Super-Resolution Imaging
Applied to Moving Targets in High Dynamic Scenes

Olegs Mise[*a], Toby P. Breckon[b]

[a]GE Intelligent Platforms Applied Image Processing, Billerica, USA, e-mail: olegs.mise@ge.com,
[b]Applied Mathematics and Computing Group, School of Engineering, Cranfield University, UK.
e-mail: toby.breckon@cranfield.ac.uk

ABSTRACT

In modern tracking systems the ability to obtain high quality, high resolution appearance of the tracked target is often highly desirable. However, the reality of operational deployment often means that imaging systems deployed for this task suffer from limitations reducing effective image quality. These limitations can be attributed to a range of causes such as low quality video sensors, system noise, high target dynamics and other environmental noise factors. Despite the advantages of the super-resolution techniques the problem of handling complex motion still remains a challenging task for the effective super-resolution implementation. The computational complexity and large memory requirements required for the implementation of super-resolution imaging largely restrict the usage of these techniques in real-time hardware implementations. In order to improve visual quality of the tracked target and overcome these limitations, we propose a simple yet effective solution that integrates a super-resolution imaging approach based on combination of the Sum of the Absolut Differences (SAD) and gradient-descent motion estimation techniques into a novel tracking approach. In addition, the proposed approach demonstrates robustness in improved target appearance modeling that assists the overall tracking system. The presented results demonstrate this significant improvement in visual target representation whilst tracking over high dynamic scenes.

The implementation simplicity of the proposed approach makes it an attractive solution for realization on low power hardware. Such a system can be deployed on small unmanned aerial vehicles (UAV) or other hardware where size, weight and power (SWaP) is of a particular concern.

Keywords: Image super-resolution, image enhancement, image stabilization, image registration, correlation maps, template matching, target tracking.

1. INTRODUCTION

Image super-resolution is a desirable feature for many video and image processing applications. Super-resolution algorithms typically require a significant number of target samples and computational resource. These limitations make them impractical for many real time applications. As a result, super-resolution algorithms are most commonly performed offline with a significant amount of human interaction [6].

The super resolution principles using sub-pixel shifted images have been established for decades [1] [2]. However, implementational challenges, stemming from computational resource requirements as well as algorithmic reliability, still prevent such techniques from becoming commonplace. The prevalent approach for constructing such high resolution (super-resolved) images from an observed series of low resolution video frames is based on motion analysis of multiple lower resolution frames. This technique relies on precise sub-pixel motion estimation and usually fails if there is significant localized motion within the scene. Super-resolved image recovery of moving targets, as proposed here, has numerous applications in aerial imaging [13], [14], multi-modal image analysis [15], [16] and the wider automated surveillance domain [17], [18]. Here we specifically concentrate on illustrative application examples from the marine and aerial surveillance domains [19].

Super-resolution makes use of multiple scene or target samples to construct a high resolution (super-sampled) version of a given scene/target from multiple low-resolution or low-quality input images. This concept is illustrated in Figure 1 for a series of such low resolution, low quality video frames.
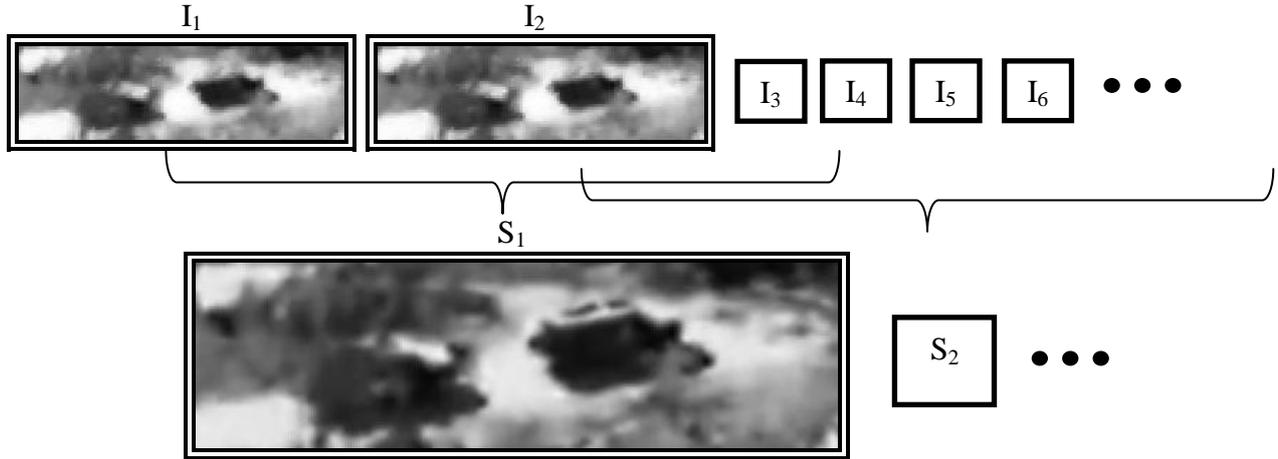
Figure 1: The process of constructing super-resolution frames from a sequential set of low resolution images. Top row $I_{1..}$ I_6 illustrates the set of low resolution images. The super-resolved result depicted at the bottom row and denote as S_1, S_2.

In this paper we introduce a novel approach that integrates a super-resolution imaging framework proposed by Hardie e*t al* [9] in to a tracking system in order to improve visual quality of the tracked target and assist the tracking system by robustly maintaining target representation. The novelty of the proposed approach lies in the combination of Sum of the Absolute Differences (SAD) and gradient-descent motion estimation techniques integrated into the tracker system whilst also offering a novel super-resolution pipeline. This assists the tracking system with distinctive advantages over prior work in the field [6], [18]. Notably, this technique effectively integrates the image super-resolution framework into the tracking system as well as robustly enhancing the core tracking components, which differs significantly from contemporary approaches [6].

