Image Understanding from Experts' Eyes by Modeling Perceptual Skill of Diagnostic Reasoning Processes

Rui Li Pengcheng Shi Anne R. Haake
Rochester Institute of Technology, 1 Lomb Memorial Drive, Rochester NY, 14623 USA
rxl5604@rit.edu, spcast@rit.edu, anne.haake@rit.edu

Abstract

Eliciting and representing experts' remarkable perceptual capability of locating, identifying and categorizing objects in images specific to their domains of expertise will benefit image understanding in terms of transferring human domain knowledge and perceptual expertise into image-based computational procedures. In this paper, we present a hierarchical probabilistic framework to summarize the stereotypical and idiosyncratic eye movement patterns shared within 11 board-certified dermatologists while they are examining and diagnosing medical images. Each inferred eye movement pattern characterizes the similar temporal and spatial properties of its corresponding segments of the experts' eye movement sequences. We further discover a subset of distinctive eye movement patterns which are commonly exhibited across multiple images. Based on the combinations of the exhibitions of these eye movement patterns, we are able to categorize the images from the perspective of experts' viewing strategies. In each category, images share similar lesion distributions and configurations. The performance of our approach shows that modeling physicians' diagnostic viewing behaviors informs about medical images' understanding to correct diagnosis.

1. Introduction

There has been significant progress in automatic algorithms for image understanding [10, 16, 13, 20, 9, 6]. However, when the cues in images are not sufficient to generate a good interpretation automatically, active learning methods are necessary in terms of incorporating human perceptual capability into this process [23, 14, 1, 15, 8].

On the other hand, image understanding in knowledge-rich domains is more challenging, since complex perceptual and conceptual processing are engaged to transform image pixels into meaningful contents [12]. Active learning methods via manually marking and annotating become not only labor intensive for experts but also ineffective because of the variability and noise of experts' performance [15, 7]. To address this problem, We propose to combine perceptual expertise as effortless yet valuable cognitive resources into image understanding. This requires the ability of extracting and representing experts' perceptual expertise in a form that is ready to be applied in active learning schemes. In this work, our contributions are: first, we summarize and represent expertise-related eye movement patterns shared among

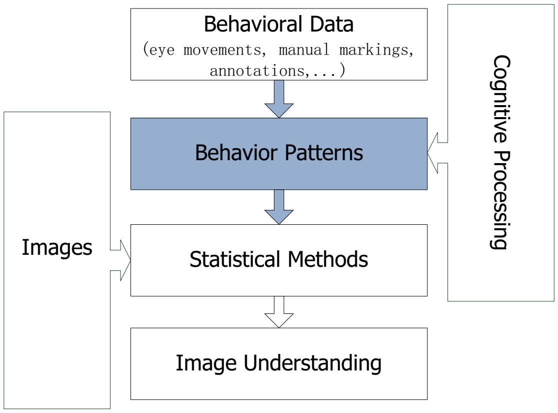


Figure 1: Paradigm of our approach. Automatic image understanding approaches attempt to interpret images solely based on statistical or optimization analysis of image pixel values [10, 16, 13, 20, 9, 6]. Recently researchers start incorporating human interactions into image understanding through active learning methods [23, 14, 1, 15, 8]. For domain-knowledge-required images, active learning methods are ineffective because of the variability and noisy nature of the human behavioral data. We thus propose that novel approach to extract tacit knowledge from experts engaging in these observable behaviors will be a more effective way to incorporate human capabilities. The extracted behavior patterns are not only more robust and consistent but also shed light on latent cognitive processing.

multiple experts in an objective and unbiased way; second, based on a subset of distinctive patterns shared across images, we semantically categorize medical images at diagnostic level bypassing the segmentation and the processing of individual lesions or regions.

Perceptual skill is considered to be the crucial cognitive factor accounting for the advantage of highly trained experts [12]. Experts generate distinctively different perceptual representations when they view the same medical images as novices [19]. Rather than passively ”photocopying” the visual information directly from sensors into minds, visual perception actively interprets the information by altering perceptual representations of the images based on experience and goals. Without guidance of perceptual skill, medical images cannot be interpreted effectively solely based on image visual features. This motivates us to investigate how to formalize perceptual skill and reason about image contents from experts' points of view as shown in Figure 1. In our work we focus on medical images [1] where domain knowledge and perceptual expertise are in demand. We elicit and model physicians' perceptual skill from their diagnostic reasoning process while inspecting medical images. Physicians examine and diagnose medical images, and their eye movements are recorded. In order to summarize the stereotypical and idiosyncratic eye movement patterns shared among these physicians, we develop a hierarchical dynamic model. It allows us to discover eye movement patterns exhibited by physicians' time-evolving eye movement sequences, and each eye movement pattern essentially characterizes a particular statistical regularity of the temporal-spatial properties inferred from multiple eye movement sequences. In particular, we specify a subset of distinctive patterns corresponding to image visual-spatial structures at pathological level. Based on the exhibitions of these patterns by physicians viewing a particular image, we are able to put the image into a category associated with a particular lesion distribution.

