ACCURATE IMAGE REGISTRATION USING APPROXIMATE STRANG-FIX AND AN APPLICATION IN SUPER-RESOLUTION

Adam Scholefield and Pier Luigi Dragotti

Electrical and Electronic Engineering Department, Imperial College London.

ABSTRACT

Accurate registration is critical to most multi-channel signal processing setups, including image super-resolution. In this paper we use modern sampling theory to propose a new robust registration algorithm that works with arbitrary sampling kernels. The algorithm accurately approximates continuous-time Fourier coefficients from discrete-time samples. These Fourier coefficients can be used to construct an over-complete system, which can be solved to approximate translational motion at around 100-th of a pixel accuracy. The over-completeness of the system provides robustness to noise and other modelling errors. For example we show an image registration result for images that have slightly different backgrounds, due to a viewpoint translation. Our previous registration techniques, based on similar sampling theory, can provide a similar accuracy but not under these more general conditions. Simulation results demonstrate the accuracy and robustness of the approach and demonstrate the potential applications in image super-resolution.

Index Terms— Image registration, sampling methods, super-resolution, translational motion.

1. INTRODUCTION

Multi-view image super-resolution aims to combine multiple low-resolution (LR) images of the same scene into one high-resolution (HR) image. A crucial first step in this process is to register the images onto a common coordinate system with sub-pixel precision. The LR images can then be combined into a single image, which needs to be restored to obtain the final super-resolved (SR) result. This restoration step is an ill-posed inverse problem and its output quality depends heavily on the accuracy of the registration. In fact, registration is normally considered to be the most critical part of the process.

A wide range of strategies exist to register images in both the spatial and frequency domains, which are summarised in a number survey articles [1, 2]. In this paper we propose a frequency based technique utilising the normalised cross power spectrum (NCPS). This is a well-known technique that is based on the fact that a translation in the spatial domain corresponds to a phase shift in the frequency domain. It follows that the phase-only correlation (POC) function, defined as the inverse Fourier transform of the NCPS, consists of an impulse at the location of the translation. By processing the POC function in log-polar coordinates, it is possible to additionally register rotational and scaling transformations [3, 4].

Traditionally, the frequency spectra are calculated using the FFT, resulting in a discrete-time approximation of the POC function, which when maximised results in registration at around pixel precision. Extensions, to achieve sub-pixel precision, have been developed that approximate the location of the maximum more accurately. These methods usually rely on fitting an interpolation function to the POC approximation [5, 6, 7], or finding approximate zeros of its gradient [8].

In contrast we propose to use modern sampling theory to extract Fourier information of the original continuous-time signal. Baboulaz et al [9] used similar sampling theory to register images at a very high precision. They assumed that the image was sampled by a polynomial reproducing kernel allowing them to, in theory, extract the exact geometric moments of the continuous-time signal from the discrete-time samples. These moments could be used to register an affine transformation at, in theory, infinite precision.

By replacing a polynomial reproducing kernel with an exponential reproducing kernel, Fourier coefficients can be calculated instead of geometric moments. Urigüen et al [10] have also shown that it is possible to relax the assumptions on the sampling kernel and still accurately approximate Fourier coefficients of the original continuous-time signal. Futhermore, this approximate framework allows many more Fourier coefficients to be approximated than if we restricted ourselves to the exact reproduction case.

In this paper, we propose to apply this approximate exponential reproduction theory to the registration problem. This offers a number of advantages: firstly by dropping assumptions on the sampling kernel we can more accurately model realistic acquisition devices and secondly the ability to approximate more exponentials provides increased robustness.

The rest of this paper is organised as follows. In Section 2 we explain the single-channel acquisition model and show how continuous-time Fourier information can be found from the discrete-time samples. This relies on the theory of exact and approximate exponential reproduction theory,

This work is in part supported by the European Research Council (ERC) starting investigator award Nr. 277800 (RecoSamp).

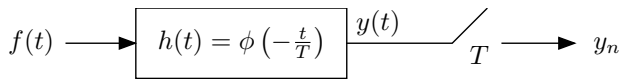


Fig. 1. Acquisition model in 1D. Here $f(t)$ is the continuous-time signal, $h(t)$ the impulse response of the acquisition device, T the sampling period and y_n the discrete-time samples.

which is also explained. In Section 3 we show how translational registration can be achieved at very high precision using the Fourier information found in Section 2. In Section 4 we present a super-resolution example using the proposed registration algorithm and in Section 5 we conclude.

