

A PRACTICAL GUIDE TO MULTI-IMAGE ALIGNMENT

Cecilia Aguerrebere[†]

Mauricio Delbracio^{*}

Alberto Bartesaghi[§]

Guillermo Sapiro[‡]

[†] Fundación Ceibal

^{*}Instituto de Ingeniería Eléctrica
Universidad de la República

[§]National Cancer Institute
National Institutes of Health

[‡] Electrical and Computer Engineering
Duke University

ABSTRACT

Multi-image alignment, bringing a group of images into common register, is an ubiquitous problem and the first step of many applications in a wide variety of domains. As a result, a great amount of effort is being invested in developing efficient multi-image alignment algorithms. Little has been done, however, to answer fundamental practical questions such as: what is the comparative performance of existing methods? is there still room for improvement? under which conditions should one technique be preferred over another? does adding more images or prior image information improve the registration results? In this work, we present a thorough analysis and evaluation of the main multi-image alignment methods which, combined with theoretical limits in multi-image alignment performance, allows us to organize them under a common framework and provide practical answers to these essential questions.

Index Terms— Multi-image alignment, Bayesian estimators

1. INTRODUCTION

Multi-image alignment consists in bringing a group of images into a common reference. It is an ubiquitous problem, being the first step of many applications, such as high dynamic range (HDR) imaging [1, 2], super-resolution [3, 4, 5], burst deblurring [6, 7] and burst denoising [8]. It appears in a wide variety of domains such as computational photography, biomedical imaging, astronomy, and many other remote sensing applications, where alignment errors highly affect the final result. For instance, in computational photography applications such as HDR imaging, ghosting artifacts that appear from incorrect alignment are extremely perturbing to the observer, often voiding completely the utility of the main technique. In biomedical imaging, alignment quality often limits the resolution of structures reconstruction, detection, or segmentation [9, 10, 11, 12].

Several techniques have been proposed to tackle the multi-image alignment problem and a great amount of effort is being invested in improving results even further. Still, other fundamental practical aspects of the multi-image registration problem have received much less attention. For instance, what is the comparative performance of the existing methods? how much improvement can we expect if we add more images or incorporate image priors? In this work, we present a comprehensive study of the main multi-image alignment methods that allows us to organize them under a common framework and provide answers to these essential questions. We focus on 2D rigid translations, which, despite being simple, are the basis of many models used in practice. For instance, in very complex scenes where objects move independently, we often make the sound hypothesis that the background suffers a 2D global translation, and even non-rigid transformations can be locally approximated by translations.

Multi-image shift estimation is interlaced with the estimation of the underlying image. In fact, a straightforward and the most popu-

lar approach for recovering the latent image, is by first estimating the multiple shifts and then averaging the *unshifted* noisy observations. There are other approaches that circumvent the shift estimation step and directly rely on shift-invariant features for estimating the latent image [13]. In this work, we focus on the problem of multiple shift estimation solely, even if as a side product we get an estimation of the underlying image. Indeed, there are many applications that estimating the shifts is an end in itself, since, for example, one could opt to combine the images later on in the image processing pipeline.

The simplest approach to multi-image shift estimation, is to do pairwise alignment between each image in the set and one chosen as reference. However, given that all the input images share the same underlying scene, they are not independent and doing a joint alignment, mostly under low signal-to-noise ratio (SNR) conditions where the pairwise alignment is very noisy, may improve the results. The main approaches to multi-image alignment include: the maximum likelihood estimator (MLE) with different optimization strategies [4, 14], the Bayesian MLE [15], the maximum a posteriori (MAP) estimator with different optimization strategies [4, 15, 16], and constrained alignment [17, 18].

Most of these algorithms were originally introduced for the super-resolution problem. For the multi-image alignment setting, we show that most of these approaches are mathematically equivalent in the sense that they optimize a very similar functional, the main difference being whether they include a prior image model or not, and which is the chosen optimization technique.

