2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA) | 978-1-6654-1914-7/20/$31.00 ©2021 IEEE | DOI: 10.1109/MeMeA52024.2021.9478607

Deep Convolutional Feature-Based Fluorescence-to-Color Image Registration

Xingxing Liu
Department of Electrical and Computer Engineering
Iowa Technology Institute
The University of Iowa
Iowa City, Iowa, USA
xingxing-liu@uiowa.edu

Tri Quang
Department of Electrical and Computer Engineering
Iowa Technology Institute
The University of Iowa
Iowa City, Iowa, USA
tri-quang@uiowa.edu

Wenxiang Deng
Department of Electrical and Computer Engineering
Iowa Technology Institute
The University of Iowa
Iowa City, Iowa, USA
dengwenxiang@gmail.com

Yang Liu
Department of Electrical and Computer Engineering
Iowa Technology Institute
The University of Iowa
Iowa City, Iowa, USA
yang-liu-ece@uiowa.edu

***Abstract*—Fluorescence imaging has been widely utilized in various clinical applications. As a functional imaging modality, NIR fluorescence imaging often does not offer sufficient structural details. Therefore, structural imaging such as color reflectance overlaid with fluorescence imaging represents a superior approach for surgical visualization. Image registration of color reflectance and NIR fluorescence is needed for accurate overlay. In this study, we have implemented a deep convolutional algorithm for feature-based fluorescence-to-color image registration. Software-hardware codesign was conducted. Several sets of experiments were performed on biological tissues to compare the performance of our algorithm and traditional methods. We have demonstrated the feasibility of deep convolutional feature-based fluorescence-to-color image registration. To our best knowledge, this is the first demonstration of deep learning-based image registration between fluorescence and color imageries.**

***Keywords*—Deep learning, image registration, fluorescence imaging, computer vision, intraoperative imaging, multimodal imaging**

I. Introduction

Fluorescence imaging has been widely utilized in various clinical applications. For example, surgeons use fluorescence imaging to guide tumor resection and sentinel lymph node mapping [1-2]. Indocyanine green (ICG) is the most popular fluorophore used for fluorescence imaging owing to its low toxicity and high quantum yield. As a functional imaging modality, NIR fluorescence imaging often does not offer sufficient structural details. Structural imaging such as color reflectance overlaid with fluorescence imaging represents a superior approach for surgical visualization [3].

Image registration of color reflectance and NIR fluorescence is needed for accurate overlay. Conventional fluorescence imaging systems use a beam splitter calibrated for precise geometrical alignment of different image sensors, as a hardware-based approach [1-3]. However, beam splitters are large, heavy, and expensive optical components, making them inappropriate for application in compact imaging systems. Software-based image registration methods are promising alternatives.

Feature-based image registration finds features such as edges, corners, lines, curves, regions, templates, and patches from images to establish point-by-point correspondences and derives transformation for image registration [4]. Traditional feature-based image registration algorithms utilize Scale-Invariant Feature Transform (SIFT) [5], Speeded Up Robust Features (SURF) [6], Binary Robust Invariant Scalable Keypoints (BRISK) [7], and Oriented FAST and Rotated BRIEF (ORB) [8]. More specifically, keypoints with location, scale, and orientation information are extracted and described by a feature descriptor for discriminative feature matching between the reference and target images.

Deep learning is a powerful tool for computer vision and image processing tasks, such as object detection, segmentation, and registration. Many of deep learning-based image registration algorithms train a network to learn the transformation of the target image to the reference image. For example, in [9], DeTone et al. proposed a Regression HomographyNet that learns the homography and the CNN model parameters simultaneously in an end-to-end fashion. However, the groundtruth homography between target images and reference images in the training set needs to be available, which is not the case for the task of registering fluorescence images to color images. Therefore, application of deep networks for feature extraction is a more practical idea. Yang et al. proposed a non-rigid registration method that uses intermediate layers of a pre-trained VGG network to generate a feature descriptor that keeps both convolutional information and localization capabilities for remote sensing [10]. Though it outperforms SIFT in the registration of multi-temporal remote sensing images, its performance is not satisfactory for our fluorescence image registration task. We have

optimized this algorithm to improve the performance in fluorescence-to-color image registration. We have made two main improvements to the algorithm: 1) feature maps with higher resolution were used to represent feature points, which improved the accuracy of keypoint correspondence; 2) a filtering strategy was deployed to remove subpar keypoint matches.

