Another possible way to overcome the problem of resolution is to accept the degradations of the image and use signal processing in order to post process the captured images, to avoid computational and hardware costs. These techniques are called Super Resolution (SR), and in some literature, the process is referred to as Resolution Enhancement (RE) [3].

In robotics applications, Solving the Simultaneous Localization And Mapping problem (SLAM) is one of the fundamental problems in robotics but has mainly been applied to Automated Guided Vehicles (AGVs), and it has recently received a lot of attention in research [8]. Due to the market price factor, high resolution cameras have not been introduced as a good choice for solving the localization problem for robots, low cost alternatives such as Kinect sensor cameras was introduced in some recent research [9], the sensors succeeded in navigating the robot motion, however they failed in providing an accurate information about the location in the presence of unprepared environments [10].

1.2 Super resolution techniques

The SR techniques can be defined as a process of obtaining a High Resolution (HR) image from low resolution observations (Figure 1). The principle of these techniques was based on the fact that the change in the relative motion between the camera and the scene leads to a change in the information of each single Low Resolution (LR) frame, so by combining and fusing all frames via a reconstruction process, an SR image of the true scene can be generated [11].

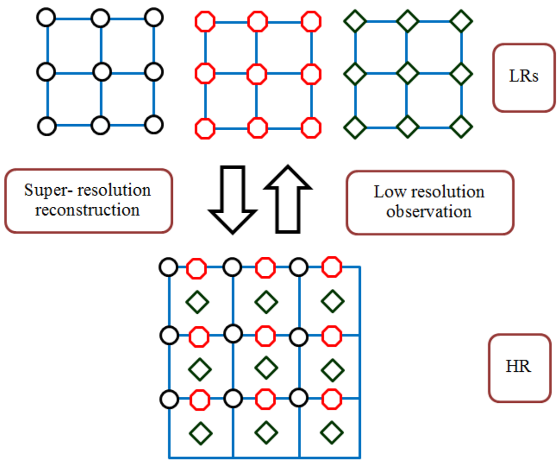

Figure 1: SR reconstruction from LR frames.

According to Katsaggelos et al [12], the topic of super resolution was addressed in the early 80s by Tsai and Huang [13] in one of the first papers in the signal processing community, the work was aimed to improve the resolution of Landsat images, and since then the SR process has been applied to a variety of fields, such as video printing, medical images, and improvement of the quality of images obtained by one CCD. The problem of SR was also described by Negroponte [14] at the media Lab when a salient still was obtained from video sequences. The SR problem arises when the high resolution image needs to be created from a video sequence, and this problem is more difficult when the video sequence has been compressed [15].

The technique of Projection onto Convex Sets (POCS) was first introduced by Stark[16]for solving the SR problem, and the suggested solution was based on a set of constraints, the modified techniques was later applied for multi camera surveillance image [17]. The advantage of the algorithm is in the simplicity of techniques and the flexibility in including a priori information. However, the proposed method is applicably limited because of a slow convergence rate. Also, the final solution basically depends on the initial guess. Adaptive filtering techniques have also been applied for super resolution reconstruction [18], suggested algorithms are based on least squares and pseudo Recursive Least Squares (RLS). Later, a Kalman filter was proposed [19, 20] as a promising technique, but it is still in a development state. The idea of these algorithms was built on the assumption that the information regarding the motion between the obtained images and the blur operators is known. However, the main drawback of these techniques is in the associated accuracy due to the assumptions and the high computational cost.

Irani and Peleg [21] proposed a Iterative Back-Projection technique for enhancing monochrome and colour low resolution images. The principle of the algorithm is based on determining the difference between the simulated and observed low resolution images. The difference value represents the back projection error which is used as initial guess in the next iteration. With increasing number of iterations, the error will be minimized and hence a high resolution output will be obtained. The advantage of this technique is the ability to solve the issue of noise and blur. However, the approach does not provide an explicit solution.

This paper investigates a new technique for improving the effective spatial resolution of images using a standard camera. The proposed technique is based on generating a high resolution (HR) image from a set of low resolution (LR) images or a set of video frames by utilising the averaging of offset images to reconstruct a high resolution image with additional information about the scene. The simplicity of the proposed technique will make the implementation of a robot-tracking system easier compared to the implementation process for mechanical trackers. Moreover, the reliability of obtained information, the low computational and hardware cost can also be added to the advantages of the introduced technique.

