

## Using Experts' Perceptual Skill for Dermatological Image Segmentation

Rui Li<sup>1</sup>, Pelz J<sup>1</sup>, Haake AR<sup>1</sup><sup>1</sup>Golisano College of Computing and Information Science, Rochester Institute of Technology, Rochester, NY.**Corresponding Author:** Rui Li, Golisano College of Computing and Information Science, Rochester Institute of Technology, Rochester, NY. **Tel:** +1 585-475-7203; **Email:** rxlics@rit.edu**Received Date:** 18 Apr 2016**Copyright** © 2016 Li R**Accepted Date:** 17 May 2016**Citation:** Li R, Pelz J and Haake AR. (2016). Using Experts' Perceptual Skill for Dermatological Image Segmentation. *M J Derm.* 1(1): 007.**Published Date:** 21 Jul 2016

### ABSTRACT

There is a growing reliance on imaging equipment in medical domain, hence medical experts' specialized visual perceptual capability becomes the key of their superior performance. In this paper, we propose a principled generative model to detect and segment out dermatological lesions by exploiting the experts' perceptual expertise represented by their patterned eye movement behaviors during examining and diagnosing dermatological images. The image superpixels' diagnostic significance levels are inferred based on the correlations between their appearances and the spatial structures of the experts' signature eye movement patterns. In this process, the global relationships between the superpixels are also manifested by the spans of the signature eye movement patterns. Our model takes into account these dependencies between experts' perceptual skill and image properties to generate a holistic understanding of cluttered dermatological images. A Gibbs sampler is derived to use the generative model's structure to estimate the diagnostic significance and lesion spatial distributions from superpixel-based representation of dermatological images and experts' signature eye movement patterns. We demonstrate the effectiveness of our approach on a set of dermatological images on which dermatologists' eye movements are recorded. It suggests that the integration of experts' perceptual skill and dermatological images is able to greatly improve medical image understanding and retrieval.

### KEYWORDS

Perceptual Expertise; Dermatological Image Understanding; Probabilistic Modeling; Gibbs Sampling; Eye Tracking Experiments; Superpixels.

### INTRODUCTION

Image understanding in knowledge-rich domains is particularly challenging, because experts' domain knowledge and perceptual expertise are demanded to transform image pixels into meaningful contents [1-3]. This motivates using active learning methods to incorporate experts' specialized capability into this process in order to improve the segmentation performance [4-6]. However, traditional knowledge acquisition used by active learning methods, such as manual markings, annotations and verbal reports, poses series of significant problems [7, 2]. Because tacit (implicit) knowledge as an integral part of expertise is not consciously accessible to experts,

it is difficult for them to identify exactly the diagnostic reasoning processes involved in decision-making. On the other hand, empirical studies suggest that eye movements, as both direct input and measurable output of real time signal processing in the brain, provide us an effective and reliable measure of both human cognitive processing and perceptual skill [8, 9].

Recently, studies try to incorporate perceptual skill into image understanding approaches [4, 7, 2, 6, 10]. Experts' eye movement data are projected into image feature space to evaluate feature saliency by weighting local features close to each fixation. Salient image features are then mapped back to spatial space in order to highlight regions of interest and at-

tention selection [4]. Furthermore, a conceptual framework is developed to measure image feature relevance based on corresponding patterned eye movement deployments defined by pair-wise comparison between multiple experts' eye movement data [7, 2]. In an image retrieval study [10], images are segmented generically based on image features first. In the later step of image matching, the similarity measures between query image segments and candidate images are weighted by subjects' fixation data. There are significant limits associated with these approaches. Some of them treat eye movements as a static process by solely using fixation locations without taking the dynamic nature into account [11]. Other studies apply dynamic models to capture sequential information of eye movements, but the cardinality is heuristic and the eye movement descriptions are limited. Since these studies directly use the observed eye movement data, which are noisy and inconsistent, to evaluate image features, their methods' reliability and robustness are undermined.

