2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Perspective Distortion Correction for Multi-Modal Registration between Ultra-Widefield and Narrow-Angle Retinal Images

Junkang Zhang[1], Yiqian Wang[1], Dirk-Uwe G. Bartsch[2],
William R. Freeman[2], Truong Q. Nguyen[1], and Cheolhong An[1]

***Abstract*— Multi-modal retinal image registration between 2D Ultra-Widefield (UWF) and narrow-angle (NA) images has not been well-studied, since most existing methods mainly focus on NA image alignment. The stereographic projection model used in UWF imaging causes strong distortions in peripheral areas, which leads to inferior alignment quality. We propose a distortion correction method that remaps the UWF images based on estimated camera view points of NA images. In addition, we set up a CNN-based registration pipeline for UWF and NA images, which consists of the distortion correction method and three networks for vessel segmentation, feature detection and matching, and outlier rejection. Experimental results on our collected dataset shows the effectiveness of the proposed pipeline and the distortion correction method.**

I. Introduction

Multi-modal retinal registration aligns fundus images of a same eye which are captured by different instruments, in order to provide a complete view of pathologies for ophthalmological diagnosis. In recent years, Ultra-Widefield (UWF) imaging becomes a popular option for retinal imaging, due to its larger field of view than conventional narrow-angle (NA) images. Therefore, UWF images can provide a more comprehensive view of retina and help with diagnosis and early screening of a variety of diseases [1]. In this paper, we investigate the multi-modal registration task between 2D UWF and NA images.

There have been extensive works on multi-modal retinal registration. Some approaches proposed complete registration pipelines [2], [3] consisting of steps for feature detection, description, and outlier rejection, while others improved only the feature detectors [4], descriptors [5], [6] or outlier rejection modules [7] in existing pipelines. Recently, multiple Convolutional Neural Networks (CNN) approaches have been proposed. Some works [8], [9] replaced certain modules in conventional pipelines with CNN, and other approaches [10]–[12] set up fully-CNN-based pipelines for this task. Nevertheless, none of the existing works have proposed to align retinal images with large differences in view angles (*e.g.*, 200° Optos UWF Colormaps and 55° MultiColor images).

A major challenge in aligning UWF and NA images comes from the perspective distortions in the UWF modality. Since retina is on the surface of a sphere (eyeball), it involves a projection process to capture and visualize the 3D retina on a 2D flat array, which introduces distortions. Although retina images always suffer from distortions from the projection, the distortions of UWF images are significantly more visible than those of NA images. For example, the stereographic projection in Optos's UWF system sets the cornea as its camera view point, which leads to significant distortions in peripheral retina area, such that peripheral patterns appear larger than their actual sizes. On the other hand, NA images are projected based on more distant view points, and thus bear less perspective distortions. Therefore, different projection methods between UWF and NA make registration more difficult, since the peripheral distortions cannot be corrected by a conventional 2D-to-2D planar transformation model (e.g. perspective transformation).

The distortion correction for UWF images is also related to retinal curvature estimation and 3D reconstruction. For example, Chanwimaluang *et al.* [13] proposed to estimate retinal curvature from NA image sequences through Structure From Motion which incorporates constraints on ellipsoid surface and lens distortions. Ataer-Cansizoglu *et al.* [14] set up a 3D fundus reconstruction method from image sequences, where the 3D surface parameters and camera poses are estimated by minimizing re-projection errors. They also explored different 3D models for reconstruction. Probst *et al.* [15] applied a 3D reconstruction model to stereo microscope for retinal microsurgery. Dan *et al.* [16] set up a 2D registration and 3D reconstruction pipeline, where their reconstruction model was merely based on existing camera parameters and was apart from their registration pipeline. However, these methods mainly focus on building 3D models from single-modal NA images, and the multi-modal registration between 2D UWF and NA images has not been investigated.

In order to reduce the projection distortion of UWF images in multi-modal registration, we propose a distortion correction method on UWF images. Specifically, the 2D UWF pixels are first projected back to their 3D positions on the eyeball through the inverse stereographic projection, and then remapped to a 2D plane based on the average projection parameters for NA images. Since most of the UWF and NA images are centered on the fovea, we assume that they share a same optical axis in projection. The assumption simplifies the 3D-to-2D remapping process, so that we only need to estimate the view point of the NA images, as shown in Fig. 3.

[1]Junkang Zhang, Yiqian Wang, Truong Q. Nguyen, and Cheolhong An are with the Electrical and Computer Engineering Department, University of California San Diego, La Jolla, CA 92093 USA. juz007@eng.ucsd.edu

[2]Dirk-Uwe G. Bartsch, and William R. Freeman are with the Department of Ophthalmology, Jacobs Retina Center at Shiley Eye Institute, University of California San Diego, La Jolla, CA 92093 USA.