Fluorescence imagery usually does not share similar image features with color imagery. To improve the common features between these two modalities for an accurate feature-based registration, we previously developed an approach that implemented additional reflectance information to the fluorescence image [11]. In the current study, we have further developed the hardware to enable the separation of added reflectance component from the true fluorescence component. This can facilitate feature-based image registration and preserve a high signal-to-background ratio of fluorescence imaging.

II. Methods

A. Hardware Setting

Fig. 1 illustrates the system setup. Two board-level 640×480 CMOS image sensors are used as image detectors. One sensor is used for color reflectance imaging, and the other is filtered by a bandpass filter (832 ± 37 nm) (Edmund Optics, NJ, USA) for NIR fluorescence imaging. A white LED (Edmund Optics, NJ, USA) and a NIR LED (Edmund Optics, NJ, USA) are mounted on a breadboard, which can be placed on a tripod. The system can capture NIR fluorescence and color reflectance images concurrently.

We characterized the field of view (FOV) of the sensors and the excitation light power distribution of the NIR light source. We measured the horizontal and vertical edges of image frames captured by the sensor A at three different working distances, including 200 mm, 300 mm, and 400 mm to depict the FOV of the sensor. Sensors A and B are the same, so their FOVs should be the same. In addition, we studied fluorescence excitation light distribution at working distances of 200 mm, 300 mm, and 400 mm with a USB power meter (PM16-120, Thorlabs, Newton, NJ, USA). The excitation power was measured along a pair of horizontal and vertical lines parallel to the imaging frame, in which the cross point was identified at the peak value of the NIR illumination.

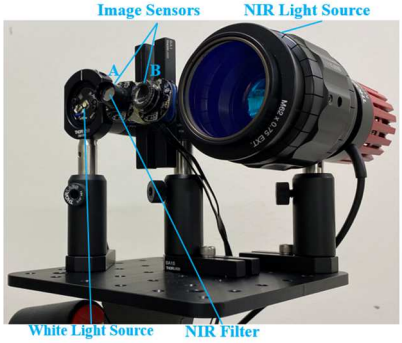

Fig. 1. Custom dual-modal imaging system.

B. Deep Convolutional Feature-Based Remote Sensing Image Registration

In [10], Yang et al. proposed a deep convolutional feature-based image registration algorithm to align remote sensing images. They first detected two sets of keypoints *X* and *Y* from the reference image *IX* and the target image *IY*, respectively. Subsequently, they used an expectation maximization (EM)-based procedure to obtain the transformed locations of *Y* (referred to as *Z*). *Y* and *Z* are used to solve a thin plate spline (TPS) interpolation for image transformation.

For each keypoint detected from the input image, the authors constructed a deep convolutional feature descriptor based on the output of certain layers of a pretrained VGG-16 network [12]. VGG is a family of deep convolutional networks trained on ImageNet with more than 1.2 million images, which are classified into 1000 categories. It is relatively deep and trained on a large dataset, thus achieves excellent performance on feature extraction. VGG is frequently used for feature extraction in various computer vision tasks. Fig. 2 shows the architecture of a slightly modified VGG-16 network. The original output of VGG should be a 1000×1 vector indicating categories the objects contained in the input image are classified. Since only the output of several intermediate layers is used for constructing the descriptor, layers after the last pooling layer 'pool5_1' are ignored.