To address these limitation, we adopt *Signature Patterns* to represent experts' perceptual skill [12-14]. *Signature Patterns* summarize a set of eye movement patterns consistently exhibited by dermatologists while examining various dermatological images in Fig. 1. *Signature Patterns* encode rich information of experts' eye movements including fixation locations, fixation durations, and saccade amplitude. Recent eye tracking studies suggest that all these descriptions characterize important diagnostic reasoning processes [15, 16, 9]. Based on these eye movement descriptions, *Signature Patterns* shed light on both diagnostic significance of image regions and their relative spatial structures. We believe that the patterned behaviors underlying observed human behavioral data is a critical intermediate component to advance domain-specific image understanding. From this perspective, we utilize the spatial-temporal properties discovered in experts' eye movement data to guide medical image segmentation. In particular, we propose a principled generative model to combine experts' *Signature Patterns* with image properties in order to improve image segmentation by estimating diagnostic significance and varying visual spatial structures of the medical images.

### Perceptual Expertise Based Segmentation

Instead of directly evaluating image features with noisy unstable fixations, we formulated this problem into a probabilistic inference one: given the image features and *Signature Patterns* from experts, we are inferring diagnostic significance of image regions and their spatial structures [17]. Our method simultaneously grouped dense image features into spatially coherent segments via color and texture cues, and refined

these partitions using *Signature Patterns*, as depicted in the probabilistic generative model in Figure 2.

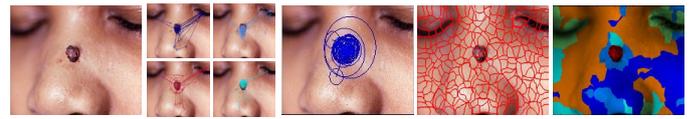


Figure 1: One dermatological image segmentation using our method which combines experts' *Signature Patterns* and superpixels. From left to right, we show an original image, four experts' eye movement sequences super-imposed onto the image (each circle center represents a fixation location and the radius is proportional to the duration time on that particular fixation, a line connecting two fixations represents a saccade), one instantiation of the *Concentrating Pattern* learned from multiple experts' eye movement sequences which is characterized by a group of eye movement sequence segments with similar spatio-temporal properties [12, 13], an over-segmentation of 400 superpixels, and segmentation result generated from our model with inferred primary abnormality highlighted in red.

Superpixels were adopted as the finest information scale [18, 19]. Each image was first segmented into a group of locally smoothed regions with the normalized cuts spectral clustering algorithm [18] as shown in Fig. 1. These regions contain more contextual information than at the pixel grid level. We described the texture of each superpixel using a local texton histogram by quantizing band-pass filter responses into 64 bins. A color histogram was computed by quantizing the Hue-Saturation-Value (HSV) color space into 60 bins. For each image  $n$ , the feature vector of  $m$ -th superpixel was represented with a variable  $x_{nm}^{(i)}$ .

An eye movement sequence is described by a series of eye movement units, each one of which is composed of one fixation and its successive saccade. The features describe an eye movement unit are fixation location, fixation duration, and saccade amplitude.

Similarly, let  $x_{nm}^{(e)}$  be the  $m$ -th eye movement unit.

Our representation of a dermatological image  $n$  was characterized by a collection of regions with various diagnostic significance levels. Each region was composed of a number of superpixels. The goal is to infer the diagnostic significance (e.g., primary morphology) of each region both locally from the set of its internal superpixels and globally from experts' signature eye movement patterns. Given the perceptual category [12],  $o_n$  which the image  $n$  belong to, we parameterized these dependencies with a significance variable  $s^{(i)}$  that indicates the

diagnostic importance of the superpixel  $i$ 's internal visual information, and a variable  $t_n$  which indicate spatial structures of the Signature Pattern  $X^{(e)}$ . The joint distribution was defined as

$$p(x^{(i)}, x^{(e)}, o_n, s^{(i)}, t_n) = \prod_{n=1}^N p(o_n) p(s^{(i)} | o_n, \pi) p(t_n | o_n, \lambda) \quad (1)$$

$$\cdot \prod_{n=1}^N \prod_{m=1}^{M_n^{(i)}} p(x_{nm}^{(i)} | s^{(i)}, \theta^{(i)}) \prod_{n=1}^N \prod_{m=1}^{M_n^{(e)}} p(x_{nm}^{(e)} | s^{(i)}, t_n, \theta^{(e)}) \quad (2)$$

