Study on Correlation Between Subjective and Objective Metrics for Multimodal Retinal Image Registration

YIQIAN WANG, JUNKANG ZHANG, (Graduate Student Member, IEEE), MELINA CAVICHINI, DIRK-UWE G. BARTSCH, WILLIAM R. FREEMAN, TRUONG Q. NGUYEN, (Fellow, IEEE), AND CHEOLHONG AN

1Electrical and Computer Engineering Department, University of California at San Diego, La Jolla, CA 92093, USA
2Jacobs Retina Center, Shiley Eye Institute, University of California at San Diego, La Jolla, CA 92093, USA
Corresponding author: Yiqian Wang (yiw018@ucsd.edu)
This work was supported in part by the U.S. National Science Foundation under Grant IIS-1522125.

ABSTRACT Retinal imaging is crucial in diagnosing and treating retinal diseases, and multimodal retinal image registration constitutes a major advance in understanding retinal diseases. Despite the fact that many methods have been proposed for the registration task, the evaluation metrics for successful registration have not been thoroughly studied. In this article, we present a comprehensive overview of the existing evaluation metrics for multimodal retinal image registration, and compare the similarity between the subjective grade of ophthalmologists and various objective metrics. The Pearson’s correlation coefficient and the corresponding confidence interval are used to evaluate metrics similarity. It is found that the binary and soft Dice coefficient on the segmented vessel can achieve the highest correlation with the subjective grades compared to other keypoint-supervised or unsupervised metrics. The paper established an objective metric that is highly correlated with the subjective evaluation of the ophthalmologists, which has never been studied before. The experimental results would build a connection between ophthalmology and image processing literature, and the findings may provide a good insight for researchers who investigate retinal image registration, retinal image segmentation and image domain transformation.

INDEX TERMS Multimodal retinal imaging, image registration, image similarity, Pearson correlation, subjective metric, objective metric, dice coefficient.

I. INTRODUCTION
Retinal diseases, including age-related macular degeneration, diabetes retinopathy, and vascular occlusion, are leading causes of multiple retina pathologies, and have systemic implications. The role of imaging in diagnosing and treating retina diseases is crucial [1], where the ophthalmologists have access to a large variety of retinal imaging devices, including color fundus (CF) imaging, scanning laser ophthalmoscopy (SLO), ultra-wide-field imaging, optical coherence tomography (OCT) angiography, and dye-based angiograms. Each device uses different wavelength, angiographic dyes, optical systems, and has different appearance and angles of view.

A major advance in understanding retinal disease would require the ability to register the multimodal images and integrate such functional and structural evaluations into one co-localizable database [2]. There is critical information available on different instruments, and the diagnostic benefit of accurately overlaying different imaging modalities has recently been demonstrated by different groups [3]–[5]. However, this is complicated by the fact that the fields of view, lens systems, light sources and manufactures of such devices are all different. Fortunately, the retinal vessels can be observed by all of these instruments in different ways, and are key features to align and overlay various diagnostic modalities.

Conventional approaches for multimodal retinal images registration can be roughly divided into three categories: area-based methods, feature-based methods, and learning-based methods. The area-based methods are designed to minimize the mutual information [6] or the entropy correlation coefficient [7] between the images to be aligned, but they are sensitive to texture variations across modalities and also require intensive computation.
The feature-based methods overcame this limitation by detecting sparse feature correspondences in both images to estimates the transformation. GDB-ICP [8] and ED-DB-ICP [9] were initially applied for multimodal retinal image registration, but they were sensitive to changes in scale. Keypoint detection and feature description was later improved by descriptors designed for retinal images [10]–[12] or domain-specific landmarks [13], combined with better vessel extraction algorithms [11], [13], [14]. The robustness of keypoint matching was also improved by graph-based matching [15], spherical model [16], robust point matching algorithm [17], asymmetric Gaussian mixture model [18], and so on. However, these conventional methods are still not robust enough for image pairs affected by disease or poor imaging quality. The recently proposed learning-based methods of the third category utilize the convolutional neural network (CNN) to achieve improved robustness and accuracy upon conventional methods [19]–[22].

However, the evaluation metrics for the multimodal retinal image registration have not been thoroughly studied and compared. The ophthalmologists use their own subjective grade to assess the accuracy of registration based on the aligned images, while most papers in image processing literature adopt several objective metrics for their fairness and simplicity. This divergence in evaluation metrics may cause performance variations from laboratory to clinical application, and poses potential barriers to improve current registration methods.