2. Related Work

Image understanding is approached broadly from two levels of description. From one level, a scene is viewed as configuration of objects, so a better performance can be achieved through recognizing objects and their spatial arrangement [21]. The other perspective considers a scene as a holistic representation with a unitary shape [18]. Besides locating and identifying the objects of interest in an image by bounding boxes or image segments with semantic labels, recent image understanding studies also aim at exploring the underlying scene structure by estimating a qualitative 3D layout of the scene to recover the spatial relationships among multiple objects in the original 3D space [10, 16, 13, 20]. These geometric approaches approximate the 3D space by planar surfaces or volumes from monocular images and some of them extend the idea to combine global consistency constraints [9]. Dynamic 3D scene reconstruction is another focus. Various computation approaches such as Markov random field and generative non-parametric graphical models are developed to robustly infer the 3D layout of the roads, the location of the buildings as well as dynamic traffic in the scene [6, 4].

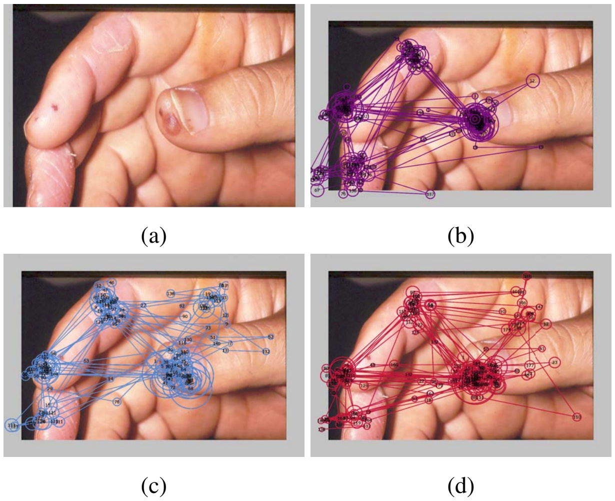


Figure 2: An illustrated dermatological image examined by the subjects. (a) The original image. (b)-(d) Three subjects' eye movement sequences super-imposed onto the image. Each circle center represents a fixation location and the radius is proportional to the duration time on that particular fixation. A line connecting two fixations represents a saccade. Fixations are numbered according to their order. Image used with permission from Logical Images, Inc.

There is significant success with the above automatic algorithms. However, human interaction becomes critical when key information such as strong edges and lines cannot be detected easily [15, 9]. To borrow human perceptual power, active learning was proposed and benefited a broad range of computer vision applications [23, 14, 1, 15, 8]. Essentially, the advantage of active learning methods relies on combining human capability of image understanding with rich information from images using machine learning approaches. This is particularly important when images are domain-specific. Some of these active learning studies attempted to maximize the knowledge gain from users while valuing their effort [23]. Others strived to simplify human interaction by fully utilizing the automatic algorithms and providing intuitive scribbles [15, 1].

Human viewing behaviors are valuable yet effortless resources worth of exploiting through active learning schemes

[1] Some dermatological images may be disturbing. Readers' discretion is required.

for image understanding. In particular, in specific domains experts perceptual expertise is considered to be more consistent and informative than their manual markings. Human vision is an active dynamic process in which the viewer seeks out specific information to support ongoing cognitive and behavioral activity [11]. Since visual acuity is limited to the fovea region and resolution fades dramatically in the periphery, we move our eyes to bring a portion of the visual field into high resolution at the center of gaze. Studies have shown that visual attention is influenced by two main sources of input: bottom-up visual attention driven by low-level saliency image features and top-down process in which cognitive processes, guided by the viewing task and scene context, influence visual attention [17, 24, 3]. Growing evidence suggests that top-down information dominates the active image viewing process and the influence of low-level salience guidance is minimal [2, 18].

These theoretical outcomes provide us with the possibility to facilitate image understanding by incorporating experts' viewing strategies through active learning paradigm. What's more, we combine multiple experts' strength by summarizing their shared eye movement patterns and decode the patterns' semantic meanings. These results can also be applied as semantic labeling without manually hand-marking with respect to image understanding.

3. Hierarchical Dynamic Model

A hierarchically-structured dynamic model was developed to capture the stereotypical and idiosyncratic eye movement patterns shared among multiple expertise-specific groups of subjects, as well as to provide the flexibility of learning new patterns from observed eye movement data in a non-parametric way.