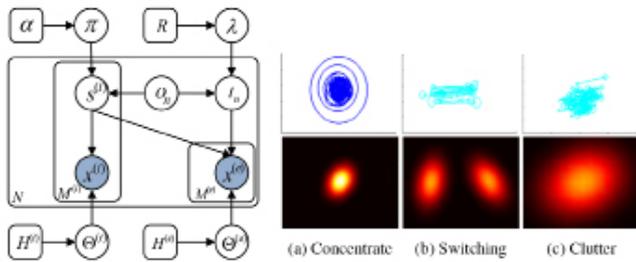


Figure 2: The proposed graphical model. *Left*: the generative probabilistic process is described using graphical models. See section 2 for details. *Right*: In the upper row, we show one centered instantiation for each type of *Signature Patterns*. In the lower row, we plot the Gaussian-distributed spatial structures  $t_n$  learned by our model from multiple instantiations of corresponding *Signature Patterns*. The combinations of these three types of *Signature Patterns* form four perceptual categories of dermatological images [12].

Where  $N$  is the number of images,  $M_n^{(i)}$  is the number of superpixels in the  $n$ -th image, and  $M_n^{(e)}$  is the number of eye movement units of the signature pattern. Thus  $X^{(i)} = \{x_{nm}^{(i)}\}$  and  $X^{(e)} = \{x_{nm}^{(e)}\}$  are the sets of all image and eye movement features.  $t_n$ ,  $s^{(i)}$ , and  $o_n$  specify spatial structures of signature patterns, the diagnostic significance level of superpixel  $i$ , and the perceptual category of the image, respectively. Note that the model has different dimensionality for different perceptual category as the dimensionality of  $t_n$  and  $s^{(i)}$  depends on the perceptual category  $o_n$ .

We defined  $K$  distinct diagnostic significance levels which give rise to superpixels with different typical features  $X^{(i)} = \{x_{nm}^{(i)}\}$ . The particular significance level  $s^{(i)}$  with each superpixel are independently sampled from a category-specific multinomial distribution:

$$p(s^{(i)} | o_n, \pi) = Mult(s^{(i)} | \pi_{o_n}) \quad (3)$$

Where  $\pi_{o_n}$  is regularized by a Dirichlet prior  $\pi_{o_n} \sim Dir(\alpha)$ . After

we independently selected a diagnostic significance level for each superpixel  $i$ , we sampled features using parameters determined by  $s^{(i)}$ :

$$p(x_{nm}^{(i)} | s^{(i)}, \theta^{(i)}) = Mult(x_{nm}^{(i)} | \theta_{s^{(i)}}^{(i)}) \quad (4)$$

We used  $t_n$  to denote spatial structures of *Signature Patterns'* eye movement units which are scale-normalized and centered, as shown in Fig. 2. Depend on the perceptual category of an image  $o_n$ ,  $t_n$  either parameterizes a unimodal Gaussian distribution for Concentrating and Clutter patterns, or a bimodal Gaussian distribution for Switching pattern.  $t_n$  is sampled from a normal-inverse-Wishard distribution parameterized by

$\lambda = \{\mu, \Lambda\}$ . Furthermore,  $\theta^{(e)}$  specifies  $t_n$ 's positions relative to an image-specific reference transformation. These transformations are independently sampled from some common prior distribution. We assigned each perceptual category unimodal Gaussian transformation densities:

$$p(\theta^{(e)} | \lambda, o_n) = N(\theta^{(e)} | \zeta_{o_n}, \gamma_{o_n}) \quad (5)$$

