

# Vessel Optimal Transport for Automated Alignment of Retinal Fundus Images

Danilo Motta, Wallace Casaca, Afonso Paiva

**Abstract**—Optimal transport has emerged as a promising and useful tool for supporting modern image processing applications such as medical imaging and scientific visualization. Indeed, the optimal transport theory enables great flexibility in modeling problems related to image registration, as different optimization resources can be successfully used as well as the choice of suitable matching models to align the images. In this paper, we introduce an automated framework for fundus image registration which unifies optimal transport theory, image processing tools and graph matching schemes into a functional and concise methodology. Given two ocular fundus images, we construct representative graphs which embed in their structures spatial and topological information from the eye’s blood vessels. The graphs produced are then used as input by our optimal transport model in order to establish a correspondence between their sets of nodes. Finally, geometric transformations are performed between the images so as to accomplish the registration task properly. Our formulation relies on the solid mathematical foundation of optimal transport as a constrained optimization problem, being also robust when dealing with outliers created during the matching stage. We demonstrate the accuracy and effectiveness of the present framework throughout a comprehensive set of qualitative and quantitative comparisons against several influential state-of-the-art methods on various fundus image databases.

**Index Terms**—Retinal Image Registration, Image Alignment, Blood Vessel Detection, Optimal Transport.

## I. INTRODUCTION

In 2012, the World Health Organization (WHO) estimated that 39 million people in the world are blind, 285 million are visually impaired and 246 million have low vision degree [1]. Considering all these disorders, *glaucoma*, a serious disease that affects the eyes, is considered the second leading cause of blindness worldwide [2]. According to the American Academy of Ophthalmology, glaucoma is a complicated condition that damages the optic nerve. It occurs when a fluid (called aqueous) builds up in the front part of the eye, increasing the pressure on it. In general, the glaucoma pathology can be broadly classified into two types: the “open-angle”, and the “closed-angle” (or “angle closure”), both of them described regarding the angle delimited between the iris and cornea. The open-angle case, more drastic, appears suddenly leading to the loss of vision quickly while the closed-angle tends to advance at a slower rate progressively. As the medical diagnosis is mostly accomplished by the human inquiry for glaucoma and other eye disorders, the use of image processing algorithms became a necessity [3], [4], especially when ophthalmologists need to manage a large set of fundus images.

Such computing apparatus has paved the way for clinicians and medical specialists to cover more patients while still seeking for greater diagnostic accuracy. However, in practice, medical inspections of retinal images are still manually performed in an attempt to carefully identify and track the evolution of eye diseases. Moreover, these visual inspections are quite time-demanding, as they depend on the physician experience in order to succeed [5], noticing that some pathologies can require intensive examination over many years to be finally detected and treated.

Regarding medical imaging tools, there are several solutions devoted to identifying eye diseases and other related disorders [6], [7], [8], [9], [10], being a small group directly aimed to register particular changes in pairs of retinal images. As physicians capture a set of images from the same patient over time, techniques designed to align retinal images help these experts monitor any sign of disease manifestation. Although existing algorithms for image registration produce satisfactory results under certain conditions, they can deform the sequence of fundus images in many practical situations. For instance, if the overlap between two fundus images is not so apparent, the relative changes in the images become computationally unfeasible to be detected [11]. Additionally, retinal images often present textureless regions that are irregularly illuminated, hampering the correct matching between the blood vessels in the acquired images.

Most recently, the Optimal Transport (OT) theory [12], [13] has figured as a robust and flexible tool to address sophisticated problems in many fields of study, as specific optimization models can be built to solve the desirable problem effectively. It has appeared in different contexts of applications including *Economic Theory* [14], *Geometry Processing* [15] and *Pattern Analysis* [16], just to cite a few examples. Despite its effectiveness and high capability in modeling complex problems, only a very reduced number of OT-inspired approaches have been devised to tackle medical imaging applications specifically. Among these particular ones, Harker et al. [17], [18] apply OT concepts to formulate a first order PDE which gives a deformation map (a grid) used to align images of specific categories (e.g., sky and MIR images). Another interesting work is the survey compiled by Papadakis [16], which offers a comprehensive overview of modern OT-based methods, not specific for medical imaging, but actually for general image processing purposes.

In this paper, we present an automated registration framework for aligning blood vessels in retinal fundus images. The present approach relies on the stable theory of OT in conjunction with graph-based models to precisely match retinal blood vessels. The performance and usefulness of

D. Motta and A. Paiva are with ICMC, University of São Paulo (USP), São Carlos, Brazil. Email: {daniloam, apneto}@icmc.usp.br

W. Casaca is with São Paulo State University (UNESP), Rosana, Brazil. Email: wallace@icmc.usp.br

the proposed methodology are validated by a comprehensive set of comparisons against competing and well-established image registration methods, by using real medical datasets and classic measures typically employed as a benchmark by the medical imaging community. In contrast to most existing OT-based methods that consider deformation grids or deal with specific kind of image classes, our methodology provides a new facet of OT as a constrained graph matching problem to align vascular vessels on retinal images. Graph representatives are constructed from the pair of the acquired images so that their nodes are viewed as a multi-valued set of features and evaluated as keypoints to achieve the registration. Once the keypoints are obtained, we compute their direct correspondence between the images, by solving the proposed OT problem. The designed optimization model allows us to establish the matching from a customized cost function that penalizes outliers as long as the matches are determined.

**Key aspects and main contributions.** The proposed framework, which we call **VOTUS** (*Vessel Optimal Transport for fundUS image alignment*), retains some desirable characteristics when registering images such as registration accuracy for pairs of images contaminated by specular noise, high-quality performance for high-resolution retinal images, and stability w.r.t. the systematic registration task, as shown in our battery of experimental tests. Another peculiarity of VOTUS is that its output can enhance the perception of anatomy changes between the retinal images so that it may assist ophthalmologists in preventing and diagnosing eye diseases and other related impairments, as reported by two domain experts in Section V. Particularly, we can summarize the technical contributions of this paper as follows:

- An automatic registration framework which relies on Optimal Transport theory as a *Mixed Integer Linear Programming Formulation* to match and align vessels extracted from a pair of retinal fundus images;
- A new OT cost function for establishing blood vessels correspondences built in terms of our customized set of dissimilarity measures and discriminative features;
- A fully deterministic outlier removal filter which benefits from our OT cost matrix, becoming more assertive and effective against outliers (noise);
- A discriminative saliency-based fundus image feature invariant under translation and scale transformations.

## II. RELATED WORK

Due to the vast amount of research on medical image registration, several approaches have been proposed over the last years, ranging from image alignment of the same modality to multimodal registration systems [19], [20]. In order to better contextualize our contributions, we focus our discussion on techniques devoted to aligning images from a unique imagery response. Since there are a huge amount of work in the literature, an in-depth survey about all those registration methods is beyond the scope of this paper.

We group fundus image registration methods into three basic and well-established classes: *vessel-based*, *intensity-based*, and *keypoint-based*. The first category, vessel-based, relies on

graph-based representations of the eye's retina to accomplish the registration, while the intensity-based one performs the eye's alignment by exploiting specific attributes of the images such as colors, contrasts, and gradients. Finally, the last group of methods aims at establishing a correspondence between sets of keypoints extracted from the retinal images.

**Vessel-based** techniques attempt to generate a tree-like model representing the vascular structures of the retina so that the topological maps of these structures are detected and captured. In fact, techniques from this category typically comprise two core steps: *vessel extraction* and *vascular registration*. For instance, Can et al. [21] accomplished both tasks by combining curvature attributes of the retina surface with a motion eye predictive algorithm. Similarly, Chen et al. [22] proposed a structure-matching mechanism that encodes bifurcation areas located at the vessels, embedding into characteristic vectors the length and the branching angles of these regions. By measuring the similarity between those vectors, they define a transformation model to match the similar structures of the image. Unlike [22], an elastic scheme was introduced by Fang and Tang [23] to reconstruct "vascular trees", where an optimization function is computed until the optimality matching convergence between rearranged and reference vascular models is achieved.

Later, Palraj and Venilla [24] formulated a hybrid approach for which the registration is performed considering mutual information-based metrics and a simulated annealing scheme. Deng et al. [11] proposed a graph matching method based upon topological structures and vessel shapes as feature points wherein the image registration is converted into an edge-to-edge correspondence problem. Similar to [11], Chen et al. [25] presented a vascular image registration algorithm based on network structures and circuit simulation. Serradell et al. [26] introduced a method called *Active Testing Search for Robust Graph Matching* (ATS-RGM) for matching 2D and 3D graph structures got from microscopy image stacks, angiography data, and retinal fundus images. Still regarding vessel-driven methods, Ghassabi et al. [27] presented an interesting structure-based registration approach which is invariant under rotation and small-scale changes. First, the method extracts enclosed regions from the vessels using watershed-based segmentation. Next, a boundary-based region descriptor computes the similarity between these regions in order to determine their final correspondences. In a more recent work, Braun et al. [28] proposed a new method called *EyeSLAM* for real-time retinal image registration applied to robot-assisted intraocular microsurgery. Their approach combines a correlative scan-matching model based on vasculature trees to register motion.

**Intensity-based** methods encode brightness properties of the image such as gradients, color contrast, and saturation into either a cost function or specific decision rules to proceed with the vessel alignment. They make use of optimization apparatus or customized decision trees such as *simulated annealing* [29], *genetic algorithms* [30], [31] and *particle swarm optimization* [32] to find a suitable minimizer to the corresponding registration problem. Although mathematically well-posed, an important issue when using methods purely based on the image intensity is their computational cost, as

finding the globally optimal solution may be time-consuming, especially for high-resolution images. To overcome this, the framework proposed by Zaslavskiy et al. [33] is an authentic intensity-based method that addresses the problem of finding a global optimum by reaching near-optimal solutions. More specifically, they combine several image processing filters so that a graph matching problem is solved to promote the final alignment. Chen et al. [34] also applied different image processing operators, giving rise to a particular feature called *Partial Intensity Invariant Feature Descriptor* (PIIFD). This framework integrates the classic *Harris-Stephen corner detector* [35], gradient-based filters and image histograms to produce a set of PIIFD-derived features. The resulting feature vectors are then matched according to a nearest-neighbor criterion. Another representative intensity-based method is the one presented by Adal et al. [36], where a hierarchical retina image registration scheme combining intensity and vascular structural data was proposed. Firstly, an image normalization is accomplished in order to reduce the local luminosity effects and contrast variations on the retina during the image acquisition. Finally, the image alignment is performed using a similarity metric that is based on the vascular-weighted mean squared difference between the images.

**Keypoint matching-based** methods face the registration problem via point cloud perspective. In this kind of approach, locally invariant features are extracted from the images so that a correspondence between two sets of feature points is determined and a geometric transformation is applied. One of the most influential algorithms in this category is the *Iterative Closest Point* (ICP) [37]. It is the basis of the Yang et al.'s method [38], where the authors presented the so-called *General Dual-Bootstrap ICP* (GDB-ICP). Starting in a bootstrap region (i.e., a small area of overlap between the images), GDB-ICP computes the well-known *Scale Invariant Feature Transform* (SIFT) descriptor [39] to obtain an initial transformation, improved by providing edge and corner keypoints within a growing bootstrap region to an ICP-based algorithm. Although quite versatile, GDB-ICP can generate inconsistent outputs under high contrast variation in high-resolution retinal images, as previously reported in [40]. Tsai et al. [41] improved the GDB-ICP as the *Edge-Driven Dual-Bootstrap ICP* (ED-DB-ICP). This algorithm incorporates vascular features taken from the retinal images into the SIFT keypoints to increase the method's accuracy for the registration task.

Wang et al. [42] provided an automatic multimodal retinal image registration framework called *Robust Point Matching* (RPM) method, wherein the technique benefits from Gaussian models to reduce the number of outliers. Their approach relies on image gradients as the PIIFD descriptor combined with a robust local feature detector called *Speeded-Up Robust Features* (SURF) [43]. In this spirit, Cattin et al. [44] have integrated SURF with the classic outlier detection algorithm *Random Sample Consensus* (RANSAC) [45] for retina mosaicing. Outlier removal has also been the goal of Ong et al. [46], but for multimodal image registration instead, where a *Least Trimmed Square*-based estimator is applied to prune inconsistent matches effectively.