In this article, we propose a method to mathematically compare the similarity of the subjective grade and the commonly used objective evaluation metrics, and establish an objective evaluation metric that is most correlated with the subjective evaluation of the ophthalmologists. To the best of our knowledge, this work is the first extensive study on various evaluation metrics for multimodal retinal image registration. We will review the subjective and objective metrics in section II, and then introduce our method for comparing their similarity in section III. In section IV, we will compare these metrics, and provide extensive analysis. Finally, section V concludes the paper.

## II. RELATED WORKS

### A. SUBJECTIVE

Ophthalmologists adopt a subjective grading method, where the aligned multimodal images are analyzed in $5 \times 5$ blocks and overlaid in two forms of checkerboard, as shown in Fig. 1. The neighboring blocks in the checkerboard image show the registered images from two different modalities. If the images are perfectly aligned, the vessels are expected to be continuous on the edge across the neighboring blocks. In this subjective grading method, a score from 0 to 5 is assigned to each block in the checkerboard image, characterizing the ratio of overlap in vessels on the edge closest to the optic nerve as specified in Fig. 1. The grading criteria is shown in Table 1 and illustrated in Fig. 2, where grade 1 and 2 are considered as poor alignment, 3 as reasonable, 4 and 5 as good/excellent matches, and grade 0 is assigned to ungradable blocks due to absence of vessels, or no visible vessels to detect because of noise or out-of-focus. Note that in Fig. 1 and 2, the deformations on the edges of the CF image is caused by the deformable registration method applied to align the images, which will be explained in section IV.

### B. OBJECTIVE

Several objective metrics have been used in image processing literature, and they can be divided into two groups:
1) SUPERVISED

Supervised evaluation metrics require manually labeled keypoint correspondences between the multimodal images. Popular metrics includes maximum error (MAE) [10], [11], [17], [18], median error (MEE) [10], [11], [17], [18], root mean square error (RMSE) [11], [12], [17], [18], and percentage of correct keypoint pairs (PCK) [23].

Denote keypoint coordinates with \( p = (x, y)^T \) and the set of \( M \) pairs of manually labeled point correspondences with \( P = \{ (p_1, q_1), \ldots, (p_M, q_M) \} \). The MAE (maximum error, different from mean absolute error used in other fields) calculates the maximum L2 norm error on the selected point correspondences

\[
\text{MAE} = \max_{(p,q) \in P} ||F(p) - q||, \tag{1}
\]

where \( F(p) \) warps source keypoint \( p \) towards the target location using the transformation \( F(\cdot) \) estimated by the registration algorithm.

Using similar notation, MEE calculates the median L2 norm error on the selected points

\[
\text{MEE} = \text{median}_{(p,q) \in P} ||F(p) - q||, \tag{2}
\]

and RMSE calculates the root mean square error on the selected points

\[
\text{RMSE} = \sqrt{\text{mean}_{(p,q) \in P} ||F(p) - q||^2}. \tag{3}
\]

PCK sets a threshold \( T \) on the L2 norm to determine whether a pair of keypoints are correctly matched, and calculates the percentage of correct keypoints

\[
\text{PCK} = \frac{|\{(p, q) \mid ||F(p) - q|| < T, (p, q) \in P\}|}{|P|} \times 100%. \tag{4}
\]

The choice of threshold \( T \) is task dependent. For retinal image registration tasks, RMSE less 5 pixels is usually considered as success registration [11], [12], [17], [18], and the threshold \( T \) can be set to 5 pixel. To compute these metrics, we need to first manually label pairs of keypoint correspondences (generally 6 or more [10]–[12], [17], [18]) for all the multimodal images, where the keypoint locations should accurately lie on salient landmarks like vessel bifurcations, and uniformly distributed in the overlapping area. However, this is not an easy task. Even with the help of user friendly software (e.g. GUI), it is still difficult for human to select points with pixel-level accuracy and to make the points uniformly distributed in the overlapping area. Besides, labeling point correspondences by hand is very time-consuming for larger datasets with more than hundreds of image pairs.

2) UNSUPERVISED

Unsupervised evaluation metrics, on the contrary, do not require manually labeled keypoint correspondences, and only take the registered images as input.

