submitted for review to *IEEE Trans. on Pattern Analysis and Machine Intelligence*

A Feature-Based Technique for Joint, Linear Estimation of High-Order Image-to-Mosaic Transformations: Mosaicing the Curved Human Retina

Ali Can[1*] Charles V. Stewart[2] Badrinath Roysam[1] Howard L. Tanenbaum[3]

[1]Dept. of Electrical, Computer, and Systems Engineering
Rensselaer Polytechnic Institute
Troy, New York 12180–3590
cana,roysab@rpi.edu

[2]Dept. of Computer Science
Rensselaer Polytechnic Institute
Troy, New York 12180–3590
stewart@cs.rpi.edu

[3]The Center for Sight
349 Northern Blvd.
Albany, NY 12204
how1@albany.net

August 9, 2001

Abstract

An algorithm for constructing image mosaics from multiple, uncalibrated, weak-perspective views of the human retina is presented and analyzed. It builds on a previously described algorithm for registering pairs of retinal images using a non-invertible, 12-parameter, quadratic image transformation model and a hierarchical, robust estimation technique. The major innovation presented here is a linear, feature-based, non-iterative method for jointly estimating consistent transformations of all images onto the mosaic "anchor image". Constraints for this estimation are derived from pairwise registration both directly with the anchor image and indirectly between pairs of non-anchor images. An incremental, graph-based technique constructs the set of registered image pairs used in the joint solution. The joint estimation technique allows images that do not overlap the anchor frame to be successfully mosaiced, a particularly valuable capability for mosaicing images of the retinal periphery. Experimental analysis of the algorithm on data sets from 16 eyes shows the average overall median transformation error in final mosaic construction to be 0.76 pixels. Overall, the technique is simpler, more accurate, and offers broader coverage than currently published methods.

*Dr. Can is now an Assistant Professor at Inonu University, Turkey

Index terms: Robust estimation, image mosaic, image montage, transformation estimation, retinal imaging, joint estimation.

1 Introduction

Building a mosaic image from a sequence of partial views is a powerful means of obtaining a broader view of a scene than is available with a single view. Research on automated mosaic construction has been underway for at least 25 years (see early work in [17, 18, 20, 29]). The range of applications includes panoramic image formation [22, 21], virtual reality [8, 28], image compression [15, 14], change detection, superresolution [7], tracking [10], navigation [11], indexing and key frame identification [15], document compositing [31], and cartography [30]. One application domain in which mosaics are particularly valuable is in the diagnosis and treatment of diseases of the retina [1, 3, 9, 16]. A seamless mosaic formed from multiple fundus camera images aids in diagnosis, provides a means for monitoring the progression of diseases, and may be used as a *spatial map* during surgical treatment [1, 3].

Several issues must be addressed in designing a mosaic construction technique. First, the coordinate system in which to build the mosaic must be established. Second, an appropriate mathematical model of the image-to-mosaic transformation must be developed. Most work in mosaic construction uses low-order, invertible transformations derived by assuming that the scene is roughly planar [2, 15] or that the images are taken by a camera rotated about its optical center [28]. Third and most important, the parameters of this transformation must be estimated for each image that is to be mosaiced. Finally, the intensity values of the transformed images must be combined to produce the final mosaic.

In a companion paper [6], we have addressed the problem of registering a pair of retinal images, effectively solving the two-image mosaicing problem. Highlights of the algorithm include derivation of a new 12-parameter image-to-image transformation model that accounts for the curvature of the retina and a hierarchical, robust, feature-based parameter estimation technique. The registration accuracy of this algorithm on 1024×1024 images is 0.83 pixels, greatly improving upon prior efforts (see discussion in [6]). A mosaic formed by registering a pair of images is shown in Figure 1. The current paper addresses the problem of forming a mosaic from an arbitrary number of retinal images.

At first glance, it appears that the ability to register a pair of images is sufficient for the problem of forming a mosaic of the entire retina from multiple partial views. Hypothetically, if one image can be established as the "anchor image" on which to base the mosaic, then the transformation of each remaining image onto this anchor may be estimated using pairwise registration, and the transformed images may