The overview of the tracking system that used in our work described in Section 2. Our main concept of super-resolution image framework is introduced in Section 3 whilst Section 4 further outlines the proposed improved super-resolution assisted tracking approach. Here we will show the advantage of incorporating super-resolution technique into the tracker framework. Furthermore, we highlight the main concepts of the proposed image super-resolution algorithm that results from this framework with results shown over a variety of environmental conditions (Section 5). We present a number of strong conclusions and additionally suggest avenues for future work (Section 6).

2. IMAGE SUPER-RESOLUTION

Image super-resolution techniques have been studied for an extensive period of time. The work done by Tsai/Hung [2], and subsequently by others [5], [6], [7], pointed out that there are two main factors that limit image resolution:

1) *Image sensors* – limit the spatial resolution of the image;

2) *Front-end optics* – limit frequency response and produce unknown lens blur associated with the sensor point spread function (PSF).

It is evident that the image resolution can be enhanced by improving one of the factors mentioned above (image sensors and/or front-end optics). Usually, technological difficulties and the system cost make high resolution image-processing systems challenging to implement. Most of the time, such systems use alternative post-processing techniques for improving image resolution. One of the most common techniques is super-resolution [6].

Super-resolution techniques commonly construct high-resolution images from several observed low-resolution images, but there are several approaches for the creation of a super-resolved image from a single low resolution image based on prior information [6]. This paper focuses on a multi-frame super-resolution reconstruction technique that does not inject *a priori* information into the process. Figure 2 shows a simplified diagram of the processes involved in the super-resolution technique. The set of low resolution images are effectively registered with sub-pixel accuracy and then are remapped into a high resolution grid.

Low resolution video frames

Sub-pixel

image registration

Super-Resolution reconstruction
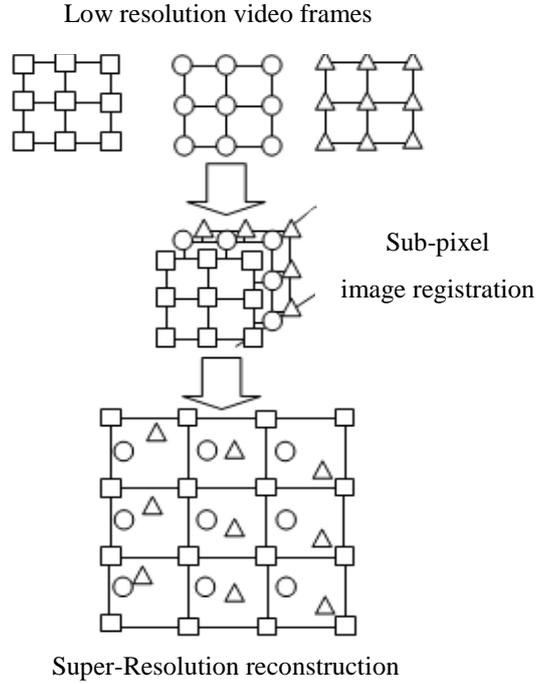
Figure 2: The top level overview of the super-resolution technique using sub-pixel image registration.

One of the important factors that affect super-resolution reconstruction using a set of low resolution images is the low resolution images must contain non-redundant information [6], [2]. This is typically achieved by sub-pixel frame-to-frame motion registration.

Most commonly super-resolution is modeled as follows:

$$Y_k = D_k H_k F_k X + V_k, k = 1,2,....K \tag{1}$$

where Y_k is a k-th observed low resolution image; X denotes high resolution image; D_k represents a down-sampling operator [24]; H_k represents an optical blurring effect; F_k represents the image geometrical transformation and V_k is noise for the k-th frame respectively. Despite the trivial mathematical representation, it is observed that the sequence of the frames/images is represented using the linear system of equations as follows:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_k \end{bmatrix} = \begin{bmatrix} D_1 H_1 F_1 \\ D_2 H_2 F_2 \\ . \\ . \\ D_k H_k F_k \end{bmatrix} X + \underline{V} \tag{2}$$

This linear system of equations shows a very important aspect that affects the super-resolution (SR) reconstruction technique. The term X must be unchanged during the SR reconstruction cycle. It means that this representation is only valid if the term X is a subject to global transformations of D_k, H_k, F_k (down-sampling, blurring and geometric transformation). Any local motions (objects) that disagree with the global motion of the entire picture will not fit in the model described in Equation 2. In the real imaging system terms D_k, H_k, F_k are mostly unknown and need to be estimated directly from the low resolution images. Subsequently, the direct solution of the linear system of equations (Eq. 2) is not applicable due to ill-condition of the system. As a result, prior regularization knowledge for the high resolution is essential [2], [3].

For the implementation of the super resolution algorithm we use an approach proposed by Hardie *et al* [9]. The main advantage of this super-resolution approach is that this technique does not require a lot of computational and memory resource, which makes it very attractive in many practical systems. Hardie *et al* [9] proposed an idea of using a fast Adaptive Weiner Filter (AWF) integrated to the SR algorithm. This work [9] described the "blur-then-warp" observation model, as a non-uniform interpolation, followed by a uniform sampling with the de-blurring restoration process performed on the final super-resolved image. The non-uniform interpolation and restoration processes are performed in one step using a weighted sum operation within the sliding rectangular window. The algorithm based on the "blur-then-warp" observation model is illustrated in Figure 3.
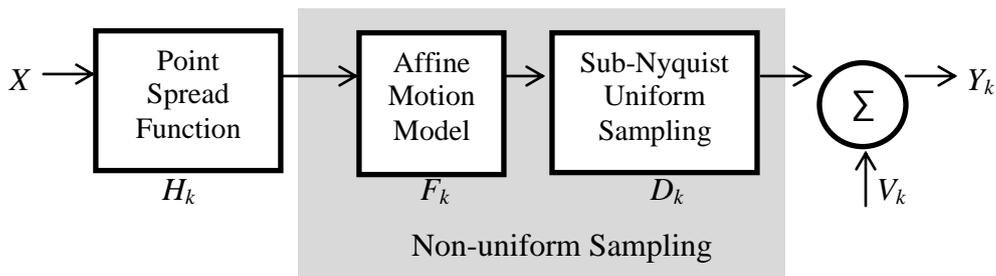
Figure 3: Blur-then-warp observation model. Y_k is a k-th observed Low Resolution image; X denotes High Resolution image; D_k is the down-sampling operator; H_k represents optical blurring effect; F_k shows image geometrical transformation; and V_k is noise for the k-th frame respectively.