In [19], we presented a theoretical analysis of the fundamental limits in multi-image alignment and analyzed the performance of the MLE with respect to these bounds. Although, MLE achieves maximum performance in high SNR, there is a gap between theory and practice in medium to low SNR conditions. In this work, we conducted a thorough experimental analysis that help us close this performance gap on a wide range of SNR conditions. We show that more images and an image prior are extremely useful in low SNR, enabling alignment in very challenging conditions where it is otherwise not possible. Indeed, the compared methods that rely on prior image information perform very close to the theoretical bounds, showing that there is little room left for improvement.

This article is organized as follows. Section 2 details the analyzed methods, while Section 3 presents an empirical evaluation on real and synthetic data. Section 4 summarizes the conclusions.

2. MULTI-IMAGE ALIGNMENT METHODS

Let us consider the image acquisition model

$$z_i(\mathbf{x}) = u(\mathbf{x} - \boldsymbol{\tau}_i) + n_i(\mathbf{x}), \quad i = 0, \dots, K, \quad (1)$$

where $z_i(\mathbf{x})$ is the observed i -th image at pixel position $\mathbf{x} = [x, y]^T$, u is the underlying continuous image, $\boldsymbol{\tau}_i = [\tau_{ix}, \tau_{iy}]^T$ is the 2D translation vector of frame i with respect to u ($\boldsymbol{\tau}_0 = 0$), and $n_i(\mathbf{x})$ is independent additive Gaussian noise with variance σ^2 . In practice, we have access to the digital images $\mathbf{z}_0, \dots, \mathbf{z}_K$ sampled on a

Work partially supported by the ARO, ONR, NGA, AFORS, and NSF.

discrete grid. We will assume that all the images are band-limited and sampled according to the Nyquist sampling theorem. Let $\boldsymbol{\tau} = [\boldsymbol{\tau}_1^T, \dots, \boldsymbol{\tau}_K^T]^T$ be the concatenation of all 2D unknown translations, and $\mathbf{z} = [\mathbf{z}_0^T, \dots, \mathbf{z}_K^T]^T$ be the concatenation of the $(K+1)$ observed images. The goal of multi-image alignment techniques is then to estimate $\boldsymbol{\tau}$ from $\mathbf{z}$. Equivalently in the Fourier domain, the model becomes: $\tilde{z}_i(\boldsymbol{\omega}) = \tilde{u}(\boldsymbol{\omega})e^{-i\boldsymbol{\omega} \cdot \boldsymbol{\tau}_i} + \tilde{n}_i(\boldsymbol{\omega})$, where $\sim$ denotes 2D image Fourier transforms, $\boldsymbol{\omega} = [\omega_x, \omega_y]^T$ represents the 2D Fourier spatial frequency and $\cdot$ denotes the inner product operation.

The main methods for 2D multi-image translation estimation can be classified into two categories, depending on whether they include prior image information or not. Among methods that do not include an image prior we find: the MLE, with different optimization strategies [4], and constrained alignment methods [17, 18]. Among those including an image prior, we find the MAP estimator, with different optimization strategies [4, 15, 16], and the Bayesian MLE [15].

Maximum Likelihood Estimator: Given $(K+1)$ independent images following Model (1), and assuming u is an unknown deterministic image, the MLE of $[u, \boldsymbol{\tau}]^T$ is the value that maximizes the log-likelihood of the samples,

$$[u, \boldsymbol{\tau}]_{\text{MLE}} = \arg \max_{u, \boldsymbol{\tau}} [\log p(\mathbf{z}|u, \boldsymbol{\tau})] = \arg \min_{u, \boldsymbol{\tau}} \|\mathbf{z} - \mathbf{B}(\boldsymbol{\tau})u\|^2, \quad (2)$$

with the shift operator $\mathbf{B}(\boldsymbol{\tau}) = [B(\boldsymbol{\tau}_1)^T, \dots, B(\boldsymbol{\tau}_K)^T]^T$, where $B(\boldsymbol{\tau}_i)$ is the Shannon 2D interpolation operator [15] that verifies: $B(\boldsymbol{\tau}_i)u(\mathbf{x}) = u(\mathbf{x} - \boldsymbol{\tau}_i)$. The functional in (2) is an example of a separable non-linear least-square problem. Given the shifts $\boldsymbol{\tau}$, the unknown underlying image u is given by the least squares solution $\hat{u} = (K+1)^{-1}\mathbf{B}^T\mathbf{z}$, which is the average of the aligned frames. Inserting $\hat{u}$ back into (2), the functional depends on the shifts only

$$\boldsymbol{\tau}_{\text{MLE}} = \arg \min_{\boldsymbol{\tau}} \|\mathbf{z} - (K+1)^{-1}\mathbf{B}\mathbf{B}^T\mathbf{z}\|^2. \quad (3)$$