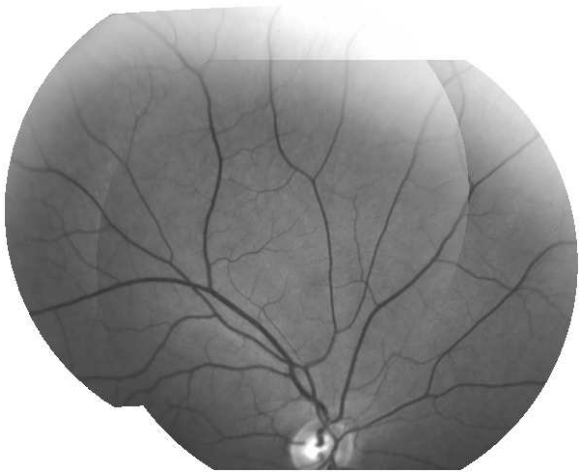


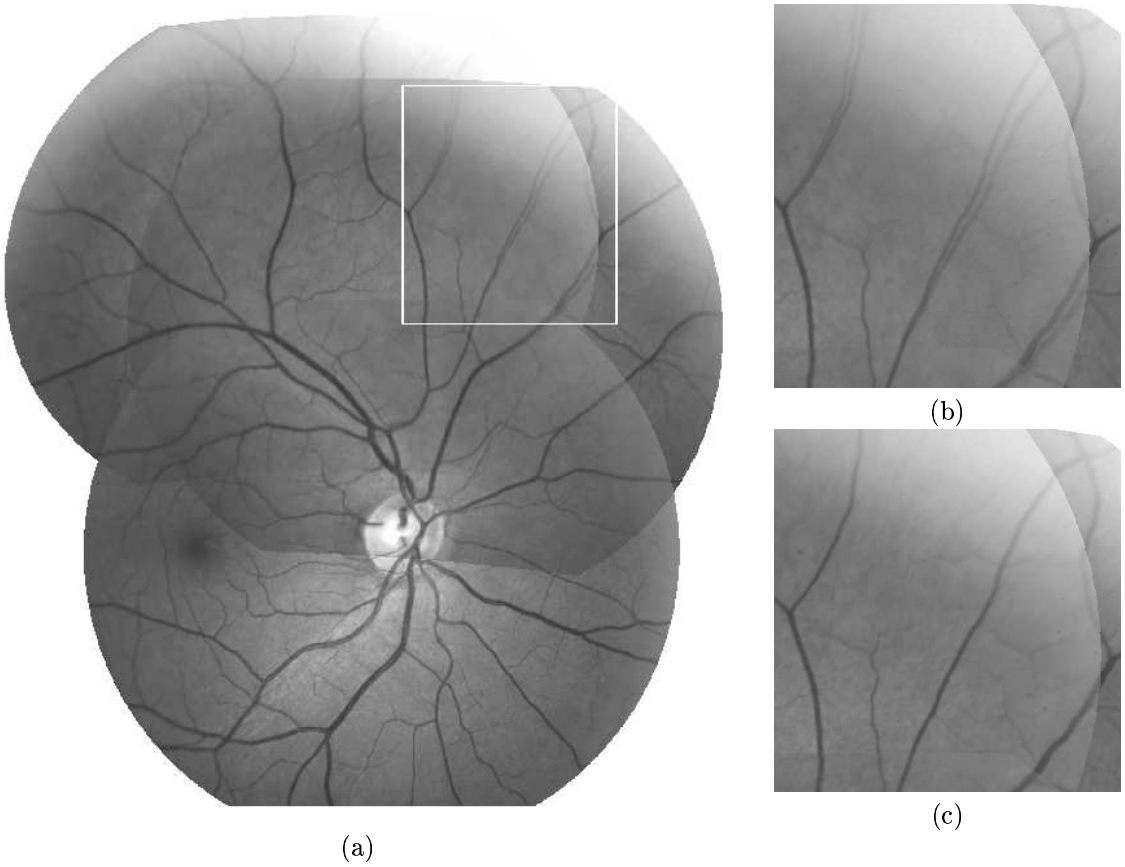
Figure 1: A mosaic of two retinal fundus images combined using the pairwise registration algorithm [6]. The blood vessels are almost perfectly aligned. The boundary between the images is visible because of image illumination differences.

be combined. Unfortunately, as observed in [24, 27] for planar, spherical, and panoramic mosaics, this approach does not ensure proper multi-image alignment. In the particular case of retinal image mosaicing, two problems arise:

Non-overlapping images: Some images may not overlap the anchor image at all, making direct computation of the transformation impossible. Other images may have insufficient overlap to compute a stable transformation. The straightforward solution is to compose transformations using an "intermediate" image [1]. This is problematic, however, both mathematically and practically: our transformation model is not closed under composition [6], and repeated application of transformations magnifies errors [24].

Mutually inconsistent transformations: The other problem arises even when image-to-anchor transformations can be estimated. Each image may individually register accurately with the anchor image and the non-anchor images may even register accurately with each other, but this does not ensure that the transformations onto the anchor image are mutually consistent! Outside the region where the images overlap the anchor image, there are no direct constraints to ensure that the transformations closely align the blood vessel structure (see Figure 2).

Our goal in this paper is to develop an algorithm to construct extremely accurate, seamless mosaics of $N > 2$ retinal images. We introduce a technique that uses feature correspondence sets resulting from pairwise registration [6] of non-anchor image pairs to jointly estimate the transformations of all images onto the anchor image. To form the feature correspondence sets without having to apply the pairwise



(a) (b) (c)

Figure 2: Illustrating mutually inconsistent transformations produced by pairwise registration, and the impact of the joint estimation technique. An example mosaic formed by independently computing the transformation of each of the two images in Figure 1 onto a third (anchor) image is shown in (a). The resulting mosaic is not seamless, especially in the outlined region, which does not overlap the anchor image (see (b)). The joint estimation technique produces near perfect alignment in the outlined region (c) and throughout the mosaic.

registration algorithm to all image pairs prior to the joint estimation, we describe a graph-based technique to incrementally determine which pairs might overlap sufficiently, and apply the algorithm only to these pairs. Experimental analysis of the overall mosaic construction algorithm on 16 retinal image data sets will demonstrate its effectiveness.

The resulting algorithm is similar in many respects to the important recent work of Sawhney, Hsu and Kumar [24] and of Shum and Szeliski [27, 26] on globally consistent mosaics, yet it has important differences as well. Most notably, our algorithm is feature-based throughout; it works on a 12-parameter, non-invertible transformation; the final estimation of image-to-anchor transformations is linear; and the incremental, graph-based technique efficiently determines which images to match pairwise without relying on the ordering imposed by a video sequence. Finally, the algorithm is developed for the important application of constructing retinal image mosaics, going beyond prior work in the area by being fully

automatic, by producing subpixel alignment accuracy in the mosaic, and by allowing broader coverage of the retina.

2 Pairwise Registration

This section summarizes our pairwise registration algorithm which forms the background for the current work [6, 5].

2.1 Transformation Model

In deriving the image-to-image transformation model, the retina is modeled as a quadratic surface, rigid transformations between views are assumed, and a weak-perspective camera model is used [19]. Let I_m be one image frame, let I_n be a second image frame, and let $\mathbf{p} = (x, y)^T$ be a pixel location in I_m. Define

$$\mathbf{X}(\mathbf{p}) = (x^2, xy, y^2, x, y, 1)^T. \tag{1}$$

Combining camera, surface, and motion models, we have derived the following equation for the transformed pixel location $\mathbf{p}'$ in frame I_n:

$$\mathbf{p}' = \mathbf{\Theta}_{m,n}\, \mathbf{X}(\mathbf{p}), \tag{2}$$

where $\mathbf{\Theta}_{m,n}$ is a 2×6 parameter matrix. This model generalizes earlier affine, planar or pure rotation motion models used in other mosaic construction work [15, 28, 24]. It is the second-order Taylor expansion of the general interframe mapping function, it is not closed under composition, and it is not invertible. Despite the approximations used and the non-invertibility, the model is accurate to less than a pixel on 1024×1024 retinal images [6].

2.2 Robust, Hierarchical Estimation of $\mathbf{\Theta}_{m,n}$

Given two images, I_m and I_n, the 12 parameters of $\mathbf{\Theta}_{m,n}$ can be estimated using a hierarchical, feature-based technique that simultaneously estimates the transformation parameters and the feature correspondences. For the joint estimation, the resulting feature correspondences are the most important results. The features are image locations, $\mathbf{p}_m$ in I_m and $\mathbf{p}_n$ in I_n, of vascular landmarks — branching and cross-over points of the retinal vasculature — detected using an exploratory vascular tracing and intersection detection algorithm [4, 25].

