ON THE AMOUNT OF REGULARIZATION FOR SUPER-RESOLUTION INTERPOLATION

Yann Traonmilin, Saïd Ladjal and Andrés Almansa

Telecom ParisTech, LTCI

ABSTRACT

Super-resolution (SR) aims at combining several aliased images of the same scene into a higher resolution image by using the difference in sampling caused by camera motion. As the problem of SR is generally ill-posed, techniques developed in the literature often rely on hypotheses on the regularity of the image. In this paper, we try to minimize these assumptions for the interpolation part of super-resolution. We describe situations where SR interpolation is invertible and/or well conditioned. We first study the interpolation problem for large numbers of images, when motions are pure translations. Then, we look at the more generic problem of super-resolution interpolation with translations and rotations. We give a simple condition on the number of images and zoom factor for perfect recovery of the high resolution image. We also study the conditioning in the critical case and propose a regularization method which adapts to local sampling variations.

Index Terms— Super-resolution, image processing, assumptions minimization, regularization degree, perfect reconstruction condition

1. INTRODUCTION

Super-resolution is the recovery of a high resolution (HR) image from several low resolution (LR) images taken from different positions. If we do not restrict the problem to a particular case, we need to estimate the camera blur, the motions and the HR image simultaneously, which is an ill-posed problem. Super-resolution has been reviewed a number of times [1] [2] in the literature. Most techniques rely on a regularized minimization of a data fit functional [3] [4] [5]. There is a wide choice of functionals that we can try to minimize to recover the HR image, e.g., L^2 norm with total variation (TV) regularization, L^1 norm with TV regularization. The choice of a regularizer is an implicit hypothesis (or a priori information) on the content of the image. For example, perfect reconstruction with TV regularization is not possible if the image contains too much texture [6]. Our aim is to avoid or minimize the assumptions made on the HR image, and consequently minimize the amount of regularization.

To reduce the complexity, motion is often restricted to a composition of translations and rotations. In this paper, we will also limit ourselves to the interpolation aspect of SR (motion parameters are given and camera blur is not inverted). With these restrictions, SR can be viewed as an irregular sampling problem which could be solved with dedicated techniques [7] [8]. However, we can use the fact that each LR image is acquired on a regular grid to obtain more powerful methods. For example, the pure translational case can be viewed as a multichannel sampling problem. Papoulis [9] showed that if we have a super-resolution factor M, only M^2 LR images are needed to recover perfectly the HR image in a noiseless set-up. In this case, no assumption on the image is needed. It naturally leads to the study of approaches where we do not regularize or regularize as little as possible.

In this paper, we first describe our theoretical context. We then justify that a non-regularized approach is valid for translational SR interpolation when many images are available. The conditioning of the system only becomes a problem when the number of images becomes close to the critical case. We then show that we have the same condition as Papoulis for the invertibility of the problem when we allow for rotations. We finally propose a local TV regularization scheme in the near-critical case, when rotations are small, to reduce the noise generated by badly conditioned areas.

2. THE SR INTERPOLATION PROBLEM

Let u be the HR image. We represent u as a discrete image (of sampling step 1) because the optical system of a camera acts as a low-pass filter on the continuous image before sampling. The N LR images w_i are formed by:

$$w_i = SV_iu$$

where the operator S is the sub-sampling by a factor M and V_i is the motion of each LR image. We restrict this motion to a translation in the next part, and will extend it to a composition of a translation and a rotation in the rest of the paper. The SR interpolation problem is the inversion of the linear map:

$$A : u \to (SV_iu)_{i=1,N} = (w_i)_{i=1,N} = \mathbf{w} \tag{1}$$

Throughout the paper, we study this problem under several angles. First, we look at the conditioning when $N >> M^2$ for pure translations. Then we show that $N = M^2$ is sufficient for a perfect theoretical interpolation when motions include rotations in the noiseless case. Finally, we study the case where $M^2 \leq N < M^2 + \epsilon$, ($\epsilon \in \mathbb{N}$) for rotation-translation motions, i.e. invertible cases were it is more likely to have a bad conditioning.

