Generation of Super-Resolution Stills from Video

Bernardt Duvenhage
Defence, Peace, Safety and Security
Council for Scientific and Industrial Research
Pretoria, South-Africa
Email: bduvenhage@csir.co.za

Abstract—**The real-time super-resolution technique discussed in this paper increases the effective pixel density of an image sensor by combining consecutive image frames from a video. In surveillance, the higher pixel density lowers the Nyquist rate of the sensor which improves the detection, recognition and identification (DRI) task performance of the system.**

When a sensor lingers on a stationary target or tracks a moving target then the image of the target would with time move around slightly on the focal plane. If one accurately registers the image of the target on the focal plane to some reference then one can increase the effective sensor pixel density by stacking or appropriately combining the registered images.

The super-resolution technique operates on the focal plane array after the image has been degraded by the Modulation Transfer Function (MTF) of the lens and atmosphere. Any high frequencies lost due to the atmosphere or lens cannot be recovered. However, if the MTFs of the lens and atmosphere are good enough to cause aliasing on the focal plane then the sharp stack algorithm discussed here can at least double the resolving power of the sensor.

I. Introduction

The focus of this paper is on combining consecutive video frames to create a super-resolved still image. The video frames are first registered and then appropriately stacked to reconstruct the still image signal from the multiple low resolution image signals.

The scope of the work includes analysing the effectiveness of the well-known average stack operation and presenting a new sharp stack operation that results in improved resolution enhancement. The super-resolution technique was developed for real-time execution on small form factor CPU-only computers. The Lucas-Kanade optical flow algorithm is used to do real-time sub-pixel accurate registration of the low resolution images.

The next section, Section II gives the background to multi-frame super-resolution followed by an overview of the related work in Section III. Section IV analysis the average stack operation and proposes the new sharp stack operation. The Lucas-Kanade image registration implementation is then briefly described in Section V. Section VI puts forward a proposal on how to measure the resolution enhancement of a super-resolution algorithm. Section VII shows some comparative as well as additional results. Section VIII reflects on the success of the undertaking, highlights limitations and points ahead to future work.

II. Background

In Surveillance the detection, recognition and identification (DRI) task performance is dependent on the number of resolved spatial cycles on target [1]. The number of resolved spatial cycles on target is in turn dependent on the surveillance system's frequency response (SFR). The SRF is a product of the atmosphere's modulation transfer function (MTF), the lens' MTF and the frequency response of the focal plane array (FPA). The ideal lens, for example, would have a modulation transfer of one for all frequencies which would contribute to a good SFR.

If one assumes that the lens is good enough and that the atmospheric effects are negligible then the SFR is dominated by the frequency response of the FPA. The frequency response of the FPA is a function of the spatial resolution (or spatial sampling rate) of the FPA. This sampling rate therefore becomes an important contributor to the resolving power of the system. Nyquist rate is defined as the lower bound on the sampling rate required to effectively reproduce a signal. Nyquist frequency is the reciprocal of Nyquist rate and normally *half* of the sampling frequency of the FPA.

Below two pixels per cycle (above 0.5 cycles/pixel) the image signal usually becomes aliased. The original high resolution (HR) image can then no longer be reconstructed from a single low resolution (LR) input frame. The term *HR image* is reserved for the continuous image function. The reconstructed digital image is referred to as the super-resolution (SR) image.

However, if one could somehow add more pixel samples to the FPA then the Nyquist rate of the sensor would be lowered. This is true even if the pixel samples overlap. Lowering the Nyquist rate of the sensor of course has the desired effect of increasing its resolving power.

If video frames are first registered and then appropriately stacked such that image features are aligned between consecutive frames then the stacking operation computationally adds pixel samples to the FPA which still lowers the Nyquist rate of the sensor. This is the basis for the multi-frame super-resolution image processing discussed in this paper.

The super-resolution technique described here was developed for real-time execution on small low power processors which places some additional limitations on the super-resolution technique. The paper therefore presents the determined effort to apply a fast image registration algorithm in combination with only an image stacking operation for generating super-resolution stills from video.

