Published in final edited form as:
Med Image Anal. 2014 August ; 18(6): 903–913. doi:10.1016/j.media.2013.09.009.

Landmark Matching Based Retinal Image Alignment by Enforcing Sparsity in Correspondence Matrix

Yuanjie Zheng[1], **Ebenezer Daniel**[2], **Allan A. Hunter III**[2], **Rui Xiao**[3], **Jianbin Gao**[5], **Hongsheng Li**[5], **Maureen G. Maguire**[2], **David H. Brainard**[4], and **James C. Gee**[1]

[1]Department of Radiology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

[2]Department of Ophthalmology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

[3]Department of Biostatistics and Epidemiology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

[4]Department of Psychology, School of Arts and Sciences at the University of Pennsylvania, Philadelphia, PA, USA

[5]University of Electronic Science and Technology, Chengdu, Sichuan, China

Abstract

Retinal image alignment is fundamental to many applications in diagnosis of eye diseases. In this paper, we address the problem of landmark matching based retinal image alignment. We propose a novel landmark matching formulation by enforcing sparsity in the correspondence matrix and offer its solutions based on linear programming. The proposed formulation not only enables a joint estimation of the landmark correspondences and a predefined transformation model but also combines the benefits of the soft assign strategy (Chui and Rangarajan, 2003) and the combinatorial optimization of linear programming. We also introduced a set of reinforced self-similarities descriptors which can better characterize local photometric and geometric properties of the retinal image. Theoretical analysis and experimental results with both fundus color images and angiogram images show the superior performances of our algorithms to several state-of-the-art techniques.

Keywords

Image alignment; sparsity; retinal image; point matching; transformation; image registration

1. Introduction

Retinal images are widely used in diagnosing and monitoring the progress of a variety of eye diseases, such as diabetic retinopathy (Winder et al., 2009), age-related macular degeneration (Wood et al., 2000) and glaucoma (Quigley et al., 1982). As a process of establishing spatial correspondences between two retinal images, retinal image alignment (Kolar et al., 2008a; Treigys et al., 2008; Dhillonb et al., 2006) is fundamental to applications as diverse as detecting locations of scars and burns, tracking the progress of diseases (Narasimha-Iyer et al., 2006), mosaicing or constructing montage to provide a wider view of the retina, fusing images from different modalities (Cao, 2008) and creating tools to assist ophthalmologists during laser surgery or other procedures (Can et al., 2002; Tobin et al., 2007).

However, automatic retinal image alignment has remained a difficult problem due to several challenges. *First*, transformation for modelling the spatial correspondences between two retinal images can be very complicated. Two contributing factors here are that the retina is a curved surface and that the retinal images can be projections of this surface from a wide range of view points with an un-calibrated camera. *Second*, there exist large homogeneous areas in retinal images. Nonvascular surface occupies much of the retina and can be textureless for healthy retina but exhibit a variety of pathologies for unhealthy retinas. *Third*, large appearance variations may be observed between the images to be registered. It is because the pathologies can appear and disappear over time, the effects of disease and poor image quality can obscure the vasculature (Stewart et al., 2003) and images may be captured under vastly different illumination conditions or acquired using different modalities. To handle these challenges, there are two keys to a successful retinal image alignment algorithm: selection of optimization strategy to more efficiently estimate a predefined transformation model and extraction of reliable features robust to image variations.

Landmark matching (Can et al., 2002; Zana and Klein, 1999; Chanwimaluang et al., 2006; Stewart et al., 2003; Lee et al., 2007) is more attractive in retinal image alignment compared with pixel-wise based image registration (a brief review to the existing image registration techniques is provided in Sec. 2). It can deal with the large homogenous areas and appearance variations in retinal images as mentioned above. Moreover, landmarks are also used in practice by ophthalmologist (Can et al., 2002). Disregarding certain encouraging results produced by some previous methods (Can et al., 2002; Stewart et al., 2003; Chanwimaluang et al., 2006), at least two weaknesses were recognized in our research. *First*, features describing landmarks are not always distinguishable. Most previous features (Chanwimaluang et al., 2006) are extracted from the binary vascular image in which a lot of information of the local image pattern is lost. Moreover, landmarks are not uniquely identified by the widely used features characterized with vessel orientations (Can et al., 2002; Zana and Klein, 1999; Stewart et al., 2003) because branching angles at bifurcations can be similar across the retina. *Second*, optimization strategies for estimating the predefined transformation model are less efficient. Most previous methods perform an iterative optimization process by repeating the estimations of correspondences and transformation model, which requires a good initialization and might lead to an inferior local estimation. As discovered by us and also mentioned in (Can et al., 2002), joint estimation of the correspondences and an appropriately predefined transformation model can result in a superior estimation. Unfortunately, no effective optimization approach has been described yet to accomplish this joint estimation for effectively aligning retinal images.

This paper addresses the problem of landmark matching based automatic retinal image alignment. We propose a novel landmark matching formulation by enforcing sparsity in the correspondence matrix and offer its solutions which are obtained using linear programming (LP) (Zheng et al., 2011). The proposed formulation not only enables a joint estimation (as shown in Fig. 1) of the landmark correspondences and the predefined transformation model but also combines the benefits of the soft assign strategy (Chui and Rangarajan, 2003) and the combinatorial optimization of LP. We also introduce a set of reinforced self-similarities descriptors (Shechtman and Irani, 2007) to better characterize the photometric and geometric properties of the retinal image. Our theoretical analysis and experimental results using both fundus color images and angiogram images show the superior performances of our methods to sate-of-the-art techniques in sense of not only the optimization ability but also the distinguishing capacity of landmark descriptors.

Our sparsity constraint enforced to the correspondence matrix is inspired by the sparsity prior (Donoho, 2006; Candes and Romberg, 2007) which has been widely used in the communities of signal processing, computer vision, multimedia and medical imaging. This matrix sparsity is accomplished by minimizing a proposed measurement called "to-centroid deviation". Minimizing this measurement can encourage the matched reference points (to each floating point) to be as spatially concentrated as possible. Obviously, this strategy is very different from the L_1 norm minimization (Donoho, 2006; Candes and Romberg, 2007; Huang et al., 2009; Zhang et al., 2012; Waters et al., 2011; Candes and Tao, 2005) which is able to enforce sparsity but incapable of obliging the spatial concentration.

2. Previous Work

Image registration has been extensively involved in a wide variety of practical tasks, including retinal image analysis, medical image analysis in a more general sense, various tasks in computer vision, etc. There exist a huge number of related publications (Gholipour et al., 2007; Modersitzki, 2004). These previous methods can basically be categorized as pixel/voxel-wise based, landmark/feature based or their combinations. Detailed review to all of them is beyond the scope of this paper. In this paper, we only provide a brief review to methods of retinal image registration and to the optimization techniques of landmark matching based image registration.

For retinal image registration, both landmark matching based methods (Can et al., 2002; Zana and Klein, 1999; Chanwimaluang et al., 2006; Stewart et al., 2003) and pixel-wise strategies (Kolar et al., 2008b; Matsopoulos et al., 1999) were explored. However, landmark matching has attracted more interests due to several reasons as explained in Sec. 1. In these works, landmarks are mostly placed at cross points or bifurcations of vessel, and characterized by for example coordinates, orientations or widths of the surrounding vessels. However, it is well known that these features are not robust and not always well distinguishable in practice, which may introduce severe matching ambiguities. For the interimage transformation, parametric models used in previous work include weak affine (Zana and Klein, 1999), affine (Can et al., 2002; Kolar et al., 2008b; Matsopoulos et al., 1999), bilinear/projective (Matsopoulos et al., 1999) and the 12-parameter quadratic model (Can et al., 2002). To estimate these transformation models, optimization techniques include iterative closest point (ICP) algorithm (Chanwimaluang et al., 2006), dualbootstrap ICP (Stewart et al., 2003), Hough Transform (Zana and Klein, 1999), and traditional global optimization techniques such as simulated annealing and genetic algorithm (Matsopoulos et al., 1999).

In those more general fields of medical image analysis and computer vision, landmark matching based image registration has been treated as a fundamental task (Gholipour et al., 2007; Berg et al., 2005; Zhan et al., 2006). To obtain a high matching accuracy, optimization also plays a vital role. There are basically three classes of optimization techniques in previous work. The first class assumes the correspondences to be pre-known and only resolves the transformation model (Shen and Davatzikos, 2002; Brown and Lowe, 2007; Rohr et al., 2001). The second class does not explicitly assume any specific transformation model but solely solves the correspondences with various optimization techniques such as the Hungarian method (Belongie et al., 2001) and the quadratic programming (Berg et al., 2005). The third class handles the joint estimation of the correspondences and transformation model. However, for simplicity, an iterative process by repeating the estimations of each of them was more widely used, e.g. the EM like technique in (Chui and Rangarajan, 2003) and the iteratively updating process in (Zheng and Doermann, 2006). With this iterative strategy, only a local optimization can be attained and the accuracy is conditioned on the initialization. In contrast, recently explored joint estimation with combinatorial optimization techniques (Jiang and Yu, 2009; Li et al., 2013, 2011) is free from the iterative process and can simultaneously solve the correspondences and the transformation model. These methods are free from initialization and able to produce a global optimization. Unfortunately, only some simple spatial transformations have been investigated (Jiang and Yu, 2009; Li et al., 2013, 2011).