Denote the aligned images with \( I_1 \) and \( I_2 \in \mathbb{R}^{H \times W} \), then the mean square error (MSE) is defined as

\[
\text{MSE} = \text{mean}(||I_1 - I_2||^2). \tag{5}
\]

The structural similarity index (SSIM) [24] is designed to improve MSE, and it is averaged on all the windowed patches \( W_1, W_2 \) from two images \( I_1, I_2 \)

\[
\text{SSIM}(I_1, I_2) = \frac{2\mu_1\mu_2 + c_1(2\sigma_{12} + c_2)}{\mu_1^2 + \mu_2^2 + c_1(\sigma_1^2 + \sigma_2^2 + c_2)}, \tag{6}
\]

where \( \mu_j \) is the mean value of window \( W_j \) (\( j = 1, 2 \)), and \( \sigma_j^2 \) and \( \sigma_{12}^2 \) are the variance of \( W_j \) and the covariance of \( W_1 \) and \( W_2 \) respectively. \( c_1 = (k_1L)^2 \), \( c_2 = (k_2L)^2 \) are two variables to stabilize the division with small denominator, \( L \) denoting the dynamic range of pixel intensity and \( k_1 = 0.01, k_2 = 0.03 \) by default. SSIM is a value between 0 and 1, and the higher SSIM indicates higher similarity between two images. Note that SSIM is only designed for single-channel images image, and RGB images need to be first converted to grayscale.

The Dice coefficient is frequently used for evaluating the overlapping region between two segmentation maps [14], [19], [20], [22]. With the aid of vessel segmentation algorithms, the Dice coefficient can also be used to evaluate the accuracy of registration. Let \( S_j \) denote the binary vessel segmentation of image \( I_j \) (\( j = 1, 2 \)), with 1 assigned to vessels and 0 assigned to background, the Dice coefficient for binary segmentation is defined as

\[
\text{Dice}(S_1, S_2) = \frac{2 \times \sum(S_1 \odot S_2)}{\sum S_1 + \sum S_2}, \tag{7}
\]

where \( \odot \) denotes element-wise product. The Dice coefficient ranges between 0 and 1, and higher number indicates larger overlap.

The soft Dice coefficient [20] introduces a differentiable counterpart of the binary Dice coefficient for vessel probability maps, and it is defined as

\[
\text{Dice}_s(P_1, P_2) = \frac{2 \times \text{ele_min}(P_1, P_2)}{\sum P_1 + \sum P_2}, \tag{8}
\]

with \( P_j \) denoting the vessel probability map of \( I_j \) (\( j = 1, 2 \)). The python code for binary and soft Dice coefficient can be found in Table 2.

III. METHOD

In order to build a connection between ophthalmology and image processing literature, in this article, we compute the Pearson correlation coefficient [25] between the subjective grade of ophthalmologists and various objective evaluation methods to compare their degree of similarity.

A. PEARSON CORRELATION COEFFICIENT

The Pearson correlation coefficient (PCC) for random variable \( X \) and \( Y \) is defined as

\[
\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}, \tag{9}
\]
where the covariance of $X$ and $Y$ is
\[
\text{cov}(X, Y) = E[(X - E[X])(Y - E[Y])], \tag{10}
\]
and $\sigma_X$ and $\sigma_Y$ are the standard deviation of $X$ and $Y$, respectively. For finite $N$ samples $\{(x_1, y_1) \cdots (x_N, y_N)\}$, PCC can be estimated by
\[
x_{x,y} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}}, \tag{11}
\]
where $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ denotes the sample mean, and similarly for $\bar{y}$.

The PCC value $\rho$ ranges between $[-1, 1]$, where higher absolute PCC value $|\rho|$ implies higher linear dependency. Positive $\rho$ value implies that $Y$ increases as $X$ increases, while negative value implies that $Y$ decreases as $X$ increases. For example, the correlation between MAE and the subjective grade should be negative, while the correlation between PCK and the subjective grade should be positive. Based on Cohen’s interpretation [26], an absolute PCC value of 0.1 is small, 0.3 is medium, and 0.5 is large. An important property of the PCC value is that it is invariant to separate linear transforms in $X$ and $Y$
\[
\rho_{x,y} = \rho(aX+b,cY+d), \text{ where } ac > 0, a, b, c, d \in \mathbb{R}, \tag{12}
\]
which means that PCC can still be applied even when the scale and range are different for different criteria.