Assuming that the optical distortion of the system is spatial invariant and identical for all low resolution images the common super-resolution workflow would be: *Low resolution image registration*, *Non-uniform interpolation*, *De-blurring and noise removal*. In this work we used the gradient descent technique described in [3] for the sub-pixel motion estimation. This approach allows registering images with large degrees of freedom. The main drawback of this approach is that its initial calculation position must be close to the final solution. This problem is mainly dictated by the gradient descent algorithm being prone to finding wrong local solution. Therefore, this technique is unreliable with large object frame to frame motion, and, especially, if the object size is small or corrupted by image noise. In order to overcome these limitations we used the Sum of the Absolute Differences (SAD) technique [4] to pre-calculate large object displacement. Subsequently, we applied the gradient descent approach for a more precise image registration. Following the image registration, it is remapped to the high resolution grid as outlined previously in Figure 2. If two or more mapped low resolution samples are positioned in the same place in the high resolution grid, they are averaged. The amount of the high resolution grid population depends on the number of processed low resolution frames and their relative motion. Figure 4 shows the process of filling the high resolution grid with registered image samples. Each time when a new frame arrives, it is registered with the reference frame and re-projected into common high resolution grid. It is important to note, if no motion between low-resolution frames occurred, all samples will be placed in the exactly identical place in the high resolution grid, therefore, making it impossible to construct super-resolution images.
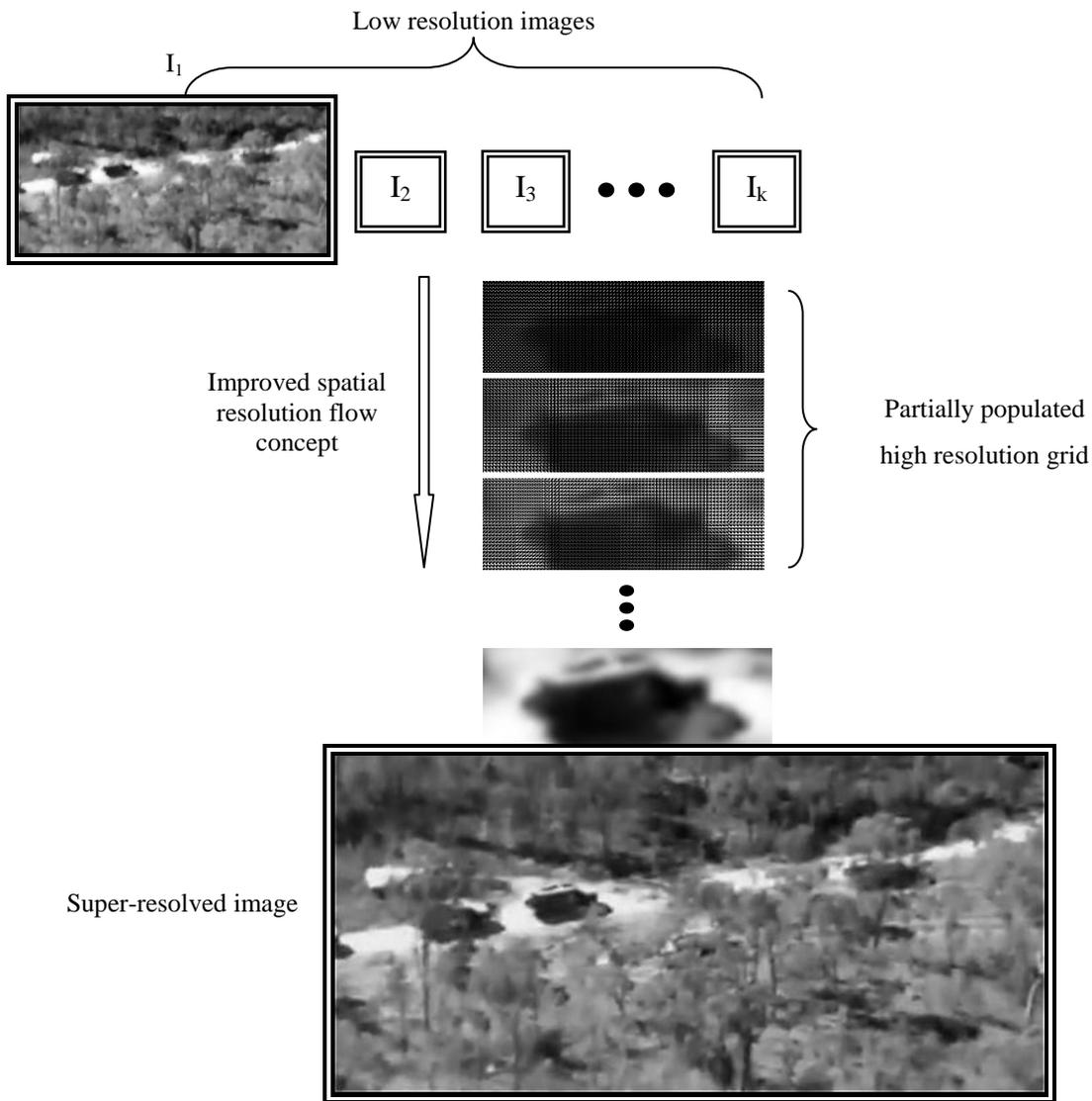
Figure 4: Concept of the super-resolution reconstruction technique. Video footage obtained from the video-sharing website YouTube by LordZearo (2010) [11].

