EXEMPLAR-BASED SEGMENTATION OF PIGMENTED SKIN LESIONS FROM DERMOSCOPY IMAGES

Howard Zhou[*] *and James M. Rehg*

School of Interactive Computing / GVU Center
Georgia Institute of Technology

Mei Chen

Intel Labs Pittsburgh

ABSTRACT

Automated segmentation of pigmented skin lesions (*PSLs*) from dermoscopy images is an important step for computer-aided diagnosis of skin cancer. The segmentation task involves classifying each image pixel as either lesion or skin. It is challenging because lesion and skin can often have similar appearance. We present a novel exemplar-based algorithm for lesion segmentation which leverages the context provided by a global color model to retrieve annotated examples which are most similar to a given query image. Pixel labels are generated through a probabilistic voting rule and smoothed using a dermoscopy-specific spatial prior. We compare our method to three competing techniques using a large dataset of dermoscopy images with hand-segmented ground truth, We show that our exemplar-based approach yields significantly better segmentations and is computationally efficient.

***Index Terms*—** segmentation, dermoscopy image, pigmented skin lesion, exemplar-based, spatial prior

1. INTRODUCTION

Skin cancer is the most common form of malignancy that occurs in humans in the United States [1]. Dermoscopy, a non-invasive imaging technique, is widely used for early detection of many forms of skin cancer. It involves using an incident light magnification system and a liquid medium applied at the skin-scope interface to enable clinicians performing detailed examination of pigmented structures beyond what would be visible to the naked eye. Studies have shown that using dermoscopy can improve the diagnostic accuracy of dermatologists by as much as 30% over clinical examination [2]. However, this improvement is seen primarily when dermoscopy is used by trained experts; in the hands of inexperienced dermatologists, dermoscopy may actually lower the diagnostic accuracy [3]. In an effort to reduce errors due to the difficulty and subjectivity of human visual interpretation, there has been increasing interest in the development of computational analysis of dermoscopy images.

The first step of such analysis is often the segmentation of a lesion from its surrounding skin. The resulting border structure provides a basis for the calculation of important clinical features, such as lesion size and symmetry axes. In addition, it is crucial for the extraction of some of the most discriminating dermoscopic features such as radial streaming and pseudopods. In recent years, a number of methods have been developed for the automatic segmentation of PSLs in dermoscopy images. See [4] and Sec. 2 for a review.

A majority of these methods assume the lesion and skin pixels define two separable distributions in some chosen feature space, in which case a good segmentation can be found by maximizing the separation between distributions. Unfortunately, this assumption does not hold in practice. Pigmented skin lesions can vary widely in appearance, resulting in lesion color distributions that are not compact and exhibit substantial overlap with skin. Four common sources of variability are depicted in Fig. 4, from the top down: (1) low contrast between the lesion and surrounding skin, (2) compound characteristics in lesion appearance, i.e. differences in appearance within the lesion are greater than the differences between lesion and skin, (3) fragmentation due to regression or depigmentation, and (4) complex background skin appearance. When methods based on the maximum separation assumption are applied to such images, they are likely to produce erroneous results.

The key idea in this paper is that the contextual information provided by the overall appearance of the lesion and its surrounding skin can be exploited to adaptively adjust the classification criteria on a per-pixel basis. This is inspired by the ability of experienced dermatologists to recognize the various conditions under which a particular pixel could be lesion, based on having seen a large number of examples of PSLs. More concretely, we hypothesize that PSLs which have similar morphology are likely to be similar in appearance, and are therefore likely to be "neighbors" with respect to a distance measure such as the Chi-squared distance between color histogram vectors. In this context, a lesion can be segmented by labeling each histogram bin as either lesion or non-lesion. Therefore, we would expect "nearby" histograms to agree on their labeling. On the basis of this insight, we have developed a novel exemplar-based pixel classification algorithm which indexes into a database of previously-segmented dermatoscope images to identify relevant examples, and then

[*]The first author performed the work at Intel Labs Pittsburgh

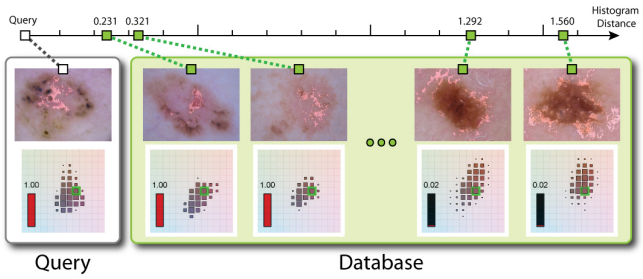

Fig. 1: Illustration of exemplar-based segmentation algorithm with adaptive context.