The hierarchical estimation technique has three levels. It starts by forming all possible correspondences between landmarks, and then estimates just a translation vector using a similarity-weighted histogram. Correspondences roughly consistent with the peak of the histogram are saved. At the next level, an affine transformation between images is estimated based on the reduced correspondence set using least-median of squares (LMS) [23]. The resulting transformation is used to initialize an M-estimator [12] to estimate the quadratic transformation parameters, $\mathbf{\Theta}_{m,n}$. When there is at most one correspondence $(\mathbf{p}_{m,i}, \mathbf{p}_{n,j})$ for each landmark this estimate is

$$\hat{\mathbf{\Theta}} = \underset{\mathbf{\Theta}}{\operatorname{argmin}} \sum_{(\mathbf{p}_{m,i}, \mathbf{p}_{n,j})} \rho(\|\mathbf{p}_{n,j} - \mathbf{\Theta}_{m,n}\, \mathbf{X}(\mathbf{p}_{m,i})\|/\hat{\sigma}). \tag{3}$$

Here, ρ is a "robust loss function" that grows subquadratically and $\hat{\sigma}$ is a robust scale estimate. Minimization uses an iteratively reweighted least squares technique [13], with weight function $w(u) = \rho'(u)/u$. This minimization is augmented as described in [6, 5] to account for non-unique correspondences.

Following hierarchical estimation, two more steps are applied before the final transformation estimate is determined. First, for each surviving, unique correspondence, $(\mathbf{p}_{m,i}, \mathbf{p}_{n,j})$, the position of $\mathbf{p}_{n,j}$ in I_n is refined to subpixel accuracy using normalized sum-of-squared-differences (SSD) matching. Second, unmatched features that fall within the region of overlap in either image are matched against the other image using SSD, producing additional correspondences. The final M-estimate of the quadratic transformation is computed using the refined feature positions and added correspondences. These positions and correspondences are preserved for the joint estimation. A sample registration result is shown in Figure 1.

After computing the transformation estimate, an acceptance/rejection decision is made. A subset of the blood vessel centerline points in I_m, detected during recursive tracing, is transformed onto I_n using the estimated transformation, and for each one that falls within image I_n, the distance to the closest centerline in I_n is computed. The median of these distances is taken as the matching error, referred to as the centerline error measure (CEM). Experimental evaluation indicates that a 1.5 pixel error threshold is sufficient to prevent false acceptance of incorrect registration results [6].

3 Joint Estimation of the Transformations

Given an anchor image, I_0, on which to build the mosaic, the main problem is estimating the quadratic transformations (2) from the remaining images, $I_1, \ldots, I_N$, onto the coordinate system of I_0. In other words, the goal is to estimate image-to-anchor transformation parameter matrices $\mathbf{\Theta}_{1,0}, \ldots, \mathbf{\Theta}_{N,0}$. The

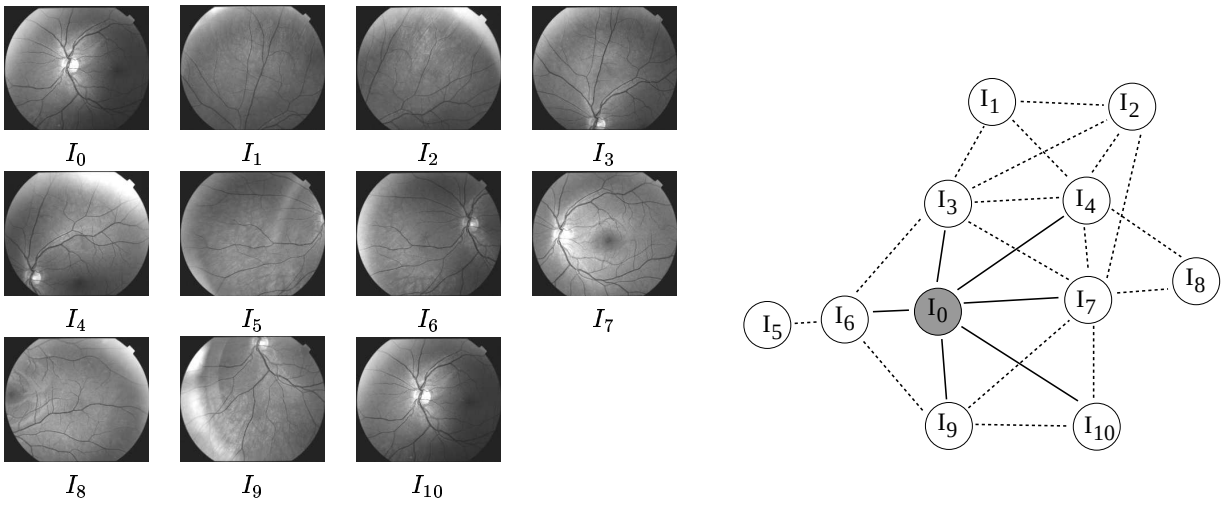


Figure 3: Pairwise image registrations represented as a graph. The solid edges in the graph show registration pairs involving the anchor image, I_0. The corresponding image features for each pair form a set of "direct constraints." The broken edges show registration pairs between non-anchor images. Corresponding image landmarks for each pair form a set of "indirect constraints."

simple method would be to register each image I_i individually with I_0 using the technique described in the previous section, but we have shown (Figure 2) that this fails to produce accurate alignment outside the original anchor image. To ensure accurate alignment throughout the mosaic, we need to use constraints from pairwise registration of non-anchor images in estimating the transformations (see Figure 3). The issues then become which non-anchor pairs to register and how to use the matching information obtained to constrain the image-to-anchor transformations. This section addresses the second issue, creating what we refer to as the "joint solution", and the next section addresses the first. The order of presentation is reversed because the joint solution is used incrementally in establishing the pairwise registrations.

3.1 Notation

Let the image index set be $\mathcal{I} = \{0, 1, \ldots, N\}$, with I_0 the desired anchor image. Let $\mathcal{P}$ be the set of pairwise registrations: $\{m, n\} \in \mathcal{P}$ if a pairwise registration has been established between images I_m and I_n. No ordering is assumed in these pairs since both the mappings, I_m onto I_n, and I_n onto I_m, may be estimated from corresponding landmark locations, despite the fact that the transformation model (2) itself is not invertible.[1] Finally, let $\mathcal{P}_D = \{\{m, 0\} \in \mathcal{P}\}$ denote the set of registration pairs involving the anchor image and let $\mathcal{P}_I = \{\{m, n\} \in \mathcal{P} \mid m, n > 0\}$ denote the other registration pairs. The former set contains what we term "direct constraints," while the latter contains what we call the "indirect constraints".

[1] This follows from the accuracy of the transformation model itself.

