PARAMETRIC REGISTRATION FOR MOBILE PHONE IMAGES

Xinxin Zhang[1], Christopher Gilliam[2] and Thierry Blu[1]

[1]Department of Electronic Engineering
The Chinese University of Hong Kong, Hong Kong

[2]School of Engineering
RMIT University, Melbourne, Australia

ABSTRACT

Image registration is a significant step in a wide range of practical applications and it is a fundamental problem in various computer vision tasks. In this paper, we propose a highly accurate and fast parametric registration method for mobile phone photos. The proposed algorithm is based on a fast and accurate elastic registration algorithm, the Local All-Pass (LAP) algorithm, which performs in a coarse-to-fine manner. At each iteration, the LAP displacement field is fitted by a parametric model. Thus the image registration problem is equivalent to finding a few parameters to describe the displacement field. The fitting step can be performed very efficiently by solving a linear system of equations. In terms of the fitting model, it is easy to change the type of models to do the parametric fitting for specific applications. Experimental results on both synthetic and real images demonstrate the high accuracy and computational efficiency of the proposed algorithm.

Index Terms— Image registration, Local All-Pass filters, Parametric fitting, Elastic registration, Quadratic polynomial function

1. INTRODUCTION

Image registration is the process of finding a geometric transformation between two Cartesian coordinate systems so as to align two or more images. Image registration plays a vital role in medical analysis [1, 2], remote sensing [3], computer vision [4] and many other fields. It is a fundamental problem for various vision tasks, such as image mosaic, image fusion, image super-resolution and video stabilization. In our daily life, this process is necessary when we want to compare or integrate information from a sequence of images captured from mobile phones.

Under the brightness constancy hypothesis [5], a pair of images to be registered are related by a displacement field,

$$I_1(x,y) = I_2(x + u_x(x,y), y + u_y(x,y)), \qquad (1)$$

where I_1 is regarded as the target image, I_2 is the source image, the complex number $u(x,y) = u_x(x,y) + iu_y(x,y)$ is the displacement field. Thus, the key task of image registration is to estimate $u(x,y)$ such that I_2 is aligned to I_1. The variables of (1) are shown in Fig. 1. The target image and the source image are shown in parts (a) and (b), respectively.

(a)   (b)   (c)   (d)

$I_1(x,y)$ $I_2(x,y)$ blending of I_1 and I_2 $\begin{pmatrix} u_x(x,y) \\ u_y(x,y) \end{pmatrix}$
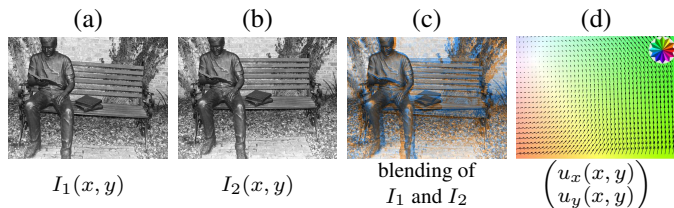
Fig. 1. The target image (a) is related to the source image (b) through the displacement field (d) according to Equation 1. An alternative representation to (a) and (b) is to blend these images as in (c): target is orange-toned, source is blue-toned, and perfect alignment is achieved when the blended image is tone-neutral (i.e., gray).

A blending of these images is then illustrated in part (c) - blue and orange tones illustrate mis-alignment. Finally, part (d) illustrates the displacement field; the colour indicates the direction of the displacement and the intensity of the colour indicates it's magnitude.

In this work, we target images captured by mobile phones under a similar lighting condition and camera setting. So we can safely assume the two images satisfy the brightness constancy. Taking photos from different view points is the main factor that generates geometric transformations. Because of the rigid motion of phone cameras, the transformations are global and rigid.

