A Convolutional Neural Network Pipeline For Multi-Temporal Retinal Image Registration

Chi-Jui Ho, Yiqian Wang, Junkang Zhang, Truong Nguyen and Cheolhong An
Department of Electrical and Computer Engineering, UC San Diego, La Jolla, USA

Abstract— **A sequence of images is usually captured to observe the change of health status in medical diagnosis. However, an image sequence taken over year usually suffers from severe deformation, making it time-consuming for physicians to match corresponding patterns. In this paper, we propose a coarse-to-fine pipeline for retinal image registration based on convolutional neural network. By leveraging the three components of the pipeline: feature matching, outlier rejection, and local registration, we recover the deformation and accurately align multi-temporal image sequences. Experimental results show that the proposed network is robust to severe deformation as well as illumination and contrast variations. With the proposed registration pipeline, the change of image patterns over time can be identified through visual analysis.**

Keywords: Image temporal registration, retinal imaging, CNN

I. INTRODUCTION

Physicians usually rely on a sequence of retinal images to observe the change of retina health over time. However, a multi-temporal retinal image sequence usually suffers from misalignment due to the variation of measuring device placement, making it challenging for physicians to match pathological patterns in multi-temporal. Moreover, the variations of illumination and contrast in image sequence further increase the difficulties. To address these issues, image registration algorithms have been developed [1]–[3]. It takes an image pair as input and estimates a transformation function that characterizes the deformation. A pipeline of coarse image registration usually involves feature extraction and outlier rejection [2]. The former extracts key points from pair of images, and the latter generates a transformation matrix accordingly. To increase the flexibility to local deformations, a fine-tuning step is usually integrated to the registration pipeline and performs fine registration [4].

Image registration has been developed in classical and deep learning approaches. Classical approaches include speed up robust features [5], and random sample consensus [6]. Most deep learning frameworks are based on convolutional neural network (CNN), which has been applied to feature description [8], outlier rejection [2, 7], and fine registration [1]. However, the integration of various components is in early development. In this paper, we propose a pipeline that integrates coarse and fine registration for multi-temporal image. It consists of a super point network (SPN) [8], an outlier rejection network, and an encoder-decoder network. The experimental results show that the proposed pipeline outperforms existing method for multi-temporal image registration. It also shows the variations of patterns over time can be identified by visual analysis.

TABLE I. DICE COEFFICNET OF IMAGE REGISTRATION METHODS

Methods	Dice Coefficient
w/o registration	0.1819
VGG-based [9]	0.2676
IC [3]	0.2038
Coarse registration (ours)	0.5685
Coarse-to-fine registration (ours)	**0.6604**

II. RELATED WORK

Multi-Temporal Image Registration: To observe the temporal variation, multi-temporal registration has been developed to aerial imaging, where Yang *et al.* adopted a pre-trained classification network to describe image features [9]. The application also includes magnetic resonance images, where Zhang proposed a loss function that constraints inverse consistency in unsupervised training [3].

III. PROPOSED APPROACH

A. Coarse Registration: Feature Matching: To alleviate the impact of changes of illuminance and contrast on image registration, we employ a segmentation network [1] to extract the vessel structure while eliminating the irrelevant information of retinal images. Then, we apply SPN to detect interest points and corresponding features from source and target images, called source and target features. A source and target feature are collected in a set of corresponding pair if they are the nearest neighbor of each other in latent space. Therefore, a set of corresponding pairs $\{p_1, \ldots, p_n\}$ is collected where each pair p_i is a 4-dimensional vector that consists of coordinate $\{x_i, y_i\}$ and $\{v_i, u_i\}$ from source and target images, respectively.

B. Coarse Registration: Outlier Rejection: To estimate transformation function from corresponding pairs, an outlier rejection network is applied [2]. It estimates a score ranging from 0 to 1 for each inlier pair according to its significance. Denote a pair of key points and the score as $p = (x, y, u, v)$ and s, respectively. A corresponding matrix m is computed by

$$m = diag([s, s])A, \tag{1}$$

where A is a 2×9 matrix

$$\begin{bmatrix} -x & -y & -1 & 0 & 0 & 0 & vx & uy & u \\ 0 & 0 & 0 & -x & -y & -1 & vx & vy & v \end{bmatrix}$$

Given n valid corresponding pairs, a 2n×9 matrix M is obtained by concatenating $\{m_1, \ldots, m_n\}$ and the transformation function, a 3×3 homography matrix H, is obtained by

$$H = \text{argmin}_{\text{H}} ||MVec(H)||, \quad (3)$$

where $Vec(\cdot)$ denotes vectorized operation.

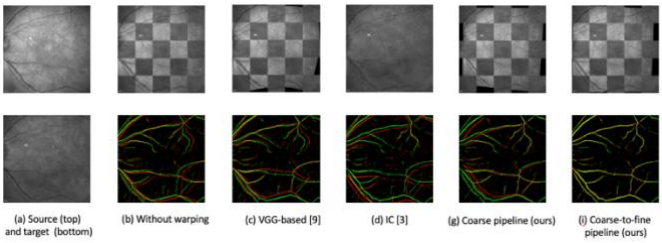

Figure 1. Registration results of (a) source and target images from different approaches. Except for the left most column, the top and bottom rows show the aligned raw image and segmentation map, respectively (zoom in to see more details).

