Published in final edited form as:

Conf Proc IEEE Eng Med Biol Soc. 2016 August ; 2016: 1312–1315. doi:10.1109/EMBC.2016.7590948.

Toward a Severity Index for ROP: An Unsupervised Approach

Peng Tian¹, Esra Ataer-Cansizoglu², Jayashree Kalpathy-Cramer³, Susan Ostmo⁴, Karyn Jonas⁵, R.V. Paul Chan⁵, J. Peter Campbell⁴, Michael F. Chiang⁴, and Deniz Erdogmus¹

¹ECE Dept, Northeastern University, Boston, MA, USA

²MERL, Cambridge, MA, USA

³Athinoula A. Martinos Center for Biomedical Imaging, Dept of Radiology, Massachusetts General Hospital, Charlestown, MA, USA

⁴Dept of Ophthalmology, Casey Eye Institute, Oregon Health Science University, Portland, OR, USA

⁵Dept of Ophthalmology and Visual Sciences, Illinois Eye and Ear Infirmary, University of Illinois at Chicago, Chicago, IL, USA

Abstract

Retinopathy of prematurity (ROP) is a disease affecting low birth-weight infants and is the major cause of childhood blindness. Although accurate diagnosis is important, there is a high variability among expert decisions mostly due to subjective thresholds. Existing work focused on automated diagnosis of ROP. In this study, we construct a continuous severity index as an alternative to discrete classification. We follow an unsupervised approach by performing nonlinear dimensionality reduction. Instead of extracting several statistics of image features, each image is represented by the probability distribution of its features. The distance between distributions are then used in manifold learning methods as the distance between samples. The experiments are carried out on a dataset of 104 wide-angle retinal images. The results are promising and they reflect the challenges of the discrete classification.

I. INTRODUCTION

Retinopathy of prematurity is a disease that affects premature infants, and is a leading cause of childhood blindness throughout the world. It is diagnosed by retinal examination of at-risk infants in neonatal intensive care units [1], and examination findings are described using parameters defined by an international classification system that was originally published in 1984 [2]. *Plus disease* is a binary parameter (plus or normal) of the international classification system. Its presence is defined as tortuosity of arteries and dilation of veins in the retina greater than or equal to that in a standard published photograph selected by expert consensus during the 1980s [2], [3]. A revised international classification system in 2005 defined a new *pre-plus disease* entity as retinal vessels which are more dilated and tortuous than normal, but less than the standard photograph (plus, pre-plus, normal) [4]. The studies show that *plus disease* is the most important factor for identifying infants with severe ROP who require treatment to prevent blindness [5]. Therefore, accurate and consistent identification of plus disease is essential.

However, it is well established that there is significant inter-expert variability in diagnosis of plus disease in ROP [6], [7], [8]. These diagnostic errors may lead to under-treatment (resulting in permanent blindness) or over-treatment (resulting in permanent retinal scars from laser photocoagulation) of infants. Improving diagnostic accuracy and consistency in plus disease will lead to improved ROP management and prevention of visual impairment in children.

Since the obvious inter-expert disagreement exists in the diagnosis of ROP, which is produced by different subject thresholds for each expert, the discrete decision for binary or three-level classification cannot satisfy the accurate diagnosis demand. This paper presents an unsupervised technique to develop a continuous severity index for plus disease diagnosis as an alternative to discrete classification. We perform nonlinear dimensionality reduction based on the distance between samples. Each retinal image is represented by the probability distribution of its extracted image features. Thus, we construct the distance between images as the distance between their probability distributions. The distance matrix is then input to various manifold learning techniques for a comparative analysis. Manifold learning can be regarded as identifying a nonlinear mapping from the original higher dimensional data space to a lower dimensional representation. Existing methods aim to preserve global or local properties in the low-dimensional representation. Multidimensional scaling (MDS) is one of the global methods that finds a projection of the original data while preserving the pairwise Euclidean distances [9]. Similarly, in Isomap, one uses a geodesic distance estimation to use with MDS [10]. On the other hand, local methods construct the lower dimensional data using the local linear relations in the original space. Local linear embedding (LLE) [11] and Laplacian eigenmaps [12] aim to preserve local neighborhood information. In [13] a family of diffusion maps is learned to construct local coordinates.

The contributions of this paper are as follows: (i) unsupervised construction of a continuous severity index by nonlinear dimensionality reduction using manifold learning, (ii) comparative analysis of different manifold learning techniques using an information theoretic evaluation measure. These results extend our previous work using MDS, and lay the foundation for future work in which we will explore supervised nonlinear dimension reduction techniques that can cope with multilabeler uncertainty and various expert label forms, including class labels and pairwise comparisons that yield inequality constraints.