At the same time, the sparsity prior (Donoho, 2006; Candes and Romberg, 2007) on a signal, image or volumetric data has inspired many research methods (Huang et al., 2009; Zhang et al., 2012; Waters et al., 2011) in various fields (e.g. computer vision, multimedia processing and medical imaging). Numerous greedy algorithms have been developed to address the sparsity optimization problem, including greedy techniques for solving an L_0 norm problem such as the orthogonal matching pursuit (Chen et al., 1989) and stagewise OMP (Donoho, 2006) and methods for dealing with a relaxed L_1 norm problem such as LARS-Lasso (Tibshirani, 1996) and interior point (Koh et al., 2007). Recovering a sparse matrix with the LP optimization technique (Candes and Tao, 2005) has also attracted a lot of research interests.

3. Landmark Matching by Enforcing Sparsity in Correspondence Matrix

We provide a new formulation for the landmark matching problem, which is obtained by both using the softassign strategy (Chui and Rangarajan, 2003) and enforcing sparsity property in the correspondence matrix. We also present solutions of the formulation by integrating the LP optimization technique. Compared with previous work, the proposed algorithms can simultaneously solve the landmark correspondences and the predefined transformation model. This joint estimation (as shown in Fig. 1) is able to result in superior accuracy and robustness due to the fact that mutual aids of the correspondences and transformation model can be maximized in the joint optimization process compared with estimating them independently. Moreover, the proposed formulation combines the benefits of the soft assign strategy and the combinatorial optimization of LP, which helps to generate better optimized estimations.

3.1. Problem Formulation

Suppose we have one reference-landmark set $\Omega^r=\{\xi_i^r, i_r=1,2,\cdots,N_r\}$ and one floating-landmark set $\Omega^m=\{\xi_i^m, i_m=1,2,\cdots,N_m\}$ where N_r and N_m represent the number of reference-landmarks and floating-landmarks, respectively. Taking the reference-landmark i as an example, ξ_i^r is expressed by the features (denoted by vertical vector $\mathbf{v}_i^r$) extracted to describe the visual appearance of this point and its coordinates, i.e. $\xi_i^r=[\mathbf{v}_i^{rT} x_i^r y_i^r]^T$ where T means transpose of the vector. For brevity, we represent $\chi_i^r=[x_i^r y_i^r]^T$, then $\xi_i^r=[\mathbf{v}_i^{rT} \chi_i^{rT}]^T$. For later usages, we provide vector ${\chi'}_i^r=[x_i^r y_i^r 1]^T$ to denote the corresponding homogeneous coordinates, vector

$${\chi''}_i^r=[x_i^{r2} x_i^r y_i^r y_i^{r2} x_i^r y_i^r 1]^T$$

to be used later for representing the 12-parameter transformation model in (Can et al., 2002), matrix $\chi^r=[\chi_1^r \chi_2^r \cdots \chi_{N_r}^r]^T$, matrix ${\chi'}^r=[{\chi'}_1^r {\chi'}_2^r \cdots {\chi'}_{N_r}^r]^T$, and matrix ${\chi''}^r=[{\chi''}_1^r {\chi''}_2^r \cdots {\chi''}_{N_r}^r]^T$. We omit the similar symbols for the floating-landmarks for brevity.

Point matching is defined as estimating a mapping γ which indicates that a floating-landmark i_m corresponds to a reference-landmark i_r, i.e. $i_r = \gamma(i_m)$. Mapping γ can be represented by a correspondence/assignment matrix (denoted by E) (Jiang and Yu, 2009) or a pre-defined transformation model (denoted by $\mathscr{T}$). $\mathscr{T}$ is usually represented by a parametric model in order for robustness to noise/outliers and for reducing the involved computational burden.

Our point matching scheme is formulated as jointly estimating the correspondences and transformation model by maximizing the feature matching quality and the transformation model compliance quality. Feature matching quality represents how similar the features $\mathbf{v}^m$ of each floating-landmark are to $\mathbf{v}^r$ of its corresponding reference-landmark. Transformation model compliance quality evaluates how well the coordinates $\mathbf{x}^m$ of each floating point and $\mathbf{x}^r$ of its corresponding reference-landmark comply with the estimated transformation model.

3.1.1. Correspondence Matrix—We first define the correspondence matrix E as a binary matrix and then relax it later. E is of size $N_m \times N_r$, for which each row contains exactly one 1. If a floating-landmark i_m is matched to the reference-landmark i_r, then E_{i_m,i_r} =

1 and other elements of the i_m th row all equal to 0. These can be stated alternatively as the following two constraints:

$$E \in \{0,1\}^{N_m \times N_r}, \quad (1)$$

$$E\mathbf{1}=\mathbf{1}, \quad (2)$$

where **1** is a vertical vector of ones. Obviously, E is a huge but sparse matrix.

Discreteness of the values taken by the elements of E in Eq. (1) introduces hardships into designing an efficient optimization algorithm for landmark matching. Many papers (Chui and Rangarajan, 2003; Jiang and Yu, 2009) tried to relax it to a continuous value within [0, 1]. This softassign strategy (Chui and Rangarajan, 2003) can make the resulting energy function better behaved. This relaxation can be guaranteed by Eq. (2) together with the below constraint

$$E \geq 0. \quad (3)$$

Unfortunately, this softassign strategy can result in a non-sparse correspondence matrix in practice, which may cause ambiguities in the matching process. Imagine that one floating-landmark is matched to five reference-landmarks and the matching likelihood value equals to 0.2 for each of them. We then have no idea to decide the assignment. To deal with this dilemma and inspired by the sparsity prior (Donoho, 2006; Candes and Romberg, 2007), we propose to to get a sparse correspondence matrix in which the number of nonzero elements in each row of E is as small as possible. Instead of minimizing the L_1 norm as in existing papers using sparsity (Donoho, 2006; Candes and Romberg, 2007; Huang et al., 2009; Zhang et al., 2012; Waters et al., 2011; Candes and Tao, 2005), we propose to penalize the to-centroid spatial deviations of the matched reference-landmarks. As shown in Fig. 2, the proposed penalization strategy can not only generate a sparse matrix but also enforce a spatial concentration on the matched landmarks. In contrast, to minimize the L_1 norm can only guarantee sparsity on E.

To define the to-centroid spatial deviations, we first locate the centroid of the matched reference-landmarks by computing its coordinates with

$$\bar{\chi}^r = E\chi^r. \quad (4)$$

We then use $\bar{\chi}_x^r$ and $\bar{\chi}_y^r$ to denote the vertical vectors concatenated by the x-axis and y-axis coordinates of all centroids, respectively, and denote $\bar{\chi}^r = [\bar{\chi}_x^r \bar{\chi}_y^r]$. Then, the to-centroid deviation matrix is written as

$$D=\sqrt{\left(\bar{\chi}_x^r \mathbf{1}^T - \mathbf{1}(\chi_x^r)^T\right)^2 + \left(\bar{\chi}_y^r \mathbf{1}^T - \mathbf{1}(\chi_y^r)^T\right)^2} \quad (5)$$

where length of **1** is set to guarantee matrix D_r is of size $N_m \times N_r$.

Penalizing the to-centroid deviations amounts to minimizing the function

$$\mathscr{O}_c = tr(D^T E), \quad (6)$$

where *tr*() denotes the trace of matrix. Minimization of the objective function in Eq. (6) with the constraints in Eq. (2) and Eq. (3) and other criteria explained below will result in a correspondence matrix E taking a continuous values in [0, 1] but with a significant bias towards 0 or 1.

3.1.2. Feature Matching Quality—

Feature matching quality measures how similar the visual appearance of each floating-landmark is to the matched reference-landmark(s). We use the reinforced self-similarities (to be proposed in Sec. 4) to describe the landmarks, which is invariant to local affine deformation, radially increasing non-rigid deformation, and rotation. We use the negative of the correlation between two series of features as the matching cost. Similar to (Jiang and Yu, 2009), for each landmark, we compute the features with different angles. The similarity between any possible pair of floating-landmark and reference-landmark is measured by the minimal cost value of features across all angles. We then obtain a feature matching cost matrix C in size $N_m \times N_r$, for which the element $C_{i,j}$ means the cost matching the ith floating-landmark to the jth reference-landmark. Note that C can be computed beforehand.

Maximization of feature matching quality is then expressed as the minimization of the below objective function:

$$\mathscr{O}_f = tr(C^T E). \quad (7)$$

3.1.3. Transformation—

We propose to apply the constraints of the predefined transformation model on the position's relations of each floating-landmark and the corresponding centroid of its matched reference-landmarks. As an example, the 12-parameter transformation model in (Can et al., 2002) can be enforced with equation

$$\bar{\chi}^r = \chi''^m \mathbf{\Theta}^T, \quad (8)$$

where $\bar{\chi}^r$ and χ''^m are defined previously and Θ is a 2×6 matrix. In Eq. (8), only the elements of Θ are unknowns. This transformation model is obtained in (Can et al., 2002) for retinal imaging by assuming a quadratic surface for the retinal, a rigid transformation between two viewpoints, and a weak-perspective camera projection model. All involved parameters are combined in Θ.