We turn now to the landmarks and the matches. For each image I_m, let $\mathbf{p}_{m,i}$, $1 \leq i \leq L_m$ be the landmark location vectors detected during the recursive tracing procedure. For each $(m,n) \in \mathcal{P}$, let $\mathcal{C}(m,n)$ be the correspondence set resulting from pairwise registration. Specifically, $(i,j) \in \mathcal{C}(m,n)$ if the feature at location $\mathbf{p}_{m,i}$ in image I_m corresponds to the feature at location $\mathbf{p}_{n,j}$ in image I_n. This notation must be extended to describe the changes in landmark feature positions and increases in the number of features occurring just before the end of pairwise registration. Each time a new feature is added, say to I_m, it is appended to the end of the list, and L_m is incremented. (These features are not currently used when I_m is subsequently registered with a different image.) For existing features, let $\mathbf{p}^n_{m,i}$ and $\mathbf{p}^m_{n,j}$ denote refined landmark positions as a result of registering I_m to I_n. In general only one of these positions will be altered from the original. Finally, let $w_{m,n;i,j}$ be the final robust weight for the matching pair, scaled by $1/\sigma^2_{m,n}$, where $\sigma_{m,n}$ is the robust scale estimate determined while matching I_m to I_n. This ensures that less reliable matches receive lower weights individually, and overall.

3.2 Linear Joint Solution

The intuition behind the joint solution is that any corresponding pair of landmark feature locations should be mapped onto the same mosaic location. One method for doing this starts by establishing equivalence classes between landmark features in the different images, effectively identifying all features in the images that correspond to the same retinal feature location. Constraints may then be built into the estimation equations to ensure that these map to the same location on the mosaic. To do this, the actual location on the mosaic must be estimated for each equivalence class at the same time that the image-to-anchor transformations are being estimated. While this will work, it leads to a coupled, nonlinear estimation problem (see discussion in [27]). Instead, we choose a second and much simpler method.

The idea is to eliminate any attempt to establish equivalence classes between features beyond pairs of correspondences. If $\mathbf{p}^n_{m,i}$ and $\mathbf{p}^m_{n,j}$ are corresponding (refined) landmark locations from pairwise matching of non-anchor images I_m and I_n, then the estimated image-to-anchor transformation parameters $\hat{\Theta}_{m,0}$ and $\hat{\Theta}_{n,0}$ must map $\mathbf{p}^n_{m,i}$ and $\mathbf{p}^m_{n,j}$ to (approximately) the same anchor frame location. We do not try to directly enforce consistency with other correspondences for $\mathbf{p}_{m,i}$ or other correspondences for $\mathbf{p}_{n,j}$. As a result, the final location for $\mathbf{p}_{m,i}$ and $\mathbf{p}_{n,j}$ need not be estimated explicitly.

This discussion leads to the following objective function for joint estimation:

$$\mathcal{E}(\mathbf{\Theta}_{1,0},\ldots,\mathbf{\Theta}_{N,0}) = \sum_{\{m|\{m,0\}\in\mathcal{P}_D\}} \sum_{\{(i,j)\in\mathcal{C}(m,0)\}} w_{m,0;i,j}\|\mathbf{\Theta}_{m,0}\mathbf{X}(\mathbf{p}^0_{m,i}) - \mathbf{p}^m_{0,j}\|^2 + \sum_{\{\{m,n\}\in\mathcal{P}_I\}} \sum_{\{(i,j)\in\mathcal{C}(m,n)\}} w_{m,n;i,j}\|\mathbf{\Theta}_{m,0}\mathbf{X}(\mathbf{p}^n_{m,i}) - \mathbf{\Theta}_{n,0}\mathbf{X}(\mathbf{p}^m_{n,j})\|^2. \quad (4)$$

The first summation includes the individual image-to-anchor constraint terms — the direct constraint sets. The second summation represents the coupling from pairwise constraints — the indirect constraint sets (Figure 3). Images I_m that were not directly registered to the anchor image only contribute indirect constraint sets. If there are no direct constraint sets, then a trivial solution (all 0's in each $\mathbf{\Theta}_{m,0}$) minimizes $\mathcal{E}$.

3.3 Minimization

Minimizing $\mathcal{E}$ to estimate $\mathbf{\Theta}_{1,0},\ldots,\mathbf{\Theta}_{N,0}$ is a linear problem. This can be seen by calculating the derivative of $\mathcal{E}$ with respect to each term in each $\mathbf{\Theta}_{i,0}$. Estimation of the terms in the 1^{st} row of the parameter matrices $\mathbf{\Theta}_{i,0}$ can be separated from estimation of the terms in the 2^{nd} row. Each depends on the inverse of the same $6N \times 6N$ matrix.

This can be shown clearly and in more detail by rewriting (4) in matrix notation. Let $\mathbf{\Phi}_1$ be a $6N$ element column vector formed by concatenating the first row of each matrix $\mathbf{\Theta}_{i,0}$, in order, and then taking the transpose. Let $\mathbf{\Phi}_2$ be formed similarly from the second row of each $\mathbf{\Theta}_{i,0}$. The goal is to estimate the parameters of these vectors. We form two different data matrices, $\mathcal{X}_D$ and $\mathcal{X}_I$, two diagonal weight matrices $\mathbf{W}_D$ and $\mathbf{W}_I$, and column vectors $\mathbf{u}$ and $\mathbf{v}$. For each direct constraint term $w_{m,0;i,j}\|\mathbf{\Theta}_{m,0}\mathbf{X}(\mathbf{p}^0_{m,i}) - \mathbf{p}^m_{0,j}\|^2$ a row of the form

$$(0\cdots 0\ \mathbf{X}(\mathbf{p}^0_{m,i})^T\ 0\cdots 0)$$

is entered into $\mathcal{X}_D$, where $6(m-1)$ 0's precede $\mathbf{X}(\mathbf{p}^0_{m,i})^T$ and $6(N-m)$ 0's follow. Also, $w_{m,0;i,j}$ is entered into the corresponding diagonal entry of $\mathbf{W}_D$, and the x and y terms of $\mathbf{p}^m_{0,j}$ are entered into the corresponding rows of $\mathbf{u}$ and $\mathbf{v}$. For each indirect constraint term, $\mathbf{\Theta}_{m,0}\mathbf{X}(\mathbf{p}^n_{m,i}) - \mathbf{\Theta}_{n,0}\mathbf{X}(\mathbf{p}^m_{n,j})$, a row of the form

$$(0\cdots 0\ \mathbf{X}(\mathbf{p}^n_{m,i})^T\ 0\cdots 0\ -\mathbf{X}(\mathbf{p}^m_{n,j})^T\ 0\cdots 0)$$

