



ELSEVIER

Computerized Medical Imaging and Graphics 27 (2003) 65–78

Computerized
Medical Imaging
and Graphics

www.elsevier.com/locate/compmedimag

Towards a computer-aided diagnosis system for pigmented skin lesions

Philippe Schmid-Saugeon^a, Joël Guillod^b, Jean-Philippe Thiran^{a,*}

^aSignal Processing Institute—ITS, Swiss Federal Institute of Technology—EPFL, 1015 Lausanne, Switzerland

^bDermatology Department, Lausanne University Hospital—CHUV, 1011 Lausanne, Switzerland

Received 17 November 2000; revised 6 August 2002

Abstract

This paper presents a computer-aided diagnosis system for pigmented skin lesions, with solutions for the lesion boundary detection and for the quantification of the degree of symmetry. Lesion detection results were validated by expert dermatologists, who also provided hand-drawn boundaries of the lesions. These reference boundaries were not used as a gold standard, but were allowed to statistically determine the accuracy of the boundaries provided by computerized techniques. We could show that the dermatologists were not able to reproduce their results, and that the boundaries of any expert taken alone showed higher divergence from those of the set of remaining experts than the automatic techniques we developed. Feature extraction is restricted in this paper to the quantification of degree of symmetry, even though it is clear that many other features will be necessary for a complete diagnosis system. The symmetry quantification step provides a six-dimensional feature vector that can be used to classify pigmented skin lesions as being benign or malignant. We demonstrate that our scheme outperforms methods based on the principal component decomposition, which is widely used for this kind of application.

© 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Computer-aided diagnosis; Malignant melanoma; Dermoscopy; Segmentation; Boundary detection; Symmetry

1. Introduction

Malignant melanoma (MM) is the third most frequent type of skin cancer and one of the most malignant tumors [1]. It is therefore worrying that the number of new cases has increased much more than other types of cancer. The incidence is on the order of 10–12 per 100,000 in Europe, five times more than 50 years ago, 18–20 per 100,000 in the USA, and 30–40 per 100,000 in Australia, where the climate and sun exposure are especially hazardous [1].

A malignant melanoma diagnosed at an early stage can be cured without complications. The *early diagnosis of malignant melanoma* is therefore a crucial issue for dermatologists. Knowing that the rate of accurate diagnosis is around 75% (true positives) [2,3], it is important to develop efficient schemes for the clinical diagnosis and to support the dermatologists with *computer-aided diagnosis* (CAD) systems. The main objective of such systems is to assist the physician in different analysis steps, such as the lesion boundary detection, the quantification of diagnostic

features, the classification into different types of lesions, the visualization, the storage, the database management, etc.

In this study we processed magnified images which use oil immersion to render the *stratum corneum*, i.e. the outer layer of the skin, translucent. This technique is called *epiluminescence microscopy* (ELM or *dermoscopy*) [2]. This imaging technique reveals most of the pigmented structures by allowing the light to penetrate deeper into the skin. Different color shades become visible, depending on the depth of the pigment, and usually the boundary between the lesion and the surrounding skin is very smooth. Since ELM is used for the diagnosis of skin cancer, the visual evaluation of diagnostic features has become a difficult problem, because the information content of dermatoscopic images is very complex. Efficient image processing techniques must therefore be developed to help physicians making a diagnosis.

Stolz et al. established a diagnosis scheme for dermatoscopic images known as the *ABCD rule of dermoscopy* [2]. The different letters stand for the following criteria:

- *Asymmetry*: in zero, one, or two orthogonal axes (0–2 points). Color, texture, and shape must be taken into account.

* Corresponding author. Tel.: +41-21-693-46-23; fax: +41-21-693-76-00.

E-mail address: jp.thiran@epfl.ch (J.P. Thiran).

- *Border*: the lesion is divided into eight radial pieces which are then labeled as showing a sharp cut-off with the surrounding skin or not (0–8 points).
- *Color*: the presence of up to six known colors must be detected—white, red, light brown, dark brown, slate blue, and black (1–6 points).
- *Differential structures*: five patterns are relevant for specific types of lesions (1–5 points).

A weighted combination of the different scores provides a *total dermatoscopic score* that is used to classify the lesions. The reader should refer to Refs. [2,4] for additional information about this scheme and the medical background.

The ABCD rule of dermatoscopy shows the weakness of the analysis done by physicians: the different criteria have a finite number of states and are evaluated subjectively. There is no defined limit between the different states and the feature extraction depends entirely on the physician’s experience. It is therefore important to design extraction schemes which rely on objective criteria and to provide the physicians with real-valued measurements.

Unfortunately, the feature extraction is often subject to error. In most of the publications dealing with this topic, many features are extracted to feed a sophisticated classifier, but there is little discussion about the real meaning of those features and about objective ways to measure them.

The design of classifiers was investigated in many studies, most of them being artificial neural networks [3,5–12]. In Ref. [13], a hard threshold is applied to the histogram of the blue component in order to obtain the lesion boundary. Very simple parameters, such as area and perimeter, are extracted in Ref. [14]. Some measurements on the different color components are also used, and the lesion is once more detected using a global thresholding technique. In Ref. [15] the authors try to critically examine image analysis techniques. However, they use very simple feature extraction methods, such as the principal component of a binary mask of the lesion to quantify asymmetry. Finally, similar work can be found in Refs. [16,17]. In the latter case the authors are concerned with texture classification, but without relevant breakthrough.

This paper is organized as follows: the scheme of our CAD system is presented in Section 2. The different processing algorithms are explained in Section 3, and the boundary detection results are validated in Section 4. Classification results are presented in Section 5, and finally we draw some conclusions in Section 6.

2. A computer-aided diagnosis system

The first step in this research was to establish the scheme of a CAD system for pigmented skin lesions. This scheme is

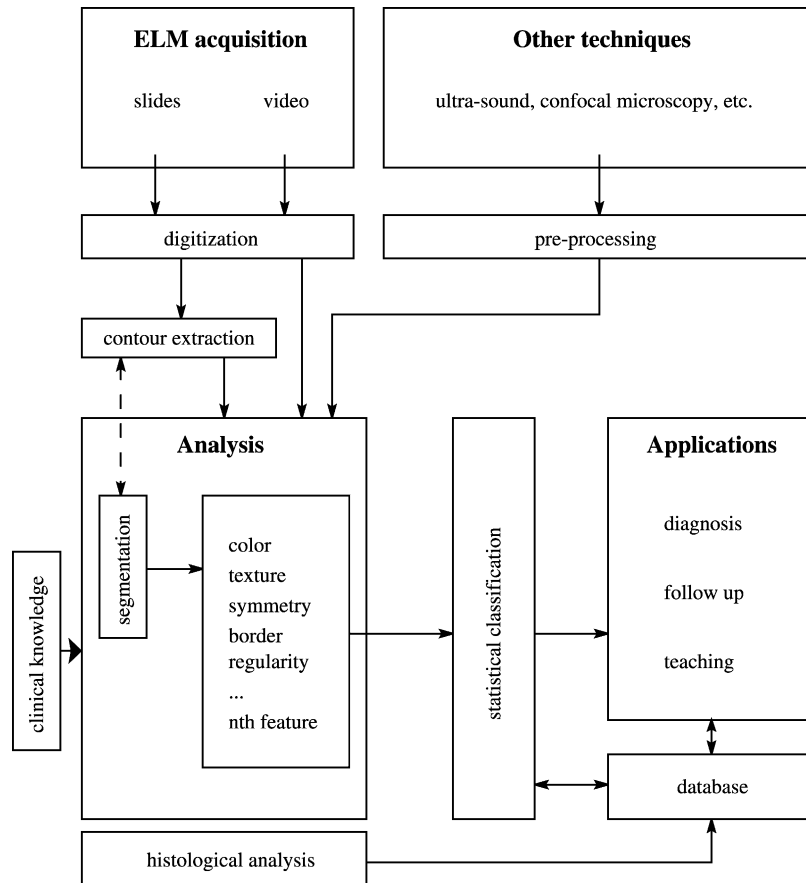


Fig. 1. CAD system for pigmented skin lesions.