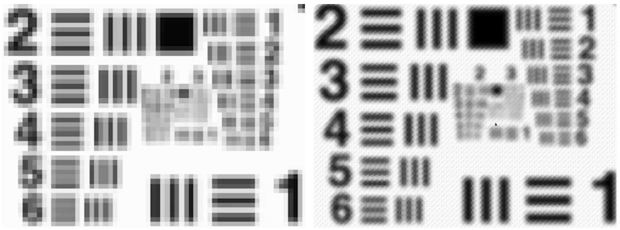

Fig. 1. An Example of a Low Resolution (LR) Input image (left) and the Resulting Average Image (right) using Eight LR Images

III. Related Work

This section captures an overview of existing research related to multi-frame super-resolution techniques. Research on other types of super-resolution such as single frame super-resolution and image enhancement in the presence of atmospheric effects such as scintillation are not discussed.

The image formation process causes image degradation due to the lens Modulation Transfer Function (MTF), the integrative FPA sampling and some quantisation and read-out noise from the FPA sensor. The classic spatial reconstruction based techniques build an image formation model, invert it and then use the inverse model to estimate the original HR image from the LR input images as discussed by among others Keren et al. [2], and Irani and Peleg [3]. Such a super-resolution algorithm then has the following steps:

- Sub-pixel accurate registration of the input LR images.
- Stacking of the registered LR images to create the SR prior.
- An iterative refinement of the SR image using the forward and inverse image degradation models.

Generation of the SR image from the image degradation model is an ill posed problem. Schultz and Stevenson [4], and others [5] [6] have used regularisation to introduce the additional information required to solve the ill-posed reconstruction problem. This information is in the form of restrictions on smoothness and a philosophical justification for regularisation is that it attempts to impose Occam's razor on the solution. In other words the simplest SR result is the most probable one. Finding this solution is however very time consuming.

The SR prior is usually the average image. He et al. [7] do a median stack instead of an average stack of the LR images to remove sampling and registration outliers, but there is still an inherent spatial filter when combining the different stacked LR images. This paper investigates a filter-less stacking that removes the need for the iterative refinement step.

IV. Stacking Low Resolution Images

This section analysis the average stack operation and proposes a new sharp stack operation. In the average stack the most recent N frames of a video stream are registered against a reference frame and upscaled by some factor. The aligned pixels are then averaged to produce a single upscaled image called the average image. Figure 1 shows an example of a low resolution input image and the resulting average image. The average image already produces quite a significant improvement over the low resolution input image.

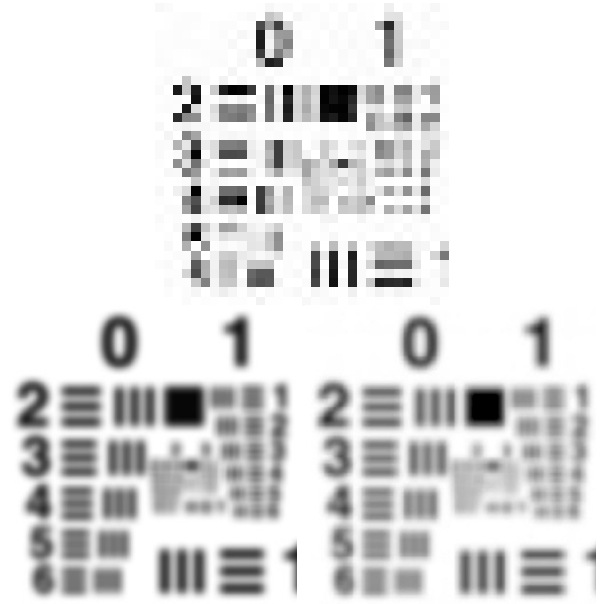

Fig. 2. Another Example of a Low Resolution (LR) Input image (top), the result of the Double Box Filtered Model (bottom-left) and the Actual Average Image (bottom-right) using Many LR Images