To reduce notational load, we present the one-dimensional (1D) case whenever possible, since it is easy to extend it to higher dimensions as is needed for image registration. For example, in the following section, instead of reproducing a 1D exponential as

$$\sum_{n\in\mathbb{Z}} c_{m,n}\phi(t-n) = e^{-j\omega_m t},$$

we would reproduce a two-dimensional (2D) exponential as

$$\sum_{n_x\in\mathbb{Z}}\sum_{n_y\in\mathbb{Z}} c_{m_x,m_y,n_x,n_y}\phi(x-n_x, y-n_y) = e^{-j(\omega_{m_x}x+\omega_{m_y}y)}.$$

Additionally, we will use uppercase letters to denote the Fourier transform of their lowercase counterparts.

2. SINGLE-CHANNEL ACQUISITION SYSTEM

2.1. Acquisition model

The one dimensional (1D) acquisition model depicted in Fig. 1 results in the following expression for the samples y_n:

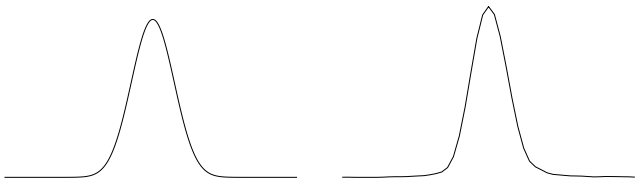
$$y_n = \int_{-\infty}^{\infty} f(t)\phi\left(\frac{t}{T}-n\right)dt = \left\langle f(t), \phi\left(\frac{t}{T}-n\right)\right\rangle,$$

where f is the continuous-time signal, ϕ is the sampling kernel and T is the sampling period; i.e., the samples are given by a uniform sampling of $y(t) = f(t) * \phi(-t/T)$. Throughout this paper we will assume, without loss of generality, that the sampling period $T = 1$.

The acquisition model can be trivially extended to 2D and used to model the image acquisition process. In this case, f is the 2D continuous-time image, ϕ is the 2D distortion introduced by the lens, and potentially other atmospheric effects, and y is the discrete-time image given by the pixel values. Note that in this case the sampling kernel is specified by the acquisition device and cannot be freely chosen.

2.2. Exponential reproducing sampling kernels

Figure 2(a) shows an example of an exponential reproducing kernel ϕ; i.e. properly weighted linear combinations of shifts



(a) Exponential reproducing kernel. (b) A real sampling kernel.

Fig. 2. Examples of different sampling kernels.

of ϕ reproduce one of M exponentials that, in general, take the form $e^{\alpha_m t}$, where $\alpha_m \in \mathbb{C}$ and $m \in \{1, 2, \dots M\}$. In this paper we are interested in the special case of purely imaginary α_m. In this case the reproduction can be written as

$$\sum_{n\in\mathbb{Z}} c_{m,n}\phi(t-n) = e^{-j\omega_m t}, \tag{1}$$

where $c_{m,n}$ are the coefficients that provide the proper weighting to reproduce the exponential $e^{-j\omega_m t}$. It is well known that kernels that satisfy (1) for $m \in \{1, 2, \dots M\}$ also satisfy the generalised Strang-Fix conditions [11] for the same values of m:

$$\Phi(\omega_m) \neq 0 \quad \text{and} \quad \Phi(\omega_m + 2\pi l) = 0 \quad \forall l \in \mathbb{Z}\setminus\{0\}. \tag{2}$$

Recall that the uppercase Φ denotes the Fourier transform of the time-domain kernel ϕ.

The correct coefficients, required for the reproduction, are given by the inner product of the desired exponential with a function $\tilde{\phi}$ that forms a quasi-biorthonormal set with ϕ [12, 13]:

$$c_{m,n} = \int_{-\infty}^{\infty} e^{-j\omega_m t}\tilde{\phi}(t-n)dt.$$

It can be shown that these coefficients are discrete-time exponentials: $c_{m,n} = c_{m,0}e^{-j\omega_m n}$, where

$$c_{m,0} = \frac{1}{\sum_{l\in\mathbb{Z}}\Phi(\omega_m + 2\pi l)e^{-j2\pi lt}} = \frac{1}{\Phi(\omega_m)}. \tag{3}$$

Note that the second equality, in (3), follows from the fact that the kernel, Φ, satisfies the Strang-Fix conditions, given in (2).

