Using Local, Contextual, and Deep Convolutional Neural Network Features in Image Registration

Raju Shrestha
raju.shrestha@oslomet.no
OsloMet - Oslo Metropolitan University
Oslo, Norway

ABSTRACT

Image registration is a well-known problem that arises in many applications in the fields of computer vision, remote sensing, and medical imaging. Many registration methods have been proposed in the literature. However, no single method works well in all kinds of images. In this work, local features and context-based augmented features are used in order to improve the accuracy of the image registration. Furthermore, an attempt has been made to use deep convolutional neural network features on top of those features for further improvement. The paper presents comparative results on image registration with and without feature augmentation and the deep convolutional neural network features. The results from the methods on a widely used benchmark dataset from the University of Oxford confirm improvement in the accuracy of image registration when local and augmented features are used.

CCS CONCEPTS

- **Computing methodologies** → **Neural networks**; **Matching**.

KEYWORDS

image registration; neural network; SIFT feature; Local features; Regional features; Contextual features; CNN features

ICICA '20, February 26–28, 2020, Brisbane, Australia

ACM ISBN 978-1-4503-7703-4. . . $15.00
https://doi.org/10.1145/12345.67890

1 INTRODUCTION

Images of the same scene, when acquired by different imaging devices (or sensors) or at different times or from different positions, may not be well aligned. Accurately aligning such images, the process known as image registration, is crucial in many applications in various fields such as computer vision, remote sensing, medical imaging.

Image registration is a fundamental task for many computer vision applications, such as object tracking, image stitching, image fusion, 3D reconstruction, etc [24]. High precision image registration is critical in military applications such as to improve precision strike capabilities [7, 8], in agriculture applications such as to record vegetation growth conditions [37], and in medical imaging such as to obtain the exact location of organs and tissues from multiple modalities such as CT, MR, or PET [20, 28].

The image registration process involves spatially transforming the test image(s), referred to as moving or source, to align them with the reference image, referred to as the target or fixed image. Ideally, mage registration methods should be invariant to scaling, illumination, noise, and geometric deformation [26]. Basically, image registration algorithm can be classified into two major classes: intensity-based and feature-based [14]. Intensity-based methods compare intensity patterns in images, while feature-based methods find correspondence between image features such as keypoints, lines, and contours. Because of their superior performance, we focus mainly on feature-based methods in this article. Based on the transformation model, the image registration algorithm can also be classified into two classes: rigid and non-rigid. Rigid image registration is global in nature, which uses linear transformations that include scaling, translation, rotation, and other affine transforms. Non-rigid image registrations are used when there is a possibility of localized complex deformations in the images and global rigid image registration doesn't work well [31]. This article is focused on rigid image registration of planar images as this has many applications, mainly in computer vision and remote sensing. Here on, the term image registration is used to refer to feature-based rigid image transformation, unless otherwise explicitly stated.

An image registration algorithm basically involves four steps: 1. Keypoint (point of interest) detection in each image, 2. Invariant feature description, a vector containing the keypoints' essential characteristics 3. Feature matching between the test and the reference images using some dissimilarity measure between descriptors and finding a projective transformation matrix, called homography matrix, which is used to spatially transform the test image [15], and 4. Warp the test image using the homography matrix to register with the reference image.

Many different methods are being proposed in the literature for keypoint detection and feature description [19, 32, 34]. The scale-invariant feature transform (SIFT) [19] is arguably the dominant algorithm for both keypoint detection and feature description. Bay et al. [4] proposed the speed up robust feature (SURF), which uses Hessian matrices and distributed descriptions without needing image sub-sampling, thus effectively reducing descriptor dimensionality and significantly improving speed. Rosten and Drummond [29] proposed the features from the accelerated segment test (FAST) algorithm, which was aimed at real-time detection of features. FAST compares surrounding pixels to obtain keypoints using machine learning. Calonder et al. proposed the binary robust independent elementary features (BRIEF) by using feature dimensionality reduction methods such as principal component analysis (PCA) and linear discriminant analysis (LDA), reducing the time needed to generate feature descriptors [6]. Rublee et al. combined FAST and BRIEF and proposed the oriented FAST and rotated BRIEF (ORB) as an efficient alternative to SIFT or SURF [30]. ORB is rotation invariant and robust to noise. Alcantarilla et al. [1] proposed the Accelerated-KAZE (AKAZE), a fast multi-scale feature detection and description approach for non-linear scale space, based on KAZE features [2]. AKAZE is also scale and rotation invariant.