Illustrative result using this technique is shown in Figure 5. Figure 5 illustrates the results obtained under well controlled scenarios where data was generated using "EIA resolution chart 1956" [12]. Using the EIA resolution chart a generated set of ten low resolution images was obtained using a digital camera. During this experiment the camera was randomly shifted in order to create a motion that is necessary for the super-resolution reconstruction process. The results shown in Figure 5 illustrate the comparisons between original low resolution original image (Figure 5 a), original 2x bilinear interpolated (Figure 5 b), an averaged version using ten registered images (Figure 5 c), and the result obtained by our chosen the super-resolution approach [9] (Figure 5 d).
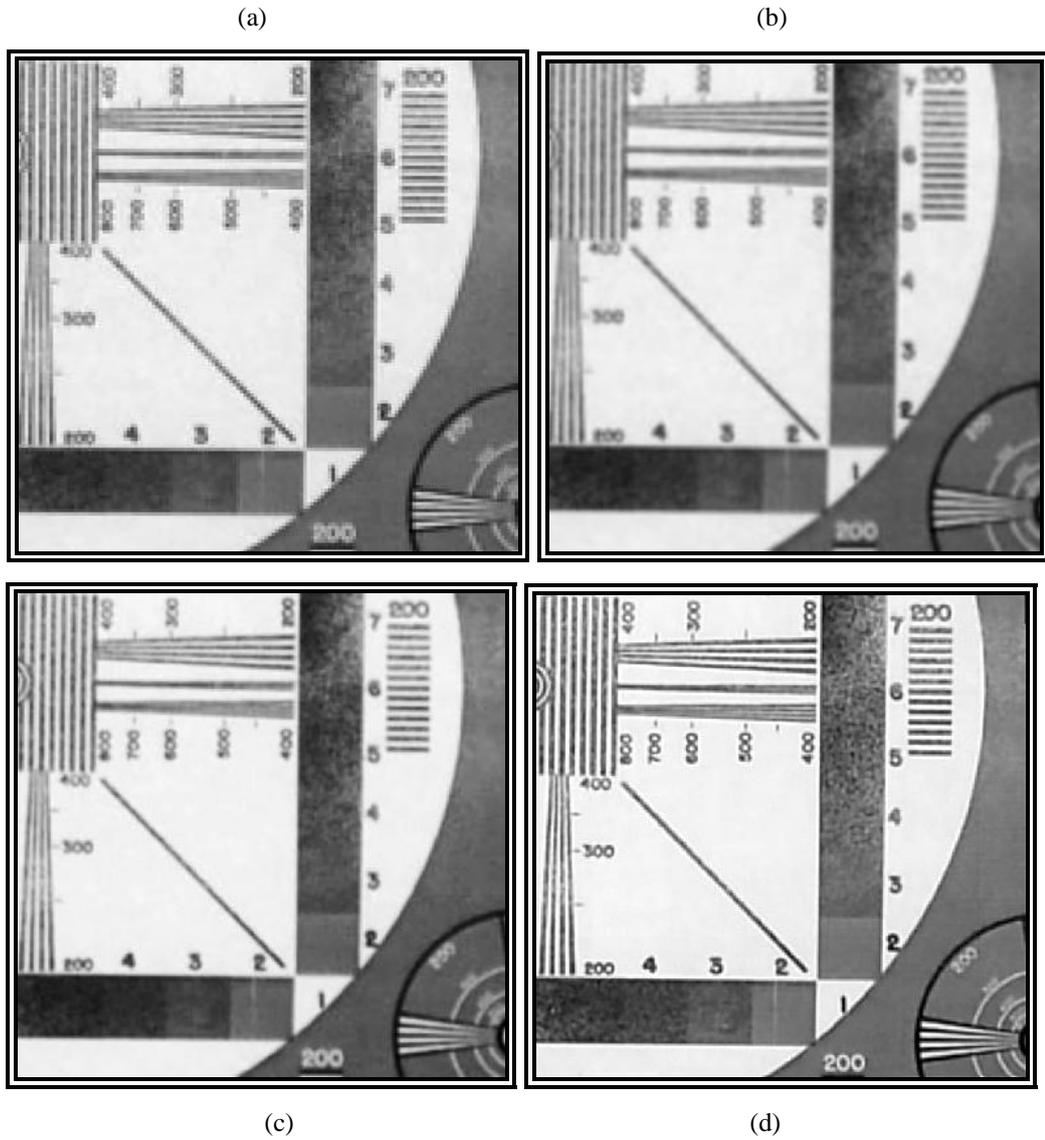
Figure 5: EIA RESOLUTION CHART 1956 original low resolution image (a), original 2x bilinear interpolated (b), original 2x bilinear interpolated and averaged using ten registered images (c) and super-resolved image (d).

