33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Augmenting the Classification of Retinal Lesions using Spatial Distribution

Elizabeth M. Massey, Andrew Hunter

Abstract— **This paper introduces SAGE – an algorithm that uses the spatial clustering of objects to enhance their classification. It assumes that discrete objects can be identified and classified based on their individual appearance, and further that they tend to appear in spatial clusters (for example, circinate exudates). The algorithm builds spatial distribution maps for objects and confounds for a given image, and adjusts individual object confidence levels to reflect their spatial clustering. SAGE may be combined with a wide range of object identification and classification methods; we demonstrate it using a Multi-Layered Perceptron (MLP) Neural Network and a Support Vector Machine (SVM) classifier types for both dark and bright retinal lesions. Using ROC analysis SAGE improves classifier performance as much as 83%.**

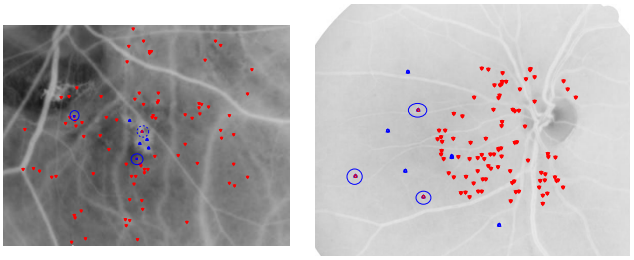
I. INTRODUCTION

This paper introduces SAGE (Spatial Augmented classification using GMM-EM) - an algorithm to augment the classification of discrete objects in an image by modeling their spatial clustering; see Fig. 1. The approach is inspired by the human capacity to exploit contextual cues to help classify objects [1]. SAGE is used when the requirement is to locate and identify multiple objects of the same type that tend to cluster together (for example, retinal lesions). It assumes that there exist techniques to find candidates (potential objects) and to estimate the probability that each is an object of interest based on its individual appearance.

Typically, there is a processing pipeline that finds candidates (e.g. by interest point detection or sliding window application), extracts a set of features reflecting their appearance (this may involve segmentation as an intermediate step), and uses a classifier to estimate the probabilities. SAGE uses the candidates' probability estimates and image positions to estimate two spatial distribution maps: one for objects of interest and one for confounds, then updates the candidates' probability estimates by using these maps. In effect, clusters of likely objects have upweighted probabilities, whereas isolated objects are downweighted. SAGE is a novel version of the Expectation Maximization (EM) algorithm for Gaussian Mixture Models (GMMs), modified to simultaneously estimate probability distributions for two separate classes.

We apply to and evaluate the algorithm for the identification of bright and dark retinal lesions and demonstrate improvements in the classification of individual objects by taking into account spatial clustering. Retinal lesions tend to appear in groups due to underlying local pathology (e.g. leaky microvasculature), thus, the presence of several high-probability objects in a region is mutually reinforcing.

This paper makes two main contributions. First, although spatial clustering is very common in many problem domains, there seems to be very little published work on algorithms that exploit the clustering behavior of objects for discrete object identification; (there are published segmentation methods exploiting spatial coherency at a pixel level but not at object level, and others exploiting figurative spatial relationships between visual features corresponding to parts of objects but not loose relationships between separate objects). Second, our implementation introduces a novel modification of GMM-EM to estimate probability density functions in a multi-class environment (GMM-EM is normally used to model a single probability density function for unlabeled data, [2], [3]), and to combine *a priori* probability estimates from outside the EM framework with this distribution.