Functional (3) is non-convex and different approaches can be followed to find a local minimum [4]. We consider here two approaches: cyclic coordinate descent and variable projections. Cyclic coordinate descent consists in optimizing (3) on one coordinate at a time. Two main steps are iterated: first, compute the average of the frames aligned with the current estimate of the shifts (given by $\hat{u}$); second, find each shift by minimizing the Euclidean distance of the corresponding image against the average. Variable projections makes use of the fact that minimizing (3) is equivalent to

$$\boldsymbol{\tau}_{\text{MLE}_{vp}} = \arg \max_{\boldsymbol{\tau}} \mathbf{z}^T \mathbf{B} \mathbf{B}^T \mathbf{z}. \quad (4)$$

Robinson *et al.* [4] proposed to compute the MLE maximizing (4), a method they named variable projections. For the super-resolution problem, they claim that this optimization has several advantages compared to the cyclic coordinate descent approach since it converges in fewer iterations and its minima are better defined.

Bayesian Maximum Likelihood Estimator: Let us assume that the underlying image u can be modeled as a stationary zero-mean Gaussian process with spectral density $\mathbf{S}_u = \mathbf{F}^T \boldsymbol{\Sigma}_u \mathbf{F}$, where $\boldsymbol{\Sigma}_u$ is the covariance matrix and $\mathbf{F}$ the Fourier operator. Under this hypothesis, u can be considered a hidden variable and be marginalized from the samples log-likelihood. The unknown shifts can then be computed maximizing the marginal likelihood $p(\mathbf{z}|\boldsymbol{\tau}) = \int p(\mathbf{z}|u, \boldsymbol{\tau})p(u)du$, which is also a Gaussian function (see e.g., [20, Eq. (2.115)]), having zero mean and covariance matrix $\boldsymbol{\Gamma} = \sigma^2 \mathbf{I} + \mathbf{B}\boldsymbol{\Sigma}_u \mathbf{B}^T$. Then,

$$\boldsymbol{\tau}_{\text{BMLE}} = \arg \max_{\boldsymbol{\tau}} \tilde{\mathbf{z}}^T \tilde{\mathbf{B}} \tilde{\mathbf{W}} \tilde{\mathbf{B}}^T \tilde{\mathbf{z}}, \quad (5)$$

where $\tilde{\mathbf{B}} = \mathbf{F}\mathbf{B}\mathbf{F}^T$ is the Fourier equivalent of the shift operator, $\tilde{\mathbf{W}} = (\mathbf{S}_u(K+1) + \mathbf{I}\sigma^2)^{-1}\mathbf{S}_u$ is the Wiener filter, and $\mathbf{I}$ is the identity matrix. Notice that the only difference between the Bayesian maximum likelihood estimator (BMLE) (5) and the MLE (4) is the Wiener filter, which appears with the introduction of the image prior.

Maximum a Posteriori: An alternative way to incorporate an image prior is to compute the MAP estimator. Given $(K+1)$ independent samples following Model (1), and assuming u is drawn from a zero-mean Gaussian process with spectral density $\mathbf{S}_u$, the MAP of $[u, \boldsymbol{\tau}]^T$ is the value that maximizes the posterior probability

$$\begin{aligned} [u, \boldsymbol{\tau}]_{\text{MAP}} &= \arg \max_{u, \boldsymbol{\tau}} [\log p(\mathbf{z}|u, \boldsymbol{\tau}) + \log p(u)] \\ &= \arg \min_{u, \boldsymbol{\tau}} \frac{1}{2\sigma^2} \|\mathbf{z} - \mathbf{B}^T u\|^2 + \frac{1}{2} u^T \boldsymbol{\Sigma}_u^{-1} u. \end{aligned} \quad (6)$$