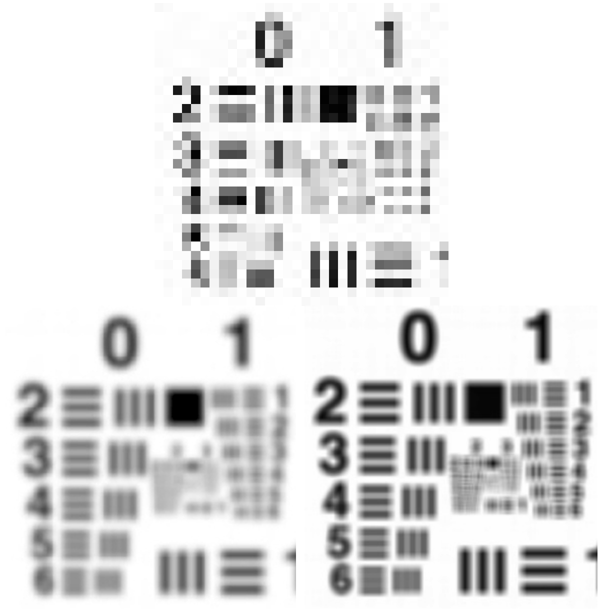

Fig. 3. A Low Resolution (LR) Input image (top), the Actual Average Image (bottom-left) using Many LR Images and the result of the Single Filter Sharp Stack Model (bottom-right)

Keren et al. [2] suggested that the registered average stack operation (using many LR images) generates a Gaussian blurred version of the HR image. However, in this paper it is proposed that the effect of the stacking operation is separable into two sequential box filters. Of course, by the central limit theorem, two sequential box filters approximate a Gaussian filter to some degree, but the motivation for the separation lies deeper than this theorem.

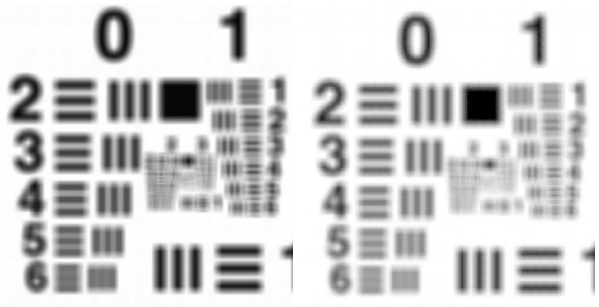

Fig. 4. The Result of the Single Filter Sharp Stack Model (left) and the Actual Sharp Stack Result (right) using Many LR Images

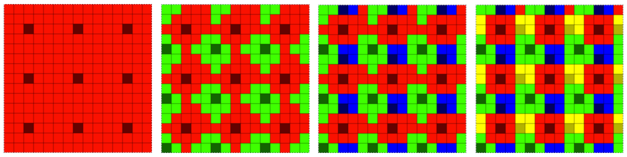

Fig. 5. The Nearest Neighbour Filter Result for the First Four Input Frames

The first filter is proposed to be due to the integrative sampling of the FPA which results in the values of the LR image pixels. The integrative sampling of the FPA is effectively a sparse box filter of the HR image. It is sparse because one convolution is done per LR pixel centre only.

The second filter is proposed to be due to the averaging inherent in the average stack operation. The value of a pixel in the stacked image is influenced by a local neighbourhood of LR pixel *splats*. The shape of this second filter is somewhat dependent on the spatial distribution of the registered LR pixels, but approximately a box filter in practice.

Figure 2 shows the result of the double box filtered model of the average image. The bottom-left image shows the double box filter of the HR image and the bottom-right image shows the actual average image given many input LR images. Note the similarity between the double filtered model and the actual average image. Recall that the size of the box filter is the same as the LR pixel size.

One cannot avoid the first filter. It is due to the physical size of the LR sensor pixels on the focal plane array. One might however be able to minimise the effect of second filter which is inherent in the average stack operation.

If one could avoid the second filter by smart stacking then it follows that the stacked image may be modelled by only one (the first) box filter. Such a simulated result is shown in Figure 3. I call the resulting single filter result the model of the sharp stack. Note the somewhat sharper appearance of the bar targets.

