Automated Melanoma Recognition

Article *in* IEEE Transactions on Medical Imaging · March 2001

DOI: 10.1109/42.918473 · Source: DBLP

6 authors, including:

Harald Ganster
Joanneum Research Forschungsgesellschaft mbH
55 PUBLICATIONS **1,475** CITATIONS
SEE PROFILE

Harald Kittler
Medical University of Vienna
333 PUBLICATIONS **18,207** CITATIONS
SEE PROFILE

Correspondence

Automated Melanoma Recognition

Harald Ganster*, Axel Pinz, Reinhard Röhrer, Ernst Wildling, Michael Binder, and Harald Kittler

***Abstract*—A system for the computerized analysis of images obtained from ELM has been developed to enhance the early recognition of malignant melanoma. As an initial step, the binary mask of the skin lesion is determined by several basic segmentation algorithms together with a fusion strategy. A set of features containing shape and radiometric features as well as local and global parameters is calculated to describe the malignancy of a lesion. Significant features are then selected from this set by application of statistical feature subset selection methods. The final kNN classification delivers a sensitivity of 87% with a specificity of 92%.**

***Index Terms*—Classification, dermatology, feature selection, fusion, segmentation.**

I. Melanoma—A Deadly Threat!

Malignant melanoma is nowadays one of the leading cancers among many white-skinned populations around the world [1]. Change of recreational behavior together with the increase in ultraviolet radiation cause a dramatic increase in the number of melanomas diagnosed. The raise in incidence was first noticed in the United States in 1930, where one person out of 100 000 per year suffered from skin cancer. This rate increased in the middle of the eighties to six per 100 000 [2] and to 13 per 100 000 in 1991. The numbers are also comparable to the incidence rates observed in Europe. In 1995, in Austria the incidence of melanoma was about 12 per 100 000, which reflected an increase of 51.8 % in the previous ten years [3], and the incidence of melanoma shows a still increasing tendency.

But on the other hand investigations have shown that the curability of skin cancer is nearly 100%, if it is recognized early enough and treated surgically [4]. Whereas the mortality rate caused by melanomas in the early sixties was about 70 %, now a survival rate of 70% [2] is achieved, which is mainly the result of early recognition.

Because of the higher incidence of malignant melanoma, researchers are concerned more and more with the automated diagnosis of skin lesions. Many publications report on isolated efforts into the direction of automated melanoma recognition by image processing. Complete integrated dermatological image analysis systems are hardly found in clinical use, or are not tested on a significant number of real-life samples.

Manuscript received February 25, 2000; revised December 2, 2000. This work was supported by the Austrian "Fonds zur Förderung der wissenschaftlichen Forschung" under Grant P11735-MED. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was M. W. Vannier. *Asterisk indicates corresponding author*.

*H. Ganster was with the Institute for Computer Graphics and Vision, Graz University of Technology, Austria. He is now with the Institute of Electrical Measurement and Measurement Signal Processing, Graz University of Technology, Kopernikusgasse 24, A-8010 Graz, Austria (e-mail: {ganster@emt.tu-graz.ac.at).

A. Pinz was with the Institute for Computer Graphics and Vision, Graz University of Technology, Austria. He is now with the Institute of Electrical Measurement and Measurement Signal Processing, Graz University of Technology, Austria.

R. Röhrer and E. Wildling were with the Institute for Computer Graphics and Vision, Graz University of Technology, Austria.

M. Binder and H. Kittler are with the Department for Dermatology, University of Vienna, Austria.

Publisher Item Identifier S 0278-0062(01)02781-1.

II. Computer Aided Image Analysis in Melanoma Research

1) Imaging Process: Epiluminescence microscopy (ELM) has proven to be an important tool in the early recognition of malignant melanoma [5], [4]. In ELM, halogen light is projected onto the object, thus rendering the surface translucent and making subsurface structures visible. A study of Schindewolf *et al.* [6] comparing direct acquisition of images by a charge-coupled device (CCD) camera [7]–[9] with digitizing color slides by means of a scanner [10], [11] concluded, that there is no significant difference in the recognition rates of malignant melanomas.

2) Segmentation: For the special problem of skin lesion segmentation, mainly region-based segmentation methods are applied, and within this category the thresholding operation is most often used [12], [7], [13]. Green *et al.* [7] derived usable segmentations in 171 of 204 lesion images (83.8% correct segmentations). Hance *et al.* [14] compare six different color segmentation algorithms (adaptive thresholding [7], [1] fuzzy-c-means [15], SCT/center split [10], PCT/median cut [10], [16], split and merge, and multiresolution segmentation). On a test set of 66 images the lowest average error could be achieved by adaptive thresholding (40 of 66 correct segmentations) and the PCT/median-cut algorithm (46 good segmentations). Combining the different methods resulted in further improvement of correctly identified tumor boundaries (57 correct segmentations).

3) Feature Calculation: In automated diagnosis of skin lesions, feature design is based on the so-called ABCD-rule of dermatoscopy [5], [17]. ABCD represent the *asymmetry*, *border* structure, variegated *color*, and *dermatoscopical* structures and define the basis for a diagnosis by a dermatologist. Pehamberger *et al.* [4] and Menzies *et al.* [18] define further eight standard ELM criteria—*pigment network*, *brown globules*, *black dots*, *radial streaming*, *pseudopods*, *overall pigmentation*, and *depigmentation*—for the diagnosis with the help of ELM. In order to represent *asymmetry* different shape features (e.g., fragmentation index [19], [7], [11], circularity factor [20], [8], asymmetry index [21], [11], bulkiness score [22]) are calculated. Some groups use the sharpness of the transition from the lesion interior to the skin [20], [8] as descriptors of the structure and irregularity of the *border*. Green *et al.* [7] and Aitken *et al.* [19] furthermore build mean value and standard deviation of these gradient values, whereas Hall *et al.* [23] calculate fractal dimensions to represent border irregularity. The descriptors of *color* are mainly statistical parameters calculated from different color channels, like average value and standard deviation of the *RGB* [19], [7], [8], [11] or *HSV* color channels [20]. Cotton and Claridge [24] use an optical model of the skin to interpret the colors occurring in a lesion. They found that all normal skin colors lie on a two–dimensional surface patch within a three-dimensional (3-D) color space (*CIE–LMS*). Atypical skin structures result in color coordinates that deviate from the normal surface patch. Parameters for the description of *dermatoscopic* structures and ELM criteria are difficult to find in literature. Franke *et al.* [25] show an example how to calculate a feature for the reticular pattern (pigment network) of a lesion.