Once the keypoints are detected and features are described, they are then used to match between the test and the reference images, the process called feature-matching. Feature matching methods traditionally use the k-nearest neighbor (kNN) [3] or the brute force algorithm [18] to match the features points. Then an outlier detection algorithm such as random sample consensus (RANSAC) [13] is used to eliminate mismatches or outliers. Zhang et al. [36] proposed a geometric-constraint based feature matching method. Kahaki et al. [17] proposed an invariant feature matching method, which is effective under different image deformations, by measuring the dissimilarity of the features through the path based on eigenvector properties. Lu et al. [20] considered a similarity transformation for the misalignment between two images for robust keypoint mappings on multispectral images.

Keypoints corresponding to inlier features in the two images are then used to estimate the homography matrix, which is then used to transform the test image to obtain the registered image. Many image registration methods have been proposed in the literature [38]. With the success of deep learning in many computer vision tasks such as image classification, object detection, and segmentation, many researchers have tried to use it in image registration as well. Dosovitskiy et al. [11] proposed an unsupervised feature learning with a convolutional neural network (CNN). Yang et al. [35] proposed a CNN feature-based multi-temporal remote sensing image registration method, which is designed to gradually increase the selection of inliers to improve the robustness of feature points registration. DeTone et al. [10] also proposed a deep image homography estimation, called HomographyNet, using the VGG-like network (VGG is a deep learning neural network architecture used in image recognition [33]), with 8 convolutional layers and two fully connected layers. As it uses a supervised approach, this requires labeled pairs of data (ground truth), which is not easy to obtain on real data. Nguyen et al. [27] proposed an unsupervised deep homography estimation model using VGG network, which uses a photometric loss function adapted to the unsupervised approach. Despite many image registration methods being available, no single method works perfectly in all kinds of images.

In this article, we proposed a novel registration method that combines keypoint features from the original image and deep convolutional neural network (DCNN) features. Furthermore, we use a new feature matching method, called ContextDesc, which is based on cross-modality context-based local descriptor augmentation, proposed by Luo et al. [21]. This method augments local and high-level regional features by aggregating the cross-modality contextual information, including visual context from high-level image representation, and geometric context from 2D keypoint distribution. The augmented feature helps to find more inlier matches, which in turn can help in accurate image registration. ContextDesc method performed the best in the recent image matching challenge at the image matching workshop, Computer Vision and Pattern Recognition (CVPR) conference, 2019. The method is briefly described in the next section.

2 CONTEXTUAL FEATURE AUGMENTATION

ContextDesc method has two main modules: preparation and augmentation. In the preparation module, it extracts K keypoints from the input image of dimension $H \times W \times 3$ using SIFT and $K \times 128$ local features using a lightweight 7-layer CNN from Luo et al. [22]. Furthermore, $\frac{H}{32} \times \frac{W}{32} \times 2048$ regional features are extracted using ResNet-50 [16], a DCNN. In the augmentation module, a geometric context encoder submodule generates 128-d feature vectors from the K keypoints. Another submodule, visual context encoder, produces K augmented features from the regional features. Finally, an aggregated K augmented features are obtained by combining the three different features: local features, geometric context features, and visual context features by element-wise summation and L2-normalization, thus keeping the feature dimensionality unaltered. For more details about the method, we refer to the original paper [21].

3 PROPOSED IMAGE REGISTRATION METHOD

The proposed method uses the ContextDesc feature augmentation method, which augments the standard SIFT features with geometric and visual contextual features in order to achieve better feature matching between two images. The method also uses CNN features from a DCNN architecture for further improvement in image registration accuracy. DCNN has a cognitive capability close to the human level and exhibited remarkable performance in image recognition and object tracking.