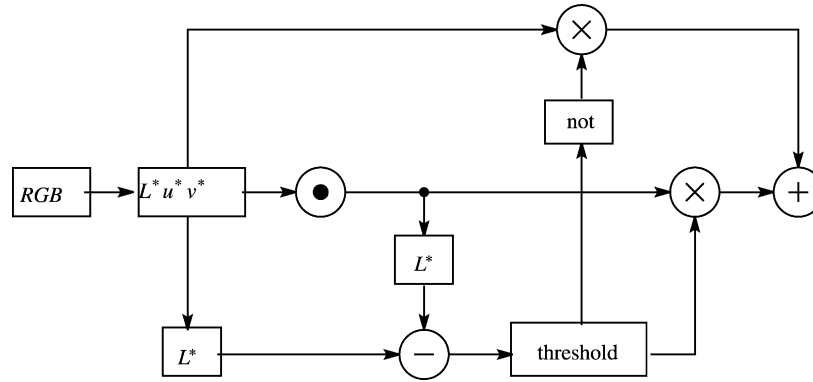


Fig. 2. Hair removal algorithm.

shown in Fig. 1. The input to our system are images obtained by ELM, with the possibility, in future extensions, to add other acquisition system such as ultra-sound [18] or confocal microscopy [19]. The analysis part relies on algorithms that can be derived from clinical knowledge gained by physicians. The first processing step is the detection of the lesion, which can be obtained either with a boundary detection or an image segmentation technique. Once the lesion is localized, different features can be quantified and used for the classification. In order to train a classifier, images of benign and malignant lesions must be collected along with their histopathology¹ and stored in a database. This collection of images will also serve as a reference database, where cases with known pathologies can be consulted by the physicians.

Besides providing information to the physicians during the diagnosis phase, a CAD system must also manage the different computing resources. It must provide an interface between the dermatologist and the image processing routines, allow storage and retrieval of unclassified and diagnosed cases, give access to the network and to a database server, and control the digitization process or a digital camera connected to the computer.

In this project, images are digitized photographic slides of dermatoscopic images. In practice, it is difficult to obtain high-quality images of pigmented lesions. Therefore, image processing algorithms must be able to show a certain insensitivity to image quality problems.

3. Image processing algorithms

3.1. Hair removal

Everything that might corrupt the image and hence the image processing results must be localized and then removed, masked, or replaced. One of the most undesirable elements that is sometimes present in dermatoscopic images is hair. While blond hair can be left without problem for

most of our processing schemes, dark hairs must be masked out. They appear as long and thin dark lines that sometimes cross the whole image. Any algorithm that uses gradient information, such as nonlinear isotropic diffusion [20,21] for example, would be influenced by the presence of hair. Therefore, the image pixels belonging to hair must be replaced by values obtained with some interpolation technique.

A scheme to remove hair from images was proposed by Lee et al. [22]. They argued that a software solution to the hair problem is better than shaving the lesion area before the acquisition. It is true that from a practical point of view the software solution is certainly more convenient. However, in our opinion, the presence of hair should be avoided. The obvious reason is that the skin behind the hair cannot be recovered. Lesions that are almost entirely covered with hair, as some of those shown in Ref. [22], are useless for diagnosis and should be rejected. The presence of some hair, however, is difficult to avoid, and the scheme we suggest is intended to solve the hair problem in those situations. We strongly believe that in the case of massive hair occlusions a 'physical' razor should be used.

Fig. 2 shows a flow chart of the hair removal scheme. A morphological closing operator [23] is applied to the three components of the $L^*u^*v^*$ uniform color space [24]. A spherical structuring element is used. We chose the $L^*u^*v^*$ color space because the luminance component was best suited to differentiate hair from dark pigment, unlike most of the linear color space transformations that provide a so-called intensity component. A hard threshold is then applied to the difference between the luminance before and after morphological closing. Hair is a light-absorbent body and we found that a constant threshold is adequate to produce satisfactory results. The selected pixels are replaced by their value after morphological closing while other pixels are left unchanged. It may happen that dark pigmented structures are masked out as well, usually inside the lesion. The luminance component L^* made the difference between hair and pigmented structures more pronounced, and it was not necessary to add any further processing step, as is the case in Ref. [22].

¹ The histopathology is obtained after excision and microscopic inspection.



Fig. 3. Image corrupted with hair.

Figs. 3 and 4 show an example of hair removal: the original image is shown in Fig. 3, and the result after hair removal is shown in Fig. 4. Upon close examination of the result, very small color changes are visible where hair was originally present. However, this did not corrupt the boundary detection and the visual inspection done by the dermatologists. We did not perform any study on the accuracy of the hair removal scheme for two main reasons: it is an exhaustive task to mask out by hand hair in those images, and the hair removal algorithm must be used in few cases only, especially when great care was taken during the acquisition.

3.2. Detection of the lesion boundary

Boundary detection is a difficult problem in dermatoscopic images because the transition between the lesion(s) and the surrounding skin is smooth and difficult to see accurately, even for a trained dermatologist. In the literature, most of the proposed techniques work with macroscopic, i.e. clinical, images [25–28]. In this kind of image, the contrast between a pigmented lesion and the surrounding skin is higher and simple methods such as histogram thresholding can be sufficient to detect the lesion. In the case of dermatoscopic images, more sophisticated

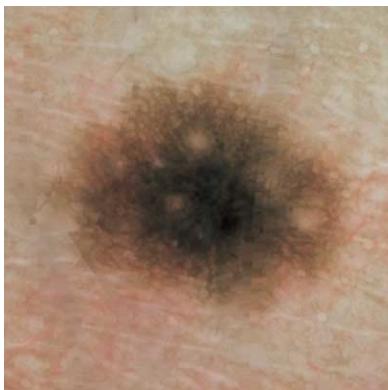


Fig. 4. Image after hair removal.

techniques must be developed to achieve this result. Several schemes taken from the literature were evaluated in Refs. [29,30] and an edge focusing algorithm in [31]. However, the different approaches do not perform well when the transition between the lesion and the surrounding skin is very smooth, or when the lesion contains different shades of color. We developed two unsupervised algorithms that can be used to obtain the lesion boundary: a segmentation technique and a boundary detection technique.

The segmentation technique is based on the clustering of a two-dimensional color space. A modified fuzzy c-means (FCM) [32] technique called *orientation-sensitive fuzzy c-means* (OS-FCM) was introduced in a previous publication [33] and used to perform the clustering. The two *principal components* [23] of the $L^*u^*v^*$ uniform color space are used to construct a two-dimensional histogram. The principal component decomposition, also called Karhunen-Loève (KL) transform, projects the data onto orthogonal directions computed according to a maximum variance criterion. The two components with highest variance are kept. A hierarchical tree representation of the different maxima along with a *strength* value are used to select those maxima corresponding to valid clusters [33].

Let c be the number of clusters decided by the maxima selection algorithm, \mathbf{x}_k the color vector of a pixel, and \mathbf{v}_j a cluster center randomly initialized. The OS-FCM is a two-step algorithm: first, the so-called fuzzy membership u_{ik} of each data sample to each cluster is computed with the following equations:

$$u_{ik} = \left(1 + \sum_{\substack{j=1 \\ j \neq i}}^c \left(\frac{(\mathbf{x}_k - \mathbf{v}_i)^T \mathbf{A}_i (\mathbf{x}_k - \mathbf{v}_i)}{(\mathbf{x}_k - \mathbf{v}_j)^T \mathbf{A}_j (\mathbf{x}_k - \mathbf{v}_j)} \right)^{\frac{1}{m-1}} \right)^{-1}, \quad (1)$$

where the matrices \mathbf{A}_i used for the distance computation must take into account the local orientation of the i th cluster. For each cluster we define a *fuzzy covariance matrix* \mathbf{C}_i :

$$\mathbf{C}_i = \frac{\sum_{k=1}^n (u_{ik})^m (\mathbf{x}_k \mathbf{x}_k^T - \mathbf{v}_i \mathbf{v}_i^T)}{\sum_{k=1}^n (u_{ik})^m}.$$

Let \mathbf{L}_i be a diagonal matrix containing the inverse of the eigenvalues of \mathbf{C}_i and \mathbf{V}_i the matrix whose lines are the corresponding eigenvectors. We can then define the following distance matrix:

$$\mathbf{A}_i = \mathbf{V}_i^T \mathbf{L}_i \mathbf{V}_i. \quad (2)$$

In a second phase, cluster centers are updated according to:

$$\mathbf{v}_i = \frac{\sum_{k=1}^n (u_{ik})^m \mathbf{x}_k}{\sum_{k=1}^n (u_{ik})^m}. \quad (3)$$

The iterative process using Eqs. (1)–(3) is called the *orientation-sensitive fuzzy c-means* (OS-FCM) [33]. Once the clustering converges, each pixel is labeled with the index

of its highest fuzzy membership, leading to the segmentation result.

