Thermal Image Super Resolution via Compressed Sensing

Epameinondas Th. Rontogiannis[1]

Abstract— **This thesis deals with the problem of resolution enhancement (*super resolution*) of thermal images using compressed sensing methods. We solve the super resolution problem in four stages. First, we seek a sparse representation of a low-resolution image with respect to two statistically-learned overcomplete dictionaries (for high and low resolution images respectively) and then we use the coefficients of this representation to calculate the high resolution image. Then, we calculate the high resolution image using methods requiring multiple low resolution images aligned with subpixel accuracy (conventional approach). We compare the results of each method using broadly acclaimed metrics regarding reconstruction quality standards.**

Index Terms— **super resolution, compressed sensing, thermal imaging, sparse representations**

I. INTRODUCTION

Nowadays, lots of applications require high resolution (HR) images, since higher resolution means more detail is included in an image. Resolution describes the level of details contained in an image and can be a critical feature of thermal images, as it enables us to locate energy losses of buildings or identify tactical targets in the darkest of nights. Thermal images are captured using special equipment that senses the infrared range ($9-14\mu m$) of the electromagnetic spectrum. They can be used in many applications such as UAV surveillance for automatic target detection, rescuing operations for locating people in danger in low-light conditions and building diagnostics for detection of energy losses and leakage.

Thermal cameras may be quite similar to optical cameras, however they do have significant differences in the way their sensors are built and their lenses act as natural polarization filters, allowing the sensors to capture the emitted thermal radiation. The materials used in manufacturing of thermal sensors do not allow integration on a very large scale (resulting in a small chip) while giving high resolution. This concept may be common with CCD and CMOS sensors but here, instead of silicon, we use elements such as InSb, InGaAs, HgCdTe which have the ability to respond to longer wavelengths. A common resolution for a thermal camera may be around 320×240 pixels, while cameras offering slightly higher resolution have a dramatically higher cost. The aforementioned resolution is not at all adequate for today's applications (some of them requiring ultra high definition (4K) images) and is significantly lower than the average resolution offered by cell phone cameras manufactured in the previous decade.

The concept of resolution enhancement is not something new and is used in many cases when it is naturally impossible to achieve better resolution through the available hardware. Higher resolution images are estimated from available low resolution images by aligning the low res. images with subpixel accuracy and then combining all the details they contain. The first methods were published around 1990 and since then they have multiplied and evolved significantly. More information about such methods can be found in [1], while we can distinguish 4 major approaches: single-frame super resolution, multiple-frame super resolution, methods operating on spatial domain, and methods operating on frequency domain.

The first paper about super resolution methods was published in 1984 by Tsai and Huang [7] and the proposed method (operating on frequency domain) exploited aliasing and the shifting property of continuous and discrete Fourier transform. Later, this method was altered [8] in order to deal with noise and blurring using recursive least squares (RLS) algorithm.

The first method operating on spatial domain used an Iterative Back-Projection algorithm [2]. More advanced methods on this domain use stochastic models [3] and the theory of Projection Onto Convex Sets (POCS) [4].

Whilst the field of super resolution methods on optical images has been thourougly researched on, there is surprisingly little progress made on thermal image super resolution. Most of the available methods follow the multiple frame approach and two notable papers using this approach are [5] and [6].

Our proposed method for super resolution on thermal images is motivated by recent work on optical images [9], following the single-frame super resolution approach and exploiting the theory of compressed sensing. A significant advantage of this method, compared to conventional ones, is that only one low resolution image is required to generate the high resolution output. Compressed sensing methods require significantly fewer samples than those defined by the sampling theorem, known as *Shannon-Nyquist*. To make signal reconstruction possible, we try to express a signal in a base using a small number of coefficients. In this thesis we use two overcomplete and coupled dictionaries as bases.

In the next section, we describe both the theoretical background of *compressed sensing* and the shift-add fusion method. In section III, we describe our proposed method for super resolution on thermal images. Next,in section IV, we evalutate the results generated by the two methods which helps us conclude (section V) and discuss the effectiveness

[1]Epameinondas Rontogiannis is currently studying at Dept. of Electrical and Computer Engineering at University of Patras. He is with the laboratory of Digital Signal and Image Processing and his main academic interests include applications of sparse representations on signal processing, image processing and real-time signal processing. e-mail: eparon at outlook.com

of each method.