2 The principle of the proposed technique

The proposed technique (Figure 2) employs similar steps to other super resolution constructions methods except that a controlled measured physical shift, smaller than the basic spatial resolution of the camera, is used to generate extra information for the SR image generation, reducing the need to make assumptions. The first step is the registration process; the aim of this step is to estimate the motion and correct the differences between LR frames. The proposed motion estimation algorithm detects circular objects (fiducial points) in each LR image by utilizing image information on colour and shape and hence allows detection of the position of each object (centres and radius); the change in object position from a frame to the next frame is basically equivalent to the camera motion. The fact that the image offsets are smaller than the resolution means that some, but not all, of the pixels register changes. Moreover, each object in the experimental work consists of two nominally concentric circles, the object offset can be calculated from the small difference between the centres of the two circles and measuring a series of these with different offsets will indicate the effective resolution.

The next step is the interpolation and shift process. The interpolation process will be first performed, the advantage of this step is to increase the number of pixels in the raw images so that phase shifting can occur at the spatial resolution required to match the small physical shifts so that the sequence of images line up. However, the magnification factor in the presented work is restricted by two factors; the resolution of the image and the camera shift value. The reason behind restricting the value of interpolation is because the reconstruction process will not provide much useful information if the magnification factors are large or mismatched [22]. After applying the interpolation, different sub pixel shifts on LR images will be performed. The advantage of the proposed technique is the ability of the algorithm to shift images at a resolution less than one pixel.

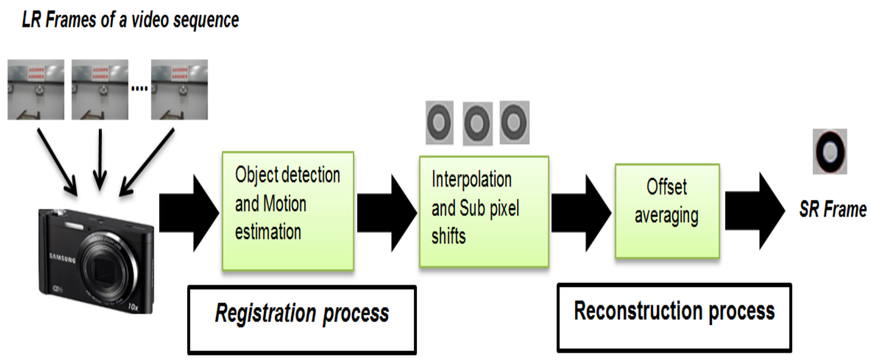

Figure 2: Block diagram of the proposed technique.

The last step is the reconstruction process is where the offset LR images will be averaged to build a high resolution image. The offset between the nominally concentric circles of the averaged (reconstructed) image will show the quality of enhancement from the proposed method.

3 The experiment set up

In order to evaluate the feasibility of the proposed image averaging technique to improve image resolution, a standard camera typically used for personal use (Samsung ST200F) was used to track multiple objects (coloured red). The camera has an effective resolution of 16.1 megapixels for still images. However, because the camera was used in video mode during the experiments, the resolution was reduced to almost 1 mega pixel (0.9216 megapixels). The objects differ in the value of the offset between the centres of two nominally concentric circles. The ability to detect these offsets will indicate effective resolution.

In Figure 3, it can be seen that the object is divided into two rows; the upper row involves objects with vertical displacements, and the lower row for objects with horizontal displacements. The camera was mounted on a numerically controlled axis and a Renishaw XL80 laser interferometer was used to measure and record actual positions. The experiment has been repeated 20 times, and 400 images have been captured. Referring to the aforementioned principle, if the camera moves a fixed and small amount relative to the base resolution, the image will be recorded slightly differently on the camera. Repeating this process, realigning the images and averaging them should improve the effective clarity and resolution.

