

Accurate Correspondence of Cone Photoreceptor Neurons in the Human Eye Using Graph Matching Applied to Longitudinal Adaptive Optics Images

Jianfei Liu, HaeWon Jung, and Johnny Tam^(✉)

National Eye Institute, National Institutes of Health, Bethesda, MD, USA
johnny@nih.gov

Abstract. Loss of cone photoreceptor neurons is a leading cause of many blinding retinal diseases. Direct visualization of these cells in the living human eye is now feasible using adaptive optics scanning light ophthalmoscopy (AOSLO). However, it remains challenging to monitor the state of specific cells across multiple visits, due to inherent eye-motion-based distortions that arise during data acquisition, artifacts when overlapping images are montaged, as well as substantial variability in the data itself. This paper presents an accurate graph matching framework that integrates (1) robust local intensity order patterns (LIOP) to describe neuron regions with illumination variation from different visits; (2) a sparse-coding based voting process to measure visual similarities of neuron pairs using LIOP descriptors; and (3) a graph matching model that combines both visual similarity and geometrical cone packing information to determine the correspondence of repeated imaging of cone photoreceptor neurons across longitudinal AOSLO datasets. The matching framework was evaluated on imaging data from ten subjects using a validation dataset created by removing 15% of the neurons from 713 neuron correspondences across image pairs. An overall matching accuracy of 98% was achieved. The framework was robust to differences in the amount of overlap between image pairs. Evaluation on a test dataset showed that the matching accuracy remained at 98% on approximately 3400 neuron correspondences, despite image quality degradation, illumination variation, large image deformation, and edge artifacts. These experimental results show that our graph matching approach can accurately identify cone photoreceptor neuron correspondences on longitudinal AOSLO images.

Keywords: Adaptive optics · Split detection · Graph matching · Sparse coding · Cone photoreceptor neurons

The rights of this work are transferred to the extent transferable according to title 17 § 105 U.S.C.

1 Introduction

Adaptive optics scanning light ophthalmoscopy (AOSLO) [2, 7] provides microscopic access to individual neurons of the retina directly in the living human eye. Critical to the phenomenon of human vision are specialized neurons called cone photoreceptors. These neurons can be noninvasively imaged using AOSLO (protrusions in Fig. 1). The loss of cone photoreceptors is a critical feature of many blinding retinal diseases. Therefore, longitudinal monitoring of these neurons can provide important information related to the onset, status, and progression of blindness.

Currently, longitudinal monitoring of individual neurons within AOSLO images across different visits has only been attempted manually, which is not only labor-intensive, but also prone to error, and applicable over only small retinal regions [4, 8]. Existing algorithms for cell tracking from microscopy videos require uniform illumination and small time intervals. For example, Dzyubachyk [3] utilized a coupled level-set method to iteratively track cells where overlapping regions in previous video frames were used for initialization. Padfield [6] modeled cell behaviors within a bipartite graph, and developed a coupled minimum-cost flow algorithm to determine the final tracking results. Longitudinal AOSLO imaging datasets contain inherent challenges due to non-uniform illumination, image distortion due to eye motion or montaging of overlapping images, and a time interval between subsequent imaging sessions that can be on the order of several months.

To address these unique challenges, we developed a robust graph matching approach to identify neuron correspondences across two discrete time points. The main contributions are three-fold. First, a local intensity order pattern (LIOP) feature descriptor is exploited to represent neuron regions, robust against non-uniform changes in illumination. Second, a robust voting process based on sparse coding was developed to measure visual similarities between pairs of neurons from different visits. Third, a global graph matching method was designed to identify neuron correspondences based on both visual similarity and geometric constraints. Validation on longitudinal datasets from ten subjects demonstrated a matching accuracy over 98%, which is promising for potential clinical implementation.

2 Methodology

2.1 Longitudinal Matching of Cone Photoreceptor Neurons

Step 1: Detection of cone photoreceptor neurons. The first step is to identify neurons on images from multiple visits. A simplified version of a cell segmentation algorithm [5] was implemented, using the multi-scale Hessian matrix to detect neurons, and the convex hull algorithm to determine neuron regions (Fig. 1A).