When the transformation between the two landmark sets is deformable, the TPS model (Chui and Rangarajan, 2003) can be employed, as expressed by

$$\bar{\chi}^r = \chi^m + \chi'^m A^T + \mathbf{\Phi}\Delta^m \quad (9)$$

where A is a 2×3 matrix containing the parameters of an affine transformation, Φ is a $N_m \times N_m$ symmetric matrix containing the information about the floating-landmark set's internal structural relationships and its elements can be pre-computed as $\Phi_{i,j}=\|\chi_i^m-\chi_j^m\|^2\log\|\chi_i^m-\chi_j^m\|$, and Δ^m is an $N_m \times 2$ matrix for which each row denotes the non-affine deformation of the corresponding floating-landmark. In Eq. (9), A and Δ^m are unknowns and need to be estimated.

Eq. (8) and Eq. (9) are all linear to the unknowns and this linearity leads to an LP based solution as detailed in next section. All transformation models with this linearity property can be incorporated in our matching scheme, which can cover the widely used general models as diverse as affine, elastic, quadratic, and deformable represented by Thin-plate Spline (TPS) or the FFD-BSpline (Rueckert et al., 1999), and some specific models designed for retinal imaging, e.g. the 12-parameter model in (Can et al., 2002). In the followings, we provide a discussion on the model in (Can et al., 2002) and the TPS based deformable model as examples.

3.2. Optimization

The optimization scheme for our point matching task can be formulated as a constrained minimization problem of an integrated objective function:

$$\mathcal{O}=\mathcal{O}_f+\lambda\mathcal{O}_c=tr\left((C+\lambda D)^T E\right) \quad (10)$$

where $\mathcal{O}_f$ and $\mathcal{O}_c$ are defined by Eq. (7) and Eq. (6), respectively, and λ is an adjusting parameter to control the "softness" of the estimated correspondence matrix E. For this optimization scheme, constraints are expressed in Eq. (2), Eq. (3), Eq. (8) or Eq. (9), and unknowns are the correspondence matrix E and the transformation parameters: Θ in Eq. (8), A and Δ^m in Eq. (9).

One benefit of our optimization scheme is that it can be accomplished with linear programming (LP). It is obvious because the objective function and the constraints involved in the optimization scheme are all linear to the unknowns if removing $\mathcal{O}_c$ from Eq. (10) (i.e. setting $\lambda = 0$). With $\mathcal{O}_c$, the objective function is not linear any more due to the nonlinearity of $\mathcal{O}_c$ to E in Eq. (6). Fortunately, $\mathcal{O}_c$ can be linearized by computing D with Eq. (5) while fixing E in Eq. (4). Hence, our optimization scheme is an iterative process. At each iteration, the estimation of E in previous iteration is used to compute D. For the first iteration, E can be estimated by removing $\mathcal{O}_c$ from Eq. (10). *This means that our landmark matching technique does not require any manual initialization*.

3.2.1. Formulation as Linear Programming—We provide expressions in a canonical form for our LP based optimization scheme by taking the transformation in Eq. (8) as an example. Other models can be handled with a similar way.

We first define some symbols. We treat all elements in the correspondence matrix E, matrix Θ in Eq. (8), matching cost matrix C, and the to-centroid deviation matrix D as a single column and denote them with vertical vectors **e**, θ, **c**, and **d**, respectively.

We then create a binary permutation matrix P_e in size $N_m \times (N_m \cdot N_r)$, such that $P_e\mathbf{e} = E\mathbf{1}$. We then have the reformulation of Eq. (2):

$$P_e\mathbf{e}=\mathbf{1}. \quad (11)$$

We further divide Eq. (8) into two parts: *x*-mapping and *y*-mapping, and represent each of them as a linear function of Θ. The *x*-mapping is written as

$$\bar{\chi}_x^r=\chi_x^m+\chi'^m\mathbf{\Theta}_x^T \quad (12)$$

where $\bar{\chi}_x^r$ is the first column corresponding to the *x*-coordinates of $\bar{\chi^r}$, χ_x^m is the first column of χ^m, and Θ_x is the first row of Θ and in charge of *x*-mapping. It is easy to further create a permutation matrix P_θ^x in size of 6 × 12, such that $\mathbf{\Theta}_x^T=P_\theta^x\theta$. Therefore, Eq. (12) can be rewritten as

$$\chi'^m P_\theta^x\theta=\bar{\chi}_x^r-\chi_x^m. \quad (13)$$

With $Q_\theta^x=\chi'^m P_\theta^x$ and $b_\theta^x=\bar{\chi}_x^r-\chi_x^m$, Eq. (13) can be rewritten concisely as

$$Q_\theta^x\theta=b_\theta^x. \quad (14)$$

Similarly *y*-mapping is expressed as

$$Q_\theta^y\theta=b_\theta^y. \quad (15)$$

We are now ready to provide the linear program:

$$\min\begin{bmatrix}\mathbf{c}+\lambda\mathbf{d}\\ \mathbf{0}\end{bmatrix}^T\begin{bmatrix}\mathbf{e}\\ \theta\end{bmatrix} \quad (16)$$

subject to

$$\begin{bmatrix}P_e & \mathbf{0}\\ \mathbf{0} & Q_\theta^x\\ \mathbf{0} & Q_\theta^y\end{bmatrix}\begin{bmatrix}\mathbf{e}\\ \theta\end{bmatrix}=\begin{bmatrix}\mathbf{1}\\ b_\theta^x\\ b_\theta^y\end{bmatrix} \quad (17)$$

and

$$-\mathbf{e}\leq 0. \quad (18)$$

In this linear program formulation, **e** and θ are unknowns.

There are different ways to solve a LP problem and various solvers are available online. We chose the interior point solver of GLPK[1] for its efficiency.

3.2.2. Algorithm Overview—Optimization of our landmark matching approach is briefed as the following five steps:

[1] http://gnuwin32.sourceforge.net/packages/glpk.htm

1. Solving the LP problem specified by Eq. (16), Eq. (17), and Eq. (18) without considering the to-centroid deviations (i.e. set $\lambda = 0$ in Eq. (16)).
2. Updating the to-centroid deviation matrix using Eq. (5) with current estimation of the correspondence matrix.
3. Increasing λ by $\delta\lambda$ (e.g. 0.2).
4. Solving the LP problem specified by Eq. (16), Eq. (17), and Eq. (18) with considering the to-centroid deviation matrix.
5. Go to step 2 until the stopping criterion is satisfied.

Our algorithm can produce high accuracies due to at least four unique characteristics. *First*, step 1 can offer a very nice initialization without any manual initialization. *Second*, step 1 and step 4 can both result in an optimal solution with the LP techniques. *Third*, the correspondences and transformation model are estimated simultaneously in step 4. *Fourth*, gradually increasing λ in step 4 approaches a binary correspondence matrix with less ambiguity in the assignment. *Finally*, the transformation model can help to guide the estimation of correspondences and resist noise/outliers better.

We set the stopping criterion as repeating steps 2–5 for 3 times and from our experiments we found that an accurate estimation can be generated in most cases. The computational complexity of our algorithm in solving the linear program in Eq. (16) is $O(n^3L)$ (Wright, 2005) where n is the number of unknowns in Eq. (16) and L is the number of encoded input bits when solving the LP. The optimization process can be finished within 1.5 seconds to match about 180 landmarks using a Dell PC with 2.39 GHz Intel Core 2 CPU. As a postprocessing, each floating-landmark is finally assigned to the reference-landmark with the largest correspondence likelihood value.

4. Landmark Detection and Feature Extraction

Landmark detection and feature extraction are also two important procedures of our retinal image registration approach. For landmark detection, the vascular-landmark detection technique in (Can et al., 2002) is employed. Several steps are involved in the detection process: specification of seed points of vessel using 1D edge detection along a series of evenly spaced rows and columns, tracing out vascular structures with an exploratory procedure, detection of landmarks by locating intersections of traces where three or more traces meet.

To characterize the local photometric and geometric appearance of each landmark, we introduce a reinforced self-similarities (SS) descriptor. The SS descriptor was recently proposed in (Shechtman and Irani, 2007) to measure similarity between two visual entities of images or videos. SS has been shown to bear high differentiating ability. Compared with SS, most other widely known descriptors (e.g. SIFT (Brown and Lowe, 2007)) assume the existence of a common underlying visual property, e.g. intensities or filter responses between images to be matched. SS relaxes this assumption and tries to capture the geometric layouts of local intensity-pattern's repetitions in nearby image locations. In other words, SS describes the local internal layouts of the self-similarities instead of the intensity pattern itself. This is very important for a robust descriptor. For example, landmarks of retinal images from different modalities can have very different local intensity patterns but the geometric layout of pattern's repetitions in the vicinity of the landmark are very similar. As a result, SS can be used in both intra-modality and inter-modality retial image registrations.

Specifically, an SS descriptor for an arbitrary pixel q is obtained by normalizing the values of a correlation surface $S_q(x, y)$ (obtained in an Euclidean space with its origin at q) in the corresponding log-polar coordinate space followed by stretching the resulting values into [0 1]. The correlation surface is computed with the following equation

$$S_q(x,y)=\exp\left(\frac{\|\mathscr{P}(q)-\mathscr{P}(x,y)\|^2}{\max(\sigma_{\mathscr{N}}^2,\sigma_{\mathscr{P}}^2)}\right) \quad (19)$$

where $\mathscr{P}(q)$ and $\mathscr{P}(x, y)$ denote a small image patch (5 × 5) surrounding q and (x, y), respectively. The operator $\|\cdot\|^2$ represents the sum of square differences of the two patches in their CIE L∗a∗b∗ color space. $\sigma_{\mathscr{N}}^2$ means image noise variance (determined automatically with the technique in (Liu et al., 2006)) and $\sigma_{\mathscr{P}}^2$ implies the maximal variance of the difference of all small patches within a neighborhood (of radius 1) relative to the small patch centered at q. In Eq. (19), (x, y) takes values in an image region surrounding q with a radius 40.