3. WELL-POSED TRANSLATIONAL SR INTERPOLATION

In [10], it was shown that the condition number of the system grows with the SR factor M. In this part, we show that the condition number of the translational ($V_i = T_i$, where T_i is a pure translation) SR interpolation problem converges to one as N grows, a fact which was experimentally illustrated in [11] in terms of the Cramer-Rao bound for HR image estimation. A similar result was demonstrated in [12] for the reconstruction error. We give a quick demonstration for our formulation relying on an argument from [12]. We consider the 1D case for the simplicity of the formulas. If the 1D condition holds ($N \geq M$ and translations are all different), the HR image can be perfectly recovered with a conditioning given by the condition number of $A^H A$. We show the following:

Proposition 3.1. *If the translations follow a uniform distribution (the translations $(t_i)_{i=1,N}$ for each LR image are i.i.d. random variables uniformly distributed in $[0, M]$), the conditioning κ of the system converges to 1 (in the distribution sense) as the number of images grows.*

Proof. To recover $\hat{u}$ (Fourier transform of u) at a particular pulsation ω, we first write $\hat{w}_i(\omega)$ (Fourier transform of w_i) as a linear combination of the $\hat{u}(\omega)$:

$$\hat{w}_i(\omega) = \frac{1}{M} \sum_{k \in \mathbb{Z}} \hat{u}(\omega + \frac{2\pi k}{M}) e^{j(\omega + \frac{2\pi k}{M})t_i}$$

Only M terms in the sum are non-zero. If there is more than M LR images, it is an overdetermined system of size $N \times M$ for each pulsation:

$$\Delta B \hat{u}_{al} = w$$

where $\hat{u}_{al}(k) = \hat{u}(\omega + \frac{2\pi k}{M})$, $w(i) = \hat{w}_i(\omega)$, $\Delta = \frac{1}{M} diag(e^{j\omega t_i})$ and $B_{i,k} = e^{j\frac{2\pi k}{M}t_i}$. The conditioning of the system is consequently the conditioning of $R = B^H B$ which is a Toeplitz matrix with term $R_{r,s} = \sum_i e^{j\frac{2\pi(s-r)}{M}t_i}$. A direct application of the central limit theorem shows that R converges to a multiple of identity because the complex numbers $e^{j\frac{2\pi(s-r)}{M}t_i}$ converge to a uniform distribution on the unit circle (for $s \neq r$). By continuity of the condition number, the condition number $\kappa(R)$ converges to 1. □

In Fig.1, we plot the logarithm of the condition number $\kappa(R)$ of random realizations of R as $N \to \infty$ for $M = 2$ and $M = 4$. It converges to 0 (i.e. the condition number converges to 1) for large N. As R converges to a multiple of identity, no inversion is needed when $N \to \infty$. We just

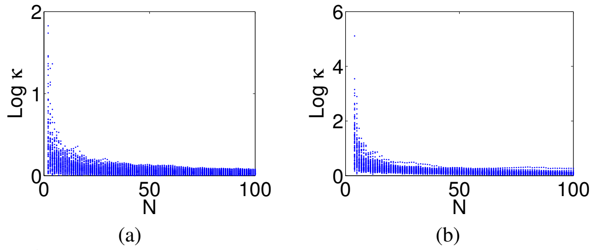

Fig. 1. Convergence of the conditioning of a 1D SR problem with respect to the number of images. (a) 80 realizations of the experiment with $M = 2$. (b) 80 realizations of the experiment with $M = 4$.