Hernandez-Matas et al. [47], [48] introduced an interesting keypoint-based registration technique that takes advantage of the geometry present on the eye fundus, by exploiting a spherical model representation of the eye and the relative pose estimation handled by acquisition cameras. Hernandez-Matas et al. [49] later proposed an improved version of their previous method, where both the SIFT descriptor and the vessel bifurcations are used in order to extract more refined keypoints. Keypoints were also successfully used in [50] in conjunction with graph-based models as a general image alignment framework based on quadratic optimization. Although mathematically elegant, the aforementioned model does not focus on fundus image registration specifically, it is actually best used for face matching. Recently, Wang et al. [51] proposed a registration method called *Gaussian Field Estimator with Manifold Regularization* (GFEMR). This technique introduces a manifold regularization to preserve the intrinsic geometry of the retina and a Gaussian probabilistic model to determine a non-rigid transformation.

In overall, the main advantage of keypoint matching-based methods is their flexibility in handling various image classes, including the ocular fundus one, however, most existing methods rely on specific random mechanisms such as RANSAC whose outputs can differ for the same image pair every time it is executed.

The technique described in this paper provides a fully automated Optimal Transport-driven framework for fundus image registration. As discussed in the experiments, the method gives deterministic solutions for the matching task, is computationally efficient when dealing with high-resolution retinal images, and achieves high registration accuracy, even for images containing critical degenerations such as specular noise, lack of focus and low image quality. Since the core contribution of our technique is the OT optimization pipeline, it can be understood as a vessel-based method, but other traits commonly found in the other two categories also appear as part of our unified framework. Finally, the paper also covers theoretical advances in the OT theory for blood vessel correspondence and an efficient outlier removal filter.

### III. THE UNDERLYING THEORY OF OPTIMAL TRANSPORT

In 1781, the mathematician Gaspar Monge published his famous problem: “*Assuming that a certain amount of sand must be conveyed from the soil to construction sites, were input and output locations are known, what is the best way to perform such a transportation?*”. In summary, we need to discover the optimal map that assigns to the construction sites portions of sand coming from different resources.

Considering that the material transportation has a cost, the optimal map should have a minimum total cost. Formally, this problem can be relaxed throughout the classical *Kantorovich model*, also called *Monge-Kantorovich formulation* [52]. Let  $\mathcal{X}$  and  $\mathcal{Y}$  be two metric spaces in which two measures  $\mu$  and  $\nu$  are defined. Then, for two measurable subsets  $\mathcal{A} \subset \mathcal{X}$  and  $\mathcal{B} \subset \mathcal{Y}$ ,  $\mu(\mathcal{A})$  provides the measure of the mass quantity located in  $\mathcal{A}$ , while  $\nu(\mathcal{B})$  gives the second measure, but rather related to  $\mathcal{B}$ . We are looking for

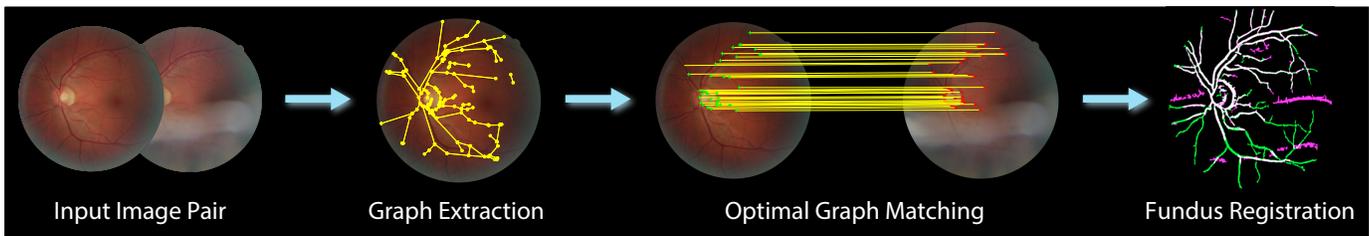


Fig. 1: Pipeline overview of the proposed approach, VOTUS.

the optimal transportation plan taking into account those ones that are modeled in terms of a *measure coupling*  $\pi$  on the product space  $\mathcal{X} \times \mathcal{Y}$ . In fact, a valid transportation plan must preserve measure (or mass), i.e.,  $\pi(\mathcal{A} \times \mathcal{Y}) = \mu(\mathcal{A})$  and  $\pi(\mathcal{X} \times \mathcal{B}) = \nu(\mathcal{B})$ .

Let  $C(\mathbf{x}, \mathbf{y})$  be the cost to transport one unit of mass, from location  $\mathbf{x} \in \mathcal{X}$  to  $\mathbf{y} \in \mathcal{Y}$ . Denoting the set of valid transportation plans as  $\Pi(\mu, \nu)$ , the optimal transport problem can be formulated as follows:

$$\min_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} C(\mathbf{x}, \mathbf{y}) d\pi(\mathbf{x}, \mathbf{y}), \quad (1)$$

where  $d\pi(\mathbf{x}, \mathbf{y})$  denotes the integration against the measure  $\pi$ .

Taking  $\mathcal{X}$  and  $\mathcal{Y}$  as discrete spaces, where the points have  $\mu$  and  $\nu$  as corresponding discrete measures, the transportation plans can be represented by a matrix  $\pi_{ij}$  which holds:

$$\begin{cases} \forall \mathbf{x}_i \in \mathcal{X}, \sum_j \pi_{ij} = \mu(\mathbf{x}_i) \\ \forall \mathbf{y}_j \in \mathcal{Y}, \sum_i \pi_{ij} = \nu(\mathbf{y}_j) \end{cases}. \quad (2)$$

Then, the discrete optimal transport problem takes the form:

$$\min_{\pi \in \Pi'} \sum_{i,j} C(\mathbf{x}_i, \mathbf{y}_j) \pi_{ij}, \quad (3)$$

where  $\Pi'$  states that all transportation plans must satisfy Equation (2). When the cost  $C(\mathbf{x}, \mathbf{y})$  is chosen as  $\|\mathbf{x} - \mathbf{y}\|_p$ , the optimization problem (3) is known as the *p-Wasserstein distance* [13], while for the particular case in which  $p = 1$ , this problem is called as *Earth Mover's distance* [53].

**Optimal transport vs graph matching.** In computer vision and image processing, the matching of feature points from graphs is typically a *Quadratic Assignment Problem* (QAP), known to be an NP-hard task [54]. Usually, QAP is approximated using *Semidefinite Programming* (SDP) [55]. However, such an approximation is not always feasible in practice, especially for large graphs, turning the use of SDP very prohibitive in several imaging applications. In contrast, the discrete OT formulation (3) provides an approximation of the graph matching problem adapted to the geometry of the feature space (input density). This relaxation leads to a linear program model which works efficiently while still remaining easy-to-solve for many imaging problems [56], [16].

In our application, the fundus registration task is modeled as a graph matching problem that can be successfully solved as a discrete OT problem, by setting a valid cost function  $C(\mathbf{x}, \mathbf{y})$  built in terms of a specific set of similarity measures. As shown in Section V.C, our OT-based model tends to generate more accurate results when compared to other graph matching

methods, especially in cases where the graphs have artifacts, a harder occurrence efficiently addressed by our formulation.

#### IV. REGISTRATION BY OPTIMAL TRANSPORT

As shown in Figure 1, the designed framework comprises three main steps: *Graph extraction*, *Optimal graph matching*, and *Fundus registration*. First, we construct representative graphs for the vessels in the images so that the key attributes of the images (e.g., local shapes and topological relations) are captured and embedded into these graphs. Next, the graphs produced are used as input by our optimal transportation model to establish the correspondence between their sets of nodes. Potential outliers are then automatically detected and removed from the matches previously originated, while in the last step, geometric transformations are finally applied on the images to proceed with the fundus registration. The details of each above-mentioned step are presented below.

##### A. Graph extraction

**Graph construction.** Given two retinal fundus images  $I$  and  $R$ , the first step of our framework consists in generating graph-based representations of  $I$  and  $R$ , more precisely, to represent the retinal veins of the images (see Figure 2, right). In order to construct such reference graphs, we initially use the technique proposed by Bankhead et al. [57], to create binary images  $B_I$  and  $B_R$ . These binarizations reflect into their structures vessel attributes such as lines, shapes and topological groupings. From the vessel binaries  $B_I$  and  $B_R$ , we generate skeleton images  $S_I$  and  $S_R$  by applying the iterative thinning method [58], which shrinks large pixelated regions while still removing possible outliers. After the skeletons generation, we build two reference graphs  $G_I(V_I, E_I)$  and  $G_R(V_R, E_R)$ , by associating each segmented pixel from  $S_I$  (resp.  $S_R$ ) as a node of  $V_I$  (resp.  $V_R$ ). The edges of  $E_I$  (analogously for  $E_R$ ) are defined by connecting two nodes  $\mathbf{v}_i^I$  and  $\mathbf{v}_j^I$  from  $V_I$ , since inequality  $\|p_i^I - p_j^I\|_\infty \leq \varepsilon$  holds, wherein  $p_i^I$  and  $p_j^I$  are the pixels of  $B_I$  assigned to the nodes  $\mathbf{v}_i^I$  and  $\mathbf{v}_j^I$ . Quantity  $\varepsilon$  is a parameter which establishes how local the edges should be. In our experiments, we set  $\varepsilon = 1$ . Figure 2 shows the above-described substeps to a retinal fundus image.

**Graph simplification.** In order to take off duplicate information carried out by look-alike nodes in  $G_I$  and  $G_R$ , we take the branch points of the skeleton images as keypoints, which are morphologically extracted from the pixels neighborhoods using MATLAB's `bwmorph` function [59], pulling out from

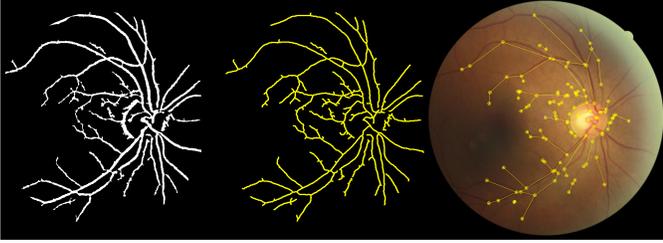


Fig. 2: Graph extraction step (from left to right): blood vessel binarization  $B_I$ , blood vessel skeleton  $S_I$ , and the final reference graph drawn on the retinal image.

the graphs any other internal node. As stated by Lee [60] and Chen et al. [22], these keypoints correspond to significant structures of the blood vessels. Next, the remaining non-branch nodes are progressively removed by a simple set of heuristic rules which relies on the neighborhoods of these nodes and the distances to the pre-specified keypoints. Denoting the 1-ring neighborhood of a node  $\mathbf{v}_I^i$  by  $\mathcal{N}_I^i$ , the simplification rules consist in the following:

- Rule 1:** if  $|\mathcal{N}_I^i| = 1 \Rightarrow$  remove  $\mathbf{v}_I^i$ ;
- Rule 2:** if  $|\mathcal{N}_I^i| = 2 \Rightarrow$  merge  $\mathbf{v}_I^i$  into a node of  $\mathcal{N}_I^i$ ;
- Rule 3:** if  $|\mathcal{N}_I^i| > 2 \Rightarrow$  merge  $\mathbf{v}_I^i$  into its nearest branch point using  $L_\infty$  norm.

Figure 3 depicts the rules of our simplification algorithm. Note that the branch points are preserved during the simplification process so that all nodes of the simplified graphs are considered as keypoints after this stage. Besides the ambiguity reduction, the reduced number of nodes also alleviates the computational burden during the optimization step (Section IV-B). Figure 2 (last column) illustrates the final graph generated after the simplification process is terminated.

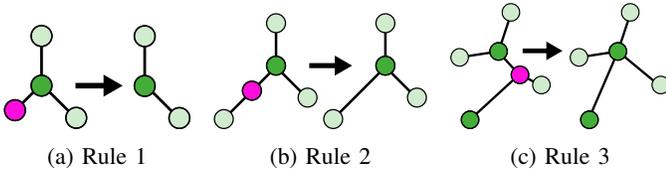


Fig. 3: Simplification rules for a node  $\mathbf{v}_I^i$  (magenta) given in terms of  $\mathcal{N}_I^i$  and its distance to a branch point (dark green).