2.3. Continuous-time information from discrete-time samples

In the previous subsection we saw both the conditions for a kernel to be exponential reproducing and an expression for the coefficients that produce this reproduction. We are interested in this class of kernels because they allow us to obtain continuous-time Fourier information from the discrete-time samples. To see this, let ϕ be an exponential reproducing kernel that reproduces the exponentials $e^{-j\omega_m t}$ with properly chosen coefficients $c_{m,n}$ ($m \in \{1, 2, \dots M\}$). Also let y_n be the samples obtained from acquiring $f(t)$, using the acquisition model depicted in Fig. 1. Then, the following weighted combinations of the samples, y_n, are exactly equal

to the Fourier transform of the original continuous-time signal at the frequencies ω_m:

$$\begin{aligned} s_m &:= \sum_{n=0}^{N-1} c_{m,n} y_n = \sum_{n\in\mathbb{Z}} c_{m,n} \left\langle f(t), \phi\left(\frac{t}{T} - n\right)\right\rangle \\ &= \int_{-\infty}^{\infty} f(t) \sum_{n\in\mathbb{Z}} c_{m,n}\phi(t-n)dt \\ &= \int_{-\infty}^{\infty} f(t) e^{-j\omega_m t} dt = F(\omega_m). \end{aligned}$$

In the following two subsections we will show how these results can be extended to arbitrary sampling kernels.

2.4. Arbitrary sampling kernels

As we have seen, exponential reproducing kernels provide a very attractive way to infer continuous-time information from the discrete-time samples; however, in practical situations the sampling kernel is, typically, specified by the acquisition device and cannot be chosen. For example Fig. 2(b) shows a 1D plot of a cross section of the sampling kernel of a real digital camera. One could find an exponential reproducing kernel that closely approximates this real kernel, however there is a better approach.

Urigüen et al [10] have recently shown sampling schemes utilising approximate Strang-Fix theory. This new approach drops any assumptions on the kernel ϕ and instead looks to find the best coefficients that approximately reproduce exponentials given the arbitrary kernel:

$$\sum_{n\in\mathbb{Z}} c_{m,n}\phi(t-n) \simeq e^{-j\omega_m t}.$$

The advantage of this approach is two-fold: firstly we can work with the actual sampling kernel specified by the acquisition device and secondly we can (approximately) reproduce many more exponentials.

If the kernel could exactly reproduce exponentials we would use the coefficients $c_{m,n} = \frac{1}{\Phi(\omega_m)} e^{-j\omega_m n}$. Let us see what happens if we use the same coefficients with a general kernel ϕ. We are no longer guaranteed to reproduce an exponential and will instead produce

$$\begin{aligned} g_m(t) &:= \sum_{n\in\mathbb{Z}} c_{m,n}\phi(t-n) \\ &= \frac{1}{\Phi(\omega_m)} \sum_{n\in\mathbb{Z}} e^{-j\omega_m n}\phi(t-n) \\ &= e^{-j\omega_m t} \frac{1}{\Phi(\omega_m)} \sum_{n\in\mathbb{Z}} e^{j\omega_m (t-n)}\phi(t-n) \\ &= e^{-j\omega_m t} \frac{1}{\Phi(\omega_m)} \sum_{l\in\mathbb{Z}} \Phi(\omega_m + 2\pi l) e^{-j2\pi l t}, \end{aligned} \quad (4)$$

where the last equality comes from the Poisson summation formula.

In previous works, there has not been a specific way to decide if (4) is a satisfactory approximation of $e^{-j\omega_m t}$. We can formulate such a test by looking at the error of this approximation:

$$\begin{aligned} \epsilon(t) =& e^{-j\omega_m t} - g_m(t) \\ =& e^{-j\omega_m t}\left[1 - \frac{1}{\Phi(\omega_m)} \sum_{l\in\mathbb{Z}} \Phi(\omega_m + 2\pi l) e^{-j2\pi l t}\right]. \end{aligned}$$

This will be small if

$$\frac{1}{|\Phi(\omega_m)|} \sum_{l\in\mathbb{Z}\setminus\{0\}} |\Phi(\omega_m + 2\pi l)| \le \gamma, \quad (5)$$

where γ is some small non-negative scalar value. In practice (5) is satisfied for a large number of frequencies, ω_m. In addition, $\Phi(\omega_m + 2\pi l)$ usually approaches zero quickly as $|l|$ increases, therefore only a few terms of (5) need to be evaluated to understand the quality of the exponential approximation. In fact, in practice, it is often sufficient to just consider the first.