Similarly to functional (2), functional (6) is an example of a separable non-linear least-square problem. Following the same steps as for the MLE, it can be shown that minimizing (6) is equivalent to [4]

$$\boldsymbol{\tau}_{\text{MAP}} = \arg \max_{\boldsymbol{\tau}} \tilde{\mathbf{z}}^T \tilde{\mathbf{B}} \tilde{\mathbf{W}} \tilde{\mathbf{B}}^T \tilde{\mathbf{z}}, \quad (7)$$

where $\tilde{\mathbf{W}}$ is the Wiener filter as defined before. Hence, the MAP and the BMLE optimize the same cost function.

Common Framework. While often presented as different techniques, we have just showed that the previously presented multi-image alignment approaches, all optimize the cost function¹ $E(\boldsymbol{\tau}) = \tilde{\mathbf{z}}^T \tilde{\mathbf{B}} \tilde{\mathbf{W}} \tilde{\mathbf{B}}^T \tilde{\mathbf{z}}$, whether with $\tilde{\mathbf{W}} = \mathbf{I}$ for the methods without image prior or $\tilde{\mathbf{W}}$ equal to the Wiener filter for methods including an image prior. Nevertheless, given that this functional is non convex, results may vary depending on the optimization strategy and the initialization procedure.

Constrained Alignment. Another way of aligning multiple images is to use all possible pairwise estimations and use their redundancy to get more accurate and self-consistent shifts [17]. For this purpose, all the pairwise shifts b_{ij} between any two frames $i < j$ are first computed by locating the maximum value of their correlation map. Then, these estimates are combined making use of the fact that b_{ij} is the summation of the shift vectors of all intermediate adjacent frames, $r_i + r_{i+1} + \dots + r_{j-1} = b_{ij}$. The shifts r_i can then be found solving the over-determined set of linear equations $(K(K+1)/2)$ equations and K unknowns) obtained by determining all possible shifts between any two frames. A similar approach was proposed by Farsiu *et al.* [18], where they incorporate the coherence as a constraint in the optimization of the shifts directly.

Theoretical Performance Limits. In [19] we presented a theoretical study deriving different statistical performance bounds for the translations estimation accuracy in multi-image alignment. The Cramér-Rao bound, under the assumption that the underlying image follows a natural image prior (CRB) was derived. The CRB gives a lower bound on the mean square error (MSE) of any unbiased estimator of the shifts $\boldsymbol{\tau}$ (see [19, Eq. (44)]). Different behaviors for the alignment accuracy are identified, depending on the SNR of the input images. For very high SNR, the performance bound is independent of the number of images K or prior information (image or shifts prior), and the MLE attains the bound. Doing pairwise alignment using the MLE is optimal in this case. For high to moderate SNR, increasing the number of images does improve performance. Interestingly, theory predicts the existence of an SNR threshold below which performance degrades briskly and a lower limit SNR value

¹This equivalence is not necessarily valid when addressing the super-resolution problem (including blurring and subsampling operators).

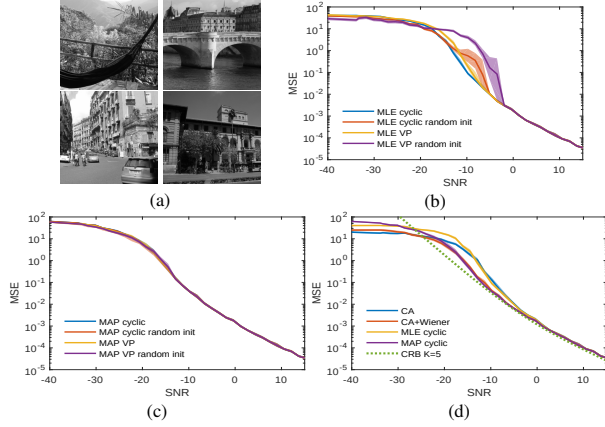


Fig. 1. (a) Natural images used as ground-truth (images from [19]). Comparison of optimization and initialization strategies of MLE (b) and MAP (c) with $K = 5$ images. (d) Performance comparison of MLE, MAP and the constrained alignment applied to Wiener filtered images. The CRB is displayed as a performance benchmark [19].