It seems plausible then that one could generate an actual sharp stack result by reducing each LR pixel to a single SR pixel (instead of a box covering multiple SR pixels) before stacking. Figure 4 shows the result of the sharp stack model and an actual sharp stack result generated in this way given many input LR images.

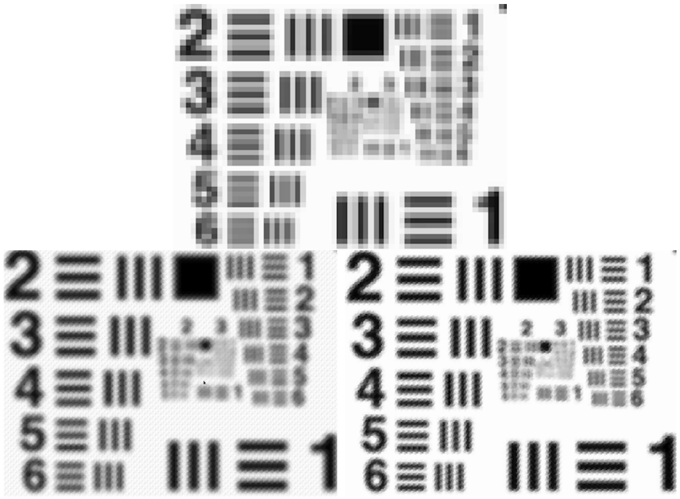

Fig. 6. The Low Resolution Input (top), Average Stack (bottom-left) and Sharp Stack (bottom-right) with Nearest Neighbour Filter Given Eight Input Frames

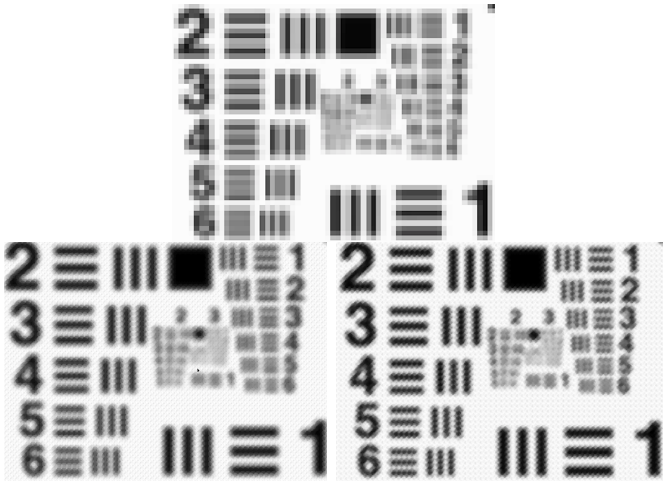

Fig. 7. The Low Resolution Input (top), Average Stack (bottom-left) and Sharp Stack (bottom-right) with Kernel Filter Given Eight Input Frames

The fact that the double box filter and single box filter match the average image and sharp stack results strengthens the double box filter conjecture. The sharp stack operation then has the ideal modulation transfer function of one. Section VI will show that the sharp stack is probably the best that one can do and as good as alternative super-resolution techniques.

However, although the sharp stack is simple and efficient to implement, if one is not able to linger long enough on a target or if the target changes appearance too quickly then the sharp stack would not produce a dense result. The rest of this section discussed two filters that are useful for producing a dense result in such cases.

The two filters investigated to produce a dense result from sparse samples are:

- Nearest neighbour Voronoi filter.
- Kernel filter.

Figure 5 demonstrates the operation of the nearest neighbour filter for the first (red), second (green), third (blue) and fourth (yellow) registered and upscaled LR frames. The darker pixels indicate the centres of the registered Low Resolution (LR) input pixels for each frame. The super-resolution up scaling is five in this case.

A radius image of the same resolution as the SR image is used to keep track of the pixel distances from the LR pixel centres. Each entry in the radius image holds the distance/radius to the closest LR pixel.

For each registered LR pixel centre added to the SR image a local neighbourhood of SR pixels are processed. A SR pixel is given the value of the LR pixel if the LR pixel centre is closer than the radius given in the radius image. The radius image is then updated as the new LR pixel is nearer than the previous LR neighbour was. The radius image is initially filled with the diagonal radius of the LR pixel.