### B. Optimal graph matching

**Graph matching by optimal transportation.** Let  $G_I$  and  $G_R$  be the blood vessel graphs computed as in the previous stage of our pipeline. We formulate the graph matching task as an OT problem that leads to “one-to-one” correspondences between the graph nodes while still eliminating noticeable outliers. Our formulation is based on a linear program which solves an optimal assignment problem, an adapted case of the minimum cost flow problem [61]. Specifically, we want to determine the optimal transportation plan  $\mathbf{X}$ , represented by a matrix of size  $N_I \times N_R$ , with  $N_I = |V_I|$  and  $N_R = |V_R|$ , that matches the set of nodes  $V_I$  to  $V_R$ . In other words, each coefficient  $X_{ij}$  of  $\mathbf{X}$

provides the transfer route for conveying a certain quantity of mass from a node  $\mathbf{v}_I^i \in V_I$  to another one  $\mathbf{v}_R^j \in V_R$ .

Mathematically, we can model the transportation problem by solving the following optimization problem:

$$\begin{aligned} & \arg \min_{\mathbf{X}} \sum_{i=1}^{N_I} \sum_{j=1}^{N_R} C_{ij} X_{ij} \\ & \text{subject to} \quad \begin{cases} \sum_j X_{ij} = \mu(\mathbf{v}_I^i), & \forall \mathbf{v}_I^i \in V_I \\ \sum_i X_{ij} = \nu(\mathbf{v}_R^j), & \forall \mathbf{v}_R^j \in V_R, \\ 0 \leq X_{ij} \leq 1 \end{cases} \end{aligned} \quad (4)$$

where  $\mu(\mathbf{v}_I^i)$  and  $\nu(\mathbf{v}_R^j)$  are the masses located at the nodes  $\mathbf{v}_I^i$  and  $\mathbf{v}_R^j$ , respectively. Each value  $C_{ij}$  represents the coefficient of a *cost matrix*  $\mathbf{C}$ , which accounts for the transportation cost to move a portion of mass from  $\mathbf{v}_I^i$  to  $\mathbf{v}_R^j$ . In our approach, Problem (4) is converted to a *Mixed Integer Linear Programming* (MILP) model, by setting the masses as unitary values:

$$\mu(\mathbf{v}_I^i) = \nu(\mathbf{v}_R^j) = 1, \quad \forall i, j.$$

Consequently, all entries of the optimal solution matrix  $\mathbf{X}$  will be binaries, 0 or 1. In our implementations, we use the MILP solver as provided by Vielma [62] to get a solution for the resulting optimization problem.

The set of linear constraints in Problem (4) is imposed to manage the transport of mass between  $G_I$  and  $G_R$ . More precisely, the mass located in  $\mathbf{v}_I^i$  is allowed to be reconducted to some node  $\mathbf{v}_R^j$ , and vice-versa, but assuring the system will preserve the total mass. Indeed, this optimization problem requires that  $V_I$  and  $V_R$  must have the same total mass, and since every single node is always placed as one mass unit, such a condition would imply that  $N_I = N_R$ . In practice, this situation rarely occurs. Another issue is that outliers can be created in the previous processing stages of our framework.

In order to equalize the mass difference between the graphs while still removing possible outliers, two *ghost nodes*  $\mathbf{g}_I \in V_I$  and  $\mathbf{g}_R \in V_R$  are introduced into the optimization process. Without loss of generality, we assume that  $N_I \geq N_R$ , so the portions of mass in the ghost nodes  $\mathbf{g}_I$  and  $\mathbf{g}_R$  are given by  $\alpha$  and  $N_I - N_R + \alpha$ , respectively. Quantity  $N_I - N_R$  regulates the difference of masses between  $G_I$  and  $G_R$ , while  $\alpha$  is a fixed threshold which estimates the number of outliers in the mass transportation model. Mathematically, the inclusion of the ghost nodes in the optimization problem means adding the following constraints into the Problem (4):

$$\sum_{j=1}^{N_R} X_{N_I+1,j} \leq \alpha \quad \text{and} \quad \sum_{i=1}^{N_I} X_{i,N_R+1} \leq N_I - N_R + \alpha, \quad (5)$$

with  $\alpha = \omega \min(N_I, N_R)$ ,  $\omega \in [0, 1]$ . The value  $\omega$  depends on the number of outliers in the graphs, which varies from image to image. Regarding the computation of the parameter  $\omega$ , we test a range of values and select the best estimate according to the gain registration measure recently introduced by Motta et al. [63] (see Section IV-C).

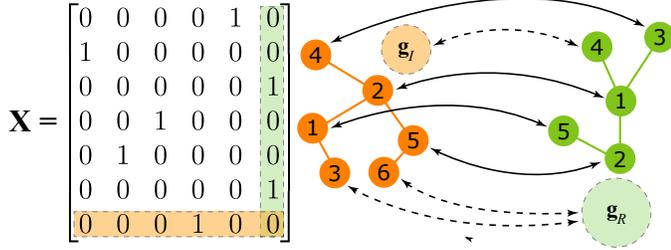


Fig. 4: Optimal graph matching: matrix  $\mathbf{X}$  encodes the assignment between the nodes  $V_I$  (rows/orange) and  $V_R$  (columns/green), including their associated ghost nodes  $\mathbf{g}_I$  (light orange) and  $\mathbf{g}_R$  (light green).

In summary, the coefficients of the optimal transportation plan  $\mathbf{X}$  provide the correspondences between the nodes  $V_I$  and  $V_R$  (see Figure 4). The additional constraints given by Eq. (5) allow us to deal with outliers located at  $G_I$  and  $G_R$ , as they will be attracted to the ghost nodes  $\mathbf{g}_I$  and  $\mathbf{g}_R$  carrying up to  $\alpha$  unities of mass. The transportation cost for the ghost nodes is also required to be established into the model, and for this reason, we chose it as the minimum value of  $\mathbf{C}$ . Once the optimization is performed, all nodes matched with ghost nodes are interpreted as outliers and then discarded, i.e.,  $\mathbf{X}$  is updated by removing the rows and columns related to the ghost nodes. Finally, the set of correspondences of the matched nodes  $\mathbf{v}_I^j$  and  $\mathbf{v}_R^j$  is given by  $\mathcal{M} = \{(i, j) \mid X_{ij} = 1\}$ .

**Cost matrix assignment.** The cost matrix  $\mathbf{C}$  operated into Equation (4) is generated by summing the matrices of *node-to-node dissimilarities*  $\mathbf{D}^k$  and *neighborhood-to-neighborhood dissimilarities*  $\mathbf{S}^k$ , which are built according to a prefixed feature  $k$  of the image. Let  $F$  be the number of image features. The cost matrix  $\mathbf{C}$  is therefore defined as:

$$\mathbf{C} = \sum_{k=1}^F \mathbf{D}^k + \mathbf{S}^k, \quad (6)$$

where the elements of  $\mathbf{D}^k$  and  $\mathbf{S}^k$  are normalized in the range  $[0, 1]$ .

Each coefficient  $D_{ij}^k$  from matrix  $\mathbf{D}^k$  accounts for the dissimilarity between the nodes of  $V_I$  and  $V_R$ . More specifically, given an image feature  $k$ , the cost to transport a unit of mass from the current node  $\mathbf{v}_I^i$  to a second one  $\mathbf{v}_R^j$ , disregarding their neighborhoods, is computed by:

$$D_{ij}^k = \text{diss}_k(\text{desc}_k(\mathbf{v}_I^i), \text{desc}_k(\mathbf{v}_R^j)),$$

where  $\text{desc}_k(\cdot)$  denotes the feature descriptor  $k$  for the pixel associated with its corresponding graph node, while  $\text{diss}_k(\cdot, \cdot)$  represents the dissimilarity measure in the feature space  $k$  (e.g., negative cross-correlation, cosine or Euclidean distance).

We also take advantage of the neighborhoods of each node when building matrix  $\mathbf{S}^k$  as part of the cost matrix  $\mathbf{C}$ , by exploiting the adjacency structures given by the edge sets  $E_I$  and  $E_R$ . Since computing the dissimilarity between two sets is computationally expensive, we propose the Algorithm 1 as a rapid alternative to getting  $\mathbf{S}^k$ . Algorithm 1 is “greedy” and eliminates the use of a more costly and complicated

---

**Algorithm 1:** Dissimilarity between neighborhoods.

---

**Input:**  $V_I$ ,  $V_R$  and an image feature  $k$ .

- 1 **for**  $i \in \{1, \dots, N_I\}$  **do**
- 2     **for**  $j \in \{1, \dots, N_R\}$  **do**
- 3         **for**  $\mathbf{v}_I^m \in \mathcal{N}_I^i$  **do**
- 4             **for**  $\mathbf{v}_R^n \in \mathcal{N}_R^j$  **do**
- 5                  $f_I \leftarrow \text{desc}_k(\mathbf{v}_I^m)$ ;
- 6                  $f_R \leftarrow \text{desc}_k(\mathbf{v}_R^n)$ ;
- 7                  $\mathbf{F}(m, n) \leftarrow \text{diss}_k(f_I, f_R)$ ;
- 8              $D \leftarrow 0$ ,  $r \leftarrow |\mathcal{N}_I^i|$ ,  $c \leftarrow |\mathcal{N}_R^j|$ ;
- 9              $N_{\min} \leftarrow \min(r, c)$ ,  $N_{\max} \leftarrow \max(r, c)$ ;
- 10             **for**  $p \in \{1, \dots, N_{\min}\}$  **do**
- 11                  $[\text{value}, \text{row}, \text{column}] \leftarrow \min(\mathbf{F})$ ;
- 12                  $D \leftarrow D + \text{value}$ ;
- 13                  $\mathbf{F}(\text{row}, 1 : c) \leftarrow \infty$ ;
- 14                  $\mathbf{F}(1 : r, \text{column}) \leftarrow \infty$ ;
- 15              $\mathbf{S}^k(i, j) \leftarrow \frac{D}{N_{\min}} \left(1 + \frac{N_{\max} - N_{\min}}{N_{\max}}\right)$ ;
- 16 **return**  $\mathbf{S}^k$

---

procedure to measure the dissimilarity between sets. The resulting dissimilarity (line 15) is normalized by dividing the cumulative dissimilarity  $D$  by  $N_{\min}$ , while neighborhoods  $\mathcal{N}_I^i$  and  $\mathcal{N}_R^j$  with different cardinalities are penalized by the factor  $(N_{\max} - N_{\min})/N_{\max}$ , where  $N_{\min} = \min\{|\mathcal{N}_I^i|, |\mathcal{N}_R^j|\}$  and  $N_{\max} = \max\{|\mathcal{N}_I^i|, |\mathcal{N}_R^j|\}$ .

Next, we describe in details the features and the dissimilarity measures used to build the matrix  $\mathbf{D}^k$  in Equation (6).

**Feature extraction.** We now provide the different types of image features employed to achieve a context-relevant cost matrix. All features are computed regarding the input retinal images and collected at the nodes or from image patches around these nodes. In our experiments, we set patches of size  $21 \times 21$  pixels centered at each evaluated node.

*Gray-image patches* are directly collected varying a pre-defined template over the gray-scaled version of the images. Therefore,  $\text{desc}_1(\mathbf{v})$  is a “vected patch” with pixel intensities around node  $\mathbf{v}$  on the gray-scaled fundus image.

*Gabor filter patches* are computed to capture texture information and time-frequency location from the given images. It provides robustness to our system even when varying the brightness/contrast content of the images as well. In our experiments, we set four Gabor filter-based features, ranging from  $0^\circ$  to  $90^\circ$  with a wavelength of 4 and 8 pixels/cycle. Hence,  $\text{desc}_2(\mathbf{v})$  returns four patches related to  $\mathbf{v}$  on the Gabor filtered fundus image. For a detailed survey of Gabor filters, we refer the reader to [64].

*Saliency-based vector* is a novel, scale and translation invariant feature suited for fundus images. In order to define the most salient region, we modify the saliency detection algorithm proposed by Zhu et al. [65] for the context of fundus image registration. After detecting the most salient region on the green channel of the image, a representative vector encoding the saliency cues of a given node  $\mathbf{v}$  is derived

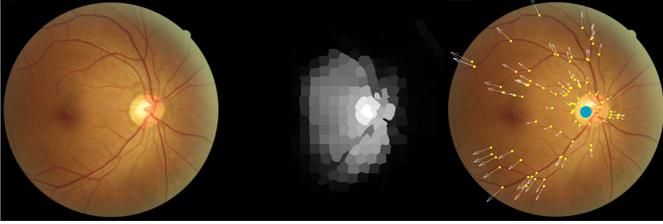


Fig. 5: Saliency-based vector field (from left to right): input image, the most salient region (gray) and the vector field (yellow) builds in terms of the ONH (blue) and graph nodes.

by  $\mathbf{u}_v = \mathbf{v} - \mathbf{o}$ , where  $\mathbf{o}$  denotes the centroid of this region. In other words, a vector field is defined by the centroid of the most salient region and the graph nodes of the corresponding image, as shown in Figure 5.