Various registration algorithms have been developed by researchers in recent years. Pixel-based methods use the intensity information of the whole image. Feature- or landmark-based methods extract a number of corresponding feature points or landmarks between a pair of images to be registered. The SIFT-based feature matching methods [6, 7, 8, 9], such as SURF [10], SIFT flow [11], CSIFT [12] are popular because of their invariance to translation, rotation and scaling. In general, the performance of these methods depends heavily on the feature types, the accuracy of the feature detection, the number of feature points or landmarks as well as the feature matching strategies. Feature-based methods may reduce the computational complexity and be able to cope with images with large intensity differences, but they usually perform poorly on feature-less images.

Image registration in essence is a mapping problem. Thus the mapping model selection is a quite important step. Most

existing models involve a small number of parameters, often modeling affine geometric transformations (shifts, rotations, scale changes, shears) [13, 14, 15]. Hence, the registration problem reduces to calculating very few parameters of the model which is fast. However, the accuracy of these low-order parametric methods is limited in real applications, even for the image pairs taken by mobile phones, because they do not model non-linear deformation effects. In order to cope with local and complex transformations, non-parametric (or high-order parametric) methods [16, 17, 18, 19, 20, 21, 22] are adopted to generate dense displacement fields, but the algorithms are prone to become unreliable and time consuming.

In most registration works, it is essential to optimize non-linear objective functions to estimate the parameters. Hence, the similarity measures and regularization are very important. There are many kinds of similarity measures, including sum-of-squared differences (SSD), normalized correlation coefficient (NCC) [23], correlation ratio (CR) and mutual information (MI) [24, 25]. In terms of the SSD measurement, intensity changes affect the similarity function. Mutual information is one of the most common methods for multimodal registration which needs heavy computation. The computation time and estimation accuracy are also closely related to the regularization terms and the optimizing algorithm.

In this paper, we propose a highly accurate and fast parametric registration method for photos taken by mobile phones. The proposed algorithm is based on a fast and accurate elastic registration algorithm, the local all-pass filtering algorithm (LAP), which performs in a coarse-to-fine hierarchical framework [26, 27]. At each iteration, we use a a parametric model to fit the displacement field estimated by the LAP. Thus the image registration problem is equivalent to finding a few parameters to describe the displacement field. The fitting step can be performed very efficiently by solving a linear system of equations. In terms of the parametric model, it is flexible to choose simple or complex models to do parametric fitting so as to deal with different types of displacement fields. The proposed method improves the accuracy of registration for mobile phone images comparing with the LAP algorithm.

2. LOCAL ALL-PASS REGISTRATION

The main idea of the LAP is that a constant shift is equivalent to filtering with an all-pass filter under the condition of brightness constancy, see Fig 2. The local displacement within a window is estimated using a local all-pass filter. A dense deformation estimate is then obtained by repeating this process for every pixel in the image. The LAP algorithm is implemented in a coarse-to-fine manner with respect to the filter size to deal with both large and small deformations. This algorithm is termed as poly-filter LAP.

Compared with the state-of-the-art elastic registration methods, the poly-filter LAP algorithm has significant advantages. First, this algorithm is very fast and accurate when the brightness constraint is exactly satisfied. In addition, it
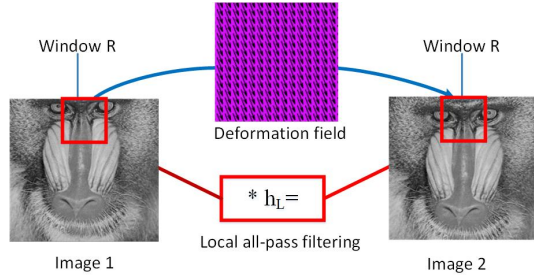
Fig. 2. Diagram illustrating the equivalence of constant deformation field and filtering with an all-pass filter in a window. The symbol $*$ denotes the convolution operator and h_L is the local all-pass filter of the window. Note that the locations of the two windows on the right-hand side and left-hand side are exactly the same.

is robust to noise. However, it becomes inaccurate when the illumination change violates the brightness constancy and the deformation is very large. Moreover, pre- and post-processing such as high-pass filtering, inpainting and smoothing are necessary for this algorithm.