To incorporate CNN features, an appropriate, pretrained DCNN architecture can be used. Then selected CNN layers from the network are used to extract augmented features from each CNN feature image from these layers. These CNN features are combined with the features from the original image. The pseudo-code of the algorithm is given below.

```
features ← Get_features(given_image)
for layer_l in selected_layers
  features_l←empty
  for cnn_image_i in CNN_images
    features_i←Get_features(cnn_image_i)
    features_l← features_l + features_i
features←Combine(features, features_l)
Remove_redundancy(features)
```

In this work, one of the most recent DCNN architectures, the Xception [9], which uses cross-channel (or cross-feature map) correlation, is used. Three batch normalized convolution layers namely block1_conv1_bn, block1_conv2_bn, and block2_sepconv1_bn are used. As the size of the feature map images from the layers further deep down is small and found to have no significant contribution, they are not used. Both reference and test input images are pre-processed and resized to 299 × 299 × 3 according to the format required by Xception before feeding them into the network.

The combined features thus extracted from test and reference images are then used in feature matching between the two images, and outliers are removed using the random sample consensus (RANSAC) algorithm [13]. The keypoints corresponding to the remaining inliers are then used to estimate the homography matrix [12]. Using the estimated matrix, the test image is warped and registered to the reference image.

4 EXPERIMENTS

The performance of the proposed method is evaluated using the Oxford dataset [25], a benchmark dataset from the Visual Geometry Group, University of Oxford, and four different metrics. The dataset and the evaluation metrics are described below.

Dataset: The Oxford dataset [25] consists of eight different sets of images as shown in Figure 1. Each set of images contains one reference image and five different test images acquired under various different imaging conditions: viewpoint, scale, blur, illumination, and JPEG compression, from low to high changes. The dataset includes ground truth homography matrices for each reference-test image pair.

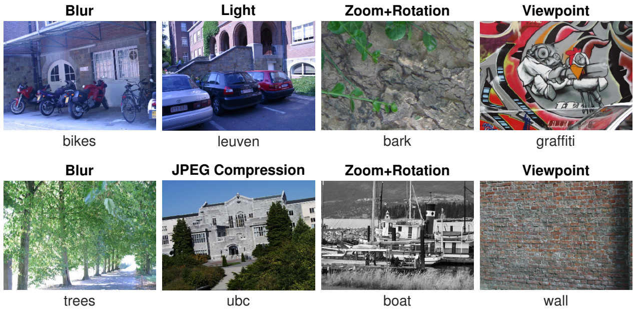


Figure 1: Eight sets of images used in experiments. Texts below the images show the names of the image set and above show the changes in the imaging conditions in the image set.

Evaluation metrics: The performance of an image registration method is measured based on the number and correctness of match points. A match point is considered as correct if it is within four pixels distance (in line with [5, 23] from the ground truth point, which is obtained using the ground truth homography matrix. The following four evaluation metrics are used to evaluate the performance.

Number of match points and inliers: Number of match points is the total number of match points detected by a method, and inlier is the remaining number of match points after the outliers are removed from the total.

Accuracy: is calculated as a ratio of the correct match points to the total number of match points.

Error: is measured as a median of the Euclidean distance (in pixels) of the match points from the ground truth.

Results from the methods which use SIFT features, Local features (LF), and Augmented features (AF) both with and without CNN features are used to compare their performance. The software program for the experiment was developed in Python 3.7 and executed in an Apple MacbookPro with Intel Core i7 CPU 2.7GHz and 32GB RAM. Code from [21] was adapted for the extraction of local and augmented features and used the same ratio of 0.8 for SIFT, and 0.89 for LF and AF in finding good matches.

5 RESULTS AND DISCUSSION

Table 1 shows the average values of the four metrics: error, accuracy, match points, and inlier, produced by the six different registration methods (SIFT, LF, AF, SIFT+CNN, LF+CNN, and AF+CNN) with the eight sets of images in the dataset. The results are also shown graphically in Figures 2, 3, 4, and 5.

