Embedded video rate super-resolution in the infrared with a low-cost multi-aperture camera

Roi Méndez-Rial, Álvaro Souto-López, Antón García-Díaz

Aimen Technology Centre, Porriño, Spain;

ABSTRACT

We present an embedded imaging approach based on low cost sensors that span a long spectral range in the infrared. A system has been implemented with 12 apertures that combine unique uncooled FPAs in the mid infrared domain -2 to 5 microns wavelength- with very low cost microbolometers in the thermal infrared -7 to 14 microns wavelength-. Both FPA technologies are uncooled and low cost, manufactured as monolithic devices. The system is made of two modules, one LWIR, other MWIR. Each module has a system-on-chip GPU/ARM board that carries out all the image processing required for image reconstruction. This includes the calibration of the system, the registration of the images acquired with the many apertures, and the reconstruction of the super resolved image. Besides, the board performs all the operations and transformations required for noise correction. The output of each of the modules is a video stream at 30 frames per second. Each frame is a super resolved image with a resolution 2.5x compared to the images acquired by the FPAs used. Furthermore, the modules may be integrated and the acquired images combined in a single one in the embedded processing boards. Moreover, the boards may also combine and fuse this output with a visible range video stream. The use of low cost FPAs facilitates the deployment in a broad range of applications that an benefit from imaging in the infrared, particularly in the MWIR range in which existing commercial cameras based on hybrid technology are very expensive. The system is being tested in different applications, including surveillance in variable lighting conditions and monitoring in firefighting scenarios.

Keywords: Super-resolution, multi-aperture, mid-wavelength infrared, long-wavelength infrared

1. INTRODUCTION

Thermal imaging has been growing fast and it is playing an important role in a wide range of civil applications thanks to the reduction of costs and evolution of the uncooled technology. Microbolometers are a mature technology for the long-wavelength infrared (LWIR) range from 7 to $14\mu m$. The microbolometer market offers from cheap and miniaturized focal plane arrays (FPAs) with reduced resolution 80×80, to the most advanced ones with resolutions up to 1024×768. In the mid-wavelength infrared (MWIR) range from 2 to $5\mu m$, there is a lack of uncooled detectors, with the low resolution PbSe imagers[1] (32×32 and 128×128) the only solution available in the market to the best of the author's knowledge. The lenses play a major role on the price and weight of the thermal cameras. The manufacturing costs are reducing, but there is still a tradeoff cost/performance that has not yet open the way of the uncooled IR technology to consumer applications.

The general aim of multi-aperture imaging design is the substitution of a single lens aperture by an array of lenses with shorter focal length,[2] reducing the requirements of objective lens (cheaper optics), allowing the capture of depth information, achieving better color separation (less crosstalk) and providing scalable resolution. A multi-aperture array produces a set of low-resolution images that may be computationally combined to produce a single high-resolution image with super-resolution algorithms.

Here, we describe the embedded implementation of a super-resolution method for a multi-aperture camera array in a broad IR range using low cost and low resolution FPAs. The super-resolution techniques is based on maximum a posteriori (MAP) approach using a regularized image prior. We achieve super-resolution at video rate for the LWIR and MWIR range using a camera array with six apertures in each band. Combining six low resolution images in the LWIR of 160×120 pixels, the embedded system is able to deliver higher resolution 400×300 images at video rate. In the MWIR, the system increase the native resolution of the PbSe FPAs from 32×32 to 76×76 pixels. Besides, the computational imaging technique allows refocusing of the scene.

Figure 1: Broad band multi-aperture snapshot camera in the IR (0.4-14μm). Left: MWIR module, centre Visible and NIR, right LWIR.