II. METHOD

We first trace retinal vessels and construct the vasculature tree given a binary image of retinal vessels. Second, tortuosity features are extracted from each point of the vessels which yields a pool of tortuosity values for each image. Third, we represent the pool of numbers with probability distributions instead of using statistics such as mean and maximum. Fourth, pairwise distances between images are computed as the distance between the probability distributions. Lastly, resulting distance matrix is fed into various manifold learning algorithms in order to perform a dimensionality reduction. We obtain 1-dimensional coordinates for each image as a candidate of its severity index.

A. Feature Extraction and Representation

We carry out principal spanning forest algorithm [14] to trace the vessels and construct the vasculature tree. Figure 1 displays tracing result on an example image from our dataset. Based on the observation that an image can contain both tortuous and straight vessels, an image is represented with two-component Gaussian Mixture Model (GMM) distribution of its image features [15]. This notion is demonstrated in Figure 1 where we compute tortuosity for each vessel segment and fit a two-component GMM to the extracted feature values. The figure shows the histogram of tortuosity features along with the vessels that are associated with each component. As can be seen tortuous and straight vessels are well distinguished with the probabilistic representation.

Based on our previous analysis [16] on the performance of various features with the aforementioned GMM representation, we extract acceleration, which is a point-based image feature that accounts tortuosity of a point on a vessel.

B. Distance between Images

Since each image is represented with the GMM distribution of its image features, we compute the distance between samples based on the distance between their distributions. Let $p^i(f)$ denote the GMM distribution of the feature in image i with means μ_1^i and μ_2^i , variances σ_1^i and σ_2^i , component priors π_1^i and π_2^i . Then, $p^i(f) = \pi_1^i \mathcal{N}(f; \mu_1^i, \sigma_1^i) + \pi_2^i \mathcal{N}(f; \mu_2^i, \sigma_2^i)$. The inner product between the distributions of image i and image j is $\langle p^i, p^j \rangle = \int p^i(f) p^j(f) df$. After some derivations, this is simplified as

$$\langle p^i, p^j \rangle = \sum_{t=1}^2 \sum_{k=1}^2 \pi_t^i \pi_k^j \mathcal{N}(\mu_k^j; \mu_t^i, \sigma_t^i + \sigma_k^j) \quad (1)$$

Using the inner product in equation (1) Euclidean distance satisfies $D^2(p^i, p^j) = \langle p^i, p^j \rangle + \langle p^j, p^i \rangle - 2 \langle p^i, p^i \rangle$.

C. Dimensionality Reduction

We employ six different manifold learning techniques that are well-established and popular in the literature: principal component analysis (PCA), multidimensional scaling (MDS) [9], Isomap [10], local linear embedding (LLE) [11], Laplacian eigen-maps [12] and diffusion maps [13]. These methods use a distance matrix between samples and aim to preserve either local or global distance relations in the reduced dimension. We plug in our GMM-based distance matrix, that measures pairwise distance as the distance between the GMMs. We obtain 1-dimensional coordinates as a result of these techniques, which can be regarded as a severity index. Moreover, these coordinates can help to order images with respect to the amount of tortuosity observed in the vessels.

D. Evaluation

In order to compare the results of different manifold learning algorithms, we compute mutual information (MI), a measure of dependency between two random variables. Assume X and Y are two random variables. The MI between them is $I(X; Y) = H(X) - H(X|Y)$ where $H(\cdot)$ and $H(\cdot|Y)$ denote entropy and conditional entropy respectively.

We use the Kernel Density Estimation plug-in MI estimator. Assume there are N samples, $x_i, i = \{1, 2, \dots, N\}$. With a Gaussian kernel, the probability density function is

$p(x) = N^{-1} \sum_{i=1}^N G_{\sigma^2}(x - x_i)$ where $G_{\sigma^2}(x_i) = (\sqrt{2\pi}\sigma_i)^{-1/2} \exp\{-(x_i^2)/(2\sigma_i^2)\}$ and σ_i^2 is the variable kernel covariance for the i th sample. The plug-in estimate of entropy is

$H(X) = -N^{-1} \sum_{i=1}^N \log p(x_i)$. The probability density given class label is $p(x|c_k) = \sum_{i \in C_k} G_{\sigma^2}(x - x_i)$ where C_k is the set of indices for the samples from class c_k . Then, the estimate

of the conditional entropy can be written as $H(X|Y) = -\sum_{k=1}^K p(c_k) m_k^{-1} \sum_{i=1}^{m_k} \log p(x_i|c_k)$ where m_k is the number of samples in class c_k . We compute kernel size by Silverman rule [17] where a fixed isotropic kernel is used for all samples.

To quantify the consistency between resulting coordinates and diagnosis, we compute the MI between resulting reduced coordinates of each algorithm and the class labels.