As a preliminar study to the current paper, we showed how to increase the robustness of the registration by fitting iteratively a parametric expression to the output of the poly-filter LAP and then repeating the process to refine the estimate of model [28]. Such an approach however increases the computation time significantly and is relatively inefficient. Accordingly, it is these shortcomings we wish to remedy in the current paper.

3. REGISTRATION BY PARAMETRIC FITTING

In this section, we develop a method based on the LAP algorithm for mobile phone image registration. The main idea is to fit the output of the LAP at each iteration by a parametric model. Compared to [28] where the fitting is performed after an application of the full poly-filter LAP, our fitting is performed at each scale and the extrapolatory nature of the parametric representation allows to get rid of all the pre-processing, inpainting and post-processing steps that are otherwise needed in the poly-filter LAP algorithm. Thus the developed fitting algorithm is much more computationally efficient and, moreover, we observe that the accuracy achieved is at least as high as the one in [28].

3.1. Displacement field fitting

In general, the displacement field can be expressed by a combination of basis functions,

$$u(x,y) = \sum_{k=1}^{K} c_k u_k(x,y), \qquad (2)$$

where K is the total number of basis functions, c_k are coefficients and $u_k(x,y)$ are a basis of elementary displacements. Examples of such representations are global geomet-

ric transformations like linear transformations (shifts, rotations, shears, scaling), more general polynomial or Fourier series expressions, but also local transformations (e.g., expressed onto a local basis like a uniform B-spline basis [29]) and wavelet decompositions.

When it comes to registering images taken by a mobile phone, the displacement field becomes much simpler. The view point change causes rigid transformation which can be modelled as a constant or a first order polynomial function. To deal with more complicated situations, such as lens distortion and slight perspective, a quadratic polynomial function is sufficient to represent the transformation:

$$u(x,y) = c_1 + c_2 x + c_3 y + c_4 x^2 + c_5 y^2 + c_6 xy, \quad (3)$$

where c_1, c_2, c_3, c_4, c_5 and c_6 are unknown complex-valued coefficients.

As far as the displacement field is concerned, estimating the displacement on a global level boils down to calculating the six coefficients in (3). They are obtained from minimizing the difference between the displacement from the LAP, u_{LAP}, and u:

$$\min_{c_1,c_2,c_3,c_4,c_5,c_6} \sum_{x,y \in \Omega} |u(x,y) - u_{\text{LAP}}(x,y)|^2, \quad (4)$$

where Ω is the fitting region and it is described in the next part in detail. The solution to (4) is equivalent to solving a linear system of six equations which is quite fast, further avoid the risk of being stuck into local minimums. Once the coefficients are obtained, the dense displacement field is known.

The next step is to warp the source image closer to the target image according to the estimated parametric displacement field. Since the estimated displacement is non-integer, it is essential to build a continuous model of image for image warping. We adopt the shifted linear interpolation [30] for all the LAP windows larger than 5×5 pixels to obtain a high quality warped image while ensuring the high speed. However, at finer filter resolution, i.e. for LAP windows smaller than or equal to 5×5 pixels, we use cubic-OMOMS interpolation [31]—more accurate, albeit somewhat slower. The warped image is the new source image in the next iteration.

3.2. Fitting region determination

It is necessary to find the overlap regions between two images to maximize the similarity measures. However, determining reliable fitting regions and optimizing an objective function is a chicken-and-egg problem since the accurate overlap region is only known after a successful registration. Thank to the LAP algorithm, the errors in the displacement field are detected, because the LAP algorithm tags the displacement larger than the LAP resolution. Moreover, the pixels go out of the boundary of target image after warping are also excluded from the fitting region. Hence, the adaptive fitting region at each iteration is determined, which avoids the influence of outliers, shadows, artifacts and occlusions through the strat-

egy of erroneous exclusion. Although at the first iteration, i.e. the coarsest level, the determined fitting region is quite different from the true overlap region. As the iteration goes, the determined fitting region becomes closer to the true overlap region. The global parametrization makes it possible to extrapolate the displacement field from a limited region, and this is particularly useful in the case of large deformations where many algorithms usually fail. This strategy also has the same effects as pre- and post-processing in the poly-LAP algorithm.