4) Feature Selection and Classification: Green *et al.* [7] show by calculating correlation coefficients that the size of the lesion is the most

important feature in their system. Roß *et al.* [26] perform a feature selection by application of the *sequential forward selection* algorithm. They achieve a tremendous reduction to five features starting with 87 features calculated from surface profiles of skin lesions. Hintz-Madsen *et al.* [27] achieve feature selection with a neural network by application of *node pruning*. They start with a fully connected feed-forward network with 21 input units that represent 21 skin lesion features (color and texture parameters), one hidden layer with four hidden units, and one output unit for the categorization into benign or malignant class. After training and application of node pruning the hidden layer could be reduced to three units with a small amount of connections to the remaining six input units.

The actual classification is the final step in automated analysis of skin lesions. In some of the skin lesion applications mentioned in previous sections, some classical statistical methods like *nearest neighbor* classification [26], [12] or *discriminant analysis* [7], [1], [9], as well as neural networks [28]–[31], [11], [32] have been applied for the purpose of skin lesion classification.

III. Data Description

The skin lesion data used in this project has been provided by the *Department of Dermatology* at the *Vienna General Hospital*. In this department, the imaging is performed by a hand-held CCD camera (one chip color sensor) that is combined with an epiluminescence microscope in order to produce digitized ELM images of skin lesions. The images have a spatial resolution of 632×387 pixels with 1 pixel $\approx 22\ \mu\text{m}^2$. They are available as true color images in the *RGB* color system with a radiometric depth of eight bits per color channel. No normalization of color during the imaging process is performed.

Overall 5393 skin lesion images categorized into three classes have been put at our disposal. All images were captured during routine clinical examinations to reflect the *a priori* probabilities of the routine diagnosis in a specialized dermatology clinic. 4270 of these lesions are classified as clearly benign lesions (referred as *class 1* throughout this project). In order not to cause any needless distress to the patient, the lesion is not surgically excised, if the diagnosis (benign) is achieved by a consensus of three experienced dermatologists. The category of dysplastic lesions (*class 2*) is represented in the dataset by 1011 lesion samples. This category still defines a benign class, but these kind of lesions are so-called precursors to malignant melanoma [33]–[35]. Since this category represents skin lesions with an increased risk to turn into a melanoma, the category receives an own class label in the clinical diagnosis. 108 lesions are categorized as malignant melanomas (*class 3*). Four lesions do not have any clinical diagnosis in the database due to their small size. The lesions of the classes 2 and 3 were all surgically excised and the ground truth was generated by means of histopathological diagnosis.

IV. Segmentation

The first aim of this project was to build an efficient robust automatic segmentation tool for skin lesion images (Fig. 1). It can be noticed that the lesions have large variations in size as well as in color and contrast to the surrounding skin. In Fig. 1(d), hairs covering the skin lesion can be noticed, which would have disturbing influence on the derivation of the correct boundaries. Since hairs appear as thin dark elongated structures their influence is diminished by applying the grey-scale morphologic operation of closing. In order not to loose any important structures within the lesion, grey-scale morphology is just used to derive the segmentation. In further steps of feature calculation, the information is taken again from the original lesion image, but now with better defined lesion boundaries.

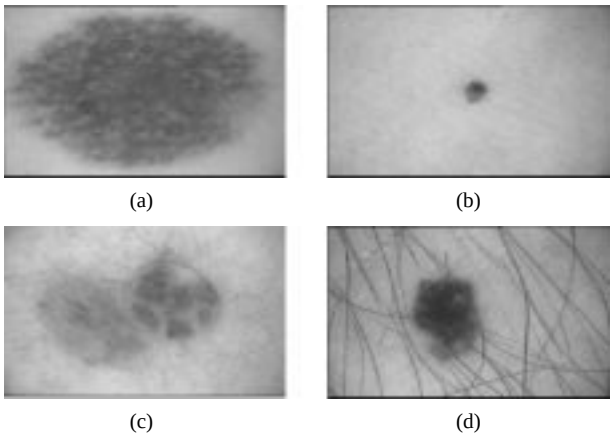


Fig. 1. Different possible appearances of skin lesions.

A disadvantage of many segmentation algorithms is the need for user interaction to tune some parameters. Tests on several sample images [36] have shown that it is possible to derive a small amount of segmentation techniques with appropriate parameter settings that allow the segmentation of each individual skin lesion image. The final segmentation is then derived in a *fusion* process (cf. Section IV-A) [37], [38].

In the skin lesion segmentation system, basically three different algorithms are used to segment a lesion, namely *global thresholding*, *dynamic thresholding*, and a *3-D color clustering concept* [39]. Best results were achieved with the following color channels: thresholding with the *blue channel of the RGB color model* and *b channel of the CIE–Lab color model*, 3-D color clustering with the *X, Y, and Z channels of the CIE–XYZ color model*. After morphologic processing of the segmentation results, in order to remove tiny artifacts, the mask of the lesion is derived as the largest region within the image.