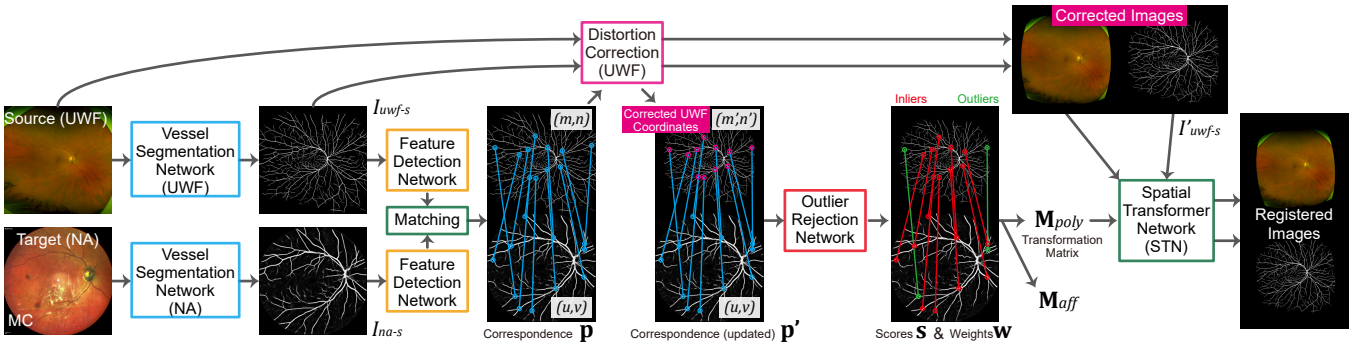


Fig. 1. Proposed registration pipeline for UWF and NA retinal images.

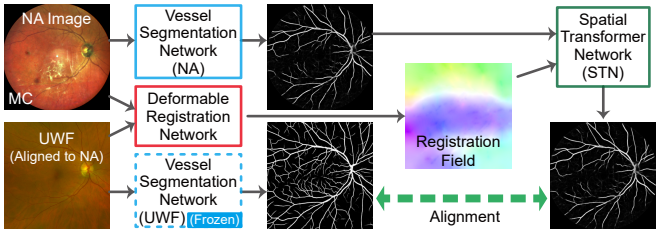


Fig. 2. Training scheme of the vessel segmentation network for NA images.

Besides, we propose a complete CNN-based registration pipeline for UWF and NA images, which consists of networks for vessel segmentation, feature detection, and outlier rejection in addition to the distortion correction. We also set up a new training scheme for the vessel segmentation network on NA images, and adopt the polynomial transformation model to improve registration performance.

II. Proposed Registration Pipeline

Fig. 1 shows the proposed registration pipeline where the vessel structures from both input images are first extracted using two independent segmentation networks. Next, keypoints from both images are detected by a feature detection network, and then paired through a matching process. Furthermore, the UWF image and its coordinates in the matched keypoints are corrected through our proposed distortion correction module, which is detailed in Section III. Finally, a transformation matrix is estimated by an outlier rejection network from the corrected matched keypoints, and the UWF image can be aligned with the NA image.

A. Vessel Segmentation Network

We adopt the pre-trained vessel segmentation network [17] for the UWF modality. In order to train the other segmentation network for NA images, we build a modified learning framework based on [18] as shown in Fig. 2, where the weights of the segmentation network for UWF are frozen during the training process. The input UWF images are coarsely aligned to NA images based on manual labels, and then cropped to the overlapped area as in the NA images. We train the vessel segmentation network for NA image with a combination of two loss terms as

$$\mathcal{L}_{seg} = \lambda_{pc}\mathcal{L}_{pc} + \lambda_{sm}\mathcal{L}_{sm}. \quad (1)$$

where $\mathcal{L}_{pc}$ and $\mathcal{L}_{sm}$ are photometric consistency loss and smoothness loss in the optical flow networks [19], respectively. The loss function enables the segmentation network to predict NA's vessel maps that can be aligned with the UWF vessels. As shown in Fig. 1, the vessel segmentation results are denoted as $I_{uwf\text{-}s}$ and $I_{na\text{-}s}$ for UWF and NA images, respectively. The UWF vessel map $I_{uwf\text{-}s}$ will be corrected to remove distortion in the following distortion correction module, which is denoted as $I'_{uwf\text{-}s}$.