The proposed reinforced SS descriptor is based on the original SS descriptor but tailored to be invariant to large image rotations. Although the original SS features are robust to local affine deformation and radially increasing non-rigid deformation (Shechtman and Irani, 2007), they become fragile when a large rotation happens between the matching images. It is because both the order of the entries in the SS descriptor vector and the partitioned bins (Shechtman and Irani, 2007) can change with rotation. The demand to make SS invariant to large rotations is particularly necessary for retinal images considering the large range of view points involved in the retinal imaging process (Chanwimaluang et al., 2006). This large rotation invariance can be simply accomplished by computing the Hu's seven moment invariants (Hu, 1962) of the outputs of the original SS descriptors. Therefore, the reinforced SS results in a feature vector with seven entries.

5. Results and Discussions

We evaluated our retinal image alignment algorithms quantitatively with both synthetic image transformations artificially produced with certain models and real transformations generated during the imaging process. We also show results for a set of retinal image analysis tasks performed with the image alignment solutions of our algorithms.

We ran our LP based method with the features in (Can et al., 2002) (denoted by "LP-Can"), our LP based method with the reinforced SS features introduced in this paper (denoted by "LP-SS") and our LP based method with the SIFT features (Brown and Lowe, 2007) (denoted by "LP-SIFT").We compared our algorithms with the hierarchical method (Can et al., 2002) (denoted by "Hierarchical"), the hybrid method (Chanwimaluang et al., 2006) (denoted by "Hybrid"), and the EM like technique (Chui and Rangarajan, 2003) using features in (Can et al., 2002) (denoted by “EM-Can”), our reinforced SS features (denoted by “EM-SS”) and the SIFT features (Brown and Lowe, 2007) (denoted by “EM-SIFT”). Our LP based algorithms can work on all the affine transformation, 12-parameter transformation model (Can et al., 2002) and the TPS based nonlinear transformation. In contrast, “Hierarchical” can work on the affine transformation and the 12-parameter transformation model. “Hybrid” can work only on the affine transformation model. The EM based techniques are able to handle the TPS based nonlinear transformation.

Our testing images were chosen from the CATT study (Martin et al., 2012) and CAPT study (CAPT Study Group et al., 2004). They include stereoscopic color fundus photographs and angiogram grayscale images of the macula (DRS Field 2) of oculus dexter (OD) and oculus sinister (OS) at the initial visit of 1030 participants age 50 years and older, enrolled may 1, 1999 through March 31, 2001. All patients signed a patient consent form approved by the CAPT Coordinating Center and local Institutional Review Board. Each patient bears at least 10 large (125μm in diameter) drusen in each eye. High quality color fundus photographs were acquired by establishing and maintaining high photographic standards. Zeiss 30^o or Topcon 35^o fundus photograph cameras with a 2.5× to 3× magnification and either Kodachrome or Ektachrome color slide films were used. Study photographs were digitized using a Nikon Coolscan slide scanner (Nikon LS 4000ED; Nikon, Melville, NY) at a resolution of 7.5 MB.

5.1. Evaluation

5.1.1. Registration of Image Pairs with Synthetic Transformations—In order to empirically specify $\delta\lambda$, which is the only parameter of our main algorithm (as described in Sec. 3.2.2), and quantitatively compare our method with state-of-the-art, we chose a digitized fundus retinal image (as shown in Fig. 3), warped it with 50 synthetic affine transformations. These synthetic affine warps were created by randomly choosing a rotation angle in [0 2π], translation in [−8 8] pixels for both directions, scale in [0.3 3.5], and shear in [−1.8 1.8] for both directions. Note that for each pair of images, the ground-truth position in the warped image of each pixel in the original image is known. We employed the mean square error (MSE) between the ground-truth position and the estimation over all pixels in the original image to measure the registration accuracy.

To empirically determine an optimal value for $\delta\lambda$ (a learning based method can also be used), we ran our “LP-SS” algorithm with different $\delta\lambda$ values and compared the corresponding errors. As shown in Fig. 4, the algorithm works best when $\delta\lambda = 0.2$, and with this value, we executed our LP based algorithms in other experiments.

From the error bars in Fig. 4, we can see that our LP based algorithms (with affine model) produced higher accuracies than “Hierarchical” (with affine model) and “Hybrid” (with affine model). We can also find that the reinforced SS features outperformed both the features in (Can et al., 2002) and the SIFT features (Brown and Lowe, 2007). Therefore, our optimization technique is more effective and the proposed SS features have better differentiating abilities. Between the features in (Can et al., 2002) and the SIFT features (Brown and Lowe, 2007), the latter performed better.

From the up-left graph of Fig. 4, we can also find that the sparsity constraint (e.g. the second part of Eq. (10)) can help to improve matching accuracy when the parameter $\delta\lambda$ is set properly. The sparsity constraint is actually neglected when $\delta\lambda = 0$ because our algorithm is always initialized with $\lambda = 0$. Matching errors generated when $\delta\lambda = 0$ are relatively larger than most of other testing $\delta\lambda$ values as shown in Fig. 4.

We also tested our algorithms on 50 nonlinear deformation fields simulated with the TPS model. We first manually specified 15 bifurcation/crossover points in the image and treated them as the control points of the TPS model. We then synthesized these fields by randomly setting the affine transformation with the same way as above and randomly displacing control points in a range [−4 4] pixel for both directions. The registration errors of “LP-Can”, “LP-SS”, “LP-SIFT”, “EM-Can”, “EM-SS”, “EM-SIFT” are shown in Fig. 4 (all algorithms were carried out with the TPS model). We can see that our LP based algorithms performed better than the EM based techniques no matter what features were used. Moreover, our reinforced SS features performed best and the SIFT features (Brown and Lowe, 2007) produced higher accuracies than the features in (Can et al., 2002).

5.1.2. Registration of Image Pairs with Real Transformations—We randomly chose 40 pairs of retinal images from the CATT study (Martin et al., 2012) and CAPT study (CAPT Study Group et al., 2004) to evaluate “LP-Can” (with the 12-parameter model), “LP-SS” (with the 12-parameter model), “LP-SIFT” (with the 12-parameter model), “Hierarchical” (with the 12-parameter model) and “Hybrid” (with affine because only affine can be accomplished by this algorithm). The centerline error described in (Can et al., 2002) was used as the error measure for validation. From the error values in Fig. 4, we found that our LP based algorithms performed better than “Hierarchical” and “Hybrid”. Moreover, the reinforced SS features outperformed other features.

5.2. Application

We ran our “LP-SS” algorithm with the 12-parameter model to carry on tasks of pairwise image mosaic (Cattin et al., 2006), retinal image fusion (Cao, 2008) and retinal montage (Becker et al., 1998) using a set of angiogram images and color fundus images from the CATT study (Martin et al., 2012) and CAPT study (CAPT Study Group et al., 2004). Through visual inspections, we found that our algorithm performed well regardless of the modality difference, image bias artifact and image changes caused by retinal diseases. We show some example results in Fig. 5, Fig. 6 and Fig. 7.

Our algorithm performs well on retinal images with different properties mainly due to three reasons. First, it enables a joint estimation of the landmark correspondences and the predefined transformation model. Second, it combines the benefits of the softassign strategy and the combinatorial optimization of linear programming. Finally, the reinforced SS descriptor is pretty robust.

6. Conclusion and Future Work

We proposed a novel landmark matching based retinal image alignment formulation by using the softassign strategy (Chui and Rangarajan, 2003) and enforcing the sparsity prior to the correspondence matrix. Our formulation bear several advantages. First, a joint estimation of the landmark correspondences and the predefined transformation model can be achieved. Joint estimation is important in obtaining more accurate and robust alignment solutions. It is because of the fact that the mutual benefits between searching landmark matches and confining the transformation field with a prior-known model can be maximized in a joint estimation. Second, our formulation enables linear programming (LP) (a combinatorial optimization technique) to be used to generate a better optimization. Third, our formulation employs the softassign strategy (Chui and Rangarajan, 2003) to eliminate the hardships in estimating a binary correspondence matrix and at the same time is capable of handling the potential matching ambiguities. Fourth, the proposed reinforced self-similarities (SS) descriptors are novel and were validated to outperform state-of-the-art features used in retinal image alignment. These descriptors capture the geometric layouts of local intensity-pattern's repetitions in nearby image locations. At the same time, they are invariant to global rotation and local affine deformations. Experimental results of this paper generated with both synthetic and real image transformations show that the LP based joint estimation and the proposed reinforced self-similarities (SS) descriptors help our algorithms outperform state-of-the-art techniques of retinal image alignment.

The proposed sparsity prior enforced to the correspondence matrix is inspired by recent works on the sparsity prior (Donoho, 2006; Candes and Romberg, 2007; Huang et al., 2009; Zhang et al., 2012; Waters et al., 2011; Candes and Tao, 2005). To carry it out, we penalize a to-centroid deviation measurement which is defined on the correspondence matrix in order not only to guarantee the correspondence matrix to be sparse but also to make the matched reference points to be as spatially concentrated as possible.