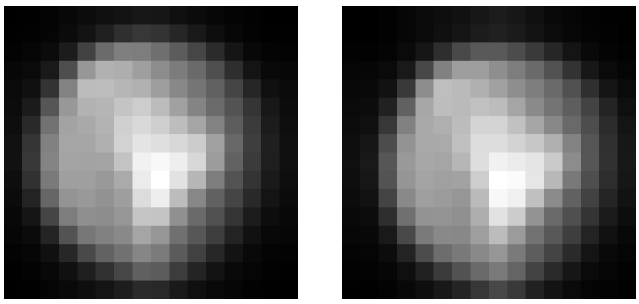
Note that if ϕ satisfies the generalised Strang-Fix conditions, given in (2), the left hand side of (5) and $\epsilon(t)$ are both zero.

In [10], Urigüen et al propose other possible choices for the coefficients $c_{m,n}$ that provide different exponential approximations. Specifically, they derive the coefficients that minimise the mean squared error (MSE) and the coefficients that produce an approximation that interpolates the true exponential $e^{-j\omega_m t}$ through the points $t \in \mathbb{Z}$. However, as the authors note, in practice the coefficients $c_{m,n} = \frac{1}{\Phi(\omega_m)} e^{-j\omega_m n}$, provide the best trade-off of accuracy and simplicity. For this reason they are the only coefficients considered in this paper and we refer the interested reader to [10] for the details of other possibilities.

2.5. Approximating Fourier coefficients with arbitrary sampling kernels

We now have all the information to approximate the continuous-time Fourier transform of the original signal at the frequencies ω_m. In the exact reproduction case the sum $s_m := \sum_{n=0}^{N-1} c_{m,n} y_n$ was exactly equal to $F(\omega_m)$. In the approximate reproduction case, the same sum yields

$$\begin{aligned} s_m &:= \sum_{n=0}^{N-1} c_{m,n} y_n = \int_{-\infty}^{\infty} f(t) g_m(t) dt \\ &= \int_{-\infty}^{\infty} f(t) e^{-j\omega_m t} \frac{1}{\Phi(\omega_m)} \sum_{l\in\mathbb{Z}} \Phi(\omega_m + 2\pi l) e^{-j2\pi l t} dt \\ &= \frac{1}{\Phi(\omega_m)} \sum_{l\in\mathbb{Z}} \Phi(\omega_m + 2\pi l) \int_{-\infty}^{\infty} f(t) e^{-jt(\omega_m + 2\pi l)} dt \\ &= \frac{1}{\Phi(\omega_m)} \sum_{l\in\mathbb{Z}} \Phi(\omega_m + 2\pi l) F(\omega_m + 2\pi l) =: \hat{F}(\omega_m). \end{aligned}$$



(a) LR image one. (b) LR image two.

Fig. 3. Two translated LR images.

Here we have used $\hat{F}$ to denote an approximation of the Fourier transform of f.

We can calculate the above approximation for different frequencies ω_m; however, it is impossible to obtain a precise error in each approximation, since we have no knowledge of F. However, in practical situations one can expect F to be smooth, so we use (5) to evaluate if the approximation $\hat{F}(\omega_m)$ is reliable. Specifically, to find Fourier information of a signal, $f(t)$, over a region of length L, we evaluate (5) for

$$\omega_m = \frac{2\pi m}{L}, \quad \forall m \in \mathbb{Z}. \tag{6}$$

When (5) is satisfied we treat $\hat{F}(\omega_m)$ to be an accurate approximation of $F(\omega_m)$, otherwise we discard that frequency. Choosing the frequencies given in (6) ensures that the exponentials contain whole periods across the interval.