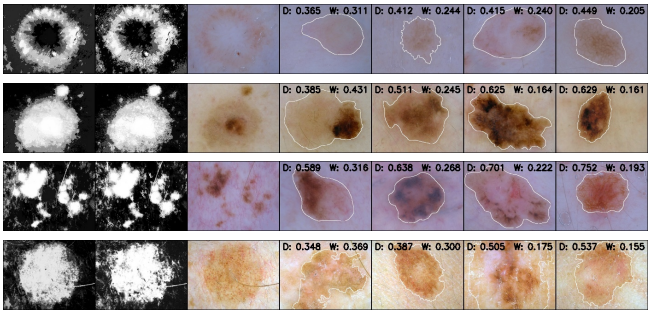

Fig. 2: Algorithm performance on four difficult example lesions (one per row). Cols 1 and 2 are smoothed and unsmoothed segmentation results. Col 3 is query image, while cols 4-7 show the four nearest exemplar neighbors.

generates segmentation labels through a probabilistic voting method. Our approach produces significantly better results than methods which employ a single global segmentation criteria, and it is highly efficient and simple to implement. In addition, we show that an existing technique for enforcing spatial constraints on dermoscopy image segmentations [5] can be utilized to smooth the segmentation map, resulting in a more uniform output and an additional decrease in segmentation error. The performance of our approach is compared to state-of-the-art methods on a hand-segmented dataset of more than 2300 PSL images, which is among the largest in the published literature. Our experiments quantify the benefits of our method in comparison to existing techniques.

2. RELATED WORK

Existing segmentation methods can be roughly classified into three categories: edge/contour-based [6], region-merging and clustering-based [7, 8, 5], and thresholding/classification [9, 10] methods (see [4] for a detailed review). Unsupervised methods that attempt to cluster the feature data into lesion and skin distributions are unlikely to succeed due to the overlap between the two classes. Our approach is most similar to supervised learning methods which train a classifier using labeled examples (e.g. the neural network approach in [9]). In fact, our technique is most related to SVM classifiers, which employ a kernel distance measure related to ours. The primary advantage of our approach in comparison to SVM and other supervised methods is the ability to *interactively* adjust the set of exemplars on-the-fly, effectively tuning the neighborhood definition for a query image. For example, in the context of an interactive retrieval system, a user could easily refine the segmentation by modifying the relevant exemplars. Such flexibility would be difficult to obtain using an offline-trained supervised learning method.

3. EXEMPLAR-BASED PIXEL CLASSIFIER

Fig. 1 illustrates our exemplar-based pixel classification approach. We start with a gallery of pre-segmented exemplars. Each exemplar has an associated color histogram, which defines the "context" for image matching. First, we identify the neighboring exemplars for a given query image, via chi-squared histogram matching. The exemplars are shown in the top row of the green box in Fig. 1, ordered by distance. Below each exemplar is a 2D rendering of its color histogram, where the size of each colored square is proportional to the number of counts.

Our goal is to classify each histogram bin as lesion or skin, thereby labeling all pixels in the query image. A representative bin, whose correct label in the query image is "lesion," is shown in the figure with a green border. The pixels belonging to this bin are "highlighted" in each of the images.[1] Note that this bin arises as both lesion and skin across the exemplar set in Fig. 1, underscoring the limitation of the separable assumption. The highlighted bin has a probability of being lesion in each of the exemplars, as indicated by the red/black bars in the figure. These exemplar probabilities are combined using weights derived from the histogram distances to obtain the predicted lesion probability in the query image histogram (in this example, correctly classified as lesion). The leftmost exemplars, being the closest to the query image, determine the vote, while the rightmost exemplars are ignored. Note that this approach is completely general and could be applied to any image-based segmentation problem.