Our LP based algorithms can not only work on retinal images but also be extended to other landmark based 2D/3D image alignment/registration tasks. Its simultaneous estimation of the predefined transformation model and the correspondence matrix can help to search correspondences whin a more reasonable space which is much smaller than the space without constraint of the transformation model. At the same time, the softassign strategy and the sparsity constraint provides robustness to noise and outliers. We found that our algorithms can deteriorate especially when the predefined transformation model is not well chosen.

Our future work would include extension of our formulation to 3D image registration and its application to other images and other modalities (e.g. brain MR images and CT lung images). This extension can be simply achieved by extending the reinforced SS descriptor to 3D and including a 3D transformation model. Moreover, we are interested in research on automatic selection of an appropriate transformation model from a set of candidates.

Acknowledgments

This work was made possible through support from the National Institutes of Health (NIH) via grants NS045839, HHSN276201000492P and P30 EY001583.

References

Becker DE, Can A, Turner JN, Tanenbaum HL, Roysam B. Image processing algorithms for retinal montage synthesis, mapping, and real-time location determination. IEEE Transactions on Biomedical Engineering. 1998; 45:105–118. [PubMed: 9444845]

Belongie S, Malik J, Puzicha J. Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2001; 24:509–522.

Berg, AC.; Berg, TL.; Malik, J. Shape matching and object recognition using low distortion correspondence; IEEE Conference on Computer Vision and Pattern Recognition; 2005. p. 26-33.

Brown M, Lowe DG. Automatic panoramic image stitching using invariant features. International Journal of Computer Vision. 2007; 74:59–73.

Can A, Stewart CV, Roysam B, Tanenbaum HL. A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002; 24:347–364.

Candes E, Romberg J. Sparsity and incoherence in compressive sampling. Inverse problems. 2007; 23:969.

Candes EJ, Tao T. Decoding by linear programming. Information Theory, IEEE Transactions on. 2005; 51:4203–4215.

Cao, H. Ph.D. thesis. Louisiana State University; 2008. A Novel Automated Approach of Multi-Modality Retinal Image Registration and Fusion.

CAPT Study Group. The complications of age-related macular degeneration prevention trial (capt): Rationale, design and methodology. Clin Trials. 2004; 1:91–107. [PubMed: 16281465]

Cattin, PC.; Bay, H.; Van Gool, L.; Székely, G. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006. Springer; 2006. Retina mosaicing using local features; p. 185-192.

Chanwimaluang T, Fan G, Fransen SR. Hybrid retinal image registration. IEEE Transactions on Information Technology in Biomedicine. 2006; 10:130–142.

Chen S, Billings SA, Luo W. Orthogonal least squares methods and their application to non-linear system identification. International Journal of control. 1989; 50:1873–1896.

Chui H, Rangarajan A. A new point matching algorithm for non-rigid registration. Computer Vision and Image Understanding. 2003; 89:114–141.

Dhillonb B, Eikelboomf RH, Yogesana K, Constablea IJ. Retinal image analysis: concepts, applications and potential. Progress in retinal and eye research. 2006; 25:99–127. [PubMed: 16154379]

Donoho DL. For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution. Communications on pure and applied mathematics. 2006; 59:797–829.

Gholipour A, Kehtarnavaz N, Briggs R, Devous M, Gopinath K. Brain functional localization: a survey of image registration techniques. IEEE Transactions on Medical Imaging. 2007; 26:427–451. [PubMed: 17427731]

Hu MK. Visual pattern recognition by moment invariants. IRE Transactions on Information Theory. 1962:179–187.

Huang, J.; Zhang, T.; Metaxas, D. Learning with structured sparsity; Proceedings of the 26th Annual International Conference on Machine Learning, ACM; 2009. p. 417-424.

Jiang, H.; Yu, SX. Linear solution to scale and rotation invariant object matching; IEEE Conference on Computer Vision and Pattern Recognition; 2009.

Koh K, Kim SJ, Boyd SP. An interior-point method for large-scale l1-regularized logistic regression. Journal of Machine learning research. 2007; 8:1519–1555.

Kolar R, Kubecka L, Jan J, et al. Registration and fusion of the autofluorescent and infrared retinal images. International journal of biomedical imaging 2008. 2008a

Kolar R, Kubecka L, Jan J. Registration and fusion of the autofluorescent and infrared retinal images. International Journal of Biomedical Imaging 2008. 2008b:513478.

Lee S, Abràmoff MD, Reinhardt JM. Feature-based pairwise retinal image registration by radial distortion correction. Medical Imaging, International Society for Optics and Photonics. 2007:651220–651220.

Li, H.; Huang, J.; Zhang, S.; Huang, X. Optimal object matching via convexification and composition; Computer Vision (ICCV), 2011 IEEE International Conference on, IEEE; 2011. p. 33-40.

Li H, Huang X, He L. Object matching using a locally affine invariant and linear programming techniques. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2013; 35:411–424. [PubMed: 22529322]

Liu, C.; Freeman, WT.; Szeliski, R.; Kang, SB. Noise estimation from a single image; Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, IEEE; 2006. p. 901-908.

Martin D, Maguire M, Fine S, Ying G, Jaffe G, Grunwald J, Toth C, Redford M, Ferris F 3rd. Comparison of age-related macular degeneration treatments trials (catt) research group: Ranibizumab and bevacizumab for treatment of neovascular age-related macular degeneration: two-year results. Ophthalmology. 2012; 119:1388–1398. [PubMed: 22555112]

Matsopoulos GK, Mouravliansky NA, Delibasis KK, Nikita KS. Automatic retinal image registration scheme using global optimization techniques. IEEE Transactions on Medical Imaging. 1999; 3:47–60.

Modersitzki, J. Numetical Methods for Image Registration. New York, NY, USA: Oxford University Press; 2004.

Narasimha-Iyer H, Can A, Roysam B, Stewart V, Tanenbaum HL, Majerovics A, Singh H. Robust detection and classification of longitudinal changes in color retinal fundus images for monitoring diabetic retinopathy. IEEE Transactions on Biomedical Engineering. 2006; 53:1084–1098. [PubMed: 16761836]

Quigley HA, Addicks EM, Green WR, et al. Optic nerve damage in human glaucoma. iii. quantitative correlation of nerve fiber loss and visual field defect in glaucoma, ischemic neuropathy, papilledema, and toxic neuropathy. Archives of Ophthalmology. 1982; 100:135. [PubMed: 7055464]

Rohr K, Stiehl HS, Sprengel R, Buzug TM, Weese J, Kuhn MH. Landmark-based elastic registration using approximating thin-plate splines. IEEE Transactions on Medical Imaging. 2001; 20:526–534. [PubMed: 11437112]

Rueckert D, Sonoda LI, Hayes C, Hill DL, Leach MO, Hawkes DJ. Nonrigid registration using free-form deformations: application to breast mr images. Medical Imaging, IEEE Transactions on. 1999; 18:712–721.

Shechtman, E.; Irani, M. Matching local self-similarities across images and videos; IEEE Conference on Computer Vision and Pattern Recognition; 2007.

Shen D, Davatzikos C. Hammer: Hierarchical attribute matching mechanism for elastic registration. IEEE Transactions on Medical Imaging. 2002; 21:1421–1439. [PubMed: 12575879]

Stewart CV, ling Tsai C, Roysam B. The dual-bootstrap iterative closest point algorithm with application to retinal image registration. IEEE Transactions on Medical Imaging. 2003; 22:1379–1394. [PubMed: 14606672]

Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological). 1996:267–288.

Tobin KW, Chaum E, Govindasamy VP, Karnowski TP. Detection of anatomic structures in human retinal imagery. IEEE Transactions on Medical Imaging. 2007; 26:1729–1739. [PubMed: 18092741]

Treigys, P.; Dzemyda, G.; Barzdziukas, V. Computational Science–ICCS 2008. Springer; 2008. Automated positioning of overlapping eye fundus images; p. 770-779.

Waters AE, Sankaranarayanan AC, Baraniuk RG. Sparcs: Recovering low-rank and sparse matrices from compressive measurements. Neural Information Processing Systems (NIPS). 2011:1089–1097.

Winder R, Morrow P, McRitchie I, Bailie J, Hart P. Algorithms for digital image processing in diabetic retinopathy. Computerized medical imaging and graphics: the official journal of the Computerized Medical Imaging Society. 2009; 33:608. [PubMed: 19616920]

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Wood AJ, Fine SL, Berger JW, Maguire MG, Ho AC. Age-related macular degeneration. New England Journal of Medicine. 2000; 342:483–492. [PubMed: 10675430]

Wright M. The interior-point revolution in optimization: history, recent developments, and lasting consequences. Bulletin of the American mathematical society. 2005; 42:39–56.

Zana F, Klein JC. A multimodal registration algorithm of eye fundus images using vessels by hough transform. IEEE Transactions on Medical Imaging. 1999; 18:419–428. [PubMed: 10416803]

Zhan, Y.; Feldman, M.; Tomaszeweski, J.; Davatzikos, C.; Shen, D. Medical Image Computing and Computer-Assisted Intervention–MICCAI, 2006. Springer; 2006. Registering histological and mr images of prostate for image-based cancer detection; p. 620-628.

