AN EPIPOLAR-CONSTRAINED PRIOR FOR EFFICIENT SEARCH IN MULTI-VIEW SCENARIOS

Ignacio Bosch[†] [‡] *Jordi Salvador*[†] *Eduardo Pérez-Pellitero*[†] *Javier Ruiz-Hidalgo*[‡]

[†] Technicolor R&I Hannover

[‡] Image Processing Group
Universitat Politècnica de Catalunya

ABSTRACT

In this paper we propose a novel framework for fast exploitation of multi-view cues with applicability in different image processing problems. In order to bring our proposed framework into practice, an epipolar-constrained prior is presented, onto which a random search algorithm is proposed to find good matches among the different views of the same scene. This algorithm includes a generalization of the local coherency in 2D images for multi-view wide-baseline cases. Experimental results show that the geometrical constraint allows a faster initial convergence when finding good matches. We present some applications of the proposed framework on classical image processing problems.

Index Terms— Super resolution, deblurring, epipolar line, approximate nearest neighbor

1. INTRODUCTION

Multi-view settings for image processing provide richer information thanks to the redundancy and disocclusions inherent to its data. This kind of data has improved the performance of a great number of different applications in the image processing field, such as reconstruction [**?**], accurate tracking [**?**] or pose estimation [**?**].

The redundancy of data among different views of the same scene is particularly useful. It is known that if an image is divided in small regions (called *patches*) it is very likely to find similar ones in the same image and across scales [**?**] but also between different multi-view captures of the same scene [**?**]. As it will be seen in Section 3, this redundancy can be combined with *geometric constraints* to formulate a non-parametric compact prior that can be useful to solve a grand variety of classical image processing problems.

As usual in exploitation of non-parametric priors, the use of efficient search algorithms to find similar patches is required. A naive approach to exploit the multi-view information would be to perform a random search on the other views using efficient methods as approximate nearest neighbor by either random sampling and propagation [**?**], propagation of good matches with kd-trees [**?**] or seeding initial matches with hashing and propagation by using image coherency [**?**]. However, none of them is designed to take advantage of the information redundancy present in multi-view environments. We propose on Section 4 an efficient search algorithm based on the compact prior.

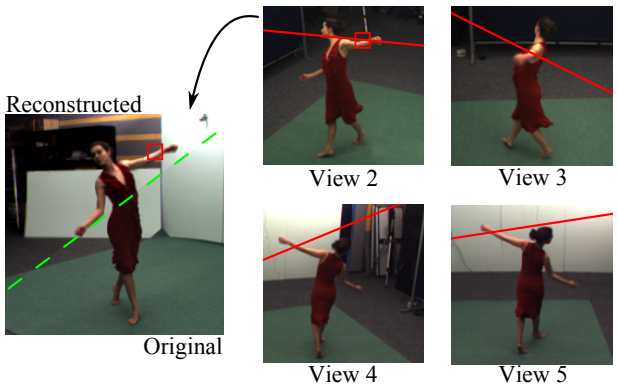


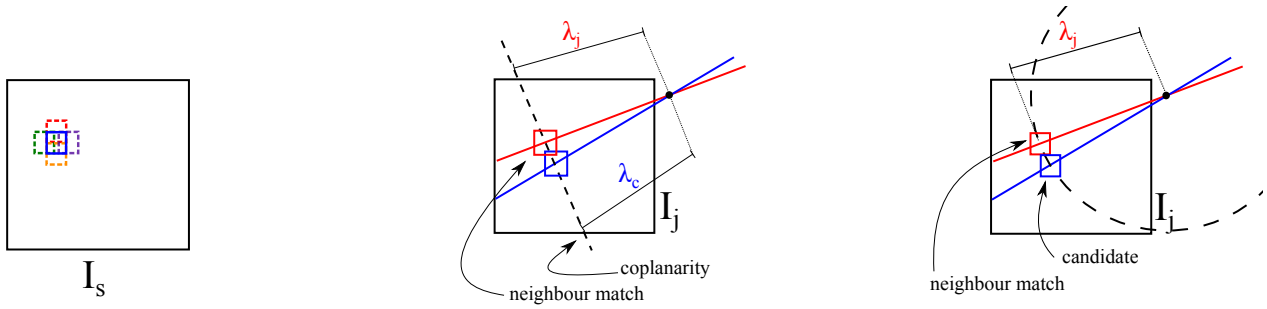
Fig. 1. Patch-based reconstruction by choosing the best match on epipolar lines in the other available views. The left image of this figure shows the reconstruction (top-left half) along with the original image (bottom-right part of the image).