Where  $\lambda = \{\zeta, \gamma\}$ , and these transformation distribution are regularized by conjugate normal-inverse-wishart prior  $\lambda \sim NIW(R)$ . We then modeled the eye movement units of *Signature Patterns* as a Gibbs distribution:

$$p(x_{nm}^{(e)} | s^{(i)}, t_n, \theta^{(e)}) \propto exp(\beta f(x_{nm}^{(i)}, s^{(i)}, t_n, \theta^{(e)})) \quad (6)$$

Where  $f$  denotes the correlation between spatial structures of *Signature Patterns* and diagnostic significance. And  $\beta$  is a free parameter to control the overall reliability of *Signature Patterns*.

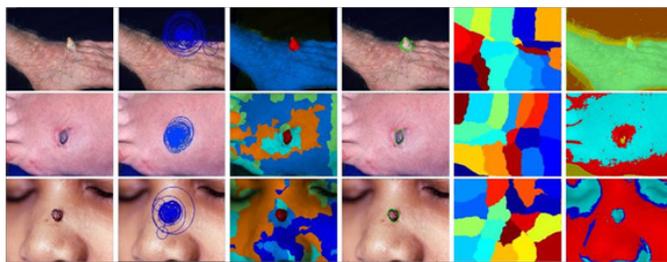
A Gibbs sampler was developed to learn the parameters of our model. We assumed that hyperparameter values are fixed. We also employed conjugate prior distributions for the diagnostic significance and spatial structure densities. These allowed us to learn the parameters by iteratively sampling significance  $s^{(i)}$  using likelihoods which analytically marginalize significance probabilities  $\pi$  and parameters  $\theta^e$ . This Rao-Blackwellized Gibbs sampler approximates the posterior distribution  $p(s^{(i)} | X^{(e)}, X^{(i)}, o_n)$  by sampling from it. This posterior can then be used to estimate the underlying parameters.

### Experimental Evaluation

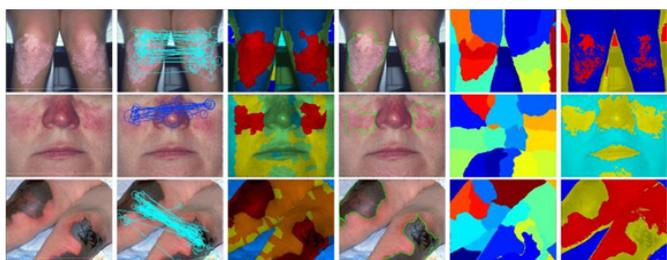
Eye tracking experiments were conducted to collect eye movement sequences from eleven board-certified dermatologists while they were examining and diagnosing a set of 70 dermatological images, each representing a different diagnosis. A SMI (SensoMotoric Instruments) eye tracking apparatus, which was running at 50 Hz sampling rate and has reported

accuracy of 0.5° visual angle, was applied to display the stimuli at a resolution of 1680 by 1050 pixels. The subjects viewed the medical images binocularly at a distance of about 60 cm. They were instructed to examine the medical images and make a diagnosis. The experiment started with a 9-point calibration and the calibration was validated after every 10 images. Based on the inferred *Signature Patterns*, we ran our model to do segmentation.

We compared our results to the normalized cuts spectral clustering method (NCuts), and a Gaussian mixture model (GMM) based on the same set of color and texture features as used in our model. Our model consistently detected variability in the lesion spatial distributions, and capture both primary and secondary abnormalities. In contrast, NCuts tends to generate regions of equal size, which produces distortions. If the lesion scales are comparable with the body parts' in the images, NCuts is relatively effective (4th row and 6th row, 5th column in Fig. 3). Otherwise, it tends to fail. Without incorporating domain-specific expertise, GMM discards the dependencies among the superpixels at diagnostic level. So it fails to either differentiate the regions with similar appearance (3rd row and 6th column in Fig. 3).



(a) Singular lesion segmentation based on Concentrating Pattern.



(b) Symmetric lesion segmentation based on Switching Pattern.