The most salient region tends to be located at the *Optic Nerve Head* (ONH) [66], [67]. Thus,  $\text{desc}_3(\mathbf{v})$  returns a two-dimensional vector with origin in  $\mathbf{v}$ . Once the vector field is obtained, the saliency-based features are estimated by using dissimilarity measures, such as the negative cross-correlation for the patch-based features (gray-images and Gabor filter) and the cosine distance.

**OT-based outlier removal filtering.** Aiming at eliminating possible outliers originated during the previous steps, we adapt the method proposed by McIlroy et al. [68], the so-called DeSAC (*Deterministic Sample And Consensus*), for our specific application. Since DeSAC requires a matching probability  $P_{ij}$  as input for each correspondence  $(i, j) \in \mathcal{M}$ , here this value is computed by a convex combination and then parameterized regarding the coefficients of our OT cost matrix  $\mathbf{C}$ , i.e.:

$$P_{ij} = 0 \cdot t_{ij} + 1 \cdot (1 - t_{ij}) = 1 - t_{ij} \quad \text{with} \quad t_{ij} = \frac{C_{ij} - m}{M - m},$$

where  $M$  and  $m$  account for the maximum and minimum values of the OT cost set  $\{C_{ij} \mid (i, j) \in \mathcal{M}\}$ . Notice that, a low transportation cost  $C_{ij}$  leads to a high matching probability  $P_{ij}$ , and vice-versa. Figure 6 shows the result of our improved DeSAC to a pair of fundus images.

### C. Fundus registration

Once the outliers have been removed, the image registration is finally performed through the scheme proposed by Motta et al. [63], called *Gain Coefficient* (GC). Firstly, from the corresponding matched keypoints  $\mathbf{v}_I^i$  and  $\mathbf{v}_R^j$ , GC computes a set of standard transformations between the binary vessel images, i.e.,  $T: B_I \mapsto B_R$  such that  $T(\mathbf{v}_I^i) \approx \mathbf{v}_R^j, \forall (i, j) \in \mathcal{M}$ . The candidates for  $T$  are similarity, affine, projective and quadratic models. Notice that both deformable and non-deformable mappings can be chosen by the GC decision rule, including the quadratic one, which has been widely employed in many deformable registration tasks [21], [47], [69], [51]. The best geometric transformation given by GC is the one that maximizes the following measure [63]:

$$GC(B_I, B_R, T) = \frac{|B_R \cap T(B_I)|}{|B_R \cap B_I|}. \quad (7)$$

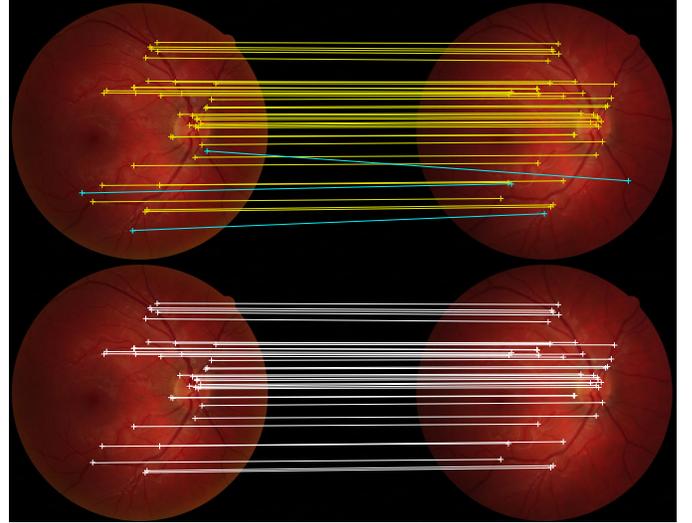


Fig. 6: Outlier filtering: before outlier removal (outliers in cyan), and after with the improved DeSAC filtering.

Equation (7) measures the amount of pixel aligned after the mapping by the geometric transformation  $T$ . If more pixels are overlapped after applying  $T$  than the trivial overlap, then GC will be higher than 1.

## V. RESULTS, COMPARISONS AND EVALUATION

In this section, we present an extensive set of comparisons involving our framework against several state-of-the-art methods traditionally used in the context of fundus image registration. In the following, we first detail the measures and datasets used to conduct the experiments.

### A. Registration error

In order to quantitatively assess the quality of the registration results, we employ a well-established registration error measure which requires the corresponding ground truth points for each image pair.

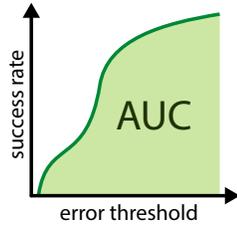
**Mean Registration Error (MRE)** [48] computes the error between the available prescribed points (ground truth) in image  $R$  and the corresponding ones in the transformed image  $T(I)$ . More precisely, given the sets of reference points  $\mathbf{Y}_I = \{\mathbf{y}_I^1, \dots, \mathbf{y}_I^N\} \subset \mathbb{R}^2$  and  $\mathbf{Y}_R = \{\mathbf{y}_R^1, \dots, \mathbf{y}_R^N\} \subset \mathbb{R}^2$  initially marked as corresponding points in  $I$  and  $R$ , we calculate the MRE as follows:

$$MRE(\mathbf{Y}_I, \mathbf{Y}_R, T) = \frac{1}{N} \sum_{i=1}^N \|\mathbf{y}_R^i - T(\mathbf{y}_I^i)\|_2, \quad (8)$$

where  $\|\cdot\|_2$  denotes the Euclidean norm. Since MRE is an error measure, values closer to 0 indicate more accurate results. Our choice by MRE instead of the *Root Mean Square Error* (RMSE) is due to its robustness against outliers, which is not the case for RMSE, as observed by Pontius et al. [70].

### MRE's Area Under the Curve (AUC) [49]

provides a useful index to summarize the registration success for a given dataset. Despite the versatility of MRE, computing the mean and the standard deviation of entire sets of registered image pairs may lead to biased results, as any of the evaluated methods can fail when registering some image pairs, giving no output or resulting in images projected far from the starting pose. In order to avoid these issues, an alternative is to plot the curve of *success rate* (the quotient between the number of successfully registered pairs by the total number of pairs) as a function of target accuracy (with a threshold error varying from 0 to 25, similar to [49], [71], [63]). The abscissa gives an error threshold, and the ordinate, the success rate for MRE lower than the corresponding threshold. Next, the AUC is computed by the trapezoidal rule and scaled to a standard reference range of [0, 100]. Higher scores lead to better results.



**Median Error (MEE) and the Maximal Error (MAE)** over all distances  $\|y_R^i - T(y_I^i)\|_2$  also provide a reliable criterion to evaluate the performance of retinal image registration methods [34], [42]. As suggested by Wang et al. [51], we can combine MEE and MAE so that the following classification is derived: *acceptable* ( $MAE \leq 20$  pixels and  $MEE \leq 3.5$  pixels), *inaccurate* ( $MAE \leq 20$  pixels and  $MEE > 3.5$  pixels), and *incorrect* ( $MAE > 20$  pixels). In this case, the success rate is defined by the percentage of image pairs for both acceptable (AR) and inaccurate (IR) results, i.e., alignments without meaningful errors.

### B. Datasets

We run all the evaluated techniques in three fundus image datasets with retinal pairs differing from each other in terms of geometric transformations as well as rendering aspects such as noise occlusion and low image quality. Involuntary movements of the patient eyes can generate translation, rotation and scale variability during the acquisition of the images. So, among the evaluated images, some of them have low sharpness caused by lack of focus while a full set of image pairs presents anatomical changes, a behavior typically found in patients with some degenerative disease. Therefore, all the datasets aim at reflecting practical situations as described above.

**FIRE dataset** comprises 134 pairs of digital retinal images made publicly available by Hernandez-Matas et al. [71]. FIRE is composed of three major classes of images: A (14 image pairs), P (49 image pairs) and S (71 image pairs). Category A contains images that present anatomical changes caused by degenerative eye conditions while Category S holds the opposite situation. Finally, Category P covers the case wherein the pairs are recorded as mosaic representations of the eyes. To be more precise, mosaics are sets of images used to provide a combined view of the retina by applying image stitching on different fundus images from the same patient; hence, these images carry low overlap. FIRE also provides ground truths

and their corresponding points annotated by a specialist for each image pair in the dataset.

**Dataset 1** contains 18 pairs of high-resolution images available in a public repository provided by Kohler et al. [72]. This dataset presents twin-like images extracted from the same patient so that the images vary by an RST (Rotation-Scale-Translation) and a large amount of noise. Moreover, each retinal pair is composed of a good quality image with its counterpart impaired by specular noise or decreased sharpness.

**Dataset 2** is formed by 85 image pairs collected by an ophthalmologist during his preventive exams for different patients. It covers real situations where undesirable artifacts were created during the capture of the images. This dataset reveals an essential overview of real cases faced by ophthalmologists and other eye specialists in their daily routine, so it is expected that any registration algorithm useful in practice should be able to cope with cases gathered in this dataset.

Finally, to compute the registration error from a given set of ground truth points, we follow the FIRE protocol [71] towards delivering ten reference points and their correspondences for each pair of images for both Datasets 1 and 2. In order to validate the ground truth, we invited two experienced ophthalmologists (both with more than 20 years of experience) to guarantee the reliability of the marked points. These points have been carefully annotated, corrected, and finally approved by these experts to ensure high accuracy and validity of the correspondences. Figure 7 shows the ground truth points chosen on an exemplar fundus image pair. These points are placed on vessel bifurcations and around of ONH, widespread them in both images to improve the measure of the overlapping between the images.

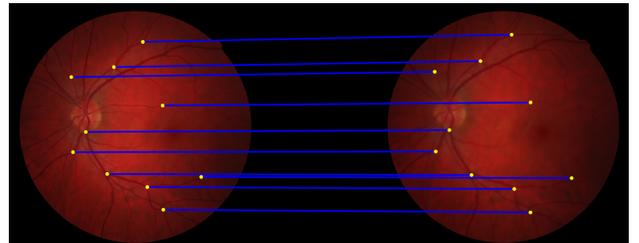


Fig. 7: The marked ground truth points (yellow dots) and their matchings (blue lines) in an image pair from Dataset 1.

### C. Ablation study

To better point out the individual contributions of this work while still validating the overall efficacy of our framework, we first compare our pipeline considering separate modules using the AUC and AR scores.

We start by analyzing our approach, VOTUS, against four relevant state-of-the-art methods for graph matching when they are tested in the context of fundus image registration. The implementations of these methods were provided by Zhou and De la Torre [54] and includes: *Graduated Assignment* (GA) [73], *Spectral Matching with Affine Constraints* (SMAC) [74], *Re-weighted Random Walk Matching* (RRWM) [75] and *Factorized Graph Matching* (FGM) [54]. In this experiment, we

only modify the graph matching scheme into our pipeline (Section IV-B), replacing our OT-derived modulus (OT-GM) by each one of the graph matching methods under evaluation. This means that the same graphs have been used as input for all methods, besides preserving the optimal parameters as given in the original codes. From the results reported in Table I, we can observe that our optimization scheme delivers the best scores in all datasets, especially on images from FIRE and Dataset 1, which covers the presence of noise from the structural changes caused by eye diseases (FIRE-A and Dataset 1) and a large number of graphs with topological changes (FIRE-P). Note that the AR scores highly pronounce the difference between the variations of our pipeline in these datasets. It demonstrates that the node assignment provided by OT-GM does not have a strong dependence with the graph topology. For the well-behaved datasets (i.e., collected from healthy patients) FIRE-S and Dataset 2, all flavors of our pipeline work satisfactorily as expected.

TABLE I: Comparative results w.r.t. AUC and AR (IR in parentheses) for different graph matching models when they are run as part of our pipeline. Best results in bold.