The boundary detection technique is based on color diffusion and morphological flooding [34]. Let $v(x, y)$ be the pixel value at location (x, y) . Nonlinear isotropic diffusion is controlled by the following equation:

$$\frac{\partial v_i}{\partial t} = \nabla \cdot (g(|\nabla v_1|, \dots, |\nabla v_n|) \nabla v_i), \quad i = 1, \dots, n,$$

where n is the number of color components and $g(\cdot)$, called the *conductivity function*, controls the diffusion process according to the underlying image gradient. This function is given by:

$$g(\cdot) = \begin{cases} 1 & \text{if } x \leq k_0 \\ k_0/x & \text{if } x > k_0 \end{cases} \quad (4)$$

where k_0 is a threshold value. The diffusion of color images is done in the $L^*u^*v^*$ color space with $\sqrt{\sum_i |\nabla v_i|^2}$ as the argument to the conductivity function g . Note that the diffusion is done separately for each component.

The main advantage of the color diffusion is that every frame is marked with the edge information contained in the other frames. Since the pigment is the main light-absorbing component of skin, structures are best rendered by the luminance component. Therefore only the luminance component L^* is kept for the selection of the highest energy boundary. *Morphological flooding* [34] is used to extract the iso-level boundaries, and the final selection is done by minimizing an energy function based on the mean and variance of the gradient values along each boundary [34].

The results obtained with these two algorithms are validated in Section 6.

3.3. Quantification of the degree of symmetry

In this study we have focused on symmetry quantification to classify pigmented skin lesions. Symmetry of shape, color, and texture were computed separately because there is no proof that these three components are correlated, especially in almost-symmetrical and asymmetrical objects. According to clinical research, it is clear that symmetry (or asymmetry) alone is not sufficient to discriminate between benign and malignant lesions. However, we want to show that the symmetry quantification technique we developed gives better classification results than other existing approaches.

Symmetry has never been seriously exploited and only very simple approaches were used. In Ref. [35] the authors use the principal component decomposition of a binary mask of the lesion to compute two orthogonal symmetry axes. An asymmetry index is computed for both axes using the following definition:

$$\text{Asymmetry index} = \frac{\Delta A}{A}, \quad (5)$$

where ΔA is the nonoverlapping area between the original and reflected masks, and A is the area of the original mask. This index is computed for both axes. In Ref. [35] only the minimum value is kept.

Another approach based on the principal component decomposition is used in Ref. [13]. The image is rotated so that its principal component coincides with the x axis. The asymmetry index is then computed as follows [13]:

$$\text{Asymmetry index} = \frac{\sum |v(\mathbf{x}) - \tilde{v}(\mathbf{x})|}{\sum v(\mathbf{x})}, \quad (6)$$

where $v(\mathbf{x})$ is the original image and $\tilde{v}(\mathbf{x})$ the reflected version. The index is computed for both axes and only for pixels inside the lesion.

Most of the existing techniques for the detection of symmetries in images or objects cannot be used here, because the lesions usually have poor symmetry characteristics. The symmetry problem is considered here as an optimization problem, where a given symmetry measure must be maximized. This symmetry measure is a function of the *mean square error* (MSE) between an image and its reflected version, given a particular symmetry axis [36,37]. If we suppose that any object, for a given axis, is the combination of a symmetric object and an asymmetric object, and that the latter is considered as a noise component, then the *peak signal to noise ratio* (PSNR) can be used as symmetry measure:

$$\text{PSNR} = 10 \log_{10} \left(\frac{(N_q - 1)^2}{\text{MSE}} \right),$$

where N_q is the number of quantization levels in the image and MSE is the mean square error between the original image and its reflected version. It is given by:

$$\text{MSE} = E[\|\mathbf{v}(\mathbf{x}) - \mathbf{v}(\tilde{\mathbf{x}})\|^2],$$

where $\mathbf{v} : \mathbb{D}_v \mapsto \mathbb{R}^n$ is the vector-valued input image, n the data dimension, $\mathbb{D}_v \subset \mathbb{Z}^2$ is the object of interest domain, $\mathbf{x} \in \mathbb{D}_v$ is the pixel coordinate in the original image and $\tilde{\mathbf{x}} \in \mathbb{R}^2$ is the symmetric coordinate. Note that the MSE is proportional to the energy of the asymmetric part of the image, which is given by $\sum_x \|\mathbf{v}(\mathbf{x}) - \mathbf{v}(\tilde{\mathbf{x}})\|^2/4$. Finally, the PSNR is used to define the following symmetry coefficient:

$$\psi = \psi(\varphi, \rho, \mathbf{v}) = 1 - \frac{1}{1 + \text{PSNR}}, \quad (7)$$

where φ and ρ are two axis parameters. We restricted our measurements to the set of pixels belonging to a particular lesion. Pixels surrounding the lesion belong to a homogeneous background and are therefore highly symmetrical. They would corrupt significantly the symmetry value if they were used. Therefore, we assume here that the segmentation problem was previously solved and that pixels belonging to a lesion were masked out.

In order to take into account shape, color, and texture information, the algorithm needed for the symmetry

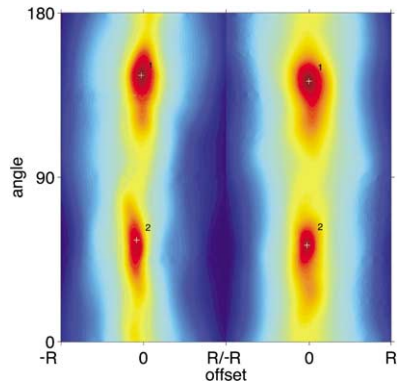


Fig. 5. Texture (left) and color (right) symmetry maps of symmetric lesion.

analysis of pigmented skin lesion must be able to process those sources of information. To this end, multidimensional descriptors are used for the color and texture components [36,37]. In the former case filtered versions of the L^* , u^* , and v^* color components were used, and in the latter case a bank of five isotropic Gaussian filters was used to describe texture information. Intensive simulations showed that these descriptors perform well even in a noisy environment [36, 37]. For the shape symmetry, a binary mask of the lesion was processed in place of the image.

Several approaches were proposed to reduce computation time needed for the optimization process. The axis that gives the highest symmetry coefficient can be computed with a genetic algorithm or an algorithm based on self-organizing maps [36,37]. In this paper, we will focus our attention on the computation of *symmetry maps* (SM) [37]. We define a symmetry map as a 2D representation of the axis parameters. They allow characterization of the symmetry of any object. Multiple local symmetry axes, even those having a poor symmetry value, can be detected and used for later analysis. To compute these maps, two multiresolution approaches were used: one is in the image domain and the other is in the symmetry map domain [36,37]. In the former case, a pyramidal image representation is used, while in the latter approach the sampling grid of the symmetry map is iteratively refined in regions of the parameter space that have high symmetry value.

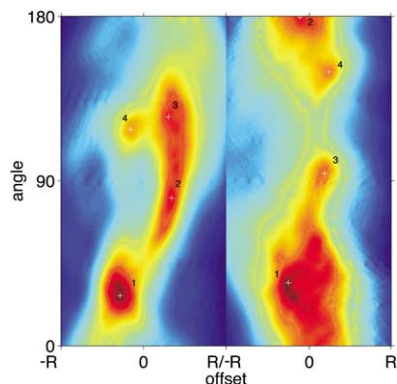


Fig. 6. Texture (left) and color (right) symmetry maps of asymmetric lesion.

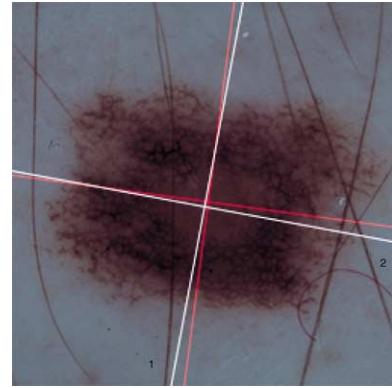


Fig. 7. Axes obtained from the map of Fig. 5 (red for color, white for texture).

Figs. 5 and 6 show two examples of *texture* symmetry maps, where the maxima were numbered in decreasing order according to their symmetry value. The axes corresponding to the local maxima are superimposed on the original images in Figs. 7 and 8, respectively. In the former case the lesion is benign and symmetrical, while in the latter case it is malignant and asymmetrical. In both cases, at least one axis is found, namely the axis that minimizes the MSE between the original and reflected images. Therefore, even asymmetrical images have a symmetry axis, but the symmetry value is low. It is interesting to note that when lesions are asymmetrical they have multiple local minima, as seen in Fig. 8. While we cannot yet give any scientific explanation of that phenomenon, this characteristic might be exploited in future studies.

Section 5 gives classification results obtained with this approach.