A. Fusion of Segmentations

Experiments with 40 selected lesions showed, that the dynamic thresholding algorithms achieve a good segmentation result in about 80% of the images. The 3-D-clustering approach has in general a bad performance (about 8%), but turns out to deliver good results in cases, where the other approaches deliver wrong segmentations. Thus, we applied the following rules to perform the fusion of the individual segmentation results:

1) segmentation results with the lesion mask growing into the image border are rejected;
2) segmentation results without any detected region are rejected;
3) segmentation results comprising fragments at the image borders are rejected;
4) the segmentation result with the smallest mask is rejected;
5) segmentation results, whose mask areas differ too much to the other segmentation results are rejected;
6) the remaining segmentation results are combined to yield a common segmentation.

Since the database of lesions is built without considering lesions that are larger than the field of view, rule 1 removes wrong segmentation results, that do not fit into the image window [e.g., Fig. 2(d)]. Due to bad lighting conditions or camera placement in the imaging process artifacts near the image border may appear, which are eliminated by rule 3. We also had to reject the smallest detected mask (rule 4), since in some situations not the desired lesion, but another dark spot beside the lesion area was detected.

After rejection of the outliers, several different operators can be applied to the remaining segmentations to build the final segmentation. Experiments [36] showed that best results are achieved by a simple

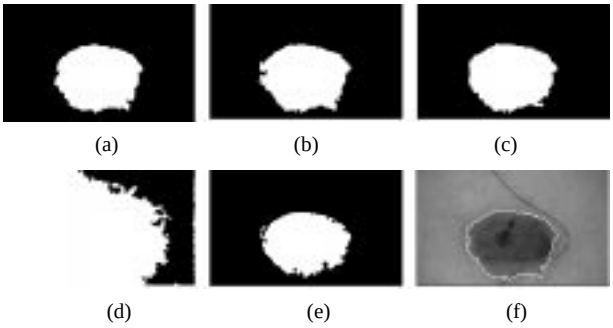


Fig. 2. Example for the fusion of five initial segmentation (a)–(e) results of the initial segmentation algorithms, (f) original lesion image with superimposed lesion boundary after fusing (a)–(e); image (d) was rejected by rule 1 and image (e) was rejected by rule 4; all other images are used in the fusion process.

OR-combination (Fig. 2). Majority voting caused some problems, due to the fact that the actual number of remaining initial segmentations was not known in advance.

B. Results and Discussion

Verification of the segmentation results (Fig. 3) was done in cooperation with trained dermatologists at the Vienna General Hospital.

The derived lesion borders were compared to lesion borders marked by a specialized dermatologist and the differences were evaluated in three steps. First a set of 150 images was selected and lesion borders were marked by hand by dermatologists. Several features (e.g., area, perimeter, shape descriptors) were calculated for both sets of lesion masks (hand drawn by a dermatologist and automated extraction by the system) and statistically evaluated. Comparison of the lesion areas gave an average absolute pixel difference of 4183 pixel with a standard deviation of 6771 pixel. The average lesion size was 53 217 pixel in the manual segmentation. In the second step of the evaluation, the masks of a set of 400 lesion images were visually inspected by four experienced dermatologists and the segmentation was said to be correct if the majority of the dermatologists agreed on the derived lesion boundaries. The rest (3450) of the lesion images was visually evaluated by only one dermatologist.

From a set of 4000 ELM images of skin lesions, the segmentation algorithm in 19 cases was not able to detect any lesion, which amounts to less than 0.48%. The fact that segmentation was not successful was noticed by the algorithm and reported back to the user. It is hard even for the experienced human eye to detect lesion borders in the given example [Fig. 4(a)] and it is not possible to define them exactly, since the colors within the lesion do not or just slightly differ from the colors in the surrounding skin area. In 140 cases (3.5%), the segmentation of the lesions was not satisfactory, since the mask did not represent the complete lesion and, therefore, some important areas for the feature calculation process were not covered [Fig. 4(b)].

Overall there are 159 lesions that have to be rejected due to segmentation failures within the set of 4000 lesion images. The achieved performance, therefore, is about 96% of correctly segmented skin lesion images.

V. Feature Calculation

Overall 122 different parameters are calculated, which try to reflect the parameters used in medical diagnosis. There are on one hand global features calculated for the representation of the entire lesion. On the other hand also local descriptors are calculated in order to allow for recognizing differences in several sectors of the lesion.

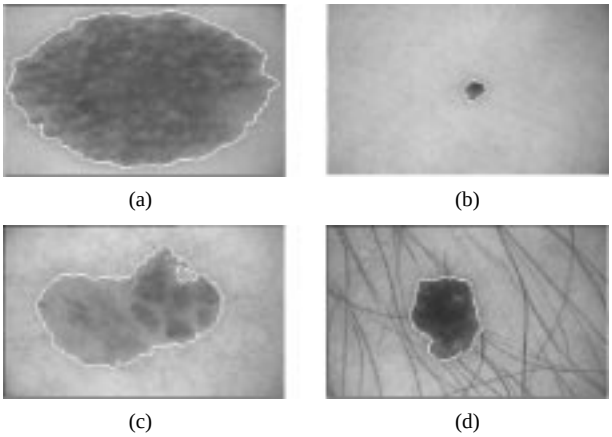


Fig. 3. Segmentation results of the images in Fig. 1 achieved with suggested segmentation technique.

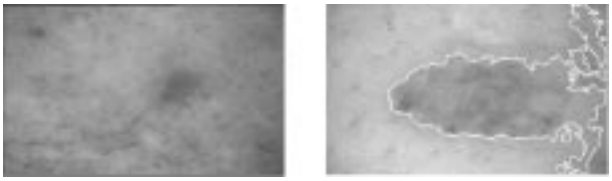


Fig. 4. Examples of wrong segmentations; (a) The segmentation failed, but was detected by the system and reported back to the user. (b) The segmentation delivered a mask that could not represent the original lesion.

The calculated **global features** cover a wide range of categories, starting with simple *size* (area, perimeter, polar measures, and bounding rectangle) and *shape* descriptors [40] (formfactor, roundness, compactness, extent). Further shape descriptors are derived from the bounding rectangle: aspect ratio of length and width as well as the distance between the center of the bounding rectangle and the center of mass of the lesion.