In order to show the applicability of our framework, in Section 5 we exploit our prior in two applications: super resolution and deblurring, in which we recover the lost sharpness of a blurred image by using the information contained in other views of the same scene.

2. NOTATION

We denote 2D image positions as lowercase bold letters ($\mathbf{x}$), 3D positions in uppercase bold letters ($\mathbf{X}$), and $\tilde{\mathbf{x}}$, $\tilde{\mathbf{X}}$ their homogeneous counterparts.

Uppercase letters (I) refer to images. In any of the images, a subscript indicates the image number, while the superscript indicates whether the image contains low-frequency (I^L) or high-frequency (I^H) information. If there is no superscript, we refer to a sharp image (containing both low- and high-frequency components).



(a) Source patch and its neighbours (b) Propagation by coplanarity (c) Relaxation of the coplanarity

Fig. 2. The proposed mechanism for multi-view propagation. (b) presents a generalization for wide-baseline multi-view cases of the 2D local coherency by coplanarity propagation to determine the candidate patch (blue) from the neighbor patch (red). Then we propagate the position parameter λ_j between neighbor patches with *circular propagation*, as shown in (c).

3. EPIPOLAR-CONSTRAINED PRIOR

In [**?**], targetting a *super-resolution* (SR) problem, it is shown that the quality of SR images can be improved by using *multi-view* captures of the same scene, as similar patches are likely to be found between them. This implies that there is a high likelihood of finding relevant patches on projections of the same surface on each view introducing suitable geometrical constraints.

We use the *epipolar lines* $\boldsymbol{\ell}$ to establish the relations between the different images of the scene. An epipolar line is the 2D projection of a 3D ray that connects a point $\mathbf{X}$ in the world with the *optical center* of the camera through the position $\mathbf{x}$ in the image [**?**]. It can therefore be used to restrict the search space to a single line when looking for the matching point in another view. This constraint enhances the search speed when compared to a random search on the entire image. Assuming an image with a number N of pixels, our proposal requires in the worst case (i.e., when the epipolar line goes from the top-left corner to the bottom-right corner of the image) a computation time of $\mathrm{O}(\sqrt{N})$, while the full image search would require $\mathrm{O}(N)$.

To obtain the parameters of the epipolar lines we need the *fundamental matrices* $\mathbf{F}$ that relate each pair of images. They can be obtained by estimation or from a calibrated sequence, in which case we extract the *intrinsic* ($\boldsymbol{\Lambda}$) and *extrinsic* parameters ($\boldsymbol{\Omega}$, $\boldsymbol{\tau}$) from the *projection matrices* $\mathbf{P}$. Said parameters are used to find the *center of projection* $\mathbf{X}_{CoP}$ in world coordinates of the different cameras in the dataset. Then, by projecting the $\tilde{\mathbf{X}}_{CoP}$ of a camera to the remaining cameras of the setting, we obtain the *epipoles* $\tilde{\mathbf{e}}$. Therefore, the fundamental matrix between a source image I_s and a target image I_t can be found by applying $\mathbf{F}_{s,t} = \tilde{\mathbf{e}}_t \times \mathbf{P}_t\mathbf{P}_s^+$, where $\tilde{\mathbf{e}}_t$ is the projection of the source camera $\tilde{\mathbf{X}}_{CoP}$, and $\mathbf{P}_s^+$ is the *pseudoinverse* of $\mathbf{P}_s$.

With $\mathbf{F}_{s,t}$ we can find the epipolar line $\boldsymbol{\ell}_t$ in I_t for each point $\tilde{\mathbf{x}}_s$ of I_s using the equation $\boldsymbol{\ell}_t = \mathbf{F}_{s,t}\tilde{\mathbf{x}}_s$. More details can be found in [**?**].

To compute each of the points contained in the epipolar line, we use a parameterization of the line in order to make the prior suitable for all the possible lines that can be generated in a multi-view scenario. The calculation of a point $\mathbf{x} = (x, y)^T$ with the above mentioned parameterization is

$$\mathbf{x} = \mathbf{e} + \lambda\hat{\boldsymbol{u}}, \tag{1}$$

being $e_{x,y} = \tilde{e}_{x,y}/\tilde{e}_z$ the coordinates of the epipole $\mathbf{e}$, $\hat{\boldsymbol{u}}$ a unitary direction vector determined by $(b, -a)^T/\sqrt{a^2 + b^2}$, where a, b are the two first parameters of the epipolar line, and λ the depth parameter.