Table 1: Coarse bands

Coarse band	Wavelength (μm)	Sensor technology
VIS/NIR	0.4 to 1.0	CMOS
MWIR	2.0 to 5.0	PbSe
LWIR	7.0 to 14.0	MEMS microbolometer

2. BROAD BAND MULTI-APERTURE CAMERA

We have designed and assembled a broad band multi-spectral camera for the IR range (0.4-14μm). The camera combines the only three IR technologies that are monolithically compatible with Si-CMOS: PbSe detectors for the MWIR range[3],[1] CMOS-based microbolometers[4] for the LWIR, and a standard CMOS detector for the visible and NIR (Table-1). The camera has three main modules: a multi-aperture (MA) MWIR array composed of six low resolution (32×32) PbSe detectors, a MA LWIR array composed of six QQVGA (160×120) microbolometers, and a CMOS VIS/NIR module with one aperture and a beamsplitter (each band with a resolution of 1088×2048). Each coarse band module has a system-on-chip ARM/GPU board (Jetson TX1) for synchronized acquisition, pre-processing, image reconstruction based on computational imaging and band fusion. Each module can work as a standalone unit, providing a control interface and video stream in standard format, or be synchronized to provide a unique fused video stream. The interfaces are based on USB3 for the VIS/NIR and LWIR module and USB2 for MWIR (see Figure 2). The camera provides multispectral video stream embedded in a mosaic in standard video format at 25 frames per second (Fps). One of the key features of the SoC board is the shared memory between the CPU and GPU, which makes it an efficient processor for streaming applications with low latency.

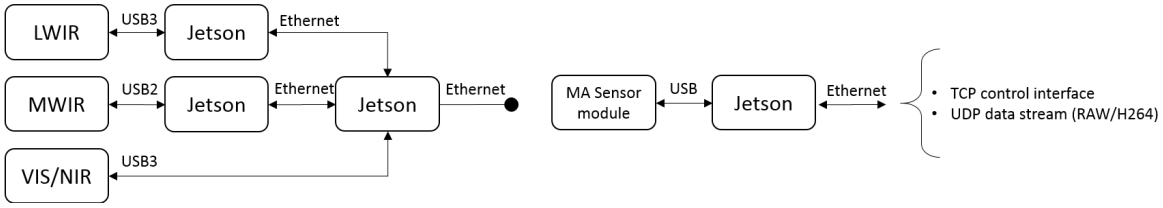

Figure 2: Communication scheme between the broad band modules (left). Single sensor module interface (right).

3. SUPER-RESOLUTION

Super-resolution is a procedure that combines noisy and low resolution images to provide a super-resolved image with higher spatial resolution and improved image quality. In this section describe the implementation of the embedded super-resolution approach for the LWIR and MWIR ranges providing video-rate. The method combines six low resolution images

Email: maximino.mendez@aimen.es.

simultaneously acquired with the camera array to generate a super-resolved image with higher resolution (up to $2.5\times$) with respect to the native resolution of the FPAs.

3.1 System model

Consider a high-resolution image scene $\boldsymbol{x} \in \mathbf{R}^{N\times 1}$ in lexicographical order, and a set of K low resolution images $\boldsymbol{y}_i \in \mathbf{R}^{M\times 1}$ with $i \in [1, K]$. The warping, blurring and downsampling of the scene is model by a sparse matrix $\boldsymbol{W}^i \in \boldsymbol{R}^{M\times N}$. The simplest forward model for one low resolution image generation is

$$\boldsymbol{y}_i = \boldsymbol{W}^i \boldsymbol{x} + \boldsymbol{e}_i, \tag{1}$$

where $\boldsymbol{e}_i$ is assumed to be zero-mean Gaussian Noise. Concatenating all the low resolution images in a vector $\boldsymbol{y} = [\boldsymbol{y}_1^{\mathrm{T}}, \ldots, \boldsymbol{y}_K^{\mathrm{T}}]^{\mathrm{T}}$, and defining the system matrix as

$$\boldsymbol{W} = \begin{bmatrix} \boldsymbol{W}^1 \\ \boldsymbol{W}^2 \\ \vdots \\ \boldsymbol{W}^K \end{bmatrix}, \tag{2}$$