Table 1: Average metric values from different methods on the eight image sets.

Metric	Method	bark	bikes	boat	graffiti	leuven	trees	ubc	wall	Average
Error	SIFT	2.85	0.21	2.10	250.81	0.25	1.69	0.25	72.10	41.28
	LF	2.81	0.23	3.78	170.30	0.27	1.43	0.23	1.54	22.57
	AF	2.73	0.28	1.70	5.01	0.28	1.72	0.17	1.41	1.66
	SIFT+CNN	2.83	0.24	2.05	211.68	0.22	0.97	0.23	1.09	27.41
	LF+CNN	2.84	0.21	1.36	137.86	0.24	1.08	0.17	1.30	18.13
	AF+CNN	3.15	0.23	0.78	6.73	0.26	1.38	0.30	1.14	1.75
Accuracy	SIFT	0.95	1.00	0.86	0.51	1.00	0.96	1.00	0.80	0.88
	LF	0.96	1.00	0.74	0.85	1.00	1.00	1.00	0.92	0.93
	AF	0.94	1.00	0.89	0.69	1.00	0.98	1.00	0.90	0.92
	SIFT+CNN	0.96	1.00	0.87	0.52	1.00	0.98	1.00	0.92	0.91
	LF+CNN	0.96	1.00	0.82	0.68	1.00	0.95	1.00	0.93	0.92
	AF+CNN	0.92	1.00	0.94	0.80	1.00	0.96	1.00	0.94	0.95
Match points	SIFT	481	600	1179	419	811	482	1727	2208	
	LF	563	756	1757	633	1002	706	2311	2857	
	AF	521	773	1684	604	1014	743	2405	2807	
	SIFT+CNN	484	610	1181	438	818	500	1743	2226	
	LF+CNN	585	782	1779	679	1028	756	2373	2911	
	AF+CNN	545	808	1699	647	1041	803	2473	2863	
Inliers	SIFT	26	30	30	14	74	11	179	99	
	LF	29	35	29	13	68	11	173	110	
	AF	28	34	30	13	73	11	173	122	
	SIFT+CNN	26	28	29	14	71	11	179	103	
	LF+CNN	27	31	27	13	74	11	156	112	
	AF+CNN	26	33	30	14	71	11	174	112	

Results show that LF and AF methods perform better than SIFT with all the eight image sets both in terms of accuracy and error values. Performance of SIFT is close to these methods in some image sets, however, it failed badly in some other image sets, particularly with graffiti and wall, where there are big viewpoint changes. Among LF and AF methods, AF performs equally or better on the average and in almost all image sets. Also, LF and AF methods detect the similar number of match points, but significantly higher compared to by SIFT. All three methods produced a similar number of inliers. This infers that a higher number of match points detected by LF and AF helps to achieve more correct inliers and that leads to better accuracy and reduced error.

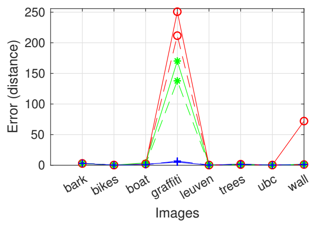


Figure 2: Error.

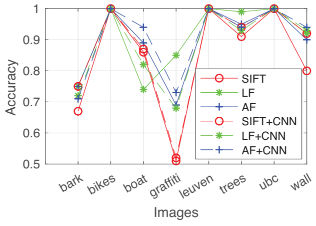


Figure 3: Accuracy.

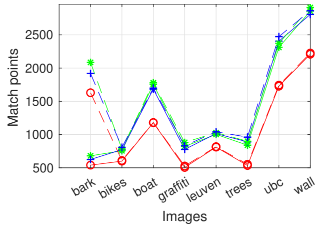


Figure 4: Number of match points.