In order to make the prior robust to illumination changes between projections, we compare an illumination-invariant representation of the texture of the patch, i.e., the pixel values of the patch after subtracting its mean value.

As a demonstration of the validity of the prior, we reconstructed an entire image from just patches on the epipolar lines of all views. In Figure 1 we show the original and reconstructed image along with the other images of the dataset. In this example we find the best match for the marked patch in the line of the second view (red square), so we use its information to reconstruct the patch. As we can see, the reconstructed part is coherent with the rest of the image, proving that the epipolar prior is able to find good matches.

4. EFFICIENT SEARCH

Our goal is to find good matches from our multi-view prior with a reduced number of searches, in order to make the algorithm efficient towards real-time applications. Thus, a randomized search approach is better suited than an exhaustive search. To reduce the number of searches without missing potential good matches, we exploit the spatial coherency in a way that can be interpreted as a generalization of 2D propagation schemes in existing randomized search strategies.

The core algorithm in our proposed framework is based on a randomized search strategy similar to [**?**] with the patch

search constrained to the epipolar lines (our multi-view prior). As in the mentioned algorithm, our approach makes a random initialization to allow independent uniform samples across the different images of the dataset and along the epipolar lines. Then, it is followed by a number of iterations of our *circular propagation* and random-search steps in order to reach convergence of the results. The proposed circular propagation is effective when the different views of the scene have a wide baseline. If the baseline between the views is small, the epipolar lines are almost parallel and the search is more efficiently done with the classical propagation scheme in existing randomized search approaches.

Initialization. Let $\{\mathrm{I}_1, \mathrm{I}_2, \ldots, \mathrm{I}_K\}$ be our image dataset where we take I_s as the source image. We want to match every $m \times m$ patch $\mathcal{P}$ in I_s with another one in any of its epipolar lines in the rest of the images. Therefore, we give to every $\mathcal{P}$ of I_s a random depth (λ) value and a random image number *i*. The image number *i* determines the epipolar line where the patch is located by means of the line parameters *a* and *b*, and the λ its exact point inside this line (Equation 1).

Circular propagation. Let $\mathcal{P}_1$ be a source patch in the source image I_s with its central pixel located at $\mathbf{x}_s$ and its match $\mathcal{P}_1'$ determined by I_i and λ_i. Through a comparison of $\mathcal{P}_1'$ with the matches of the neighbors of $\mathbf{x}_s$, we want to improve the previously mentioned match. This comparison is made with the four neighboring patches of $\mathcal{P}_1$, which are the ones located one row above, one row below, one column before and one column after $\mathbf{x}_s$ (Figure 2(a)).

Let $\mathcal{P}_2$ be any of the neighboring patches of $\mathcal{P}_1$ and its current match $\mathcal{P}_2'$, determined by λ_j and located in I_j inside the corresponding epipolar line $\boldsymbol{\ell}_{2,j}$. We need to determine the position of the *candidate* patch $\mathcal{P}_c$ for $\mathcal{P}_1$.

Based on the assumption that a 3D scene is locally planar, we consider that a good neighbor match can be found in a local neighborhood of a detected good match. However, the difference in perspective between the cameras in wide-baseline setups require a generalization of the local coherency in 2D for the multi-view cases. Since we have the epipolar lines $\boldsymbol{\ell}_{2,j}$ and $\boldsymbol{\ell}_{1,j}$ we can determine the position of $\mathcal{P}_c$ by assuming the coplanarity of neighbor patches, effectively generalizing the previously mentioned 2D local coherency for these multi-view cases. The determination of the candidate's position by coplanarity can be seen in Figure 2(b). As the epipolar lines between neighbors are close to each other, we can relax coplanarity to equidistance from the epipole in order to allow a faster implementation. In Figure 2(c) we can see how the equidistance is applied as a relaxation of the coplanarity, resulting in a circular propagation.