4. EXPERIMENTAL RESULTS

In the experiments, registration methods are tested on synthetic images, the Oxford affine dataset [32] and real images taken by mobile phones. We compared our method with several existing state-of-the-art registration methods: three global parametric methods including a method using enhanced correlation coefficient maximization (ECC) [33], a parametric method optimizing normalized total gradient (NTG) [34], a method using a smoothly varying field to warp and orient features between two images (LAFP) [39], and seven well-known elastic image registration methods including the poly-filter LAP algorithm [27], an intensity-based image registration using residual complexity minimization from the Medical Image Registration Toolbox (MIRT) [35], the improved Demons algorithm based on the implementation in [36] and the elastic registration using a cubic B-spline free form displacement model implemented in ImageJ (bUnwarpJ) [18], an algorithm incorporating both geometric and intensity transformation (GIT) [16], and a feature-based image registration method, the SIFT flow method [11] and the implementation freely available in the Elastix toolkit that optimizing mutual information [37]. We implemented these methods based on the publicly available codes and their default parameters.

For our algorithm, the iteration number was set to 3 at each scale. The window size W of the finite impulse response filters of the LAP is determined according to the size of images, specifically, it is set to one quarter of $\min(\text{width}, \text{height})$ of the image. All algorithms were run on an Intel Core i7-5930K CPU @ 3.50 GHz with 64 GB RAM, using MATLAB R2013b.

To validate the robustness and accuracy of our algorithm, we first tested it on synthetic images (image source of size 512×768, taken from the LIVE data set [38]) deformed by various smooth displacement fields like affine transformations and transformations described by quadratic polynomial functions (maximum amplitude = 83 pixels). We also distorted the warped images by adding noise (PSNR = $20\,\text{dB}$), or by blurring them (variance = $5\,\text{pixel}^2$). Typically, the Median Absolute Error (MAE) and the Average Absolute Error (AAE) achieved by our algorithm were of the order 0.1 pixels in noisy conditions, while staying at 0.01 pixels in blurry conditions. When the images where not distorted, our algo-

**Table 1**. Comparison on the Oxford affine dataset [32]

		Bikes			Trees			Leuven		
		E_{Med}	E_{Mean}	Time	E_{Med}	E_{Mean}	Time	E_{Med}	E_{Mean}	Time
Global algorithms	This paper	**0.45**	**0.55**	**9.7**	**1.39**	**1.68**	**8.7**	**0.20**	**0.27**	**7.1**
	ECC [33]	0.47	0.60	17.8	2.05	3.32	14.5	0.21	**0.27**	13.0
	NTG [34]	0.87	0.98	15.5	1.71	1.97	15.6	0.62	0.89	32.9
	LAFP [39]	1.79	2.58	1933.4	4.31	5.24	2330.9	1.14	1.57	1446.2
Elastic algorithms	LAP [27]	0.63	0.78	11.3	1.69	3.22	9.4	0.33	0.43	7.9
	Demons [36]	21.64	22.10	27.4	3.92	7.28	12.7	0.56	1.58	25.8
	MIRT [35]	19.14	24.65	206.9	25.95	26.99	76.1	2.95	5.34	106.8
	bUnwarpJ [18]	1.02	1.12	18.0	2.15	3.00	24.3	0.34	0.40	24.7
	GIT [16]	1.38	3.84	982.9	2.18	4.66	962.0	0.48	0.96	948.7
	SIFT flow [11]	0.85	1.94	84.7	2.50	5.70	86.2	0.55	0.80	66.3
	Elastix [37]	6.69	12.38	299.9	3.78	10.17	319.5	0.44	1.46	307.0

(1) Bold values indicate the best results. (2) The results are calculated in the common region and averaged over 5 different image pairs in each subset.

rithm was almost exact (10^{-6} pixels) when we were using the exact same interpolation method as the one used to warp the images—and 0.01 pixels. The computation time was 8 seconds.