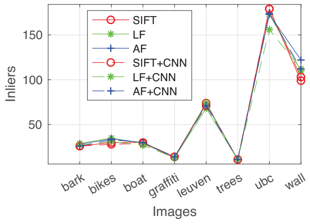
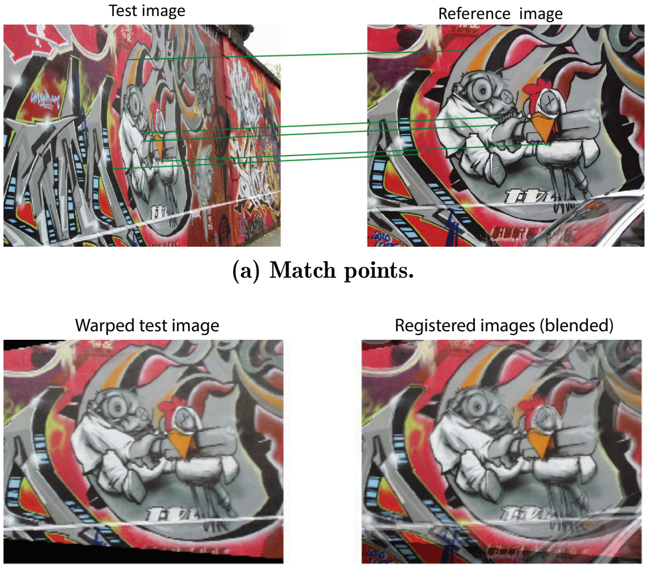


Figure 5: Number of inliers.

Adding CNN features to SIFT, LF, and AF methods, in general, improves the results. However, the improvement is not so large except in some cases such as SIFT+CNN reduced average error in case of wall image set from 72 down to 1, and increased accuracy from 0.8 to 0.92. The small improvement with LF+CNN and AF+CNN could be because LF and AF already incorporate some CNN features as they use high-level regional features extracted using ResNet-50 [16].

Among the five different imaging condition changes: blur, zoom+rotation, light, JPEG compression, and viewpoint, most of the methods are capable of dealing reasonably well with the first four changes. But SIFT failed completely with large viewpoint changes. In the meantime, LF, AF, LF+CNN, and AF+CNN have shown robustness towards viewpoint changes as well. As an illustration, Fig. 6 shows matching and registration results from AF+CNN in case of a big viewpoint change (image5) in the graffiti image set.

Even though accurate image registration is the main concern in this work, computation time can be an important issue in many applications, particularly in real-time applications. It is worth noting here that extraction of CNN features takes significantly long time, and considering small improvement, additional CNN features may not be practicable in time-critical applications.



(a) Match points.

(b) Warped and registered images.

Figure 6: Illustration of matching and registration of a complex image from the graffiti image set.

6 CONCLUSION

The research results show that feature-based rigid image registration methods, which use invariant features such as SIFT, work well in many images, but they are not robust enough as they may fail in case of complex image pairs with extreme imaging condition changes, viewpoint change in particular. Use of local features and contextual augmented features have shown improved and robust results. Further adding deep convolutional neural network features takes significantly long computation time, however with not so significant improvement in the accuracy of image registration.

REFERENCES

[1] Alcantarilla, P., Nuevo, J., and Bartoli, A. 2013. Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces. In *Proceedings of the British Machine Vision Conference*. BMVA Press.

[2] Alcantarilla, P. F., Bartoli, A., and Davison, A. J. 2012. KAZE Features. In *Computer Vision – ECCV 2012*, Andrew Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato, and Cordelia Schmid (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 214–227.

[3] Altman, N. S. 1992. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. *The American Statistician* 46, 3 (1992), 175–185. https://doi.org/10.1080/00031305.1992.10475879

[4] Bay, H., Ess, A., Tuytelaars, T., and Gool, L. V. 2008. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding* 110, 3 (2008), 346 – 359. https://doi.org/10.1016/j.cviu.2007.09.014 Similarity Matching in Computer Vision and Multimedia.