(c) High-density lesion segmentation based on Clutter Pattern.



(d) Multiple-morphology lesion segmentation based on Concentrating and Switching Patterns.

Fig.3: Qualitative comparison of dermatological image segmentation for each perceptual categories. Column 1: actual photographs. Column 2: Signature Patterns. Column 3 and 4: Our algorithm (red=primary abnormality). Column 5: NCuts spectral clustering method. Column 6: Gaussian mixture model. **Best viewed in color.**

Using Experts' Perceptual Expertise for Dermatological Image Segmentation 7

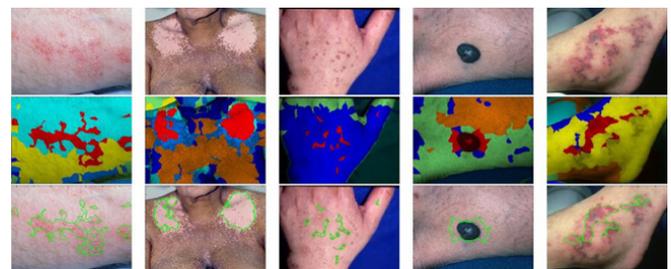


Fig.4: Qualitative test results for our perceptual expertise based segmentation. An even sampling from least to highest pixel error are shown from left to right. Original images are shown in the top row, the detected region labels in the middle row, and estimated primary abnormalities in the bottom row. The diagnostic significance estimates suffer if some abnormalities of high-density lesions are peripheral in terms of Signature Patterns' spatial distributions.

To quantitatively evaluate results, we measured overlap with dermatologists' segments via the Rand index in Fig. 5. Although GMM performs well for some images with unambiguous features, our method was often substantially better. *Signature Patterns* enabled us to improve segmentation performance without learning images layers. The perceptual skill encoded in *Signature Patterns* helped to capture global spatial structures of lesions among dermatological images. Note that superpixels being fixated on were not necessarily to be diagnostically significant, and those without fixations on them may be equivalently important as parts of primary abnormalities. Our model captured this subtlety by inferring the correlations between *Signature Patterns* and superpixel properties as well as taking *Signature Patterns*' spatial transformation across different images into consideration.

## CONCLUSION

We developed a principled generative model for dermatological image segmentation which integrates the extracted experts' perceptual expertise represented by their patterned eye movement behaviors and image superpixel properties to evaluate diagnostic significance of image regions. Our experiments show that the effective representation of visual perceptual expertise allows better domain-specific image segmentation. This provides a promising starting point for jointly performing experts' perceptual skill eval-

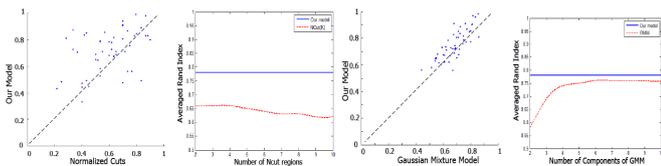


Fig.5: Quantitative comparison of segmentation results to human segments using the Rand index. *Left*: Scatter plot of our method and NCut (4) Rand indexes for 60 dermatological images, and average Rand indexes for NCut(K) where K refers to the number of NCuts regions. *Right*: Scatter plot of our method and GMM Rand indexes for the images, and average Rand indexes for GMM with various number of Gaussian components.

uation and image understanding. We believe that such integration will be able to greatly improve medical training and medical image retrieval.

## REFERENCES

1. Palmeri TJ, Wong AC and Gauthier I. (2004). Computational approaches to the development of perceptual expertise. *Trends in Cog. Sci.* 8(8), 378-386.
2. Dempere-Marco L, Hu X-P and Yang G-Z. (2011). A Novel Framework for the Analysis of Eye Movements during Visual Search for Knowledge Gathering. *J. Cognitive Computation.* 3(1), 206-222.
3. Hoffman R and Fiore MS. (2007). Perceptual (Re) learning: A Leverage Point for Human-Centered Computing. *J. Intelligent Systems.* 22(3), 79-83.
4. Hu X-P, Dempere-Marco L and Yang G-Z. (2003). Hot Spot Detection based on Feature Space Representation of Visual Search. *IEEE Trans Med Imaging.* 22(9), 1152-1162.
5. Lotenberg S, Gordon S, Long R, Antani S, et al. (2009). Automatic Evaluation of Uterine Cervix Segmentations using Ground Truth from Multiple Experts. *Computerized Medical Imaging and Graphics.* 33(3), 205-216.