(*a*) (*b*)

Fig. 1. SAGE correctly identifies clustered/standalone lesion (blue) and non-lesion (red) objects (a)Partial clusters, (b)Standalone lesion objects

II. RELATED WORK

As noted above, a substantial amount of research has been reported using spatial proximity in segmentation, at the level of individual pixels [4], [5], [6], [7], [8] using information from neighboring pixels. However, these approaches do not take into account the contextual cues provided by separate lesions in the vicinity. Approaches that do consider contextual information such as 'convex area', solidity' and 'orientation', to classify pixels, show success with clustering data ([6], [9]); however, they make the assumption that related pixels are homogenized within a region. In [10] *co-occurring signs of pathology* characterize ischemic retinal regions using pixel-by-pixel comparisons of appearance, temporal and contextual features in an AdaBoost framework. These methods largely perform pixel-level segmentation, and assume color homogeneity of objects. They are therefore unsuitable for

E.M.Massey & A.Hunter are both with the School of Computer Science, University of Lincoln, Brayford Pool, Lincoln, UK, LN6 7TS bmassey,ahunter@lincoln.ac.uk

identification of objects with complex appearance (e.g. objects whose foreground looks similar to the background). In addition, they segment regions but do not exploit contextual cues from separate, nearby objects. The *constellation* model [11] classifies an object based on its constituent parts. It identifies a number of visual distinctive object parts assuming a rigid (figurative) spatial relationship. The main difference is that SAGE identifies visually similar objects with non-rigid proximity. Thus, while [11] is suitable for modeling single objects with a complex composite appearance, SAGE is suited to modeling multiple objects of similar appearance that tend to appear in groups.

III. METHOD

In this paper, we apply SAGE to bright and dark retinal lesion identification. We assume that objects of interest and confounds are spatially distributed according to two separate, unknown spatial distributions (in particular, similar objects are more likely to occur in groups than not), and that both types are independently drawn from their respective spatial distributions. SAGE is applied independently to each image, and so does not model consistency of object positions across multiple images. We further assume that the appearance of objects and confounds are each independent of spatial location and image, given the class. Within the domain of retinal lesions we utilize maximum likelihood (ML) classifiers that estimates the class of candidates based on their appearance alone; these classifiers are typically trained off-line using a dataset of labeled candidates taken from a set of training images. For each new image, the candidates are first located and the ML classifier used to estimated the probability that each is an object of interest, given the appearance. Although retinal lesions tend to be either lighter or darker than the surrounding retinal tissue, there are many confounds that have similar levels of local contrast. These can be distinguished from lesions via a more sophisticated analysis of the color, texture and shape. We therefore use intensity peak detection to locate candidate lesions, level sets to segment the potential lesion around the peak [12], extract appearance features from the segmented area and then use a classifier to estimate the probability that the candidate is a lesion based on its local appearance. The 100 candidates with highest local intensity contrast in an image, and the lesion probabilities provided by the classifier together with the location are fed to SAGE to estimate spatial distribution functions for both lesions and confounds. We can model the distribution of objects using a GMM, optionally including a uniform distribution to account for isolated objects when they do not always occur in groups. Depending on the domain we may model confounds simply using a uniform distribution or using both a GMM and uniform.

The model described is a novel augmentation to the standard GMM-EM algorithm in two important respects: first, to integrate the probability estimates of the separate appearance-based classifiers; second, to simultaneously update two mixture models (one for objects, and one for confounds). The full derivation can be found in [13].

Consider an image with a number of candidate objects. Let $\mathbf{X} = \{\mathbf{x}_1..\mathbf{x}_N\}$ be the set of candidate's feature vectors, with $\mathbf{x}_i = \{\mathbf{q}_i, \mathbf{v}_i\}$ the i^{th} candidate; $\mathbf{q}_i$ represents the i^{th} candidate's spatial position (x,y) in the image and $\mathbf{v}_i$ represents the i^{th} candidate's appearance feature vector. Let c_i be an indicator variable giving the class of the i^{th} candidate where $c_i = 1$ implies that the i^{th} candidate is an object of interest, $c_i = 0$ that it is a confound. We wish to determine the probability that each candidate is an object, given the observed data, $p(\mathbf{x}_i) = p(c_i = 1|\mathbf{X})$.