In the remainder of this thesis we use the following notation: $\boldsymbol{X}$ denotes a high resolution image and $\boldsymbol{x}$ its patches, $\boldsymbol{Y}$ denotes a low resolution image and $\boldsymbol{y}$ its patches, $\boldsymbol{Y}_i$ denotes a low resolution image set, $\boldsymbol{D}$ denotes dictionaries used for sparse representations, lowercase bold letters denote vectors, uppercase bold letters denote arrays and plain lowercase letters denote scalars.

II. OVERVIEW OF THE PROBLEM

In this section we present a reconstruction method called *shift-add fusion*, generating a high resolution image using a low resolution image set of the same scene. Next, we give an introduction to the theory of compressed sensing, and continue with the presentation of our proposed method in section III.

A. Shift-add fusion method/Multiple-Image SR

Multiple-image super resolution methods generate a high resolution image using a set of low resolution images of the same scene.

When capturing images with camera, we assume that there is usually motion (either due to camera motion or motion of the subject) between any pair of the observation set and the image is blurred in some way as it passes through the lenses of the camera system. The final recorded images are sampled at a relatively low spatial frequency and we also assume the pressence of additive noise, corrupting the image observation. The aforementioned *observation model* can be summarized for each observation using three linear operators and the addition of a noise term. Supposing that a high resolution image $\boldsymbol{X}$ (which we are going to estimate) of the scene exists, it is typical to model the low resolution image set as:

$$\boldsymbol{Y}_i = \boldsymbol{S}_i \boldsymbol{T}_i \boldsymbol{H}_i \boldsymbol{X} + n_i \tag{1}$$

where we define for the i-th low resolution image $\boldsymbol{Y}_i$ the following operators applied on $\boldsymbol{X}$: $\boldsymbol{S}_i$ denotes undersampling, $\boldsymbol{T}_i$ denotes the geometric transformation relative to the 1st image of the set, $\boldsymbol{H}_i$ denotes the effect of blurring and n_i denotes the additive noise.

For this observation model, we make the following assumptions:

- the effect of blurring is spatially-invariant across the image plane and the blurring kernel is known to the algorithm,
- the noise is white Gaussian with the same variance for all images in the set,
- the geometric transformation of each image (also refered to as *motion estimates*) is constained to global translation

When applying the shifts suggested by $\boldsymbol{T}_i$ to $\boldsymbol{X}$, for each pixel shift on the high resolution grid, we get a subpixel shift on the low resolution grid. This is happening, as we undersample the high resolution image. To make the concept of subpixel shifts more clear, let us consider a random pixel on the high resolution grid, (x,y). If we try to move 1 pixel left on the high resolution grid, like $(x+1,y)$, the observed shift on the low resolution grid is divided by the factor of downsampling, giving a subpixel shift of $(x+\frac{1}{S},y)$.

The theoretical basis of super resolution is best explained in the frequency domain [7], however there are lots of methods operating on the spatial domain, featuring significantly lower computational complexity.

The concept of Shift-add fusion relies on a simple fact: By knowing the motion estimates for a low resolution image set, we are able to calculate the original location on the high resolution grid for every pixel of every low resolution image. Motion estimates are critical to this algorithm, as the quality of the reconstructed image depends on their accurate estimation. The estimation procedure can be either algorithmic, using methods (e.g. RANCAC [23]) that are able to detect and calculate such shifts, or we may know them *a priori*.

Relying on the fact that $\boldsymbol{S}_i$ of (1) is invertible, and by knowing the motion estimates $\boldsymbol{T}_i$, we know that the combined operator $\boldsymbol{S}_i\boldsymbol{T}_i$ maps every pixel of each low resolution image to one (and only) specific location on the high resolution grid. This allows us to calculate the high resolution image by putting all pixels of low resolution images into their correct locations on the high resolution grid. The aforementioned procedure is demonstated on Algorithm 1, where we assume the following: $\boldsymbol{dx}$, $\boldsymbol{dy}$ to be the motion estimates for every observation, W, H to be the dimensions of the high resolution image, $\boldsymbol{Y}_i$ to be the low resolution image set, N to be the number of available low resolution images and $\boldsymbol{S}$ to be the scaling operator. This version of the algorithm does not take into consideration the presence of noise or blurring.