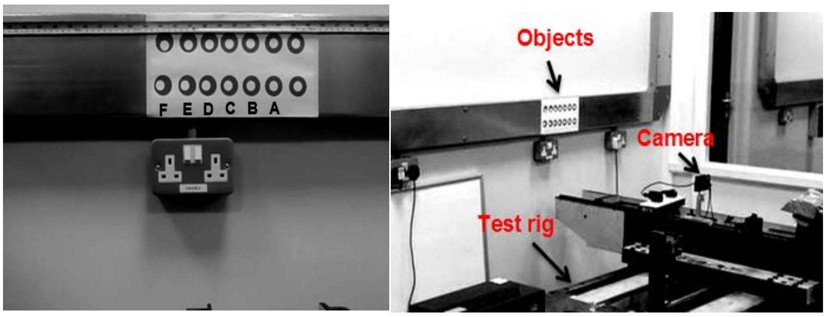

Figure 3: Board of objects hanged on the wall (left) and Experimental setup (right).

Figure 3(left) shows the first step of tracking objects. A measurement tape was used to measure the width of the image at the image depth in order to determine the approximate resolution of the image. The base resolution at this distance was almost 800µm which is wholly insufficient for tracking, for example, a robot.

$$Horizontal\ resolution \cong \frac{990.60 \times 1000}{1280} = 774\ \mathrm{\mu m}/pixel$$

The next step is setting up the experiment as shown in Figure 3(right). The camera was mounted on the test rig in front of the objects, and the axis programmed to provide a horizontal displacement of the camera. Two step sizes of 50µ, and 100µ were chosen to compare with unprocessed images to provide three points on a chart to see if there is a linear or exponential relationship between the improvement and number of averaged images. The laser interferometer (XL80) will record the position errors during the experiment (see Figure 3). This was used as a backup so that if variations were experienced it could help correlate with position error. In reality, the axis was very accurate within 2µm.

In this experiment, we analyse 6 objects in front of the moving camera. Objects (A to F see Figure 3) have real horizontal offsets in of 100, 1000, 2000, 3000, 4000, and 5000µm respectively. By using Matlab programming, our proposed algorithm extracts the motion information. The feature-based tracking algorithm is based on variation of object features in images (shape and colour). The measured offsets and standard deviations for tracked objects have been measured (in image units) for each experiment, and the results obtained classified according to the value of camera shift. Therefore, we have three cases: Case one: raw images with no camera shifts. Case two: images with a horizontal camera shift of 100µ*m*, and the number of processed images is 10 images. Case three: images with a horizontal camera shift of 50µ*m*, and the number of processed images is 20. These the three image cases were applied on 20 experiments to obtain sufficient data for statistical analysis.

Table 1: A comparison made between real and measured offsets of the objects.

Objects	Nominal offsets in µ*m*	Nominal offsets in pixels	Measured offsets in pixels (case1)	Measured offsets in pixels (case2)	Measured offsets in pixels (case3)
A	100	0.13	0.04	0.11	0.12
B	1000	1.29	0.89	5.54	10.83
C	2000	2.58	2.15	11.10	22.98
D	3000	3.88	2.82	16.95	34.13
E	4000	5.17	3.64	22.72	45.27
F	5000	6.46	4.67	28.37	57.36

Note: real offsets in Table 1 are nominal values, the accuracy of the laser printer that generated the artefact needs confirming although a basic microscope check indicated that it was within 30µm and suitable for this application. Table 1 shows a proportional relationship between the real offsets of the objects, and the measured offsets from images in image units (pixels).

In case two, and three, when the camera moves with a displacement of 100µ*m*, and 50µ*m*, the offsets of objects in images will increase 5 and 10 times (respectively) compared to the offsets in case one. Moreover, Table 2 shows that the increase in number of averaged offsets (case two and three) minimized the detectable pixel size in the image and hence increased the effective spatial resolution of the image. On the other hand, the relationship between the value of camera displacement and standard deviation is inverse, and that can be clearly notified from Table 3, the highest standard deviations (poorest performance) is obtained when there is no movement for the camera in front of the object.

Table 2: Effective pixel ratio in the averaging cases.