The input of the VGG-16 network is a concatenated image of the reference image *IX* and the target image *IY*. More specifically, *IX* and *IY* are first resized to 224×224×3 RGB images, which are then concatenated into a 2×224×224×3 vector and passed into the network. The biggest advantage of this is that the output feature

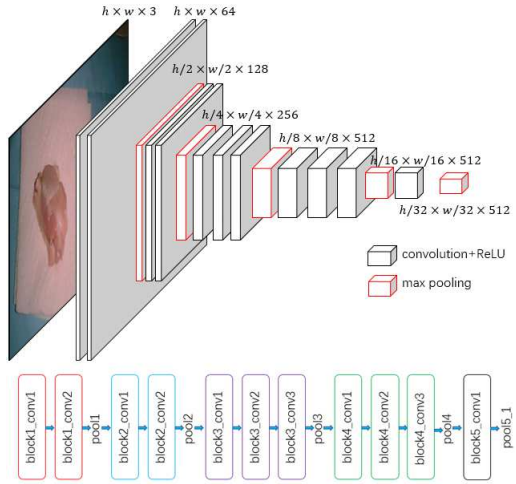

Fig. 2. Modified VGG-16 network architecture. *h* and *w* represent the height and width of the input image [10].

maps of *IX* and *IY* can be obtained simultaneously. Output of layers 'pool3', 'pool4', and 'pool5_1' are chosen to construct the descriptor; their corresponding sizes are 28×28×256, 14×14×512, and 7×7×512, respectively. F_1 denotes the output feature map of layer 'pool3'. Expanded feature map F_2 and F_3 can be obtained by Equation (1) and Equation (2), respectively. Where Kronecker product is denoted by '⊗', I denotes a matrix of subscripted shape and filled with 1, and O_{pool4} and O_{pool5_1} denote the output feature maps of layers 'pool4' and 'pool5_1', respectively.

$$F_2 = O_{pool4} \otimes I_{2\times2\times1} \tag{1}$$

$$F_3 = O_{pool5\_1} \otimes I_{4\times4\times1} \tag{2}$$

F_1, F_2, and F_3 are normalized into unit variance by Equation (3). Where $\sigma(\cdot)$ computes the standard deviation of elements in a matrix. The 'pool3', 'pool4', and 'pool5_1' descriptors of point *x* are denoted by $D_1(x)$, $D_2(x)$, and $D_3(x)$, respectively.

$$F_i \leftarrow \frac{F_i}{\sigma(F_i)} \tag{3}$$

Given two points *x* and *y*, feature distance *d*(*x*, *y*) is defined as

$$d(x,y) = \sqrt{2}d_1(x,y) + d_2(x,y) + d_3(x,y) \tag{4}$$

Where $d_i(x,y)$ denotes the Euclidean distance of $D_i(x)$ and $D_i(y)$. The weight $\sqrt{2}$ is applied to $d_1(x,y)$ since D_1 is 256-d, whereas D_2 and D_3 are 512-d.

For feature points *x* and *y*, if the following conditions are satisfied:

1) *d*(*x*, *y*) is the smallest of all *d*(·, *y*).
2) There does not exist a *d*(*z*, *y*) such that *d*(*z*, *y*) < $\theta \cdot d(x, y)$. θ is the matching threshold, greater than 1.

then *x* is matched to *y*. But this may lead to one-to-many mapping. This matching process is called prematching.

With a low threshold θ_0, a larger number of feature points are selected in the prematching stage. After that, a large starting threshold $\hat{\theta}$ is set to choose highly corresponding feature points only. An EM algorithm is used to get initial *Z* (the transformed locations of *Y*) based on these highly corresponding feature points. Then, threshold θ is subtracted by a step-length δ in every *k* iterations, allowing a few more feature points to affect the transformation and update *Z* iteratively. Such practice enables strongly matched feature points to determine the overall transformation while other feature points optimize registration accuracy.