B. Feature Detection and Matching Network

In this paper, we adopt the SuperPoint [20] network as the feature detection network. The network is pre-trained on a multi-modal retinal image dataset with color fundus and infrared reflectance images [12]. It takes an uncorrected vessel map ($I_{uwf\text{-}s}$ or $I_{na\text{-}s}$) as an input, and outputs a keypoint heatmap and its corresponding descriptor tensor. Then, the keypoints are obtained through Non-Maximum-Suppression on the heatmap, and their coordinates are denoted as (m, n) and (u, v) in the UWF and NA images, respectively. The correspondence $\mathbf{p} = [(m, n, u, v), ...] \in \mathbb{R}^{N\times 4}$ between the two images is established by a bi-directional matching algorithm for the keypoints, where N is the number of matched keypoint pairs. In a matched pair, the UWF's feature should be a best match for the NA's feature, and vice versa. Readers can refer to [12], [20] for more details.

After keypoint matching and before the outlier rejection network, the UWF keypoints' coordinates (m, n) are corrected by the distortion correction method as (m', n'). Therefore, the correspondence is updated as $\mathbf{p}' = [(m', n', u, v), ...]$.

C. Outlier Rejection Network

We use the outlier rejection network structure [21], which was pre-trained on the other retinal dataset [12] and then fine-tuned on our dataset. The network takes the correspondence $\mathbf{p}'$, and outputs scores $\mathbf{s} \in \mathbb{R}^{N\times 1}$ for all correspondences, which is similar to RANSAC [22]. Then, the scores are translated into weights as $\mathbf{w} = tanh(ReLU(\mathbf{s}))$, where $w_i \in [0, 1)$. Finally, transformation matrices are estimated based on $\mathbf{p}'$ and $\mathbf{w}$ using weighted least square methods, where both affine transformation matrix $\mathbf{M}_{aff} \in \mathbb{R}^{2\times 3}$ and 2nd-order polynomial transformation matrix $\mathbf{M}_{poly} \in \mathbb{R}^{2\times 6}$

are obtained. The affine matrix is only used in $\mathcal{L}_r$ of Eq. (5) during training, because we only have ground-truth for it. Meanwhile, the polynomial matrix is used for image warping, since it has shown advantages in retinal image registration [23].

During network training, we define the loss function as

$$\mathcal{L}_{outlier} = \lambda_c \mathcal{L}_c(\mathbf{p}', \mathbf{s}, \mathbf{M}_{gt}) + \lambda_r \mathcal{L}_r(\mathbf{M}_{gt}, \mathbf{M}_{aff}) + \lambda_d \left(1 - Dice\left(\mathrm{STN}(\mathbf{M}_{poly}, I'_{uwf\text{-}s}), I_{na\text{-}s}\right)\right) \quad (2)$$

where $\mathbf{M}_{gt} \in \mathbb{R}^{2\times 3}$ is a ground-truth affine matrix, and $\mathrm{STN}(\cdot)$ is an image warper [24]. $\mathcal{L}_c$ is a classification loss (Binary Cross Entropy Function) between the estimated scores $\mathbf{s}$ and ground-truth inliers

$$\mathcal{L}_c(\mathbf{p}', \mathbf{s}, \mathbf{M}_{gt}) = \frac{1}{N}\sum_{i=1}^{N} \gamma_i \mathrm{BCE}\big(y_i, \sigma(s_i)\big) \quad (3)$$

where $\sigma(\cdot)$ is a sigmoid function, γ_i is a balancing factor for positive and negative samples, and $y_i \in \{0, 1\}$ is the inlier ground-truth. The inliers are computed from $\mathbf{M}_{gt}$ as

$$y_i = \begin{cases} 1, & \big|\big|T\big((m', n'), \mathbf{M}_{gt}\big) - (u, v)\big|\big| \leq 5 \quad \text{pixels} \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where $T(\cdot)$ translates a UWF keypoint (m', n') into NA's coordinate based on $\mathbf{M}_{gt}$. A keypoint pair with distance less than 5 pixels in NA's coordinate is considered as an inlier. Next, $\mathcal{L}_r$ is a regression loss between $\mathbf{M}_{aff}$ and $\mathbf{M}_{gt}$, which is defined as

$$\mathcal{L}_r = \mathrm{MSE}(\mathbf{M}_{gt} - \mathbf{M}_{aff}). \quad (5)$$

Finally, $Dice$ is derived to measure the overlapping degree between two vessel maps, which is written as

$$Dice(I_1, I_2) = \frac{2 \cdot \sum \big(\mathrm{ele\_min}(I_1, I_2)\big)}{\sum I_1 + \sum I_2}. \quad (6)$$

Readers can also refer to [12] for more details.

D. Learning Process

First, we train the segmentation network for NA images using manually aligned image pairs as described in section II-A. Then, the outlier rejection network in Section II-C is trained without the distortion correction module, *i.e.*, using the uncorrected UWF vessel maps $I'_{uwf\text{-}s} \leftarrow I_{uwf\text{-}s}$ and keypoints $\mathbf{p}' \leftarrow \mathbf{p}$. Finally, with all networks' weights frozen, the view point of the NA images is estimated in the distortion correction module on the training dataset, which will be detailed in the next section.

