Pigmented skin lesion segmentation based on sparse texture representations

César E. Martínez\textsuperscript{a} and Enrique M. Albornoz\textsuperscript{a, b}

\textsuperscript{a}Research Institute for Signals, Systems and Computational Intelligence (\textit{sinc(ı)}), Faculty of Engineering and Water Sciences, UNL-CONICET, CC 217, Ciudad Universitaria Paraje El Pozo, S3000, Santa Fe, Argentina.

\textsuperscript{b}CONICET, Argentina.

\textbf{ABSTRACT}

Among the most dangerous cancers, there is the Melanoma that affects millions of people. As this is a type of malignant pigmented skin lesion and it can be recognized by medical experts, computer-aided diagnostic systems are developed in order to assist dermatologists in clinical routine. One of the more difficult tasks is to find the right segmentation of lesions whose precision is very important to distinguish benign from malignant cases. In this work, we propose a new method based on sparse representation. First, an alternative representation of the image is obtained from the texture information. A sparse non-negative dictionary is computed and every image is projected onto this space. The reconstruction is calculated using only the most active atoms, which allows to obtaining an enhanced version of the texture where the morphological post-processing can effectively extract the lesion area. The experiments were carried out on a publicly available database and performance was evaluated in terms of segmentation error, accuracy, and specificity. Results showed that this first approach performs better than methods reported in the literature on this same data.

\textbf{Keywords:} Pigmented skin lesions, Texture, Segmentation, Sparse representations

\section{1. INTRODUCTION}

Researchers have worked in developing diagnostic systems for skin lesions from 80’s years. Proposed teledermatology systems allow the expert to receive images of suspicious lesions and provide a diagnosis remotely; while the computer-aided diagnosis systems can classify an acquired lesion; and combined systems use both in conjunction.\textsuperscript{1} Melanoma is the most aggressive form of skin cancer and as it has high levels of metastasis, it has the highest mortality rate.\textsuperscript{2} As was mentioned in,\textsuperscript{3} in 2008 the Melanoma was the 19th most common cancer worldwide, about 200,000 new cases were estimated. These numbers widely justify the dedication of thousand of researchers.

Identifying melanomas from benign pigmented skin lesions (called nevi or moles) can be a hard task. In the last years, researchers have done great efforts to develop systems for detection and analysis of skin lesions from images.\textsuperscript{4–6} These systems can be used to monitor benign skin lesions in order to diagnose the malignant at a early stage, during which the patient has a higher probability of cure.

Many computer-aided diagnosis system were proposed to help the physicians analyzing the dermoscopy images,\textsuperscript{7–9} whereas other systems dealing with standard camera images also have been proposed with the motivation of easy access to health-care. In general, a complete computer-aided diagnosis systems can be defined in four steps: preprocessing where the image is enhanced and restaurated; segmentation where the lesion borders are delimited; feature extraction where descriptors of the lesion are generated using the lesion segment; and classification where the lesion are labeled as benign or malignant.

The segmentation step has the more critical role in the effectiveness of the complete system, where suitable results can be seen in previous works.\textsuperscript{5, 10, 11} Some of the difficulties of standard camera images of pigmented skin lesions are given by the shading areas,\textsuperscript{12, 13} which may be attenuated in the preprocess step; the hairs that are usually eliminated using the DullRazor algorithm\textsuperscript{14} but can be eliminated after the segmentation.\textsuperscript{13} The
Methods commonly used in literature for the segmentation of pigmented skin lesions in images, can be grouped in the edge-based methods,\textsuperscript{3,15} thresholding-based methods,\textsuperscript{16–19} region-based methods and active contours,\textsuperscript{20,21} and methods based on artificial intelligence.\textsuperscript{20,22,23}

In this work we present a new approach to lesion segmentation based on sparse representations. Taking the texture information as input, the method models the image using an optimal set of atoms that capture features of the texture images. Then, the system reconstructs an alternative version of the image where the robust representation hopefully enhances the important characteristics decreasing the incidence of particularities such as color differences or hairs. Finally, the border and lesion area are found by a morphological postprocessing.