Data: $\boldsymbol{dx}$, $\boldsymbol{dy}$, $\boldsymbol{Y}_i$, N, W, $\boldsymbol{H}$, S
Output: High resolution reconstructed $\boldsymbol{X}$
$\boldsymbol{X} \leftarrow 0$;
for $n \leftarrow 1$ ***to*** N **do**
  $\tilde{\boldsymbol{Y}} \leftarrow \boldsymbol{Y}_i$;
  $\boldsymbol{i} \leftarrow 1 : W/S$;
  $\boldsymbol{j} \leftarrow 1 : H/S$;
  $\boldsymbol{px} = \boldsymbol{i} * S + \boldsymbol{dx}_n$;
  $\boldsymbol{py} = \boldsymbol{j} * S + \boldsymbol{dy}_n$;
  $\boldsymbol{X}_{px,py} = \tilde{\boldsymbol{Y}}_{i,j}$;
end

Algorithm 1: Shift-add fusion reconstruction

B. Compressed sensing method/Single Image SR

The main idea of compressed sensing methods suggests that the linear relationships among high resolution signals can be accurately recovered from their low-dimensional projections [10], [11]. This problem is extremely ill-posed, making accurate recovery nearly impossible. However, we are going to examine under which constraints and conditions we are able to efficiently solve this problem and get accurate results.

Let $\boldsymbol{D} \in \mathbb{R}^{n \times K}$, be an overcomplete dictionary of K atoms ($K > n$) and suppose that a signal $\boldsymbol{x} \in \mathbb{R}^{n \times N}$ can be represented as a sparse linear combination with respect to $\boldsymbol{D}$.

This suggests that the signal $\boldsymbol{x}$ can be expressed as $\boldsymbol{x} = \boldsymbol{D}\boldsymbol{\alpha}_0$ where $\boldsymbol{\alpha}_0 \in \mathbb{R}^K$ is a sparse vector, containing very few ($\ll n$) non-zero entries. Practically, due to undersampling, we observe a small set of measurments $\boldsymbol{y}$ of $\boldsymbol{x}$

$$\boldsymbol{y} \doteq \boldsymbol{L}\boldsymbol{x} = \boldsymbol{L}\boldsymbol{D}\boldsymbol{\alpha}_0$$

where $\boldsymbol{L} \in \mathbb{R}^{k \times n}, k < n$, is a projection matrix. Its application to another matrix it *projects* (keeps) specific elements. In this super resolution context $\boldsymbol{x}$ denotes a high resolution image patch, whereas $\boldsymbol{y}$ denotes its low resolution couter part. The overcompleteness of dictionary $\boldsymbol{D}$ makes the equation $\boldsymbol{x} = \boldsymbol{D}\boldsymbol{\alpha}$ undetermined and the equation $\boldsymbol{y} = \boldsymbol{L}\boldsymbol{D}\boldsymbol{\alpha}$ even more undetermined, for the unknown coefficients of $\boldsymbol{\alpha}$. Under mild conditions, however, the sparsest solution $\boldsymbol{\alpha}_0$ will be unique. Moreover, if $\boldsymbol{D}$ satisfies the near isometric property, then for various matrices $\boldsymbol{L}$ any sparse representation of a high resolution image patch $\boldsymbol{x}$ with respect to $\boldsymbol{D}$, can be (almost) accurately recovered from the low resolution image patch [12].

In this paper, we are not directly computing the sparse representation of the high resolution image patch. Instead, we work with two coupled dictionaries – $\boldsymbol{D}_h$ for high resolution patches and $\boldsymbol{D}_l$ for low resolution patches. The coupling of the dictionaries implies that, their atoms are related in a way that the first atom of the high resolution dictionary is matched with the first atom of the low resolution dictionary and so on. Taking this into consideration, we use the sparse representation of the low resolution image patch in terms of $\boldsymbol{D}_l$ to directly compute the high resolution image patch from $\boldsymbol{D}_h$. The coupling of the dictionaries suggests that, if we have the sparse representation of a low resolution image patch in terms of $\boldsymbol{D}_l$, we can use the same coefficients from $\boldsymbol{D}_h$ in order to recover the high resolution image patch.