Methods	FIRE Dataset			Datasets		
	A	P	S	1	2	
AUC	GA	63.42	48.85	92.08	<b>78.66</b>	92.91
	SMAC	64.28	58.08	92.25	78.44	92.89
	RRWM	63.71	54.00	92.42	78.00	92.56
	FGM	60.85	54.53	92.59	78.44	93.24
	our OT-GM	<b>68.14</b>	<b>67.18</b>	<b>93.43</b>	<b>78.66</b>	<b>94.56</b>
AR (IR)	GA	14.3 (64.3)	16.3 (20.4)	95.8 (4.20)	27.8 (66.7)	96.5 (0.00)
	SMAC	14.3 (57.1)	30.6 (24.5)	95.8 (4.20)	27.8 (66.7)	96.5 (1.20)
	RRWM	14.3 (57.1)	18.4 (24.5)	<b>97.2 (2.80)</b>	27.8 (66.7)	96.5 (0.00)
	FGM	21.4 (50.0)	22.5 (24.5)	95.8 (4.20)	22.2 (72.2)	97.7 (0.00)
	our OT-GM	<b>42.9 (35.7)</b>	<b>51.0 (14.3)</b>	<b>97.2 (2.80)</b>	<b>100 (0.00)</b>	<b>100 (0.00)</b>

Next, we analyze the accuracy of the proposed outlier removal scheme, the improved DeSAC, against RANSAC [45], and the robust outlier removal filter *Weighted Graph Transformation Matching* (WGTM) [76], keeping the other modules of our pipeline intact. The key idea of the WGTM relies on using an angular distance between the  $K$ -nearest-neighbor ( $K$ -NN) features. In particular, we use the best parameters obtained for WGTM ( $K = 20$ , and stopping threshold  $\varepsilon = 10^{-3}$ ). From Table II, we can attest that our improved DeSAC achieves the best scores in all the measurements in comparison with RANSAC and WGTM, besides having the benefit of being fully deterministic. Therefore, the current outlier removal scheme is essential to enhance the registration quality, noticing that our DeSAC improves the results significantly on images with the presence of noise, providing AR scores three times higher in FIRE-A and four times higher in Dataset 1.

TABLE II: Comparative results w.r.t. AUC and AR (IR in parentheses) for different strategies of outlier removal.

Methods	FIRE Dataset			Datasets		
	A	P	S	1	2	
AUC	RANSAC	63.29	47.95	92.37	78.59	91.67
	WGTM	52.42	63.63	93.04	78.00	94.47
	our DeSAC	<b>68.14</b>	<b>67.18</b>	<b>93.43</b>	<b>78.66</b>	<b>94.56</b>
AR (IR)	RANSAC	14.3 (64.3)	42.9 (26.5)	<b>97.2 (2.80)</b>	27.8 (66.7)	98.8 (0.00)
	WGTM	14.3 (28.6)	34.7 (28.6)	<b>97.2 (1.40)</b>	22.2 (72.2)	<b>100 (0.00)</b>
	our DeSAC	<b>42.9 (35.7)</b>	<b>51.0 (14.3)</b>	<b>97.2 (2.80)</b>	<b>100 (0.00)</b>	<b>100 (0.00)</b>

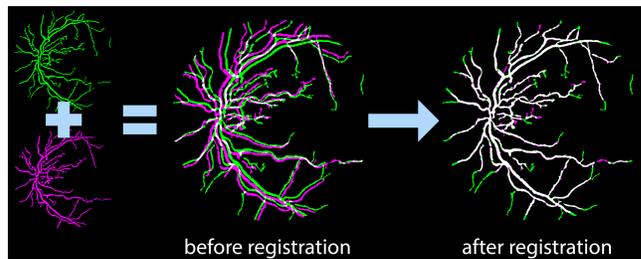


Fig. 8: Color labeling-based inspection: the composition of two binary vessel images  $B_I$  (green) and  $B_R$  (magenta) are used to drive the visual analysis. A large amount of white pixels means better registration outcomes.

#### D. Comparisons with existing image registration methods

We compare VOTUS against ten existing image registration methods on all datasets (Section V-B), ranging from well-known keypoint-based methods such as SIFT and SURF when they are equipped with accurate outlier removal filters such as WGTM, to well-established and modern techniques strictly designed to accomplish fundus image registration.

The implementations of most of these methods are available online or were provided by the authors on request. For each method, we keep their best parameter settings as stated by the authors in their papers or as specified in the reference codes. For SIFT and SURF, the point correspondences are determined by a brute force scheme (i.e., computing the pairwise distance between all feature vectors) as coded by MATLAB's function `matchFeatures` [59], while the registration is performed identically as in our pipeline.

Table III gives more details of the registration methods and implementations used in our qualitative/quantitative analysis and the statistical validation of the results as well.

TABLE III: Image registration methods.

Method	Authors/Year	Ref.	Impl.
SIFT	Lowe, 2004	[39]	[77]
SURF	Bay et al., 2006	[43]	[59]
GDB-ICP	Yang et al., 2007	[38]	[78]
Harris-PIIFD	Chen et al., 2010	[34]	*
ED-DB-ICP	Tsai et al., 2010	[41]	*
RIR-BS	Chen et al., 2011	[22]	[79]
ATS-RGM	Serradell et al., 2015	[26]	[26]
H-M'17	Hernandez-Matas et al., 2017	[49]	[71]
EyeSLAM	Braun et al., 2018	[28]	[28]
GFEMR	Wang et al., 2019	[51]	[51]

\* Implementation provided by the authors (on request).

**Qualitative analysis.** We start our discussion showing how visual inspections can be properly conducted when viewing the registration results. Figure 8 illustrates the process of inspecting the matching between the fundus images, by counting the amount of overlapped pixels in the registered images. More specifically, we decompose into RGB channels the binary vessels of the fundus images  $B_I$  and  $B_R$  so that the green component  $(0, 1, 0)$  is taken from  $B_I$ , while its counterpart, the magenta color  $(1, 0, 1)$ , is designated to the binary image  $B_R$ . These cropped components are then overlaid so that the resulting image will have white pixels  $(1, 1, 1) = (0, 1, 0) + (1, 0, 1)$ , denoting a perfect matching, and green/magenta colors, otherwise.

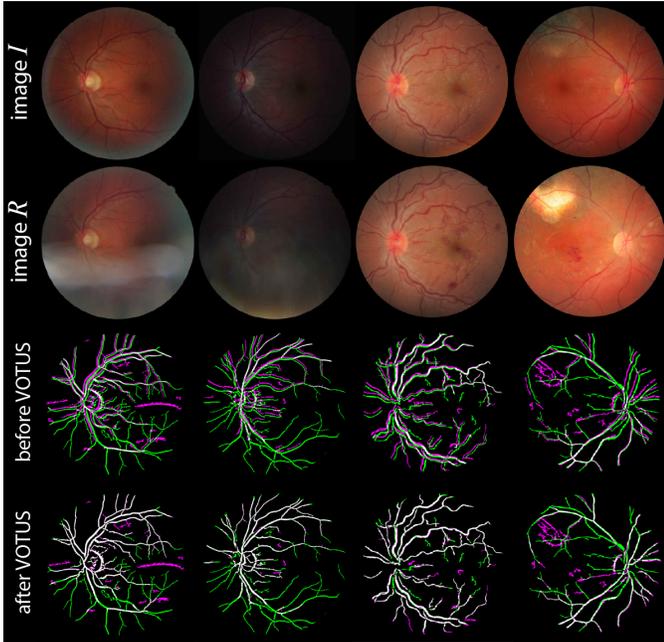


Fig. 9: Visual results generated by VOTUS when applied to arbitrary images with different losses of quality from Dataset 1 (first two columns), while the remaining ones ticked with anatomical changes were taken from FIRE (last two columns).

First, the flexibility of VOTUS to cope with different kind of images are illustrated in Figure 9. Notice from the results that our approach accurately assigns a significant number of vessels, attesting its accuracy for the cases where the primary and the secondary images differ from each other as to color variation, the eminence of artifacts and presence of geometric deformations. Even for the more severe situations, the method aligns a substantial portion of the vessel structures. Considering the visual inspection for the particular cases representing diseases, pathogenic disorders or even unexpected structural changes on the retinal blood vessels, VOTUS provides a useful perception of the eye fundus alignment as seen by the two rightmost examples. For instance, the number of green veins located in the northwest region in the last sequence of eye images could be successfully used to detect unexpected vessels changes, as a portion of veins has disappeared from the secondary validated image.

As verified by Ghassabi et al. [27], an essential aspect that registration techniques must address is their ability in handling high-resolution images. For instance, by examining a certain image pair in such a way that the resulting alignment is inspected at its original scale, one could track the progress of eye diseases such as *diabetic retinopathy*<sup>1</sup>. Other benefits are the use of compatible and more advanced technologies to store and retrieve the results and powerful visual exploratory systems to accurately audit the images. Bearing this in mind, Figure 10 brings qualitative comparisons of VOTUS against seven registration methods over four high-resolution images taken from Dataset 1. EyeSLAM, GDB-ICP, ED-DB-ICP, and

<sup>1</sup>Diabetic retinopathy is an eye disease that occurs as a complication of diabetes. It can eventually lead to blindness [80].

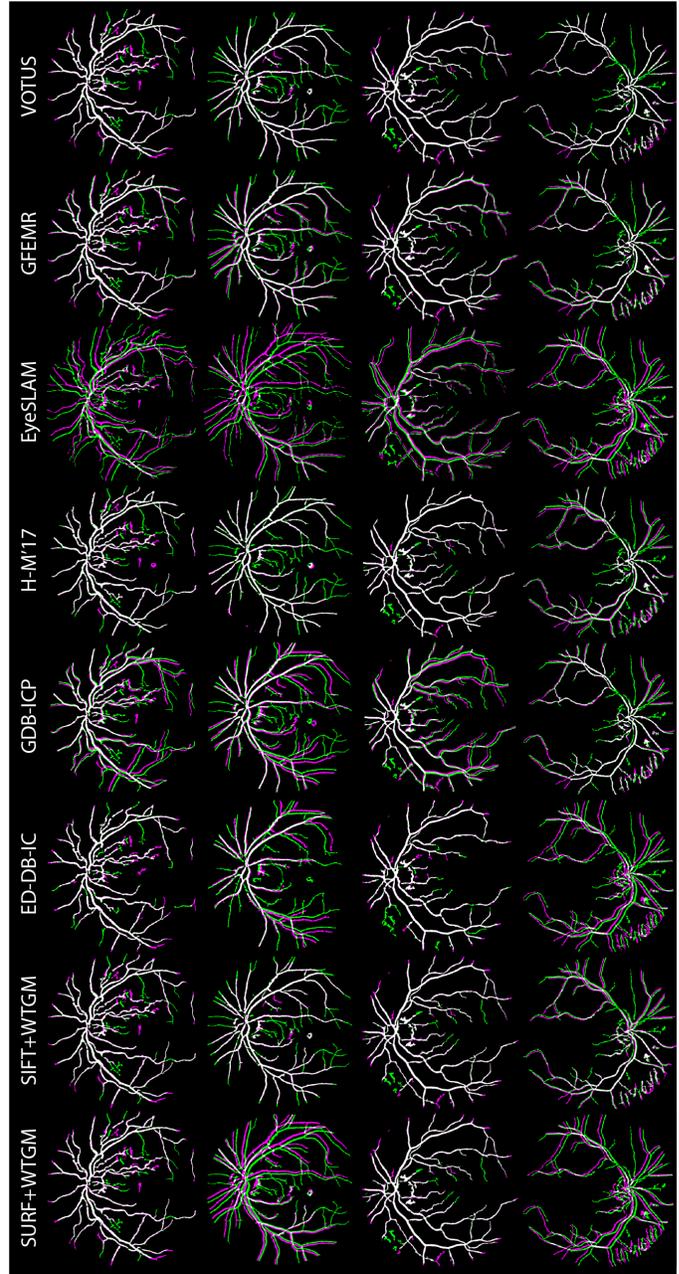


Fig. 10: Qualitative comparisons in Dataset 1.

SURF+WTGM produce visually acceptable results, but they carry some miss-alignments (noticed by magenta and green pixels rather than white). Also, most competitors generate misfitted alignments at least on the first-row image for which the original image had some focus noise. In contrast, VOTUS achieves more consistent and pleasant results in all cases, mainly regarding the quality of matching refinement as shown by a large amount of white color in the resulting montages.