3.1. Hierarchical prior

The hierarchical beta-Bernoulli processes proposed by Thibaux *et al.* [22] is a suitable tool to describe the situation where multiple groups of subjects are defined by countable infinite shared features following the Levy measure. We utilize this combinatorial stochastic process in the following specification based on our problem scenario, so that we can treat the number of shared eye movement patterns as a random number which is learned from the observed eye movement data.

Let B_0 denote a fixed continuous random base measure on a space Θ which represents a library of all the potential eye movements patterns. For multiple groups to share patterns, let B denote a discrete realization of a beta process given the prior $BP(c_0, B_0)$. Let $\{G_j\}_{j=1}^{N}$ be a discrete random measure on Θ drawn from B following the beta process which represents a random measure on the eye movement patterns shared among multiple subjects within the group j. Let $\{P_{ij}\}_{i=1}^{N_j}$ denote a Bernoulli measure given the beta process G_j. P_{ij} is a binary vector of Bernoulli random variables representing whether a particular eye movement pattern exhibited in the eye movement data of subject i within group j. This hierarchical construction can be formulated as follow:

$$B|B_0 \sim BP(c_0, B_0) \quad G_j|B \sim BP(c_j, B) \tag{1}$$

$$P_{ij}|G_j \sim BeP(G_j) \quad P_{ij} = \sum_k p_{ijk}\delta_{\theta_{jk}} \tag{2}$$

where $G_j = \sum_k g_{jk}\delta_{\theta_{jk}}$. This term shows that G_j is associated with both a set of countable number of eye movement patterns $\{\theta_{jk}\}$ drawn from the eye movement pattern library Θ and their corresponding probability masses $\{g_{jk}\}$ given group j. The combination of these two variables characterizes how the common eye movement patterns shared among subjects within expertise-specific group j. Thus P_{ij} is a Bernoulli process realization from the random measure G_j where p_{ijk} as a binary random variable denotes whether subject i within group j exhibits eye movement pattern k given probability mass g_{jk}. Based on the above formulation, for $k = 1...K_j$ patterns we readily define $\{(\theta_{jk}, g_{jk})\}$ as a set of common eye movement patterns shared among group j and $\{(\theta_{jk}, p_{ijk})\}$ as subject i's personal subset of eye movement patterns given group j.

The transition distribution $\pi_{ij} = \{\pi_{z_t^{(ij)}}\}$ of the Hidden Markov Model (HMM) at the bottom level governs the transitions between the i^{th} subject's personal subset of eye movement patterns θ_{jk} of group j. It is determined by the element-wise multiplication between the eye movement subset $\{p_{ijk}\}$ of subject i in group j and the gamma-distributed random variables $\{e_{ijk}\}$:

$$e_{ijk}|\gamma_j \sim Gamma(\gamma_j, 1) \quad \pi_{ij} \propto E_{ij} \bigotimes P_{ij} \tag{3}$$

where $E_{ij} = [e_{ij1}, ...e_{ijK_j}]$. So the effective dimensionality of π_{ij} is determined by P_{ij}.

3.2. Dynamical likelihoods

We apply one autoregressive HMM as the likelihood to describe the dynamics of each subject's eye movement sequence. This model is proposed to be a simpler but often effective way to describe dynamical systems [5]. Let $y_t^{(ij)}$ denote the observation unit of the eye movement sequence at time step t of the i^{th} subject in the j^{th} group. We associate each time-step's observation with one fixation and its successive saccade as one observation unit. Let $x_t^{(ij)}$ denote the corresponding latent dynamic mode. We have