The dictionaries used have been trained in a statistical manner [15] and in order to ensure that a high resolution patch will have the same sparse representation with its low resolution patch, all atoms of the dictionaries have been properly normalized.

Last but not least, we should mention that super resolution is not the only problem solved using sparse representations. Other well-known applications of sparse representations in the field of image processing are for *denoising* [13] and *restoration* [13].

III. PROPOSED METHOD

The compressed sensing reconstruction method treats differently thermal and optical images, as it operates on the intensity channel of an image. Raw thermal images can be considered as grayscale images and, as a result, we can directly work on the intensity channel. Optical images, however, consist of three channels *(R,G,B)*. In order to be able to process them, we need to transform them to a suitable color space that gives an intensity-like channel, such as Y,Cb,Cr.

Successfull recovery of the high resolution image using sparse representations, we have to take into consideration the following two constraints:

1) Reconstruction constraint: The estimated high resolution image should be consistent with the low resolution input, with respect to the image observation model. This implies that the low resolution image is considered to be a blurred and downsampled version of the high resolution image $\boldsymbol{X}$, as

$$\boldsymbol{Y} = \boldsymbol{S}\boldsymbol{H}\boldsymbol{X} \tag{2}$$

2) Sparsity prior: The aforementioned constraint is not sufficient to successfully solve the reconstruction problem, as for every low resolution image $\boldsymbol{Y}$, there exists an infinite set of high resolution images $\boldsymbol{X}$ satisfying the constraint defined by (2). In order to further regularize the problem, we demand that the patches $\boldsymbol{x}$ of the high resolution image $\boldsymbol{X}$ can be expressed as a sparse linear combination with respect to the dictionary $\boldsymbol{D}_h$:

$$\boldsymbol{x} \approx \boldsymbol{D}_h\boldsymbol{\alpha}, \text{ where } \boldsymbol{\alpha} \in \mathbb{R}^K \text{ and } \|\boldsymbol{\alpha}\|_0 \ll K \tag{3}$$

The term $\boldsymbol{\alpha}$ for the high resolution image patch will be recovered using the sparse representation of the low resolution image patch, with respect to the low resolution dictionary $\boldsymbol{D}_l$, which is coupled with $\boldsymbol{D}_h$.

The reconstruction process is divided in two steps. First, we seek the sparse representation of all patches of the low resolution image with respect to $\boldsymbol{D}_l$ and recover an estimation of the high resolution image. Next, we further refine the estimated image in order to satisfy (2). Thus, in order to get the final result, first we work *locally* on image patches and then we work *globally* on the entire estimated image. This approach helps us to remove prossible artifacts inserted from the first step, and also make the generated image look more consistent and natural.

A. Working locally (local model)

We begin by splitting the low resolution image $\boldsymbol{Y}$ in (3×3) patches (starting from the upper-left corner) and we require those patches to overlap by 1 pixel in every direction. For each patch, we subtract the mean pixel value from all pixels in order to work on the image's texture rather than on absolute intensities.

For each patch, we seek a sparse representation in terms of $\boldsymbol{D}_l$, which we use to obtain the high resolution patch from $\boldsymbol{D}_h$, as the two dictionaries are coupled. Finding the sparsest representation of $\boldsymbol{y}$, can be expressed as:

$$\min \|\boldsymbol{\alpha}\|_0 \text{ s.t. } \|F\boldsymbol{D}_l\boldsymbol{\alpha} - F\boldsymbol{y}\|_2^2 \leq \varepsilon \tag{4}$$

where F is a linear operator extracting *features* (e.g. edges) of image patches, and we will further analyze its role later in this text.