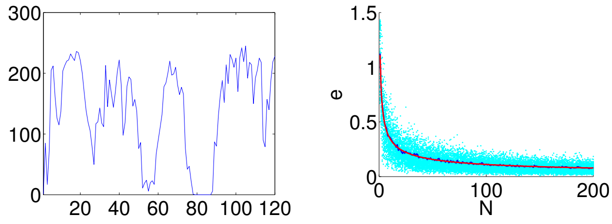

Fig. 2. Convergence of the estimator $\tilde{u}_N$. (a) Reference signal. (b) Reconstruction errors (light blue) for M=3 with predicted convergence speed in red. In dark blue is the mean of the error for each N value (hidden under the prediction).

need to apply a normalized version of the adjoint operator of A to $\mathbf{w}$. This operation gives the mean of back-shifted zero-padded LR images. We call S^T the operation of up-sampling by zero-padding by a factor M. We have:

$$\tilde{u}_N = \frac{1}{N} \sum_{i=1,N} T_i^{-1} S^T w_i \xrightarrow[N \to \infty]{} u \qquad (2)$$

The central limit theorem for $\tilde{u}_N$ gives a convergence speed of $\frac{1}{\sqrt{N}}$ for the L^2 norm error of this estimator. In Fig.2, we generate several SR experiments with the same reference signal (line extracted from a natural image). We then plot the relative reconstruction error of the estimator $\tilde{u}_N$ from equation (2) overlaid with the expected convergence speed.

We showed in this part that, with a large number of image, a direct inversion is possible with a high probability. To achieve this result, we did not make any assumption on the HR image. It leads to an intuitive take on super-resolution: it is very interesting to take more images than the critical case to try to avoid regularization in the inversion process.

4. PERFECT INTERPOLATION IN A ROTATION-TRANSLATION SET UP

In this part, we extend the result on the critical number of images for perfect interpolation to motions composed of rotations and translations. We consequently write the transformation of the LR images as a composition of rotations and translations. We will consider the finite discrete case :

$$A : (\mathbb{C}^{ML\times ML}) \to (\mathbb{C}^{L\times L})^N$$
$$u \to (SR_iT_iu)_{i=1,N}$$

Here, the T_i and R_i are the translations and rotations by Shannon interpolation. A is a linear map on the vector space of discrete images. Thus, perfect recovery is possible if $N \geq M^2$ and A is full rank.

We now show that $N \geq M^2$ is a sufficient condition under a hypothesis on the transformations. Let us name the following sampling grids: $\Gamma^{hr} = [1, ML]^2 \subset \mathbb{Z}^2$ and $\Gamma = M.[1, L]^2$. Γ^c is the complement of Γ in Γ^{hr}, i.e. the support of images in the kernel of S. We write r_i the rotation on the coordinates for LR image i and t_i the value of its translation. The following theorem states a sufficient condition for A to be invertible: none of the displacement coordinates between two different transform of two points of Γ^c should be an integer.

Theorem 4.1. *If $N \geq M^2$ and $\forall p_i, p_j \in \Gamma^c, 1 \leq k_1 < k_2 \leq N, ||r_{k_1}^{-1}p_i - r_{k_2}^{-1}p_j + (t_{k_1} - t_{k_2})\, mod\, 1||_0 = 2$, A is injective.*

Proof. We show by induction over N that adding a LR image decreased the dimension of the kernel of the function A by a factor L^2. We demonstrated the necessary lemmas in the appendix for clarity purpose. Let:

$$A_n : (\mathbb{C}^{ML\times ML}) \to (\mathbb{C}^{L\times L})$$
$$u \to SR_nT_nu$$

We prove : $\forall 1 < n \leq M^2, dim \cap_{k=1,n} kerA_k = (M^2 - n)L^2$

For $n = 2$: let $p_i \in \Gamma^c$. Let $v_i = \mathbf{1}_{p_i}$. Let $u_i = T_1^{-1}R_1^{-1}v_i$. We have $Sv_i = 0$. Consequently $A_1u_i = 0$ and $u_i \in kerA_1$. We just defined $(M^2 - 1)L^2$ independent u_i generating $kerA_1$: $span(u_i)_{i=1,(M^2-1)L^2} = kerA_1$. Similarly we construct $span(u'_i)_{i=1,(M^2-1)L^2} = kerA_2$. We calculate the dimension of the intersection with Lemma A.2 ($kerA_1 + kerA_2 = \mathbb{C}^{ML\times ML}$):

$$\begin{aligned} &dim(kerA_1 \cap kerA_2) = \\ &\quad dim(kerA_1) + dim(kerA_2) - dim(kerA_1 + kerA_2) \\ &\quad = 2(M^2 - 1)L^2 - dim(kerA_1 + kerA_2) \\ &\quad = (M^2 - 2)L^2 \end{aligned}$$