In order to calculate simple *color features*, the original images are transformed from *RGB* color space into *HSI* space [41]. The features derived are the *minimum*, *maximum*, *average*, and the *variance* of the *intensity (I)* and *hue (H)* channels. Due to the lighting conditions used for the digital ELM camera *saturation (S)* does not give any useful feature values.

Since the color of a skin lesion is dependent on the skin type it is desirable to normalize the color channels. The normalization of the intensity channel is achieved by subtracting the average intensity in the skin area (i.e., outside the binary mask). The hue channel is normalized by subtracting the hue value with highest probability in the skin area. From these normalized color channels, again *minimum*, *maximum*, *average*, and the *variance* are taken as features.

Certain colors are significant features in the examination of a skin lesion by a dermatologist. The lesion images are available in *RGB* format (24-bit colors), which results in an overall amount of 16.7 million colors. Typical examples of lesions which show reddish, bluish, grey and black areas and spots are selected and by application of the *median cut color quantization* algorithm of Heckbert [16] 15 significant colors are derived. The parameters used in the automated diagnosis system are the number of different colors appearing within the lesion, as well as the percentage of each of the 15 colors (ratio between the overall area covered by this color and the entire area of the lesion).

In order to reflect the significance of the transition area from lesion to the skin, a six-pixel-wide mask was generated by morphological operations. Gradient features consist of the *minimum*, *maximum*, *average*, and *variance* of gradient values calculated by application of the Sobel

operators within the extracted mask for the border area. Further border features are the *minimum*, *maximum*, *average*, and *variance* of the normalized intensity image in the border region.

The remaining features are calculated on **local areas** of the lesion. The lesion, therefore, is separated along major and minor axis of the lesion mask into left and right half, into upper and lower half, as well as into four quadrants. The features calculated for these local segments can be categorized into three groups. The first is the category of shape and size features directly calculated from the segment mask (perimeter, area, and the formfactor). The next group consists of the quantized color features derived for the segment (percentage of individual colors and number of colors).

In the diagnosis of melanomas, the analysis of symmetry is a very important fact, since the local appearance of certain structures or colors is most indicative for a melanoma. In order to recognize differences in certain regions of the lesion, ratios R_i between different segments are calculated for selected features (perimeter, area, and formfactor) according to following equation:

$$R_i = \frac{Q_i}{\sum_{j \neq i} Q_j} \qquad i, j = 1, \ldots, 4 \tag{1}$$

where Q_i represents the feature value of quadrant i.

Further local properties of a lesion can be detected by splitting the mask into a central and a peripheral part (Fig. 5). Only the parameters concerning the quantized colors, *color percentage* and *number of colors* are calculated.

VI. Feature Subset Selection

The images used in this project are captured during routine diagnosis in the department of dermatology (cf. Section III). Thus, the dataset reflects the *a priori* probabilities of the routine diagnosis in a hospital (note: this is not the distribution of lesions observed among the population). In order to achieve an equal cardinality of each class, from each of the three categories (cf. Section III) 90 samples are selected randomly, which results in a training set of 270 lesions for the feature selection and classification. Due to the fact that only a small set of 270 lesions (out of 5363 lesions) is selected, special strategies from statistical pattern recognition, namely *cross-validation* (XVAL, [42]) and *leave one out method* (LOO, [43, p. 219 ff]) are applied to best exploit the available data by still keeping independent data for the performance evaluation.

All the features presented in Section V have different physical units (e.g., *pixel*, *pixel*^{2}) or are without any unit (e.g., feature relations between sectors) and, thus, have completely different value ranges. Due to the lack of knowledge about individual features, an objective scaling of the features is achieved by calculating z-scores [44]

$$z_{ij} = \frac{x_{ij} - \overline{m}_j}{s_j} \tag{2}$$

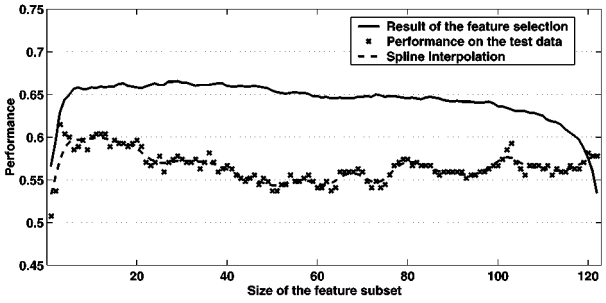
where x_{ij} represents the ith sample measure of feature j and $\overline{m}_j$ the mean value of all n samples for feature j. s_j is the *mean absolute deviation* [44]

$$s_j = \frac{1}{n} \sum_{i=1}^{n} |x_{ij} - \overline{m}_j|. \tag{3}$$

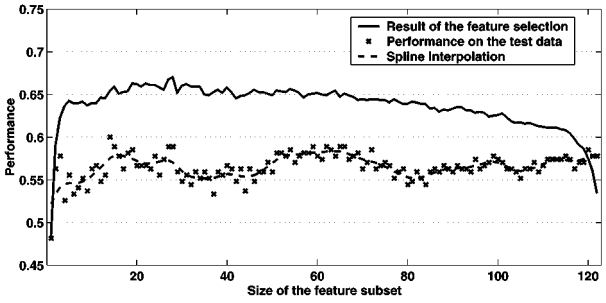
The z-scores have the statistical properties $\overline{m}_j = 0$ and $s_j = 1$, they are robust against outliers, and they are objectively derived without any knowledge about individual features.

(a) (b)

Fig. 5. Splitting of a mask into: (a) periphery and (b) center. The cut was at 65% of the lesion radius.



(a)



(b)

Fig. 6. Result of feature subset selection: (a) SFFS, (b) SBFS. The separability criterion is the LOO performance of a 24-NN classifier. Selection performances and classification results on test data are calculated by averaging five XVAL runs.