Figure 6 shows the LR input, average stack and nearest neighbour (i.e. Voronoi) filter results given only eight registered input frames. Note the improvement in target resolution which is most noticeable on the smaller 1 - 6 targets on the right-hand side of the USAF1951 target.

The kernel filter operates similar to the nearest neighbour filter except for one detail. If the SR pixel is nearer to incoming LR pixel centre than the corresponding radius in the radius image then the LR pixel is blended with the SR pixel instead of replacing it. The blending is done for the colour image as well as the radius image. The radius is referred to as the kernel radius.

The incoming LR colour is blended with the current SR pixel colour using a specific weighting. This weighting provides an immunity to noise, but causes the result to converge slower than the nearest neighbour filter.

The distance of the incoming LR pixel centre to the current SR is also blended with the current SR radius using a specific weighting. This weighting is coupled to the colour blending weight; the radius should converge slower than the pixel colour. Note that the SR pixel is only processed when closer to the LR pixel centre than the current radius. The SR kernel radius is therefore strictly decreasing.

Figure 7 shows the average stack and kernel filter results given only eight registered input frames. Note the improvement in target resolution which is most noticeable on the smaller 1 - 6 targets on the right-hand side of the USAF1951 target.

The nearest neighbour and kernel filters have similar quality performance in the limit. The nearest neighbour method is however simpler, faster and has improved quality for a small number of inputs. On the other hand when the input is very noisy then the kernel method offers an advantage due to its temporal smoothing behaviour.

V. Registration of Images

This section briefly describes the Lucas-Kanade (LK) optical flow image registration implementation. The implementation closely follows the original 1981 Lucas-Kanade paper [8] with only one optimisation as discussed below. The interested readers that are unfamiliar with the LK algorithm should please read the 1981 paper.

The LK algorithm is a coarse to fine optical flow based image registration algorithm. At each resolution level only sub-pixel movement is expected. The optical flow is tracked using a Newton-Raphson error reduction between the input and a reference image. Seven Newton-Raphson iterations offers good sub-pixel accuracy.

Generating synthetic low resolution images provides one with an accurate ground truth to compare the Lucas-Kanade registration to. A root mean square error (RMSE) of 0.045 pixels per frame was measured in this way. Even when adding uniform noise of 20% and applying a Gaussian filter with a standard deviation of 5 pixels an RMSE of 0.063 pixels per frame was achieved.

Bi-linear interpolation is used in the Newton-Raphson error reduction to calculate a smooth image function for the reference image. The bi-linear filter is one of the most expensive operations of the algorithm.

To optimise the image registration implementation the bilinear interpolation of the reference image is executed on the fly, but only for areas of approximately constant gradient. The contribution weight of the other areas to the displacement vector are simply set to zero.

VI. Gauging the Resolution Improvement

The super-resolved edge spread function of the SR image was calculated for each stacking operation. The Fourier transform of an edge spread function (of what should be an impulse edge) gives the SFR of the stacking operation and FPA. The SFR of the stacking operation and FPA along with the pixel resolution of the SR image may then be used to gauge the resolution improvement due to a super-resolution algorithm.

Figure 8 shows the measured frequency response of the input and super-resolved images. All plots except for the average stack SFR are typical of a box filter's response [9]. Notice that the frequency response of the sharp stack is the same as the frequency response F_{pixel} of the low res image. Also notice that the frequency response of the average stack is less than that of the sharp stack and is in fact ${F_{pixel}}^2$.

The Double Low Resolution (DLR) SFR shown in Figure 8 is given as a reference of what the SFR of the system would look like if the resolution of the FPA could be physically doubled. Without any processing the DLR system would provide significant modulation up to 1 cycle/LRPixel and then suffer aliasing problems.

Lin and Shum [10] found that in practice the resolution increase factor of spatial reconstruction based algorithms is typically limited to 1.6. They state that the limited resolution increase factor is mainly due to typical registration inaccuracies and noise in the LR images. However, the SFR of the FPA and the stacking operations show that resolution limit is much more fundamental. Above 1.5 cycles per pixel the modulation transfer reduces sharply. It is in theory possible to also recover information around 2.5 cycles/pixel, 3.5 cycles/pixel, etc., but probably difficult in practice due to the low transfer.