Let $n > 2$. Let us suppose that $dim \cap_{k=1,n} kerA_k = (M^2 - n)L^2$. We use Lemma A.3: $(\cap_{k=1,n}kerA_k) + kerA_{n+1} = \mathbb{C}^{ML\times ML}$. By using the same dimensions relation as for $n = 2$, we get the result. □

The hypothesis on the transformation is not a necessary condition. In the proof, we imposed the invertibility in both spatial directions, which is stronger than necessary (e.g. the case of regular sampling is excluded but is already known), but the space of excluded motion parameters has measure 0.

5. LOCAL CONDITIONING IN A ROTATION-TRANSLATION SET UP

The fusion of rotated-translated grids is a sampling grid which is generally not periodic. We expect local variations of the spatial distribution of samples leading to a spatial variability in the noise generated by the inversion of the system. In this section, we show that we can predict this conditioning when rotations are small (which would be a reasonable hypothesis for a hand held camera), and use this prediction to regularize adaptively with respect to local conditioning.

We study the conditioning in the critical case $N = M^2$ where the problem is invertible (from the previous section). We compare the reconstruction noise n_{rec} of the system and the reconstruction noise of a pure translational model at each location. When LR images are contaminated by a noise $\mathbf{n}$, the reconstruction noise is $n_{rec} = A^{-1}\mathbf{n}$.

We calculate the power of this noise locally. We restrict the image space of the application A^{-1} to one LR pixel in the HR image space to study its local behavior. Let $\mathbf{x}_0 = [x_0, y_0]$. Let $\mathbf{x} \in [x_0, x_0 + M - 1] \times [y_0, y_0 + M - 1] = D \subset \mathbb{Z}^2$. Let $\mathbf{1_x}$ be the indicator function $\mathbf{x} \in D$ in the HR image. We now consider the mapping:

$$A_{\mathbf{x}_0}^{-1} : E = A(span((\mathbf{1_x})_{\mathbf{x}\in D}) \to F = span((\mathbf{1_x})_{\mathbf{x}\in D})$$
$$w \to A^{-1}w$$

We call local conditioning at position $\mathbf{x}_0$, the conditioning of $A_{\mathbf{x}_0}^{-1}$. This conditioning is the ratio of the bounds of the quantity (greatest and smallest singular values):

$$||A_{\mathbf{x}_0}^{-1}w||, ||w|| = 1$$

We can calculate equivalently the bounds of $||A_{\mathbf{x}_0}u||, ||u|| = 1$. Let $u = \sum b_k\mathbf{1}_{\mathbf{x_k}} \in F$ with $||u|| = 1$. We have :

$$\begin{aligned} ||A_{\mathbf{x}_0}u||^2 &= ||\sum b_kA\mathbf{1}_{\mathbf{x_k}}||^2 \\ &= \sum_{k_1,k_2} \bar{b}_{k_1}b_{k_2}(\mathbf{1}_{\mathbf{x_{k_1}}})^HA^HA\mathbf{1}_{\mathbf{x_{k_2}}} \\ &= \sum_{k_1,k_2} \bar{b}_{k_1}b_{k_2} \sum_{i=1,N}\sum_{\mathbf{y}\in\Gamma} \text{sincd}(\mathbf{y} - \tau_{i,k_1})\text{sincd}(\mathbf{y} - \tau_{i,k_2}) \end{aligned}$$

where sincd is the finite discrete Shannon interpolator and $\tau_{i,k} = r_i(\mathbf{x_k} - t_i)$. Because sincd is differentiable, we can use the mean value theorem to compare this expression to a pure translational one and obtain an expression of the form:

$$|||A_{\mathbf{x}_0}u||^2 - ||A_{\mathbf{x}_0}^{tr}u||^2| \leq K||\mathbf{t} - \mathbf{t}^{tr}||$$