[5] Berengolts, A. and Lindenbaum, M. 2006. On the Distribution of Saliency. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 29, 12 (Dec. 2006), 1973–1990. https://doi.org/10.1109/TPAMI.2006.249

[6] Calonder, M., Lepetit, V., Ozuysal, M., Trzcinski, T., Strecha, C., and Fua, P. 2012. BRIEF: Computing a Local Binary Descriptor Very Fast. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34, 7 (July 2012), 1281–1298. https://doi.org/10.1109/TPAMI.2011.222

[7] Can, T., Karali, A. O., and Aytaç, T. 2011. Detection and tracking of sea-surface targets in infrared and visual band videos using the bag-of-features technique with scale-invariant feature transform. *Appl. Opt.* 50, 33 (Nov. 2011), 6302–6312. https://doi.org/10.1364/AO.50.006302

[8] Chen, J., Luo, L., Liu, C., Yu, J.-G., and Ma, J. 2016. Nonrigid registration of remote sensing images via sparse and dense feature matching. *J. Opt. Soc. Am. A* 33, 7 (Jul 2016), 1313–1322. https://doi.org/10.1364/JOSAA.33.001313

[9] Chollet, F. 2017. Xception: Deep Learning with Depthwise Separable Convolutions. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1800–1807. https://doi.org/10.1109/CVPR.2017.195

[10] DeTone, D., Malisiewicz, T., and Rabinovich, A. 2016. Deep image homography estimation. arXiv:1606.03798.

[11] Dosovitskiy, A., Springenberg, J. T., Riedmiller, M., and Brox, T. 2014. Discriminative Unsupervised Feature Learning with Convolutional Neural Networks. In *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger (Eds.). Curran Associates, Inc., 766–774. http://papers.nips.cc/paper/5548-discriminative-unsupervised-feature-learning-with-convolutional-neural-networks.pdf

[12] Dubail, M., Darzord, C., and Boust, C. 2009. Study of Contemporary Art Preservation with Digitization. In *Archiving 2009*. IS&T, Washington, USA, 47–52.

[13] Fischler, M. A. and Bolles, R. C. 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Commun. ACM* 24, 6 (June 1981), 381–395. https://doi.org/10.1145/358669.358692

[14] Goshtasby, A. A. 2005. *2-D and 3-D Image Registration: for Medical, Remote Sensing, and Industrial Applications*. Wiley.

[15] Hartley, R. I. and Zisserman, A. 2004. *Multiple View Geometry in Computer Vision* (2nd ed.). Cambridge University Press, ISBN: 0521540518.

[16] He, K., Zhang, X., Ren, S., and Sun, J. 2016. Deep Residual Learning for Image Recognition. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[17] Kahaki, S. M. M., Nordin, M. J., Ashtari, A. H., and Zahra, S. J. 2016. Invariant Feature Matching for Image Registration Application Based on New Dissimilarity of Spatial Features. *PLOS ONE* 11, 3 (March 2016), 1–21. https://doi.org/10.1371/journal.pone.0149710

[18] Kapela, R., Gugala, K., Sniatala, P., Swietlicka, A., and Kolanowski, K. 2015. Embedded platform for local image descriptor based object detection. *Appl. Math. Comput.* 267

(2015), 419 – 426. https://doi.org/10.1016/j.amc.2015.02.029 The Fourth European Seminar on Computing (ESCO 2014).

[19] Lowe, D. 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* 60, 2 (2004), 91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

[20] Lu, Y., Gao, K., Zhang, T., and Xu, T. 2018. A novel image registration approach via combining local features and geometric invariants. *PLOS ONE* 13, 1 (01 2018), 1–18. https://doi.org/10.1371/journal.pone.0190383

[21] Luo, Z., Shen, T., Zhou, L., Zhang, J., Yao, Y., Li, S., Fang, T., and Quan, L. 2019. ContextDesc: Local Descriptor Augmentation with Cross-Modality Context. *Computer Vision and Pattern Recognition (CVPR)* (2019).

[22] Luo, Z., Shen, T., Zhou, L., Zhu, S., Zhang, R., Yao, Y., Fang, T., and Quan, L. 2018. GeoDesc: Learning Local Descriptors by Integrating Geometry Constraints. In *The European Conference on Computer Vision (ECCV)*.