It is important to note that one cannot recover spatial frequencies that are suppressed due to the FPA's SFR. The recommended measure of performance of the super-resolution

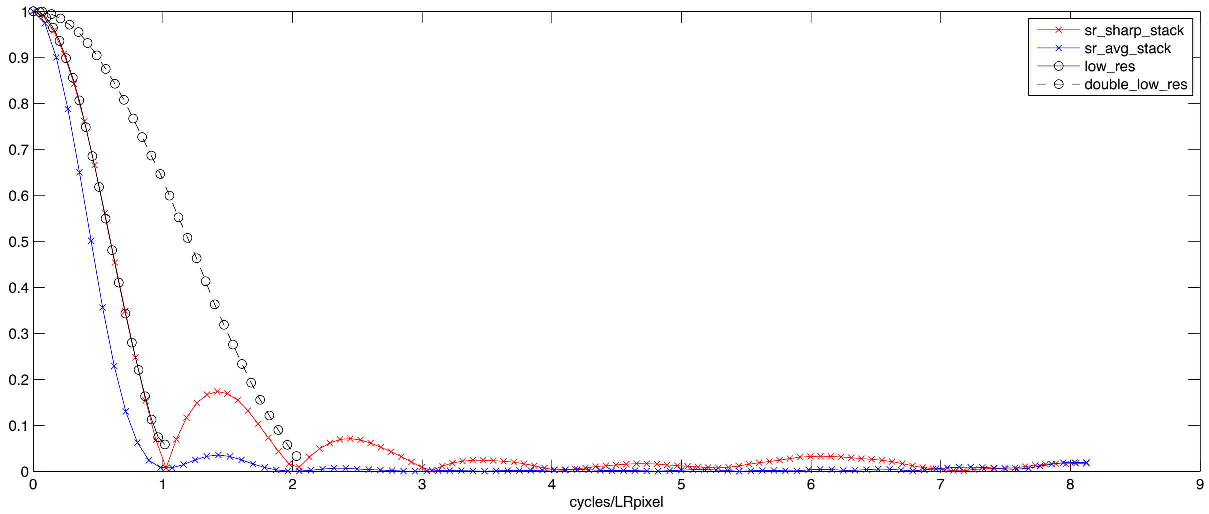

Fig. 8. The System Frequency of the Input and Super-Resolved Images (see the figure key)

algorithm is therefore how close the SR image's *measured* SFR is to the super-resolved SFR of the LR input image. In other words how much the system's frequency response deteriorates when super-resolving the image.

VII. Results

The registration is implemented in C++ and quick enough to run at 20 fps on a low power dual-core Intel i5 NUC running at 1.8 GHz for images up to 512x512 pixels. One processor core is used for the image registration and the other for video input, stacking and video output which then lags by one frame.

The sharp stack results are compared to the SupReMe reference solution developed at Stellenbosch University by Stefan van der Walt [11]. SupReMe was chosen as a reference because the source code is available for running on one's own data. Figure 9 shows the results of the sharp stack and the SupReMe reference for the USAF1951 test target. Figure 10 shows the Supreme reference results for video recorded in Simonstown, South Africa. The sharp stack result is comparable to that of SupReMe although the sharp stack executes at 20 fps while the SupReMe implementation takes about 2 minutes per frame.

Figure 11 shows a SR result when the image is degraded by 20% uniform pixel noise and a Gaussian blur with a standard deviation of 0.5 pixels. The system MTF is still good enough to cause aliasing on the FPA which is super-resolved by the sharp stack algorithm.

VIII. Conclusion and Future Work

The paper presents the determined effort to apply a fast image registration algorithm in combination with only an image stacking operation for generating super-resolution stills from video. The sharp stack algorithm is able to keep the frequency response of the system the same as the frequency response of the LR image while lowering the Nyquist rate of the sensor. This is a measurable improvement over the average stack algorithm which lowers the Nyquist rate of the sensor, but degrades the frequency response compared to that of the LR image.