where $\mathbf{t} = (\tau_{i,k})_{i,k}$ is the set of translations induced by the motion, $\mathbf{t}^{tr}$ is $\mathbf{t}$ averaged over the HR pixel (over index k) and $A_{\mathbf{x}_0}^{tr}$ is the translational SR operator associated with $\mathbf{t}^{tr}$ and K is a constant which does not depend on $\mathbf{x}_0$. Thus, for small rotations, the noise of the system will behave as in a pure translational case. Experiments showed that for rotations in

a small range (-5,+5 degrees), we can use $\kappa(\mathbf{x}_0) = cond(R)$ as a local conditioning measure, with R defined as in section 3 with the translations $\mathbf{t}^{tr}$. This measure can be calculated a priori because it only depends on motion parameters, M and the size of the image.
We propose a local total variation regularization scheme where our local conditioning measure defines weights for the TV term. We minimize the function:

$$J_2(\tilde{u}) = J(\tilde{u}) + H(\tilde{u})$$
$$H(\tilde{u}) = \int \alpha.|\nabla \tilde{u}|$$

where $\alpha(\mathbf{x}) = \lambda log(\kappa(\mathbf{x}))$. The choice of the logarithm was driven by experiments where other increasing functions were tested (identity, square-root). λ is the regularization parameter. Conventional TV regularization is achieved by taking $\alpha(\mathbf{x}) = \lambda$. The selection of an optimal λ is an issue which is not adressed here. In the following experiments, we choose the λ achieving the best reconstruction in the L^2 sense for global TV regularization and local TV regularization. We show in Fig.3 how we predict local conditioning. We generate 9 noisy LR images from a 240×240 HR image (SR with $M = 3$, rotations between -5 and 5 degrees, translation distributed in $[0, M]^2$) and perform the inversion of the system (1) without regularization. The amplitude of the reconstruction noise and the corresponding prediction of the local conditioning show a good spatial correlation. The fusion of the points of the 9 grids also illustrates the spatially varying nature of the sampling density. In Fig.4, SR interpolation without regularization, with a global TV regularization or with our local regularization scheme are compared. In Fig.4, reconstruction without regularization is perfect in well conditioned areas. Global TV regularization does not take into account the spatial variations of the sampling density. Thus, the regularization parameter λ is a trade-off. A large value for λ would be needed to fill parts where holes are big (because of the bad local conditioning of the data fit part) and a small one for well conditioned areas. Consequently, the resulting image is smoothed excessively in well conditioned areas and not enough in badly conditioned areas. With local regularization, smoothing occurs only in badly conditioned areas. Differences are mostly noticeable in the images of the residuals (Fig.4 (e) (f) displayed at the same scale).

6. CONCLUSION

We have studied super-resolution interpolation under several aspects. We showed that the conditioning depends on the number of images: if many LR images are available, the likelihood of a bad conditioning decreases. Avoiding regularization accordingly would lead to SR interpolation without hypothesis on the nature of the image. For more complex motions (rotations + translations), we showed that the critical

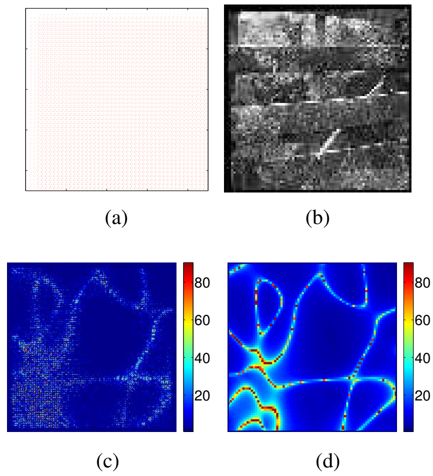

Fig. 3. Local conditioning of the SR problem. (a) Zoom on the fusion of the 9 LR grids (40×40 pixels upper left corner). (b) one LR image. (c) Amplitude of the reconstruction error normalized by the input noise variance. (d) Local conditioning $2\sqrt{\kappa(\mathbf{x})}$.