[23] Lv, G., Teng, S. W., and Lu, G. 2016. Enhancing SIFT-based image registration performance by building and selecting highly discriminating descriptors. *Pattern Recognition Letters* 84 (2016), 156 – 162. https://doi.org/10.1016/j.patrec.2016.09.011

[24] Mendelowitz, S., Klapp, I., and Mendlovic, D. 2013. Design of an image restoration algorithm for the TOMBO imaging system. *J. Opt. Soc. Am. A* 30, 6 (Jun 2013), 1193–1204. https://doi.org/10.1364/JOSAA.30.001193

[25] Mikolajczyk, K. and Schmid, C. 2005. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27, 10 (Oct. 2005), 1615–1630. https://doi.org/10.1109/TPAMI.2005.188

[26] Moravec, H. P. 1977. Techniques Towards Automatic Visual Obstacle Avoidance. (1977).

[27] Nguyen, T., Chen, S. W., Shivakumar, S. S., Taylor, C. J., and Kumar, V. 2018. Unsupervised Deep Homography: A Fast and Robust Homography Estimation Model. *IEEE Robotics and Automation Letters* 3, 3 (July 2018), 2346–2353. https://doi.org/10.1109/LRA.2018.2809549

[28] Oliveira, F. P. and Tavares, J. M. R. 2014. Medical image registration: a review. *Computer Methods in Biomechanics and Biomedical Engineering* 17, 2 (2014), 73–93. https://doi.org/10.1080/10255842.2012.670855 PMID: 22435355.

[29] Rosten, E. and Drummond, T. 2005. Fusing points and lines for high performance tracking. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, Vol. 2. 1508–1515 Vol. 2. https://doi.org/10.1109/ICCV.2005.104

[30] Rublee, E., Rabaud, V., Konolige, K., and Bradski, G. 2011. ORB: An efficient alternative to SIFT or SURF. In *2011 International Conference on Computer Vision*. 2564–2571. https://doi.org/10.1109/ICCV.2011.6126544

[31] Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L. G., Leach, M. O., and Hawkes, D. J. 1999. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging* 18, 8 (Aug 1999), 712–721. https://doi.org/10.1109/42.796284

[32] Schmid, C., Mohr, R., and Bauckhage, C. 2000. Evaluation of Interest Point Detectors. *International Journal of Computer Vision* 37, 2 (01 Jun 2000), 151–172. https://doi.org/10.1023/A:1008199403446

[33] Simonyan, K. and Zisserman, A. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556 [cs.CV].

[34] Tuytelaars, T. and Mikolajczyk, K. 2008. Local Invariant Feature Detectors: A Survey. *Foundations and Trends® in Computer Graphics and Vision* 3, 3 (2008), 177–280. https://doi.org/10.1561/0600000017

[35] Yang, Z., Dan, T., and Yang, Y. 2018. Multi-Temporal Remote Sensing Image Registration Using Deep Convolutional Features. *IEEE Access* 6 (2018), 38544–38555. https://doi.org/10.1109/ACCESS.2018.2853100

[36] Zhang, J., Chen, L., Wang, X., Teng, Z., Brown, A. J., Gillard, J. H., Guan, Q., and Chen, S. 2014. Compounding Local Invariant Features and Global Deformable Geometry for Medical Image Registration. *PLOS ONE* 9, 8 (Aug. 2014), 1–11. https://doi.org/10.1371/journal.pone.0105815

[37] Zhu, J., Wang, L., Yang, R., Davis, J. E., and pan, Z. 2011. Reliability Fusion of Time-of-Flight Depth and Stereo Geometry for High Quality Depth Maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33, 7 (July 2011), 1400–1414. https://doi.org/10.1109/TPAMI.2010.172

[38] Zitová, B. and Flusser, J. 2003. Image registration methods: a survey. *Image and Vision Computing* 21, 11 (2003), 977 – 1000. https://doi.org/10.1016/S0262-8856(03)00137-9