We now describe the approach in more detail. Let $\vec{X} = \{\vec{x}_i\}_{i \in S}$ be the observed data from an input image where S is a set of image sites to be labeled, and $\vec{x}_i \in \mathbb{R}^c$ is the feature vector at site i (in this work, x_i is pixel color and $c = 3$). Let $H(\vec{X})$ denote the histogram of feature values for the image pixels. The corresponding labels are given by $\vec{Y} = \{y_i\}_{i \in S}$, $y_i \in \{0, 1\}$, representing skin or lesion. We define the label probability distribution for pixel x_i as $P(y_i | \vec{h}_m, H(\vec{X}))$, where $\vec{h}_m$ is the color feature vector at the center of the histogram bin m which contains x_i. Note that in this model,

[1] The highlighting effect is achieved by reducing the opacity of the pixels that are not in the specified bin, causing the pixels within the bin to appear more prominent in the image.

spatially-adjacent labels are conditionally independent given the histogram. We discuss the enforcement of spatial constraints in Sec. 4.

In addition, we are given a library of exemplar images $\mathcal{X} = \{\vec{X}_j\}, j \in \{1, \cdots, J\}$ with a corresponding set of ground-truth labellings $\mathcal{Y} = \{\vec{Y}_j\}$. We use the annotated exemplars to construct the posterior probability distribution for each quantized feature vector as follows:

$$P_i(y_m|\vec{h}_m, H(\vec{X})) \equiv \frac{1}{K}\sum_{k=1}^{K} w_k P(y_m|\vec{h}_m, H(\vec{X}_k)), \quad (1)$$

where K is some fixed number of closest neighbors. Here, proximity is defined by the histogram distance between two images, $d_j = d(H(\vec{X}), H(\vec{X}_j))$. The normalized weight $w_k = d_k^{-2} / \sum_{i=1}^{K} d_k^{-2}$ determines the influence each exemplar has on the final decision. Given the probability model in Eq. 1, the final label assignment is given by $\vec{Y}^* = \arg\max_{\vec{Y}} P(\vec{Y}|H(\vec{X})) = \arg\max_{\vec{Y}} \prod_i P_i(y_m|\vec{h}_m, H(\vec{X}))$. Returning to Fig. 1, we can identify the histogram distances as d_k. For the highlighted bin m, the red/black bars correspond to $P(y_m|\vec{h}_m, H(\vec{X}_k))$. The pixels corresponding to bin m are rendered in each image with the color $\vec{h}_m$.

Fig. 2 illustrates the segmentation results and associated exemplar neighbors for four representative examples of lesions that are difficult to segment reliably. The four query images are illustrated in column 3. Columns 4-7 show the four nearest neighbors returned from the dataset. Note that the similarity is defined only in terms of color histogram, and not based on lesion shape or pigmentation pattern. Nevertheless, our experiments show that this similarity measure enables us to achieve good segmentations while maintaining computational efficiency. The images in column 2 illustrate the per-pixel lesion probabilities (high probabilities are white) produced by our method. Column 1 shows spatially-smoothed probability estimates produced using the technique that we will describe in Sec. 4.

4. ENFORCING SPATIAL SMOOTHNESS

Spatial dependencies between pixels can be used to enforce a local smoothness constraint on image segmentations. In the case of dermoscopy images, the growth pattern of PSLs results in a radiating appearance, so that pixels at the same distance from the center of the lesion are more likely to be statistically dependent. We use a recent approach from [5] as a postprocessing method to enforce smoothness; we review it briefly here for completeness: The pixel color vector is augmented with a normalized polar radius which is 0 at the center of the lesion and 1 at the corner. *k-means++* [11] clustering is then used to group the 4-D feature vectors into segments. Fig. 3 illustrates the identified segments for a typical lesion image. Notice the annular rings which follow the growth pattern. We use these segments to smooth the probability estimates: Within each segment, we compute the average posterior according to Eq. 1 and take the max to obtain the label for that segment. By using a large number of segments (32 in our examples) we obtain a detailed segmentation which respects spatial continuity in a manner that is consistent with the growth patterns of PSLs. The final segmentation boundary is obtained by post-processing the mask with connected component analysis and morphological operations.

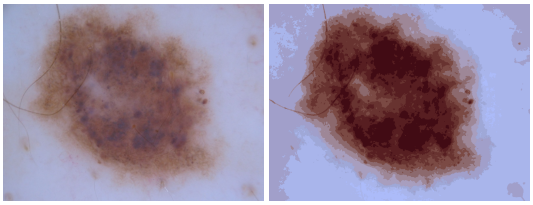

Fig. 3: Dermoscopic specific radiating appearance captured by the clustering step.