4. Validation of boundary detection results

4.1. Visual assessment

In this section different data sets are shown as *box and whisker* plots [38]. The boxes are defined by the lower

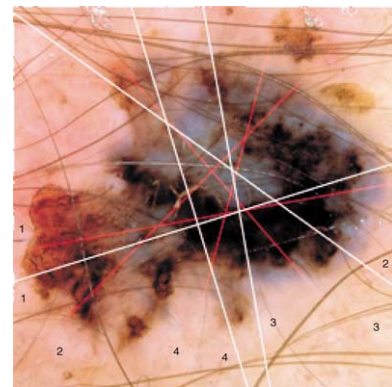


Fig. 8. Axes obtained from the map of Fig. 6 (red for color, white for texture).

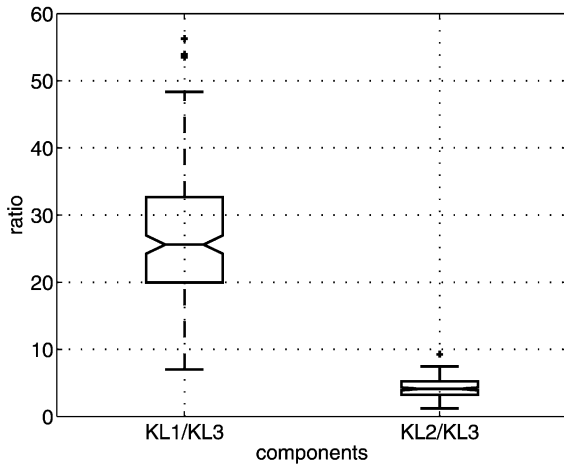


Fig. 9. Ratios between variance of the different KL components in the RGB color space.

quartile, median, and upper quartile values. The whiskers are lines beginning at each end of a box that show the extent of the rest of the data. They are computed as a fraction (here 1.5) of the difference between the upper and lower quartiles. However, the whisker limits cannot extend beyond the smallest and largest values. Outliers (+) are values outside the limits set by the whiskers. Finally, the boxes are notched. Notches represent a robust estimate of the uncertainty about the means for box to box comparison. They are computed from the median value at a distance equal to a fraction of the difference between the upper and lower quartiles, normalized by the square root of the data size.

In Figs. 9 and 10, the ratios between the variance of the different KL components taken pair-wise are plotted, where KL1, KL2, and KL3 are the variance values in decreasing order. They are computed in the RGB and $L^*u^*v^*$ color spaces, respectively, and show that when the variance is taken as an indicator of ‘information content’, then most of

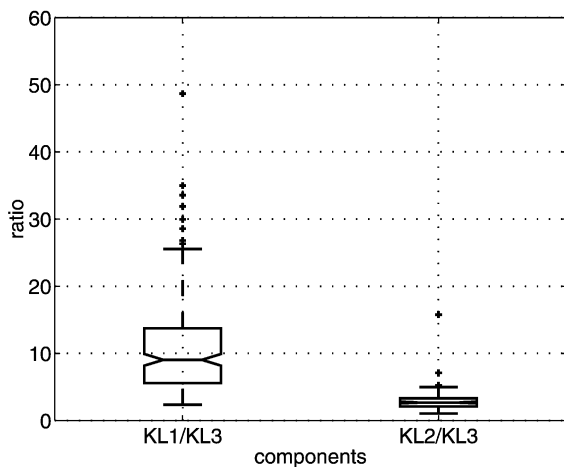


Fig. 10. Ratios between variance of the different KL components in the $L^*u^*v^*$ color space.

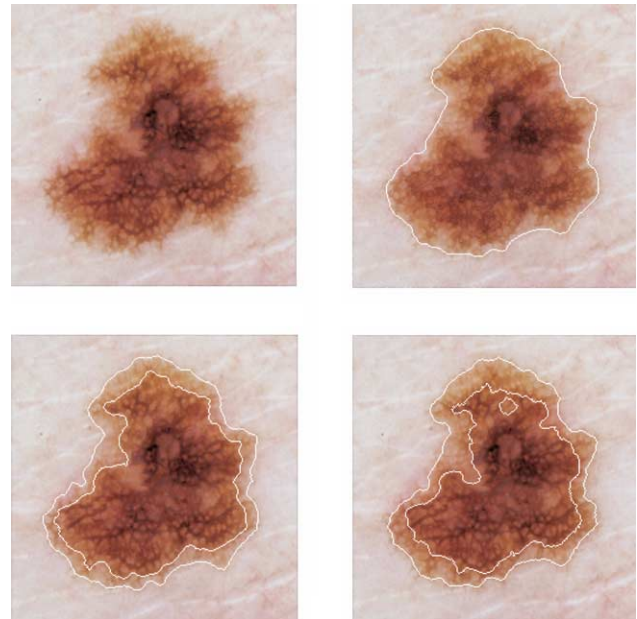


Fig. 11. Detection of a pigmented skin lesion: the original lesion is shown top-left, the boundary obtained with nonlinear isotropic diffusion and morphological flooding is shown top-right, and the left and right-bottom images were obtained with the FCM and OS-FCM clustering algorithms, respectively.

the information in projected onto the first two components. This phenomenon is less pronounced in the $L^*u^*v^*$ color space, but simulation results showed that the third component contains mainly noise. Two hundred unclassified lesions were used to produce those two plots.

Fig. 11 shows segmentation and boundary detection results. The original image is shown top-left, the boundary obtained by nonlinear isotropic diffusion and morphological flooding is shown top-right, while the left and right bottom images are the segmentation results obtained, respectively, with the FCM and OS-FCM techniques. A complex image that contains hair, crusts, and a pigmented lesion with smooth borders and several regions with different colors (bluish, dark brown, and light brown) is shown in Fig. 12. This image did not need to be pre-processed with the hair removal algorithm, because the hairs are blond and do not hinder the boundary detection. The segmentation obtained with the OS-FCM technique is shown bottom-left, the boundary detection result is shown top-right, and the bottom-right image shows the boundaries corresponding to all local minima of the energy functional used for the boundary detection [34]. This last example shows that regions inside the lesion can be detected as well with this technique. Finally, the improvement gained with the OS-FCM over the more classical FCM technique is best visible in Figs. 13 and 14, where both approaches were used. The reddish skin corresponds to an elongated cluster in the two-dimensional histogram, which can be best detected with the OS-FCM technique.

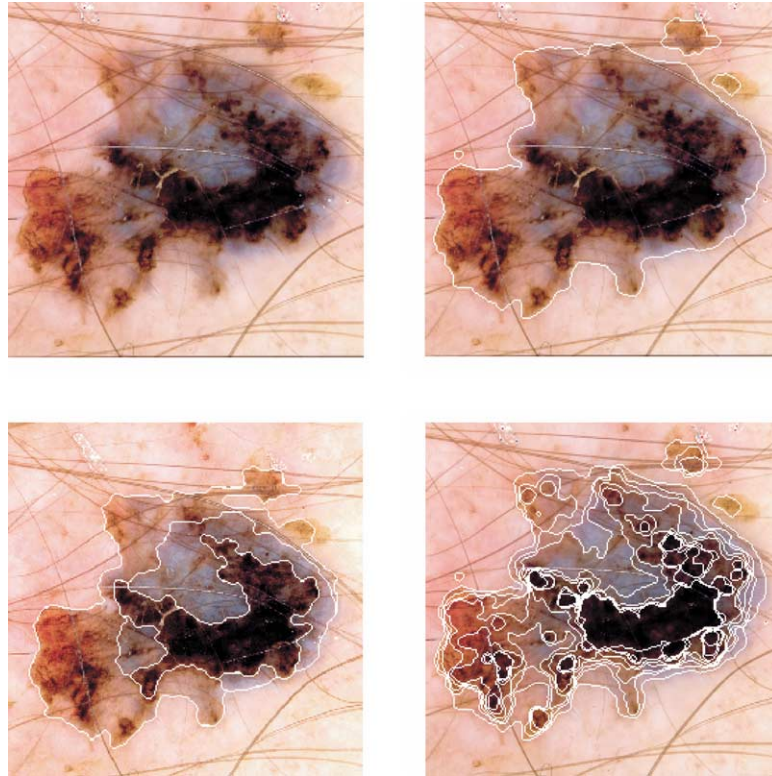


Fig. 12. Detection of a complex pigmented lesion: the original lesion is shown top-left, the boundary obtained after nonlinear isotropic diffusion and morphological flooding is shown top-right, the bottom-right image shows the different curves corresponding to all local minima of the energy functional, and the segmentation by OS-FCM clustering is shown bottom-left.