We model the spatial distributions of the lesions and non-lesions using two mixture models:

$$p(\mathbf{q}_i|c=0,\Theta_1) = \sum_{j=1}^{R} \alpha_j p_j(\mathbf{q}_i|\theta_j) \tag{1}$$

$$p(\mathbf{q}_i|c=1,\Theta_2) = \sum_{j=R+1}^{M} \alpha_j p_j(\mathbf{q}_i|\theta_j) \tag{2}$$

where p_j is a component density parameterized by θ_j, and α_j is the corresponding mixing coefficient, such that $\sum_{j=1}^{R} \alpha_j = 1$, $\sum_{j=R+1}^{M} \alpha_j = 1$, and the set of parameters Θ is the union of the disjoint subsets $\Theta_1 = \{\alpha_j, \theta_j : j \leq R\}$ and $\Theta_2 = \{\alpha_j, \theta_j : j > R\}$ for the confound and object mixture models respectively.

We posit the existence of the hidden variable vector, $\mathbf{y} = \{y_i\}$, $i \in (1,N)$, where $y_i \in (1,M)$ indicates which mixture component generated the i^{th} data point, and $\mathscr{Y}$ denotes the set of all possible values of $\mathbf{y}$, $\mathbf{y} \in \mathscr{Y}$. Thus, with the inclusion of the hidden variable $\mathbf{y}$ our data can be described in its complete form as $\mathbf{z} = \{\mathbf{x},\mathbf{y}\} = \{\mathbf{q},\mathbf{v},\mathbf{y}\}$.

The representation of y_i above is key in allowing us to use EM-GMM, a method usually associated with modeling a single unlabeled pdf. The hidden variable identifies the dominant mixture component of a candidate across the two distributions; as each component is assigned to a single class it therefore also implicitly identifies the candidate class. We can then write the complete data log-likelihood function for Θ given the dataset and hidden variables as:

$$\ln \mathscr{L}(\Theta|\mathbf{X},\mathbf{y}) = \sum_{i=1}^{N} \ln \left(p(\mathbf{v}_i) \left(\sum_{j=1}^{M} \delta_{j,y_i} \alpha_j p_j(\mathbf{q}_j|\theta_j) f(\mathbf{v}_j, j) \right) \right) \tag{3}$$

where $p(\mathbf{v}_i)$ is a prior for appearance that disappears in the optimization as it is identical for either class; δ_{j,y_i} is the Kronecker delta such that when $j = y_i$ this indicates the component that x_i belongs to, and $f(\mathbf{v},i)$ is defined by:

$$f(\mathbf{v},i) = \begin{cases} 1 - g(\mathbf{v}) & i \leq R \\ g(\mathbf{v}) & i > R \end{cases} \tag{4}$$

where $g(\mathbf{v})$ is the ML appearance-based estimator for $p(c = 1|\mathbf{v})$. This is a key innovation allowing us to integrate the probability estimate of the separate appearance-based classifier into the spatial distribution framework. It can be provided using any suitable probabilistic classifier (we use an MLP and an SVM) that is trained to distinguish lesions from non-lesions given $\mathbf{v}_i$ as an input . The definition of f

above is a second key innovation which allows us to treat the probability estimates differently for the object and confound classes (inverting the object probability to determine the confound probability), and so treating the two mixture distributions differently. We begin the EM algorithm with the form in [14]. We can iteratively estimate the parameters Θ^t at iteration t, by maximizing the function:

$$\begin{aligned} & Q(\Theta^t, \Theta^{t-1}) \\ & = \sum_{j=1}^{M}\sum_{i=1}^{N} \ln(p(\mathbf{v}_i)) p(j|\mathbf{x}_i, \Theta^{t-1}) \\ & + \sum_{j=1}^{M}\sum_{i=1}^{N} \ln\left(f(\mathbf{v}_i, j)\right) p(j|\mathbf{x}_i, \Theta^{t-1}) \\ & + \sum_{j=1}^{M}\sum_{i=1}^{N} \ln\left(\alpha_j\right) p(j|\mathbf{x}_i, \Theta^{t-1}) \\ & + \sum_{j=1}^{M}\sum_{i=1}^{N} \ln\left(p_j(\mathbf{q}_i|\theta_j)\right) p(j|\mathbf{x}_i, \Theta^{t-1}) \end{aligned} \tag{5}$$

And maximize the terms in (5) containing α_j and $p_j(\mathbf{q}_i|\theta_j)$ independently.