Zhang S, Zhan Y, Dewan M, Huang J, Metaxas DN, Zhou XS. Towards robust and effective shape modeling: Sparse shape composition. Medical image analysis. 2012; 16:265–277. [PubMed: 21963296]

Zheng Y, Doermann D. Robust point matching for nonrigid shapes by preserving local neighborhood structures. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2006; 28:643–649. [PubMed: 16566512]

Zheng, Y.; Hunter, AA., III; Wu, J.; Wang, H.; Gao, J.; Maguire, MG.; Gee, JC. Information Processing in Medical Imaging. Springer: 2011. Landmark matching based automatic retinal image registration with linear programming and self-similarities; p. 674-685.

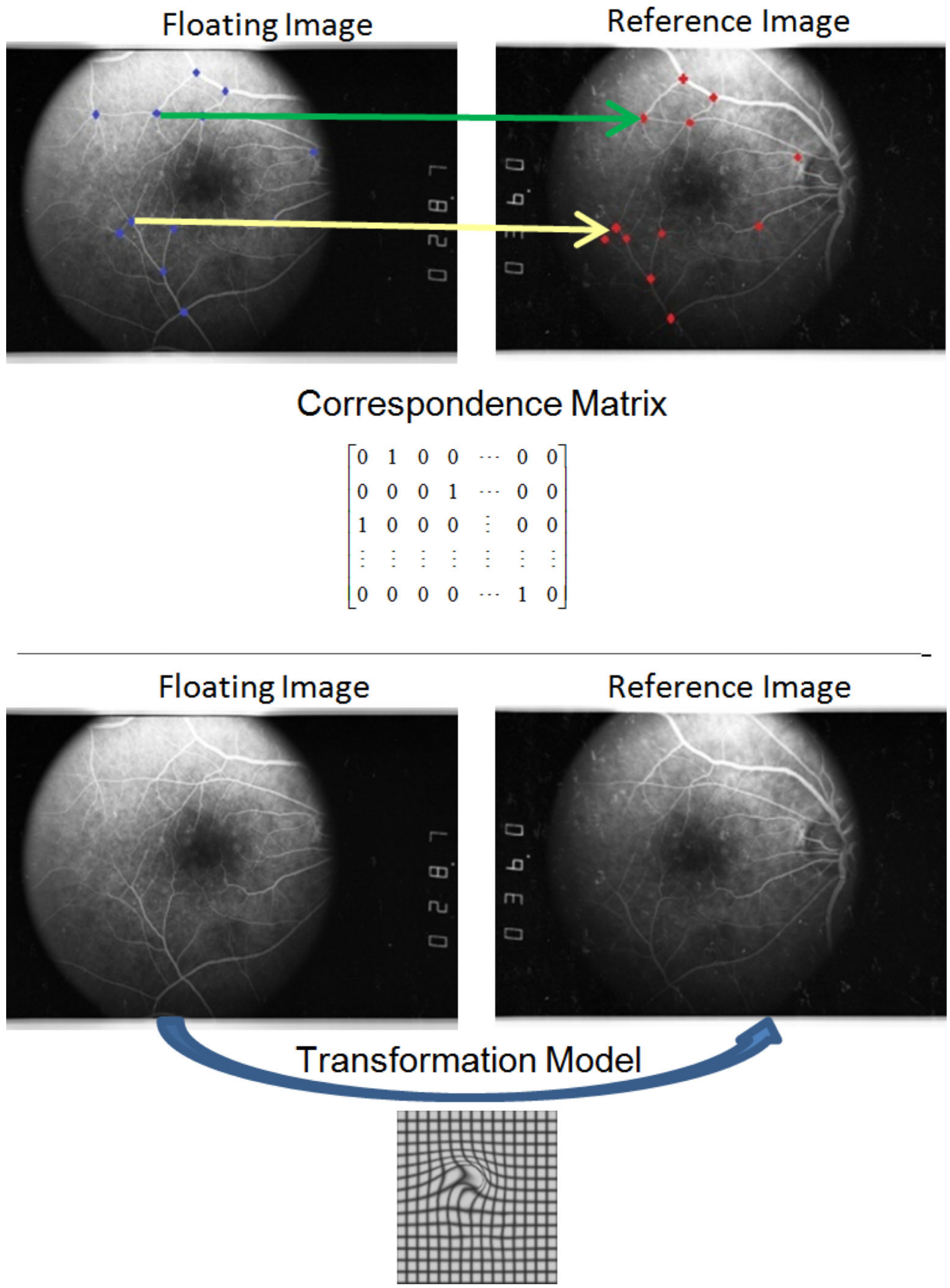

Figure 1.

Estimation of the correspondences (up) between the floating and reference landmark points and estimation of the transformation model (down) between the floating and reference images are both important in obtaining an accurate alignment of retinal images with a landmark matching algorithm. Most existing methods refine these two estimations independently and in an iterative fashion while the proposed approach offers a joint estimation.

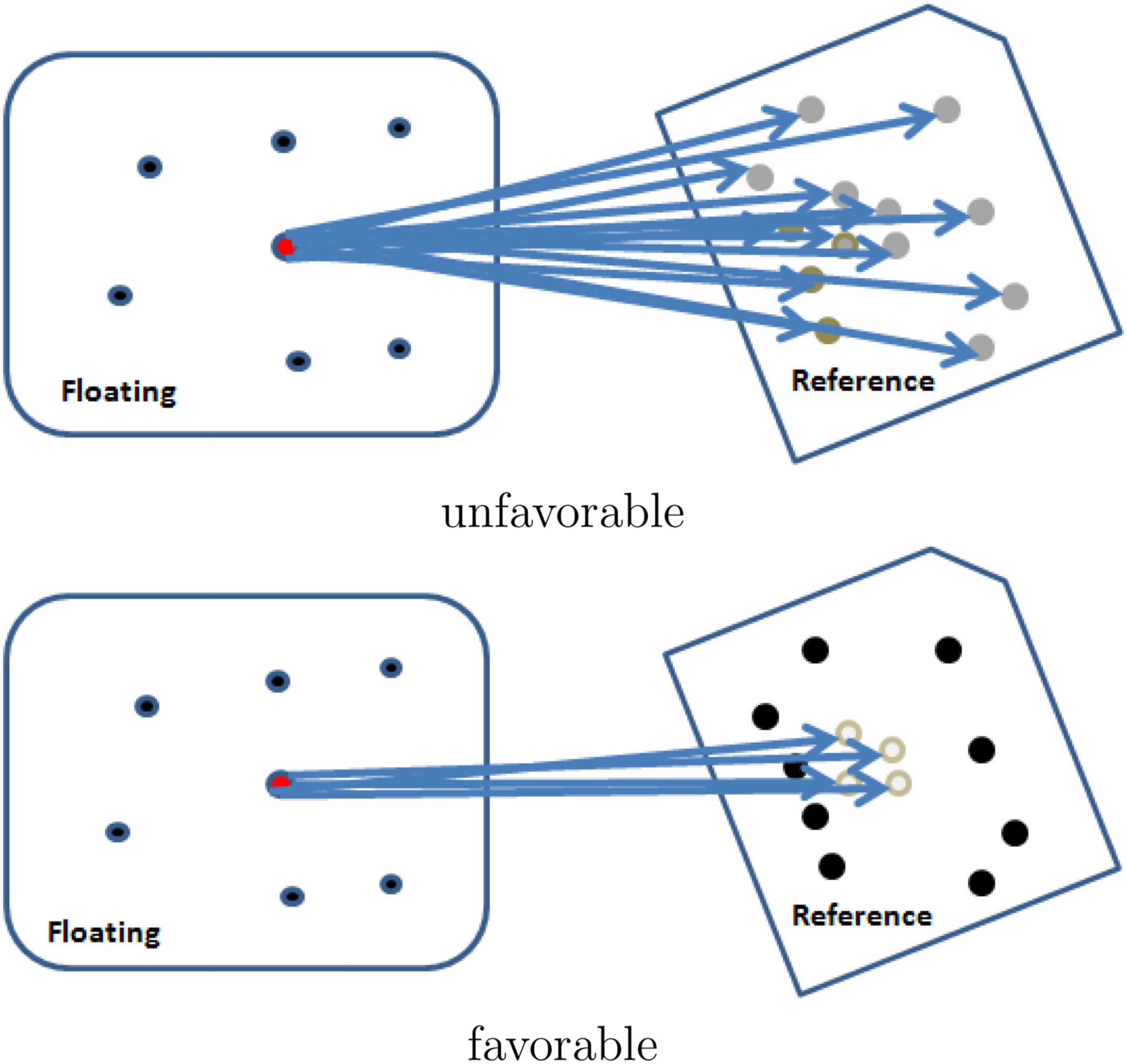

Figure 2.
Our landmark matching formulation relies on two strategies: softassignment and penalization on the to-centroid deviation. The former allows each floating point to match all reference points while the latter discourages the matched reference points to be scattered. In the above two examples, the lower is favorable because the reference points with higher matching score values (a higher gray scale of the reference points in the picture means a better matching case) are more concentrated.