3. REGISTRATION IN A MULTI-CHANNEL ACQUISITION SYSTEM

The previous analysis allows us to very accurately register two or more images that are related by translations. Consider two continuous-time images, $f_1(x, y)$ and $f_2(x, y)$, that are related by the translation $f_2(x, y) = f_1(x - s_x, y - s_y)$. This corresponds to a simple phase shift in the frequency domain:

$$F_2(\omega_x, \omega_y) = e^{-j(\omega_x s_x + \omega_y s_y)} F_1(\omega_x, \omega_y),$$

so that translation parameters can be found from the NCPS:

$$e^{j(\omega_x s_x + \omega_y s_y)} = \frac{F_1(\omega_x, \omega_y) F_2^*(\omega_x, \omega_y)}{|F_1(\omega_x, \omega_y) F_2^*(\omega_x, \omega_y)|}. \tag{7}$$

Note that, although $F_1(\omega_x, \omega_y)/F_2(\omega_x, \omega_y)$ produces the same exponential, the NCPS is much less sensitive to uniform variations of illumination and other globally constant variations.

Traditionally, (7) is solved by taking the inverse Fourier transform, which produces an impulse at the location of the translation. Since we only have access to the Fourier transforms at a limited number of frequencies, we take a different approach. We can look at the phase of (7) at each of the frequencies $(\omega_{m_x}, \omega_{m_y})$ that satisfy (5):

$$\omega_{m_x} s_x + \omega_{m_y} s_y = \arg\left(\frac{F_1(\omega_{m_x}, \omega_{m_y}) F_2^*(\omega_{m_x}, \omega_{m_y})}{|F_1(\omega_{m_x}, \omega_{m_y}) F_2^*(\omega_{m_x}, \omega_{m_y})|} \right).$$

Stacking these equations into an over-complete system and finding the least squares solution provides a robust estimation of the translational parameters. The robustness comes from the fact that we can accurately approximate many more Fourier coefficients than the two necessary to register a 2D translation.

Note that if ϕ satisfies the Strang-Fix conditions, we can set $\gamma = 0$ and, in theory, register at infinite precision, since there will be no error in the calculation of the Fourier coefficients.

Figure 3 shows two translated LR images that have been acquired using the real sampling kernel depicted in Fig. 2(b). Table 1 compares the registration performance of the proposed method with a traditional discrete approach and continuous geometric moments, as proposed in [9]. In order to approximate the continuous geometric moments the sampling kernel has been approximated by a B-spline.

	Discrete GM	Continuous GM	Proposed
RMSE (pixels)	0.7072	0.0948	0.0051

Table 1. A comparison of the root mean squared error (RMSE) when registering the two images depicted in Fig. 3. The registration techniques are discrete geometric moments (GM), continuous GM and the proposed approach.

4. APPLICATION IN IMAGE SUPER-RESOLUTION

Multi-view super-resolution aims to combine multiple LR images of the same scene into one HR image. It can be achieved using the steps shown in Fig. 4. Specifically, the multiple LR images are registered and interpolated on a HR grid. The resulting HR image is finally deconvolved to remove the blurring introduced by the sampling kernel.

Figure 5 shows an example of super-resolving an image by a factor of 10. In this simulation, a HR image was sampled with the real sampling kernel shown in Fig. 2(b) and subjected to additive white Gaussian noise with a standard deviation of 0.5, generating 100 LR images. Each LR image corresponded to a different 400×400 window of the HR image. Registration was performed as proposed and produced a root mean squared error (RMSE) of 0.0120 pixels. For comparison, approximating the kernel with a B-spline and registering from geometric moments, as proposed in [9], produces a RMSE of 0.288 pixels.

The registered images were placed on a HR grid and interpolated using bi-cubic interpolation. Restoration was performed using a regularised inverse and a state of the art denoising algorithm [14], producing the final SR image.