The next step is to measure the error between $\mathcal{P}_1$ and $\mathcal{P}_c$ by computing the Euclidean distance between them and comparing it to the error between $\mathcal{P}_1$ and its previous match, $\mathcal{P}_1'$. If the candidate's error is smaller, we store the λ_j and I_j to make $\mathcal{P}_c$ the new best match. We repeat this process for all the four surrounding neighbors of $\mathbf{x}_s$ and proceed to the second step of the iteration.

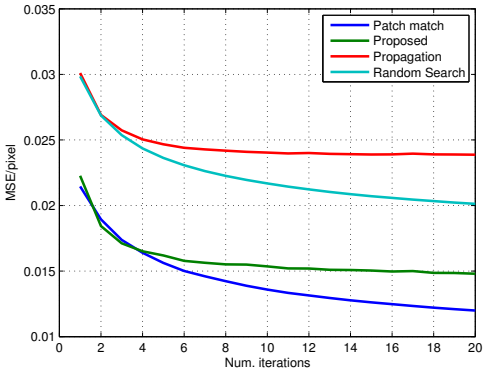


Fig. 3. MSE per pixel of the matches at each iteration when applying PatchMatch [**?**] and our approach. The *Propagation* and *Random Search* curves refer to the execution of our algorithm iterating exclusively the circular propagation or random search steps respectively.

Random search. As the circular propagation does not guarantee the found match to be the optimal in the line, we proceed to a random search in the line in order to determine if a better match can be found in there. The search is done iteratively reducing the search range in each step by α, which is a fixed ratio between search ranges (in our experiments, $\alpha = 0.7$). In each iteration, we give a new random value to λ in the equation 1 to determine the location of a new candidate patch $\mathcal{P}_c$. We keep the patch that presents the smaller error when compared to the source patch $\mathcal{P}_1$ in I_s. This process is repeated until the search range becomes smaller than 1.

5. EXPERIMENTAL RESULTS

For our experiments, we used the geometrically calibrated Inria 4D Repository datasets[1], consisting of cameras 0-4, frames 550, 641, 706 from the *dancer* dataset; cameras 4, 5, 9, 13, 15, frames 50, 108, 169 from the *martial* dataset; and cameras 4, 5, 9, 13, 14, frames 005, 107, 279 from the *dog* dataset. Due to lack of illumination, the *dog* dataset was gamma-corrected. The implementation was done in MATLAB, with the costliest functions implemented in OpenCL. The results of the applications presented in this section were obtained after 20 iterations of the corresponding algorithm (PatchMatch or the proposed) with 3×3 patches.

5.1. Validation

To show the validity of our approach, we compare the convergence of our search strategy with that of PatchMatch [**?**]. PatchMatch searches for matches in the entire image, while we only search in the points strictly contained inside the cor-

[1] http://4drepository.inrialpes.fr/public/datasets

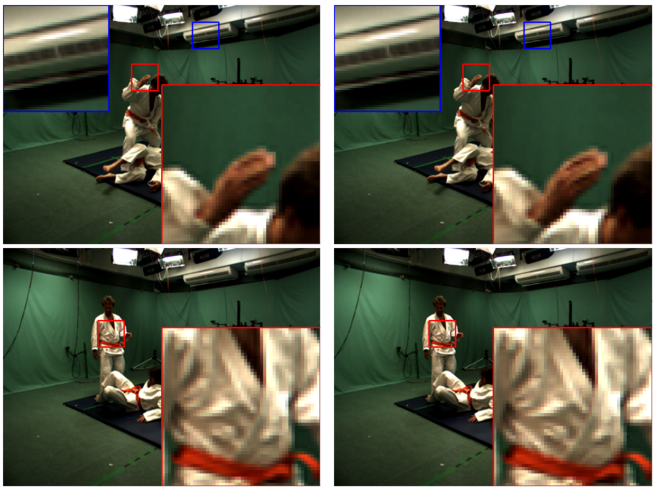


Fig. 4. Super-resolution results. First column, the result of applying a single-image search with PatchMatch; second column, the application of our search strategy through the rest of the images of the dataset. As we can see, our approach more realistically reconstructs the fingers of the raised hand in the first image or the creases in the clothing in the second image.

responding epipolar line. In order to make a fair comparison, we apply PatchMatch to all the images of the dataset excluding the source image. Then, we compare the results and search for the ones that have less error when compared to the patches of the source image. Both searches are applied up to twenty iterations in an attempt to show the asymptotic accuracy of our approach as compared to a full-range randomized search scheme. In Figure 3 we show the convergence curve for one of the scenes.