3. SUPER-RESOLUTION ENHANCED TRACKING SYSTEM

We integrated our proposed super-resolution framework described in (Section 2) into the tracking system using a simple yet effective block matching correlation algorithm based on the Sum of Absolute Differences (SAD), to recover in scene motion. Major advantages of the block-matching approach include its direct matching nature, that simplifies the tracking prediction filtering mechanism, and the computational efficiency makes it suitable for the implementation on low powered devices. Conversely, such block-matching techniques suffer from poor performance when dealing with non-rigid shapes [6] and also with the existence of repetitive scene background patterns as multiple possible solutions to the SAD determination. Figure 7 depicts the top level block diagram of the tracking system with the integrated super-resolution framework used in our work.
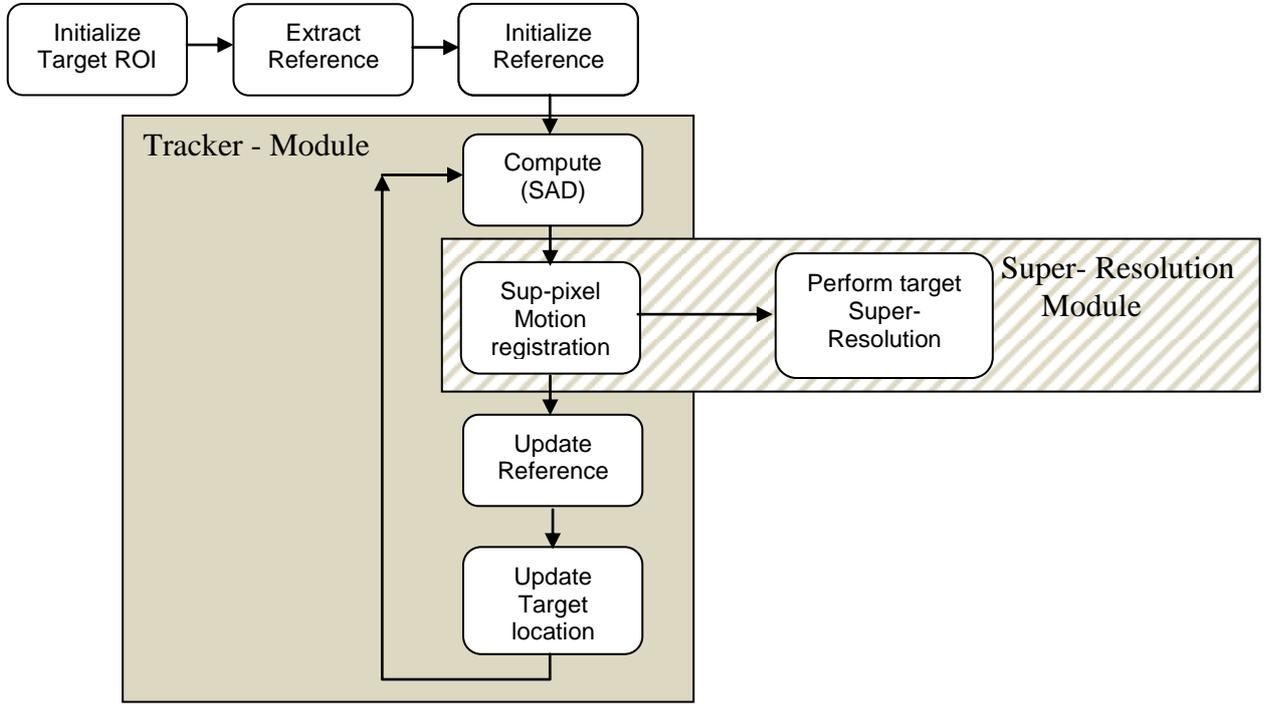
Figure 6: Tracking system with integrated super-resolution - top level block diagram.

Here the initial position of the target (initialized by the operator) is accepted as a reference image and initializes the target reference image (Section 3.2). The following SAD-based correlation approach (Section 3.1) is used to perform a full search over a limited search window (usually a few pixels around the target location). Based on a gradient-descent approach [3], sub-pixel motion registration is performed on the extracted target reference. This effectively compensates the motion according to the position in the live image. Information provided by the sub-pixel motion registration block is also used in our developed super-resolution framework in order to align the low resolution representation of the target. Subsequently, the registered result is up-sampled and placed into the up-sampled grid by the super-resolution algorithm framework (Section 2). At the same time the target reference is updated by the motion compensated target appearance estimated by the gradient-descent technique. The final stage updates the target location results reported by the (SAD) and sub-pixel motion registration.

3.1 Sum of Absolute Differences (SAD) Motion Estimation

The Sum of Absolute Differences (SAD) is well established within image processing literature [4]. This approach essentially measures similarity between image blocks by taking an absolute difference between every pixel in the reference block, R, and a corresponding pixel in the current image, I. The differences are summed producing a so-called correlation surface with the maximum value accepted as best matching position in image I for reference block R. As shown in Equation 3:

$$Cr = \sum_{(x,y) \in w} | R(x, y) - I(i + x, j + y) | \qquad (3)$$

Where: Cr is the result of the SAD operation, i.e. correlation value, R is a reference patch, I is processed image patch, w is the target/reference domain and finally i and j specify the indices over that search surface, where block matching is performed.

3.2 Target Reference Update

The target model update is most commonly performed as a weighted average of the target appearance extracted every frame [8]. Here this approach is implemented using an alpha-weighted update between subsequent frames.

$$F = \alpha I_t + (1 - \alpha) I_{t-1} \qquad (4)$$

where: F – is an alpha filter result, α – update coefficient, I – video frame.

During the initialization phase (Figure 6) the initial position of the target is accepted for the reference. This allows accepting the first incoming reference as is without incremental update. Subsequently the alpha parameter (Eq 4) is then reconfigured to facilitate slow incremental update. In general a parameter of alpha equals {0.1 and 0.025} range and mainly depend on target dynamics.

During the experiments it has been found that *faster* alpha update can lead to reference sliding, whereas *slower* alpha update can lead to a completely unusable blurred reference. In order to improve the quality of the generated reference, we applied the gradient-descent technique [3] for the sub-pixel motion estimation. This approach allows estimating affine or full projective transformation between the reference I_{t-1} and a live image I_t (Eq. 4) and, therefore, compensating motion before a new reference is passed to the alpha-filter as follows:

$$F = \alpha I_t + W^{-1} (1 - \alpha) I_{t-1} \qquad (5)$$

where: F – is an alpha filter result, α – update coefficient, I – video frame and W – is a motion parameters estimated by the gradient descent motion estimation algorithm [3].