We also quantify the total number of white pixels on the full high-resolution dataset, by computing the GC (7). Figure 11 shows the boxplots of the GC scores for the methods on the complete Dataset 1. We can observe from the results that VOTUS outperforms the evaluated methods. However, one may notice that although GC provides a reasonable estimate

for the registration quality, it is not precisely a quality metric in the sense of other measures as MRE, AUC, MAE and MEE. In the next section, we compare the methods in light of more prescriptive and accurate measures.

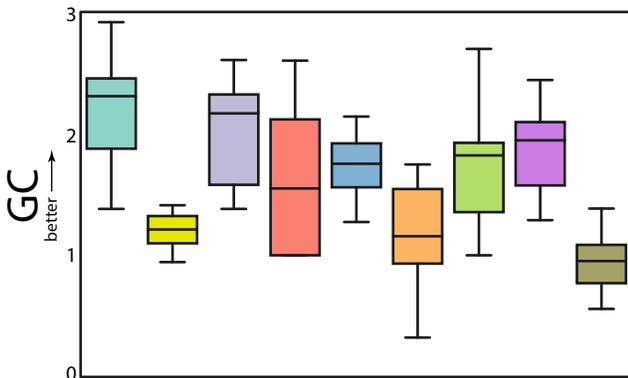


Fig. 11: Boxplots of the GC scores for the methods evaluated on Dataset 1: VOTUS (teal), H-M'17 (yellow), GDB-ICP (purple), ED-DB-ICP (red), SIFT+WGTM (blue), SURF+WGTM (orange), RIR-BS (green), GFEMR (purple) and EyeSLAM (brown).

**Quantitative analysis.** Table IV and Table V depict a full set of comparisons of our approach versus the state-of-the-art registration methods as detailed in Table III on all datasets considered. For a fair comparison among the non-deterministic methods such as GDB-ICP, ATS-RGM, and H-M'17, mimicking the analysis of [76], we run them three times and take the mean average of the errors.

TABLE IV: Comparative results of VOTUS against other image registration methods w.r.t. AUC. The best scores and the second best are shown in bold and italic, respectively.

Methods	FIRE Dataset			Datasets	
	A	P	S	1	2
VOTUS	<b>68.14</b>	<b>67.18</b>	<b>93.43</b>	<b>78.66</b>	<b>94.56</b>
H-M'17	56.86	50.29	93.38	68.56	94.09
GFEMR	47.43	60.69	81.21	54.22	94.00
EyeSLAM	26.85	22.40	30.76	03.11	30.76
GDB-ICP	29.57	30.24	79.77	23.00	89.55
Harris-PIIFD	43.57	22.00	86.54	72.44	92.73
ED-DB-ICP	49.71	44.12	60.39	43.00	62.80
SIFT+WGTM	40.71	54.40	83.69	67.33	90.63
SURF+WGTM	06.85	06.08	83.54	52.00	72.23
RIR-BS	12.42	04.93	77.21	49.55	68.00
ATS-RGM	14.71	00.00	36.85	00.00	25.17

As can be seen, VOTUS leads to higher AUC (Table IV) and AR (Table V) scores, demonstrating its accuracy and robustness in a substantial number of cases. Furthermore, VOTUS is stable for both measures, exhibiting excellent results independently of the criteria adopted for the success rate. Such a solid performance is more apparent when one checks the results from Dataset 1, which gathers noisy images, and also Category P of FIRE, which includes mosaic images wherein significant differences in quality occur due to the combination of our OT-based graph matching and the improved DeSAC algorithm. Looking at the results of Category A of FIRE (i.e., images with anatomical changes), we can perceive that

TABLE V: Success rate of the retinal image registration methods w.r.t. AR (IR in parentheses). The best AR results and the second best are shown in bold and italic, respectively.

Methods	FIRE Dataset			Datasets	
	A	P	S	1	2
	AR (IR)	AR (IR)	AR (IR)	AR (IR)	AR (IR)
VOTUS	<b>42.9</b> (35.7)	<b>51.0</b> (14.3)	<b>97.2</b> (2.80)	<b>100</b> (0.00)	<b>100</b> (0.00)
H-M'17	35.7 (14.3)	0.00 (8.20)	94.4 (5.60)	16.7 (72.2)	98.8 (1.20)
GFEMR	7.10 (28.6)	20.4 (36.7)	59.2 (28.2)	0.00 (44.4)	<b>100</b> (0.00)
EyeSLAM	0.00 (7.10)	0.00 (18.4)	0.00 (21.1)	0.00 (0.00)	4.71 (69.4)
GDB-ICP	21.4 (0.00)	30.6 (8.20)	49.3 (28.2)	0.00 (16.7)	5.88 (32.9)
Harris-PIIFD	21.4 (14.3)	0.00 (14.3)	81.7 (8.50)	11.1 (61.1)	94.1 (1.20)
ED-DB-ICP	14.3 (28.6)	49.0 (0.00)	36.6 (9.90)	11.1 (33.3)	4.70 (37.7)
SIFT+WGTM	21.4 (21.4)	30.6 (22.5)	73.2 (19.7)	16.7 (72.2)	90.6 (3.50)
SURF+WGTM	0.00 (7.10)	0.00 (2.00)	74.7 (15.5)	27.8 (33.3)	68.2 (9.40)
RIR-BS	0.00 (0.00)	0.00 (0.00)	49.3 (38.0)	0.00 (33.3)	60.0 (10.6)
ATS-RGM	0.00 (7.10)	0.00 (0.00)	7.00 (21.1)	0.00 (0.00)	0.00 (0.00)

VOTUS also yields the best scores, as well as for image pairs of Category S (image pairs without anatomical changes), where VOTUS slightly outperforms H-M'17, the second-best approach for this set of images. Finally, for Dataset 2, which is formed by image pairs of healthy patients (i.e., without lesions in the retina), VOTUS yields very high scores as well as H-M'17 and GFEMR for the considered measures.

Analyzing the results in another viewpoint, one could consider, for instance, that the H-M'17, the second best method according to Table IV, is not precisely deterministic. Indeed, H-M'17 sets a combination of non-deterministic models, such as particle swarm optimizations and RANSAC that are applied many times before reaching the best score w.r.t. its minimization score function. To better explain this, we run three times the H-M'17 code in a given fundus pair, obtaining an average score of 21.03 for the MRE with a standard deviation of 8.01 (see Figure 12). Notice that, in a few cases, H-M'17 produces suitable results (see Figure 12(b)), while in others, low-quality outputs (see Figure 12(d)). Since VOTUS is fully deterministic and can be applied under severe noise conditions satisfactorily, these aspects are also reflected in the results as a unique output as seen in Figure 12(a).

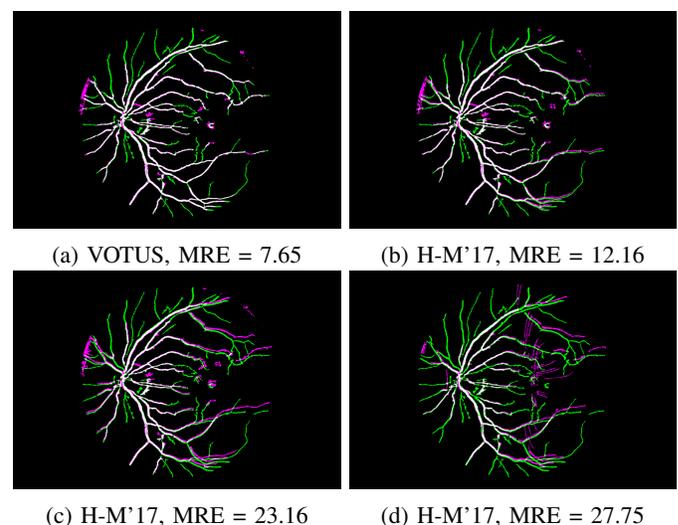


Fig. 12: MRE for a sample of fundus image pair.

From Table IV, we again select the two most accurate methods and reevaluate their general performance by collecting the MRE scores for all datasets, as displayed in Figure 13. From the obtained boxplots, one can confirm that VOTUS leads to more compact interquartile ranges, besides avoiding large uncertainties (see the upper whiskers in the boxplots).

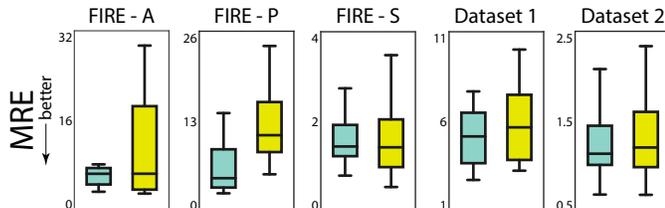


Fig. 13: Boxplots of the MRE scores for VOTUS (■) and H-M'17 (■) in all datasets.

**Statistical validation.** In order to verify whether the results afforded by VOTUS are statistically significant compared to other methods, we apply a Student's t-test using 95% confidence levels. Our null hypothesis is that  $H_0: \bar{V}_\varepsilon \leq \bar{O}_\varepsilon$ , where  $\bar{V}_\varepsilon$  is the arithmetic mean of the number of registered images by VOTUS under a threshold error  $\varepsilon$  varying from 0 to 25 as considered in the AUC index, while  $\bar{O}_\varepsilon$  accounts for the mean of a pre-specified registration method. Low  $p$ -values ( $p < 0.05$ ) indicate that the results obtained by VOTUS are statistically more accurate than the ones produced by other approaches. Particularly, we reject  $H_0$  and state that our method statistically registers more images (on average) than any other method for a given dataset if the corresponding  $p$ -value is low. Table VI presents the  $p$ -values computed per dataset, while the last column accumulates the number of times that  $H_0$  was rejected. Notice from the results that VOTUS is statistically superior in at least three datasets in comparison to each method. Moreover, we found that the results are statically significant in 43 of the 50 tests conducted, also performing well in most of the remaining cases.

TABLE VI: The  $p$ -values of the hypothesis tests conducted regarding the results listed in Table IV.

Methods	FIRE Dataset			Datasets		#Rej. $H_0$
	A	P	S	1	2	
H-M'17	0.005	0.000	0.509	0.017	0.437	3
GFEMR	0.000	0.063	0.002	0.000	0.410	3
EyeSLAM	0.000	0.000	0.000	0.000	0.000	5
GDB-ICP	0.000	0.000	0.000	0.000	0.049	5
Harris-PIIFD	0.000	0.000	0.021	0.110	0.284	3
ED-DB-ICP	0.000	0.000	0.000	0.000	0.000	5
SIFT+WGTM	0.000	0.000	0.002	0.013	0.099	4
SURF+WGTM	0.000	0.000	0.002	0.000	0.000	5
RIR-BS	0.000	0.000	0.000	0.000	0.000	5
ATS-RGM	0.000	0.000	0.000	0.000	0.000	5

## VI. COMPUTATIONAL ASPECTS

Aiming at checking the computational performance of the methods when dealing with real cases as properly faced by eye care specialists in their routine daily inspections, we tabulate in Table VII the average timings for various registration methods run on Dataset 2. One can observe that our approach is considerably faster than H-M'17 and slower than other baseline

methods. More specifically, it is 25 times faster than the competitor with the best quality accuracy, H-M'17, using CPU, and almost two times faster than the time reported in [49] for the tuned H-M'17 when it is run on GPGPU, noticing that our method could still get a lot faster by setting parallelism as well. The methods EyeSLAM, GFEMR, GDB-ICP, ED-DB-ICP, and Harris-PIIFD are faster than our framework, but one may note that, first, they tend to generate less accurate results under real-world circumstances and, second, we assumed for these experiments the best hypothetical situation for which only a RANSAC iteration is run to speed up the methods.

TABLE VII: Computational times (in seconds) and the programming language used in the implementation of each method.

Methods	Avg. Time	Prog. Lang.
VOTUS	106	MATLAB
Harris-PIIFD	13	MATLAB
GFEMR	10	MATLAB
H-M'17	2674	C/C++
ED-DB-ICP	44	C/C++
GDB-ICP	19	C/C++
EyeSLAM	7	C/C++

Despite the accuracy of VOTUS comes with the price of solving the optimizations required, the average time presented is feasible even being run along with a regular physician consultation, fully attending the *ophthalmologic protocol*<sup>2</sup>.