Objects	Nominal offsets (pixels)	Pixel ratio (case 1)	Pixel ratio (case 2)	Pixel ratio (case 3)
A	0.13	3.23	1.22	1.12
B	1.29	1.46	0.23	0.12
C	2.58	1.20	0.23	0.11
D	3.88	1.37	0.23	0.11
E	5.17	1.42	0.23	0.11
F	6.46	1.38	0.23	0.11

$$pixel\ ratio = \frac{Camera\ pixels}{Effective\ pixels}$$

Table 3: Standard deviation (µm) calculated for offset images.

Objects	STD (µm) (case 1)	STD (µm) (case 2)	STD(µm) (case 3)
A	360.9	67.5	72.8
B	313.0	93.2	72.3
C	289.2	113.9	98.6
D	241.0	102.3	55.1
E	213.6	106.7	61.4
F	245.1	88.9	47.3
RMS	281.5	96.6	69.9

The quality of the high-resolution image is obviously related to the number of Averaged LR frames used in the reconstruction process; higher the number of averaged LR frames, better is the quality of the reconstruction.

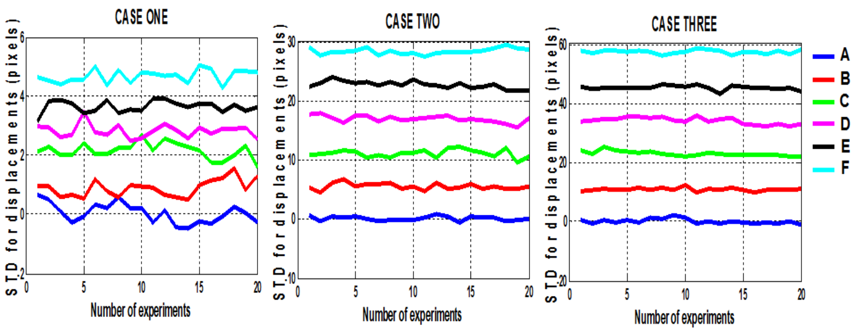

Figure 4: The relationship between the average of displacements for the objects and number of experiments.

From Figure 4, it can be seen that the offset of the objects was not constant during the iterations, the highset oscillation was in the case of raw images where there is no movement for the camera, and the oscillations are reduced considerably when the number of averaged offsets increases. Moreover, simulations showed the need to filter images from the effect of noise. The increase in the value of camera shifts leads to a reduction in the standard deviations for the measured offsets for the tracked objects in the images as shown in Figure 5.

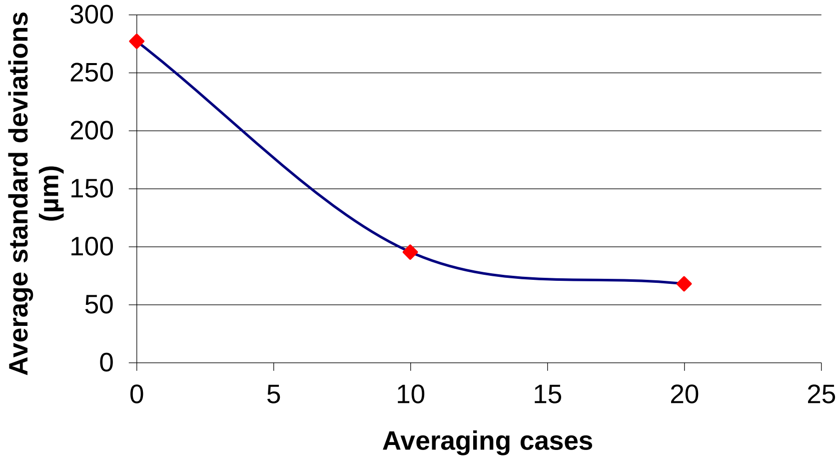

Figure 5: The relationship between the effective resolution and averaging cases.

It can be seen from Figure 5 that relationship between the averaged imaging cases and effective resolution is reasonably exponential and indicates that although the best performance (the lower standard deviation) was achieved by averaging of 20 images, the improvement in using more would be quite small. The averaging of 10 offset images improved the minimum detectable position change of a tracked object by around 3 times. However, averaging of 20 offset images improved the effective resolution by up to 4 times.