$$x_t^{(ij)} \sim \pi_{x_{t-1}^{(ij)}} \tag{4}$$

$$y_t^{(ij)} = A_{x_t^{(ij)}}\tilde{y}_t^{(ij)} + e_t(x_t^{(ij)}) \tag{5}$$

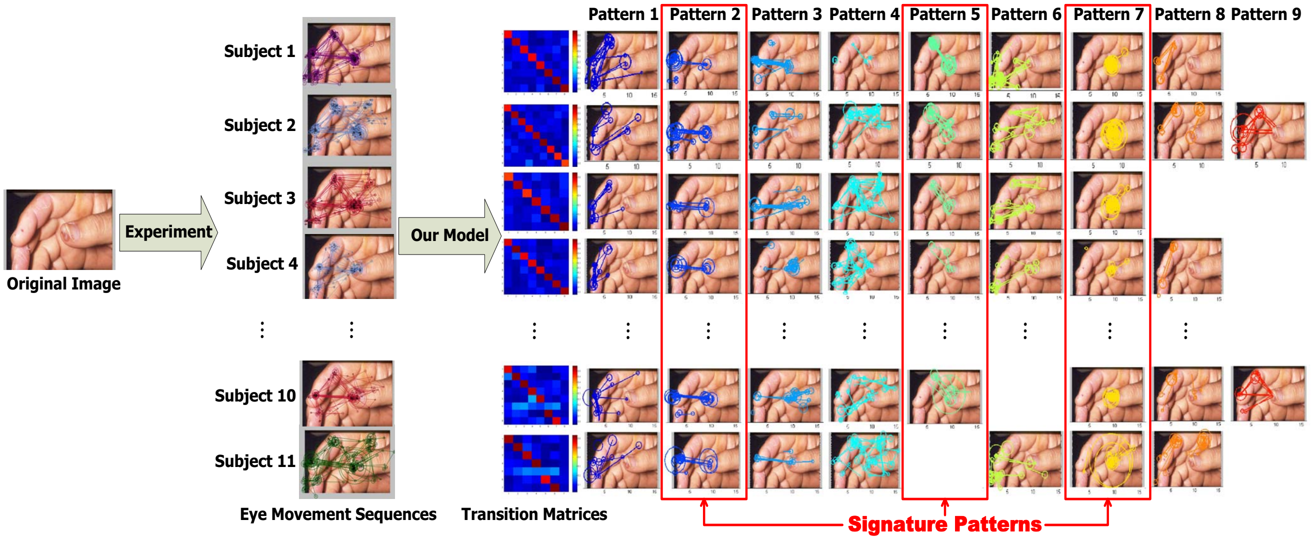


Figure 3: Six out of eleven eye movement sequences super-imposed onto one dermatological image are illustrated here. Our model respectively decomposes the eleven eye movement sequences into nine eye movement patterns (color-coded) with the transition probability matrices, so that each sequence can be represented by a certain number out of nine patterns and their corresponding transition matrix. On the right, it is the shared eye movement pattern matrix of which each row corresponds to a subject's eye movement sequence and each column indicates one shared eye movement pattern among multiple subjects. In this case, three patterns are recognized as *Signature Patterns* based on their self-transition probabilities, temporal-spatial properties and diagnostic semantics.

where $e_t^{(ij)}(k) \sim N(0, \Sigma_k)$ which is an additive white noise, $A_k = [A_{1,k}, ..., A_{r,k}]$ as the set of lag matrices, and $\tilde{y}_t^{(ij)} = [y_{t-1}^{(ij)}, ..., y_{t-r}^{(ij)}]$. In our case, we specify $r = 1$.We thus define $\theta_k = (A_k, \Sigma_k)$ as one eye movement pattern.

We use Markov chain Monte Carlo method to do the posterior inference. Based on the sampling algorithm proposed in [22], we developed a Gibbs sampling solution to sample the marginalized hierarchical beta processes part of the model. Although our model is capable of profiling eye movement patterns shared among multiple expertise-specific groups, we focus on expert group's performance and its potential contribution to image understanding.

4. Eye Tracking Experiment

Eleven board-certified dermatologists (attending physicians) with normal or corrected to normal vision participated for monetary compensation. A SMI (Senso-Motoric Instruments) eye tracking apparatus was applied to display the stimuli at a resolution of 1680x1050 pixels for the collection of eye movement data and recording of verbal descriptions. The eye tracker was running at 50 Hz sampling rate and has reported accuracy of 0.5^o visual angle. The subjects viewed the medical images binocularly at a distance of about 60 cm. The experiment was conducted in an eye tracking laboratory with ambient light.

A set of 50 dermatological images, each representing a different diagnosis, was selected for the study. These images were presented to subjects on the monitor. Medical professionals were instructed to examine and describe each image to the students while working towards diagnosis, as if teaching. The experiment lasted approximately 1 hour. The subjects were instructed not only to view the medical images and make a diagnosis, but also to describe what they see as well as their thought processes leading them to the diagnosis. Both eye movements and verbal descriptions were recorded for the viewing durations controlled by each subject. The experiment started with a 13-point calibration and the calibration was validated after every 10 images.

5. Image Analysis through Signature Patterns

We generate 387 eye movement patterns based on eleven subjects examining and diagnosing fifty dermatological images. These results allow us to analyze images from a novel perspective of experts' perceptual strategies.

5.1. Eye movement pattern estimation

In Figure 3, we illustrate one set of observed eye movement sequences and estimating processes from our model of the eleven dermatologists diagnosing a case of a skin manifestation of endocarditis. In the medical image, there are multiple skin lesions spreading over the thumb nail and tip, the two parts of index finger and the middle finger. A primary abnormality is on the thumb tip. The eye movement sequences in Figure 2 indicate that dermatologists examine the image in a highly patterned manner by fixating on the primary abnormality heavily and switching their visual attention actively between and within the primary and secondary abnormalities. Our model decomposes each eye movement sequence into several subsets of its segments. Each subset is characterized by one estimated latent state and a Gaussian emission distribution which summarizes the similar temporal-spatial properties shared among multiple sequences, as described in Equation 5. The way that the patterns are shared among the subjects is also indicated by their matrix in Figure 3. For example the first subject's eye movements evolve over time with the first eight out of nine patterns, and the eleventh subject has seven patterns except pattern 5 and pattern 9. Transition probability matrices indicated these patterns are persistent with high self-transition probabilities. Although such analysis estimates varied image-specific patterns, we discover several basic yet distinctive types of patterns shared across multiple images called *Signature Patterns*.