we obtain the image retrieval model

$$\boldsymbol{y} = \boldsymbol{W}\boldsymbol{x} + \boldsymbol{e}. \tag{3}$$

The objective of the super-resolution is to recover $\boldsymbol{x}$ from $\boldsymbol{y}$ without knowing $\boldsymbol{W}$ and $\boldsymbol{e}$. Several approaches have been proposed to simultaneously recover $\boldsymbol{y}$ and $\boldsymbol{W}$.[5] To simplify the reconstruction, however, the problem is usually split in two steps: first construct the warping matrix $\boldsymbol{W}$, and second recover the high resolution image $\boldsymbol{x}$, which is the strategy we follow here. The warping matrces $\boldsymbol{W}^i$ are parametrized by few values using motion models with different degrees of freedom and point spread functions (PSF).

We define the high resolution image plane, as a plane parallel to one of the low resolution images taken as reference, at a distance d, Figure 3. Thus, the geometric relationship between the location of a pixel in the low resolution (LR) image $\boldsymbol{u}$, and its location in the high resolution (HR) image $\hat{\boldsymbol{u}}$, may be written in homogeneous coordinates as

$$\hat{\boldsymbol{u}} = \boldsymbol{H}\boldsymbol{u}, \tag{4}$$

where $\boldsymbol{H}$ is the homography matrix that relates both image planes. The transformation matrices $\boldsymbol{H}^i$ with $i \in [1, \ldots, K]$, for each of the LR images are obtained from a calibration and registration procedure. Assumning an isotropic Gaussian PSF each element of the i^{th} warping matrix is defined as

$$\boldsymbol{W}^i_{n,m} = \exp\left(\frac{-||(\boldsymbol{v}_n - \hat{\boldsymbol{u}}^i_m)||_2^2}{2\sigma^2_{psf}}\right), \tag{5}$$

with $\boldsymbol{v}_n$ the location of the n^{th} pixel in the HR image, and $\hat{\boldsymbol{u}}^i_m$ the location of the m^{th} LR pixel from the i^{th} image when projected in the HR image.

3.2 Image registration

The warping matrices $\boldsymbol{W}^i$ are constructed from the transformation matrices $\mathbf{H}^i$ obtained from a registration procedure. Registration in thermal infrared is challenging due to the lack of definition of corners and other features like texture. Methods such as Shift or Surf usually fail. Simple motion models between images where the registration only allows translations and rotations are preferred[5,6].

To simplify the registration problem and reduce the degrees of freedom of the motion model, we exploit the geometry of the camera array. We use a full metric calibration to obtain the relative position and orientation of each camera with respect to the reference camera frame (Figure 3). From the intrinsic and extrinsic parameters we compute the homographies $\boldsymbol{H}^i = \boldsymbol{H}_1(d)\boldsymbol{H}_i$ that transform each LR image to the super-resolved reference image plane placed at a distance d. $\boldsymbol{H}_1(d)$ is the homography matrix that transform the image plane of the reference camera to the super-resolved image plane, and $\boldsymbol{H}_i$ with $i \in [2, K]$ the homography matrices between the other cameras and reference coordinate frame.