**B. CONFIDENCE INTERVAL**

In order to test the certainty in the estimated PCC $r$ value in eq (11), we further calculate its confidence interval to indicate the possible range of true PCC $\rho$ values in eq (9). The confidence interval for the population’s Pearson’s $\rho$ is can be computed with three steps [27].

Firstly, a $z$-score is computed applying the Fisher transformation $F(\cdot)$
\[
z = F(r) = \frac{1}{2} \ln \frac{1+r}{1-r} = \arctanh(r). \tag{13}
\]

Secondly, given a significant level $\alpha$ which is typically set to 95%, the critical $z$-score $z_{\alpha/2}$ for a two-tail test can be obtained from a look-up table. Then the confidence interval for $F(\rho)$ would be
\[
F(\rho) \in \left[ F(r) - z_{\alpha/2} \text{SE}, F(r) + z_{\alpha/2} \text{SE} \right], \tag{14}
\]
where SE is the standard error
\[
\text{SE} = \frac{1}{\sqrt{n-3}}. \tag{15}
\]

Finally, the confidence interval for $r$ is converted back using the inverse Fisher transformation
\[
F(\rho) \in \left[ F^{-1}\left( F(r) - z_{\alpha/2} \text{SE} \right), F^{-1}\left( F(r) + z_{\alpha/2} \text{SE} \right) \right], \tag{16}
\]
where
\[
F^{-1}(z) = \frac{e^{2z} - 1}{e^{2z} + 1} = \tanh(z). \tag{17}
\]

**IV. EXPERIMENT**

In this experiment, we test the Pearson correlation between various objective metrics and the subjective grade on the multimodal retinal image registration result. We select two automatic deformable registration methods to obtain aligned images for evaluation. The first one is a conventional method, MIND [28], and the second is a learning-based method [20]. As the deformable registration methods are designed to local misalignment, we generate the input image pairs by random deformations based on ground truth alignment [19], [20], as shown in Fig. 3. For the aligned images using two methods, ophthalmologists grade them for the subjective metric, and we derive the objective metric with MAE, MEE, RMSE, PCK, SSIM and Dice coefficients. Note that here, we use two automatic deformable registration methods [20] and [28] not because of comparing performance between two methods but because of providing consistency of correlation between the subjective and objective metrics for both the conventional and learning methods.

**A. DATASET AND GROUND TRUTH**

The dataset consists of 109 pairs of Color Fundus images (TRC-50DX color fundus images, Topcon) and infrared Scanning Laser Ophthalmoscope images (Heidelberg Engineering Spectralis SLO). The dataset includes a variety of pathologies including hemorrhages, diabetes, and macular degeneration.

The multimodal images used in the experiment include 3 different types of content change. Firstly, the pathology has different appearance in the two imaging modalities, as illustrated in the example in Fig. 4. Secondly, the images are not necessarily taken at the same time, which also makes a difference as the disease progresses. Thirdly, there are different artifacts like out-of-focus, reflection, and over or under exposed regions in the two modalities.

The source CF images are 24-bit RGB and their resolution is 3000 × 2672, and the target SLO images are 8-bit grayscale and resolution of the images is 768 × 768 or 1536 × 1536.

---

**TABLE 2. Python3 code for binary and soft Dice coefficient.**

```python
import numpy as np
def dice_bin(src, tgt):
    
    Binary dice coefficient
    src, tgt: (H,W) binary numpy array
    
    denom = src.sum() + tgt.sum()
    return 2*(src*tgt).sum() / denom

def dice_soft(src, tgt):
    
    Soft Dice coefficient
    src, tgt: (H,W) float numpy array in [0,1]
    
    denom = src.sum() + tgt.sum()
    return 2*np.minimum(src,tgt).sum() / denom
```
Y. Wang et al.: Study on Correlation Between Subjective and Objective Metrics for Multimodal Retinal Image Registration

FIGURE 3. Input generation and deformable registration pipeline. Registration methods [20] and [28] are used.

FIGURE 4. Example image pair with disease (aligned by [20]).

They are both padded to square shape and resized to 768 × 768 before registration, and the ground truth transformation matrices are derived by manually selected correspondences from the source to target images. We use the SuperPoint network [29] to detect keypoints on the segmentation result of [20] and select all possible pairs of point correspondences.