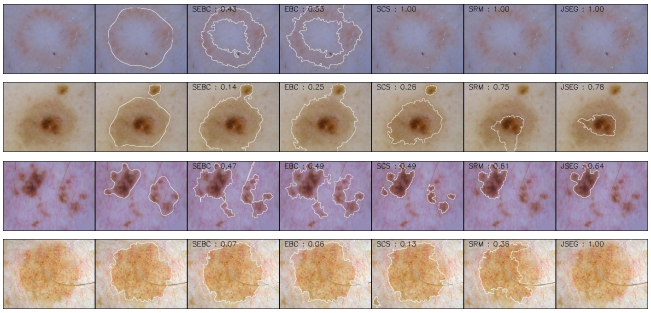

Fig. 4: Segmentation results for the images of Fig. 2. Col 1 is the image, col 2 is the groundtruth. Cols 3-7 are results generated by the SEBC, EBC, SCS, SRM, and JSEG methods.

5. EXPERIMENTAL RESULTS

We evaluated the performance of our method on three dermoscopy datasets with ground truth segmentations. Dataset D1 consists of 67 images labeled by two expert dermatologists, and was provided by the authors of [5]. Dataset D2 contains 111 images, and includes many difficult cases such as examples in the first and third rows in Fig. 4. It was labeled by two expert dermatologists. Dataset D3 combines D1 and D2 with an additional 2159 images from a variety of sources, including standard dermoscopy books [12, 13].[2] Ground truth labels for these additional images were provided by a skilled operator.

We conducted a total of five experiments in which we compared the results from our exemplar-based classifier (EBC) and its spatially-smoothed version (SEBC) to three existing methods: JSEG [14], a method that consists of color quantization and a subsequent region growing step to locate

[2] Images came as jpeg files on the companion CDs of the books.

lesion borders, SRM [8], a method based on the statistical Region Merging algorithm, and SCS [5], a clustering-based method that incorporates a dermoscopic spatial prior. These methods represent the standard approaches to PSL segmentation and were chosen for their competitive performance. Using software provided by the authors,[3] we tested each method on our three datasets. For datasets D1 and D2, we performed leave-one-out cross-validation, treating each image as a query and using the remaining images as exemplars. For D3, 550 exemplars were randomly chosen and the remaining 1787 images were used for testing. Segmentations were scored using the standard XOR metric [8], which reports the area of the XOR between the prediction and groundtruth masks, normalized by the ground truth area.

The three parameters for our method were determined empirically, by searching over sets of possible values. These include the number of histogram bins per dimension n (from $\{10, 20, 30, 40\}$), the number of neighbors K (from $\{3, ..., 10\}$), and the number of clusters for spatial smoothing N (from $\{10, 20, 30, 40\}$). In addition, we also explored two color spaces: RGB and CIE L*a*b* and four histogram distance measures: intersection, correlation, chi-square, and Bhattacharyya. The parameters were tuned on a randomly-selected subset of images from D2. The results were not sensitive to the choice of parameters. The best combination, $n = 20$, $K = 5$, $N = 20$, RGB, and Bhattacharyya, was used for all of the experiments.

Fig. 5 shows the average XOR errors for all methods. Both EBC and SEBC consistently out-performed all other methods, frequently by a large margin. In the case of datasets D1 and D2, the existence of two separate hand-segmentations allowed us to estimate the inter-operator error, which was 11.32 for D1 and 13.72 for D2. Note that on dataset D1, SEBC came within 2% of the expert inter-operator error. All experiments were performed on an Intel Pentium D processor clocked at 3.20GHz with 3GB of memory. The average execution time in seconds was computed for dataset D2 and is reported in the fifth column of Fig. 5. EBC and SRM were the fastest.

Fig. 4 shows example segmentations for the images of Fig. 2, which are difficult cases. Due to low contrast, the lesion in the first row is missed by all previous methods. Our approach is successful because it can exploit contextual information from neighbors with similar appearance (as illustrated in row 1 of Fig. 2). Similarly in row 2, the ability to retrieve neighbors with compound appearance results in improved performance. Comparing the results from SEBC and EBC (columns 2 and 3), we find that incorporating dermoscopy-specific spatial constraints improves border localization and segmentation quality.