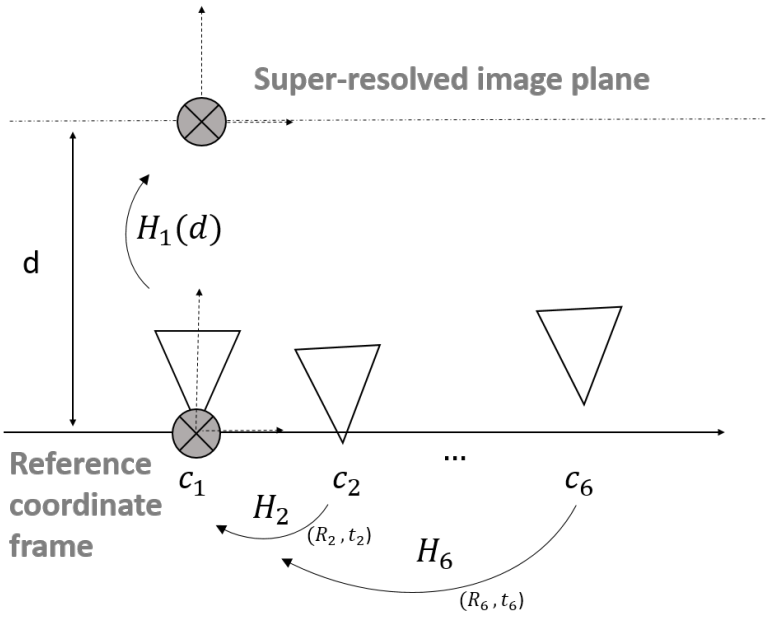

Figure 3: Camera array reference plane and super-resolved image plane. Each LR image is projected onto the super-resolved image plane using the registration information $(\boldsymbol{R}_i, \boldsymbol{T}_i)$ and the distance between the two planes d. The information is embedded in each homography matrix $\boldsymbol{H}_i$.

The homographies $\boldsymbol{H}_i$ are computed from the intrinsic and extrinsic parameters estimated from the calibration method $\boldsymbol{H}_i = \boldsymbol{f}(\text{camera matrix}, \boldsymbol{R}_i, \boldsymbol{t}_i)$.

This approach allow us to focus the HR image at an arbitrary distance varying d. It is usually more interesting to select the region in the image we want to be focused. For that, we follow a different strategy similar to.[7] We first re-project all the LR images to a reference plane at distance d_0 in the far field, removing in this way all the effects relatives to rotation and translations of the cameras. Assuming that the images are now aligned on a plane, we apply a subpixel-registration method on the region of interest, allowing only vertical and horizontal shifts between the images. The homographies are given by $\boldsymbol{H}^i = \boldsymbol{H}_{\text{offset}_i} \boldsymbol{H}_1(d_0) \boldsymbol{H}_i$, with $\boldsymbol{H}(\text{offset}_i)$ an homography matrix that applies the estimated shifts in the vertical and horizontal direction to the projected LR image i.

The full metric calibration procedure involves the acquisition of many images using a calibration pattern illuminated with a thermal heater placed at different distances and orientations. For the subpixel registration based on correlations we use the efficient algorithm in.[8]

3.3 Maximum-a-posteriori image reconstruction

To reconstruct the super-resolved image we use a maximum a posteriori (MAP) model

$$\hat{\boldsymbol{x}} = \arg min \left\{ ||\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}||_2^2 + \lambda \boldsymbol{U}(\boldsymbol{x}) \right\}. \tag{6}$$

The first term is a data fidelity item between the observed LR images and the HR image, while $\boldsymbol{U}(\boldsymbol{x})$ is an image prior that promote smooth solutions, with a control regularization parameter λ which represents a trade-off between data fidelity and smoothness. We use a Huber prior widely extended in super-resolution approaches, which benefits from penalizing edges less severely than other Gaussian or Laplacians priors[9].[10] Solving $\boldsymbol{x}$ can be a time-consuming process if the $\boldsymbol{W}$ matrix is large or have many non-zero elements. Iterative methods are available to obtain a good estimate of $\boldsymbol{x}$. We use here an efficient gradient descent algorithm for large system, the Scaled Conjugate Gradient (SCG) in.[11]