Visual assessment was performed by expert dermatologists on the 200 cases by hand. The overall impression was very positive and showed that the two techniques we used always gave results considered as acceptable boundaries, i.e. boundaries that can be considered as accurate solutions to the lesion detection problem. None of the detected boundaries was considered as being clearly wrong. In order to get a more objective impression of the performances of our segmentation and boundary detection schemes, a statistical comparison was made with boundaries drawn

by hand by five expert dermatologists. The results we obtained are presented in the following paragraph.

4.2. Objective validation

For dermatoscopic images, the lesion border is difficult to establish, and a boundary drawn manually by a dermatologist cannot be used as an absolute reference. Any human being, expert or not, is unable to reproduce precisely his drawings. However, showing that results provided by automatic techniques are statistically close to those of a group of experts is an acceptable validation.

Five expert dermatologists were asked to draw by hand the boundary of the lesion in 25 images. These images were selected for the smoothness of their border. This experiment was repeated three times by each physician at intervals of several days, in order to evaluate reproducibility of their answer to the boundary problem. Two examples of manual drawings are shown in Figs. 15 and 16. One can note that the boundary location is uncertain in regions where the transition from the lesion to the surrounding skin is very smooth and where the boundary is concave.

Let A_i be the set of all pixels inside the lesion for the i th manual drawing. In order to evaluate the consistency of the different results, the analysis was made of the following steps:

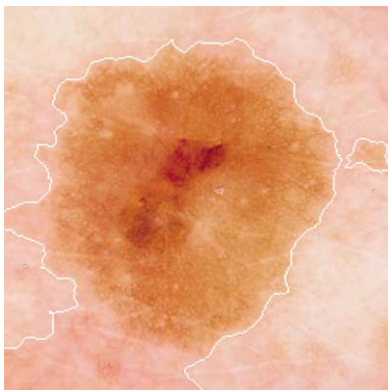


Fig. 13. Pigmented skin lesion segmented with the FCM technique.

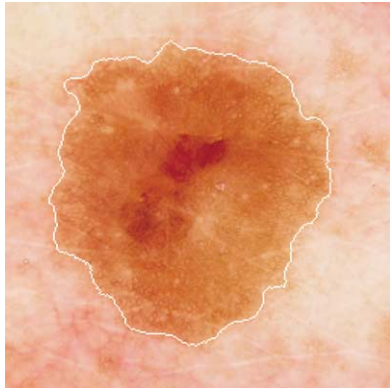


Fig. 14. Same pigmented skin lesion as shown in Fig. 13, segmented with the OS-FCM technique.

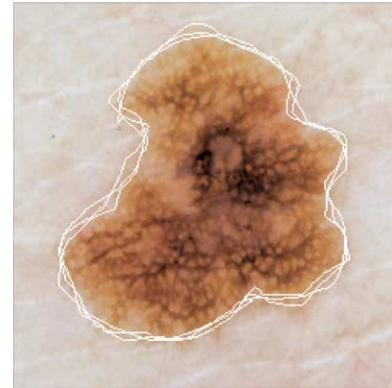


Fig. 16. Three boundaries drawn by physician (B).

- Take the sets pair-wise and compute the following two subsets: $A_i \text{ XOR } A_j$ and $A_i \text{ OR } A_j$.
- Compute the ratio of the size of the two subsets.

This procedure was repeated for the all three masks. In that case, the second subset was replaced by: $(A_1 \text{ OR } A_2 \text{ OR } A_3) \text{ AND NOT } (A_1 \text{ AND } A_2 \text{ AND } A_3)$.

These ratios are equal to zero only when the boundaries are strictly identical. Figs. 17 and 18 shows as box plots the results obtained for two dermatologist (A) and (B). These two physicians were the most and least consistent, respectively. When comparing the three days, the values are higher because the error is different from one day to another and is therefore summed up. The immediate and expected conclusion is that hand-drawn boundaries in dermatoscopic images are not consistent enough to be used as absolute references. However, it is interesting to find out how our segmentation and boundary detection schemes perform compared to the group of experts.

The approach used for the inter-physician measurements was slightly different. For each image, the probability for a pixel to be inside the lesion was computed, based on the statistics provided by the manual and automatic boundaries. Such a probability image was computed for each lesion.

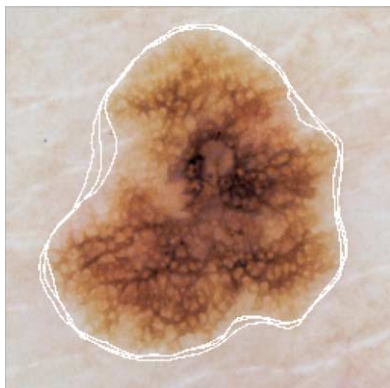


Fig. 15. Three boundaries drawn by physician (A).

An example is shown in Fig. 19, where the probability that a pixel is inside the lesion is 1 when its color is black and 0 when its color is white. This image gives a within-group statistic for each pixel and can be used to evaluate individually the different boundaries. Only those pixels which have a nonzero probability to be inside the lesion are taken into account. Then, for each lesion, day, and physician, the mean probability that pixels were misclassified is computed, based on the probability image. This measure allows one to say if a boundary is consistent with respect to the whole group of experts. Fig. 20 shows the results obtained for the two techniques we developed. The clustering technique gives results that are closer to the boundaries drawn by the physicians than the boundary detection technique using nonlinear isotropic diffusion and morphological flooding. While the latter has a low median value, the misclassification error grows significantly in some cases. This result is certainly due to the fact that the clustering technique is based on criteria close to those used by physicians.

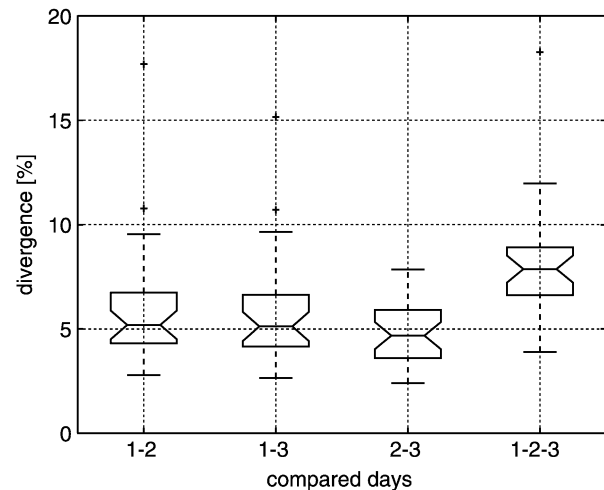


Fig. 17. Intra-physician comparison: lesion boundaries of 25 lesions are compared pair-wise according to the day they were made. The error or divergence is displayed as box plots (A).

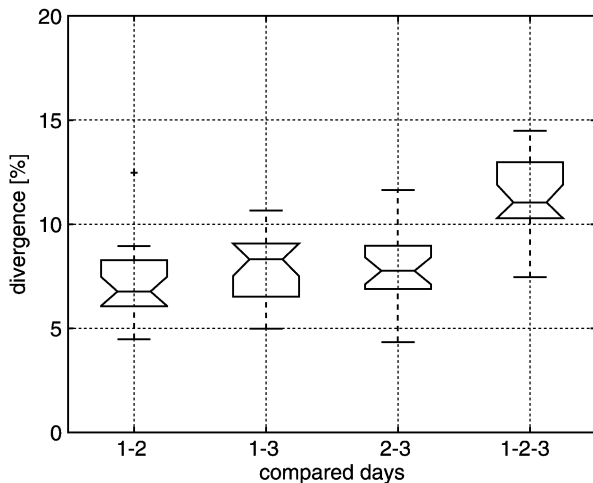


Fig. 18. Same as Fig. 17 for physician (B).

Figs. 21 and 22 show the results obtained for physicians (A) and (B), respectively. In both cases, the box plots show that the misclassification error is generally higher than that obtained for the automatic techniques. The most interesting fact is that physician (A), who was the most able to reproduce his results, obtains the worst results in this case. The criteria used by physician (A) to draw a boundary are certainly different from those used by the other physicians, but he is able to reproduce quite precisely his results. It is now difficult to say if physician (A) is a good reference because he is consistent, or a bad one because his results diverge from the others.

Finally, we will conclude that the validation of boundary detection and segmentation techniques cannot be performed with a single reference. The objective of our research is to avoid human subjectivity in specific tasks, and the use of a hand drawn boundary as a reference is therefore contradictory. However, from the results presented in this section, we can claim that our segmentation and boundary detection techniques perform as well as expert dermatologists.