A. Feature Subset Selection with Different Algorithms

Inspection of several features as well as feature combinations showed that the data does not yield clusters for the three classes in the feature space. Thus, the performance of a *nonparametric classifier* (k-nearest-neighbor) is chosen as the separability criterion for the feature subset selection. In exhaustive experiments, the high number of 24 neighbors delivered the most reliable results for several different selection strategies. In the classification experiments presented in this section the dataset of 270 lesions is split into five disjunct subsets. The selection performance is evaluated by fivefold cross validation (XVAL) in five individual runs with 4/5 of the initial data serving as training set for the selection algorithm. During the training phase in each XVAL run, i.e., the actual feature subset selection, performances are estimated by application of the leave-one-out method (LOO). The remaining subset serves as test set to estimate classification performances for the selected feature subsets on independent data. All calculated performances show the fraction of the number of correctly classified lesions over the number of all lesions.

Due to the high dimensionality of the feature space (122 features) it is not possible to derive the optimal performance by exhaustive search. The algorithms applied for feature subset selection in the experiments are *Sequential forward floating selection* (SFFS) and *Sequential backward floating selection* (SBFS) [45]. In Fig. 6, it can be noticed that

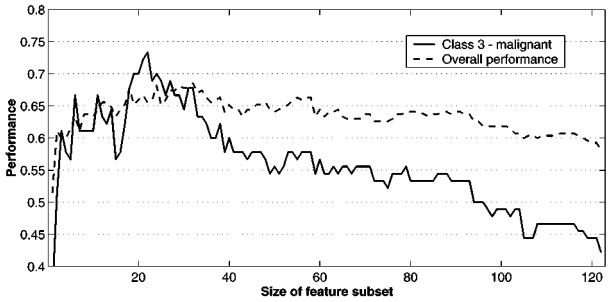


Fig. 7. Result of sequential backward floating selection: The separability criterion is the LOO performance of a 24-NN classifier. LOO performances are calculated on the entire training set of 270 lesions.

in both selection methods the classification performance of the selected feature subset reaches 65% or is slightly above. The performance achieved on the independent test data is about 5–10% less than the achieved selection performance. The performance ascent for small subsets can be detected in the selection performance as well as the performance on the test data, but for subsets with higher cardinality a large variability in the classification performances is noticed, which is due to the *over-fitting* of the classifier to the training data.

B. Size of Feature Subset

The best selection performances achieved with the SFFS algorithm start at a subset size of 10, whereas the SBFS algorithm delivers a good selection performance beginning at subset sizes of 15. The classification performance achieved on independent test data decreases for subsets with more than 20 features selected by the SFFS, and the best classification performances for small subsets selected by SBFS are in the range from 15 to 25 features. By only inspecting the overall achieved performances (Fig. 6) one even could imagine using just five to ten features to get acceptable classification results. In order to be able to inspect performances of individual classes, another feature subset selection is applied to the training set. But this time the selection is performed with a leave-one-out performance on the entire training set of 270 samples (Fig. 7). By inspecting individual sensitivities on the malignant class of several subset sizes it turns out that an acceptable performance is only achieved with subsets of more than 20 features. Since in this project feature subsets are preferred that classify the *category of melanomas* correctly with a high degree while still keeping a good overall classification performance, a subset size of 21 is chosen for further classification experiments.

VII. Classification Experiments

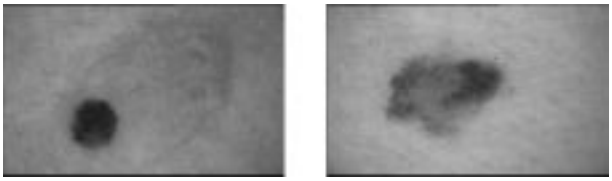
Overall feature vectors of 5363 lesions, after removal of the images that did not deliver a correct segmentation (30), are available for the classification experiments. The set of 5363 lesions is composed of the three categories benign (4257 lesions), dysplastic (1010 lesions), and malignant (96 lesions). Results are calculated for two different classification approaches: a) a classification into all three categories (benign, dysplastic, and malignant) is performed, b) classifications of a new test set into two categories are performed. All classification experiments are performed with a 24-NN classifier based on the derived 21 features.

A. Classification into Three Categories

The training set for the classifier is the set of 270 lesions already used in the feature selection. The test set of this first classification experiment is the entire database of 5363 lesions in three categories. The confusion matrix for this classification is displayed in Table I. The main aim is to achieve as good as possible classification for the category of

TABLE I
Confusion Matrix for the Classification Into Three Categories

Assigned / True	1	2	3	
1 (benign)	2500	1347	410	59%
2 (dysplastic)	324	531	155	53%
3 (malignant)	14	12	70	73%
Overall				61%



(a) (b)

Fig. 8. Two examples for the false classification of malignant lesions.

TABLE II
Confusion Matrix for the Clinical Diagnosis

Assigned / True	1	2	3	Other	
1 (benign)	4161	94	9	6	97%
2 (dysplastic)	42	960	8	1	95%
3 (malignant)	6	19	78	5	72%
Overall					88%

malignant lesions, since the false classification or the late treatment of a melanoma has severe, often deadly, consequences for the patient. The result of 73% correct classifications for malignant lesions also reflects this desired high performance, although the overall achieved performance with this training and feature set yields only a rather low classification rate of 61%. The calculation of the overall achieved performance as an average performance is preferred to calculating the diagnostic correctness, i.e., building the fraction of the number of all correctly classified lesions over the number of all lesions, since the available lesions are not equally distributed over all classes. In case of using the diagnostic correctness as estimator for the overall performance, it would be easy to tune the classification parameters to achieve a much higher overall performance with the curse of loosing performance on the malignant class.