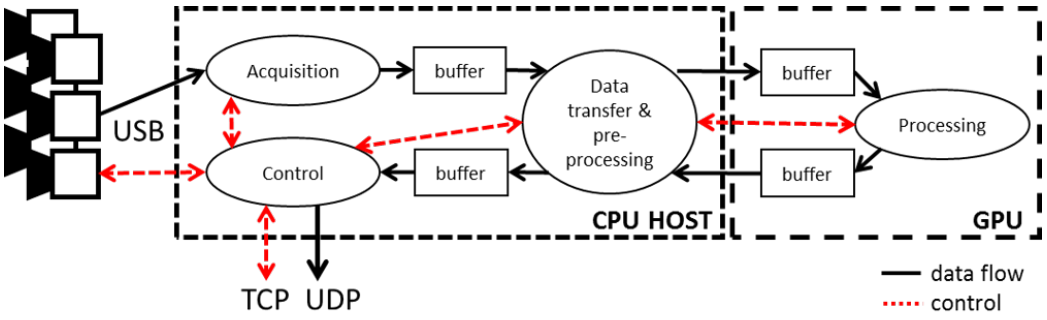

Figure 4: CPU/GPU software architecture of a MA camera module.

3.4 Implementation

To achieve super-resolution at video rate we have implemented the reconstruction algorithm exploiting the powerful GPU of the embedded board. The firmware architecture is depicted in Figure 4. The CPU is dedicated to synchronized acquisition, image pre-processing (e.g. non-uniformity correction and photometric correction) and control interface, while the GPU is exclusively devoted to the super-resolution. The SCG requires the computation of matrix-vector multiplications at each iteration which represents the most computational demanding operation. The complexity of matrix-vector multiplication is of order $O(N^2)$ with N the total number of pixels in the super-resolved image. Since $\boldsymbol{W}$ is a sparse matrix, the complexity can be reduced significantly using efficient sparse data schemes, with the complexity depending on the number of non-zero entries. The sparsity level of the system matrix directly depend on the assumptions made on the generative model and the registration. The matrix vector multiplication has been implemented in CUDA. We exploit the shared memory between CPU and GPU of the Jetson board to reduce memory transfer between host and device. Optimization of the sparse matrix vector product (SpMV) is challenging because of the irregular computation of large spare operations. The effort to accelerate the computation of SpMV is focused on the design of appropriate data formats to store the spare matrix, since the performance of the SpMV is directly related to that. From all the sparse data schemes we have tested, ELLPACK-R[12] provides the best performance.

ELLPACK-R is a variant of the ELLPACK format. The scheme uses three array; a float array $\boldsymbol{A}$ to save the non-zero entires, one integer array $\boldsymbol{j}$ to save the columns of every entry, and another integer array $\boldsymbol{r}$ to store the actual length of each row. $\boldsymbol{A}$ and $\boldsymbol{j}$ of dimension at least $KM \times$ MaxEntreisByRows, and the vector r of dimension KM. We refer the reader to[12] for the GPU implementation details. The scheme takes advantages of:

- Coalesced memory access on the GPU.
- Non-synchronized execution between different blocks of threads.
- The reduction of the waiting time or unbalance between threads of one warp.
- Homogeneous computing within the threads in the warps.

4. RESULTS

We tested the multispectral broad IR camera in the field for traffic monitoring and during a fire fighting drill. Figure 5 and 6 show reconstruction results using the LWIR module on a traffic scene. We show the six low resolution images (160×120) acquiried with the camera array (top), a interpolated LR image (bottom-left), and the super-resolved image (bottom-right). The low resolution images have good contrast but are noisy, due to the non-uniformities of the FPA. The super-resolved image (400×300) shows better image quality, mitigating the noise and providing much better definition than the median image.

The processing time for super-resolved an image is $\sim$ 42 milliseconds in the embedded system with the CPU/GPU implementation, achieving a super-resolved video stream of 24 frames per second.