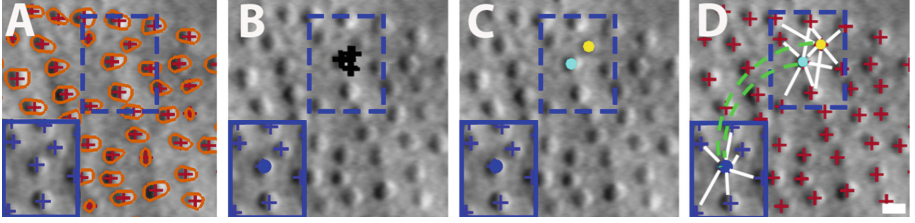


Fig. 1. Framework for neuron correspondence matching on longitudinal AOSLO images of the human eye, taken two months apart. In each panel, a portion of the image from the first visit is overlaid in the bottom left corner (solid rectangle) of the second visit image. Its corresponding location in the second visit is indicated by the dashed rectangles. (A) Identification of neurons (+’s) and convex hull regions (orange curves). (B) For each neuron from the first visit (e.g. blue dot), the LIOP feature descriptor and spare coding is used to determine candidate image points on the second visit (black +’s). (C) Based on the voting response at each candidate image point (i.e. visual similarity), candidate neurons for pairing are assigned, each with a visual similarity score (cyan and yellow dots). (D) Graph matching is used to determine correspondences based on both visual similarity (dashed green lines) and the arrangement of neighboring neurons (white lines). Scale bar = 10 μm .

Step 2: Neuron-to-region matching. The next step is to find all relevant neuron pairs between visits in order to set up graph matching, which relies on robust feature descriptors for neuron regions and an image matching process.

Since longitudinal AOSLO images often have significant illumination variation, we adapted the LIOP feature descriptor [10]. The LIOP descriptor starts by sorting all pixels in a neuron region based on their intensity values, I , in increasing order, and then equally dividing the region into M ordinal bins in terms of the intensity order. For each image point p from bin B , an N -dimensional vector $\mathbf{v} = \langle I(q) \rangle, q \in N(p)$ is established by collecting all intensity values $I(q)$ from their N -neighborhood points, and then the indices of \mathbf{v} are re-ordered based on intensity values to derive vector $\hat{\mathbf{v}}$. Let \mathbf{W} be an $N! \times N$ matrix containing all possible permutations of $\{1, 2, \dots, N\}$, and \mathbf{I} be an $N! \times N!$ identity matrix. The LIOP descriptor for point p is

$$\mathbf{d}(p) = \mathbf{I}_i, \text{ if } \hat{\mathbf{v}} = \mathbf{W}_i \quad (1)$$

The LIOP for each ordinal bin is defined as

$$\mathbf{d}(B) = \sum \mathbf{d}(p), p \in B \quad (2)$$

The LIOP descriptor of the entire neuron region is built by concatenating all sub-descriptors at each bin, which has the dimension of $N! \times M$. Note that LIOP groups image points with similar intensity in each bin, instead of their spatial neighborhood. Therefore, the LIOP descriptor is insensitive to the global illumination changes, such as when entire neuron regions become darker or brighter, which often happens in longitudinal AOSLO images.

We also developed a robust neuron-to-region matching strategy based on sparse coding to identify relevant neuron pairs. Suppose the LIOP descriptor for the neuron detection p (blue dot in Fig. 1B) in the first visit is an $N! \times M$ dimensional vector \mathbf{d}_1 . Transform p into the second visit image, and define a large image matching range Ω with size $M_1 \times M_1 > N! \times M$, centered at the transformed point. The LIOP descriptor is again established for each image point $q \in \Omega$, and combining all descriptors over Ω leads to basis matrix \mathbf{D} of size $(N! \times M) \times (M_1 \times M_1)$, which then fulfills the requirement of sparse coding that the basis matrix should be over-complete. Therefore, the image matching problem is converted into the vector \mathbf{d}_1 represented by the basis matrix \mathbf{D} , and mathematically defined as

$$\bar{\mathbf{x}} = \arg \min \|\mathbf{x}\|_1 \text{ subject to } \mathbf{d}_1 = \mathbf{D}\bar{\mathbf{x}} \quad (3)$$

where $\|\mathbf{x}\|_1 = \sum_{i=1}^{M_1 \times M_1} |x_i|$ denotes the L_1 norm of the vector \mathbf{x} . Subspace pursuit [1] was used to minimize Eq. 3, and non-zero elements of sparse vector $\bar{\mathbf{x}}$ are illustrated as black crosses in Fig. 1B. A voting process can thus be developed to find relevant neuron candidates (cyan and yellow points in Fig. 1C) in the second visit if their convex hulls have image points with non-zero sparse vector elements. Most of the black crosses are within the convex hull of actual corresponding neuron, and only a small set of relevant neuron pairs get reported from the neuron-to-region matching strategy, which significantly simplifies graph matching.