For limited run-time, the proposed restricted search can find better matches than PatchMatch. This is due to the fact that most of the zones of the used dataset are large, smooth regions and our prior presents higher probability of finding a similar patch inside a restricted line than if we search in all the image. However, searching for good matches exclusively inside the epipolar line drastically reduces the amount of matching candidates and reaches faster its limit. In the curves of the circular propagation and random search, we can see how the random search inside the line is the process giving better matches as the number of iterations is increased. However, the combination of circular propagation and random search can still achieve better matches.

5.2. Applications

We illustrate the usefulness of the proposed framework through two different applications. The first one is an example-based super resolution by high-frequency (HF) transfer [**?**], [**?**], where we enhance the single-image result by applying our method as a post-processing step. The second application consists in the deblurring of an out-of-focus image by using the information of the non-blurred images of the dataset.

Super resolution. To test the super resolution application, we start with a set $\{\mathrm{I}_1, \mathrm{I}_2, \ldots, \mathrm{I}_K\}$ of sharp images (that means, high- and low-frequency components) of the same scene. We take I_1 as the source image and apply a bicubic upscaling with magnification factor $s = 1.5$, resulting in the image U_1^L. The superscript L points that it is a low-frequency (LF) image. Then we use both search algorithms (PatchMatch and our proposal) to search the best match for each patch of U_1^L, taking as data images the LF spectrum of I_1 (I_1^L) for the patch match search (resulting in a *single-image* search) and the LF of $\mathrm{I}_2^L, \ldots, \mathrm{I}_k^L$ for our proposal (*multiple-image* search).

The super resolution image is built by taking U_1^L as the LF background and adding the best HF patches of the single-image search and the multiple-image search. That allows us to enhance the results of a single-image search using the additional information found in the other cameras. Some examples of the improvement in the HF components of the image can be found in Figure 4.

In Table 1 we show the average PSNR and SSIM of each upscaling method we applied when compared to the sharp image. Both PSNR and SSIM show that our proposal improves the results of a single-image approach.

Deblurring. We start with a set $\{\mathrm{B}_1, \mathrm{I}_2, \ldots, \mathrm{I}_K\}$, being B_1 a damaged (blurred) image that we want to fix with deblurring. In order to find the best matching patches to obtain I_1 (i.e., a sharp version of B_1), we compute a blurred

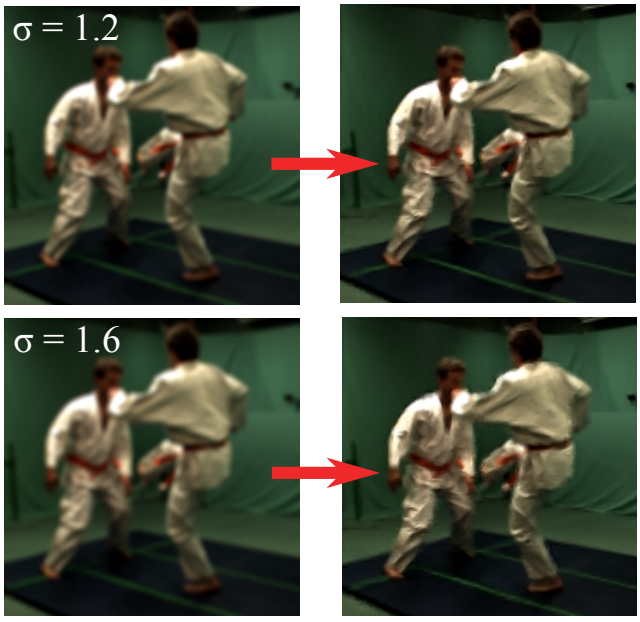


Fig. 5. Detailed view of deblurring results. We present the blurred images (left) with different Gaussian PSF and the deblurred result obtained applying our algorithm.