In the E-step, we calculate the values $p(j|\mathbf{x}_i, \Theta^{t-1})$, using the parameter estimates from the previous iteration, thus:

$$p(j|\mathbf{x}_i, \Theta^{t-1}) = \frac{p_j(\mathbf{q}_i|\Theta^{t-1}) p(j|\mathbf{v}_i, \Theta^{t-1}) p(\mathbf{v}_i)}{p(\mathbf{x}_i|\Theta^{t-1})} \tag{6}$$

Noting that $p(j|\mathbf{v}_i, \Theta^{t-1}) = \alpha_j f(\mathbf{v}_i, j)$, expanding the denominator $p(\mathbf{x}_i|\Theta^{t-1})$ and canceling the common factor $p(\mathbf{v}_i)$ from the numerator and denominator, we get:

$$p(j|\mathbf{x}_i, \Theta^{t-1}) \propto \alpha_j p_j(\mathbf{q}_i|\Theta^{t-1}) f(\mathbf{v}_i, j) \tag{7}$$

The novel inclusion of the term $f(\mathbf{v}_i, j)$ takes into account the appearance of the candidate, and whether the mixture component is a member of the object or confound mixture. In the M-step, we optimize the mixture component parameters, α_j and θ_j. To find the expression for α_j, we consider the confound and object components separately. Considering first the confounds ($j \leq R$), we introduce the Lagrange multiplier λ, with the constraint $\sum_{j=1}^{R} \alpha_j = 1$, and solve the following:

$$\frac{\partial}{\partial \alpha_j}\left[\sum_{j=1}^{M}\sum_{i=1}^{N} \ln\left(\alpha_j\right) p(j|\mathbf{x}_i, \Theta^{t-1}) + \lambda\left(\sum_j \alpha_j - 1\right)\right] = 0 \tag{8}$$

which yields:

$$\sum_{i=1}^{N} \frac{1}{\alpha_j} p(j|\mathbf{x}_i, \Theta^{t-1}) + \lambda = 0 \tag{9}$$

Multiplying by α_j, summing over $j \in (1, R)$, and rearranging we obtain:

$$\lambda = -\sum_{j=1}^{R}\sum_{i=1}^{N} p(j|\mathbf{x}_i, \Theta^{t-1}) \tag{10}$$

and therefore, for $j \leq R$:

$$\alpha_j = \frac{\sum_{i=1}^{N} p(j|\mathbf{x}_i, \Theta^{t-1})}{\sum_{j=1}^{R}\sum_{i=1}^{N} p(j|\mathbf{x}_i, \Theta^{t-1})} \tag{11}$$

By a similar derivation, for $j > R$:

$$\alpha_j = \frac{\sum_{i=1}^{N} p(j|\mathbf{x}_i, \Theta^{t-1})}{\sum_{j=R+1}^{M}\sum_{i=1}^{N} p(j|\mathbf{x}_i, \Theta^{t-1})} \tag{12}$$

With a Gaussian Mixture Model, the parameters are $\theta_j = \left\{\mu_j, \Sigma_j\right\}$, with μ_j the centroid and Σ_j the covariance matrix. The derivation then follows the standard EM-derivation for GMMs, yielding:

$$\mu_j^t = \frac{\sum_{i=1}^{N} \mathbf{q}_i p(j|\mathbf{x}_i, \Theta^{t-1})}{\sum_{i=1}^{N} p(j|\mathbf{x}_i, \Theta^{t-1})} \tag{13}$$

$$\Sigma_j^t = \frac{\sum_{i=1}^{N} p(j|\mathbf{x}_i, \Theta^{t-1})(\mathbf{q}_i - \mu_j^t)(\mathbf{q}_i - \mu_j^t)^T}{\sum_{i=1}^{N} p(j|\mathbf{x}_i, \Theta^{t-1})} \tag{14}$$