Step 3: Similarity assignment of neuron pairs. Using the sparse vector $\bar{\mathbf{x}}$, the similarity of a selected neuron pair can be computed as

$$\theta^v = 1.0 - \sum_j |\bar{x}_j| / \|\bar{\mathbf{x}}\|_1, \bar{x}_j \neq 0 \quad (4)$$

Here, \bar{x}_j denotes a non-zero sparse element associated with an image point which is within the convex hull of the neuron in the second visit. Utilizing Eq. 4, we can obtain discriminative assignments for all selected neuron pairs (e.g. blue to cyan and blue to yellow pairings in Fig. 1C).

Step 4: Graph matching. We now describe the graph matching model for finding neuron correspondences on longitudinal AOSLO images. Let P_1 and P_2 be the sets of neuron detections in two visits (blue and red crosses in Fig. 1D), and $A \subseteq P_1 \times P_2$ be the set of neuron pairs found from step 2. A matching configuration between P_1 and P_2 can be represented as a binary valued vector $\mathbf{m} = \{0, 1\}^A$. If a neuron pair $\alpha \in A$ is a true neuron correspondence, $m_\alpha = 1$; otherwise, $m_\alpha = 0$. Therefore, finding neuron correspondences is mathematically equivalent to calculating \mathbf{m} for all possible neuron pairs.

The first constraint is that the matching graph should contain the similarity assignments of the selected neuron pairs from the previous step depicted as dashed green curve in Fig. 1D, given by

$$E^v(\mathbf{m}) = \sum_{\alpha \in A} \theta^v \cdot m_\alpha \quad (5)$$

The second important constraint in the matching graph is the similarity of the adjacent neuron packing of neuron pairs (S), which is modeled as

$$E^g(\mathbf{m}) = \sum_{\alpha, \beta \in A} \theta^g \cdot m_\alpha \cdot m_\beta \quad (6)$$

S contains all adjacent neuron pairs defined over neighboring neurons

$$S = \{((p_1, p_2), (q_1, q_2)) \in A \times A \mid p_1 \in N^K(q_1) \wedge q_1 \in N^K(p_1) \wedge p_2 \in N^K(q_2) \wedge q_2 \in N^K(p_2)\} \quad (7)$$

N^K indicates the set of K -nearest neighborhood in the graph structure. In this paper, we set $K = 6$ as illustrated with white lines in Fig. 1D, motivated by the hexagonal packing arrangement observed for human cone photoreceptors. The similarity of adjacent neuron packing is calculated by combining both distance and direction constraints:

$$\begin{aligned} \theta^g &= (\exp(\delta_{\alpha, \beta}^2 / \sigma^2) - 1) + (\exp(\gamma_{\alpha, \beta}^2 / \sigma^2) - 1) \\ \delta_{\alpha, \beta} &= \frac{\|p_1 - q_1\| - \|p_2 - q_2\|}{\|p_1 - q_1\| + \|p_2 - q_2\|} \\ \gamma_{\alpha, \beta} &= \arccos\left(\frac{p_1 - q_1}{\|p_1 - q_1\|}, \frac{p_2 - q_2}{\|p_2 - q_2\|}\right) \end{aligned} \quad (8)$$

We set $\sigma = 2$ in our experiments.

The third term in our graph matching model is to ensure unique one-to-one neuron correspondence, which can be used to identify neuron appearance and disappearance.

$$E^p(\mathbf{m}) = 1 - \sum_{\alpha \in A} m_\alpha / \min\{|P_1|, |P_2|\} \quad (9)$$

$|P_1|$ and $|P_2|$ denote the number of neuron detections in the two visits, respectively.

Combining Eqs. 5, 6, and 9 leads to our graph matching model:

$$E(\mathbf{m}) = \lambda_v E^v(\mathbf{m}) + \lambda_g E^g(\mathbf{m}) + \lambda_p E^p(\mathbf{m}) \quad (10)$$

Here, λ_v , λ_g , and λ_p are weights set to 2, 1, and 10, respectively, in our experiments. Equation 10 was minimized by a dual decomposition approach [9], which leads to the final neuron correspondences for longitudinal AOSLO images.