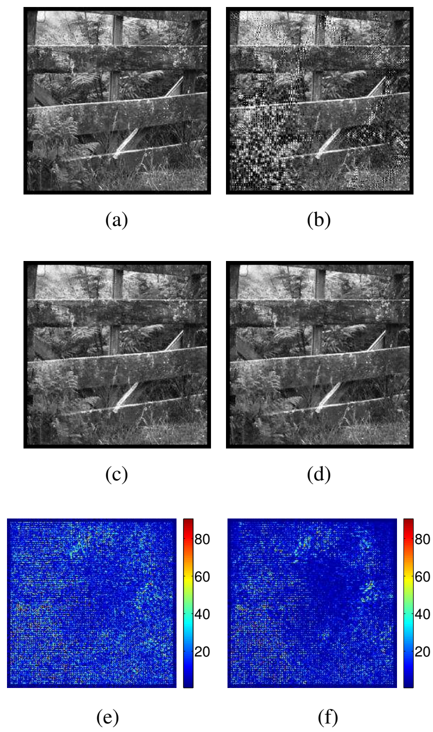

Fig. 4. Local TV regularization for critical SR. (a) HR image. (b) Reconstruction without regularization, PSNR=11.38db. (c) Reconstruction with best global TV regularization, PSNR=27.11db. (d) Reconstruction with our local regularization, PSNR=29.23db. (e) Reconstruction error with best global TV regularization. (f) Reconstruction error with our local regularization.

condition on the number of images still holds and is sufficient almost surely. To fill the gap between well conditioned invertible and ill-posed SR, we propose a way to regularize locally the reconstruction problem when camera rotation is small. With these developments, this paper completes the available set of SR interpolation methods to the well posed and badly conditioned invertible cases.

7. REFERENCES

[1] Jing Tian and Kai-Kuang Ma, "A survey on super-resolution imaging," *Signal, Image and Video Processing*, pp. 1–14, Feb. 2011.

[2] Sina Farsiu, Dirk Robinson, Michael Elad, and Peyman Milanfar, "Advances and challenges in super-resolution," *Int. J. Imaging Syst. Technol.*, vol. 14, no. 2, pp. 47–57, 2004.

[3] Kim-Hui Yap, Yu He, Yushuang Tian, and Lap-Pui Chau, "A Nonlinear-Norm Approach for Joint Image Registration and Super-Resolution," *Signal Processing Letters, IEEE*, vol. 16, no. 11, pp. 981–984, Nov. 2009.

[4] M. D. Robinson, C. A. Toth, J. Y. Lo, and S. Farsiu, "Efficient Fourier-Wavelet Super-Resolution," *Image Processing, IEEE Transactions on*, vol. 19, no. 10, pp. 2669–2681, Oct. 2010.

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

[6] Yann Gousseau and Jean-Michel Morel, "Are Natural Images of Bounded Variation?," *SIAM J. on Mathematical Analysis*, vol. 33, pp. 634+, 2001.

[7] K. Gröchenig and T. Strohmer, "Numerical and theoretical aspects of nonuniform sampling of band-limited images," in *Nonuniform sampling : Theory and Practice*, Inf. Technol. Transm. Process. Storage, pp. 283–324. Kluwer/Plenum, New York, 2001.

[8] Gabriele Facciolo, Andres Almansa, Jean-Francois Aujol, and Vicent Caselles, "Irregular to Regular Sampling, Denoising, and Deconvolution," *Multiscale Modeling & Simulation*, vol. 7, no. 4, pp. 1574+, 2009.

[9] A. Papoulis, "Generalized sampling expansion," *Circuits and Systems, IEEE Transactions on*, vol. 24, no. 11, pp. 652–654, Nov. 1977.

[10] Simon Baker and Takeo Kanade, "Limits on Super-Resolution and How to Break Them," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167–1183, 2002.

[11] D. Robinson and P. Milanfar, "Statistical performance analysis of super-resolution," *Image Processing, IEEE Transactions on*, vol. 15, no. 6, pp. 1413–1428, June 2006.