In our experiments, we run our non-tuned MATLAB-built code against the optimized versions implemented in C/C++ or GPGPU of existing methods for fundus image registration. More specifically, our prototype was implemented in MATLAB R2015a on a laptop equipped with Intel Core i7 processor of 2.40GHz with 8Gb of RAM. We believe that the execution time of our method can decrease significantly via MEX-C, parallel processing, or GPGPU architectures as well.

## VII. FEEDBACK FROM MEDICAL EXPERTS

We have asked for two experienced ophthalmologists (both with more than 20 years of experience, working with eye disease diagnostics and monitoring) their feedbacks regarding the registration results obtained by VOTUS for both Dataset 1 and Dataset 2, i.e., the “smaller” sets of images examined in this work.

In terms of clinical benefits, they explained that the designed method might assist the identification of specific anatomical changes in lesions present in the retina, allowing the patient to start the treatment early in order to avoid damage to his/her eyesight. So, a potential application is related to a preventive protocol for identification and monitoring anatomical changes that can be judged by ophthalmologists as a probable lesion. Notice that the goal here is not to provide a comprehensive analysis of existing methods by an expert's judgment or perform a clinical control group study, which would be time-consuming and very laborious for the invited specialists, but

<sup>2</sup>Ophthalmologists say they usually spend an average of 15-20 minutes per patient during eye exams [81], [82].

instead to estimate the impact reached by VOTUS in the day-to-day practice of the physicians. After operating with VOTUS and examining the results, ophthalmologists have sent us the following feedback:

**Ophthalmologist 1:** *“I was invited to know the project in question, actually to evaluate its practical utility. After knowing the project and seeing the presented results, I can state that it appears as a valuable tool, for detection and evaluation of disease progression, and in the execution’s accuracy of specific treatments of pathologies that affect the retina. In ophthalmology, we work with parameters ranging from millimeters to micrometers, with very restricted error margins, so any tool that helps in the measurement accuracy and location of the pathologies can be extremely useful in the therapeutic success against these pathologies. I hope we can include this registration tool in our propaedeutic arsenal as soon as it is available.”*

**Ophthalmologist 2:** *“I was impressed with these results. This software allows for much more precise monitoring of subsequent examinations, making possible a more faithful comparison between size/progression of lesions and ocular pathologies, hence providing more reliability for the indication of treatments and evaluation of the evolution of diseases.”*

### VIII. DISCUSSION AND LIMITATIONS

The comparisons presented in Section V has demonstrated the accuracy and effectiveness of VOTUS in dealing with a wide variety of retinal fundus images on representative medical datasets. The technique is sensitive to color variations, being also capable of generating pleasant results regarding many image quality aspects such as lack of focus, illumination distortion, and blood vessel motion — a particular and desirable behavior not always present in most existing image registration methods. The alignments obtained from our framework turned out to be clean, well assembled, and easy to check visually. Regarding the mathematical facet of our method, the designed OT scheme has proven to be robust and flexible, providing accurate matchings while still guaranteeing a unique solution to the associated mathematical problem.

Despite its consistency, versatility and solid mathematical foundation, there are two aspects that must be considered when applying the method on very deteriorated images (e.g., see Figure 14). First, the registration performance of the retinal images may be affected in the absence of a salient region to be taken as a referential, such as the ONH. VOTUS assumes this pattern as a natural feature to build the cost matrix, and in the extreme cases without ONH, the method may fail. However, it is worth mentioning that, in some pragmatic cases, VOTUS has successfully registered images even without the ONH data. The second usability aspect is related to the non-availability of the retinal blood vessels to generate the graph, and such essential missing data could hamper the direct use of our method in these images.

### IX. CONCLUDING REMARKS AND FUTURE WORK

In this paper, we introduced VOTUS, an automated and flexible technique based on optimal mass transport theory to

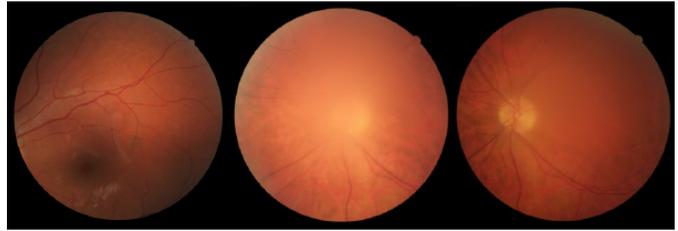


Fig. 14: Limitation of VOTUS: retinal images without ONH as reference, and missing retinal veins.

tackle the problem of fundus image registration. The designed framework yields high-accuracy registrations including many cases difficult to be handled in real circumstances such as potential changes in the geometry of the vessel structures, lack of focus, and the presence of specular noise. Moreover, we also verified the effectiveness of each core modulus of VOTUS separately, by analyzing them individually against existing well-established approaches. As demonstrated by a comprehensive set of experiments, VOTUS has produced highly accurate results even when compared against baseline as well as modern registration methods. All those properties render VOTUS a useful and compelling retinal registration technique, begin also capable of aiding physicians and ophthalmologists under real circumstances, as raised by experienced specialists.

As future work, we are currently adapting our approach to addressing multimodal fundus images registration [40], [42], [83]. Originally, VOTUS was not designed to deal with multimodal images, however, competitive results have already been achieved just by setting a minor modification into our pipeline. This particular change simply consists in providing as input to the feature extraction step of our framework the gradients of both color and grayscale multimodal images instead of the raw images. Figure 15 shows some preliminary results involving patients diagnosed with diabetic retinopathy (dataset provided by Hajeb et al. [84]), aiming at eventually contributing towards other medical applications such as multimodal registration. In addition, we intend to extend our approach to compute 3D stereoscopic reconstruction of fundus images, which is another underlying problem related to the context of diagnostic assistance as well.

### ACKNOWLEDGMENTS

We would like to thank the anonymous reviewers for their suggestions, and the Center for Mathematical Sciences Applied to Industry (CEPID-CeMEAI) for providing the computational resources. The authors thank the ophthalmologists (Gladston Pereira Paiva and Raquel Nunes) by their valuable comments and the provided dataset. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES), National Council for Scientific and Technological Development – Brasil (CNPq) under grant #301642/2017-6, and São Paulo Research Foundation (FAPESP) under grants #2019/13165-4, #2014/09546-9, and #2013/07375-0.

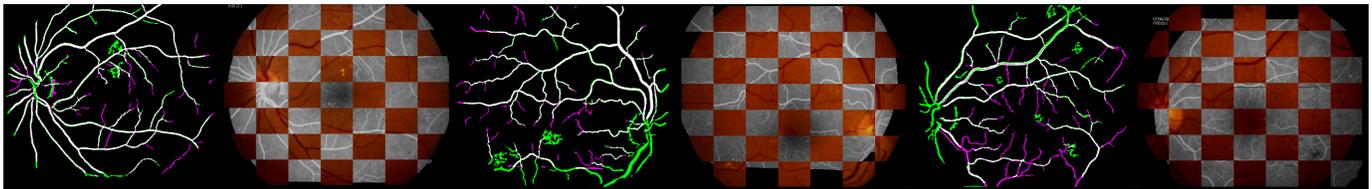


Fig. 15: Extension to multimodal fundus images. Binary vessels overlapping and their usual checkerboard representation.

## REFERENCES

- [1] World Health Organization. (2014) Visual impairment and blindness. [Online] Available: <http://www.who.int/mediacentre/factsheets/fs282/en/>.
- [2] H. A. Quigley and A. T. Broman, "The number of people with glaucoma worldwide in 2010 and 2020," *Brit. J. of Ophthalmol.*, vol. 90, no. 3, pp. 262–267, 2006.
- [3] F. Laliberté, L. Gagnon, and Y. Sheng, "Registration and fusion of retinal images: an evaluation study," *IEEE Trans. Med. Imaging*, vol. 22, pp. 661–673, 2003.
- [4] Y. Know, M. Adix, M. Zimmerman, S. Piette, E. Greenlee, W. Alward, and M. Abramoff, "Variance owing to observer, repeat imaging, and fundus camera type on cup-to-disc ratio estimates by stereo planimetry," *J. Glaucoma*, vol. 18, no. 4, pp. 305–310, 2009.
- [5] E. Karali, P. Asvestas, K. S. Nikita, and G. K. Matsopoulos, "Comparison of different global and local automatic registration schemes: An application to retinal images," in *LNCS*, 2004, pp. 813–820.
- [6] T. Kauppi, "Eye fundus image analysis for automatic detection of diabetic retinopathy," Ph.D. dissertation, Lappeenranta University of Technology, 2010.
- [7] Z. Zhang, R. Srivastava, H. Liu, X. Chen, L. Duan, D. Kee, C. Kwok, T. Wong, and J. Liu, "A survey on computer aided diagnosis for ocular diseases," *BMC Med. Inform. Decis. Mak.*, vol. 14, no. 1, pp. 1–80, 2014.
- [8] B. N. Kumar, R. P. Chauhan, and N. Dahiya, "Detection of glaucoma using image processing techniques: A review," in *MCCS 2016*, 2016, pp. 1–6.
- [9] D. S. Sisodia, S. Nair, and P. Khobragade, "Diabetic retinal fundus images: Preprocessing and feature extraction for early detection of diabetic retinopathy," *Biomedical and Pharmacology Journal*, vol. 10, no. 2, pp. 615–626, 2017.
- [10] H. Liao, Z. Zhu, and Y. Peng, "Potential utility of retinal imaging for Alzheimer's disease: A review," *Front. Aging Neurosci.*, vol. 10, pp. 1–12, 2018.
- [11] K. Deng, J. Tian, J. Zheng, X. Zhang, X. Dai, and M. Xu, "Retinal fundus image registration via vascular structure graph matching," *Int. J. Biomed. Imaging*, vol. 2010, pp. 14:1–14:13, 2010.
- [12] F. Santambrogio, *Optimal Transport for Applied Mathematicians*. Springer, 2015.
- [13] J. Solomon, "Optimal transport on discrete domains," arXiv, Tech. Rep., 2018, URL: <https://arxiv.org/abs/1801.07745>.
- [14] P.-A. Chiappori, R. J. McCann, and L. P. Nesheim, "Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness," *Economic Theory*, vol. 42, no. 2, pp. 317–354, 2010.
- [15] J. Solomon, F. de Goes, G. Peyré, M. Cuturi, A. Butscher, A. Nguyen, T. Du, and L. Guibas, "Convolutional Wasserstein distances: efficient optimal transportation on geometric domains," *ACM Trans. Graph.*, vol. 34, no. 4, pp. 1–11, 2015.
- [16] N. Papadakis, "Optimal transport for image processing," Ph.D. dissertation, Université de Bordeaux, 2015.
- [17] S. Haker, L. Zhu, A. Tannenbaum, and S. Angenent, "Optimal mass transport for registration and warping," *Int. J. Comput. Vis.*, vol. 60, no. 3, pp. 225–240, 2004.
- [18] T. ur Rehman, E. Haber, G. Pryor, J. Melonakos, and A. Tannenbaum, "3D nonrigid registration via optimal mass transport on the GPU," *Medical Image Analysis*, vol. 13, no. 6, pp. 931–940, 2009.
- [19] W. Zhou, L. Zhang, Y. Xie, and C. Liang, "A novel technique for prealignment in multimodality medical image registration," *Biomed Res. Int.*, pp. 1–17, 2014.
- [20] P. A. Legg, P. L. Rosin, A. D. Marshall, and J. E. Morgan, "Feature neighbourhood mutual information for multi-modal image registration: an application to eye fundus imaging," *Pattern Recognition*, vol. 48, no. 6, pp. 1937–1946, 2015.
- [21] A. Can, C. V. Stewart, B. Roysam, and H. L. Tanenbaum, "A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 347–364, 2002.
- [22] L. Chen, Y. Xiang, Y. Chen, and X. Zhang, "Retinal image registration using bifurcation structures," in *IEEE ICIP*, 2011, pp. 2169–2172.
- [23] B. Fang and Y. Y. Tang, "Elastic registration for retinal images based on reconstructed vascular trees," *IEEE Trans. Biomed. Eng.*, vol. 53, pp. 1183–1187, 2006.
- [24] P. Palraj and I. Venilla, "A hybrid retinal image registration using mutual information," *American J. of Applied Sciences*, vol. 10, no. 11, pp. 1386–1391, 2013.
- [25] L. Chen, Y. Lian, Y. Guo, Y. Wang, T. S. Hatsukami, K. Pimentel, N. Balu, and C. Yuan, "A vascular image registration method based on network structure and circuit simulation," *BMC Bioinformatics*, vol. 18, no. 1, p. 18:229, 2017.
- [26] E. Serradell, M. A. Pinheiro, R. Sznitman, J. Kybic, F. Moreno-Noguer, and P. Fua, "Non-rigid graph registration using active testing search," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 625–638, 2015, source code: <http://cvlab.epfl.ch/software/biomed/corr>.
- [27] Z. Ghassabi, J. Shanbehzadeh, and A. Mohammadzadeh, "A structure-based region detector for high-resolution retinal fundus image registration," *Biomed. Signal Process. Control*, vol. 23, pp. 52–61, 2016.
- [28] D. Braun, S. Yang, J. N. Martel, C. N. Riviere, and B. C. Becker, "EyeSLAM: Real-time simultaneous localization and mapping of retinal vessels during intraocular microsurgery," *Int. J. Med. Robotics Comput. Assist. Surg.*, vol. 14, no. 1, p. e1848, 2018, source code: <https://github.com/brianbecker/eyeslam>.
- [29] N. Ritter, R. Owens, J. Cooper, R. H. Eikelboom, and P. P. van Saarloos, "Registration of stereo and temporal images of the retina," *IEEE Trans. Med. Imaging*, vol. 18, no. 5, 1999.
- [30] G. K. Matsopoulos, N. A. Mouravliansky, K. K. Delibasis, and K. S. Nikita, "Automatic retinal image registration scheme using global optimization techniques," *IEEE Trans. Inf. Technol. Biomed.*, vol. 3, no. 1, pp. 47–60, 1999.
- [31] Y. Yang and S. Huang, "Retinal image mosaic based on genetic algorithm and automated blood vessel extracting approach," in *7th World Congress on Intelligent Control and Automation*, 2008, pp. 7751–7756.
- [32] P. Palraj and I. Vennila, "Retinal fundus image registration via blood vessel extraction using binary particle swarm optimization," *J. Med. Imaging Health*, vol. 6, no. 2, pp. 328–337, 2016.
- [33] M. Zaslavskiy, F. Bach, and J.-P. Vert, "A path following algorithm for the graph matching problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2227–2242, 2009.
- [34] J. Chen, J. Tian, N. Lee, J. Zheng, R. T. Smith, and A. F. Laine, "A partial intensity invariant feature descriptor for multimodal retinal image registration," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1707–1718, 2010.
- [35] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. of the 4th Alvey Vision Conf.*, 1988, pp. 147–151.
- [36] K. M. Adal, R. M. Ensing, R. Couvert, P. van Etten, J. P. Martinez, K. A. Vermeer, and L. J. van Vliet, "A hierarchical coarse-to-fine approach for fundus image registration," in *Int. W. Biomedical Image Registration*, 2014, pp. 93–102.
- [37] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, 1992.
- [38] G. Yang, C. Stewart, M. Sofka, and C.-L. Tsai, "Registration of challenging image pairs: Initialization, estimation, and decision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 11, pp. 1973–1989, 2007.
- [39] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