2.2 Data Collection and Validation Method

To the best of our knowledge, there are no algorithms or publicly-available datasets utilizing this recently-developed AOSLO instrumentation [7] that could be used for comparison to our proposed method. Therefore, we acquired imaging data from ten subjects (5 male, 5 female; age: 26.3 ± 5.4 years, mean \pm SD) by repeatedly imaging the same retinal regions over several months. To construct

larger regions of interest, overlapping images were acquired and then montaged together. Imaging data was used to construct two types of datasets from ten subjects to evaluate the robustness and accuracy of the matching framework. For the first dataset (“validation dataset”), from each subject we collected multiple images of a retinal region within a time period of several hours and generated two different sets of images of the same retinal region, each with unique distortions due to eye motion (300×300 pixels; approximately 100×100 microns). Then, two different modifications were performed on the artificial image pairs: neuron removal on one image to simulate cell loss/gain, and artificial image translation to simulate mismatches in alignment between visits. The second dataset (“test dataset”) consisted of two sets of images collected several months apart from the same retinal region of each subject (500×500 pixels; approximately 170×170 microns). The matching accuracy was estimated as:

$$F = 1.0 - \frac{\text{number of errors}}{\text{maximum number of possible matches}} \quad (11)$$

Here, the errors include two different types: type 1, incorrect pairings between two neurons visible across both visits (this type of error usually leads to at least one additional error due to the one-to-one mapping) and type 2, incorrect pairings where one neuron was only visible on one of the visits (typically due to alignment errors at the boundaries).

3 Experimental Results

3.1 Validation Dataset

The number of neuron correspondences of each subject varied from 48 to 137 due to subject-to-subject anatomical differences (total: 713 neuron pairs). To test whether the proposed methods could detect cases of newly-appearing or disappearing neurons, 10 neurons were artificially removed from one image of each pair of images, resulting in a net increase in number of neurons of 8.0% to 26.3% ($18.0 \pm 5.5\%$), or conversely, a net loss of 7.3% to 21.4% ($15.1 \pm 3.8\%$) neurons (by reversing the order of visits; all numbers in this paper reported as mean \pm SD). In the case of adding neurons, 7 of 10 subjects maintained an accuracy of 100%, while the remaining 3 subjects had one error due to a mis-connection of one of the erased neurons. The overall matching accuracy in the presence of appearing neurons was 99.5% over 713 neuron correspondences. In the case of neuron removal, 6 of 10 subjects maintained an accuracy of 100%, while the remaining 4 subjects had one error which occurred at a site of artificial neuron removal. The overall accuracy in the presence of disappearing neurons was 98.2% over 713 correspondences. In both cases, the matching accuracy for the neuron pairs which were not removed was 100%, demonstrating that the algorithm was robust to different sets of distortion due to eye motion. The average computation time for the 300×300 pixel images which all contained different numbers of cells was 90 ± 28 s (Intel i7-3770 CPU, 16 GB RAM).

The matching accuracy after artificial translation, which effectively reduces the area of overlap between two visits, was no lower than 99.5% for a range of translations tested (from 0 to up to 150 pixels, corresponding to overlaps ranging from 100% down to 50%). These validation results establish that the proposed methods performed well even in the presence of disappearing/appearing neurons, artifacts due to eye motion distortion, and alignment mismatches resulting in a significant reduction in the amount of overlap between image pairs.

3.2 Test Dataset

Across 20 image pairs in the test dataset, the total number of neurons from the first and second visits were 3905, and 3900, respectively. Our matching framework determined that there were 3399 correspondences between the two visits. To evaluate accuracy, images were manually examined to detect all matching errors, including type 1 (black circle, Fig. 2K), and type 2 (black circle, Fig. 2I) errors. Across the entire test dataset, a total of 44 type 1 and 34 type 2 errors were flagged. The overall accuracy achieved was 98%.