To demonstrate the accuracy and efficiency of the our algorithm in comparison with the state of the art, we tested our method on the publicly available Oxford affine dataset (Mikolajczyk et al. [32]) and several real outdoor images captured from iPhone, Huawei and Vivo smart camera phones (`http://www.ee.cuhk.edu.hk/%7Exxzhang/welcome_files/icip_exp.html`). The Oxford dataset includes an estimate of the ground truth spatial transformation between source and target images. We considered only three subsets of this dataset, involving blur and illumination changes (the other subsets involve displacements that are too extreme—1/3 to 1/2 of the image size—for all the compared image registration algorithms, including ours, to be successful). In order for our algorithm (and the others) to deal with illumination changes, we equalized the intensity of the source and target images by matching their histogram. This pre-processing was neutral for ECC, NTG, SIFT flow and Elastix (no benefit, no loss compared to doing nothing), but was useful for the other algorithms, including ours. The errors and computation time of different methods are listed in Table 1. From this table we can find that our method outperforms others in terms of accuracy and computation time. The algorithm that achieves results that are closest to ours is ECC. In relation to our previous work [28], the proposed algorithm obtains the same accuracy for up to a fifth of the computation time.

The image pairs shown in Fig. 3 (a) were captured with three different mobile phone cameras, from diverse viewpoints. Here, apart from our algorithm, most other methods fail to align these real images, although the geometric transformations involved seem to include mainly shift, rotation, scaling and a slight perspective. These examples demonstrate the ability of our algorithm to deal with both small and large geometric transformation, and show that its implementation is very efficient.

(a)   (b)   (c)
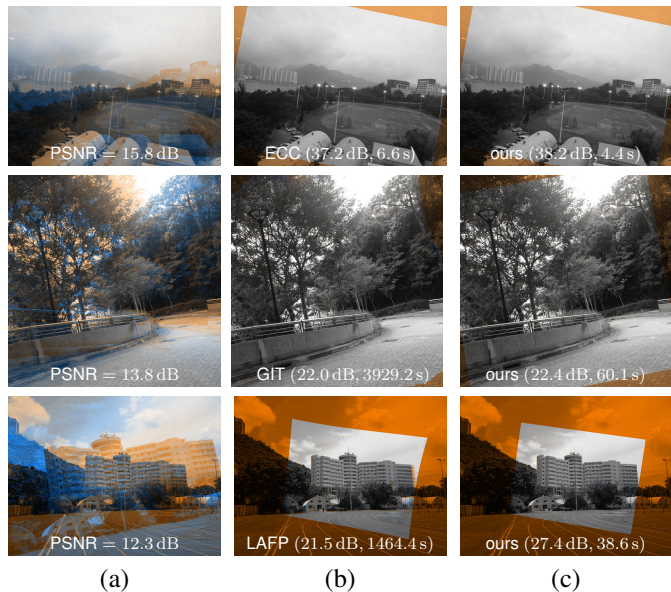
Fig. 3. Example of image registration on real images captured by three mobile phones. The images have been downscaled to 480×640 (top row), 1600×1600 (middle row), and 960×1280 (bottom rows). Column (a) shows the source and target image. In column (b), we show the result of methods that perform best in terms of PSNR, with their PSNR and computation time. (c) shows our results.