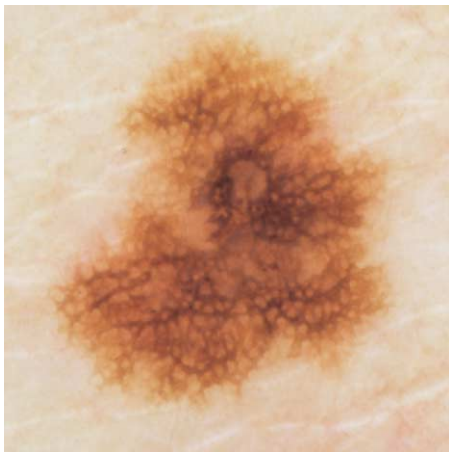


Fig. 19. Probability image obtained from the hand drawn and automatic boundaries.

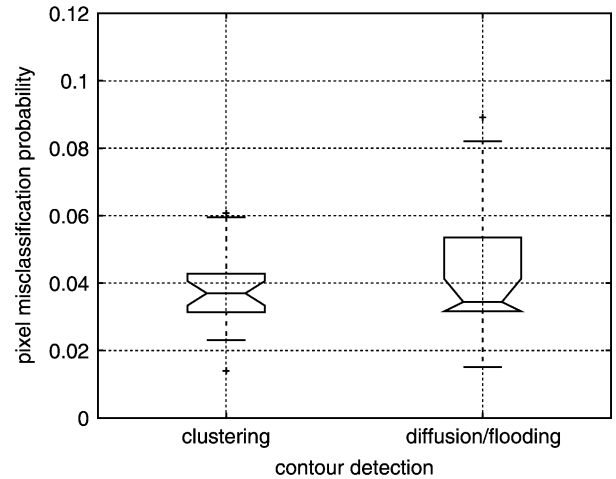


Fig. 20. Probability that an image pixel was misclassified using our schemes.

The segmentation technique seems to reduce misclassification errors when compared to the manual drawings.

5. Classification results

Linear classification results that were obtained with symmetry maps and techniques based on the principal component decomposition are presented. In the former case, three SMs are available, one for shape, one for color, and one for texture symmetry, and in the latter case, binary and masked gray-level images are used. The gray level principal component decomposition is computed for the luminance component L^* . In each case, two parameters are extracted: the two largest local maxima in the case of the SM, and the two asymmetry indexes given by Eqs. (5) and (6), respectively. In this initial study, we had 50 images of malignant melanoma and 50 images of benign nevi. In this second set, 20 lesions were *dysplastic nevi*, which are considered to be a particular class of benign lesions. For some experts, the dysplastic nevi are intermediate lesions

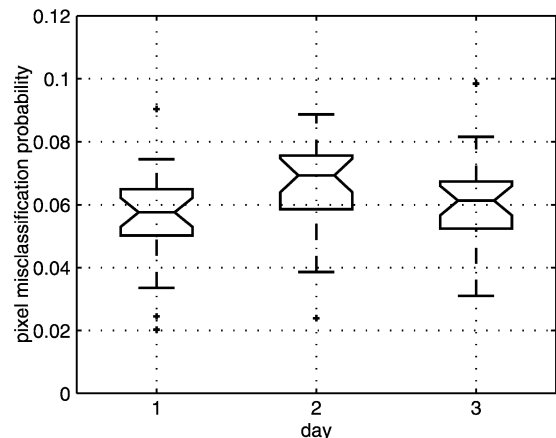


Fig. 21. Probability that an image pixel was misclassified (A).

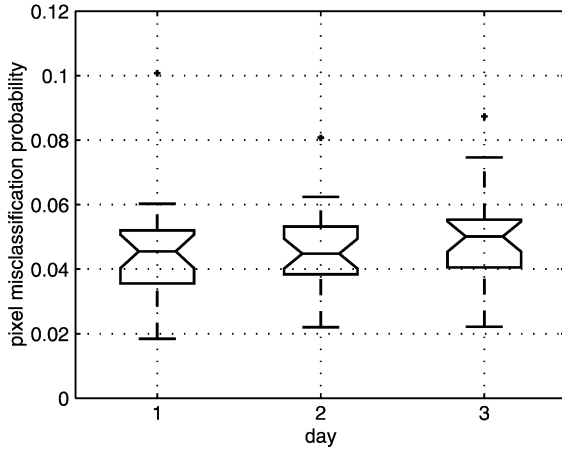


Fig. 22. Probability that an image pixel has been misclassified (B).

between melanocytic lesions and malignant melanoma, and some of them progress to become malignant melanoma [2]. This kind of lesions have different characteristics than other benign lesions.

A linear classifier with training by epoch [39] was used to classify lesions according to their symmetry values. This method uses a gradient descent approach to determine the parameters of the hyperplane that minimizes the misclassification error. We found that this classifier was adequate for our dataset. A more sophisticated classifier will be used when more features are quantified and more data becomes available.

Figs. 23 and 24 show the minimum MSE found in the texture and color symmetry maps of 50 MM (×) and 50 nevi (○), using the lesion masks obtained with our segmentation and boundary detection schemes, respectively. We notice that the boundary detection affects only the SM computation but not the classification results. Finally, we conclude from these plots that even if the limit between both classes is not sharp, texture seems to be a better discriminant.

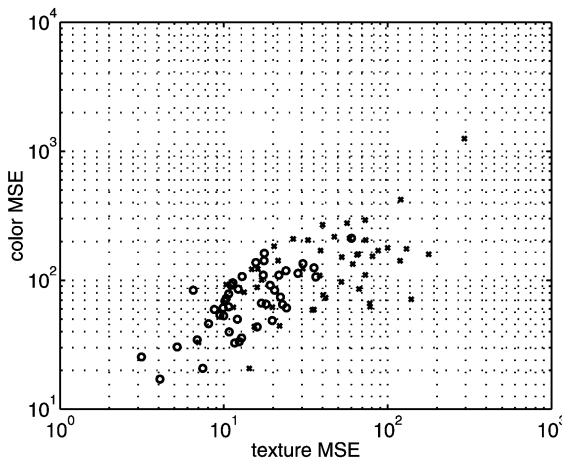


Fig. 23. Color and texture symmetry MSE of 50 malignant melanoma (×) and 50 nevi (○). The lesion masks have been obtained through the clustering technique.

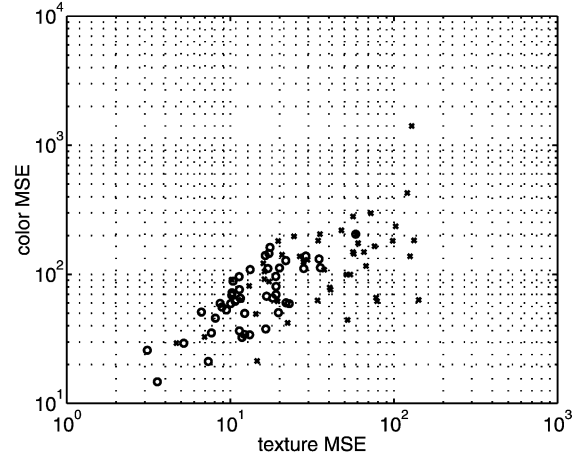


Fig. 24. Color and texture symmetry MSE of 50 malignant melanoma (×) and 50 nevi (○). The lesion masks have been obtained through the boundary detection technique.

A number of measurements can be used to evaluate the classification accuracy:

$$\text{sensitivity} = \frac{TP}{TP + FN},$$

$$\text{specificity} = \frac{TN}{TN + FP},$$

$$\text{positive predictive value} = \frac{TP}{TP + FP},$$

$$\text{negative predictive value} = \frac{TN}{TN + FN},$$

$$\text{index of suspicion} = \frac{TP + FP}{TP + FN},$$

$$\text{diagnostic accuracy} = \frac{TP}{TP + FP + FN},$$

where TP is the number of true positives, FN the number of false negatives, TN the number of true negatives, and FP the number of false positives. A *positive* is a lesion that is classified as being malignant. The *sensitivity* and *specificity* measure the percentages of accurate classifications for the malignant and benign cases, respectively. The positive predictive value is the proportion of lesions classified as MM that are really MM, and inversely for the negative predictive value. The index of suspicion gives the degree of awareness of the likelihood of a lesion to be a MM. An index higher than 100% represents an over-diagnosis while an index lower than 100% represents an *under-diagnosis*. The *diagnosis accuracy* shows the loss of accuracy due to misclassified elements.