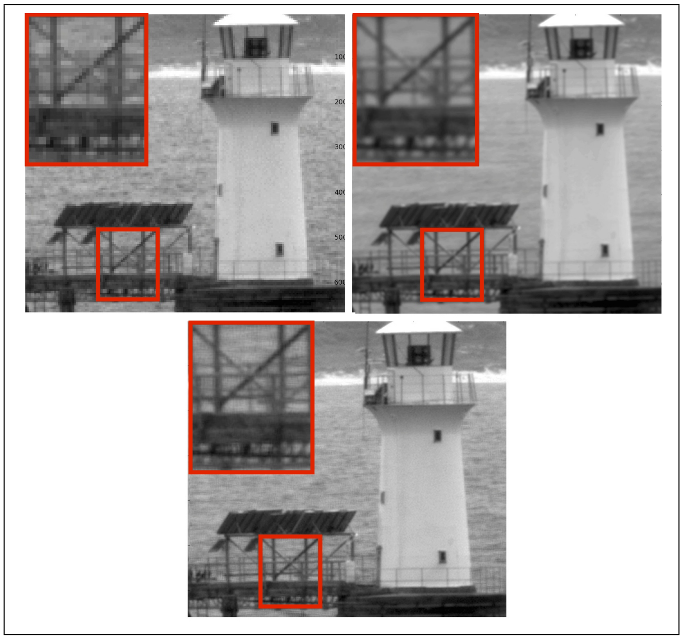

Fig. 10. An LR Image frame from a Video Sequence of Roman Rock in Simonstown is Shown at the Top Next to the Resulting Average Image Stack on the Right and the Supreme SR Solution at the Bottom

The sharp stack super-resolution algorithm lowers the Nyquist rate to one pixel per cycle which is a doubling of the resolving power. Spatial frequencies above 1 cycle per pixel experience a contrast inversion due to the FPAs frequency response, but could potentially in future be super-resolved if the inversion is factored into the problem.

The recommended measure of performance of the super-resolution algorithm is how close the SR image's is to the SFR of the LR input image. In other words how much the SFR deteriorates when super-resolving the image.