5.2. Signature pattern recognition

We define a type of signature patterns by three criteria: first, its self-transition probability, which is indicated by the transition matrix, is no less than 0.65; second, it manifests clear diagnostic regions; third, the temporal-spatial properties of signature pattern exemplars within each type are similar but distinctive from other types, which is depicted in Figure 4. In the illustrated case in Figure 3, there are three instantiations of the signature patterns recognized. Pattern 2 and Pattern 5 is characterized by fixations switching back and forth between the primary and the different secondary abnormalities with long saccade amplitudes and relatively short fixation durations. These patterns suggest that subjects compare and associate the two types of abnormalities. Pattern 7 is characterized by a series of long-duration fixations only on the primary abnormality with extremely short saccades. This pattern suggest that subjects fixate on primary abnormality to make diagnosis.

Based on the eye movement patterns generated from our model over fifty images, we are able to specify three types of signature patterns. The first type is named as *Concentrating Pattern* which is characterized by a series of long-duration fixations and short-amplitude saccades usually fixating on primary abnormalities; the second is *Switching Pattern* characterized by a series of relatively short-duration fixations and long-amplitude saccades usually switching back and forth between two abnormalities; and the third is *Clutter Pattern* characterized by a series of shorter fixations and relatively long saccades usually scanning within localized abnormal regions. To quantify the temporal-spatial properties of the three types of signature patterns, we illustrate some of their exemplars in Figure 4. The estimation of the signature patterns based on their exemplar features can be solved using different classification techniques. We first adopt quadratic discrimination analysis (QDA) by assuming a simple parametric model for the densities of the temporal-spatial properties of the eye movement units and a training set includes 217 eye movement units of 12 exemplar patterns from 10 images. Their temporal-spatial properties are shown in Figure 4. We test the validity of the classifier through comparing the image categorization performance based on QDA with K nearest neighbors (K-NN) and experts' performance.

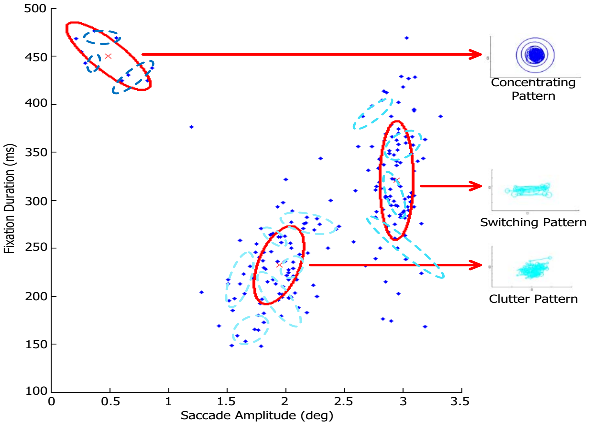


Figure 4: Distinctive temporal-spatial properties of 217 eye movement units from 12 exemplars forms the three types of signature patterns. Each blue dot represents one eye movement unit from a signature pattern exemplar. The exemplars are indicated by dash-line emission distributions estimated from our model. Both eye movement units and their corresponding exemplars are projected from a four-dimension space (including x-y coordinate, fixation duration and saccade amplitude) onto this space. The signature patterns are characterized by a three-component Gaussian mixture. The one on the upper left represents *Concentrating Pattern*, the one on the right captures *Switching Pattern*, and the one on the lower middle represents *Clutter Pattern*. For each type, we project the units back into x-y coordinate space centered on the origin and visualize them on the right side.