If we take a closer look at the classification result (Table I), it can be noticed that most of the false classifications are concerned with the dysplastic class. But for an automated diagnosis system these false classifications are not a real problem, since these kind of lesions can be detected in later routine examinations, because patients with dysplastic nevi are under permanent supervision of a dermatologist. The severe problem in this classification is the clearly benign categorization of 14 melanomas (Fig. 8). The lesion in Fig. 8(a) shows a very large area on the right side that has almost same coloration as the background skin. Thus, the segmentation algorithm only delivered a binary mask for the darker part, which is very regular in shape and coloration and, thus, is interpreted as a benign nevus. Even dermatologists had problems in the categorization of this lesion, since the clinical diagnosis was a *dermal nevus* (benign lesion type) and the malignancy of the lesion was revealed in histopathological diagnosis. Fig. 8(b) depicts a melanoma with very light coloration. Although it is rather asymmetric it does not show significant growth areas. The clinical diagnosis for the lesion was a *dysplastic nevus*. Thus, the dermatologists considered it to be a suspicious, but not malignant lesion type, and again the malignancy was detected in histopathology. If we compare the achieved

classification result of the system to results obtained in clinical diagnosis (Table II), it can be observed that the sensitivity on melanomas is in the same range for the automated classification (73%) and the clinical diagnosis (72%).

B. Classification into Two Categories, New Test Set

Since the previous experiment used a training and test set, that have some lesions in common, the Department of Dermatology at the Vienna General Hospital provided 200 more completely unknown lesion images, which are not within the presented data (cf. Section III), thus serving as completely independent test data. The training set used for this classification experiment is the already described training set for the classification approach with three classes.

Classification performances were calculated for two different strategies. In the first, more conservative, approach, the lesions of class 2 (atypical and dysplastic nevi) are combined with the malignant lesions to yield a common class representing the need for intervention. With this training set and classification strategy a sensitivity of 87% and a specificity of 92% could be achieved. In the second strategy, the dysplastic nevi are included into the benign class to form a new benign category. Thus, only the strictly malignant lesions are separated from the benign lesion categories. The classification using the same training set as in the previous strategy delivered a sensitivity of 77% with a specificity of 84%.

VIII. Summary and Conclusion

The result of more than 96% correctly segmented lesion images (in a set of 4000 skin lesions) reflects a very reliable segmentation module for the special task of skin lesion segmentation. The impressive segmentation performance is achieved by fusion of several simple segmentation algorithms (thresholding, color clustering). The fusion concept allows for the further extension of the segmentation module by integration of other segmentation methods (texture analysis, other color segmentation algorithms).

The features used in the integrated system are designed to resemble the clinical ABCD and ELM criteria, but are adjusted to a quick and easy automated calculation. They mainly focus on size and shape descriptors, as well as on color and local parameters. Some parameters are included to allow the description of the transition area between lesion and skin.

By application of feature selection strategies the number of features relevant for a good classification is reduced from 122 to 21 parameters. It is shown that this optimized set covers most of the discriminative information of the entire feature set.

Due to the nature of the data (no clusters in feature space, several subtypes for each class) the best classification performance is achieved by a nonparametric 24-NN classifier. The overall performance of 61% achieved by classification into three classes and a performance of 73% for the class of melanomas in this experiment is comparable to rates observed in clinical routine and depicts a rather good performance for an automated system, since the data (overall 5393 skin lesions) was gathered in daily clinical routine and was not especially selected for this project. In a classification approach with two categories (malignant and dysplastic lesions against benign lesions), a sensitivity of 87% for the intervention class and a specificity of 92% is observed.

In order to achieve better classification results, new parameters have to be derived and included in the feature set. Good parameters describing the lesion boundary as well as texture descriptors are not yet included in the feature set, and might yield a good starting point to improve the discriminative information in the feature set.

The system presented here is able to support the clinical diagnosis, especially in large scale screening applications, and currently undergoes comprehensive tests at the *Department of Dermatology* at the *Vienna General Hospital*.

Acknowledgment

The authors would like to thank E. Claridge for valuable discussions about melanoma recognition.

References

[1] A. Green, N. Martin, G. McKenzie, J. Pfitzner, F. Quintarelli, B. W. Thomas, M. O'Rourke, and N. Knight, "Computer image analysis of pigmented skin lesions," *Melanoma Res.*, vol. 1, pp. 231–236, 1991.

[2] K. Wolff and H. Pehamberger, "Malignes Melanom: Füherkennung und Prognose," *Wiener klinische Wochenschrift*, vol. 97, no. 10, pp. 451–455, Mai 1985.

[3] "Arbeitsgemeinschaft für chirurgische Onkologie der Österreichischen Gesellschaft für Chirurgie ACO," *Karzinome — Fakten und Statistik*, ACO-Bulletin, 1995.

[4] H. Pehamberger, A. Steiner, and K. Wolff, "*In vivo* epiluminescence microscopy of pigmented skin lesions. I. Pattern analysis of pigmented skin lesions," *J. Amer. Acad. Dermatol.*, vol. 17, no. 4, pp. 571–583, Oct. 1987.

[5] F. Nachbar, W. Stolz, T. Merkle, A. B. Cognetta, T. Vogt, M. Landthaler, P. Bilek, O. Braun-Falco, and G. Plewig, "The ABCD rule of dermatoscopy: High prospective value in the diagnosis of doubtful melanocytic skin lesions," *J. Amer. Acad. Dermatol.*, vol. 30, no. 4, pp. 551–559, Apr. 1994.

[6] T. Schindewolf, R. Schiffner, W. Stoltz, R. Albert, W. Abmayr, and H. Harms, "Evaluation of different image acquisition techniques for a computer vision system in the diagnosis of malignant melanoma," *J. Amer. Acad. Dermatol.*, vol. 31, no. 1, pp. 33–41, 1994.

[7] A. Green, N. Martin, J. Pfitzner, M. O'Rourke, and N. Knight, "Computer image analysis in the diagnosis of melanoma," *J. Amer. Acad. Dermatol.*, vol. 31, no. 6, pp. 958–964, Dec. 1994.