is entered into $\mathcal{X}_I$ (assuming $m<n$), where $6(m-1)$ 0's precede $\mathbf{X}(\mathbf{p}^0_{m,i})^T$, $6(N-m)$ 0's follow $\mathbf{X}(\mathbf{p}^m_{n,j})^T$,

and $6(n-m-1)$ 0's are in between. Also, $w_{m,n;i,j}$ is entered into the corresponding diagonal entry of $\mathbf{W}_I$. With these matrix notations, we can rewrite (4) as follows:

$$\begin{aligned}\mathcal{E}(\mathbf{\Phi}_1, \mathbf{\Phi}_2) =&(\mathcal{X}_D\mathbf{\Phi}_1 - \mathbf{u})^T\mathbf{W}_D(\mathcal{X}_D\mathbf{\Phi}_1 - \mathbf{u}) + (\mathcal{X}_D\mathbf{\Phi}_2 - \mathbf{v})^T\mathbf{W}_D(\mathcal{X}_D\mathbf{\Phi}_2 - \mathbf{v}) \\ &+ (\mathcal{X}_I\mathbf{\Phi}_1)^T\mathbf{W}_I(\mathcal{X}_I\mathbf{\Phi}_1) + (\mathcal{X}_I\mathbf{\Phi}_2)^T\mathbf{W}_I(\mathcal{X}_I\mathbf{\Phi}_2) \end{aligned} \tag{5}$$

Taking the derivatives with respect to $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$, setting each equal to $\mathbf{0}$, and solving yields the estimate:

$$\begin{aligned}\hat{\mathbf{\Phi}}_1 &= (\mathcal{X}_D^T\mathbf{W}_D\mathcal{X}_D + \mathcal{X}_I^T\mathbf{W}_I\mathcal{X}_I)^{-1}\mathcal{X}_D^T\mathbf{W}_D\mathbf{u}, \\ \hat{\mathbf{\Phi}}_2 &= (\mathcal{X}_D^T\mathbf{W}_D\mathcal{X}_D + \mathcal{X}_I^T\mathbf{W}_I\mathcal{X}_I)^{-1}\mathcal{X}_D^T\mathbf{W}_D\mathbf{v}. \end{aligned} \tag{6}$$

From these estimates, the individual transformation parameter matrices, $\hat{\mathbf{\Theta}}_{m,0}$, may be recovered immediately.

4 Forming the Correspondence Graph

Given the above linear technique for jointly estimating the transformations of all images onto the anchor frame, we still have the problem of determining which pairs of non-anchor images to register pairwise to form indirect constraints. In effect this is the "topology" problem defined in [24]. Here, we can not assume any ordering to the image sequence to aid in solving the problem.

This may be formulated as a graph theoretic problem. The image index set $\mathcal{I}$ determines the vertices of the graph, and the image correspondence set $\mathcal{P}$ defined above, when built, forms the edges. Clearly, the graph must be connected, but connectivity is not sufficient. The example shown in the introduction illustrates this (Figure 2). To obtain maximal consistency between transformations, we attempt to build the largest possible set of indirect constraints for any given set of images. In a graph theoretic sense, this means that we attempt to build a maximally connected graph.

The simple means for building a maximally connected graph is to apply the pairwise registration algorithm to all $O(N^2)$ pairs. We can do better by incrementally constructing the graph, using each image in turn as the anchor. Initially, all images are registered pairwise against the anchor. Edges are added to the graph between the vertices corresponding to successfully registered pairs. Subsequently, when a new image is chosen as the (temporary) anchor, the partially constructed graph is used in the joint solution (Equation 6) to predict the transformations onto this anchor and tell which images do

not overlap it. Pairwise registration with the anchor is not performed for such images. Overall, this procedure runs pairwise matching $O(N^2)$ times in the worst case and runs the joint solution $O(N)$ times.[2] In practice, approximately 45% of the registration pairs are discarded due to insufficient overlap without running the registration algorithm.

5 Grayscale Mosaic Synthesis Algorithm

For any image, I_0, chosen as the anchor image, the constructed correspondence graph is used to determine which constraint sets are direct and which are indirect, and the joint solution (Equation 6) is used to compute the image-to-anchor transformation estimates $\hat{\mathbf{\Theta}}_{1,0}, \ldots, \hat{\mathbf{\Theta}}_{N,0}$. In constructing the mosaic, each image I_m is mapped to a new image I_m^* in the anchor coordinate system, and then the images I_m^* are combined. The intensities of I_m are normalized prior to mapping.

The mapping of discrete pixels poses a problem because the transformation function is not algebraically invertible. The usual method would be to inverse map each pixel location (u, v) in I_m^* back to I_m and interpolate from the surrounding (discrete) pixel locations. We can do this here as well by numerically inverting the transformation $\hat{\mathbf{\Theta}}_{m,0}$. This requires an initial forward mapping, which is taken from the center pixel in I_m. Using the inverse mapping at each pixel to predict the inverse mapping at surrounding pixels, this can be made nearly as inexpensive as the forward mapping.

In combining intensities from several overlapping images in the mosaic, a number of different weighting functions can be used. The simplest one is uniform weighting. A second weights each pixel in inverse proportion to the squared distance from the center in the original image, I_m. This tends to reduce image boundary seams in the resulting mosaic. A third weights each pixel in proportion to the amount of compression in the transformation, so that pixels that are "spread out" by the transformation receive less weight. These weights are used in a final weighted-average calculation of intensity.

6 Results

We have applied our "joint solution" mosaic construction technique to image data sets taken from 16 different eyes. Each dilated eye was imaged using a TOPCON IMAGENET megapixel digital fundus camera at the Center for Sight (Albany, NY). Up to 20 images from each eye were collected in the center and around the periphery of the retina.

[2] This $O(N^2)$ complexity is not a major concern in the retinal image application because typically fewer than 20 images are combined in each mosaic.