III. Distortion Correction

The distortion correction process can be separated into two steps, *i.e.*, a 2D-to-3D projection which maps the original UWF image back onto the eyeball, and a 3D-to-2D projection that casts a 3D point on the eyeball to a 2D image plane based on a new view point.

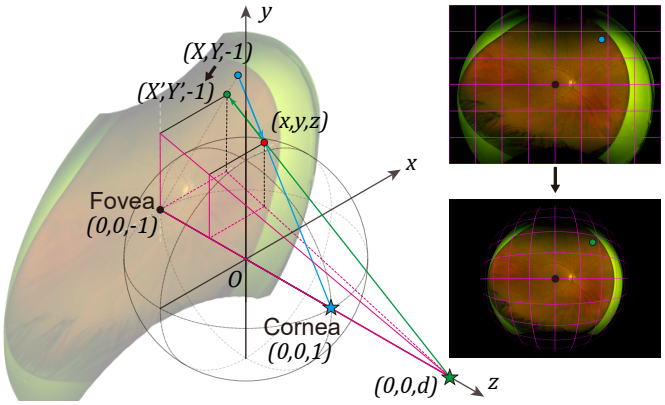


Fig. 3. Illustration for the distortion correction process. On the left side, the original UWF pixel $(X, Y, -1)$ (blue dot) is first mapped to its 3D position (x, y, z) on the eyeball (red dot), and then projected to a new 2D position $(X', Y', -1)$ on the image plane (green dot). The blue and green stars are the view points of UWF and NA images respectively. Right side shows the 2D UWF images before and after correction.

A. 2D-to-3D Projection

Optos's instruments comply with the DICOM standard for Wide Field Ophthalmic Images [25] for UWF imaging and storage. The captured 3D retinal data are transformed into the 2D UWF images based on stereographic projection. When comparing this projection to a pinhole camera projective model, the view point (lens) is located at the cornea, and the optical axis is from cornea to fovea.

We set up a 3D coordinate system as shown in Fig. 3. Specifically, we set the zero point $(0, 0, 0)$ at the sphere center and the sphere radius as 1 (i.e. a unit sphere) for convenience, which is different from the mathematical derivations in the DICOM standard [25]. The coordinate for UWF's view point (cornea) is $(0, 0, 1)$. Besides, in order to reduce peripheral patterns instead of enlarging the fovea in the correction process, we set the UWF imaging plane (sensor) at the back of the eyeball. Therefore, in the 2D-to-3D projection process, the mapping functions between a 2D point $(X, Y, -1)$ on the UWF image plane and its 3D position (x, y, z) on the sphere are written as

$$(x, y, z) = \left(\frac{4X}{4+X^2+Y^2}, \frac{4Y}{4+X^2+Y^2}, \frac{-4+X^2+Y^2}{4+X^2+Y^2}\right) \quad (7)$$

$$(X, Y) = \left(\frac{2x}{1-z}, \frac{2y}{1-z}\right). \quad (8)$$

B. 3D-to-2D Projection

We assume that the 3D-to-2D projection process uses a same optical axis as the stereographic projection, since most of our UWF and NA images are centered on fovea. We set $(0, 0, d)$ as the position of the new viewing point, *i.e.*, the lens position of the NA imaging system, where $d \geq 1$. The 3D point (x, y, z) on the sphere is remapped to $(X', Y', -1)$ on the image plane based on NA's view point $(0, 0, d)$. We have equations from two pairs of similar triangles (shown by

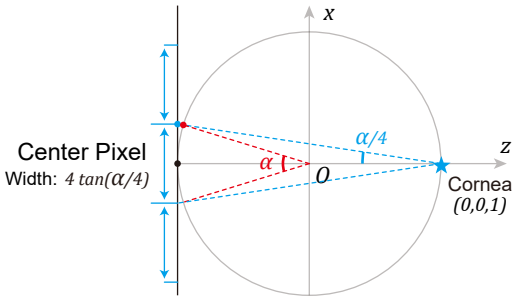


Fig. 4. Conversion between pixel coordinates and sphere coordinates (in $x-z$ plane).

red lines in Fig. 3) as

$$\frac{x}{X'} = \frac{y}{Y'} = \frac{d-z}{d+1}, \tag{9}$$

and the remapped 2D point is derived as

$$(X', Y'; d) = \left(\frac{d+1}{d-z} \cdot x, \frac{d+1}{d-z} \cdot y\right). \tag{10}$$