The rest of the article is organized as follows: Section 2 presents the technical details of the method, Section 3 shows the results of the experiments and discussion about analysis and comparisons with other systems, finally Section 4 presents the conclusions of this work and delineates future works.

2. PROPOSED SEGMENTATION APPROACH

The main steps in our proposed method comprise:

1. Preprocessing of images.
2. Dictionary learning and optimization.
4. Postprocessing.
5. Obtainment of the lesion border.

The following subsections present the technical details with figures that illustrate each step.

2.1 Preprocessing

A multichannel image representation of pigmented skin lesion images was computed here.\textsuperscript{13} The method replaces the original (normalized) color image $I^C_i$ with a new matrix $I^N_i$ with three channels containing information that maximizes the discrimination between healthy and unhealthy skin regions.

Channel 1 represents the image darkness given by the depigmented regions of the lesion and is calculated as the complement of the normalized red channel $I^N_1(x,y) = 1 - I^C_1(x,y)$.

Channel 2 represents the texture of the image. The important clue here is that local regions with high variability in texture are pointing out the lesion instead of the low variability found in healthy regions. It is obtained using the normalized luminance image $L(x,y) = \frac{1}{3} \sum_i I^C_i(x,y)$ from which its textural variability $\tau(x,y,\sigma)$ is computed as:

$$\tau(x,y,\sigma) = \tilde{L}(x,y) \frac{S(x,y,\sigma)}{\tilde{S}(x,y,\sigma)},$$

where $S(x,y,\sigma) = \tilde{L}(x,y) * G(\sigma)$ (luminance smoothed by a Gaussian filter with standard deviation $\sigma$) and $\tilde{S}(x,y,\sigma)$ being its complement. Thus, a dark region in the image obtains an emphasized textural information and vice versa. This parameter is calculated for different $\sigma$ values and the maximum value $T(x,y)$ at each pixel is saved. Then, the texture variation $T$ is normalized and channel 2 of the representation is obtained as $I^N_2(x,y) = (T(x,y) - \min(T)) / (\max(T) - \min(T))$.

Channel 3 $I^N_3(x,y)$ represents the color variation, under the hypothesis that healthy and unhealthy skin (background and lesion pixels) present different color distributions. A PCA method is applied on $I^C_i(x,y)$ and the first component is retained (maximum data variance).

Previous works have used the three channels as input to the segmentation block.\textsuperscript{5,13,24} In this work only the texture information contained in channel 2 is used ($I^N_2(x,y)$) given that our main interest relays on the proposal and analysis of sparse representations as a suitable image modeling technique for the task.
2.2 Dictionary learning and image representation

We base the modeling of the images by adapting our previous proposal of the representation of speech at the auditory cortical level. Here, considering \( x \in \mathbb{R}^{m \times n} \) as the image sliding patches from the texture information, we can have a linear combination of atoms representing the texture features, in the form

\[
x = \Phi a,
\]

where \( \Phi \in \mathbb{R}^{m \times n \times M} \) is the dictionary of \( M \) bidimensional atoms and \( a \in \mathbb{R}^{M} \) is the target representation.

The basis functions are vectorized as \( \Phi = [\tilde{\Phi}_1 \ldots \tilde{\Phi}_M] \) with \( \tilde{\Phi}_i \in \mathbb{R}^{mn \times 1} \). Thus, (2) can be alternatively written as \( \tilde{x} = \sum_{1 \leq i \leq M} \tilde{\Phi}_i a_i \). The sparse representation is obtained when the solution is restricted to

\[
\min_a \|a\|_0,
\]

where \( \| \cdot \|_0 \) is the \( l^0 \) norm. This is a known NP-complete problem so different approximations were proposed. For obtaining the representation, two problems have to be jointly solved: the estimation of a sparse representation and the inference of the dictionary of specialized atoms. Particularly in image modeling, an only-positive-values representation is desired, given the nature of the data represented in the images and the ability of the method to obtain a representation only with the additions of positive atoms. To this end, we used the non-negative singular value decomposition method (NN-K-SVD) proposed in, which is a widely adopted solution for the problem. The method also includes the possibility to set the number of sparse components to use in the approximation and solves the problem