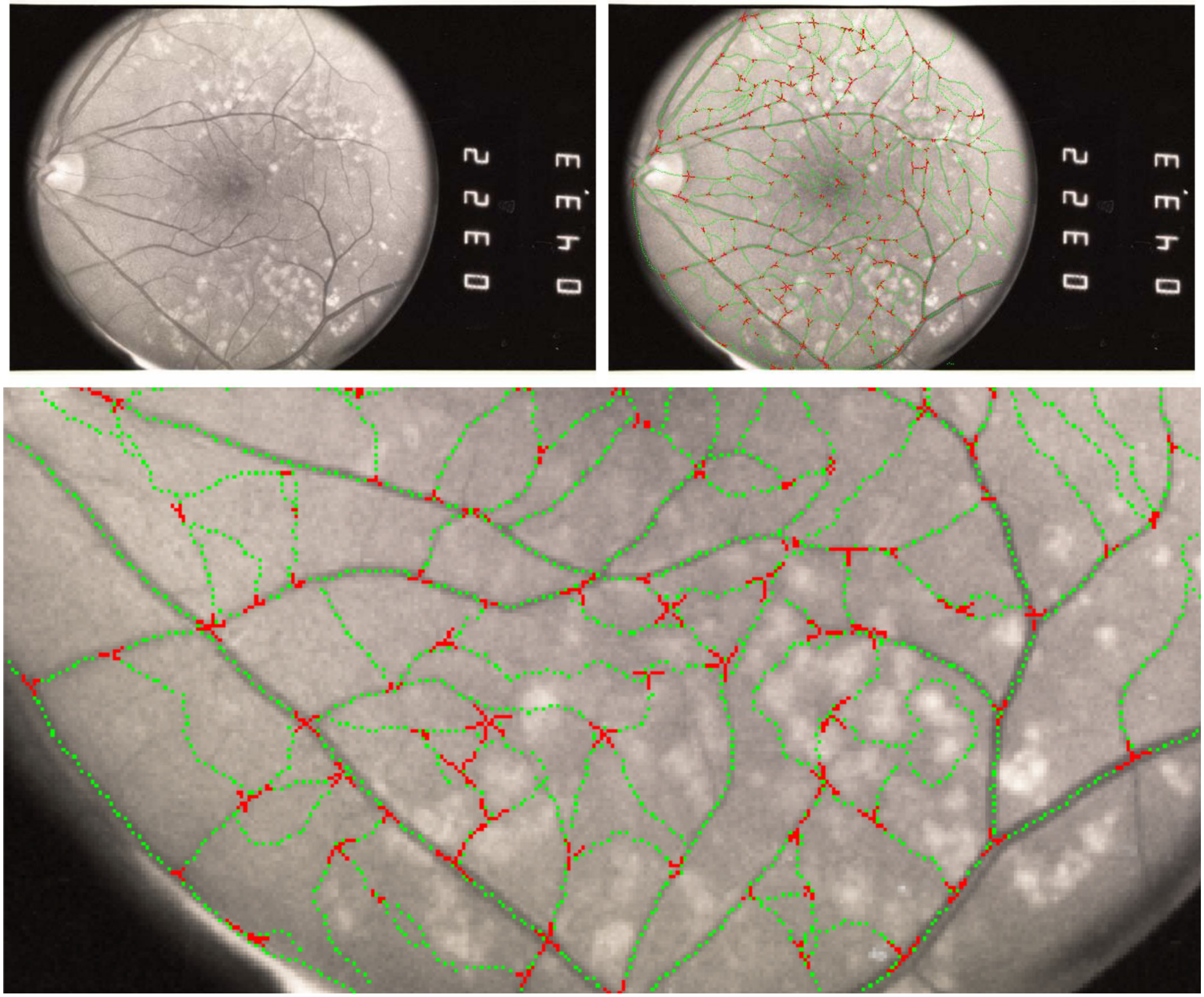

Figure 3.

Up-left: a digitized fundus retinal image. Up-right: the bifurcation/crossover landmark points (red) and vessel's centerline pixels (green) detected by the technique in (Can et al., 2002). Down: an enlarged view of the lower retina portion in up-right image.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

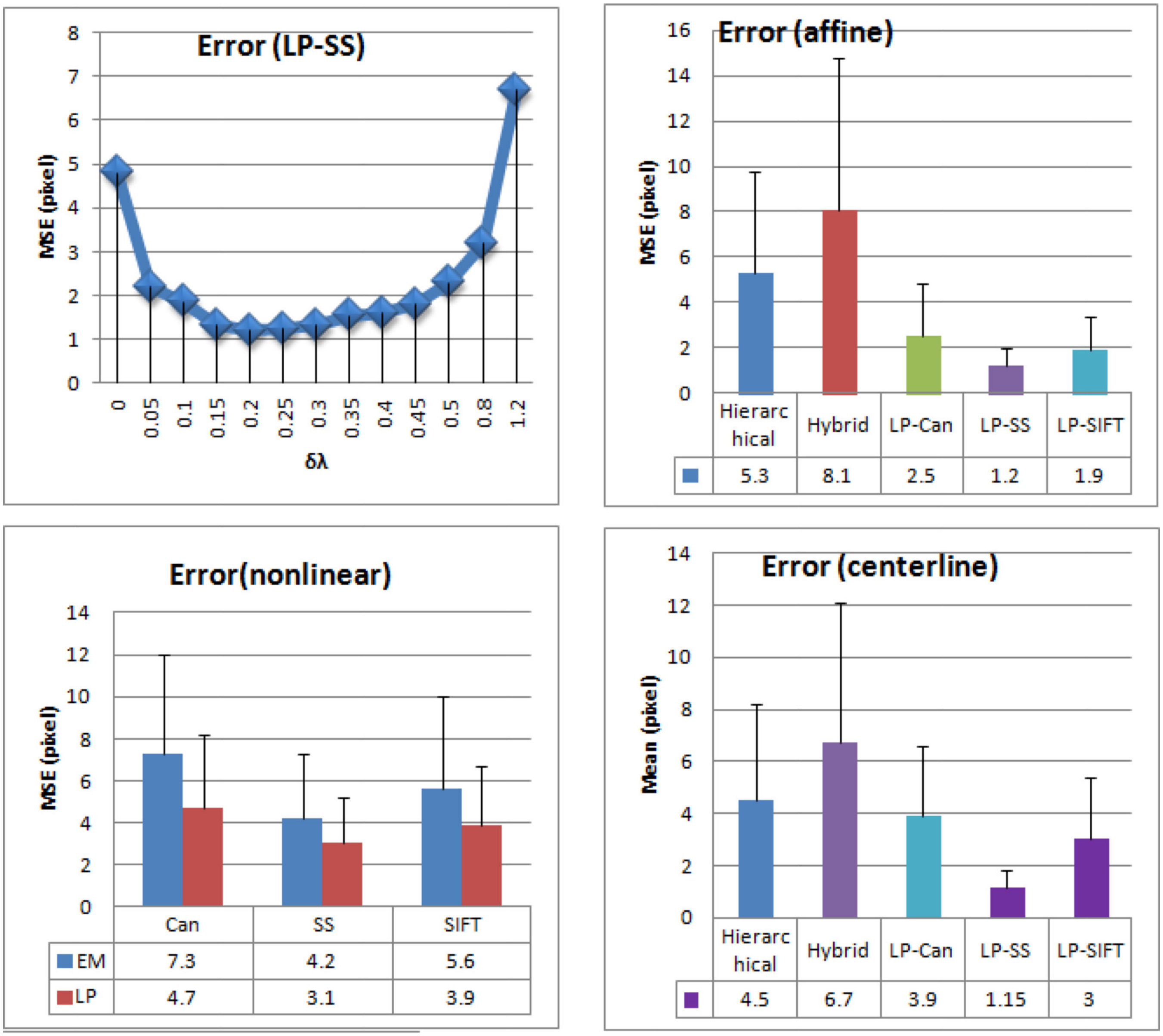

Figure 4.

Up-left: registration errors of our LP-SS algorithm with different δλ values. Up-right: registration error bars over 50 synthetic affine transformations. Down-left: registration error bars over 50 synthetic nonlinear deformations. Down-right: registration error bars over 40 real transformations.

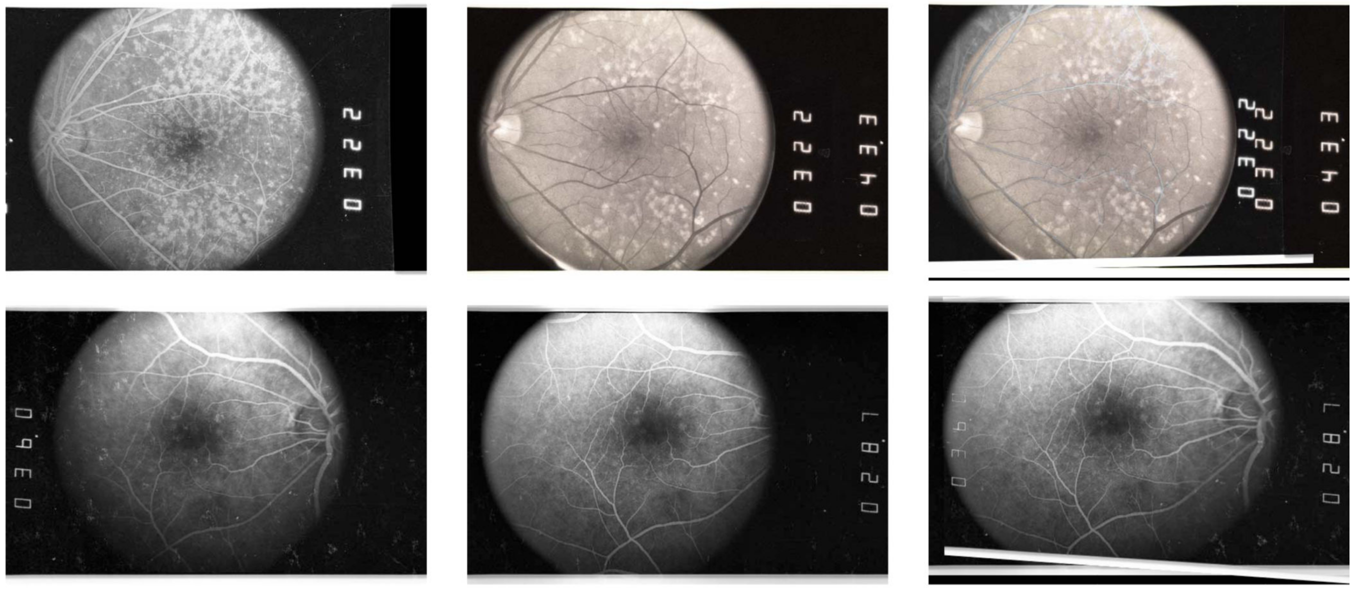

Image 1 Image 2 Mosaiced Image

Figure 5.