- [40] Z. Ghassabi, A. Sedaghat, J. Shanbehzadeh, and E. Fatemizadeh, "An efficient approach for robust multimodal retinal image registration based on UR-SIFT features and PIIFD descriptors," *EURASIP J. Image Vide.*, vol. 2013, p. 25, 2013.
- [41] C.-L. Tsai, C.-Y. Li, G. Yang, and K.-S. Lin, "The edge-driven dual-bootstrap iterative closest point algorithm for registration of multimodal fluorescein angiogram sequence," *IEEE Trans. Med. Imaging*, vol. 29, no. 3, pp. 636–649, 2010.
- [42] G. Wang, Z. Wang, Y. Chen, and W. Zhao, "Robust point matching method for multimodal retinal image registration," *Biomedical Signal Processing*, vol. 19, pp. 68–76, 2015.
- [43] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *ECCV*, 2006, pp. 404–417.
- [44] P. C. Cattin, H. Bay, L. Van Gool, and G. Székely, "Retina mosaicing using local features," in *MICCAI*, 2006, pp. 185–192.
- [45] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM.*, vol. 24, no. 6, pp. 381–395, 1981.
- [46] E. P. Ong, J. A. Lee, J. Cheng, G. Xu, B. H. Lee, A. Laude, S. Teoh, T. H. Lim, D. W. K. Wong, and J. Liu, "A robust outlier elimination approach for multimodal retina image registration," in *MICCAI*, 2015, pp. 329–337.
- [47] C. Hernandez-Matas, X. Zabulis, and A. A. Argyros, "Retinal image registration through simultaneous camera pose and eye shape estimation," in *IEEE EMBC*, 2016, pp. 3247–3251.
- [48] C. Hernandez-Matas, X. Zabulis, A. Triantafyllou, P. Anyfanti, and A. A. Argyros, "Retinal image registration under the assumption of a spherical eye," *Comput. Med. Imaging Graph.*, vol. 55, pp. 95–105, 2017.
- [49] C. Hernandez-Matas, X. Zabulis, and A. A. Argyros, "An experimental evaluation of the accuracy of keypoints-based retinal image registration," in *IEEE EMBC*, 2017, pp. 377–381.
- [50] H. Liu and S. Yan, "Common visual pattern discovery via spatially coherent correspondences," in *CVPR*, 2010, pp. 1609–1616.
- [51] J. Wang, J. Chen, H. Xu, S. Zhang, X. Mei, J. Huang, and J. Ma, "Gaussian field estimator with manifold regularization for retinal image registration," *Signal Processing*, vol. 157, pp. 225 – 235, 2019, source code: <https://github.com/JiahaoPlus/GFEMR>.
- [52] C. Villani, *Optimal transport: old and new*. Springer, 2009.
- [53] Y. Rubner, C. Tomasi, and L. J. Guibas, "A metric for distributions with applications to image databases," in *IEEE ICCV*, 1998, pp. 59–66.
- [54] F. Zhou and F. De la Torre, "Factorized graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 9, pp. 1774–1789, 2016, source code: <http://www.f-zhou.com/gm.html>.
- [55] C. Schellewald, S. Roth, and C. Schnörr, "Evaluation of a convex relaxation to a quadratic assignment matching approach for relational object views," *Image Vision Computing*, vol. 25, no. 8, pp. 1301–1314, 2007.
- [56] S. Ferradans, N. Papadakis, J. Rabin, G. Peyré, and J.-F. Aujol, "Regularized discrete optimal transport," in *Scale Space and Variational Methods in Computer Vision*. Springer, 2013, pp. 428–439.
- [57] P. Bankhead, C. N. Scholfield, J. G. McGeown, and T. M. Curtis, "Fast retinal vessel detection and measurement using wavelets and edge location refinement," *PLoS One*, vol. 7, no. 3, p. e32435, 2012.
- [58] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*, 2nd ed. Prentice Hall, 2002.
- [59] "MATLAB Image Processing Toolbox," 2017, the MathWorks, Natick, MA, USA.
- [60] S. Lee, *A study of ophthalmologic imaging distortion for retinal image registration and validation*. ProQuest, 2008.
- [61] T. L. Magnanti, R. K. Ahuja, and J. B. Orlin, *Network Flows: Theory, Algorithms, and Applications*. Pearson, 2014.
- [62] J. P. Vielma, "Mixed integer linear programming formulation techniques," *SIAM Rev.*, vol. 57, pp. 3–57, 2015.
- [63] D. Motta, W. Casaca, and A. Paiva, "Fundus image transformation revisited: Towards determining more accurate registrations," in *IEEE CBMS*, 2018, pp. 227–232.
- [64] J.-K. Kamarainen, V. Kyrki, and H. Kälviäinen, "Invariance properties of Gabor filter-based features-overview and applications," *IEEE Trans. Image Process.*, vol. 15, no. 5, pp. 1088–1099, 2006.
- [65] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in *CVPR*, 2014, pp. 2814–2821.
- [66] A. Hoover and M. Goldbaum, "Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels," *IEEE Trans. Med. Imaging*, vol. 22, no. 8, pp. 951–958, 2003.
- [67] W. Zhou, H. Wu, C. Wu, X. Yu, and Y. Yi, "Automatic optic disc detection in color retinal images by local feature spectrum analysis," *Comput. Math. Methods Med.*, vol. 2018, pp. 1–12, 2018.
- [68] P. McIlroy, E. Rosten, S. Taylor, and T. Drummond, "Deterministic sample consensus with multiple match hypotheses," in *BMVC*, 2010, pp. 1–11.
- [69] Z. Li, F. Huang, J. Zhang, B. Dashtbozorg, S. Abbasi-Sureshjani, Y. Sun, X. Long, Q. Yu, B. ter Haar Romeny, and T. Tan, "Multi-modal and multi-vendor retina image registration," *Biomed. Opt. Express*, vol. 9, no. 2, pp. 410–422, 2018.
- [70] R. G. Pontius, O. Thonteh, and H. Chen, "Components of information for multiple resolution comparison between maps that share a real variable," *Environmental and Ecological Statistics*, vol. 15, no. 2, pp. 111–142, 2008.
- [71] C. Hernandez-Matas, X. Zabulis, A. Triantafyllou, P. Anyfanti, S. Douma, and A. A. Argyros, "FIRE: Fundus image registration dataset," *J. Model. Ophthalmol.*, vol. 1, no. 4, pp. 16–28, 2017, source code: <http://www.ics.forth.gr/cvrl/fire/>.
- [72] T. Kohler, A. Budai, M. F. Kraus, J. Odstrcilik, G. Michelson, and J. Hornegger, "Automatic no-reference quality assessment for retinal fundus images using vessel segmentation," in *IEEE CBMS*, 2013, pp. 95–100, dataset: <http://www5.cs.fau.de/research/data/fundus-images/>.
- [73] S. Gold and A. Rangarajan, "A graduated assignment algorithm for graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 4, pp. 377–388, 1996.
- [74] M. Leordeanu and M. Hebert, "A spectral technique for correspondence problems using pairwise constraints," in *IEEE ICCV*, vol. 2, 2005, pp. 1–8.
- [75] M. Cho, J. Lee, and K. M. Lee, "Reweighted random walks for graph matching," in *ECCV*, 2010, pp. 492–505.
- [76] M. Izadi and P. Saeedi, "Robust weighted graph transformation matching for rigid and nonrigid image registration," *IEEE Trans. Image Process.*, vol. 21, no. 10, pp. 4369–4382, 2012.
- [77] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," in *ACM-MM*, ser. MM '10, 2010, pp. 1469–1472, source code: <http://www.vlfeat.org>.
- [78] G. Yang, C. Stewart, M. Sofka, and C.-L. Tsai, "Generalized Dual Bootstrap-ICP Algorithm," source code: <http://www.vision.cs.rpi.edu/gdbicp/>.
- [79] L. Chen, "Feature-based retinal image registration," source code: <https://www.mathworks.com/matlabcentral/fileexchange/23015-feature-based-retinal-image-registration>.
- [80] E. D. P. R. Group *et al.*, "The prevalence of diabetic retinopathy among adults in the united states," *Archives of Ophthalmology*, vol. 122, no. 4, p. 552, 2004.
- [81] "Key metrics: Assessing optometric practice performance," Practice Advancement Associates. Management & Business Academy for Eye Care Professionals, 2015, [Online] Available: [http://ecpu.com/media/wysiwyg/docs/paa\\_keymetrics\\_0415.pdf](http://ecpu.com/media/wysiwyg/docs/paa_keymetrics_0415.pdf).
- [82] "Top metric to track: Gross revenue per complete exam," Review of Optometric Business, 2014, [Online] Available: <https://reviewob.com/top-metric-to-track-complete-exams-per-od-hour/>.
- [83] J. A. Lee, J. Cheng, B. H. Lee, E. P. Ong, G. Xu, D. W. K. Wong, J. Liu, A. Laude, and T. H. Lim, "A low-dimensional step pattern analysis algorithm with application to multimodal retinal image registration," in *CVPR*, 2015, pp. 1046–1053.
- [84] S. Hajeb Mohammad Alipour, H. Rabbani, and M. R. Akhlaghi, "Diabetic retinopathy grading by digital curvelet transform," *Comput. Math. Methods Med.*, vol. 2012, 2012, dataset: [goo.gl/9Ngq8d](http://goo.gl/9Ngq8d).