below which alignment is not possible. Nevertheless, increasing the number of images can push back these thresholds several dBs, making alignment possible in much more challenging conditions. For natural images, the MLE performance is close to the CRB but it is not tight. A possible reason for this is the non-optimality of the MLE, for which a critical drawback is that it does not use image prior information. Moreover, for this SNR region, the MLE performance clearly improves with increasing number of images. Including prior image information, for example through a Bayesian approach such as minimizing the expected MSE or computing the Maximum a Posteriori, could help close the gap between the fundamental limits and the MLE performance.

3. EXPERIMENTAL ANALYSIS

In what follows, we conduct a thorough experimental analysis to evaluate the presented techniques with both synthetic and real data.

3.1. Synthetic Data

Experimental setup: Sets of images are generated following Model (1), taking as ground-truth the examples in Fig. 1 (a) (50×50 pixels), with different noise levels and number of shifted images. Two types of motion are considered: uniformly distributed independent shifts and drift-driven trajectories (each image is shifted from the previous image position, according to prior angle and speed distributions), often observed in biomedical applications (e.g., cryo-electron microscopy) and image bursts capture with hand-held cameras, among others. The image prior for the MAP estimator is a stationary zero-mean Gaussian process with a spectral density that decays as the inverse of the squared Fourier frequency, a prior widely used for locally modeling natural images [19, 21, 22, 23, 24]. The experiments are repeated 100 times for each SNR level and the mean and 95% confidence intervals of the MSE are reported. The methods are almost unbiased (the squared bias was on average, in all experiments, three orders of magnitude smaller than the variance). The SNR is defined as the ratio between the total energy of the derivative of the ground-truth image and the noise power [19].

Optimization and initialization: The MLE and MAP were solved using both cyclic coordinate descent and variable projections. All

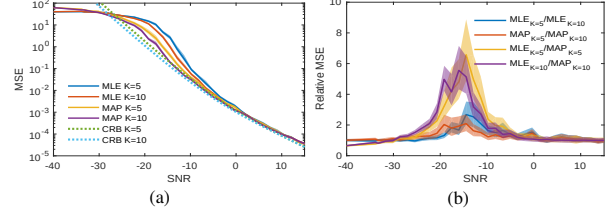


Fig. 2. (a) MLE versus MAP performance as a function of the input image SNR when using $K = 5$ and $K = 10$ images, and the respective CRB bounds. (b) Ratio of the performance of the different configurations shown in (a).

optimizations were initialized using random shifts as well as pairwise alignment (each image is aligned to the first image using classical correlation). Fig. 1 shows the results for the MLE (b) and MAP (c) with $K = 5$ images with a random shifts trajectory. For the MLE, in high SNR conditions, there are no significant performance differences between the optimizations nor the initialization strategies. For lower SNR levels, however, the cyclic coordinate descent optimization seems to be slightly better than variable projections, and the pairwise alignment initialization considerably outperforms the random initialization. For the MAP approach, no significant differences are found between the tested configurations at all SNR levels.

Image prior: Fig. 1 (d) compares the performance of the cyclic coordinate descent MLE and MAP, both initialized using pairwise alignment, and the constrained alignment estimator applied to Wiener prefiltered images (assuming the same image prior as for MAP). The results are compared to the Cramér-Rao lower bound (CRB) [19, Eq. (44)], which gives a lower bound on the variance of any unbiased estimator of the shifts, therefore establishing a performance benchmark. The CRB is computed assuming the same image prior as for the MAP estimator (see [19] for details). For very high SNR, all methods perform very similarly and close to the CRB, meaning that the extra information provided by the image prior is not useful. For low to moderate SNR levels, however, a clear performance improvement is observed with the MAP estimator which reduces the gap between the MLE and the limit predicted by the CRB. Even more importantly, including the image prior pushes back several dBs the SNR threshold after which alignment performance degrades dramatically (about -10dB for MLE to -14dB for MAP, see Fig. 1 (d)). Hence, including the image prior enables alignment in very challenging noise conditions where it is otherwise not possible.