Pairwise image mosaics accomplished by our approach on two exampling pairs of angiogram images.

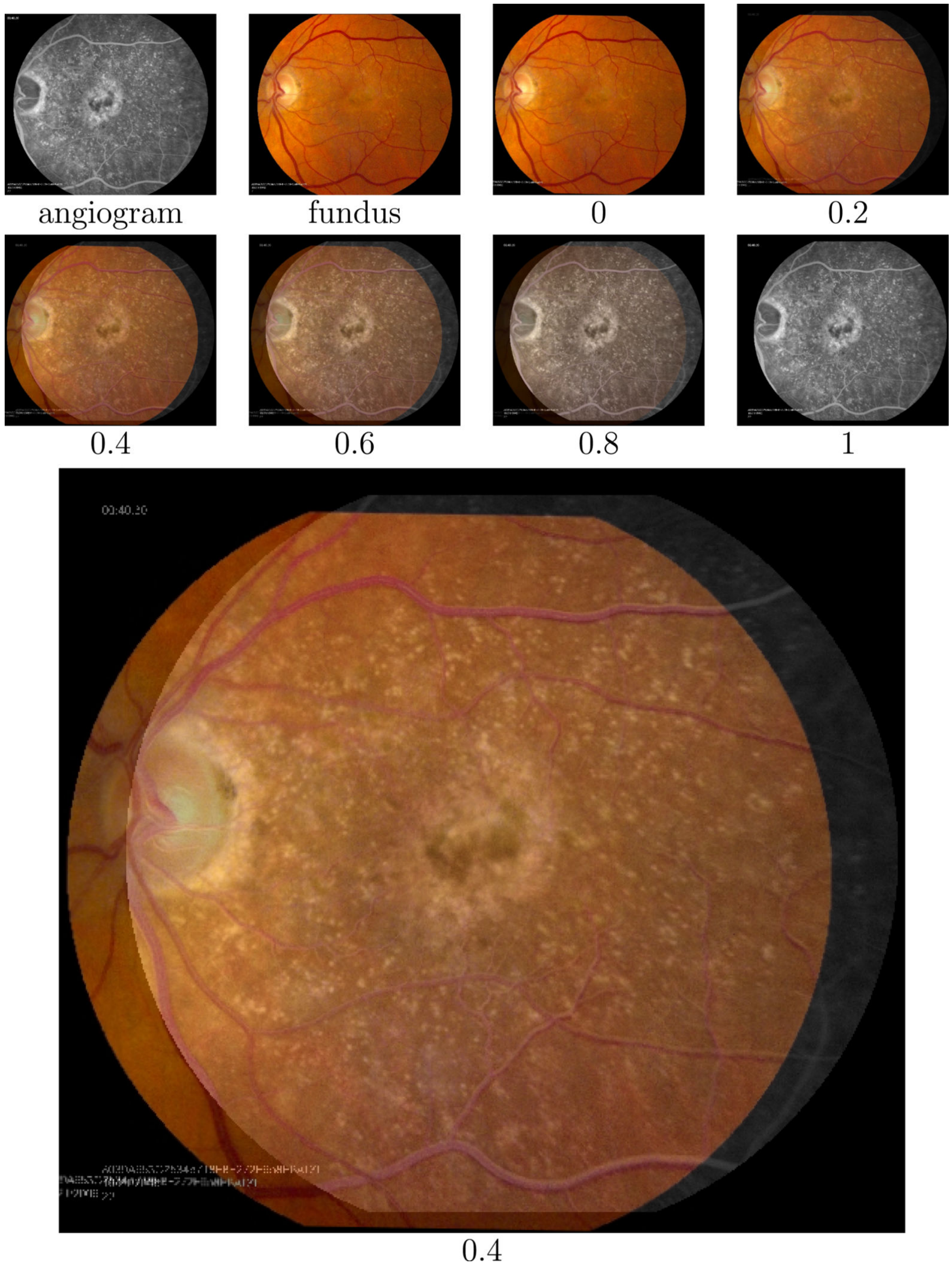

Figure 6.

Retinal image fusion accomplished by our algorithm with one angiogram grayscale image and one fundus color image. We show the fusion results obtained by linearly combing the angiogram and fundus images with a set of weight values (as shown by the numbers under picture).

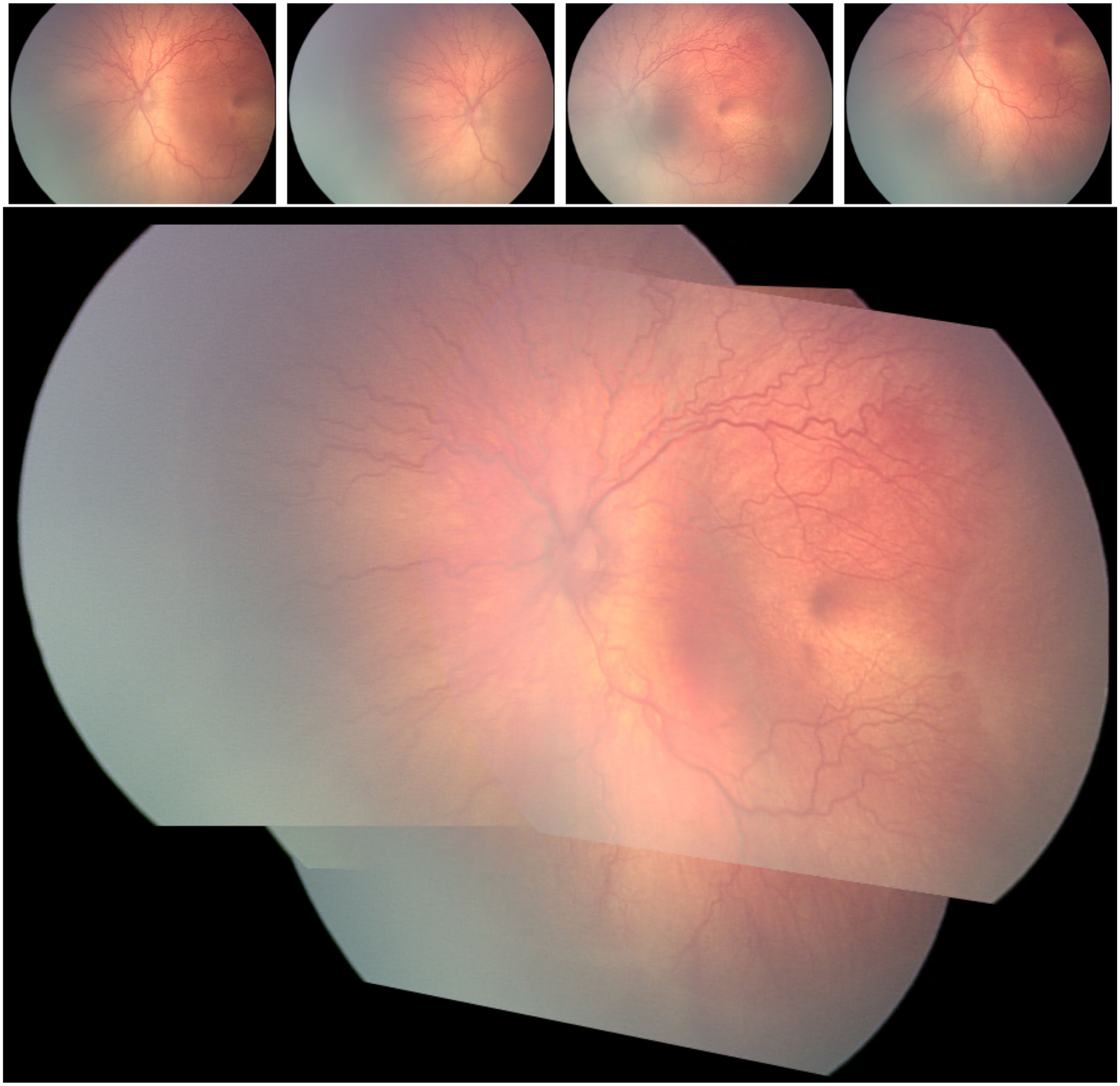

Figure 7.
Retinal montage accomplished by our algorithm with four fundus color images (above) captured from one retina and at different angles.