The result is then processed through the reference update process where the final appearance of the target reference is formed. Figure 7 depicts the live raw image (a), the result obtained from the alpha filter without motion compensation (b) and the result with motion compensation (c). The blurred image shown in Figure 7 (b) occurred due to the fact that the images were not properly allied (motion compensated) prior being processed by the alpha-filter. Therefore, this resulted in the accumulative blurred image. By contrast, the motion compensated accumulative result in Figure 7 (c) is free from the blurring artifacts.

| (a) | (b) | (c) |
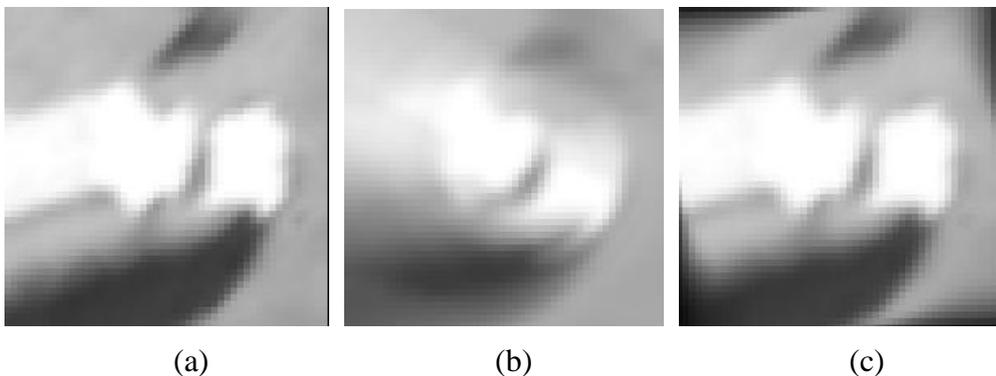
Figure 7: Reference update results. The live raw image (a), the averaged reference without motion compensation (b) and motion compensated reference (c).

4. EXPERIMENTAL RESULTS

Over a series of experiments we examined the robustness of the proposed approach of the super-resolution and applied it to the moving target. The full algorithm implementation was realized in the Matlab in order to determine its performance. In our experiments we used uncompressed video sources obtained from [10]. The main goal for the conducted experiments was to reliably track the object through a sequence of the video frames and at the same time to produce a super-resolved image of the tracked object. The first experiment shown in Figure 8 illustrates the super-resolution results while tracking the boat. In this experiment 22 sequential frames were processed by the Hardie *et al* super-resolution technique [9] and formed a 4x super-resoled image. Figure 8 (a), (b) and (c) shows the 100th, 200th and 230th raw video frames. The frames in Figure 8 (a), (b) and (c) were 4x bilinear up-sampled in order to visually compare them to the 4x super-resolution output. Figure 8 (d), (e) and (f) respectively depict the super-resolved results for the frames 100, 200 and 230. The cropped section of the boat shown in Figure 8 (g) and (h) shows the raw frame and the super-resolved result.

(a) (b) (c)

(d) (e) (f)

(g) (h)
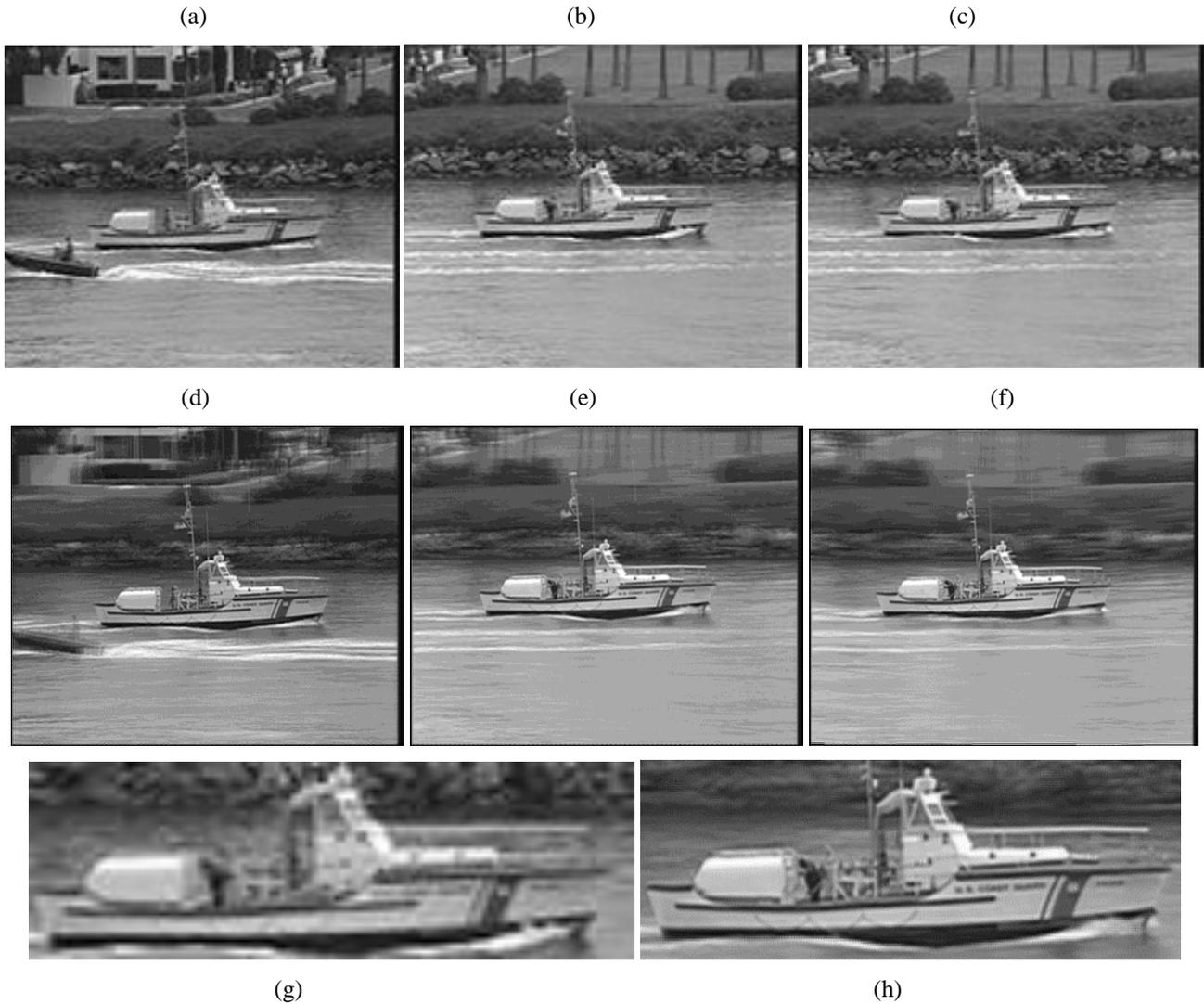
Figure 8: 4x bilinear interpolated raw video frames shown in (a), (b) and (c). Super-resolved output presented in (d), (e), (f). The cropped raw video frame depicted in (g) and cropped super-resolved image presented in (h).