In every iteration, *M* feature points from *IX* and *N* feature points from *IY* are chosen. An $M \times N$ probability matrix P_R can be built, which is then taken by a Gaussian mixture model (GMM)-based transformation solver. $P_R[m,n]$ denotes the putative probability of $x[n]$ and $y[m]$, in which $x[n]$ is corresponding to $y[m]$. The large probability would further lead to a conspicuous transformation over $y[m]$ by which the corresponding pair can be aligned.

C. Our Algorithmic Improvements

Although in [10], the original algorithm showed good performance on registering remote sensing images compared with SIFT-based methods, it does not work well on fluorescence-to-color image registration. We proposed two algorithmic improvements to address this challenge.

Our first improvement is filtering out subpar feature point matches based on the slope and length of the line connecting two corresponding feature points. Specifically, given two sets of feature points *X* of *IX* and *Y* of *IY*, we add a horizontal position shift on each feature point in *Y* to obtain shifted feature points Y'. The correspondences of feature points are transferred to *X* and Y'. Then, the length and slope of each line connecting a feature point in *X* and its corresponding feature point in Y' are calculated. Based on the assumption that slopes and lengths of the lines with high correspondence fall into a small range, we can filter out the feature point matches with low correspondence. More specifically, we calculate the histogram of these slopes and distances and set the range at their maximum values, as a filtering band to get good feature point matches. Fig. 3 shows the histogram and how to filter out subpar matches.

Our second improvement is using high-resolution multiscale feature maps. The multiscale feature maps are based on four output feature maps of four layers in the VGG-16 network. Observing the extracted feature point matches, we found that the locations of these corresponding feature points were not very accurate due to low resolution, impeding accurate image registration. Then, we used layers 'pool2', 'pool3', 'pool4' and 'pool5_1' to construct a high-resolution feature descriptor. F_0', O_{pool3}, O_{pool4}, and O_{pool5_1} denote output maps of layers 'pool2', 'pool3', 'pool4' and 'pool5_1',

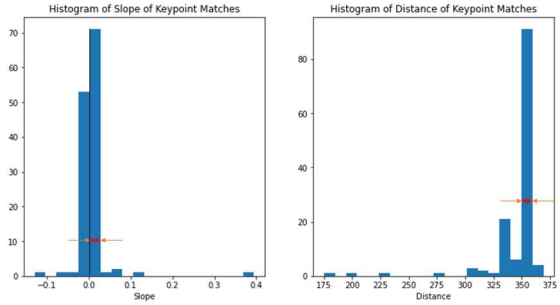

Fig. 3. The histogram of slopes and distances. Only feature point matches falling into the red double-headed arrow will be chosen for further registration.

respectively. F_1', F_2', and F_3' are defined by the following equations.

$$F_1' = O_{pool3} \otimes I_{2\times2\times1} \quad (5)$$

$$F_2' = O_{pool4} \otimes I_{4\times4\times1} \quad (6)$$

$$F_3' = O_{pool5\_1} \otimes I_{8\times8\times1} \quad (7)$$

Subsequently, the output maps are normalized, as depicted in Equation (3). The 'pool2', 'pool3', 'pool4', and 'pool5_1' descriptors of point x are denoted by $D_0'(x)$, $D_1'(x)$, $D_2'(x)$, and $D_3'(x)$, respectively. The definition of $d(x, y)$ in Equation (4) evolves into $d'(x, y)$ in Equation (8)

$$d'^{(x,y)} = 2d_0'(x,y) + \sqrt{2}d_1'(x,y) + d_2'(x,y) + d_3'(x,y) \quad (8)$$

We used a 56×56 grid to represent the input image. Compared with the 28×28 grid in the original algorithm, the resolution was increased.