5. CONCLUSION

In this paper, we are proposing a very efficient and accurate parametric registration algorithm for images acquired by consumer cameras (e.g., mobile phones). The proposed algorithm is based on a fast and accurate elastic registration algorithm, the local all-pass filtering algorithm (LAP), which is iterated from large to small filter sizes, in a coarse-to-fine manner. At each iteration, the LAP displacement field is fitted by a parametric model. Hence, the image registration problem boils down to finding a few parameters to describe the displacement field. The fitting step can be performed very efficiently by solving a linear system of equations. An important advantage is that our image registration algorithm does not need any extra pre-processing of the images nor postprocessing of the displacement field. We have demonstrated that, on synthetic and real experiments our algorithm reaches high sub-pixel alignment accuracy and is significantly faster than other state-of-the-art algorithms.

6. REFERENCES

[1] A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey," *IEEE Trans. MI*, vol. 32, pp. 1153–1190, 2013.

[2] B. Zitova and J. Flusser, "Image registration methods: a survey," *Img. Vis. Comp.*, vol. 21, pp. 977–1000, 2003.

[3] S. Dawn, V. Saxena, and B. Sharma, "Remote sensing image registration techniques: A survey," in *Int. Conf. Img. Sig. Process. (ICSIP)*, 2010, pp. 103–112.

[4] X. Shen, L. Xu, Q. Zhang, and J. Jia, "Multi-modal and multi-spectral registration for natural images," in *Proc. ECCV*, 2014, pp. 309–324.

[5] S. Baker, D. Scharstein, J.P. Lewis, S. Roth, M. J Black, and R. Szeliski, "A database and evaluation methodology for optical flow," *Int. J. Comput. Vis.*, vol. 92, no. 1, pp. 1–31, 2011.

[6] Q. Li, G. Wang, and J. Liu, "Robust scale-invariant feature matching for remote sensing image registration," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, pp. 287–291, 2009.

[7] J. Ma, W. Qiu, and J. Zhao, "Robust L_2E estimation of transformation for non-rigid registration," *IEEE Trans. SP*, vol. 63, pp. 1115–1129, 2015.

[8] A. Kelman, M. Sofka, and C.V. Stewart, "Keypoint descriptors for matching across multiple image modalities and non-linear intensity variations," in *Proc. IEEE CVPR*, 2007, pp. 1–7.

[9] J. Chen, J. Tian, N. Lee, J. Zheng, R.T. Smith, and A.F. Laine, "A partial intensity invariant feature descriptor for multimodal retinal image registration," *IEEE Trans. BME*, vol. 57, no. 7, pp. 1707–1718, 2010.

[10] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Proc. ECCV*, 2006, pp. 404–417.

[11] C. Liu, J. Yuen, A. Torralba, J. Sivic, and W.T. Freeman, "SIFT flow: Dense correspondence across different scenes," in *Proc. ECCV*, 2008, pp. 28–42.

[12] A.E. Abdel-Hakim and A.A. Farag, "CSIFT: A SIFT descriptor with color invariant characteristics," in *Proc. IEEE CVPR*, 2006, vol. 2, pp. 1978–1983.

[13] R. Woods, S. Cherry, and J. Mazziotta, "Rapid automated algorithm for aligning and reslicing PET images," *J. Comp. Ass. Tomo.*, vol. 16, pp. 620–633, 1992.

[14] R. Woods, S. Grafton, and C. Holmes, "Automated image registration: I. General methods and intrasubject, intramodality validation," *J. Comp. Ass. Tomo.*, vol. 22, pp. 141–154, 1998.

[15] F. Albu, "Low complexity image registration techniques based on integral projections," in *Int. Conf. Syst. Sig. Img. Process. (IWSSIP)*, 2016, pp. 1–4.

[16] S. Periaswamy and H. Farid, "Elastic registration in the presence of intensity variations," *IEEE Trans. MI*, vol. 22, pp. 865–874, 2003.

[17] R. Bajcsy and S. Kovacic, "Multiresolution elastic matching," *CVGIP*, vol. 46, pp. 1–21, 1989.

[18] I. Arganda-Carreras, C. O. S. Sorzano, and R. Marabini, "Consistent and elastic registration of histological sections using vector-spline regularization," in *Proc. Int. Worksh. Comput. Vis. App. to Med. Img. Anal. (CVAMIA)*, 2006, pp. 85–95.