The next experiment shown in Figure 9 demonstrates the results of the algorithm applied to a moving ship. The super-resolution result was generated using 65 consecutive low resolution frames. The output of the super-resolved image is set to be 4x of the original low resolution video frame. Figure 9 (a) shows the 4x bilinear interpolated low resolution frame, Figure 9 (b) depicts the super-resolved result and the corresponding cropped section shown in Figure (c) and (d) respectively.

(a)

(b)

(c)   (d)
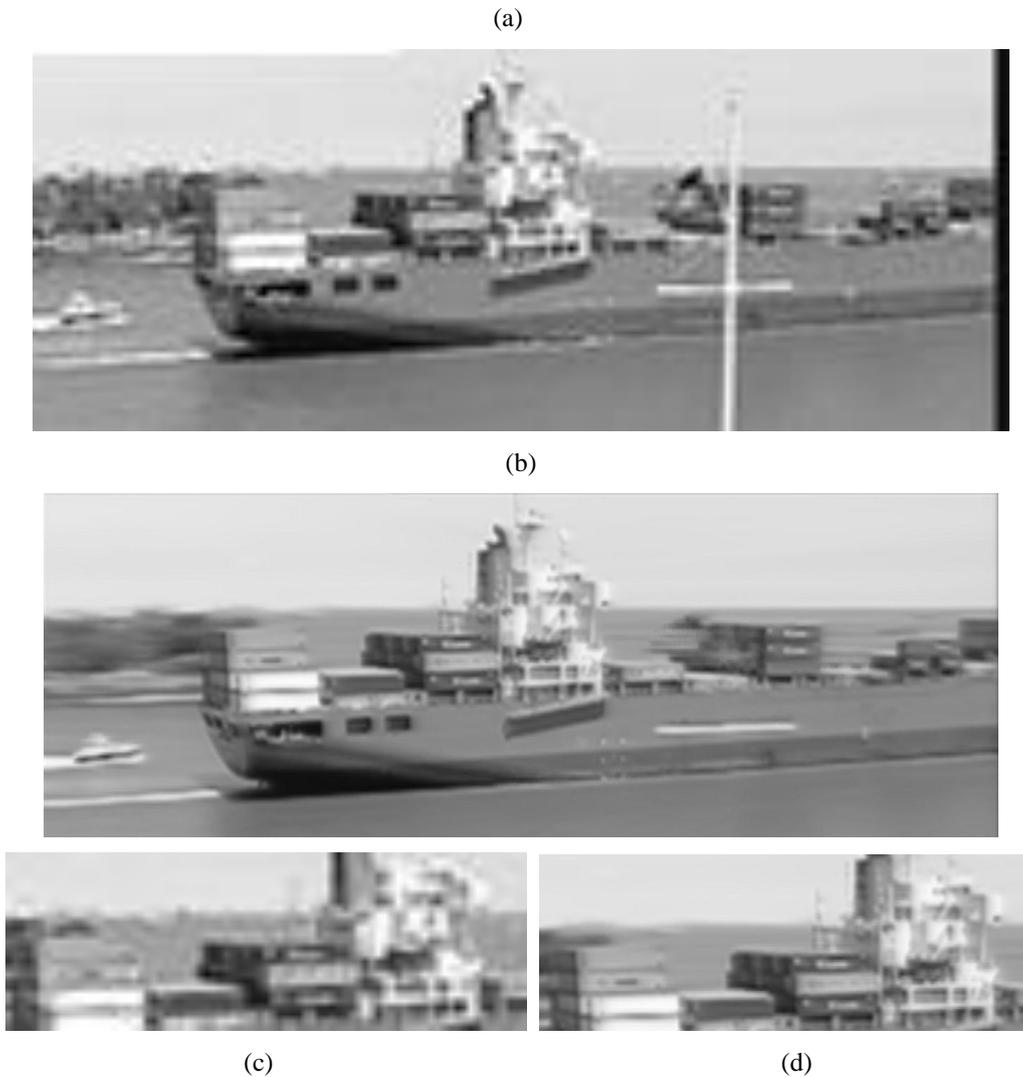
Figure 9: 4x bilinear interpolated image (a), super-resolved result (b) and corresponding cropped section (c) and (d).

5.   CONCLUSIONS

In this work we presented a novel framework where a super-resolution technique developed by Hardie *at al* [9] was integrated into the tracking system. This approach significantly improves the overall robustness of the tracking system by reliably maintaining the target appearance under large transformation degrees. In order to improve the image registration process, we propose a simple yet effective solution that integrates a super-resolution imaging approach based on the combination of the Sum of the Absolut Differences (SAD) and gradient-descent motion estimation techniques [3] into a novel tracking approach. The presented results demonstrate this significant improvement in visual target representation whilst tracking over high dynamic scenes. The proposed technique also shows an attractive solution for realization on

low power hardware. The work we presented here differs from the approach developed by Hardie *at al* [9] as it primarily designed to work and assist the tracking system whilst it is tracking the target.