B. EXPERIMENT SETTING

As shown in Fig. 3, the input is generated by randomly deforming the image pair aligned by the ground truth transformation matrix, where the random deformation field is upsampled from a 4 × 4 × 2 matrix that follows normal distribution with 0 mean and standard deviation of 5 pixels. The input pair is then registered using each of the two registration methods to generate the output pair, where MIND [28] is implemented in MATLAB, and [20] is implemented in PyTorch.

The output of the registration methods are graded independently by 2 retina specialists and 3 medical students. Each pair of aligned images are graded two times based on two forms of checkerboard overlay shown in Fig. 1. We take the average of the two grades assigned to the same block if both of them are non-zero, take the non-zero grade if only one grade is zero, and keep the grade zero if both grades are zero. To verify the reliability of the grades, each of the 5 graders first grade the same subset of 10 images, and the average intra-class correlation coefficient (ICC) on the subset calculated by SPSS [30] is 0.903, which indicates excellent similarity among graders.

In the comparison, each block is treated as an individual sample, as illustrated in Fig. 5. MSE, SSIM, and Dice coefficients are computed on the image patch within each block, where the black area in CF image is excluded. MAE, MEE, RMSE, and PCK are calculated on all correspondences within the block, as shown in Fig. 6. Especially, blocks with no valid correspondences or marked as ungradable (grade 0) are excluded when calculating the correlation coefficient, as they may affect majorly the correlation coefficients otherwise.

We set the threshold T in eq. (4) at 5 pixels for PCK, which yields the maximum correlation with grade at 0.1539, as shown in Fig. 7. MSE and SSIM are computed on the aligned images after converting to grayscale. Two segmentation methods [31] and [20] are tested for the binary and soft Dice coefficients. We use the optimal threshold at 0 and 0.5 for [31] and [20] respectively when calculating the binary Dice coefficient, which yields largest correlation with the grade.

C. RESULTS

Fig. 8 shows the correlations between the subjective grade and various objective metrics and their 95% confidence intervals, which are evaluated on each block or image pair.
The soft Dice coefficient using segmentation [20] achieves the highest correlation with the subjective grade at 0.35, which can be considered moderate based on Cohen’s interpretation [26], and the binary Dice coefficient using [20] has slightly lower correlation. The binary and soft Dice coefficient in [31] also demonstrate small correlation with the subjective grade. This coincide with the fact that [20] can obtain more accurate of vessel segmentation map compared to [31]. PCK and SSIM show small positive correlation, while the error terms including MAE, MEE, RMSE, and MSE show smaller negative correlation. The main reason that the binary and soft Dice coefficient show the highest correlation with subjective grade may lie in their common emphasis on retinal vessels. The Dice coefficient calculates the overlapping of vessel segmentation, and the subjective grade also depends on the continuity of vessels. Meanwhile, the other metrics do not explicitly depend on vessels, which leads to lower correlation with the subjective grade.

In order to evaluate the intrinsic correlation between different metrics, we calculate the absolute PCC value between each pair of metrics shown in Fig. 9. We observe that the correlation between the supervised metrics are very high, because they all operate on the selected keypoint correspondences. MSE and SSIM also have very large correlation,
as they both take the grayscale images as input. The binary and soft Dice coefficient for the segmentation [20] are nearly perfectly correlated, as the segmentation map obtained by [20] is already close to binary. Meanwhile, the binary and soft Dice coefficient make a difference for the segmentation [31], where the vessel segmentation is closer to the probability map.

We further investigate the influence of choosing different registration method. In the following experiment, correlation is calculated separately using the two registration methods [20], [28]. As shown in Fig. 10, the soft Dice coefficient under segmentation [20] still has the highest absolute correlation in both methods. Specifically, the soft Dice coefficient under segmentation [20] has higher correlation using only registration method [28], where the grade distribution, shown in Fig. 11, better covers both higher and lower grades. This experiment demonstrates that adopting different registration methods will lead to similar conclusion in this article.

V. CONCLUSION

In this article, we presented a comprehensive overview of the existing evaluation metrics for multimodal retinal image registration, and compared the Pearson correlation coefficient between the ophthalmologists’ subjective grade and several commonly used objective evaluation metrics. We found that the soft Dice coefficient using the segmentation method in [20] achieved higher correlation with the subjective grades compared to many commonly used keypoint-supervised metrics. The paper established an objective metric that is highly correlated with the subjective evaluation of the ophthalmologists. The experimental results would build a connection between ophthalmology and image processing literature, and the findings may provide a good insight for researchers who investigate retinal image registration, retinal image segmentation and image domain transformation.