Number of images: Fig. 2 shows a comparison of MLE and MAP with different number of images ($K = 5, 10$). Similarly to what was observed for the image prior, increasing the number of images has no effect in high SNR conditions. For lower SNR, increasing the number of images improves the results, both for MLE and MAP. It is interesting to remark, however, that the performance gain given by including the image prior is much larger than that of increasing the number of images. Indeed, the MAP estimator with 5 images performs considerably better than the MLE with 10 images. Fig. 2 (b) shows the ratio between the performance of the different configurations. Increasing the number of images produces a higher improvement for MLE (blue curve) than for MAP (red curve). But more importantly, the improvement is much larger when including the image prior (MLE/MAP ratio shown in the yellow and violet curves), showing a larger gain for smaller K . This is particularly interesting from a practical perspective given the increased complexity, and thus time requirements, of increasing the number of images as opposed to the almost costless inclusion of the appropriate image prior.

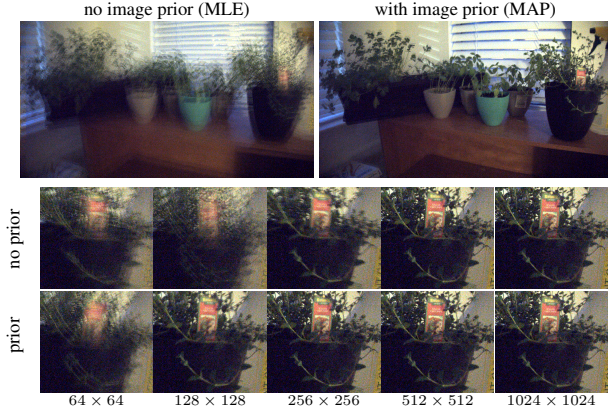


Fig. 3. Top: Comparison of the alignment with and without prior when doing the shift estimation using a patch size 128×128 and $K = 5$ images. **Bottom:** Alignment with increasing patch size (left to right) without (top) and with (bottom) image prior.

The previous results correspond to *bolivia* image (Fig. 1 (a) top left) with independent random shifts, but the same behavior is observed for the other examples and for drift-driven trajectories as well.

3.2. Real Data

To evaluate the applicability of the previous results to real data, we compare the alignment performance for the burst denoising task [8]. For burst denoising (a widely used and very powerful technique for noise reduction in low light conditions) correct alignment is essential to obtain good ghosting-free results.

Experimental setup: Image bursts are acquired in low light conditions with a hand-held Sony α -5100 camera set to ISO 20000. Burst denoising is performed by aligning and averaging groups of $K = 2, 5, 10$, raw images. Images are aligned using the MLE and MAP solved by cyclic coordinate descent, considering the same image prior as before. To have different SNR levels, alignment is performed using a sub-image (patch) of varying size: $64 \times 64, \dots, 1024 \times 1024$. The gradient content increases with the patch size, thus reproducing conditions of increasing SNR. The shifts are estimated using the red channel, then applied to all channels before demosaicking.

Effects of the image prior: Fig. 3 (top) shows an example of the results obtained with 5 images, where alignment is performed using a sub-region of 128×128 pixels. In this case, alignment is not possible without the image prior but it becomes feasible including it. Fig. 3 (bottom) shows an extract of the results obtained when using different patch sizes. The first and second rows show the result without and with image prior, respectively. For the two smallest patch sizes (64×64 and 128×128) the noise is very high and there is almost no gradient in the patch, representing cases of very low SNR. Even under these very challenging conditions, the image prior enables alignment with the 128×128 patch. This result confirms what was observed in Section 3.1, which showed that the SNR threshold below which alignment performance degrades dramatically can be pushed back by including an image prior. For the patch size 256×256 , the gradient content is already visible and alignment is possible with or without the image prior. Nevertheless, the quality difference is still clear. For patch sizes above 512, the difference is almost indistinguishable. This was also predicted by the results obtained in Section 3.1, since in high SNR all methods agree and