[12] Frédéric Champagnat, Guy Le Besnerais, and Caroline Kulcsár, "Statistical performance modeling for superresolution: a discrete data-continuous reconstruction framework," *J. Opt. Soc. Am. A*, vol. 26, no. 7, pp. 1730–1746, July 2009.

A. APPENDIX

Lemma A.1. *For $1 \leq i \leq N$, let $u_i \in \mathbb{C}^{n\times n}$, $u_i(r,s) = x_i^r y_i^s$, we call u_i 2D Vandermonde vectors with seed $[x_i, y_i]$. If $\forall 1 \leq i < j \leq N, x_i \neq x_j, y_i \neq y_j$, $dim(span(u_i)_{i=1,N}) = min(N, n^2)$.*

Proof. We show that the u_i are linearly independent if $N \leq n^2$. Let us suppose $\sum \lambda_i u_i = 0$. Let $u_i(s) = X_i y_i^s$ with $X_i = (x_i^r)_r$. $\forall s, \sum \lambda_i u_i(s) = \sum \lambda_i X_i y_i^s = 0$. The X_i form an independent family of 1D Vandermonde vectors. It implies $\forall s, \sum \lambda_i y_i^s = 0$ which we rewrite $\sum \lambda_i Y_i = 0$, but the Y_i are also independent. Consequently, $\lambda_i = 0$. □

Lemma A.2. *If $\forall p_i, p_j \in \Gamma^c, ||r_1^{-1}p_i - r_2^{-1}p_j + (t_1 - t_2)\, mod\, 1||_0 = 2$, $kerA_1 + kerA_2 = \mathbb{C}^{ML\times ML}$.*

Proof. We can construct a basis of $kerA_1$ and $kerA_2$ by taking the inverse transformations of the indicator functions of the pixels zeroed by the subsampling. In the Fourier domain, these bases are:

$$\hat{u}_i(\omega) = e^{-j<\omega, r_1^{-1}p_i + t_1>},\ \hat{u}'_i(\omega) = e^{-j<\omega, r_2^{-1}p_i + t_2>}$$

which are 2D Vandermonde vectors with seed $[e^{-j<e_x, r_k^{-1}p_i + t_k>}, e^{-j<e_y, r_k^{-1}p_i + t_k>}]$. We use Lemma A.1: $kerA_1 + kerA_2 = span((\hat{u}_i), (\hat{u}'_i)) = \mathbb{C}^{ML\times ML}$ (the seeds are all different because $\forall p_i, p_j, ||r_1^{-1}p_i - r_2 p_j^{-1} + t_1 - t_2\, mod\, 1||_0 = 2$). □

Lemma A.3. *Let $N < M^2$.If $\forall p_i, p_j \in \Gamma^c, 1 \leq k_1 < k_2 \leq N, ||r_{k_1}^{-1}p_i - r_{k_2}^{-1}p_j + (t_{k_1} - t_{k_2})\, mod\, 1||_0 = 2$ and $dim(\cap_{k=1,n} kerA_k) = (M^2 - n)L^2$ then $\cap_{k=1,n} kerA_k + kerA_{n+1} = \mathbb{C}^{ML\times ML}$.*

Proof. Let (e_i) be a basis of $\cap_{k=1,n} kerA_k$ of size $(M^2 - n)L^2$. In the basis $(u_j)_{j=1,n}$ of $kerA_1$,:

$$e_i = \sum \alpha_{i,j} u_j$$

Let u'_i a basis of $kerA_n$. With the hypothesis, any linear combination of e_i, u'_i is a linear combination of independent Vandermonde vectors. Therefore, $dim(span((\hat{e}_i), (\hat{u}'_i))) = min((ML)^2, (M^2 - n)L^2 + (M^2 - 1)L^2) = (ML)^2$. Thus, we have $\cap_{k=1,n} kerA_k + kerA_2 = span((\hat{e}_i), (\hat{u}'_i)) = \mathbb{C}^{ML\times ML}$. □