C. Fine Registration: We note that homography matrix may not characterize local deformations, especially when a branch of vessels lacks sufficient corresponding pairs. To address this issue, we rewarp the source image by a fine registration model [1]. It takes a stacked target and a warped source images as input and outputs a two-channel registration map, representing horizontal and vertical displacement. The model adopts an encoder-decoder architecture, performing four-level down-sampling and up-sampling through seven convolution and deconvolution layers.

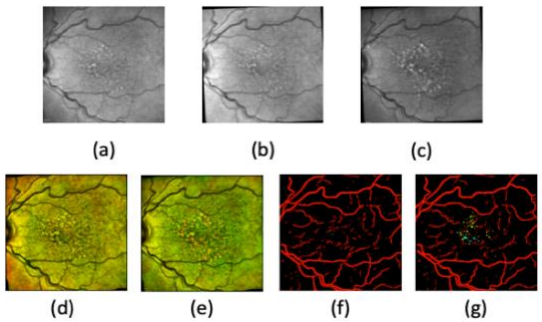

Figure 2. Visual alignment of a target image (a) and two source images (b) and (c). The stacked images and the residual images between the target image (a) and the warped images of (b) and (c) are shown in (d), (f) and (c), (g), respectively.

IV. Experiments

The dataset contains retinal image sequences collected from 10 different patients by Jacobs Retina Center at Shiley Eye Institute. For each image sequence, we set the image that was first captured as the target image, while the remaining ones as source images. We compared the results of the proposed pipeline with two multi-temporal registration approaches [3, 9] with dice coefficient, which evaluates correspondence between two segmentation maps S_1 and S_2:

$$Dice(S_1, S_2) = \frac{2\,\Sigma(S_1 \circ S_2)}{\Sigma\, S_1 + \Sigma\, S_2}, \quad (3)$$

where ° denotes Hadamard product. Table I shows that the proposed pipeline yields the highest dice and that fine registration improves the dice score by 0.0919. Qualitative analysis is also provided. Fig. 1 shows chessboard-like images that alternatively show the patches from warped source and target images. It also shows stacked warped and target segmentation maps, assigned in red and green channels, respectively. The proposed algorithm enables continuities in chessboard-like images and aligns most vessels in overlapped segmentation map, which is not the case for other approaches.

We further visualize the variation of a retina over time. Fig. 2 (b) and (c) show the images captured 3 and 28 months after the target image (a). After registration, we stacked warped and target images, assigned in red and green channels, respectively. We also show residual images: When a pixel value of warped image is higher (lower) than that of target image by over 90, we colorized it in cyan (yellow). The stacked images of (a) and registered (b) and (c) are shown in (d) and (e), respectively, while residual images are shown in (f) and (g), respectively. As observed, many cyan and yellow pixels appear in the center of (g), but not (f). It verifies that the propagation of pathological patterns is significant after 28 months, but subtle after 3 months.

V. Conclusion

We propose a CNN-based pipeline for multi-temporal retinal image registration. Unlike other methods, the proposed pipeline provides accurate alignment regardless of the change of contrast. The superior registration performance of proposed pipeline enables accurate image alignment over time and hence broadens the applications of image registration in clinical analysis.

Acknowledgment

We would like to thank Shiley Eye Center for providing retinal images for this study.

References

[1] J. Zhang *et al.* "Joint vessel segmentation and Deformable Registration on Multi-Modal Retinal Images Based on Style Transfer," in *IEEE Int. Conf. Image Proc.*, 2019, pp. 839–843.

[2] Y. Wang *et al.* "A segmentation based robust deep learning framework for multimodal retinal image registration," in *IEEE Int. Conf. Acous, Speech and Sig. Proc.*, 2020, pp. 1369–1373.

[3] Zhang, "Inverse-consistent deep networks for unsupervised deformable image registration," in *arXiv preprintarXiv:1809.03443, 2018.*

[4] Z. Li *et al.*, "Multi-modal and multi-vendor retina imageregistration," in *Biomedical optics express*, vol. 9, no. 2, pp. 410–422, 2018.

[5] H. Bay *et al.*, "Surf: Speededup robust features," in *Euro. Conf. Comp. Vis.*, 2006, pp. 404–417.

[6] M. A. Fischler *et al.* "Random sample con-sensus: a paradigm for model fitting with applicationsto image analysis and automated cartography," in *Com.* of the ACM, vol. 24, no. 6, pp. 381–395, 1981.

[7] M. Y. Kwang *et al.* "Learning to find good correspon-dences," in *Proc. IEEE Conf. Comp. Vis. and Patt. Recog.*, 2018, pp. 2666–2674.

[8] D. DeTong *et al.* "Super-point: Self-supervised interest point detection and description," in *Proc. IEEE Conf. Comp. Vis. and Patt. Recog. Work.*, 2018, pp. 224–236.

[9] Z. Yang *et al.* "Multi-temporal remotesensing image registration using deep convolutional features," in *IEEE Access*, vol. 6, pp. 38544–38555, 2018.