5.3. Perceptual category specification

Three additional experienced board-certified dermatologists as our consultants suggests four broad perceptual categories in terms of lesion distribution and configuration. We further determine the associations between the combinations of the exhibitions of these three types of signature

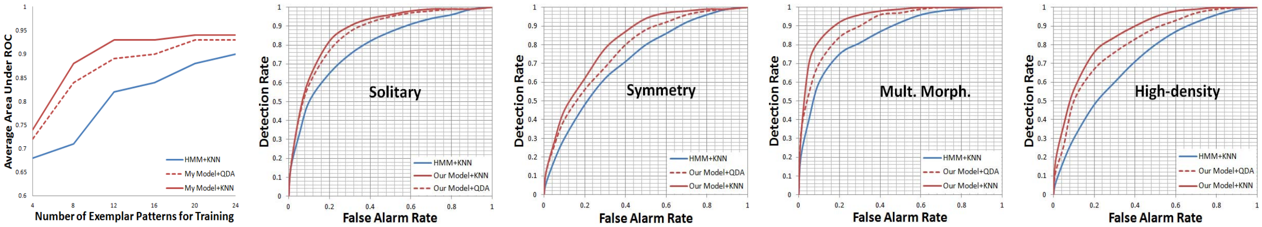


Figure 5: ROC curves summarizing categorization performance for the four perceptual categories. Left: Area under average ROC curves for different numbers of exemplar patterns. Right: We compare our model using two different classification techniques with canonical Hidden Markov Models.

patterns and the four specified categories:

- If the set of eye movement patterns exhibited on an image solely includes *Concentrating Patterns*, the image is categorized as *Solitary* which means that the image contains a solitary lesion as primary abnormality.
- If the set of eye movement patterns exhibited on an image solely includes *Switching Patterns*, the image is categorized as *Symmetry* which means that the lesions in the image are symmetrically distributed.
- If the set of eye movement patterns exhibited on an image includes both *Concentrating Patterns* and *Switching Patterns*, the image is categorized as *Multiple Morphologies* which means that the lesions in the image belong to different morphologies and usually one lesion are primary abnormalities and others are secondary ones.
- If the set of eye movement patterns exhibited on an image includes *Clutter Patterns*, the image is categorized as *High-Density Lesions* which means that the image contains multiple lesions distributed in either scattered or clustered manner.

According to the signature patterns recognized on the images, we can put them into the four categories as shown in Figure 6a, 6b, 6c, and 6d.

6. Results and Discussion

To measure the performance of our image categorization approach, we conduct an experiment following the same procedure by recruiting another ten dermatologists and using a different set of forty dermatological images as stimuli. These images are randomly selected. Our three consulting dermatologists achieve consensus to categorize the forty images into the four perceptual categories. We use 232 estimated eye movement patterns on these images and the ones from the previous experiment as a testing set. In Figure 5, we examine categorization performance given training sets containing between 4 and 24 exemplars. We assume each eye movement sequence exhibits the same set of patterns in order to implement the canonical HMMs. We see that our model lead to significant improvements in categorization performance, particularly when few training exemplars are available. The highest accuracy is achieved on detection of the *Multiple Morphologies* category. This may be caused by the requirement of detections of the two different *Signature Patterns* to determine the varied distributions and significance of the lesions. The difference between *Multiple Morphologies* images and *Symmetry* images is that the eye movement patterns exhibited on the latter do not contain *Concentrating Pattern*. This is because the symmetrical visual-spatial structures imply that lesions are equivalent important without single primary one for the subjects to concentrate their focus on as shown in Figure 6b. Since the specifications of signature patterns are heuristic, we may be able to improve the categorization performance by identifying extra meaningful and distinctive eye movement patterns, and these extra patterns may also lead to image categorization at a finer detailed level.

Since the dermatological images are collected for future diagnosis, and training purposes, the dermatologists took them in a particular way. They tend to center primary abnormalities and preserve as much related contextual information as possible, such as patients' demographic information, body parts, lesion size and so on. Nonetheless, these high-resolution images have complex backgrounds, and large appearance variations for luminance and camera angles. These factors cause some false alarms. In particular, scales of some lesions in the images tend to influence our model's performance. For instance, the solitary lesions have large scales in some images, this leads to cluttered eye movement patterns rather than concentrating ones as shown in Figure 6d. Since both the number of fixations and their durations are indicative of the depth of information processing associated with the particular image regions, the exhibition of *Concentrating Pattern* usually corresponds to a localized primary abnormality as shown in Figure 6a and 6c, which is the most important cue for correct diagnosis. The saccade amplitudes of *Switching Pattern* and *Clutter Pattern* inform dermatologists' visual comparison or association during examining images based on both the image visual-spatial structures (symmetry, *e.g.*) as in Figure 6b and distributions of multiple abnormalities (primary abnormality versus secondary abnormality, *e.g.*) as in Figure 6d.

We obtain certain aspects of experts' domain-specific knowledge by summarizing their perceptual skills from their eye movements while diagnosing images. The domain-specific knowledge unveils the meaning and significance of the visual cues as well as the relations among functionally integral visual cues without segmentation or processing of individual objects or regions. This will benefit the traditional pixel-based statistical methods for image understanding by evaluating perceptual significance and relations of the image features which spatially correspond to the eye movement patterns. This combination of expert knowledge and image features allows us to generalize our approach to images on which there is no experts' eye movements recorded.