Inversely, to obtain (x, y, z) from $(X', Y', -1)$, we can combine Eq. (9) with the unit sphere constraint $x^2+y^2+z^2 = 1$, and derive

$$az^2 + bz + c = 0, \tag{11}$$

where $a = (X'^2+Y'^2)/(d+1)^2+1$, $b = 2d(X'^2+Y'^2)/(d+1)^2$, and $c = d^2(X'^2+Y'^2)/(d+1)^2-1$. The solution to z is

$$z = \frac{-b - \sqrt{b^2 - 4ac}}{2a}, \tag{12}$$

where only the point closer to the image plane is used. Finally, we can write the 3D coordinate as

$$(x, y, z; d) = \left(\frac{d-z}{d+1}X', \frac{d-z}{d+1}Y', z\right). \tag{13}$$

C. Coordinate Conversion

In the correction process, a scaling conversion is needed between (m, n) in the image pixel coordinate and (X, Y) in the sphere coordinate (or between (m', n') and (X', Y')). α and β are the view angles (with regard to the sphere center) of the 2D image center pixel in the x and y directions respectively, as shown in Fig. 4. Then, the width and height of the center pixel are $4{\cdot}\tan(\alpha/4)$ and $4{\cdot}\tan(\beta/4)$, which are used as the scaling factors for the conversion. Consequently, the conversion is written as

$$(X, Y) \leftarrow (m \cdot 4\tan(\alpha/4), n \cdot 4\tan(\beta/4)), \tag{14}$$

$$(m, n) \leftarrow \big(X/(4\tan(\alpha/4)), Y/(4\tan(\beta/4))\big). \tag{15}$$

In Optos's UWF DICOM files, α and β are defined in tags (0028,1528) and (0028,1529) as Center Pixel View Angle of X and Y coordinates. In our dataset, we set $\alpha = \beta = 0.08596515°$ for all UWF images, since they are from a same instrument model.

D. Correction Process

To correct the correspondence, *i.e.*, $\mathbf{p} \rightarrow \mathbf{p}'$, we

(a) first scale (m, n) in $\mathbf{p}$ into (X, Y) by Eq. (14),
(b) then obtain its 3D location (x, y, z) via Eq. (7),
(c) next remap it to a new position (X', Y') for a given d via Eq. (10),
(d) finally get the corrected keypoint (m', n') by Eq. (15).

Meanwhile, to correct the UWF image pixels (*e.g.*, $I_{uwf\text{-}s} \rightarrow I'_{uwf\text{-}s}$), we need to get the interpolation position (m, n) in the original image, from every pixel (m', n') in the corrected image. We

(a) first scale (m', n') into (X', Y') via Eq. (14),
(b) then get the 3D position (x, y, z) by Eq. (13),
(c) next obtain (X, Y) via Eq. (8),
(d) finally get the sampling position (m, n) via Eq. (15).

Bilinear interpolation and $\mathrm{STN}(\cdot)$ are used in image warping.

E. View Point Estimation

The optimal view point position $d = d_{optimal}$ is derived through Algorithm 1. The NA vessel map $I_{na\text{-}s}$, the uncorrected UWF vessel map $I_{uwf\text{-}s}$ and their correspondence $\mathbf{p}$ are pre-computed on the training set to avoid repeated computation. The algorithm searches for $d_{optimal}$ in multiple loops with decreasing $step$ (*i.e.*, increasing searching accuracy). In a current $loop$, it searches over several candidate view point positions d_c. At each d_c, the algorithm corrects UWF vessel maps $I_{uwf\text{-}s}$ and correspondence $\mathbf{p}$, then estimates transformation matrix $\mathbf{M}_{poly}$ for each image pair, and warps the corrected UWF vessel map again with the matrix. An average $Dice$ value over all training pairs is computed and stored for each d_c. At the end of each $loop$, $d_{optimal}$ is updated as the d_c with the highest $Dice$. In the next loop, $loop+1$, we search around $d_{optimal}$ in a smaller range $[be, en]$ but with a finer $step$. We do not use Gradient Descent algorithms due to their low speed.

For simplicity, an optimal d is found by searching over the whole training dataset, and then applied for all testing UWF images. We initialize the algorithm by $be = 1$, $en = 3$, $step = 1/4$, $ext = 4$, and $loop = 4$.

IV. Experiments

A. Settings

We collected 116 image pairs, consisting of 200° Optos UWF Colormaps as well as 55° MultiColor (MC) images from Heiderburg's Spectralis platform for the NA modality. The resolution of the MC images is 768×768, and most MC images are centered around the fovea. The original resolution of UWF images is 4000×4000. For each pair, we manually label 3 pairs of corresponding points, and then estimate the 2D affine matrix $\mathbf{M}_{gt}$ as ground-truth.