6.1 Example Mosaics

The primary means of evaluating the joint solution is to examine resulting mosaics both visually and numerically. Example mosaics from three different eyes are shown in Figure 4. In some cases the anchor image is chosen near the center of the eye, whereas in others the anchor is near the periphery. All but mosaic (b) were constructed using the third method for image intensity combination, which preserves the original intensity structure of the images, while producing a smooth mosaic except at the image boundaries. Mosaic (b) illustrates use of the second method for intensity combination, which produces a smoother combination especially at boundaries between images. Such a method could be controversial, however, because the intensities are changed more substantially than with other methods. Example mosaics for all datasets are available on our website.[3]

The mosaics may also be evaluated numerically. The measure used is an extended version of the pairwise centerline error measure (CEM) discussed at the end of Section 2. Consider a pair of non-anchor images, I_m and I_n, connected by an edge in the pairwise registration graph. First, transform each centerline pixel $\mathbf{p}_n$ from I_n onto the anchor: $\mathbf{p}_n^* = (x_n^*, y_n^*)^T = \hat{\mathbf{\Theta}}_{n,0}\mathbf{X}(\mathbf{p}_n)$. In addition, calculate the Jacobian of the transformation at each transformed centerline pixel: $\mathbf{J}_n(\mathbf{p}_n^*) = \frac{\partial \hat{\mathbf{\Theta}}_{n,0}\mathbf{X}(\mathbf{p}_n)}{\partial \mathbf{p}}$. Next, for each centerline pixel $\mathbf{p}_m$ in I_m, transform the pixel onto the anchor to compute

$$\mathbf{p}_m^* = (x_m^*, y_m^*)^T = \hat{\mathbf{\Theta}}_{m,0}\mathbf{X}(\mathbf{p}_m),$$

and then find the closest transformed centerline pixel $\mathbf{p}_n^*$ from I_n. Finally, the normalized squared distance between these centerline pixels is

$$(\mathbf{p}_m^* - \mathbf{p}_n^*)^T \left[\mathbf{J}_n(\mathbf{p}_n^*)\right]^{-1} (\mathbf{p}_m^* - \mathbf{p}_n^*) \tag{7}$$

The normalization is needed to account for compression/expansion effects of the transformation, especially near the periphery.[4] Note that when $n = 0$, the transformation $\hat{\mathbf{\Theta}}_{n,0}$ becomes the identity, as does the Jacobian, and therefore the normalized centerline error measure reduces to the original pairwise error measure.

The distance in Equation 7 is the basis for two error measures. First, for each corresponding pair, we calculate the median error. The average of these median errors is taken as the "average CEM". Second, we combine all errors from all images and take the median, forming the "combined median CEM". The

[3] *http://www.cs.rpi.edu/~stewart/joint_sol.html*

[4] In practice, the Jacobians are nearly diagonal, so we just take the diagonal terms, simplifying the computation.

median is used in both cases because the recursive tracing algorithm does not always trace each vessel in each image.

Over all 16 datasets, the average CEM was never higher than 0.86 pixels; the combined median CEM averaged 0.76 pixels and was never higher than 0.79. These numbers demonstrate the stability of the joint solution resulting from the combined use of direct and indirect constraint sets.

6.2 Illustrating the Roles of Direct and Indirect Constraints

The significance of indirect constraints in reducing alignment error in the mosaics was shown with an example in the introduction. This section shows a more detailed example using the CEM measures.

The pairwise registration graph and joint solution results for five images are shown in Figure 5. The image labeled 'A' in the graph (a) is the anchor, and the joint solution CEMs using all pairwise registration constraints are shown. Two different reduced registration graphs are also illustrated. In one, shown in Figures 5(b) and 5(d), the image-to-anchor transformations are estimated using only direct constraints. The results indicate both numerically and visually that substantial misregistration occurs. In the second, shown in Figures 5(c) and 5(e), the transformations are estimated using only one direct constraint set and three indirect constraint sets. The results are much improved. In fact, the numerical results from this second reduced graph are slightly better in terms of the average CEM than when all constraints are used (Figure 5(a)), but the maximum pairwise CEM is slightly worse. This raises the interesting theoretical question of how many sets of constraints are necessary to obtain sufficiently accurate alignment. Since the issue is not of immediate concern for the retinal mosaicing application, we defer it to future work.

7 Discussion and Conclusions

This paper has presented a substantially improved solution to the problem of constructing mosaics from a series of images of the human retina. The main new idea is a linear technique for using both direct and indirect constraints from matching pairs of retinal images to jointly estimate the transformations of all images onto the anchor image coordinate system in which the mosaic is built. In addition, we summarized a procedure for incrementally constructing a maximally-connected graph of pairwise correspondences. This algorithm applies the joint solution estimation to initialize transformations and predict which image pairs can not be registered, thereby avoiding unnecessary computation.

Experimentally, we have demonstrated the significance of the new techniques. We showed example mosaics from three of sixteen datasets and gave summary numerical results for the others. These show

precisely aligned images, low numerical error, and almost no variation in this error across all datasets, indicating the overall stability of the joint solution. Detailed consideration of a small example demonstrated the importance of indirect constraints.

The ability to produce accurate mosaics covering a broad region of the retina is important in ophthalmic applications. The physician can now choose any image as the anchor image, see which regions have not been imaged by looking at the incrementally constructed mosaic, and need not worry about identifying a single image that will overlap all others. The protocol for imaging the retina is greatly simplified. The accuracy of the registration implies that more images can be combined and the resulting mosaic has a crisper appearance. Such mosaics are especially valuable for quantitative analysis of the retina and for retinal change detection.

Acknowledgments

This work was supported by NIH grant 1R21 RR14038-01A1, the NSF-supported Engineering Research Center for Sub-surface Imaging and Sensing (CenSSiS), and Rensselaer Polytechnic Institute. A portion of Dr. Can's time was supported by Inonu University, Turkey. The authors gratefully acknowledge the help from the professional staff at the Center for Sight, especially the professional photographers Gary Howe and Marc Fish. The retinal images were of unnamed volunteers whose assistance is greatly appreciated. Thanks are also due to Matthew Freshman for assistance with image capture. Finally, the continuing guidance provided by Dr. James N. Turner at the Wadsworth Center in Albany is greatly appreciated.

References

[1] D. E. Becker, A. Can, H. L. Tanenbaum, J. N. Turner, and B. Roysam. Image processing algorithms for retinal montage synthesis, mapping, and real-time location determination. *IEEE Transactions on Biomedical Engineering*, 45(1), 1998.

[2] J. Bergen, P. Anandan, K. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *Proceedings Second European Conference on Computer Vision*, pages 237–252, 1992.

[3] J. W. Berger and D. S. Shin. Computer vision enabled augmented reality fundus biomicroscopy. *Ophthalmology*, 106(10), Oct 1999.

[4] A. Can, H. Shen, J. N. Turner, H. L. Tanenbaum, and B. Roysam. Rapid automated tracing and feature extraction from live high-resolution retinal fundus images using direct exploratory algorithms. *IEEE Trans. on Info. Tech. for Biomedicine*, 3(2):125–138, 1999.

[5] A. Can, C. Stewart, and B. Roysam. Robust hierarchical algorithm for constructing a mosaic from images of the curved human retina. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pages 286–292, 1999.