The problem formulated by (4), is characterized as *NP-hard*. However, if vector $\boldsymbol{\alpha}$ is sufficiently sparse, then we can efficiently recover [17],[18] the coefficients $\boldsymbol{\alpha}$ by minimizing the ℓ_1 norm, as:

$$\min \|\boldsymbol{\alpha}\|_1 \text{ s.t. } \|F\boldsymbol{D}_l\boldsymbol{\alpha} - F\boldsymbol{y}\|_2^2 \leq \varepsilon \tag{5}$$

We can also express this optimization problem using Lagrange multipliers, offering an equivalent formulation:

$$\min_{\alpha} \|FD_l\alpha - Fy\|_2^2 + \lambda\|\alpha\|_1 \quad (6)$$

where λ is used to balance the sparsity of the solution and the fidelity of the approximation to y. This problem is broadly known in the literature as *"the Lasso problem"* [24].

By solving (6) for each patch of the low resolution image, does not guarantee compatibility between adjacent patches. All the patches are processed from left to right and top to bottom [19] and in order to enforce compatibility between adjacent patches we modify (4) in a way that each reconstruction $D_h\alpha$ of a patch y should be consistent with its neighbor patches:

$$\begin{aligned} \min \|\alpha\|_1 \text{ s.t. } & \|FD_l\alpha - Fy\|_2^2 \le \varepsilon_1 \\ & \|PD_h\alpha - w\|_2^2 \le \varepsilon_2 \end{aligned} \quad (7)$$

where, matrix P extracts the overlapping region of current patch and previously reconstructed high resolution image, and w contains the values of the previously reconstructed high-resolution image on the overlap. Thus, equation (7) can be written as

$$\min_{\alpha} \|\tilde{D}\alpha - \tilde{y}\|_2^2 + \lambda\|\alpha\|_1 \quad (8)$$

where $\tilde{D} = \begin{bmatrix} FD_l \\ \beta PD_h \end{bmatrix}$ and $\tilde{y} = \begin{bmatrix} Fy \\ \beta w \end{bmatrix}$, where β controls the tradeoff between matching the low-resolution input and finding a high-resolution patch that is compatible with its neighbors. Considering $\beta = 1$ in this paper, the optimal solution of α^* is given as $x = D_h\alpha^*$.

B. Working globally (global model)

Equations (5) and (7) do not require exact equality between y and $D_l\alpha$ and, additionally, the presence of noise in the high resolution image might result in violation of (2). To eliminate this lack of compatibility between X_0 and X, we project X_0 onto the solution space of (2):

$$X^* = \arg\min_X \|SHX - Y\|_2^2 + c\|X - X_0\|_2^2 \quad (9)$$

Equation (9) can be solved using a *gradient descent* method, which formulates the following iterative method:

$$X_{t+1} = X_t + \nu[H^T S^T (Y - SHX_t) + c(X_t - X_0)] \quad (10)$$

where X_t is the estimation of the high resolution image after the t-th iteration and ν is the step used from gradient descent method. The final result of this procedure is the high resolution image X^* which tries to be as close as possible to the initially estimated high resolution image X_0 while satisfying (2). This concept can be summarized in Algorithm 2.

Data: D_h, D_l, Y
Output: High resolution reconstructed X^*
for *each* 3×3 *patch* $y \in Y$ **do**
  Compute mean pixel value of y;
  Solve $\min_\alpha \|\tilde{D}\alpha - \tilde{y}\|_2^2 + \lambda\|\alpha\|_1$;
  Compute $x = D_h\alpha^*$;
  Compute $x + m$ and store it in X_0;
end
Solve gradient descent problem:
$X_{t+1} = X_t + \nu[H^T S^T (Y - SHX_t) + c(X_t - X_0)]$

Algorithm 2: SR via sparse representations

C. Feature extraction of low resolution image patches

In equation (4), we use operator F to ensure that the content of the high resolution image reflects as much as possible the content of low resolution image. Usually, F is chosen as a highpass filter, as high frequencies contribute significantly to image details which are crucial to the reconstruction procedure and its output result.

In the literature, we can find various suggestions regarding F, among them: [20], where F is selected to be a highpass filter to extract edge information, [21], where Gaussian derivative filters are selected for extraction of contours and [22], where first and second order gradients are applied in order to extract edge information. In this paper, we also follow the first and second order gradient approach, due to their simplicity and effectiveness. The filters we use to extract the derivatives are:

$$\begin{aligned} f_1 &= [-1, 0, 1], & f_2 &= f_1^T \\ f_3 &= [1, 0, -2, 0, 1], & f_4 &= f_3^T \end{aligned}$$

Applying these filters to each low resolution patch, we get four vectors which are concatenated to become the feature representation of the patch.