D. Registration of Overlaid Image with Fluorescence

We use the sensor B of our fluorescence-color dual-modal imaging system, as illustrated in Fig. 1 to capture color reflectance images I_{color} and the sensor A to capture NIR reflectance images I_{nir} and fluorescence images I_{fl} with the NIR light source turned off and on, respectively. We set I_{color} as the reference image and I_{nir} as the target image. With the aforementioned method, we can obtain the transformation from I_{nir} to I_{color}. I_{nir} and I_{fl} are captured by the same sensor, thus the transformation can also be applied to align I_{fl} to I_{color}. In this way, we can achieve fluorescence-to-color image registration. Inspired by [13], instead of directly aligning I_{fl} to I_{color}, we first extract and combine denoised fluorescence signal with I_{nir} to obtain reflectance-fluorescence composite images I_{nir_fl}, then align I_{nir_fl} to I_{color} with the transformation.

Given I_{fl} and I_{nir}, we firstly subtract I_{nir} from I_{fl} to obtain the difference image I_{diff}. The pixels in I_{diff} with negative values are set as zero, which is based on the assumption that I_{fl} has higher luminance compared with I_{nir}. By calculating the histogram of I_{diff}, we can use a threshold to filter out noise signals in I_{diff} and get $I_{diff\_denoised}$. By overlaying $I_{diff\_denoised}$ and I_{nir}, we can get the composite image I_{nir_fl}.

III. Experiments

We conducted deep convolutional feature-based fluorescence-to-color image registration of a fluorescence tube containing an ICG solution and ICG-labeled biological tissues, such as chicken wing and porcine rib. We first manually segmented the targets in both reference images I_{re} and target images I_{tgt} and got two binary mask images I_{re_mask} and I_{tgt_mask}, respectively. After determining the transformation from I_{tgt} to I_{re} with the algorithm, we applied the transformation to I_{tgt_mask} to obtain $I_{tg\_mask\_reg}$. In addition, we calculated the intersection over union (IOU) between I_{re_mask} and $I_{tg\_mask\_reg}$ as a quantitative evaluation of the registration performance of the algorithm. Equation (9) shows how to calculate the IOU between I_{re_mask} and $I_{tg\_mask\_reg}$. We used the IOU to show the improvement of our implemented algorithm.

$$IOU = \frac{|I_{re\_mask} \cap I_{tg\_mask\_reg}|}{|I_{re\_mask} \cup I_{tg\_mask\_reg}|} \quad (9)$$

IV. Results and Discussions

The system characterization results of FOV and NIR excitation light distribution is shown in Fig. 4. As the working distance increases, the FOV increases linearly in both horizontal and vertical directions. NIR light power distributions at three working distances are close to gaussian distribution.

The registration results are illustrated in Fig. 5. Row 2 of Fig. 5 shows that filtering out incorrect feature point matches is necessary and helpful to increase the registration accuracy. Using feature descriptors with a higher resolution also improved the resolution of correspondences between the reference and target images and enhanced the registration results accordingly (Fig. 5: Row 3**)**. When our both improvements were combined, accurate registration was achieved (Fig. 5: Row 4**)**. The registration results of our implemented algorithm and

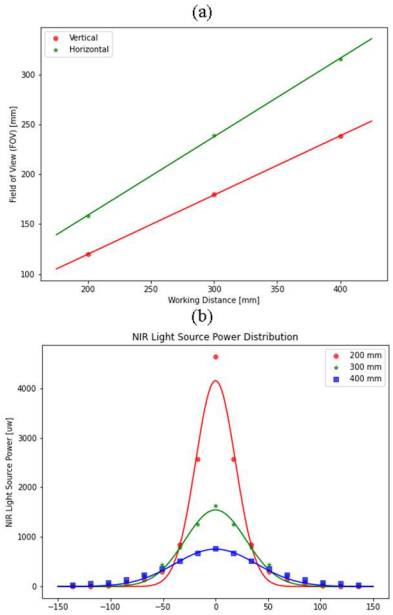

Fig. 4. System characterization of (a) the horizontal and vertical FOV and (b) the NIR light power distribution at different working distances.