III. EXPERIMENTS AND RESULTS

A. Dataset Preparation

We use a dataset of 104 wide-angle colored retinal images acquired during routine clinical care and independently graded by three experts as plus, pre-plus, or normal, and compared to the clinical diagnosis. Gold standard is obtained using methods previously described [18]. There are 19 plus, 31 pre-plus and 54 normal samples in the dataset. The tracings are obtained from manual segmentations in order to avoid the bias an automated segmentation algorithm might introduce.

B. Pairwise Distance Matrix

Figure 2 shows the color map visualization of the pairwise distance matrix. The first 19 rows display the plus images, while the following 31 rows are pre-plus and the rest are normal samples. As can be seen the block diagonals are better visible between normal and plus samples while pre-plus class seems ambiguous as some of its samples are closer to normal and the others resemble to plus samples. Hence, discrete classification on the dataset is a challenging task given the distance matrix and a continuous index might help to order samples according to their severity.

C. Performance Evaluation

We run all methods given the GMM-based distance matrix and obtain a 1-D output. However, Isomap, LLE and Laplacian eigenmaps require a local neighborhood parameter k , while diffusion map method necessitates input parameters σ and α associated to neighborhood and regularization respectively. For parameter selection, we use jackknife

estimate. We vary values for these parameters and pick the best one for each method that maximizes jackknife estimate of MI between resulting coordinates and the class labels.

Table I reports the MI values computed for each method given the distance matrix and the selected parameter values. The coordinates Diffusion map generate has the highest consistency with expert's decision. Figure 5 shows the resulting coordinates for the methods. For better visualization we also plot the rank orders such that the rank of each image is obtained by sorting the resulting 1-D coordinates. Class labels are indicated by different colors. As can be seen, for most of the methods pre-plus samples are mixed into the normal and plus classes, which is also consistent with the distance matrix visualization in Figure 2 where pre-plus samples have a diverse range of distance relation with other classes. Diffusion maps method provides a relatively better discrimination between classes. An interesting observation is that most of the methods tend to keep normal samples close to each other while plus and pre-plus samples are sparsely distributed. This might be due to higher intra-class connectivity of normal samples compared to the other classes. Moreover, this also shows that making a decision of pre-plus or plus is more challenging than determining a normal sample.

Figure 3 displays some sample images along with the ranking coordinates of diffusion maps algorithm. The numbers above each image show the number of expert decisions plus disease (+), pre-plus ($\pm$), and normal (–). We notice that from left to right, the amount of tortuous vessels increase. Furthermore, plus disease decisions among experts increase from left to right.

Figure 4 shows the severity ranking obtained with diffusion map versus that obtained with expert consensus (extracted from pairwise comparisons of images by five experts). While the unsupervised ranking exhibits some level of correlation with experts, we anticipate much better results when, in the future, we develop severity indices that consider labels in the form of diagnosis (class label) and relative severity (pairwise comparisons).

IV. CONCLUSION AND FUTURE WORK

The demand to support clinical diagnosis of ROP through automatic data analysis has imposed the need to develop computer-based diagnostic assistance systems. A continuous severity index for ROP extracted automatically from retina imagery is useful for approximately assessing the severity of a patient's image and comparing two images from two patients or from the same patient in two separate visits, for instance during the course of treatment.

We presented several unsupervised severity index propositions generated using existing manifold learning methods with novel a GMM-based pairwise distance computation. To the best of our knowledge, we present the first automatic retina image based ROP severity index assessment as an alternative to discrete plus-disease stage classification (among healthy, pre-plus, and plus). In future work, we will augment and improve the developed severity index to use class label and pairwise comparison information from multiple experts. In doing this, we

will consider the uncertainty of expert labels and resolve contradictions that may arise from inter-labeler variability.

Acknowledgments

Our work is supported by NIH (R01EY019474).