The results of this procedure are essential for dictionary training and sparse representation of patces with respect to the dictionaries used.

IV. EVALUATION OF METHODS

For the sake of completeness, we consider the dictionaries D_h, D_l as known and given quantities in the rest of this paper. More information about the creation and training of these dictionaries can be fount at [16], however, in this paper, we do not deal with the creation method nor will we compare them with other dictionaries in terms of their features. All comparisons will be made with respect to the results given (high resolution images) of the reconstruction method using each dictionary.

A. Compressed Sensing method

We run the super resolution method via compressed sensing using four different dictionaries that have been built either from thermal or from optical images, containing either 512 or 1024 atoms and are used in applications where low resolution images have been undersampled by a factor of

2. We also run the method of super resolution via image registration for different levels of *confidence* during the calculation of motion estimates using *feature points* (RANSAC method) and we compare obtained results. The metrics we are going to use for comparison are *PSNR (Peak Signal-to-Noise Ratio)* and *MSE (Mean Squared Error)* which constitute broadly acclaimed methods in quality measurment for signal reconstruction methods.

We define MSE between an image I and its approximated version K, as:

$$MSE = \frac{1}{m \cdot n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - K(i,j)]^2 \qquad (11)$$

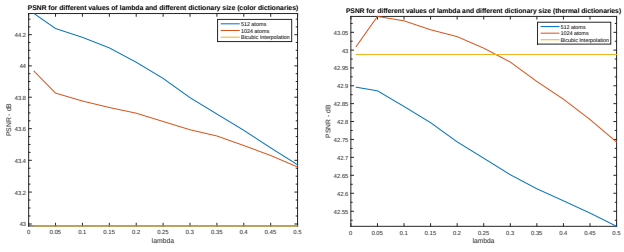
and we also define PSNR, as:

$$PSNR = 20 \cdot log_{10}(MAX_I) - 10 \cdot log_{10}(MSE) \qquad (12)$$

where MAX_I is the *maximum possible value* that a pixel of I can obtain (e.g. for grayscale images we use 255).

(a) input image (160×120) (b) output image (320×240)

Εικ. 1: *Testing image. Building at University of Patras' Campus.*



(a) optical images dict. (b) thermal images dict.

Εικ. 2: *PSNR using different dictionaries for reconstruction*

Bicubic Interpolation gives the following scores in MSE and PSNR: **PSNR:** 42.99 dB, **MSE:** 3.27.

We notice that the results of compressed sensing method are always better (in all cases). We should also consider that realtively "small" improvements of PSNR are in fact very significant as PSNR is measured on logarithmic scale.

When using dictinaries that have been built from optical images, we obtain improved results which is expected, as the dictionaries are built with respect to image *edges*, as explained at III-C. Generally speaking, optical images are rich in edges comparing to thermal images, making them a better source for training data.

TABLE I: PSNR/MSE for dictionaries built from optical images. Left: 512 atoms, Right: 1024 atoms.

λ	PSNR *dB*	MSE	PSNR *dB*	MSE
0.01	**44.34**	**2.40**	**43.97**	**2.61**
0.05	44.24	2.45	43.83	2.69
0.1	44.18	2.48	43.78	2.72
0.15	44.11	2.52	43.73	2.75
0.2	44.02	2.57	43.70	2.77
0.25	43.92	2.64	43.65	2.80
0.3	43.80	2.71	43.59	2.84
0.35	43.69	2.78	43.55	2.87
0.4	43.59	2.85	43.49	2.90
0.45	43.48	2.92	43.43	2.95
0.5	43.37	2.99	43.36	3.00

TABLE II: PSNR/MSE for dictionaries built from thermal images. Left: 512 atoms, Right: 1024 atoms.