4 Conclusions

In this paper, a new approach has been presented for improving the spatial resolution of low resolution (LR) frames taken by a standard camera, the proposed method is based on the averaging of offset LR images to build an image with high quality. Although the used camera has a low resolution (less than 1mega pixel) during our experiments, the obtained results showed the ability to improve the effective resolution up to 4 times. The proposed technique is shown to be very efficient in terms of the image enhancement. During 20 experiments, the position errors (motion errors) of the moving camera were less than 2µm, and the value of any increase in positioning error should be considered for any future work therefore this will target a Piezo controlled flexure rig which should rotate the camera at high speed to effectively offset the image repeatedly to within an acceptable amount. Due to the simplicity and cost effectiveness of the proposed technique, the implementation of a robot- tracking system will be easier and more efficient compared to the implementation of mechanical trackers.

5 References

1. Baker, S. and T. Kanade. Hallucinating faces. in Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on. 2000.
2. Chaudhuri, S. and J. Manjunath, Motion-free Super-resolution2006: Springer.
3. Yang, J. and T. Huang, Image super-resolution: Historical overview and future challenges. Super-resolution imaging, 2010: p. 20-34.
4. Park, S.C., M.K. Park, and M.G. Kang, Super-resolution image reconstruction: a technical overview. Signal Processing Magazine, IEEE, 2003. 20(3): p. 21-36.
5. Chaudhuri, S., Super-resolution imaging2001: Springer.
6. Grothe, B. and T. Park, Structure and function of the bat superior olivary complex. Microscopy Research and Technique, 2000. 51(4): p. 382-402.
7. Yilmaz, A., O. Javed, and M. Shah, Object tracking: A survey. Acm computing surveys (CSUR), 2006. 38(4): p. 13.
8. Nüchter, A., 3D robotic mapping: the simultaneous localization and mapping problem with six degrees of freedom. Vol. 52. 2009: Springer Science & Business Media.
9. Ramey, A., V. González-Pacheco, and M.A. Salichs. Integration of a low-cost RGB-D sensor in a social robot for gesture recognition. in Proceedings of the 6th international conference on Human-robot interaction. 2011. ACM.
10. Cunha, J., et al., Using a depth camera for indoor robot localization and navigation. DETI/IEETA-University of Aveiro, Portugal, 2011.
11. Bannore, V., Iterative-interpolation super-resolution (iisr), in Iterative-Interpolation Super-Resolution Image Reconstruction2009, Springer. p. 19-50.
12. Katsaggelos, A.K., R. Molina, and J. Mateos, Super resolution of images and video. Synthesis Lectures on Image, Video, and Multimedia Processing, 2007. 1(1): p. 1-134.
13. Tsai, R. and T.S. Huang, Multiframe image restoration and registration. Advances in computer vision and Image Processing, 1984. 1(2): p. 317-339.
14. Negroponte, N., Being digital1996: Random House LLC.
15. Alvarez, L.D., R. Molina, and A.K. Katsaggelos, High resolution images from a sequence of low resolution observations. Digital Image Sequence Processing, Compression and Analysis, 2004. 9: p. 233-259.
16. Stark, H. and P. Oskoui, High-resolution image recovery from image-plane arrays, using convex projections. JOSA A, 1989. 6(11): p. 1715-1726.
17. Caner, G., A.M. Tekalp, and W. Heinzelman. Super resolution recovery for multi-camera surveillance imaging. in Multimedia and

Expo, 2003. ICME'03. Proceedings. 2003 International Conference on. 2003. IEEE.

18. Elad, M. and A. Feuer, Superresolution restoration of an image sequence: adaptive filtering approach. Image Processing, IEEE Transactions on, 1999. 8(3): p. 387-395.
19. Elad, M. and A. Feuer. Super-resolution reconstruction of continuous image sequences. in Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on. 1999. IEEE.
20. Elad, M. and A. Feuer, Super-resolution reconstruction of image sequences. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1999. 21(9): p. 817-834.
21. Irani, M. and S. Peleg, Improving resolution by image registration. CVGIP: Graphical models and image processing, 1991. 53(3): p. 231-239.
22. Baker, S. and T. Kanade, Limits on super-resolution and how to break them. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2002. 24(9): p. 1167-1183.