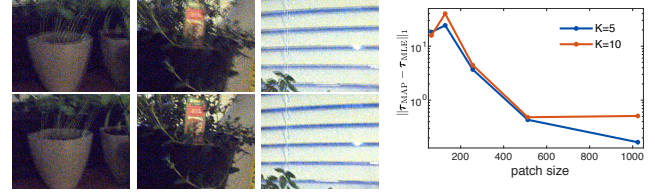


Fig. 4. Left: Extracts of the average of two images aligned using the shifts obtained with $K = 2$ (pairwise) and $K = 10$ images (top and down respectively). **Right:** Difference between MAP and MLE estimations for varying patch size and number of images.

attain the CRB [19]. Fig. 4 (right) shows the absolute value of the difference between the shift estimates obtained with and without image prior for the different patch sizes and different number of images (average of the difference in x and y directions). The estimations are very similar for patch sizes above 512.

Effects of the number of images: Fig. 4 (left) compares the quality of the estimated shifts with MAP using 2 (pairwise alignment) versus 10 images for a patch size of 128×128 . For this purpose we show the average of two images aligned using the shifts obtained with $K = 2$ (left) and using $K = 10$ (right). The result of the multi-image alignment is sharper than the pairwise alignment, showing the clear performance improvement obtained by using more images.

4. CONCLUSIONS

We conducted a thorough analysis on multi-image shift estimation methods. We showed that most of them use different optimization techniques to optimize the same functional, the main difference being the inclusion of prior image information. We then conducted an experimental analysis, that confirmed the per-region behavior depending on the SNR conditions predicted by a theoretical analysis on the fundamental limits of multi-image alignment performance [19]. In very high SNR, all the evaluated methods perform very similarly and very close to the Cramér-Rao lower bound. Hence, the simplest methods already achieve the best possible performance in this SNR condition and including prior image information or more images does not improve the alignment result. For moderate to low SNR, however, we show that there is a clear performance gain when including an image prior or using more images. This gain is twofold: the MSE is reduced and, more importantly, the threshold at which performance degrades dramatically is pushed back several dBs. Therefore, for these SNR conditions, including more images or an image prior makes alignment possible in conditions where it is otherwise not possible. The performance gain obtained by including the image prior is larger than that of increasing the number of images. This is important from a practical perspective given the increased complexity, and therefore time requirements, of increasing the number of images as opposed to the almost costless inclusion of the image prior. Indeed, we found that the methods that include the image prior perform very close to the CRB for a larger SNR range, showing that there is little room left for improvement as they close the gap between the MLE and the CRB [19] in moderate to low SNR. Regarding optimization and initialization, slight differences are observed for the MLE in low SNR conditions, but this does not seem to be a critical aspect. Finally, as predicted by theory [19], we observe the existence of an SNR threshold below which none of the evaluated methods manages to align the images, and neither more images nor an image prior can revert this situation. The only way out is increasing the SNR, e.g., increasing the image size (or patch size in case of local alignment).