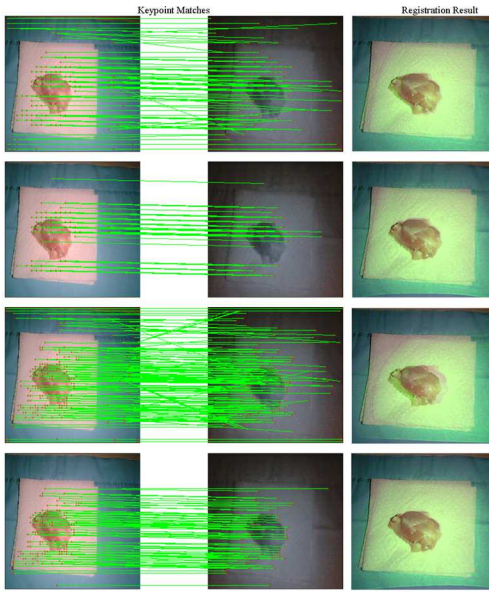

Fig. 5. Extracted feature point matches (column 1-2) & DL registration result (column 3). Row 1: original algorithm; row 2: improved algorithm with feature point filtering; row 3: improved algorithm with feature descriptor of higher resolution; row 4: algorithm with both improvements.

four traditional feature descriptor-based registration algorithms are shown in Fig. 6. The DL-56 algorithm outperformed SIFT and BRISK and achieved competitive registration performance compared to SURF and ORB. The increase in the resolution of feature descriptor helped boost the registration accuracy, especially for registering smaller objects with fine details, such as ICG tubes. These results have shown great potential to apply our deep-learning-based registration algorithm in preclinical and clinical settings.

We have computed the IOU between the object binary mask of the reference image and the transformed mask of its corresponding target image to evaluate the performance of our deep learning-based registration algorithms. The implementation of filtering has improved registration performance, as demonstrated in Table I. The IOU value of the improved deep learning-based algorithm was greater than that obtained by the original algorithm in [10]. DL-56 filter gave the best performance, especially in surgical relevant settings where biological tissues are present (chicken and porcine tissues). The results have shown great promise for intraoperative applications.

Limitation and Future Work: in this paper, we reported the algorithms and methods of a deep

Table I: Effects of Our Two Improvements on the Algorithm

	IOU		
Algorithm	A (Chicken Wing)	B (Tube)	C (Pork Rib)
DL-28_nofilter	0.852	0.032	0.769
DL-28_filter	0.848	0.129	0.816
DL-56_nofilter	0.629	0.124	0.786
DL-56_filter	**0.931**	**0.714**	**0.929**