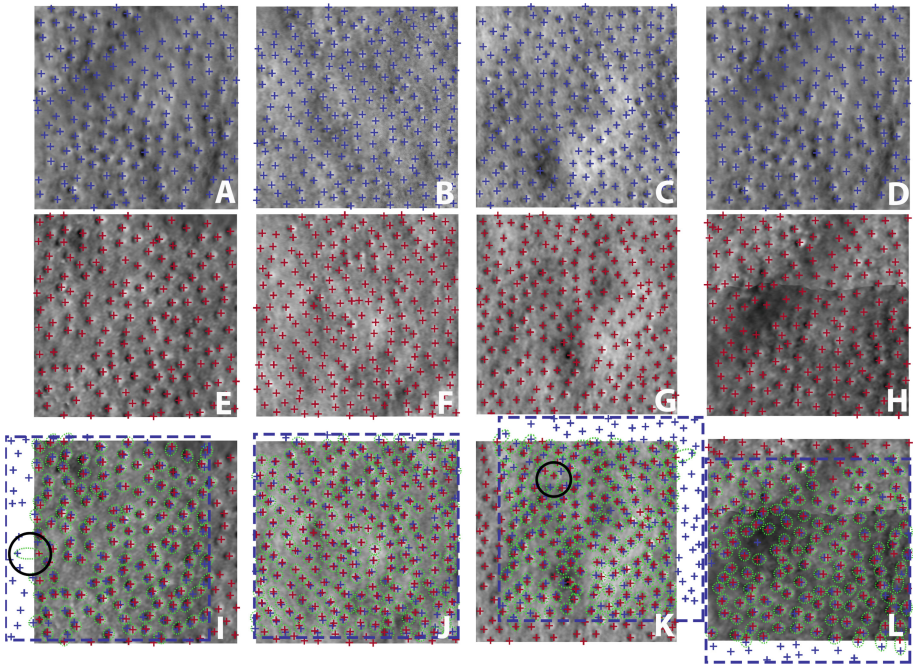


Fig. 2. Example matching results (each column is a subject), with neuron detections (+'s) from the first visit shown in the top row, second visit in the middle, and matching results overlaid on visit 2 in the bottom (dashed square indicates actual position of visit 1). In the bottom row, neuron correspondences are marked as green ellipses. Circles show examples of type 1 (K) and type 2 (I) errors.

Matching results for four subjects are shown in Fig. 2. In the first column, the image pair (A and E) exhibits significant illumination variation across visits, with most neurons in Fig. 2E being brighter than those in Fig. 2A. In addition, the contrast between neurons and background tissue is also higher in Fig. 2E. Overall, our matching framework was robust to the illumination changes. In the second column, the image quality was significantly lower across both visits, but our matching framework could still find neuron correspondences accurately. Large image distortions due to eye motion are visible in the third subject (Figs. 2C, G), but our matching framework was still able to identify most neuron correspondences. Finally, due to montaging of overlapping images, edge artifacts are sometimes present (Fig. 2H). Nevertheless, our matching framework was still able to accurately identify neuron correspondences. The average computation time for 500×500 pixel images was 430 ± 79 s.

4 Conclusion and Future Work

In this paper, we developed a robust matching framework to accurately determine cone photoreceptor neuron correspondences on longitudinal AOSLO images. The matching framework was developed based on three key contributions: application of the LIOP descriptor for neuron regions to tolerate illumination variation, a sparse-coding based voting process select relevant neuron pairs with discriminative similarity values, and a robust graph matching model utilizing both visual similarity and geometrical cone packing information. The validation dataset showed that the matching accuracy could achieve 98.2% even with about 15% neuron loss. The matching framework was able to tolerate an alignment error of at least 50% while maintaining over 99% accuracy. The matching accuracy on the test dataset was 98% over 3399 neuron correspondences, and showed high robustness to illumination variation, low image quality, image distortion, and edge artifacts. Future work will include application of our framework to additional patient datasets and optimization of computational speed.

Acknowledgements. This research was supported by the intramural research program of the National Institutes of Health, National Eye Institute.

References

1. Dai, W., Milenkovic, O.: Subspace pursuit for compressive sensing signal reconstruction. *IEEE Trans. Inf. Theory* **55**(5), 2230–2249 (2009)
2. Dubra, A., Sulai, Y.: Reflective afocal broadband adaptive optics scanning ophthalmoscope. *Biomed. Opt. Express* **2**(6), 1757–1768 (2011)
3. Dzyubachyk, O., van Cappellen, W., Essers, J., et al.: Advanced level-set-based cell tracking in time-lapse fluorescence microscopy. *IEEE Trans. Med. Imaging* **29**(3), 852–867 (2010)
4. Langlo, C., Erker, L., Parker, M., et al.: Repeatability and longitudinal assessment of foveal cone structure in CNGB3-associated achromatopsia. *Retina* (Epub Ahead of Print)

5. Liu, J., Dubra, A., Tam, J.: A fully automatic framework for cell segmentation on non-confocal adaptive optics images. In: SPIE Medical Imaging, p. 97852J (2016)
6. Padfield, D., Rittscher, J., Roysam, B.: Coupled minimum-cost flow cell tracking for high-throughput quantitative analysis. *Med. Image Anal.* **15**(4), 650–668 (2011)
7. Scoles, D., Sulai, Y., Langlo, C., et al.: In vivo imaging of human cone photoreceptor inner segments. *Invest. Ophthalmol. Vis. Sci.* **55**(7), 4244–4251 (2014)
8. Talcott, K., Ratnam, K., Sundquist, S., et al.: Longitudinal study of cone photoreceptors during retinal degeneration and in response to ciliary neurotrophic factor treatment. *Invest. Ophthalmol. Vis. Sci.* **54**(7), 498–509 (2011)
9. Torresani, L., Kolmogorov, V., Rother, C.: A dual decomposition approach to feature correspondence. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(2), 259–271 (2013)
10. Wang, Z., Fan, B., Wang, G., Wu, F.: Exploring local and overall ordinal information for robust feature description. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(11), 2198–2211 (2016)