\[
\min_a \|x - \Phi^L a\|_2^2 \text{ s.t. } a \geq 0,
\]

where a sub-matrix \( \Phi^L \) that includes only a selection of the \( L \) largest coefficients is used. In the dictionary updating, this matrix is forced to be positive by calculating

\[
\min_{\tilde{\phi}_k, a_k} \|E^k - \tilde{\phi}_k a_k\|_2^2 \text{ s.t. } \tilde{\phi}_k, a_k \geq 0,
\]

for each one of the \( k \) selected coefficients. The error matrix \( E^k \) is the residual between the signal and its approximation, while \( k \)-th atom \( \tilde{\phi}_k \) and its respective activation \( a_k \) are updated. Finally, the dictionary itself, and the activation coefficients are calculated from the SVD of \( E^k = U \Sigma V^T \), where the atoms and activations are obtained as the rank-one approximation with the first left and right singular vector as \( \tilde{\phi}_k = u_1 \) and \( a_k = v_1 \).

Once the representation is computed, we proposed here to further optimize \( \Phi \) by selecting the most active atoms and building a new dictionary \( \Phi' \), where the less active atoms are set to zero. The purpose of this step is to extract and project the patches only in representative atoms so providing the complete method of robustness against the natural variability found in different skin lesions.

During runtime, each texture image \( \overline{I}_N(x, y) \) is scanned by non-overlapping patches of \( m \times n \) pixels and vectorized to obtain the \( x \) data matrix. Then, the activations are found using the NN-K-SVD algorithm and the optimal dictionary \( \Phi' \). Finally, the “denoised” texture image is built by reconstructing the image using 2.

2.3 Postprocessing

The obtained image is post-processed to finally get the binary mask that represents the lesion area. A number of morphological operations are applied following this order:

- Threshold the image using the peak detection of the histogram. The threshold is fixed to the mean value of the main peak (dark mass of background pixels) and the light peak (lesion).
- Area opening: remove noise given by very small objects.
- Merge near connected components using the (Euclidean) distance transform with 8-connectivity followed by an erosion to recover the original size.
- Remove small areas that ideally do not correspond to lesions.
- Closening of the surviving areas (lesion candidates).
- Selection of the lesion blob whose center of mass is near (Euclidean distance) to the center of the image, followed by the morphological filling of holes to get the final binary mask.

The last step is the obtaining of the lesion border by means of a contour extraction algorithm that removes the pixels with 4-connected neighbors being 1’s. Finally, the obtained curve is smoothed using the Savitzky-Golay sliding polynomial filter and the area of the lesion is filled for building the segmented mask that allows the performance comparison with the ground truth segmentation.

3. EXPERIMENTS AND RESULTS

In our experiments, we use the same image dataset proposed by Alcón et al. It contains 152 images that have been collected from the Dermnet dataset. The dataset consists of 107 melanomas and 45 Clark nevi (or atypical nevi), a benign kind of lesion that present similar characteristics to melanomas.

During initial experiments we found that a good choice of parameters is: size of patches \(m \times n = 16 \times 16\), size of dictionary \(L = 256\) atoms, number of components selected by the NN-K-SVD algorithm \(K = 4\) atoms.

The evaluation of the segmentation results is carried out using the following parameters: error, accuracy and specificity.

The error is given by

\[
\epsilon = \frac{\sum_{x,y} S_{x,y} \oplus G_{x,y}}{\sum_{x,y} G_{x,y}},
\]

where \(S_{x,y}\) denotes every pixel in the resulting mask of our method, \(G_{x,y}\) is the expert manual segmentation for the lesion and \(\oplus\) indicates the exclusive-OR operation that gives the pixels for which both sets differ.

The accuracy represents the hit probability (lesion detection) and is calculated as

\[
\text{acc} = \frac{TP + TN}{TP + FP + TN + FN},
\]

where

- \(TP\) is the number of true positive pixels: \(S_{x,y} = G_{x,y} = 1\).
- \(TN\) is the number of true negative pixels: \(S_{x,y} = G_{x,y} = 0\).
- \(FP\) is the number of false positive pixels: \(S_{x,y} = 1, G_{x,y} = 0\).
- \(FN\) is the number of false negative pixels: \(S_{x,y} = 0, G_{x,y} = 1\).