Table 1 shows the different classification accuracy indices for three combinations of symmetry features: texture (2D), texture and color (4D), and texture, color, and shape (6D). In Table 2 the results obtained without the dysplastic nevi (considered as benign lesions) are given. These results show that the combination of shape, color, and texture improves the separation between benign and malignant

Table 1
Linear classification of symmetry values for malignant melanoma and nevi, including dysplastic nevi

Symmetry features	Texture (%)	Texture–color (%)	Texture–color–shape (%)
Sensitivity	66	64	78
Specificity	90	92	90
Positive predictive value	86.8	88.9	90.7
Negative predictive value	72.6	71.9	80.4
Index of suspicion	76	72	86
Diagnostic accuracy	60	59.3	72.2

lesions. Shape and texture are especially relevant, while the color symmetry does not improve the results significantly. Malignant melanoma is shown to be rather asymmetrical, but this criterion is not sufficient to separate malignant from benign lesions.

An important study that must follow these initial results is the evaluation of different texture and color descriptors. The features used in this study have shown to work well on synthetic images and in noisy environments [36,37] but may not be optimal for dermatoscopic images.

Table 3 gives the results obtained for the principal component approach. The binary approach gives weak results while the gray-scale approach improves significantly the classification. However, these two approaches are very sensitive to the shape of the lesion mask. This is especially the case for the binary approach, which is intuitively appealing.

It is interesting to note from the above results that a very simple approach such as the gray-level principal component decomposition gives slightly better results than the texture symmetry computed with the SM. One possible reason is that the former method uses asymmetry indexes computed from two orthogonal axes, and thus the asymmetry criterion is evaluated in two orthogonal directions. In the SM case, the axes corresponding to the two smallest MSE are used. They correspond to two axes which are far from orthogonal, especially in the case of a malignant melanoma. The evaluation of symmetry in two orthogonal directions might

Table 2
Linear classification of symmetry values for malignant melanoma and nevi, without dysplastic nevi

Symmetry features	Texture (%)	Texture–color (%)	Texture–color–shape (%)
Sensitivity	76	78	82
Specificity	83.3	80	80
Positive predictive value	88.4	86.7	87.2
Negative predictive value	67.6	68.6	72.7
Index of suspicion	86	90	94
Diagnostic accuracy	69.1	69.6	73.2

Table 3
Linear classification of malignant melanoma and nevi, including dysplastic nevi, using principal component decompositions

Symmetry features	Binary (%)	Gray-level (%)
Sensitivity	60	70
Specificity	82	92
Positive predictive value	76.9	89.7
Negative predictive value	67.2	75.4
Index of suspicion	78	78
Diagnostic accuracy	50.8	64.8

be an important issue. It is therefore important to extend this work with the extraction of all the local minima present in a SM and to quantify different characteristics such as their respective location. According to what we could observe, it may be a very promising direction to the symmetry problem in dermatoscopic images.

In order to employ a more complex classifier, such as a neural network for example, more data is needed. However, the simple classifier we used so far illustrates that this approach is encouraging and that the integration of other diagnostic features may lead to a better separation between benign and malign lesions. It is therefore necessary to give more importance to the development of efficient image processing algorithms than to the classifier itself, at least in a first phase.

6. Conclusions

The different sections propose solutions to specific problems related to CAD of pigmented skin lesions. While the lesion detection problem is solved, there is still much to do for the extraction of diagnostic features. In this study we focus on a single criterion for classification purpose, namely symmetry. We know from skin cancer research that a unique feature is not sufficient to diagnose precisely skin cancer, and that the combination of different criteria is the key to the early detection of malignant melanoma and other types of skin cancer.

For the symmetry quantification, a deeper analysis of symmetry maps is necessary, and we believe that the classification into benign and malignant lesions can be improved even more. The different maxima in the symmetry maps are important feature points, and their respective importance and location must be analyzed and measured. We could observe that benign lesions have fewer number of local maxima than malignant lesions, and that in benign cases the shape, color, and texture maps show maxima that are distributed in a similar fashion.

Among the different criteria used for the clinical diagnosis, one of the most interesting and promising is the detection of specific patterns in dermatoscopic images. This task, which is also the most difficult one, should allow the detection and analysis of different pigmented structures in terms of regularity, size, etc. We are not aware of any

relevant work in this area, and future investigations should allow us to answer this question.

An efficient user interface is an important component of any CAD system. Such an interface should allow users to access the different image processing algorithms, to visualize the results, and to make basic manipulations. Network and database management are also very important issues that must be investigated in order to build a complete system. In the framework of this research, we wrote a client/server applications in Java. The idea was to have the graphical interface on the user side, and the database and image processing routines on the server side. This approach seems to be the most promising. Another approach, which uses Matlab as platform, has been evaluated as well. In that case, a graphical user interface allows to the different image processing routines, and the results are displayed using Matlab's graphical user interface tools. This solution, however, requires the user to have a good knowledge of Matlab programming.

7. Summary

In this paper we present a CAD system for pigmented skin lesions. Dermatoscopic images obtained by ELM are processed to obtain the lesion boundary and to quantify the degree of symmetry of the lesion. Section 1 introduces the topic and Section 2 presents the flow-chart of a complete system.

To facilitate the subsequent segmentation and feature quantification steps, we introduce a novel scheme in Section 3.1 to remove hair from the images. The segmentation and boundary detection schemes developed in this study are explained and illustrated in Section 3.2. A statistical evaluation based on the comparison with hand drawn boundaries, obtained from a group of trained dermatologists, is presented in Section 4.2. We show that a physician, even when trained for this type of diagnosis, cannot be used as a gold standard. Human beings are usually not able to reproduce precisely measurements and the comparison can therefore only be done with a group of experts. In such cases, the results obtained with the developed techniques must be as close as possible to that of the experts. We show that the within-group error obtained with the automatic schemes, and especially that of the segmentation scheme, is generally lower than that obtained by each physician taken alone.

The scheme used to quantify the shape, color, and texture symmetry is evaluated in Section 5. We observe that benign lesions often have only two well separated minima (almost orthogonal), while multiple minima are present in SMs of malignant lesions. It is shown in Section 3.3 that the combination of texture, color, and shape symmetry values improves significantly the separation between benign and malignant lesions. Results obtained with principal components decompositions are used for comparison.

Acknowledgements

The authors would like to thank Dr David Guggisberg, and Dr Jean-Philippe Cerottini from the CHUV-DHURDV, Lausanne, Switzerland, Dr Ralph Braun, Dr Joakim Krischer, and Prof. Jean-Hilaire Saurat from the HUG, Geneva, Switzerland, for their participation to the evaluation of the segmentation and boundary detection techniques. The authors also thank warmly Prof. Murat Kunt for encouraging and supporting this project. This research was funded by the Swiss National Science Foundation, fund no. 3252-053175.

References

- [1] Burg G. *Das Melanom*, Serie Gesundheit. Weinheim: Piper/VCH; 1993.
- [2] Stolz W, Braun-Falco O, Bilek P, Landthaler M, Cogeneta AB. *Color atlas of dermatoscopy*. Oxford: Blackwell; 1994.
- [3] Ercal F, Chawla A, Stoecker WV, Lee HC, Moss RH. Neural network diagnosis of malignant melanoma from color images. *IEEE Trans Biomed Engng* 1994;41(9):837–45.
- [4] Argenyi ZB. Dermatoscopy (epiluminescence microscopy) of pigmented skin lesions. *Dermatol Clinics* 1997;15(1):79–95.
- [5] Binder M, Kittler H, Seeber A, Steiner A. Epiluminescence microscopy-based classification of pigmented skin lesions using computerized image analysis and an artificial neural network. *Melanoma Res* 1998;8:261–6.
- [6] Hintz-Madsen M, Kai Hansen L, Larsen J, Olesen E, Drzewiecki KT. Design and evaluation of neural classifiers application to skin lesion classification. *IEEE Workshop on Neural Networks for Signal Processing (NNSP'95)*; 1995. p. 484–93.
- [7] Kjoelen A, Thomson MJ, Umbaugh SE, Moss RH, Stoecker WV. Performance of AI methods in detecting melanoma. *IEEE Engng Med Biol* 1995;14(4):411–6.
- [8] Binder M, Steiner A, Schwarz M, Knollmayer S, Wolff K, Pehamberger H. Application of an artificial neural network in epiluminescence microscopy pattern analysis of pigmented skin lesions: a pilot study. *Br J Dermatol* 1994;130:460–5.
- [9] Ercal F, Lee HC, Stoecker WV, Moss RH. Skin cancer diagnosis using hierarchical neural networks and fuzzy systems, vol. 4. *Proceedings of the Artificial Neural Networks in Engineering conference (ANNIE '94)*; November 1994. p. 613–18.
- [10] Bostock RRJ, Claridge E, Harget AJ. Towards a neural network based system for skin cancer diagnosis. *IEE International Conference on Neural Networks*; 1993. p. 215–19.
- [11] Durg A, Stoecker WV, Cookson J, Umbaugh SE, Moss RH. Identification of variegated coloring in skin tumors; neural networks vs. rule-based induction methods. *IEEE Engng Med Biol* 1993;12(3): 71–4.
- [12] Umbaugh SE, Moss RH, Stoecker WV. Applying artificial intelligence to the identification of variegated coloring in skin tumors. *IEEE Engng Med Biol* 1991;10(4):57–62.
- [13] Gutkiewicz-Krusin D, Elbaum M, Szwaykowski P, Kopf AW. Can early malignant melanoma be differentiated from atypical melanocytic nevus by in vivo techniques? part II. Automatic machine vision classification. *Skin Res Technol* 1997;3:15–22.
- [14] Ganster H, Gelautz M, Pinz A, Binder M, Pehamberger H, Bammer M, Krocza J. Initial results of automated melanoma recognition. In: Borgefors G, editor. *Proceedings of The 9th Scandinavian Conference on Image Analysis*, Uppsala, Sweden: Swedish Society for Automated Image Analysis; 1995. p. 209–18.