There are some limitations in the experiment that can be improved by future work. For example, the dataset consists of 108 image pairs with two imaging modalities (Color Fundus and infrared SLO), and we evaluated two deformable registration methods [20], [28]. Although our dataset is one of the largest multimodal datasets in the literature, in the future work, the experiment can be improved by including larger dataset, more imaging modalities, and more registration methods.

ACKNOWLEDGMENT

The authors would like to thank Mahima Jhingan, Manuel J. Amador-Patarroyo from Jacobs Retina Center, Shirley Eye Institute, University of California San Diego, and Christopher P. Long, Alison X. Chan, Samantha Madala from School of Medicine, University of California San Diego for their help in data collection.

REFERENCES

[1] T. J. MacGillivray, E. Trucco, J. R. Cameron, B. Dhillion, J. G. Houston, and E. J. R. van Beek, “Retinal imaging as a source of biomarkers for diagnosis, characterization and prognosis of chronic illness or long-term conditions,” Brit. J. Radiol., vol. 87, no. 1040, Aug. 2014, Art. no. 20130832.
[2] T. K. Yoo, J. Y. Choi, J. G. Seo, B. Ramasubramanian, S. Selvaperumal, and D. W. Kim, “The possibility of the combination of OCT and fundus images for improving the diagnostic accuracy of deep learning for age-related macular degeneration: A preliminary experiment,” Med. Biol. Eng. Comput., vol. 57, no. 3, pp. 677–687, Mar. 2019.
[3] C. A. Toth, V. Tai, M. Pistilli, S. J. Chiu, K. P. Winter, E. Daniel, J. E. Grunwald, G. J. Jaffe, D. F. Martin, G.-S. Ying, S. Farsiu, and M. G. Maguire, “Distribution of OCT features within areas of macular atrophy or scar after 2 years of anti-VEGF treatment for neovascular AMD in CATT,” Ophthalmology Retina, vol. 3, no. 4, pp. 316–325, Apr. 2019.
[4] M. Nagpal, J. Khandelwal, R. Juneja, and N. Mehrouta, “Correlation of optical coherence tomography angiography and microperimetry (MP3) features in wet age-related macular degeneration,” Indian J. Ophthalmol., vol. 66, no. 12, p. 1790, 2018.
[5] D. Mukherjee, E. M. Lad, R. R. Vann, S. J. Jaffe, T. E. Clemens, M. Friedlander, E. Y. Chew, G. J. Jaffe, S. Farsiu, and f. the MacTel Study Group, “Correlation between macular integrity assessment and optical coherence tomography imaging of ellipsoid zone in macular telangiectasia type 2,” Investigative Ophthalmology Vis. Sci., vol. 58, no. 6, Sep. 2017, Art. no. BIO291.
[6] N. Ritter, R. Owens, J. Cooper, R. H. Eikelboom, and P. P. Van Saarloos, “Registration of stereo and temporal images of the retina,” IEEE Trans. Med. Imag., vol. 18, no. 5, pp. 404–418, May 1999.
[7] T. Chanwimaluang, G. Fan, and S. R. Fransen, “Hybrid retinal image registration,” IEEE Trans. Inf. Technol. Biomed., vol. 10, no. 1, pp. 129–142, Jan. 2006.
[8] G. Yang, C. V. Stewart, M. Sofka, and C.-L. Tsai, “Registration of challenging image pairs: Initialization, estimation, and decision,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 11, pp. 1973–1989, Nov. 2007.
[9] C.-L. Tsai, C.-Y. Li, G. Yang, and K.-S. Lin, “The edge-driven dual-bootstrap iterative closest point algorithm for registration of multimodal fluorescence angiogram sequence,” IEEE Trans. Med. Imag., vol. 29, no. 3, pp. 636–649, Mar. 2010.
[10] 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. Med. Imag., vol. 57, no. 7, pp. 1707–1718, Jul. 2010.
[11] J. A. Lee, J. Cheng, B. H. Lee, E. P. Ong, G. Xu, D. W. K. Wong, J. Liu, A. Laude, and T. H. Lim, “A low-dimensional step pattern analysis algorithm with application to multimodal retinal image registration,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 1046–1053.
[12] Z. Ghassab, J. Shanbehzadeh, A. Sedaghat, and E. Fatemizadeh, “An efficient approach for robust multimodal retinal image registration based on UR-SIFT features and PIIFD descriptors,” EURASIP J. Image Video Process., vol. 2013, no. 1, p. 25, Dec. 2013.
[13] Á. S. Hervella, J. Rouco, J. Novo, and M. Ortega, “Multimodal registration of retinal images using domain-specific landmarks and vessel enhancement,” Procedia Comput. Sci., vol. 126, pp. 97–104, Jan. 2018.
Y. Wang et al.: Study on Correlation Between Subjective and Objective Metrics for Multimodal Retinal Image Registration