In the experiments, we crop the center 2000×2000 part of UWF images to remove most non-retina patterns. We randomly separate the dataset by half as Set 1 and 2. We use Set 1 for training and Set 2 for testing, which is denoted as Set 1-2, and vice versa as Set 2-1.

Algorithm 1: Searching for optimal d

Result: An optimal $d = d_{optimal}$.
Initialization: (1) Pairs of UWF and NA vessel maps $\{(I^{(i)}_{uwf\text{-}s}, I^{(i)}_{na\text{-}s}), ...\}$, and correspondence $\{\mathbf{p}_i, ...\}$;
(2) Outlier rejection network (trained and frozen);
(3) A dictionary $D \rightarrow \{k : v\}$ for storing results;
(4) Searching range $[be, en]$, $step$, ext, and $loops$;
for $loops$ **do**
  for $d_c = be : step : en$ **do**
    if d_c exists in D **then**
      Go to Next loop;
    end
    for i-th image pair & correspondence **do**
      Image correction with d_c: $I^{(i)}_{uwf\text{-}s} \rightarrow I^{(i)'}_{uwf\text{-}s}$;
      Keypoint correction with d_c: $\mathbf{p}_i \rightarrow \mathbf{p}'_i$;
      Get transformation matrix from outlier rejection network: $\mathbf{p}'_i \rightarrow \mathbf{w} \rightarrow \mathbf{M}_{poly}$;
      Warp $I^{(i)'}_{uwf\text{-}s}$ based on $\mathbf{M}_{poly}$;
      Compute and save $Dice$ value between $I_{na\text{-}s}$ and the warped $I^{(i)'}_{uwf\text{-}s}$;
    end
    $D[d_c] =$ Average $Dice$ over all pairs;
  end
  Find $d_{optimal}$ from the largest $Dice$ value in D;
  $step = step/2$;
  $be = d_{optimal} - step \times ext$;
  $en = d_{optimal} + step \times ext$;
end

In the learning process, the networks are trained separately as described in Section II-D. (1) In the segmentation network for MC images, we adopt a network [26] pre-trained on Color Fundus images [12], [18] and finetune it on our dataset for 2000 epoches. We set learning rate as $1e$-3, batch size as 1 with $\lambda_{pc} = 1e$-3 and $\lambda_{sm} = 5e$-4. (2) In the outlier rejection network, we use the 2000×2000 UWF images for training and testing. The network is trained for 1000 epoches with learning rate as $1e$-4, batch size as 8, $\lambda_c = 1$, and $\lambda_r = \lambda_d = 0.1$. The model with the highest $Dice$ value on the training set is saved for testing. For both networks, Adam optimizer [27] is used.

In testing, we estimate three transformation models (Affine, Perspective, and Polynomial) from $\mathbf{w}$ and $\mathbf{p}$ (or $\mathbf{p}'$), with/without distortion correction, which results in six settings in evaluation results.

The networks are implemented in PyTorch and trained on a GTX 1080 Ti GPU. It takes about 18 hours to train the segmentation network and about 10 hours to train the outlier rejection network.

B. Evaluation Results

Table I shows the average $Dice$ values and their standard deviations on the two testing sets. We compare our method with a conventional registration pipeline [3] that translates multi-modal images into Monogenical Phase signals [28] for registration. The results of our proposed network (Affine) are better than those in the conventional method by over 0.06 in $Dice$ value. Besides, as the complexity of the transformation model increases (Affine $\rightarrow$ Perspective $\rightarrow$ Polynomial), the average $Dice$ also improves. This indicates that more non-linear transformation is required to correctly align UWF images with narrow-angle MC images. Finally, comparing the results with distortion correction to those without correction, the average $Dice$ improves around 0.03. Especially, the results of Affine + Correction (6+1 parameters) even rivals those of Polynomial (12 parameters), *i.e.*, $+0.03$ on Set 1-2 and -0.04 on Set 2-1. It demonstrates the effectiveness of our proposed distortion correction scheme.

TABLE I
AVERAGE DICE VALUES (STANDARD DEVIATION) ON TESTING SETS

Methods	Set 1-2	Set 2-1
Before Registration	0.1727 (0.0208)	0.1841 (0.0246)
Phase-HoG-Ransac (Affine) [3]	0.3555 (0.1065)	0.3476 (0.0982)
Ours (Affine)	0.4244 (0.0990)	0.4151 (0.0728)
Ours (Affine + Correction)	0.4651 (0.1101)	0.4462 (0.0784)
Ours (Perspective)	0.4455 (0.1087)	0.4409 (0.0759)
Ours (Perspective + Correction)	0.4791 (0.1136)	0.4644 (0.0841)
Ours (Polynomial)	0.4620 (0.1251)	0.4501 (0.0991)
Ours (Polynomial + Correction)	**0.4955** (0.1323)	**0.4818** (0.1089)