- [15] Hall PN, Claridge E, Morris Smith JD. Computer screening for early detection of melanoma—is there a future? *Br J Dermatol* 1995;132:325–38.
- [16] Green A, Martin N, Pfitzner J, O'Rourke M, Knight N. Computer image analysis in the diagnosis of melanoma. *J Am Acad Dermatol* 1994;31:958–64.
- [17] Kontinen J, Rönning J, MacKie RM. Texture features in the classification of melanocytic lesions. In: Del Bimbo A, editor. Ninth International Conference on Image Analysis and Processing (ICIAP'97), vol. 2. Berlin: Springer; 1997. p. 453–60. Florence, Italy.
- [18] Dummer W, Blaheta H-J, Bastian BC, Schenk T, Bröcker E-B, Remy W. Preoperative characterization of pigmented skin lesions by epiluminescence microscopy and high-frequency ultrasound. *Arch Dermatol* 1995;131:279–85.
- [19] Rajadhyaksha M, Zavislan M. Confocal laser microscopy images tissue in vivo. *Laser Focus World* 1997;119–27.
- [20] Perona P, Malik J. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans Pattern Anal Mach Intell* 1990;12(7):629–39.
- [21] You Y-L, Xu W, Tannenbaum A, Kaveh M. Behavioral analysis of anisotropic diffusion in image processing. *IEEE Trans Image Process* 1996;5(11):1539–53.
- [22] Lee T, Ng V, Gallagher R, Coldman A, McLean D. DullRazor: a software approach to hair removal from images. *Comput Biol Med* 1997;27(6):533–43.
- [23] Gonzalez RC, Wintz PA. Digital image processing. Reading, MA: Addison-Wesley; 1993.
- [24] Wyszecki G, Stiles WS. Color science: concepts and methods, quantitative data and formulae. New-York: Wiley; 1982.
- [25] Ercal F, Moganti M, Stoecker WV, Moss RH. Detection of skin tumor boundaries in color images. *IEEE Trans Med Imaging* 1993;12(3):624–7.
- [26] Umbaugh SE, Moss RH, Stoecker WV, Hance GA. Automatic color segmentation algorithms with application to skin tumor feature identification. *IEEE Engng Med Biol* 1993;12(3):75–81.
- [27] Golston JE, Moss RH, Stoecker WV. Boundary detection in skin tumor images: an overall approach and a radial search algorithm. *Pattern Recogn* 1990;23(11):1235–47.
- [28] Umbaugh SE, Moss RH, Stoecker WV. Automatic color segmentation of images with application to detection of variegated coloring in skin tumors. *IEEE Engng Med Biol* 1989;8(4):43–52.
- [29] Gao J, Zhang J, Fleming MG, Pollak I, Cognetta AB. Segmentation of dermatoscopic images by stabilized inverse diffusion equations. Proceedings of the International Conference on Image Processing, vol. 3; 1998. p. 823–27.
- [30] Hance GA, Umbaugh SE, Moss RH, Stoecker WV. Unsupervised color image segmentation; with application to skin tumor borders. *IEEE Engng Med Biology* 1996;15(1):104–11.
- [31] Denton WE, Duller AWG, Fish PJ. Boundary detection for skin lesion: an edge focusing algorithm. *Image Processing and its Applications*. IEEE, 4–6 July, 1995, number 410 in Conference Publication. p. 399–402.
- [32] Bezdek JC. A convergence theorem for the fuzzy ISODATA clustering algorithm. *IEEE Trans Pattern Anal Mach Intell* 1980;2(1):1–8.
- [33] Schmid Ph. Segmentation of digitised dermatoscopic images by 2D colour clustering. *IEEE Trans Med Imaging* 1999;18(2):164–72.
- [34] Schmid Ph. Lesion detection in dermatoscopic images using anisotropic diffusion and morphological flooding. Proceedings of the International Conference on Image Processing (ICIP '99), Kyoto, Japan; 1999. p. 111.
- [35] Stoecker WV, Li WW, Moss RH. Automatic detection of asymmetry in skin tumors. *Comput Med Imaging Graph* 1992;16(3):191–7.
- [36] Schmid Ph. Segmentation & symmetry measure for image analysis: application to digital dermatoscopy. Thesis no. 2045, Swiss Federal Institute of Technology at Lausanne (EPFL), Signal Processing Laboratory (LTS); 1999, can be downloaded from http://ltswww.epfl.ch/pub_files/schmid/.
- [37] Schmid Ph. Symmetry axis computation for almost-symmetrical and asymmetrical objects: application to pigmented skin lesions. *Med Image Anal* 2000;4(3):269–82.
- [38] Chase W, Brown F. General statistics. New York: Wiley; 1992.
- [39] Schalkoff R. Pattern recognition: statistical, structural and neural approaches. New York: Wiley; 1992.

Philippe Schmid-Saugeon received the Electrical Engineering degree and the PhD degree from the Swiss Federal Institute of Technology at Lausanne in 1996 and 1999, respectively. His PhD thesis was on the quantification of diagnostic features for the classification of melanocytic skin lesions. From 1999 to 2000 he was a visiting scientist at the Video and Image Processing Lab of the University of California at Berkeley, where he worked on Matching Pursuits techniques for the compression of video sequences. Since 2000, he is a Consultant and Leading Research Scientist at Medispectra Inc, a company developing devices for the diagnosis of cervical cancer and located in the Boston area. His scientific interests cover everything related to CAD of diseases from digital images or video sequences, such as image filtering, segmentation, pattern recognition, camera calibration, etc. He is a member of the IEEE.

Joël Guillod was born in 1960. He studied medicine in Lausanne and Zürich and received the MD degree in 1985. He worked in the medical informatics field and wrote a thesis related to information sharing between health care professionals for a computer-based patient record. In 1997, he graduated as a specialist in dermatology and venerology. He is currently working in a private dermatologist office and part time as Assistant Professor at the Dermatology Department at the University Hospital in Lausanne (CHUV), Switzerland. His current scientific interests include ambulatory electronic medical record, dermoscopy image processing for the early diagnosis of the malignant melanoma skin cancer, telemedicine and diving medicine.

Jean-Philippe Thiran was born in Namur, Belgium, in August 1970. He received the Electrical Engineering degree and the PhD degree from the Université catholique de Louvain (UCL), Louvain-la-Neuve, Belgium, in 1993 and 1997, respectively. His PhD thesis was related to 3D medical image registration. From 1993 to 1997, he has been the co-ordinator of the image analysis group of the Communications and Remote Sensing Laboratory at UCL, mainly working on medical image analysis. Dr Thiran is currently Senior Lecturer (Maître d'Enseignement et de Recherche) and the leader of the Computer Vision and Image Analysis Group at the Signal Processing Institute of the Swiss Federal Institute of Technology (EPFL) in Lausanne, Switzerland. His current scientific interests include image segmentation, a priori information in image analysis, PDE's in image analysis, multimodal signal processing, medical image analysis, including multimodal image registration, segmentation, computer-assisted surgery, diffusion MRI, etc. In 1996, he has been an active member of the Organising Committee of the 3rd IEEE International Conference on Image Processing (ICIP'96). He is also member of the organizing committees of the 2002 IEEE International Conference on Multimedia and Expo (ICME'02) held in Lausanne, Switzerland and of the 2002 IEEE International Workshop on Neural Networks for Signal Processing (NNSP'02) held in Martigny, Switzerland. Dr Thiran is reviewer for many international journals, Co-Editor-in-Chief of the Signal Processing Journal and a member of the IEEE.