λ	PSNR *dB*	MSE	PSNR *dB*	MSE
0.01	**42.90**	**3.34**	**43.01**	**3.19**
0.05	42.89	3.34	43.09	3.20
0.1	42.84	3.38	43.08	3.22
0.15	42.80	3.42	43.06	3.23
0.2	42.74	3.46	43.04	3.25
0.25	42.70	3.49	43.01	3.25
0.3	42.65	3.53	42.97	3.28
0.35	42.61	3.56	42.91	3.32
0.4	42.58	3.59	42.86	3.36
0.45	42.54	3.62	42.80	3.41
0.5	42.50	3.55	42.74	3.46

We notice that the results of compressed sensing method are always better (in all cases). We should also consider that realtively "small" improvements of PSNR are in fact very significant as PSNR is measured on logarithmic scale.

When using dictinaries that have been built from optical images, we obtain improved results which is expected, as the dictionaries are built with respect to image *edges*, as explained at III-C. Generally speaking, optical images are rich in edges comparing to thermal images, making them a better source for training data.

We also notice that by increasing λ, we obtain smoother results leading to lower PSNR scores, as the content of the image (e.g. edges) is "smoothed" giving an image that is not visually sharp.

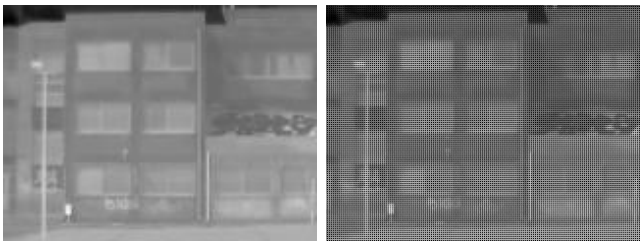
Regarding the efficiency of the compressed sensing method, we have to note that it has been implemented in MATLAB code and no speed optimizations have been taken into consideration. The execution time of the method is significantly affected by the size of the dictionary used and by the *sparsity* of the solution we want to achieve, which is defined by the value of λ. In general terms, the bigger the size of the dictionary and the lower the value of λ, will result in large execution times (5 or more minutes).

B. Shift-add fusion method

Shift-add fusion's efficiency depends on whether or not the subpixel shifts are accurately estimated. If the estimation is accurate, we can achieve a perfect high resolution image reconstruction, as we know the exact location of each pixel of the low resolution images on the high resolution grid. Most of times, however, accurate calculation of motion estimates is not an easy task, as in real world scenaria low resolution images contain noise (altering the visual content). Also the shifts we apply are not constrained to global translation

as we may introduce rotation and scaling which make the estimation process even more confusing.

As a result, when we do not get an accurate estimation, we notice *gaps* in the high resolution image as the reconstruction method puts the pixels of low resolution images in wrong locations on the high resolution grid. This is clearly seen in figure 3, where the estimation process used a small number of feature points to calculate shifts. This figure gives **PSNR:** 11.18 dB, whereas Bicubic Interpolation gives **PSNR:** 38.52 dB, magnifying a random image of the low resolution image set. Trying to "fill the gaps" by estimating their pixel values from their neighbors slightly improves the score, giving **PSNR:** 27.83 dB.



(a) input image (160×120) (b) output image (320×240)

Εικ. 3: *Shift-add fusion reconstruction*

V. CONCLUSION

Comparing the "conventional" approach of super resolution to the compressed sensing method, we notice that the latter gave optimum results in all cases, achieving a maximum score of PSNR: 44.34 dB. Bicubic interpolation scored 42.99 dB and shift-add fusion's best score was 27.83 dB (by filling the gaps of the reconstructed image). We should also mention that compressed sensing method requires only one low resolution image as input, while shift-add fusion requires a set of low resolution images.

ABOUT

This master thesis was conducted at Signal Processing and Telecommunications Laboratory of Dept. of Computer Engineering and Informatics and was supervised by Prof. Konstantinos Berberidis, Prof. Thanos Stouraitis (Dept. of ECE) and Dr. Evangelos Vlachos.

REFERENCES

[1] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: A technical overview," IEEE Signal Processing Magazine, vol. 20, no. 3, pp. 21–36, May 2003.

[2] D. Keren, S. Peleg, and R. Brada, "Image sequence enhancement using subpixel displacements," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 1988, pp. 742–746.