[8] S. Seidenari, M. Burroni, G. Dell'Eva, P. Pepe, and B. Belletti, "Computerized evaluation of pigmented skin lesion images recorded by a videomicroscope: Comparison between polarizing mode observation and oil/slide mode observation," *Skin Res. Technol.*, vol. 1, pp. 187–191, 1995.

[9] R. Pompl, W. Bunk, A. Horsch, W. Abmayr, G. Morfill, W. Brauer, and W. Stolz, "Computer vision of melanocytic lesions using MELDOQ," in *Proc. 6th Congress Int. Soc. Skin Imaging, London; Skin Research and Technology*, vol. 5, 1999, p. 150.

[10] S. E. Umbaugh, R. H. Moss, W. V. Stoecker, and G. A. Hance, "Automatic color segmentation algorithms: With application to skin tumor feature identification," *IEEE Eng. Med. Biol. Mag.*, vol. 12, no. 3, pp. 75–82, Sept. 1993.

[11] H.-C. Lee, "Skin cancer diagnosis using hierarchical neural networks and fuzzy logic," M.S. thesis, Univ. Missouri, Rolla, MO, 1994.

[12] H. Ganster, M. Gelautz, A. Pinz, M. Binder, H. Pehamberger, M. Bammer, and J. Krocza, "Initial results of automated melanoma recognition," in *Theory and Applications of Image Analysis II, Selected papers of the 9th SCIA, Scandinavian Conference on Image Analysis*, G. Borgefors, Ed. Singapore: World Scientific, 1995, pp. 343–354.

[13] L. Xu, M. Jackowski, A. Goshtasby, D. Roseman, S. Bines, C. Yu, A. Dhawan, and A. Huntley, "Segmentation of skin cancer images," *Image Vis. Computing*, vol. 17, pp. 65–74, 1999.

[14] G. A. Hance, S. E. Umbaugh, R. H. Moss, and W. V. Stoecker, "Unsupervised color image segmentation," *IEEE Eng. Med. Biol. Mag.*, vol. 15, no. 1, pp. 104–111, Jan./Feb. 1996.

[15] Y. W. Lim and S. U. Lee, "On the color image segmentation algorithm based on the thresholding and the fuzzy c-means techniques," *Pattern Recogn.*, vol. 23, no. 9, pp. 935–952, 1990.

[16] P. Heckbert, "Color image quantization for frame buffer display," *Comput. Graph. (Proc. SIGGRAPH '82)*, vol. 16, no. 3, pp. 297–307, July 1982.

[17] R. J. Friedman and D. S. Riegel, "The clinical features of malignant melanoma," *Dermatologic Clin.*, vol. 3, pp. 271–283, 1985.

[18] S. W. Menzies, K. A. Crotty, C. Ingvar, and W. H. McCarthy, *An Atlas of Surface Microscopy of Pigmented Skin Tumors*. New York: McGraw-Hill, 1995.

[19] J. F. Aitken, J. Pfitzner, D. Battistutta, P. K. O'Rourke, A. C. Green, and N. G. Martin, "Reliability of computer image analysis of pigmented skin lesions of Australian adolescents," *Cancer*, vol. 78, no. 2, pp. 252–257, July 1996.