We sometimes use a uniform distribution in the mixture model for one or other of the classes. The Uniform distribution returns a constant value $p_j(\mathbf{q}_j|\Theta^{t-1}) = 1/A$ in the E-step, where A is the area of the image in pixels. In the M-step, the α value of the Uniform distributions are optimized just as those for the Gaussians. To initialize SAGE we randomly assign the centroids μ_j to elements $\mathbf{q}_i$, with the mixture weights α_j given by $f(\mathbf{v}_i, j)$, normalized, and a standard 'reasonably broad' covariance matrix, and then use EM to iteratively determine the parameters in the usual fashion. Once the spatial distribution parameters have been determined, we can classify the candidates, thus:

$$p(c_i = 1|\mathbf{x}_i) = \frac{g(\mathbf{v}_i) p(\mathbf{q}_i|c_i = 1)}{g(\mathbf{v}_i) p(\mathbf{q}_i|c_i = 1) + (1 - g(\mathbf{v}_i)) p(\mathbf{q}_i|c_i = 0)} \tag{15}$$

with the mixture models providing the spatial probability distribution estimates.

IV. RESULTS

The main contribution of this paper is to use the spatial distribution of candidates to improve their classification, augmenting the results obtainable by using their appearance alone. To evaluate performance, for retinal lesion identification, we test SAGE with both bright and dark lesions (Fig. 1). We use a simple intensity peak detection and local contrast algorithm to identify 100 candidate lesions per image, a bank of feature extractors to produce two 96-element appearance feature vectors, one for red lesions, and one for white lesions. The 3504×2336 resolution retinal images were captured using an Canon EOS 20D camera. An expert clinician provided ground truth mark-up for candidates in a training set of 105 images. This data was used to train two classifiers for the white and red lesions, using appropriate feature vectors. For brevity, and as these aspects of the system are not the focus of this paper, we do not give any further details on this part of our system.

TABLE I
CROSS-ENTROPY ERROR COMPARISON.

Lesions	MLP	SVM	SAGE
Bright	102.57	103.63	90.74
Dark	754.61	788.52	299.79

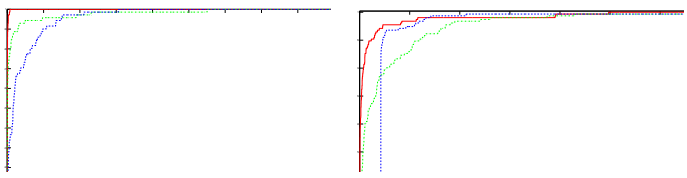
We compare the results gained using the algorithm to those if we simply use the confidence estimates of the appearance-based classifier. The performance is evaluated using a test set of 12 images. These images contain 78 red lesions,and 1122 red non-lesions, and 56 white lesions and 1144 white non-lesions, provided with a ground truth mark-up by an expert clinician.

We evaluate SAGE performance using the cross-entropy error and receiver operating characteristic curves. The cross entropy is given by (16), where t_i is the actual class label for the i^{th} case in the test set, and y_i is the probability estimate returned by (15) for SAGE, or $g(\mathbf{v}_i)$ for the NN classifier. It characterizes the overall error of the probability estimates. The Receiver Operating Characteristic (ROC) curve demonstrates the trade-off between sensitivity and specificity as the classifier decision threshold is adjusted.

$$E(\mathbf{X}|\mathbf{T}) = -\sum_i (t_i \ln(y_i) + (1-t_i)\ln(1-y_i)) \qquad (16)$$

Table I shows the cross-entropy error comparison between the appearance-based classifier and the SAGE algorithm, indicating a substantial improvement in the probability estimates.

Fig. 2 shows the ROC curves comparison of SAGE to the two classifiers. The area under curve (AUC) improves to as much as 83% with SAGE when applied to retinal lesions ($\alpha = 0.01,\ p-value = 0.0002$).



(a) (b)

Fig. 2. ROC Comparison: SAGE (Red), MLP (Green) and SVM (Blue) for: (a) Bright Lesions, (b) Dark Lesions

V. CONCLUSIONS AND FUTURE WORK

A. Conclusions

The novel algorithm introduced in this paper makes the assumption that similar objects occur in groups and if the object is in a group of similar objects that the grouping behavior is indicative of the objects' type. The algorithm uses EM to optimize mixture distribution models of the spatial distribution of objects, using an appearance-based classifier's probability estimates to form spatial distributions, and simultaneously updating the probability estimates to take account of the spatial distribution. We have evaluated the algorithm for both bright and dark retinal lesions using two different classifiers and have demonstrated significantly improved performance for both classifier and lesion types.