[6] A. Can, C. Stewart, B. Roysam, and H. Tanenbaum. A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, to appear, 2001.

[7] D. Capel and A. Zisserman. Automated mosaicing with super-resolution zoom. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pages 885–891, 1998.

[8] S. Chen. QuickTime VR — an image-based approach to virtual environment navigation. In *Proceedings of SIGGRAPH'95*, pages 29–38, 1995.

[9] P. Dani and S. Chaudhuri. Automated assembling of images - image montage preparation. *Pattern Recognition*, 28(1):431–445, March 1995.

[10] J. Davis. Mosaics of scenes with moving objects. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pages 354–360, 1998.

[11] N. Gracias and J. Santos-Victor. Underwater video mosaics as visual navigation maps. *Computer Vision and Image Understanding*, 79(1):66–91, 1999.

[12] F. R. Hampel, P. J. Rousseeuw, E. Ronchetti, and W. A. Stahel. *Robust Statistics: The Approach Based on Influence Functions*. John Wiley & Sons, 1986.

[13] P. W. Holland and R. E. Welsch. Robust regression using iteratively reweighted least-squares. *Commun. Statist.-Theor. Meth.*, A6:813–827, 1977.

[14] M. Irani and P. Anandan. Video indexing based on mosaic representations. *Proceedings of the IEEE*, 86(5):905–921, 1998.

[15] M. Irani, P. Anandan, and S. Hsu. Mosaic based representations of video sequences and their applications. In *Proceedings IEEE International Conference on Computer Vision*, pages 605–611, 1995.

[16] A. A. Mahurkar, M. A. Vivino, B. L. Trus, E. M. Kuehl, M. B. Datiles, and M. I. Kaiser-Kupfer. Constructing retinal fundus photomontages. *Investigative Ophthalmology and Visual Science*, 37(8):1675–1683, July 1996.

[17] D. Milgram. Computer methods for creating photomosaics. *IEEE Trans. on Computers*, 23:1113–1119, 1975.

[18] D. Milgram. Adaptive techniques for photomosaicking. *IEEE Trans. on Computers*, 26:1175–1180, 1977.

[19] J. Mundy and A. Zisserman, editors. *Geometric Invariance in Computer Vision*. MIT Press, 1992.

[20] S. Peleg. Elimination of seams from photomosaics. *CGIP*, 16(1):90–94, 1981.

[21] S. Peleg and M. Ben-Ezra. Stereo panorama with a single camera. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pages I:395–401, 1999.

[22] S. Peleg and J. Herman. Panoramic mosaics by manifold projection. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pages 338–343, 1997.

[23] P. J. Rousseeuw. Least median of squares regression. *Journal of the American Statistical Association*, 79:871–880, 1984.

[24] H. Sawhney, S. Hsu, and R. Kumar. Robust video mosaicing through topology inference and local to global alignment. In *Proceedings 5th European Conference on Computer Vision*, volume II, pages 103–119, 1998.

[25] H. Shen, B. Roysam, C. Stewart, J. Turner, and H. Tanenbaum. Optimal scheduling of tracing computations for real-time vascular landmark extraction from retinal fundus images. *IEEE Transactions on Information Technology in Biomedicine*, 5(1):77–91, Mar 2001.

[26] H. Shum and R. Szeliski. Systems and experiment paper: Construction of panoramic image mosaics with global and local alignment. *International Journal of Computer Vision*, 36(2):101–130, 2000.

[27] H.-Y. Shum and R. Szeliski. Construction and refinement of panoramic mosaics with global and local alignment. In *Proceedings IEEE International Conference on Computer Vision*, pages 953–958, 1998.

[28] R. Szeliski. Video mosaics for virtual environments. *IEEE Computer Graphics and Applications*, 16(2):22–30, 1996.

[29] M. Tanaka, S. Tamura, and K. Tanaka. On assembling subpictures into a mosaic picture. *IEEE Transactions on Systems, Man and Cybernetics*, 7:42–48, 1977.

[30] J. Um and R. Wright. Video strip mosaicking: a two-dimensional approach by convergent image bridging. *International Journal of Remote Sensing*, 20(10):2015, July 1999.

[31] A. Zappalá, A. Gee, and M. Taylor. Document mosaicing. *Image and Vision Computing*, 17(8):589–595, 1999.