The specificity represents the probability of doing a correct segmentation in terms of low false positive hits and is calculated as

\[
\text{spec} = \frac{TN}{TN + FP}.
\]
3.1 Analysis of the dictionary

The dictionary learning is carried out using a small set of images composed of random examples of two classes of melanomas: 5 cases of atypical nevi and 5 cases of malignant melanoma. Using the same data, the optimization (deletion of less active atoms) is performed, where 56 atoms are kept from the original 256 set.

Figure 1 shows an example of a learned dictionary \( \Phi \) where the atoms are ordered by the \( L_2 \) norm. It can be seen atoms that collect different information (some examples of each one are marked in colors in the Figure):

- directional borders, given by the green rectangles;
- background low texture areas, given by the magenta rectangle;
- some pieces of hairs, which is a common “noisy” feature in these images, given by the orange squares;
- high texture information which correspond to lesion areas, given by the cyan rectangle.

The advantages of the sparse reconstruction and the concept of robustness of the optimal representation achieved by the reduced dictionary (after deleting the less active atoms) is shown in Figure 2. Here, the texture image shows decreasing values from top to bottom, which leads to erroneous detection after thresholding. The reconstruction achieved by the sparse representation shows much better gray differences between the lesion and background, so the method effectively can delineate a border very close to the ground truth (red closed curve in the color image). Also, the figure shows a zoomed-in excerpt from the top left border of the lesion, where the reconstruction from atoms of background, borders and lesion texture can be seen (blocky effect given by the non-overlapping sliding procedure).

3.2 Segmentation results

In order to illustrate the results, Figure 3 shows examples of the segmentation obtained using the proposed method. For reference, the error, accuracy and specificity of each example are also given. The examples tried to cover the variability in the database (different skin tones, exposure, lesion type).

Figure 4 shows examples of the results obtained along with the comparison with the ground truth segmentation and the segmentation obtained in other recent proposals. The images are ordered as: ground truth in the first column, results of non-negative matrix factorization\(^5\) in the second column, results of normalized graph cuts\(^\text{24} \) in
Original color image with the segmentation obtained by the method. $\epsilon = 0.1179$, $acc = 0.9549$, $spec = 0.9939$.

Texture channel $\tilde{I}_N^2(x,y)$ (left) and sparse reconstruction obtained (right).

Zoom-in of the top left corner of the lesion.

Figure 2. Example of the reconstruction achieved and comparison with original texture information.
the third column and results of this work on the fourth column. We can see that our method obtains the lowest errors while maintaining the accuracy and specificity in high values.

Figure 5 shows examples of good segmentation results obtained on complex images. Here, the borders showed by the red curves are still good approaches to the correct segmentation, regarding the high incidence of hairs and visible differences on textures.

On the other side, Figure 6 shows examples of erroneous segmentations. The ground truth segmentation corresponds to the green curve, while the segmentation of the method is the red one. In cases (a) and (b), the errors arise by the splitting evidenced in the lesion. As the method is only searching for one isolated region, when the texture has significant changes the method cuts the masked zone (see for example the selection of the right zone in the second case). The case (c) corresponds to a lesion with a complex textural lesion (two different textures) where the method cuts only the most prominent region given by the dark inner zone. The case (d) presents a lesion with a surrounding zone also textured, making more difficult the segmentation task. In all the cases the error grows up while the accuracy decreases with respect to the images discussed above. Nevertheless, see that the specificity is very high in cases (a) to (c) due to the (almost) zero false positive detection—the segmentation is totally contained in the ground truth region.

As a final comparison in performance between our system and other approached reported in literature, Table 1 presents the mean segmentation error obtained using the same data. In our results, the mean/standard deviation for the three figures of merit, are: \( \mu_{err}/\sigma_{err} = 0.1876/0.1345, \mu_{acc}/\sigma_{acc} = 0.9667/0.0419, \mu_{spec}/\sigma_{spec} = 0.9852/0.0228 \).