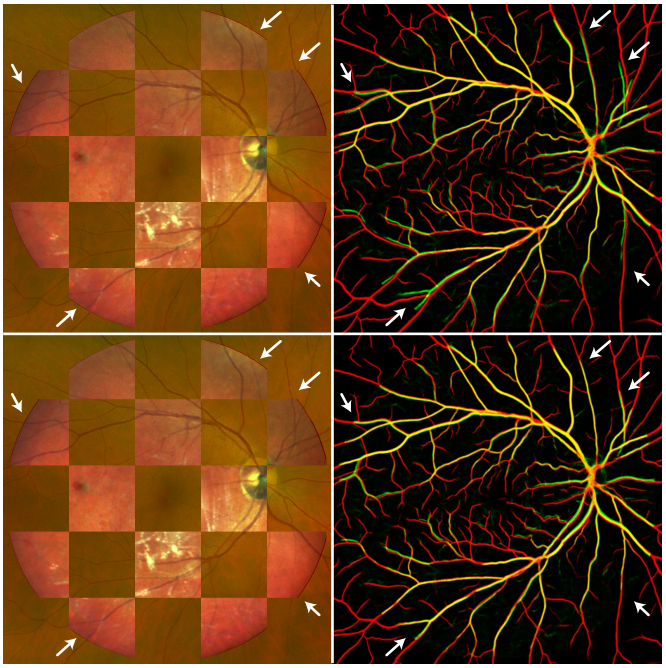
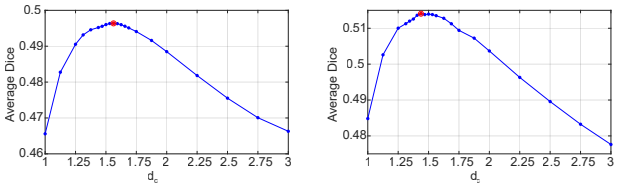


Fig. 5. Registration results (cropped 800×800). Left column shows mosaicked overlay of aligned images. Right column shows overlay of vessels (red for UWF, green for MC, and yellow for the overlapping parts).

In Fig. 5, which shows the registration results on a testing image pair, the alignment quality is improved by incorporating the proposed distortion correction module, especially in MC's peripheral areas where misalignment is reduced after correction as indicated by white arrows.

Fig. 6 shows the searching process for the optimal d on



(a) Set 1-2: $d_{optimal} = 50/32$. (b) Set 2-1: $d_{optimal} = 46/32$.

Fig. 6. Searching process for d on the training sets.

the two training sets. All searched d_c and their corresponding average $Dice$ values are plotted. As shown, the optimal d is achieved at $50/32$ for Set 1-2 and $46/32$ for Set 2-1. In addition, both curves are monotonically increasing and then monotonically decreasing in general.

V. Conclusion

In this paper, we proposed a multi-modal registration pipeline based on CNN for UWF and NA retinal images. We also proposed a distortion correction module that remaps the UWF images from the NA images' view point, so that peripheral distortions in the UWF images are reduced and the registration performance can be improved. In the future, we would extend our method into 3D space incorporating more parameters for UWF distortion correction, and find the optimal viewing distance for each image pair.

References

[1] S. N. Patel, A. Shi, T. D. Wibbelsman, and M. A. Klufas, "Ultra-widefield retinal imaging: an update on recent advances," Therapeutic advances in ophthalmology, vol. 12, pp. 1–12, 2020.

[2] Álvaro S. Hervella, J. Rouco, J. Novo, and M. Ortega, "Multimodal registration of retinal images using domain-specific landmarks and vessel enhancement," Procedia Computer Science, vol. 126, pp. 97 – 104, 2018.

[3] Z. Li, F. Huang, J. Zhang, B. Dashtbozorg, S. Abbasi-Sureshjani, Y. Sun, X. Long, Q. Yu, B. ter Haar Romeny, and T. Tan, "Multi-modal and multi-vendor retina image registration," Biomed. Opt. Express, vol. 9, no. 2, pp. 410–422, 2018.

[4] Z. Ghassabi, 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 Journal on Image and Video Processing, vol. 2013, no. 1, p. 25, 2013.

[5] 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 Transactions on Biomedical Engineering, vol. 57, no. 7, pp. 1707–1718, 2010.

[6] 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 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1046–1053.

[7] H. Zhang, X. Liu, G. Wang, Y. Chen, and W. Zhao, "An automated point set registration framework for multimodal retinal image," in 2018 24th International Conference on Pattern Recognition (ICPR), 2018, pp. 2857–2862.

[8] J. Lee, P. Liu, J. Cheng, and H. Fu, "A deep step pattern representation for multimodal retinal image registration," in 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 5076–5085.