The different viewing times of dermatologists yield length-varying eye movement sequences. Since each sequence is modeled with one HMM separately, the emission distributions of which group multiple fixation-saccadic units into one pattern exhibited repeatedly. Thus longer sequence means that its corresponding longer HMM draws more pattern samples from the prior distribution, so besides containing more repeated common patterns, it likely has some unique patterns.

7. Conclusions

This paper presents a hierarchical probabilistic dynamic framework to summarize eye movement patterns shared among dermatologists while they are examining medical images. This novel approach allows us to elicit perceptual skill as additional human capabilities to achieve image understanding at the pathological level.

References

[1] D. Batra, A. Kowdle, and D. Parikh. icoseg: Interactive co-segmentation with intelligent scribble guidance. In *CVPR*, pages 3169–3176, 2010.

[2] M. S. Castelhano, M. L. Mack, and J. M. Henderson. Viewing Task Influences Eye Movement Control during Active Scene Perception. *J. Vision*, 9(3):1–15, 2009.

[3] J. Chen and Q. Ji. Probabilistic gaze estimation without active personal calibration. In *Proc. of CVPR*, pages 609–616, 2011.

[4] D. Crandall, A. Owens, N. Snavely, and D. Huttenlocher. Discrete-continuous optimization for large-scale structure from motion. In *CVPR*, pages 3001–3008, 2011.

[5] E. B. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky. Bayesian nonparametric methods for learning markov switching processes. *IEEE Signal Processing Magazine*, 27(6):43–54, 2010.

[6] A. Geiger, M. Lauer, and R. Urtasum. A generative model for 3d urban scene understanding from movable platforms. In *CVPR*, pages 1945–1952, 2011.

[7] S. Gordon, S. Lotenberg, J. Jeronimo, and H. Greenspan. Evaluation of uterine cervis segmentations using ground truth from multiple experts. *J. Computerized Medical Imaging and Graphics*, 33(3):205–216, 2009.

[8] P. H. Gosselin and matthieu Cord. Actively learning methods for interactive image retrieval. *IEEE Trans. on Image Processing*, 17(7):1200–1211, 2008.

[9] A. Gupta, S. Satkin, A. A. Efros, and M. Hebert. From 3d scene geometry to human workspace. In *CVPR*, pages 1961–1968, 2011.

[10] V. Hedau, D. Hoiem, and D. Forsyth. Recovering the spatial layout of cluttered rooms. In *ICCV*, pages 1849–1856, 2009.

[11] J. M. Henderson and G. L. Malcolm. Searching in the Dark Cognitive Relevance Drives Attention in Real-world Scenes. *Psychonomic Bulletin and Review*, 16(5):850–856, 2009.

[12] R. Hoffman and M. S. Fiore. Perceptual (re)learning : a leverage point for human-centered computing. *J. Intelligent Systems*, 22(3):79–83, 2007.

[13] D. Hoiem, A. A. Efros, and M. Hebert. Recovering surface layout from an image. *IJCV*, 75(1):151–172, 2007.

[14] A. Kapoor, K. Grauman, R. Urtasun, and T. Darrell. Active learning with gaussian processes for object categorization. In *ICCV*, pages 1–8, 2007.

[15] A. Kowdle, Y.-J. Chang, A. Gallagher, and T. Chen. Active learning for piecewise planar 3d reconstruction. In *CVPR*, pages 929–936, 2011.

[16] D. C. Lee, M. Hebert, and T. Kanade. Geometric reasoning for single image structure recovery. In *CVPR*, pages 2136–2143, 2009.

[17] S. Marat, T. H. Phuoc, L. Granjon, N. Guyader, D. Pellerin, and A. Gurin-Dugu. Modeling spatio-temporal saliency to predicit gaze direction for short videos. *International Journal of Computer Vision*, pages 231–243, 2009.

[18] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(36):145–175, 2001.

[19] T. J. Palmeri, A. C.-N. Wong, and I. Gauthier. Computational approaches to the development of perceptual expertise. *TRENDS in Cognitive Sciences*, 8(8):378–386, 2004.