Method	PSNR	SSIM
Bicubic	39.26	0.9846
Single-image SR	40.04	0.9850
Our proposal	40.24	**0.9856**
Proposal + single-image SR	**40.31**	0.9855

Table 1. Super-resolution results. Average PSNR (dB) and SSIM of the different upscaling methods applied to the different frames of the sequence shown in Figure 4.

version of the rest of our input images to obtain the dataset $\{\mathrm{B}_2, \mathrm{B}_3, \ldots, \mathrm{B}_K\}$. This blurred version of our initial dataset is used to find matches for every patch in B_1. Considering $\mathcal{P}_{B_i}$ to be any of the matches found in any of our K-1 images, we reconstruct the image applying the next expression:

$$\mathcal{P}_{rec} = \mathcal{P}_{B_1} + \mathcal{P}_{I_i} - \mathcal{P}_{B_i}, \tag{2}$$

where $\mathcal{P}_{rec}$ is any patch of the reconstructed image and $\mathcal{P}_{B_1}$, $\mathcal{P}_{I_i}$, $\mathcal{P}_{B_i}$ are any of the patches of B_1, I_i and B_i respectively. The Equation 2 adds to every patch of B_1 the sharp information present in its best match $\mathcal{P}_{I_i}$ substracting the blurred part of it, present in B_i.

In Figure 5 we show a pair of blurred images with a Gaussian PSF of standard deviation $\sigma = 1.2$ and 1.6 respectively alongside with its corresponding deblurred results. In Table 2 we present the PSNR of different σ applied to the same input image along with the corresponding PSNR of its deblurring.

6. CONCLUSIONS

We have presented a new framework for multi-view settings which is based on a general and compact epipolar-constrained prior to efficiently exploit the geometrical properties of these scenarios.

We have also introduced a generalization of existing approximate nearest neighbor patch-search algorithms using the proposed prior. The efficient search adapts the bidimensional local coherency to wide-baseline multi-view scenarios by propagation under an assumption of local surface coplanarity, which can also be relaxed to the proposed circular propagation scheme. As an example of the applicability of our framework to different image processing problems, we show 1) a super resolution application that improves the results of single-image super resolution, and 2) how we can use the geometrical prior to deblur a damaged image in a multi-view dataset thanks to the remaining images.

As future work, we plan to study using homographies to enhance the quality of the results and the applicability of the method for other classical image processing problems such as image denoising.

	$\sigma = 1.2$	$\sigma = 1.3$	$\sigma = 1.5$	$\sigma = 1.6$
Blurred	31.28	30.87	30.28	30.07
Deblurred	33.64	33.13	32.31	31.89

Table 2. Deblurring results. PSNR (dB) of the blurred and deblurred image when compared to the original (without deblurring) for different standard deviation factors σ of a gaussian blurring kernel.

REFERENCES

[1] Y. Furukawa and J. Ponce, "Accurate, dense, and robust multiview stereopsis," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1362–1376, 2010.

[2] S. M. Khan and M. Shah, "A multiview approach to tracking people in crowded scenes using a planar homography constraint," in *European Conf. on Computer Vision*, 2006, pp. 133–146.

[3] G. Pons-Moll, A. Baak, J. Gall, L. Leal-Taixe, M. Mueller, H.P. Seidel, and B. Rosenhahn, "Outdoor human motion capture using inverse kinematics and von mises-fisher sampling," in *IEEE Int. Conf. on Computer Vision*, 2011, pp. 1243–1250.

[4] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *IEEE 12th Int. Conf. on Computer Vision*, 2009, pp. 349–356.

[5] E. Pérez-Pellitero, J. Salvador, J. Ruiz-Hidalgo, and B. Rosenhahn, "Bayesian region selection for adaptive dictionary-based super-resolution," in *Proc. of the British Machine Vision Conf.*, 2013.

[6] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, "PatchMatch: A randomized correspondence algorithm for structural image editing," *ACM Trans. on Graphics (Proc. SIGGRAPH)*, pp. 24:1–24:11, 2009.

[7] K. He and J. Sun, "Computing nearest-neighbor fields via propagation-assisted kd-trees," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2012, pp. 111–118.

[8] S. Korman and S. Avidan, "Coherency sensitive hashing," in *Proc. Int. Conf. on Computer Vision*, 2011, pp. 1607–1614.

[9] S.J.D. Prince, *Computer Vision: Models Learning and Inference*, Cambridge Uni. Press, 2012.

[10] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge Uni. Press, 2004.

[11] J. Salvador, E. Pérez-Pellitero, and A. Kochale, "Fast single-image super-resolution with filter selection," in *Proc. IEEE Int. Conf. on Image Processing*, 2013, pp. 640–644.

[12] G. Freedman and R. Fattal, "Image and video upscaling from local self-examples," *ACM Trans. Graph.*, pp. 12:1–12:11, 2011.