YIQIANG WANG: Study on Correlation Between Subjective and Objective Metrics for Multimodal Retinal Image Registration

JUNKANG ZHANG (Graduate Student Member, IEEE) received the B.E. degree in automation from Hohai University, Changzhou, China, in 2014, and the M.E. degree in pattern recognition and intelligent system from Southeast University, Nanjing, China, in 2017. He is currently pursuing the Ph.D. degree in electrical and computer engineering with University of California at San Diego, CA, USA. His research interests include image processing and computer vision.

MELINA CAVICCHINI received M.D. degree from the University Gama Filho, Rio de Janeiro, Brazil, in 2008. She did her Vitreo Retinal Surgery Fellowship with the Suel Abujamra Institute, São Paulo, Brazil, in 2014, and the Ophthalmology Residency, School of Medicine from ABC region, São Paulo, in 2011. She is currently a Retina Fellow Research with the Jacobs Retina Center, Shirley Eye Institute, University of California, San Diego. Her research interests include retinal diseases, the use of artificial intelligence in retina, and new technologies for retina.

DIRK-UWE G. BARTSCH received the bachelor’s degree from Technische Universitaet Darmstadt and the Ph.D. degree in bioengineering from the University of California at San Diego.

He received the Postdoctoral fellowship with the University of California at San Diego. He is currently an Associate Adjunct Professor and the Co-Director of the Jacobs Retina Center. His research interests include retinal imaging, scanning laser imaging-confocal/non-confocal, optical coherence tomography (OCT), indocyanine green and fluorescein angiography, and tomographic reconstruction of the posterior pole in patients with various retina diseases such as age-related macular degeneration, diabetes, and HIV-related complications.

WILLIAM R. FREEMAN is currently an Distinguished Professor of ophthalmology, the Director with the UCSD Jacobs Retina Center and the Vice Chair of the UCSD Department of Ophthalmology. He is a full time Retina Surgeon and also a Researcher who has held NIH grants for nearly 30 years. He works closely with imaging groups with the Department of Ophthalmology, UCSD School of Engineering. He has over 600 peer-reviewed publications.
TRUONG Q. NGUYEN (Fellow, IEEE) is currently a Professor with the ECE Department, UC San Diego. His current research interests include 3D video processing and communications and their efficient implementation. He is the coauthor (with Prof. Gilbert Strang) of a popular textbook, *Wavelets and Filter Banks* (Wellesley-Cambridge Press, 1997), and the author of several matlab-based toolboxes on image compression, electrocardiogram compression, and filter bank design. He has over 400 publications.

He received the IEEE *Transactions on Signal Processing* Paper Award (Image and Multidimensional Processing area) for the article he co-wrote with Prof. P. P. Vaidyanathan on linear-phase perfect-reconstruction filter banks (1992). He received the NSF Career Award in 1995 and is currently the Series Editor (Digital Signal Processing) for Academic Press. He served as the Associate Editor for the IEEE *Transactions on Signal Processing*, from 1994 to 1996, the *Signal Processing Letters*, from 2001 to 2003, the IEEE *Transactions on Circuits and Systems*, from 1996 to 1997, 2001 to 2004, and the IEEE *Transactions on Image Processing*, from 2004 to 2005.

CHEOLHONG AN received the B.S. and M.S. degrees in electrical engineering from Pusan National University, Busan, South Korea, in 1996 and 1998, respectively, and Ph.D. degree in electrical and computer engineering, in 2008. He is currently an Assistant Adjunct Professor of electrical and computer engineering from the University of California at San Diego. Earlier, he worked at Samsung Electronics, South Korea, and Qualcomm, USA. His current research interests include the medical image processing and the real-time bio image processing. His research interests include 2D and 3D image processing with machine learning and sensor technology.

* * *