6. Wang L, Merrifield R and Yang G-Z. (2011). Reinforcement Learning for Context Aware Segmentation. In: *Medical Image Computing and Computer-Assisted Intervention MICCAI 14(pt3)*, 627-634.
7. Dempere-Marco L, Hu X-P, Ellis SM, Hansell D M, et al. (2006). Analysis of Visual Search Patterns with EMD Metric in Normalized Anatomical Space. *IEEE Trans Med Imaging.* 25(8), 1011-1021.
8. Henderson J, Malcom G and Schandl C. (2009). Searching in the Dark Cognitive Relevance Drives Attention in Real-world Scenes. *Psychonomic B. & R.* 16(5), 850-856.
9. Castelhan M, Mack M and Henderson J. (2009). Viewing Task Influences Eye Movement Control during Active Scene Perception. *J. Vision.* 9(3), 1-15.
10. Liang Z, Fu H, Zhang Y, Chi Z, et al. (2010). Content-based Image Retrieval Using a Combination of Visual Features and Eye Tracking Data. In: *Symposium on Eye-Tracking Research and Applications - ETRA 2010*, ACM, Texas. 41-44.
11. Li R, Curhan J and Hoque ME. (2015). Predicting Computer-Mediated Negotiation Outcomes based on Modeling Facial Expression Synchronization. In: *Proceedings of 11th IEEE International Conference on Automatic Face and Gesture Recognition-FGR.* IEEE Press, Ljubljana. 1-6.
12. Li R, Shi P and Haake AR. (2013). Image Understanding from Experts' Eyes by Modeling Perceptual Skills of Diagnostic Reasoning Processes. In: *IEEE Conf. on Computer Vision and Pattern Recognition - CVPR 2013.* 2187-2194.
13. Li R, Pelz J, Shi P, Alm CO, et al. (2012). Learning Eye Movement Patterns for Characterization of Perceptual Expertise. In: *Symposium on Eye Tracking Research and Applications - ETRA 2012*, ACM, California. 393-396.
14. Li R, Pelz J, Shi P and Haake AR. (2012). Learning Image-Derived Eye Movement Patterns to Characterize Perceptual Expertise. In: *Proceedings of the 34th annual meeting of the Cognitive Science Society CogSci 2012*, IEEE, California. 190-195.
15. Krupinski E, Tillack A, Henderson J, Bhattacharyya A, et al. (2006). Eye-Movement Study and Human Performance using Telepathology Virtual Slides: Implications for Medical Education and Differences with Experience. *J. Human Pathology.* 37(12), 1543-1556.
16. Manning D, Ethell S, Donovan T and Crawford T. (2007). How do Radiologists do It? The Influence of Experience and

Training on Searching for Chest Nodules. *Radiography*. 12(2), 134-142.

17. Li R, Shi P, Pelz J, Alm CO, et al. (2016). Modeling Eye Movement Patterns to Characterize Perceptual Skill in Image-based Diagnostic Reasoning Processes. *Computer Vision and Image Understanding*.

18. Ren X and Malik J. (2003). Learning a classification model for segmentation. In: 9th IEEE International Conference on Computer Vision, IEEE Press, New York. 10-17.

19. Mori G, Ren X, Efros AA and Malik J. (2004). Recovering Human Body Configurations: Combining Segmentation and Recognition. In: IEEE Conf. on Computer Vision and Pattern Recognition - CVPR 2004, IEEE Press, New York. 326-333.