[3]The JSEG software failed for some images in D3, so we omitted results for that combination.

Method	D1	D2	D3	Time
JSEG	20.43	32.81	-	9.67
SRM	20.77	39.50	36.77	**0.46**
SCS	14.93	28.77	39.58	5.72
EBC	13.70	26.76	22.23	**0.45**
SEBC	**13.36**	**25.88**	**20.62**	4.37

Fig. 5: Percentage border error using the XOR grading system

6. CONCLUSION

We have described a novel lesion segmentation algorithm for dermoscopy images that leverages a set of exemplars annotated by experts and incorporates dermoscopy-specific spatial smoothing. Our method consistently outperforms three existing approaches on a large hand-segmented dataset. It is computationally efficient and simple to implement, and potentially effective in other problem domains.

7. REFERENCES

[1] National Cancer Institute, "Common cancer types," `http://www.cancer.gov/cancertopics/commoncancers`, 2009.

[2] D.S. Rigel, R.J. Friedman, and A.W. Kopf, "The incidence of malignant melanoma in the united states: Issues as we approach the 21st century," *J Am Acad Dermatol.*, vol. 34, pp. 839–847, 1996.

[3] M. Binder, M. Schwarz, A. Winkler, A. Steiner, A. Kaider, K. Wolff, and H. Pehamberger, "Epiluminescence microscopy. a useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists," *Arch. Dermatol.*, vol. 131, no. 3, pp. 286–291, 1995.

[4] M.E. Celebi, H. Iyatomi, G. Schaefer, and W. Stoecker, "Lesion border detection in dermoscopy images," *Computerized Medical Imaging and Graphics*, vol. 33, no. 2, pp. 148–153, 2009.

[5] H. Zhou, M. Chen, L. Zou, R. Gass, L. Ferris, L. Drogowski, and J. M. Rehg, "Spatially constrained segmentation of dermoscopy images," in *International Symposium on Biomedical Imaging*, May 2008, pp. 800–803.

[6] B. Erkol, R.H. Moss, R.J. Stanley, W.V. Stoecker, and E. Hvatum, "Automatic lesion boundary detection in dermoscopy images using gradient vector flow snakes," *Skin Research and Technology*, vol. 11, pp. 17–26, 2005.

[7] P. Schmid, "Segmentation of digitized dermatoscopic images by two-dimensional color clustering," *IEEE Trans. Medical Imaging*, vol. 18, pp. 164–171, 1999.

[8] M.E. Celebi, H.A. Kingravi, H. Iyatomi, J. Lee, Y.A. Aslandogan, W.V. Stoecker, R. Moss, J.M. Malters, and A.A. Marghoob, "Fast and accurate border detection in dermoscopy images using statistical region merging," in *SPIE Medical Imaging*, 2007.

[9] M. Hintz-Madsen, L. K. Hansen, J. Larsen, and K. Drzewiecki, "A probabilistic neural network framework for detection of malignant melanoma," in *Artificial Neural Networks in Cancer Diagnosis, Prognosis and Patient Management*, pp. 141–183. 2001.

[10] D.D. Gomez, C. Butakoff, B.K. Ersboll, and W. Stoecker, "Independent histogram pursuit for segmentation of skin lesions," *IEEE Trans. Biomedical Engineering*, vol. 55, no. 1, pp. 157–161, 8 2008.

[11] D. Arthur and S. Vassilvitskii, "k-means++: the advantages of careful seeding," in *SODA*, 2007, pp. 1027–1035.

[12] Scott W. Menzies, Kerry A. Crotty, Christian Ingwar, and William H. McCarthy, *An Atlas of Surface Microscopy of Pigmented Skin Lesions: Dermoscopy, 2nd Edition*, McGraw-Hill Book Company Australia, 2002.

[13] G. Argenziano, H.P. Soyer, V. De Giorgi, D. Piccolo, P. Carli, and M. Delfino et al., *Dermoscopy: a tutorial*, EDRA Medical Publishing & New Media, 2002.

[14] M. Emre Celebi, Y. Alp Aslandogan, William V. Stoecker, Hitoshi Iyatomi, Hiroshi Oka, and Xiaohe Chen, "Unsupervised border detection in dermoscopy images," vol. 13, pp. 454–462, November 2007.