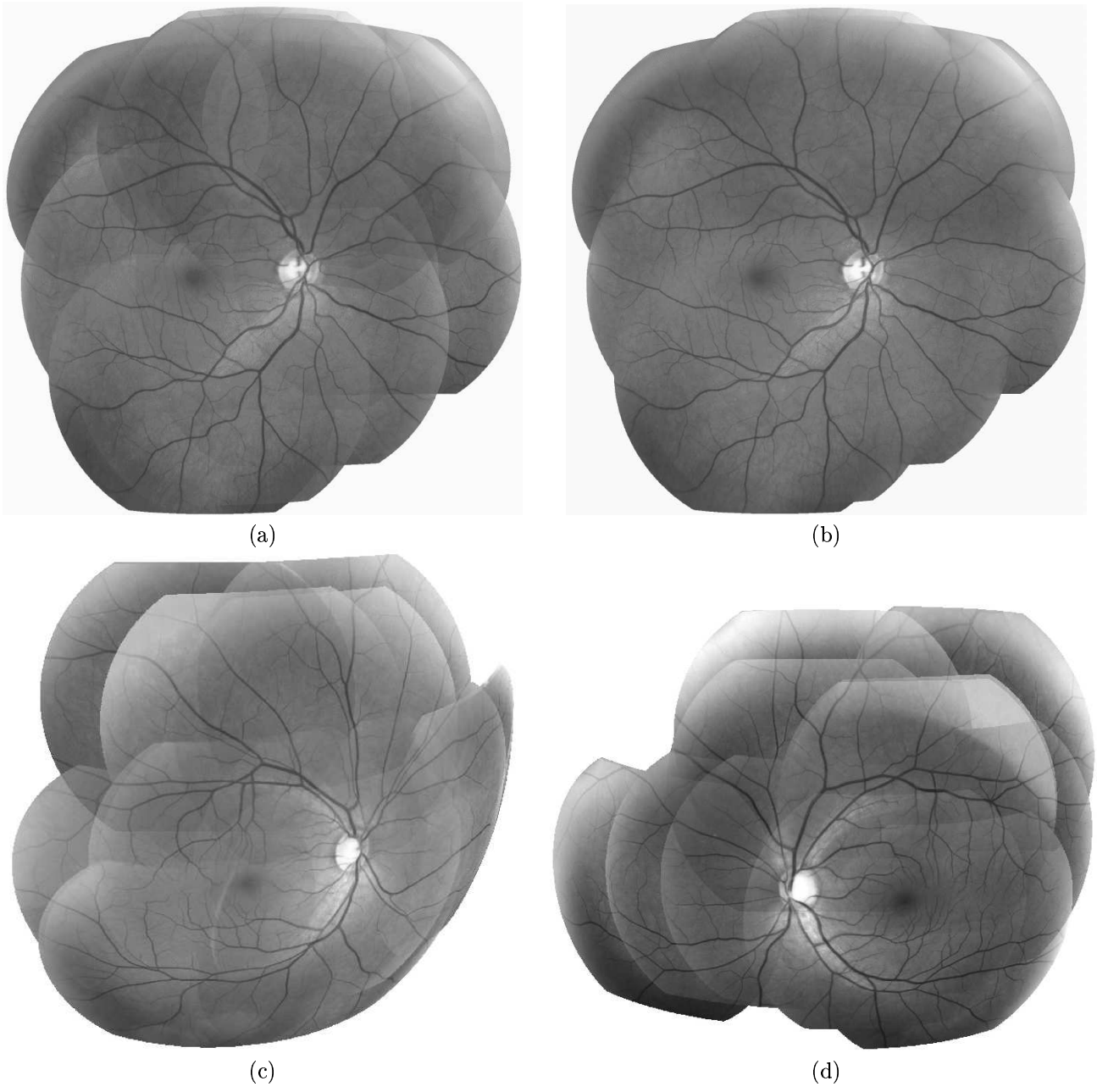


Figure 4: Example mosaics formed from image sets taken from three different eyes. Mosaics (a), (c) and (d) use intensity weighting based on the amount of compression. Mosaic (b), which is formed from the same images and transformations as (a), uses intensity weighting inversely proportional to a pixel's distance from the original image center. All mosaics have approximately 1500×1500 pixels. Transitions between images due to illumination differences are clearly visible in (a), (c) and (d), but the vasculature appears perfectly aligned.

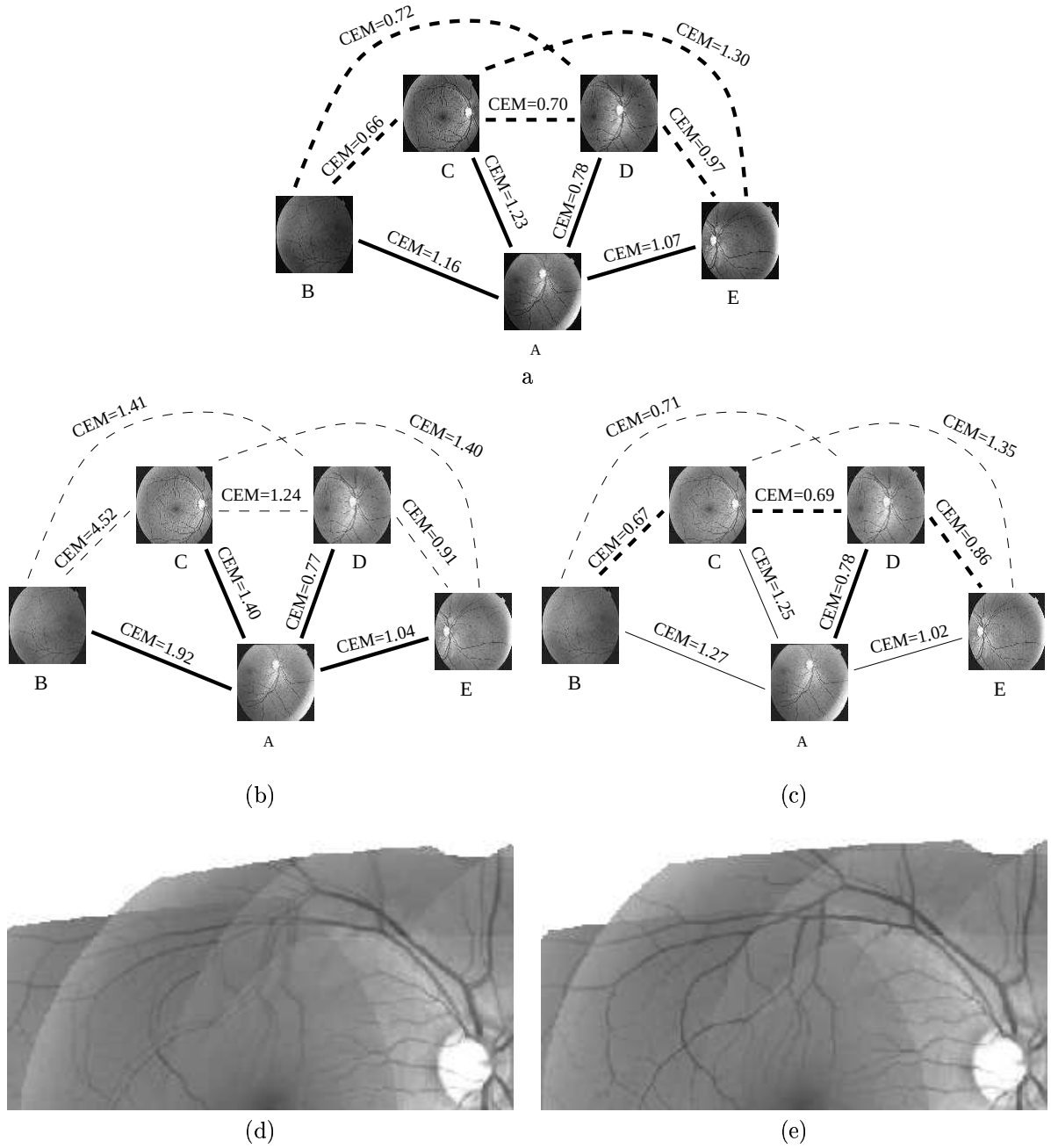


Figure 5: Constraint set graphs and mosaic regions for the five image example illustrating the roles of direct and indirect constraint sets in the joint solution. Image 'A' is the anchor. The full set of registered pairs is shown in (a) with direct constraints are shown as solid edges, and indirect constraints shown with dashed edges. Graph (b) and image (d) give the results of using four sets of direct constraints, shown as thick edges in (b). Graph (c) and image (e) give the results of using one set of direct constraints and three sets of indirect constraints, shown as thick edges in (c). The resulting CEM values, especially between images B and C, are much better in (c). The misalignment between these images using only direct constraints is clear in the cropped region of the mosaic in (d).