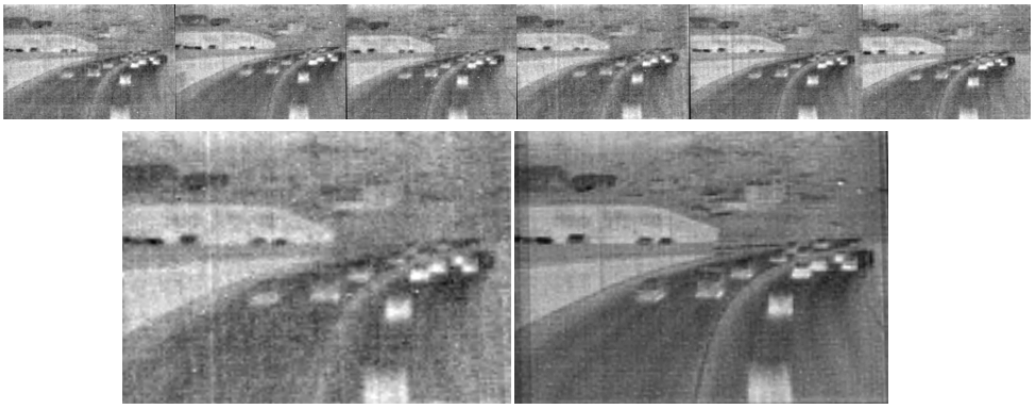

Figure 5: LWIR super-resolution results. Top: six low resolution images 160×120. Bottom: LR interpolated image (left) and super-resolved (right) with increased resolution 400×300.

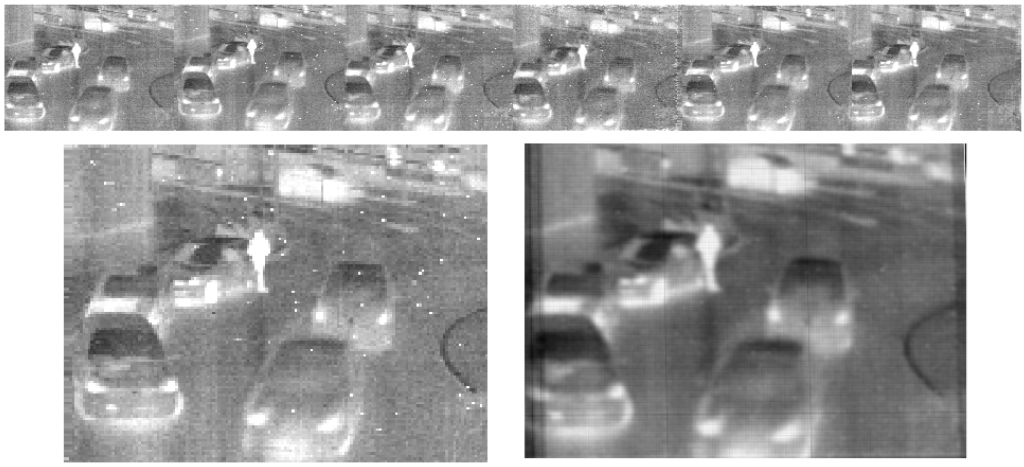

Figure 6: LWIR super-resolution results. Top: six low resolution images 160×120. Bottom: LR interpolated image (left) and super-resolved (right) with increased resolution 400×300.

In Figure 7, we show reconstruction results using the MWIR module with an image taken during a fire-fighting drill. The low resolution MWIR images (32×32) show a good SNR. The super-resolved images provides much better definition than the interpolated median image.

5. CONCLUSION

We have described the embedded implementation of a super-resolution approach for a multi-aperture camera covering a broad band IR spectrum band. The algorithms is implemented in CUDA, and runs in a powerful embedded processor integrated in the camera. The embedded system is able to provide super-resolution in LWIR and MWIR at video rate increasing the native resolution of the low cost FPAs up to a factor of x2.5.

ACKNOWLEDGMENTS

This work was supported by the European Unions Horizon 2020 research and innovation programme through SEERS project under grant agreement No 645114, and by the Spanish Ministry of Economy, Industry and Competitiveness under grant PTQ-15-07702.