5. REFERENCES

- [1] Paul E. Debevec and Jitendra Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. An. Conf. Comp. Grap. Inter. Tech. (SIGGRAPH)*, 1997, pp. 369–378.
- [2] Cecilia Aguerrebere, Julie Delon, Yann Gousseau, and Pablo Musé, "Simultaneous HDR image reconstruction and denoising for dynamic scenes," in *Proc. IEEE Int. Conf. Comput. Photogr. (ICCP)*, 2013, pp. 1–11.
- [3] Dirk Robinson and Peyman Milanfar, "Statistical performance analysis of super-resolution," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1413–1428, 2006.
- [4] Dirk Robinson, Sina Farsiu, and Peyman Milanfar, "Optimal registration of aliased images using variable projection with applications to super-resolution," *Comput. J.*, vol. 52, no. 1, pp. 31–42, 2009.
- [5] Ce Liu and Deqing Sun, "On Bayesian adaptive video super resolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 346–360, 2014.
- [6] Haichao Zhang, David Wipf, and Yanning Zhang, "Multi-image blind deblurring using a coupled adaptive sparse prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2013, pp. 1051–1058.
- [7] M. Delbracio and G. Sapiro, "Burst Deblurring: Removing camera shake through Fourier burst accumulation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2015, pp. 2385–2393.
- [8] Toni Buades, Yifei Lou, Jean-Michel Morel, and Zhongwei Tang, "A note on multi-image denoising," in *In Proc. of Local and Non-Local Approx. in Image Proc. (LNLA)*, 2009, pp. 1–5.
- [9] John L. Rubinstein and Marcus A. Brubaker, "Alignment of cryo-EM movies of individual particles by optimization of image translations," *J. Struct. Biol.*, vol. 192, no. 2, pp. 188–195, 2015.
- [10] Timothy Grant and Nikolaus Grigorieff, "Measuring the optimal exposure for single particle cryo-em using a 2.6Å reconstruction of rotavirus VP6," *eLife*, 2015.
- [11] Alberto Bartesaghi, Doreen Matthies, Soojay Banerjee, Alan Merk, and Sriram Subramaniam, "Structure of β -galactosidase at 3.2-Å resolution obtained by cryo-electron microscopy," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 111, no. 32, pp. 11709–11714, 2014.
- [12] Xueming Li, Paul Mooney, Shawn Zheng, Christopher R. Booth, Michael B. Braumfeld, Sander Gubbens, David A. Agard, and Yifan Cheng, "Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM," *Nat. Methods*, vol. 10, no. 6, pp. 584–590, 2013.
- [13] Tamir Bendory, Nicolas Boumal, Chao Ma, Zhizhen Zhao, and Amit Singer, "Bispectrum inversion with application to multireference alignment," *arXiv preprint arXiv:1705.00641*, 2017.
- [14] Afonso S. Bandeira, Moses Charikar, Amit Singer, and Andy Zhu, "Multireference alignment using semidefinite programming," in *Proceedings of the 5th conference on Innovations in theoretical computer science*. ACM, 2014, pp. 459–470.
- [15] Nathan A. Woods, Nikolas P. Galatsanos, and Aggelos K. Katsaggelos, "Stochastic methods for joint registration, restoration, and interpolation of multiple undersampled images," *IEEE Trans. Image Process.*, vol. 15, no. 1, pp. 201–213, 2006.
- [16] Russell C. Hardie, Kenneth J. Barnard, and Ernest E. Armstrong, "Joint map registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1621–1633, 1997.
- [17] Venu Madhav Govindu, "Lie-algebraic averaging for globally consistent motion estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2004, pp. 684–691.
- [18] Sina Farsiu, Michael Elad, and Peyman Milanfar, "Constrained, globally optimal, multi-frame motion estimation," in *Proc. Workshop Stat. Signal Process. IEEE*, 2005, pp. 1396–1401.
- [19] C. Aguerrebere, M. Delbracio, A. Bartesaghi, and G. Sapiro, "Fundamental limits in multi-image alignment," *IEEE Trans. Signal Process.*, vol. 64, no. 21, pp. 5707–5722, 2016.
- [20] Christopher M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.
- [21] Michael Elad and Yacov Hel-Or, "A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur," *IEEE Trans. Image Process.*, vol. 10, no. 8, pp. 1187–1193, 2001.
- [22] Rik Fransens, Christoph Strecha, and Luc Van Gool, "Optical flow based super-resolution: A probabilistic approach," *Comput. Vis. Image Und.*, vol. 106, no. 1, pp. 106–115, 2007.
- [23] Anat Levin, Yael Weiss, Frederic Durand, and William T. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2009, pp. 1964–1971.
- [24] Netalee Efrat, Daniel Glasner, Alexander Apartsin, Boaz Nadler, and Anat Levin, "Accurate blur models vs. image priors in single image super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2013, pp. 2832–2839.