References

1. Fierston WM, Saunders RA, Good W, Palmer EA, Phelps D, Reynolds J, Chiang MF, Ruben JB, Granet DB, Blocker RJ, et al. Screening examination of premature infants for retinopathy of prematurity. *Pediatrics*. 2013; 131(1):189–195. [PubMed: 23277315]
2. C. for the Classification of Retinopathy of Prematurity. An international classification of retinopathy of prematurity. *Archives of ophthalmology*. 1984; 102(8):1130–1134. [PubMed: 6547831]
3. C. for Retinopathy of Prematurity Cooperative Group et al. Multicenter trial of cryotherapy for retinopathy of prematurity: Preliminary results. *Archives of Ophthalmology*. 1988; 106(4):471–479. [PubMed: 2895630]
4. I. C. for the Classification of Retinopathy of Prematurity. The international classification of retinopathy of prematurity revisited. *JAMA Ophthalmology*. 2005; 123(7):991–999.
5. E. T. for Retinopathy of Prematurity Cooperative Group et al. Revised indications for the treatment of retinopathy of prematurity: results of the early treatment for retinopathy of prematurity randomized trial. *Archives of Ophthalmology*. 2003; 121(12):1684–1694. [PubMed: 14662586]
6. Reynolds JD, Dobson V, Quinn GE, Fielder AR, Palmer EA, Saunders RA, Hardy RJ, Phelps DL, Baker JD, Trese MT, et al. Evidence-based screening criteria for retinopathy of prematurity: natural history data from the cryo-rop and light-rop studies. *Archives of ophthalmology*. 2002; 120(11):1470–1476. [PubMed: 12427059]
7. Chiang MF, Jiang L, Gelman R, Du YE, Flynn JT. Interexpert agreement of plus disease diagnosis in retinopathy of prematurity. *Archives of ophthalmology*. 2007; 125(7):875–880. [PubMed: 17620564]
8. Wallace DK, Quinn GE, Freedman SF, Chiang MF. Agreement among pediatric ophthalmologists in diagnosing plus and pre-plus disease in retinopathy of prematurity. *Journal of American Association for Pediatric Ophthalmology and Strabismus*. 2008; 12(4):352–356. [PubMed: 18329925]
9. Kruskal JB. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*. 1964; 29(1):1–27.
10. Tenenbaum JB, De Silva V, Langford JC. A global geometric framework for nonlinear dimensionality reduction. *science*. 2000; 290(5500):2319–2323. [PubMed: 11125149]
11. Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. *Science*. 2000; 290(5500):2323–2326. [PubMed: 11125150]
12. Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*. 2003; 15(6):1373–1396.
13. Coifman RR, Lafon S. Diffusion maps. *Applied and computational harmonic analysis*. 2006; 21(1):5–30.
14. Bas E, Ataer-Cansizoglu E, Erdogmus D, Kalpathy-Cramer J. Retinal vasculature segmentation using principal spanning forests. *IEEE International Symposium on Biomedical Imaging*. 2012:1792–1795.
15. Bolon-Canedo, V., Ataer-Cansizoglu, E., Erdogmus, D., Kalpathy-Cramer, J., Chiang, M. A GMM-based feature extraction technique for the automated diagnosis of retinopathy of prematurity. *IEEE International Symposium on Biomedical Imaging; IEEE*; 2015. p. 1498-1501.
16. Ataer-Cansizoglu E, Bolon-Canedo V, Campbell JP, Bozkurt A, Erdogmus D, Kalpathy-Cramer J, Patel S, Jonas K, Chan RP, Ostmo S, et al. Computer-based image analysis for plus disease diagnosis in retinopathy of prematurity: Performance of the i-rop system and image features associated with expert diagnosis. *Translational Vision Science & Technology*. 2015; 4(6)

17. Silverman, BW. Density estimation for statistics and data analysis. Vol. 26. Chapman & Hall/CRC; 1986.
18. Ryan, MC., Ostmo, S., Jonas, K., Berrocal, A., Drenser, K., Horowitz, J., Lee, TC., Simmons, C., Martinez-Castellanos, M., Chan, RVP., Chiang, MF. Development and evaluation of reference standards for image-based telemedicine diagnosis and clinical research studies in ophthalmology. AMIA Annual Symposium; 2014.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 1.

Tracing of vessels and illustration of GMM-based feature representation: resulting vasculature tree where white squares indicate the junction/end points of vessels (left), histogram of the segment-based tortuosity feature values for the image along with the fitted GMM. On the left panel, each vessel segment is shown with its corresponding cluster color based on GMM fitting. Mean of each mixture component is shown with dashed lines with its respective color. Notice that red segments have high curvature while green segments are more straight.

Fig. 2.

Color map visualization of the pairwise distance matrix. The first 19 rows display the plus samples. Rows between 20–50 (inclusive) indicate pre-plus samples and the bottom 54 rows show the normal samples.

Fig. 3.

Rank order of images according to diffusion maps coordinates along with some example images. The numbers above each image show the number of expert decisions of plus disease (+), pre-plus ($\pm$), and normal (-).

Fig. 4.
Severity ranking with diffusion map versus consensus ranking of multiple experts.

Fig. 5. Resulting 1-D coordinates (left) and rank order of images according to these coordinates (right) for each method. The vertical axis is class label: normal (top), pre-plus (middle), plus (bottom). The same information is also coded by marker shape and color for added visual clarity.

TABLE I

Evaluation of manifold learning methods

Manifold Learning Method	Mutual Information
Diffusion Maps ($\sigma = 1.2$ $\alpha = 0$)	1.679
LLE (k=7)	1.560
Isomap (k=8)	1.521
MDS	1.515
Laplacian Eigenmaps (k=7)	1.463
PCA	1.463