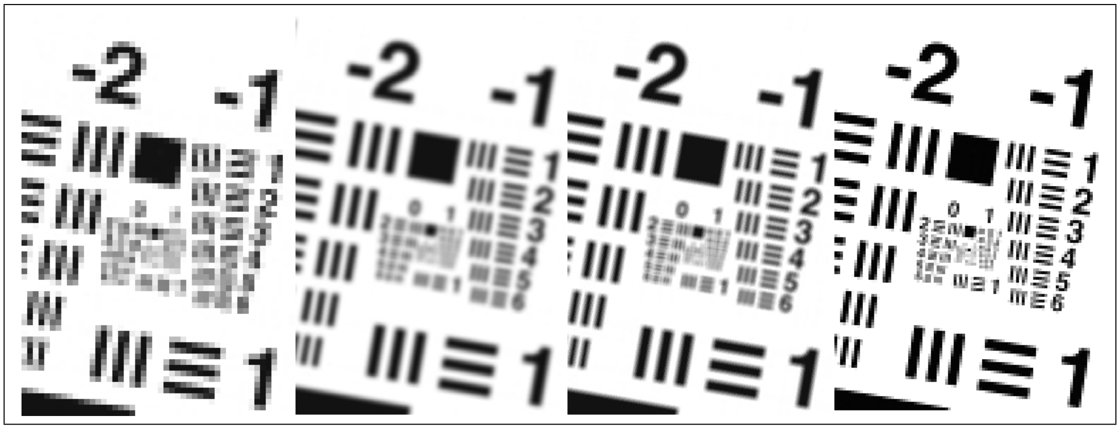

Fig. 9. An LR Image, Average Image, Sharp Stack and Reference Supreme Results

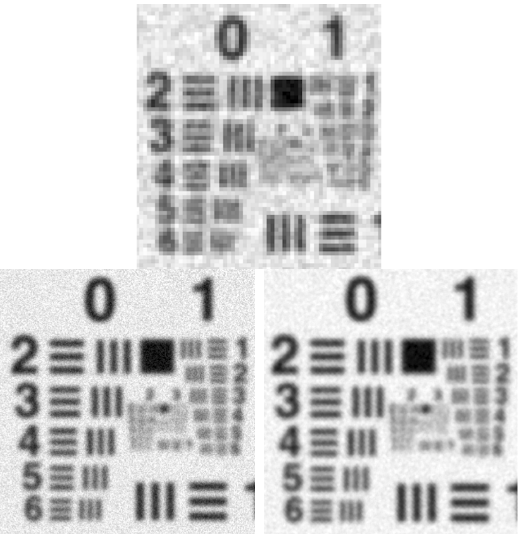

Fig. 11. The Low Resolution Input (top), Nearest Neighbour Sharp Stack (bottom-left) and Kernel Filter Sharp Stack (bottom-right) when adding 20% Noise and a Gaussian Filter with a Standard Deviation of 0.5 Pixels

When a limited number of LR frames are available then a nearest neighbour filter or kernel based filter may be used to create a dense SR result. The nearest neighbour and kernel filters have similar quality performance in the limit. The nearest neighbour method is however simpler, faster and has improved quality for a small number of inputs. On the other hand when the input is very noisy then the kernel method offers an advantage due to its temporal smoothing behaviour.

In general the registration RMSE seems to be well below 0.1 pixels per frame. In addition the systematic error (i.e. drift over multiple frames) was measured to be in the order of 0.002 pixels per frame.

This paper shows super-resolution of wide camera fields of view where the effect of atmospheric bubbling and dancing due to scintillation is not evident. When atmospheric effects start playing a major role such as in long range image enhancement then the rigid per frame image registration should to be replaced with a per-pixel optical flow technique. The LK algorithm could still be employed, but must be combined with an efficient regularisation of the resulting sparse flow fields.

Acknowledgment

The author would like to thank both the Armaments Corporation (Armscor) of South-Africa and the CSIR for supporting this research.

References

[1] R. Vollmerhausen, D. Reago, and R. Driggers, *Analysis and Evaluation of Sampled Imaging Systems*. SPIE Press, 2010.

[2] D. Keren, S. Peleg, and R. Brada, “Image sequence enhancement using sub-pixel displacements,” in *Computer Vision and Pattern Recognition, 1988. Proceedings CVPR ’88., Computer Society Conference on*, 1988.

[3] M. Irani and S. Peleg, “Improving resolution by image registration,” *CVGIP: Graph. Models Image Process.*, 1991.

[4] R. R. Schultz and R. L. Stevenson, “Extraction of high-resolution frames from video sequences,” *Image Processing, IEEE Transactions on*, 1996.

[5] A. Zomet, A. Rav-Acha, and S. Peleg, “Robust super-resolution,” in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, 2001.

[6] X. Li, Y. Hu, X. Gao, D. Tao, and B. Ning, “A multi-frame image super-resolution method,” *Signal Processing*, 2010.

[7] Q. He, R. R. Schultz, and C. H. Chu, “Efficient super-resolution image reconstruction applied to surveillance video captured by small unmanned aircraft systems,” in *Signal Processing, Sensor Fusion, and Target Recognition XVII*, 2008.

[8] B. D. Lucas and T. Kanade, “An iterative image registration technique with an application to stereo vision,” in *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2*, ser. IJCAI’81. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1981, pp. 674–679. [Online]. Available: http://dl.acm.org/citation.cfm?id=1623264.1623280

[9] R. Fisher, S. Perkins, A. Walker, and E. Wolfart, *Hypermedia Image Processing Reference*. Wiley, 1997.

[10] Z. Lin and H. Shum, “Fundamental limits of reconstruction-based superresolution algorithms under local translation,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2004.

[11] S. J. van der Walt and B. Herbst, “A polygon-based interpolation operator for super-resolution imaging,” *arXiv preprint arXiv:1210.3404*, 2012.