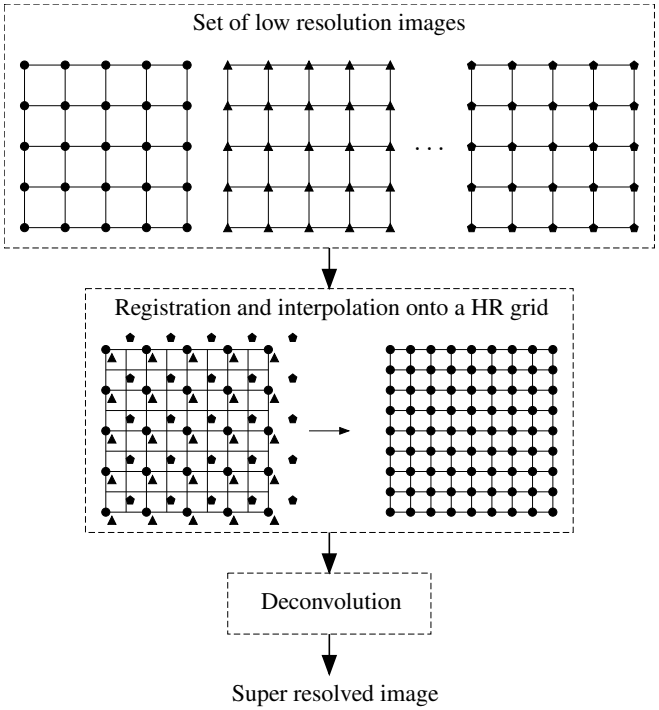


Fig. 4. Steps of super-resolution: first the LR images are registered onto a HR grid, then the samples are interpolated before a final deconvolution.

(a) One of the, 40×40, LR images. (b) The, 400×400, SR result.

Fig. 5. An example of 100 LR images being super-resolved into an image at 10 times the resolution.

5. CONCLUSIONS

In this paper accurate image registration has been achieved using phase information of the NCPS. In contrast to previous works, the NCPS has been calculated using continuous-time Fourier information that has been accurately estimated using modern sampling theory.

Simulation results demonstrates the accuracy of the registration and potential uses in applications such as super-resolution.

REFERENCES

[1] L. Brown, "A survey of image registration techniques," *ACM Computing Surveys*, vol. 24, no. 4, pp. 325–376, Dec. 1992.

[2] B. Zitová and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, no. 11, pp. 977 – 1000, 2003.

[3] B. Reddy and B. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Trans. on Image Processing*, vol. 5, no. 8, pp. 1266–1271, Aug 1996.

[4] G. Wolberg and S. Zokai, "Robust image registration using log-polar transform," in *IEEE International Conf. on Image Processing*, Sept 2000, vol. 1, pp. 493–496.

[5] I. Abdou, "Practical approach to the registration of multiple frames of video images," in *Proc. SPIE Conf. on Visual Communications and Image Processing*, Dec 1998, vol. 3653, pp. 371–382.

[6] H. Foroosh, J.B. Zerubia, and M. Berthod, "Extension of phase correlation to subpixel registration," *IEEE Trans. on Image Processing*, vol. 11, no. 3, pp. 188–200, Mar 2002.

[7] K. Takita, T. Aoki, Y. Sasaki, T. Higuchi, and K. Kobayashi, "High-accuracy subpixel image registration based on phase-only correlation," *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E86–A, no. 8, pp. 1925–1934, Aug 2003.

[8] A. Alba, R. Aguilar-Ponce, J. Vigueras-Gómez, and Edgar Arce-Santana, "Phase correlation based image alignment with subpixel accuracy," in *Advances in Artificial Intelligence*, vol. 7629 of *Lecture Notes in Computer Science*, pp. 171–182. Springer Berlin Heidelberg, 2013.

[9] L. Baboulaz and P.L. Dragotti, "Exact feature extraction using finite rate of innovation principles with an application to image super-resolution," *IEEE Trans. on Image Processing*, vol. 18, no. 2, pp. 281–298, Feb 2009.

[10] J.A. Uriguen, T. Blu, and P.L. Dragotti, "FRI sampling with arbitrary kernels," *IEEE Trans. on Signal Processing*, vol. 61, no. 21, pp. 5310–5323, Nov 2013.

[11] I. Khalidov, T. Blu, and M. Unser, "Generalized L-spline wavelet bases," in *Proc. SPIE Conf. on Mathematical Imaging: Wavelet XI*, Aug 2005, vol. 59140F, pp. 1–8.

[12] P.L. Dragotti, M. Vetterli, and T. Blu, "Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets Strang-Fix," *IEEE Trans. on Signal Processing*, vol. 55, no. 5, pp. 1741–1757, May 2007.

[13] T. Blu and M. Unser, "Approximation error for quasi-interpolators and (multi-) wavelet expansions," *Applied and Computational Harmonic Analysis*, vol. 6, no. 2, pp. 219–251, Mar 1999.

[14] A. Scholefield and P.L. Dragotti, "Quadtree structured image approximation for denoising and interpolation," *IEEE Trans. on Image Processing*, vol. 23, no. 3, pp. 1226–1239, Mar 2014.