[19] S. Klein, M. Staring, and K. Murphy, "Elastix: a toolbox for intensity-based medical image registration," *IEEE Trans. MI*, vol. 29, pp. 196–205, 2010.

[20] S. Periaswamy and H. Farid, "Medical image registration with partial data," *Med. Imag. Anal.*, vol. 10, pp. 452–464, 2006.

[21] A. Goshtasby, "Image registration by local approximation methods," *Img. Vis. Comp.*, vol. 6, pp. 255–261, 1988.

[22] J. Kybic and M. Unser, "Fast parametric elastic image registration," *IEEE Trans. IP*, vol. 12, pp. 1427–1442, 2003.

[23] M.M. Fouad, R.M. Dansereau, and A.D. Whitehead, "Image registration under illumination variations using region-based confidence weighted m-estimators," *IEEE Trans. IP*, vol. 21, no. 3, pp. 1046–1060, 2012.

[24] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. MI*, vol. 16, no. 2, pp. 187–198, 1997.

[25] P. Viola and W.M. Wells III, "Alignment by maximization of mutual information," *Int. J. Comput. Vis.*, vol. 24, no. 2, pp. 137–154, 1997.

[26] C. Gilliam and T. Blu, "Local all-pass filters for optical flow estimation," in *Proc. IEEE ICASSP*, 2015, pp. 1533–1537.

[27] C. Gilliam and T. Blu, "Local all-pass geometric deformations," *IEEE Trans. IP*, vol. 27, no. 2, pp. 1010–1025, 2018.

[28] X. Zhang, C. Gilliam, and T. Blu, "Iterative fitting after elastic registration: An efficient strategy for accurate estimation of parametric deformations," in *Proc. IEEE ICIP*, September 17–20, 2017, pp. 1492–1496.

[29] J. Kybic and M. Unser, "Fast parametric elastic image registration," *IEEE Trans. IP*, vol. 12, no. 11, pp. 1427–1442, 2003.

[30] T. Blu, P. Thévenaz, and M. Unser, "Linear interpolation revitalized," *IEEE Trans. IP*, vol. 13, pp. 710–719, 2004.

[31] T. Blu, P. Thévenaz, and M. Unser, "MOMS: Maximal-order interpolation of minimal support," *IEEE Trans. IP*, vol. 10, no. 7, pp. 1069–1080, 2001.

[32] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A comparison of affine region detectors," *Int. J. Comput. Vis.*, vol. 65, no. 1-2, pp. 43–72, 2005.

[33] G. D. Evangelidis and E. Z. Psarakis, "Parametric image alignment using enhanced correlation coefficient maximization," *IEEE Trans. PAMI*, vol. 30, pp. 1858–1865, 2008.

[34] S.J. Chen, H.L. Shen, C. Li, and J.H. Xin, "Normalized total gradient: a new measure for multispectral image registration," *IEEE Trans. IP*, vol. 27, no. 3, pp. 1297–1310, 2018.

[35] A. Myronenko and X. Song, "Intensity-based image registration by minimizing residual complexity," *IEEE Trans. MI*, vol. 29, pp. 1882–1891, 2010.

[36] H. Lombaert, L. Grady, and X. Pennec, "Diffeomorphic demons: Efficient non-parametric image registration," *NeuroImage*, vol. 45, pp. S61–S72, 2009.

[37] S. Klein, M. Staring, K. Murphy, M.A. Viergever, and J.P.W. Pluim, "Elastix: a toolbox for intensity-based medical image registration," *IEEE Trans. MI*, vol. 29, no. 1, pp. 196–205, 2010.

[38] L. Cormack H. R. Sheikh, Z. Wang and A. C. Bovik, "LIVE image quality assessment database," http://live.ece.utexas.edu/research/quality.

[39] Y. Matsushita, "Aligning images in the wild," in *Proc. IEEE CVPR*, 2012, pp. 1–8.