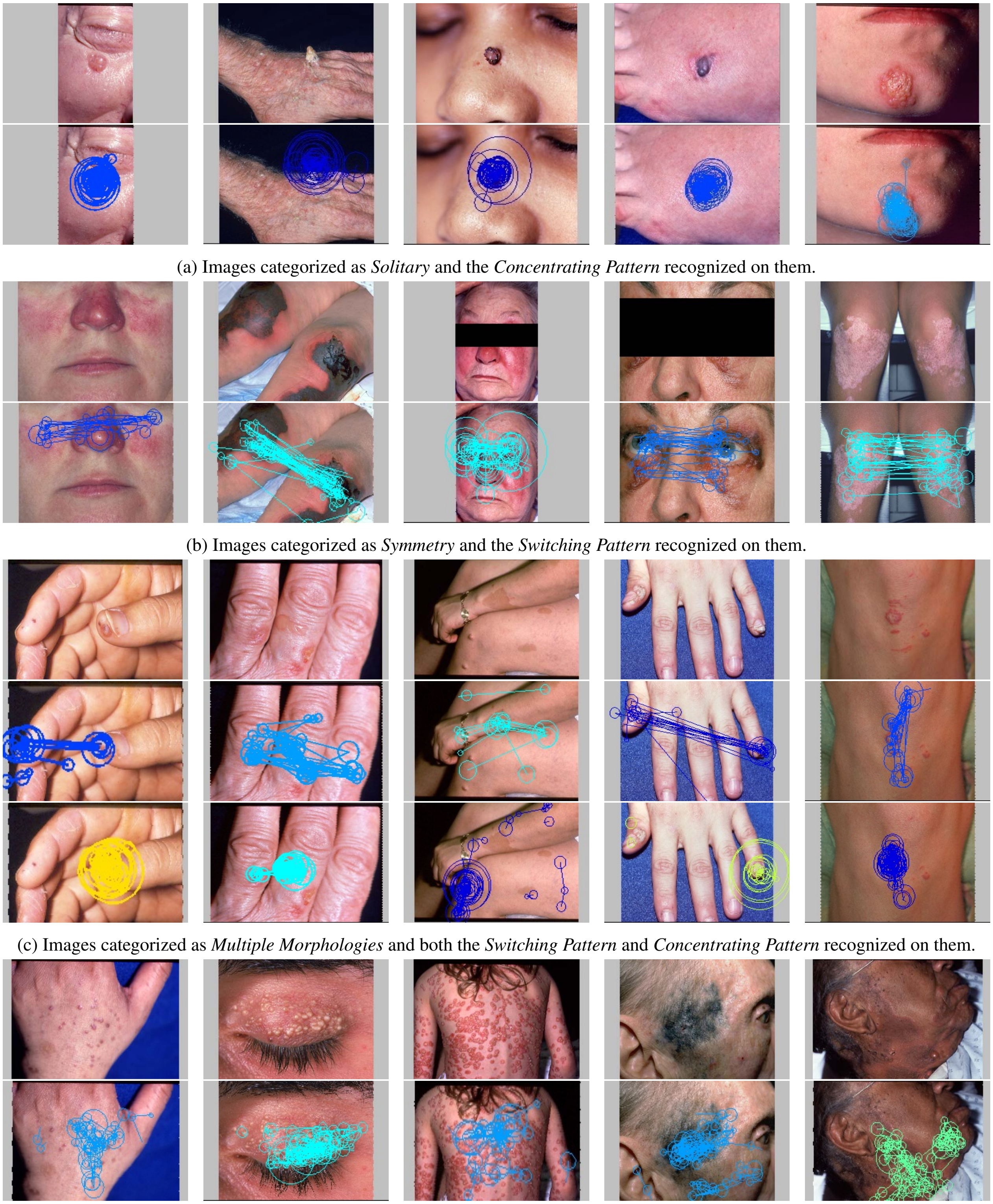
[20] A. Saxena, M. Sun, and A. Y. Ng. Learning 3d scene structure from a single still image. *PAMI*, 31(5):824–840, 2009.

[21] E. B. Sudderth, A. Torralba, W. T. Freeman, and A. S. Willsky. Describing visual scenes using transformed objects and parts. *International Journal of Computer Vision*, 77(3):291–330, 2008.

[22] R. Thibaux and M. I. Jordan. Hierarchical beta processes and the indian buffet process. *J. Machine Learning and Research*, 22(3):25–31, 2007.

[23] S. Vijayanarasimhan, P. Jain, and K. Grauman. Far-sighted actively learning on a budget for image and video recognition. In *CVPR*, pages 3035–3042, 2010.

[24] W. Wang, C. Chen, Y. Wang, T. Jiang, F. Fang, and Y. Yao. Simulating human saccadic scanpaths on natural images. In *CVPR*, pages 441–448, 2011.



(a) Images categorized as *Solitary* and the *Concentrating Pattern* recognized on them.

(b) Images categorized as *Symmetry* and the *Switching Pattern* recognized on them.

(c) Images categorized as *Multiple Morphologies* and both the *Switching Pattern* and *Concentrating Pattern* recognized on them.

(d) Images categorized as *High-density Lesions* and the *Clutter Pattern* recognized on them.

Figure 6: For each of the four categories five images are illustrated. We also demonstrate one instantiation of the signature patterns recognized from the set of subjects' eye movement patterns which is estimated by our model. Images used with permission from Logical Images, Inc.