It can be seen that our proposal lowers the segmentation error to 0.1898 (18.98%) which is the lowest error when compared with other systems. This is, up to our knowledge, the best result on this data. Although this good performance, we consider that it is just the tip of the iceberg and there is plenty of space to further investigate and improve the system. For example, the dictionary is obtained using only the texture channel and the erroneous cases could be pointing out that the use of combined dictionaries of texture and color variation might be beneficial for the segmentation. Another topic to study is the conformation of the dictionary itself. We carried out here an automatic selection of representative atoms given its activity in the representation of the data, but it also can be selected by measuring entropy, geometric features or other information that leads to obtain an initial labeling of atoms in main classes (healthy/unhealthy skin, hair, etc.).
Figure 4. Comparison with ground truth (first column), other skin lesion segmentation methods reported in literature (second and third column)\textsuperscript{5, 24} and this work (fourth column).

Figure 5. Examples of the segmentation obtained for complex images.
Figure 6. Examples of erroneous segmentations arising by lesions with very complex textures.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcón et al. Thresholding</td>
<td>Alcón et al. Thresholding</td>
</tr>
<tr>
<td>Tang’s Snake</td>
<td>Tang’s Snake</td>
</tr>
<tr>
<td>Otsu’s Thresh. on Grayscale</td>
<td>Otsu’s Thresh. on Grayscale</td>
</tr>
<tr>
<td>Otsu’s Thresh. on the R Channel</td>
<td>Otsu’s Thresh. on the R Channel</td>
</tr>
<tr>
<td>Thresh. on a Multichannel Image</td>
<td>Thresh. on a Multichannel Image</td>
</tr>
<tr>
<td>ICA-Based Active-Contours</td>
<td>ICA-Based Active-Contours</td>
</tr>
<tr>
<td>NMF Segmentation</td>
<td>NMF Segmentation</td>
</tr>
<tr>
<td>Normalized Graph Cuts</td>
<td>Normalized Graph Cuts</td>
</tr>
<tr>
<td>This work</td>
<td>This work</td>
</tr>
<tr>
<td>Average $\epsilon$</td>
<td>Average $\epsilon$</td>
</tr>
<tr>
<td>1.6531</td>
<td>0.5960</td>
</tr>
<tr>
<td>0.4233</td>
<td>0.4233</td>
</tr>
<tr>
<td>0.3858</td>
<td>0.3858</td>
</tr>
<tr>
<td>0.3483</td>
<td>0.3483</td>
</tr>
<tr>
<td>0.2834</td>
<td>0.2834</td>
</tr>
<tr>
<td>0.2599</td>
<td>0.2599</td>
</tr>
<tr>
<td>0.2112</td>
<td>0.2112</td>
</tr>
<tr>
<td><strong>This work</strong></td>
<td><strong>This work</strong></td>
</tr>
<tr>
<td><strong>0.1898</strong></td>
<td><strong>0.1898</strong></td>
</tr>
</tbody>
</table>

4. CONCLUSIONS

In this work we presented a first approach to pigmented skin lesion segmentation based on sparse representation of texture information. This representation is calculated with the non-negative K-SVD algorithm, that selects the most important atoms for representing each patch of the image. The binary lesion mask is obtained by the morphological processing of a reconstructed texture image from the sparse feature space, where the important and discriminative information is highlighted.

The segmentation error, accuracy and specificity were used as figures of merit. The results obtained by our method showed a very good performance (mean segmentation error less than 19%), and performed better than other state-of-the-art systems.

These results and the analysis of the erroneous cases encourage us to continue the research on sparse representations applied to this task. Future works will be devoted to propose the setting up of combined sparse...
dictionaries that integrate the texture with additional information. Also, convolutive sparse representations could be explored with the aim to improve the non-overlapping current limitation in data analysis.

ACKNOWLEDGMENTS

The authors would like to thank the Universidad Nacional de Litoral (with PACT 2011 #58 and CAI+D 2011 #58-511), as well as the National Scientific and Technical Research Council (CONICET), from Argentina, for their support.

REFERENCES