[9] M. Arikan, A. Sadeghipour, B. Gerendas, R. Told, and U. Schmidt-Erfurt, "Deep learning based multi-modal registration for retinal imaging," in Interpretability of Machine Intelligence in Medical Image Computing and Multimodal Learning for Clinical Decision Support, 2019, pp. 75–82.

[10] G. Luo, X. Chen, F. Shi, Y. Peng, D. Xiang, Q. Chen, X. Xu, W. Zhu, and Y. Fan, "Multimodal affine registration for icga and mcsl fundus images of high myopia," Biomed. Opt. Express, vol. 11, no. 8, pp. 4443–4457, 2020.

[11] Y. Wang, J. Zhang, C. An, M. Cavichini, M. Jhingan, M. J. Amador-Patarroyo, C. P. Long, D. G. Bartsch, W. R. Freeman, and T. Q. Nguyen, "A segmentation based robust deep learning framework for multimodal retinal image registration," in 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020, pp. 1369–1373.

[12] Y. Wang, J. Zhang, M. Cavichini, D.-U. G. Bartsch, W. R. Freeman, T. Q. Nguyen, and C. An, "Robust content-adaptive global registration for multimodal retinal images using weakly supervised deep-learning framework," IEEE Transactions on Image Processing, vol. 30, pp. 3167–3178, 2021.

[13] T. Chanwimaluang and G. Fan, "Constrained optimization for retinal curvature estimation using an affine camera," in 2007 IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.

[14] E. Ataer-Cansizoglu, Y. Taguchi, J. Kalpathy-Cramer, M. F. Chiang, and D. Erdogmus, "Analysis of shape assumptions in 3d reconstruction of retina from multiple fundus images," in 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), 2015, pp. 1502–1505.

[15] T. Probst, K.-K. Maninis, A. Chhatkuli, M. Ourak, E. V. Poorten, and L. Van Gool, "Automatic tool landmark detection for stereo vision in robot-assisted retinal surgery," IEEE Robotics and Automation Letters, vol. 3, no. 1, pp. 612–619, 2018.

[16] T. Dan, Z. Fan, Y. Hu, B. Zhang, G. Tao, and H. Cai, "Reconstruction of 3d retina from multi-viewed stereo fundus images via dynamic registration," in 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2020, pp. 980–984.

[17] L. Ding, A. E. Kuriyan, R. S. Ramchandran, C. C. Wykoff, and G. Sharma, "Weakly-supervised vessel detection in ultra-widefield fundus photography via iterative multi-modal registration and learning," IEEE Transactions on Medical Imaging, pp. 1–1, 2020.

[18] J. Zhang, C. An, J. Dai, M. Amador, D. Bartsch, S. Borooah, W. R. Freeman, and T. Q. Nguyen, "Joint vessel segmentation and deformable registration on multi-modal retinal images based on style transfer," in 2019 IEEE International Conference on Image Processing (ICIP), 2019, pp. 839–843.

[19] J. J. Yu, A. W. Harley, and K. G. Derpanis, "Back to basics: Unsupervised learning of optical flow via brightness constancy and motion smoothness," in Computer Vision – ECCV 2016 Workshops, 2016, pp. 3–10.

[20] D. DeTone, T. Malisiewicz, and A. Rabinovich, "Superpoint: Self-supervised interest point detection and description," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2018, pp. 337–33 712.

[21] K. M. Yi, E. Trulls, Y. Ono, V. Lepetit, M. Salzmann, and P. Fua, "Learning to find good correspondences," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 2666–2674.

[22] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," Commun. ACM, vol. 24, no. 6, p. 381–395, 1981.

[23] Y. Gavet, M. Fernandes, and J.-C. Pinoli, "Quantitative evaluation of image registration techniques in the case of retinal images," Journal of Electronic Imaging, vol. 21, no. 2, pp. 1 – 8, 2012.

[24] M. Jaderberg, K. Simonyan, A. Zisserman, and k. kavukcuoglu, "Spatial transformer networks," in Advances in Neural Information Processing Systems 28, 2015, pp. 2017–2025.

[25] DICOM Standards Committee. (2015) Digital imaging and communications in medicine (dicom) supplement 173: Wide field ophthalmic photography image storage sop classes. [Online]. Available: https://www.dicomstandard.org/News-dir/ftsup/docs/sups/sup173.pdf

[26] K.-K. Maninis, J. Pont-Tuset, P. Arbeláez, and L. Van Gool, "Deep retinal image understanding," in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016, 2016, pp. 140–148.

[27] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," CoRR, vol. abs/1412.6980, 2014.

[28] M. Felsberg and G. Sommer, "The monogenic signal," IEEE Transactions on Signal Processing, vol. 49, no. 12, pp. 3136–3144, 2001.