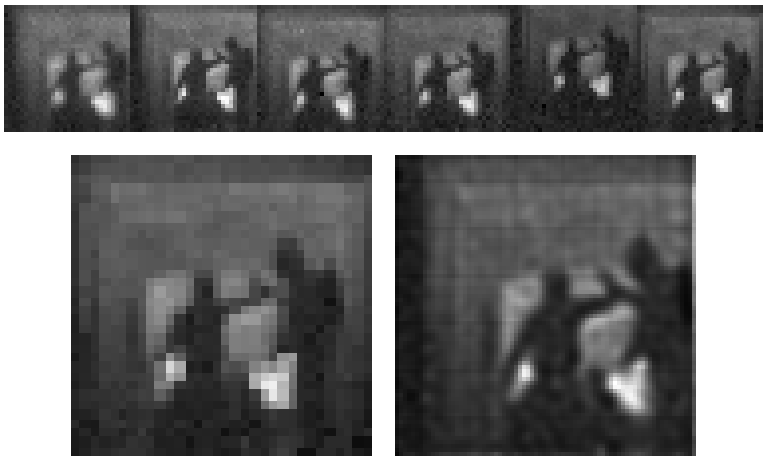

Figure 7: MWIR super-resolution results. Top: six low resolution images 32×32. Bottom: interpolated median image (left) and super-resolved (right) with increased resolution 76×76.

REFERENCES

[1] Vergara, G., Linares-Herrero, R., Gutirrez-lvarez, R., Montojo, M. T., Fernndez-Montojo, C., Baldasano-Ramrez, A., and Fernndez-Berzosa, G., "VPD PbSe technology fills the existing gap in uncooled low-cost and fast IR imagers," 80121Q (May 2011).

[2] Carles, G., Muyo, G., Bustin, N., Wood, A., Downing, J., and Harvey, A. R., "Multi-Aperture Imaging in the Visible and Thermal Infrared," in [*Computational Optical Sensing and Imaging*], CM3C–3, Optical Society of America (2013).

[3] Vergara, G., Gomez, L. J., Villamayor, V., Alvarez, M., Rodrigo, M. T., Torquemada, M. d. C., Sanchez, F. J., Verdu, M., Diezhandino, J., Rodriguez, P., Catalan, I., Almazan, R., Plaza, J., and Montojo, M. T., "Progress on uncooled PbSe detectors for low-cost applications," 279 (Aug. 2004).

[4] Tezcan, D. S., Eminoglu, S., and Akin, T., "A low-cost uncooled infrared microbolometer detector in standard CMOS technology," *IEEE Transactions on Electron Devices* **50**, 494–502 (Feb. 2003).

[5] Hardie, R. C., Barnard, K. J., and Armstrong, E. E., "Joint MAP registration and high-resolution image estimation using a sequence of undersampled images," *IEEE transactions on Image Processing* **6**(12), 1621–1633 (1997).

[6] Pickup, L. C., *Machine Learning in Multi-frame Image Super-resolution*, PhD thesis, University of Oxford (Feb. 2008).

[7] Vaish, V., Wilburn, B., Joshi, N., and Levoy, M., "Using plane+ parallax for calibrating dense camera arrays," in [*Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*], **1**, I–I, IEEE (2004).

[8] Guizar-Sicairos, M., Thurman, S. T., and Fienup, J. R., "Efficient subpixel image registration algorithms," *Optics letters* **33**(2), 156–158 (2008).

[9] Pickup, L. C., Capel, D. P., Roberts, S. J., and Zisserman, A., "Bayesian methods for image super-resolution," *The Computer Journal* **52**(1), 101–113 (2007).

[10] Zhang, L., Yuan, Q., Shen, H., and Li, P., "Multiframe image super-resolution adapted with local spatial information," *JOSA A* **28**(3), 381–390 (2011).

[11] Mller, M. F., "A scaled conjugate gradient algorithm for fast supervised learning," *Neural networks* **6**(4), 525–533 (1993).

[12] Vazquez, F., Garzon, E. M., Martinez, J. A., and Fernandez, J. J., "The sparse matrix vector product on GPUs," in [*Proceedings of the 2009 International Conference on Computational and Mathematical Methods in Science and Engineering*], **2**, 1081–1092 (2009).