[3] R. Hardie, K. Barnard, and E. Armstrong, "Joint MAP registration and high resolution image estimation using a sequence of undersampled images," IEEE Transactions on Image Processing, vol. 6, no. 12, pp. 1621–1633, December 1997.

[4] A. Patti, M. Sezan, and A. Tekalp, "High-resolution image reconstruction from a low-resolution image sequence in the presence of time-varying motion blur," in Proceedings of the IEEE International Conference on Image Processing, Austin, TX, vol. 1, 1994, pp. 343–347.

[5] R. C. Hardie, K. J. Barnard, J. G. Bognar, E. E. Armstrong, and E. A. Watson, "High resolution image reconstruction from a sequence of rotated and translated frames and its application to an infrared imaging system," Optical Engineering, vol. 37, no. 1, pp. 247–260, January 1998.

[6] M. S. Alam, J. G. Bognar, R. C. Hardie, and B. J. Yasuda, "Infrared image registration and high-resolution reconstruction using multiple translationally shifted aliased video frames," IEEE Transactions on Instrumentation and Measurement, vol. 49, no. 5, pp. 923–915, October 2000.

[7] R. Y. Tsai and T. S. Huang, "Multiframe image restoration and registration," in Advances in Computer Vision and Image Processing: Image Reconstruction from Incomplete Observations, T. S. Huang, Ed., vol. 1. London: JAI Press, 1984, pp. 317–339.

[8] N. K. Bose, H. C. Kim, and H. M. Valenzuela, "Recursive total least squares algorithm for image reconstruction from noisy undersampled frames," Multidimensional Systems and Signal Processing, vol. 4, no. 3, pp. 253–268, July 1993.

[9] J. Yang, J. Wright, T. Huang, and Yi Ma. Image super-resolution via sparse representation. IEEE Transactions on Image Processing (TIP), vol. 19, issue 11, 2010.

[10] Compressive sensing, vol. 3, 2006.

[11] D. L. Donoho, "Compressed sensing," IEEE Transactions on Information Theory, vol. 52, no. 4, pp. 1289–1306, 2006.

[12] H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed sensing and redundant dictionaries," IEEE Transactions on Information Theory, vol. 54, no. 5, May, 2008.

[13] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," IEEE Transactions on Image Processing, vol. 15, pp. 3736–3745, 2006.

[14] J. Mairal, G. Sapiro, and M. Elad, "Learning multiscale sparse representations for image and video restoration," Multiscale Modeling and Simulation, vol. 7, pp. 214–241, 2008.

[15] H. Lee, A. Battle, R. Raina, and A. Y. Ng, "Efficient sparse coding algorithms," in Advances in Neural Information Processing Systems (NIPS), 2007.

[16] Jianchao Yang, Zhaowen Wang, Zhe Lin, and Thomas Huang. Coupled dictionary training for image super-resolution. IEEE Transactions on Image Processing (TIP), vol. 21, issue 8, pages 3467-3478, 2012.

[17] D. L. Donoho, "For most large underdetermined systems of linear equations, the minimal ℓ_1-norm solution is also the sparsest solution," Communications on Pure and Applied Mathematics, vol. 59, no. 6, pp. 797–829, 2006.

[18] "For most large underdetermined systems of linear equations, the minimal ℓ_1-norm near-solution approximates the sparsest near-solution," Communications on Pure and Applied Mathematics, vol. 59, no. 7, pp. 907–934, 2006.

[19] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based superresolution," IEEE Computer Graphics and Applications, vol. 22, pp. 56–65, 2002.

[20] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, "Learning lowlevel vision," International Journal of Computer Vision, vol. 40, no. 1, pp. 25–47, 2000.

[21] J. Sun, N. N. Zheng, H. Tao, and H. Shum, "Image hallucination with primal sketch priors," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2, 2003, pp. 729–736.

[22] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in IEEE Conference on Computer Vision and Pattern Classification (CVPR), vol. 1, 2004, pp. 275–282.

[23] Martin A. Fischler and Robert C. Bolles (June 1981). "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography". Comm. of the ACM 24 (6): 381–395

[24] R. Tibshirani, "Regression shrinkage and selection via the lasso," Journal of Royal Statistical Society, Series B, vol. 58, no. 1, 1996.