The future work will be dedicated to atmospheric turbulence distortion removals which significantly degrade the performance of the super-resolution framework described in this paper. Future work will also investigate the use of super-resolution within a real-time stereo sensing context [20], [21], for future video mosaicking [22] and for the derivation of cross-spectral super-resolved scene mapping [23].

REFERENCES

[1] S. Borman and R.L. Stevenson, "Spatial resolution enhancement of low-resolution image sequences. A comprehensive review with directions for future research," Lab. Image and Signal Analysis, University of Notre Dame, IN 46556. Tech. Rep., 1998.

[2] R. Y. Tsai and T. S. Huang. "Multipleframe image restoration and registration". In Advances in Computer Vision and Image Processing", pages 317–339. Greenwich, CT: JAI Press Inc., 1984.

[3] Simon Baker , Iain Matthews." Lucas-Kanade 20 Years On: A Unifying Framework Part 1: The Quantity Approximated, the Warp Update Rule, and the Gradient Descent Approximation", International Journal of Computer Vision - IJCV , 2004

[4] C.J. Solomon, T.P. Breckon. "Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab" , Wiley-Blackwell, 2010.

[5] N. Nguyen, P. Milanfar, and G. Golub, "Efficient generalized cross-validation with applications to parametric image restoration and resolution enhancement," IEEE Trans. Image Processing, vol. 10, Pages 1299-1308, Sept. 2001.

[6] Peyman Milanfar. (2011) "Super-Resolution Imaging", CRC Press, Taylor & Francis Group.

[7]   Iain E. G. Richardson (2004). H.264 and MPEG-4 Video Compression Video Coding for Next-generation Multimedia". The Robert Gordon University, Aberdeen, UK. John Wiley & Sons Ltd.

[8] J. Kim and S. Yang, "An efficient search algorithm for block motion estimation" IEEE Workshop on Signal Processing Systems, Pages 100-109, Oct. 1999.

[9] Russell C. Hardie, Kenneth J. Barnard, Raul Ordonez, "Fast Super-Resolution with affine motion using an adaptive Wiener filter and its application to airborne imaging" 19 December 2011, OPTICS EXPRESS 26208, Vol. 19, No. 27

[10] Arizona State University uncompressed YUV videos: http://www.trace.eas.asu.edu/yuv

[11] LordZearo (2010), Real UAV footage, available at http://www.youtube.com/watch?v=M8-tu44cg1Y (accessed 5th October, 2012).

[12] EIA resolution chart 1956, Real UAV footage, available at: http://commons.wikimedia.org/wiki/File:EIA_Resolution_Chart_1956.svg (accessed 20th May, 2013).

[13] Real-time People and Vehicle Detection from UAV Imagery (A. Gaszczak, T.P. Breckon, J. Han), In Proc. SPIE Conference Intelligent Robots and Computer Vision XXVIII: Algorithms and Techniques, Volume 7878, No. 78780B, 2011.

[14] Automatic Salient Object Detection in UAV Imagery (J. Sokalski, T.P. Breckon, I. Cowling), In Proc. 25th International Conference on Unmanned Air Vehicle Systems, pp. 11.1-11.12, 2010.

[15] Consistency in Muti-modal Automated Target Detection using Temporally Filtered Reporting (T.P. Breckon, J. Han, J. Richardson), In Proc. SPIE Electro-Optical Remote Sensing, Photonic Technologies, and Applications VI, Volume 8542, No. 85420L-1, pp. 23:1-23:12, 2012.

[16] Human Pose Classification within the Context of Near-IR Imagery Tracking (J. Han, A. Gaszczak, R. Maciol, S.E. Barnes, T.P. Breckon), In Proc. SPIE Security and Defence: Optics and Photonics for Counterterrorism, Crime Fighting and Defence, SPIE, 2013. (to appear)

[17] Multi-Modal Target Detection for Autonomous Wide Area Search and Surveillance (T.P. Breckon, A. Gaszczak, J. Han, M.L. Eichner, S.E. Barnes), In Proc. SPIE Security and Defence: Unmanned/Unattended Sensors and Sensor Networks, SPIE, 2013. (to appear)

[18] Kalal, Z., Mikolajczyk, K., & Matas, J. (2012).Tracking-learning-detection. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 34(7), 1409-1422."

[19] Auslander, Bryan, Kalyan Moy Gupta, and David W. Aha. "A comparative evaluation of anomaly detection algorithms for maritime video surveillance." In SPIE Defense, Security, and Sensing, pp. 801907-801907. International Society for Optics and Photonics, 2011.

[20] An Empirical Comparison of Real-time Dense Stereo Approaches for use in the Automotive Environment (F. Mroz, T.P. Breckon), In EURASIP Journal on Image and Video Processing, Springer, Volume 2012, No. 13, pp. 1-19, 2012.

[21] A Foreground Object based Quantitative Assessment of Dense Stereo Approaches for use in Automotive Environments (O.K. Hamilton, T.P. Breckon, X. Bai, S. Kamata), In Proc. International Conference on Image Processing, IEEE, 2013. (to appear)

[22] Real-time Mosaicing from Unconstrained Video Imagery for UAV Applications (M. Breszcz, T.P. Breckon, I. Cowling), In Proc. 26th International Unmanned Air Vehicle Systems, pp. 1-8, 2011.

[23] On Cross-Spectral Stereo Matching using Dense Gradient Features (P. Pinggera, T.P. Breckon, H. Bischof), In Proc. British Machine Vision Conference, pp. 526.1-526.12, 2012.

[24] Burt and Adelson, 1983 P.J. Burt and E.H. Adelson. The Laplacian pyramid as a compact image code. IEEE Transactions on Communiations, 31 4 :532 540, April 1983.