B. Future Work

In future work we will investigate other domains where objects typically cluster spatially in groups (e.g. nesting birds), consider the influence of occlusion, integrate other known spatial relationships (e.g. to other objects or features in the scene), and consider how to model more structured spatial distributions. We will also consider extending SAGE to domains where the prior spatial distribution of objects is non-uniform (e.g. a particular street scene where pedestrians are usually to be found on pavements).

VI. ACKNOWLEDGMENTS

The authors gratefully acknowledge the contribution of Mr. David Steel and the Sunderland Eye Infirmary, for images and clinician ground truth.

REFERENCES

[1] W. Hsu, P. M. D. S. Pallawala, M. L. Lee, and K.-G. A. Eong, "The Role of Domain Knowledge in the Detection of Retinal Hard Exudates," *CVPR*, vol. 2, p. 246, 2001.

[2] A. Dempster, N. Laird, and D. Rubin, "Maximum Likelihood from incomplete data via the EM algorithm." *J. Roy. Stat. Soc.*, vol. 39, pp. 1–38, 1977.

[3] C. Ambroise, M. Dang, and G. Govaert, "Clustering of spatial data by the EM algorithm," in *geoENV*, 1996, pp. 493–504.

[4] A. Sopharak and B. Uyyanonvara, "Automatic Exudates Detection from Non-Dilated Diabetic Retinopathy Retinal Images using Fuzzy C-Means Clustering," in *WACBE*, 2007.

[5] A. Osareh, M. Mirmehdi, B. Thomas, and R. Markham, "Automatic Recognition of Exudative Maculopathy using Fuzzy C-Means Clustering and Neural Networks," in *MIUA*, July 2001, pp. 49–52.

[6] C. Jayakumari and T. Santhanam, "Detection of Hard Exudates for Diabetic Retinopathy Using Contextual Clustering and Fuzzy Art Neural Network," *Asian J. of I.T.*, vol. 6, pp. 842–846, 2007.

[7] H. Wang, W. Hsu, K. Goh, and M. Lee, "An Effective Approach to Detect Lesions in Color Retinal Images," in *CVPR*, vol. 2, 2000, pp. 181–186.

[8] H. Zhou, G. Schaefer, and C. Shi, "Fuzzy C-Means Techniques for Medical Image Segmentation." *Fuzzy Systems in Bioinformatics and Computational Biology*, vol. 242/2009, pp. 257–271, 2009.

[9] C. Ambroise and G. Govaert, "Convergence of an EM-type algorithm for spatial clustering," *Pattern Rec. Let.*, vol. 19, no. 10, pp. 919 – 927, 1998.

[10] E. Trucco, C. Burchanan, T. Aslam, and B. Dhillon, "Contextual detection of ischemic regions in ultra-wide-field-of-view retinal fluorescein angiograms," in *IEEE EMBS*, 2007, pp. 6739–6742.

[11] R. Fergus, P. Perona, and A. Zisserman, "A Sparse Object Category Model for Efficient Learning and Exhaustive Recognition," in *CVPR*, vol. 1, 2005, pp. 380–387.

[12] E. Massey, A. Hunter, and D. Steel, "A Robust Lesion Boundary Segmentation Algorithm using Level Set Methods," in *WCMBE*, vol. 25, no. 11, September 2009, pp. 304–307.

[13] E. M. Massey, "Identification and Segmentation of Retinal Lesions," Ph.D. dissertation, University of Lincoln, Brayford Pool Campus, Lincoln, UK, LN6 7TS, December 2010.

[14] J. Bilmes, "A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models," U of C, Berkeley, International Computer Science Institute, Berkeley CA, 94704, Tech. Rep. TR-97-021, April 1998.