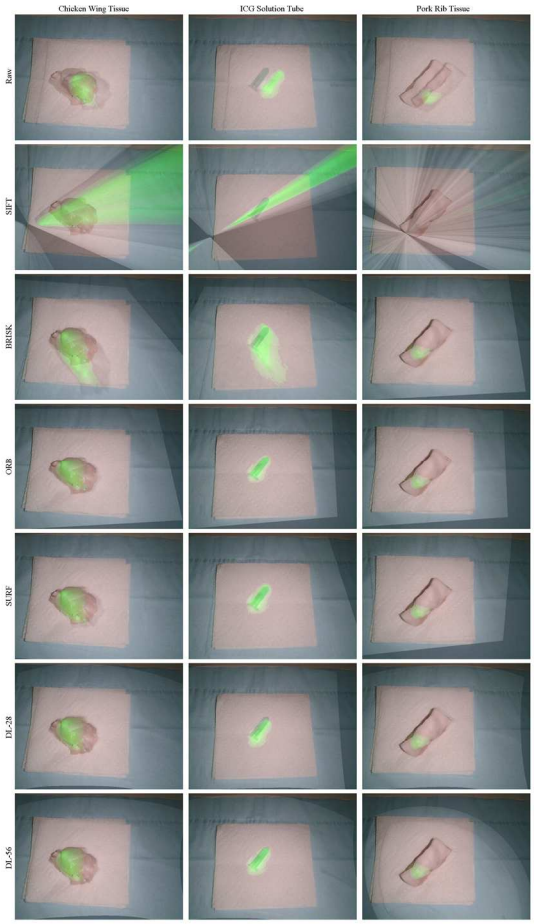

Fig. 6. Fluorescence-to-color image registration results. Fluorescence is pseudocolored in green. Column 1: chicken tissue with ICG fluorescence; column 2: centrifuge tube containing ICG solution; column 3: porcine rib containing ICG fluorescence. Row 1: direct overlay; row 2: SIFT-based method; row 3: BRISK-based method; row 4: ORB-based method; row 5: SURF-based method; row 6: deep convolutional feature-based method (28 × 28 feature descriptor) ; row 7: deep convolutional feature-based method (56×56 feature descriptor).

convolutional approach for fluorescence-to-color feature-based image registration. The algorithm has been tested in biological tissues and benchtop settings. In the future, we plan to conduct more comprehensive quantitative testing of registration accuracy of deep learning-based algorithms against other methods and apply the system and method to animal/human studies.

V. Conclusions

We have demonstrated the feasibility of deep convolutional feature-based image registration for fluorescence-to-color image registration tasks. Software-hardware codesign was conducted. To our best knowledge, this is the first demonstration of deep-learning-based image registration between fluorescence and color imageries.

REFERENCES

[1] T. Nagaya *et al.*, "Fluorescence-Guided Surgery," *Frontiers in Oncology,* vol. 7, DEC 22 2017, 2017.

[2] M. Olson, Q. Ly, and A. Mohs, "Fluorescence Guidance in Surgical Oncology: Challenges, Opportunities, and Translation," *Molecular Imaging and Biology,* vol. 21, no. 2, pp. 200-218, APR 2019, 2019.

[3] J. Watson *et al.*, "Augmented microscopy: real-time overlay of bright-field and near-infrared fluorescence images," *Journal of Biomedical Optics,* vol. 20, no. 10, OCT 2015, 2015.

[4] A. Goshtasby, "2-D and 3-D Image Registration for Medical, Remote Sensing, and Industrial Applications," *2-D and 3-D Image Registration For Medical, Remote Sensing, and Industrial Applications*, pp. 1-262, 2005, 2005.

[5] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision,* vol. 60, no. 2, pp. 91-110, NOV 2004, 2004.

[6] H. Bay *et al.*, "SURF: Speeded up robust features," *Computer Vision - Eccv 2006 , Pt 1, Proceedings,* vol. 3951, pp. 404-417, 2006, 2006.

[7] S. Leutenegger, M. Chli, and R. Siegwart, "BRISK: Binary Robust Invariant Scalable Keypoints," *2011 Ieee International Conference on Computer Vision (Iccv)*, pp. 2548-2555, 2011, 2011.

[8] E. Rublee *et al.*, "ORB: an efficient alternative to SIFT or SURF," *2011 Ieee International Conference on Computer Vision (Iccv)*, pp. 2564-2571, 2011, 2011.

[9] D. DeTone, T. Malisiewicz, and A. Rabinovich, "Deep Image Homography Estimation," arXiv:1606.03798, 2016.

[10] Z. Yang, T. Dan, and Y. Yang, "Multi-Temporal Remote Sensing Image Registration Using Deep Convolutional Features," *Ieee Access,* vol. 6, pp. 38544-38555, 2018, 2018.

[11] T. Quang *et al.*, "Fluorescence to Color Feature-Based Image Registration for Medical Augmented Reality," *2018 Ieee International Symposium on Signal Processing and Information Technology (Isspit)*, 2018, 2018.

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

[13] C. Mela, F. Papay, and Y. Liu, "Enhance Fluorescence Imaging and Remove Motion Artifacts by Combining Pixel Tracking, Interleaved Acquisition, and Temporal Gating," *IEEE Photonics Journal,* vol. 13, no. 1, pp. 1-13, 2021.