[20] N. Cascinelli, M. Ferrario, R. Bufalino, S. Zurrida, V. Galimberti, L. Mascheroni, C. Bartoli, and C. Clemente, "Results obtained by using a computerized image analysis system designed as an aid to diagnosis of cutaneous melanoma," *Melanoma Res.*, vol. 2, pp. 163–170, 1992.
[21] W. V. Stoecker, W. W. Li, and R. H. Moss, "Automatic detection of asymmetry in skin tumors," *Computerized Med. Imag. Graph.*, vol. 16, no. 3, pp. 191–197, May, June 1992.
[22] E. Claridge, P. N. Hall, M. Keefe, and J. P. Allen, "Shape analysis for classification of malignant melanoma," *J. Biomed. Eng.*, vol. 14, no. 3, pp. 229–234, 1992.
[23] P. N. Hall, E. Claridge, and J. D. M. Smith, "Computer screening for early detection of melanoma — is there a future?," *Br. J. Dermatol.*, vol. 132, pp. 325–338, 1995.
[24] S. D. Cotton and E. Claridge, "Developing a predictive model of human skin coloring," in *Proc. SPIE Medical Imaging 1996: Physics of Medical Imaging*, vol. 2708, 1996, pp. 814–825.
[25] K. H. Franke, F. Glassmann, and R. Gergs, "Früherkennung von Hautkrebs durch Farbbildverarbeitung," *Mustererkennung 1993*, pp. 407–414, 1993.
[26] T. Ross, H. Handels, J. Kreusch, H. Busche, H. H. Wolf, and S. J. Pöppl, "Automatic classification of skin tumours with high resolution surface profiles," in *Computer Analysis of Images and Patterns, CAIP 95 Proceedings*, V. Hlavac, Ed. Berlin, Germany: Springer-Verlag, 1995, vol. LNCS 970, pp. 368–375.
[27] M. Hintz-Madsen, L. K. Hansen, J. Larsen, E. Olesen, and K. T. Drzewiecki, "Design and evaluation of neural classifiers—application to skin lesion classification," in *Proc. 1995 IEEE Workshop Neural Networks for Signal Processing (NNSP'95)*, 1995, pp. 484–493.
[28] M. Binder, A. Steiner, M. Schwarz, S. Knollmayer, K. Wolff, and H. Pehamberger, "Application of an artificial neural network in epiluminescence microscopy pattern analysis of pigmented skin lesions: A pilot study," *Br. J. Dermatol.*, vol. 130, pp. 460–465, 1994.
[29] F. Ercal, A. Chawla, W. V. Stoecker, H. C. Lee, and R. H. Moss, "Neural network diagnosis of malignant melanoma from color images," *IEEE Trans. Biomed. Eng.*, vol. 41, pp. 837–845, Sept. 1994.
[30] F. Ercal, H. C. Lee, W. V. Stoecker, and R. H. Moss, "Skin cancer classification using hierarchical neural networks and fuzzy systems," *Int. J. Smart Eng. Syst. Design*, vol. 1, pp. 273–289, 1999.
[31] Z. Zhao, "Automatic Skin Tumor Segmentation and Diagnosis," Ph.D. dissertation, Univ. Missouri, Rolla, MO, 1997.
[32] M. Hintz-Madsen, L. Kai Hansen, J. Larsen, E. Olesen, and K. T. Drzewiecki, "Detection of malignant melanoma using neural classifiers," in *Solving Engineering Problems with Neural Networks, Proc. Int. Conf. Engineerings Applications on Neural Networks EANN'96*, June 1996, pp. 395–398.
[33] H. Pehamberger, "Melanompräkursoren—Risikonaevi," *Wiener klinische Wochenschrift*, vol. 99, no. 13, pp. 441–445, June 1987.
[34] D. E. Elder, M. H. Green, D. P. Guerry IV, K. H. Kraemer, and W. H. Clark, "The dysplastic nevus syndrome—Our definition," *Amer. J. Dermatol.*, vol. 4, pp. 455–460, 1982.
[35] A. S. Sober, A. R. Rhodes, C. L. Day Jr., T. B. Fitzpatrick, and M. C. Mihm Jr., "Primary melanoma of the skin. Recognition of precursor lesions and estimation of prognosis in stage I," in *Update: Dermatology in General Medicine*, T. B. Fitzpatrick, A. Z. Eisen, K. Wolff, I. M. Freedberg, and K. F. Austen, Eds. New York: McGraw Hill, 1983, pp. 98–112.
[36] E. Wildling, "Automatische Segmentierung von Pigmentläsionen sowie low-level Routinen zur Merkmalsextraktion," masters thesis, Inst. Comput. Graph. Vis., Tech. Univ. Graz, Graz, Austria, Jan. 1998.
[37] M. A. Abidi and R. C. Gonzalez, Eds., *Data Fusion in Robotics and Machine Intelligence*. New York: Academic, 1992.
[38] A. Pinz, M. Prantl, H. Ganster, and H. K. Borotschnig, "Active fusion—A new method applied to remote sensing image interpretation," *Pattern Recogn. Lett.*, pt. Special issue on "Soft Computing in Remote Sensing Data Analysis", vol. 17, no. 13, pp. 1349–1359, 1996.
[39] W. Woelker, "Image segmentation based on adaptive 3-D-analysis of the CIE-l*a*b color space," in *Proceedings of SPIE (Visual Communications and Image Processing)*. Bellingham, WA: Int. Soc. Opt. Eng., 1996, vol. 2727, pp. 1197–1203.
[40] J. C. Russ, *The Image Processing Handbook*, 2nd ed. Boca Raton, FL: CRC, 1995.
[41] J. D Foley, A. van Dam, S. K. Feiner, and J. F. Hughes, *Computer Graphics: Principles and Practice*, 2nd ed. Reading, MA, USA: Addison-Wesley, 1990.
[42] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *J. Roy. Statist. Soc. B*, vol. 36, no. 1, pp. 111–147, 1974.
[43] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. San Diego, CA: Academic, 1990.
[44] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: Wiley., Mar. 1990.
[45] P. Pudil, J. Novovicová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recogn. Lett.*, vol. 15, pp. 1119–1125, Nov. 1994.

Comparison of Low-Contrast Detail Perception on Storage Phosphor Radiographs and Digital Flat Panel Detector Images

Siegfried Peer*, Ulrich Neitzel, Salvatore M. Giacomuzzi, Regina Peer, Eva Gassner, Iris Steingruber, and Werner Jaschke

***Abstract*—A contrast detail analysis was performed to compare perception of low-contrast details on X-ray images derived from digital storage phosphor radiography and from a flat panel detector system based on a cesium iodide/amorphous silicon matrix.**

The CDRAD 2.0 phantom was used to perform a comparative contrast detail analysis of a clinical storage phosphor radiography system and an indirect type digital flat panel detector unit. Images were acquired at exposure levels comparable to film speeds of 50/100/200/400 and 800. Four observers evaluated a total of 50 films with respect to the threshold contrast for each detail size. The *numbers of correctly identified objects* were determined for all image subsets.

The overall results show that low-contrast detail perception with digital flat panel detector images is better than with state of the art storage phosphor screens. This is especially true for the low-exposure setting, where a nearly 10% higher correct observation ratio is reached. Given its high detective quantum efficiency the digital flat panel technology based on the cesium iodide scintillator/amorphous silicon matrix is best suited for detection of low-contrast detail structures, which shows its high potential for clinical imaging.

***Index Terms*—Digital radiography, low-contrast perception, observer performance.**

I. Introduction

In the last years, there was a rapid evolution of digital radiography: Digital storage phosphor plate computed radiography (CR) systems and dedicated chest systems based on solid-state selenium detectors are gradually replacing film screen radiography in clinical routine as both systems proved to produce excellent image quality [1], [2]. As a next step in the development of digital radiography systems large area electronic flat panel detectors based on either scintillators or photoconductors as X-ray conversion material [3]–[6] have been introduced into

Manuscript received August 27, 2000; revised November 21, 2000. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was M. W. Vannier. *Asterisk indicates corresponding author.*

*S. Peer is with the Department of Radiology, University Hospital, Anichstr.35, A-6020 Innsbruck, Austria. (e-mail: siegfried.peer@uibk.ac.at).

U. Neitzel is with Philips Medical Systems, Hamburg, Germany.

S. M. Giacomuzzi, R. Peer, E. Gassner, I Steingruber, and W. Jaschke are with the Department of Radiology, University Hospital, A-6020 Innsbruck, Austria.

Publisher Item Identifier S 0278-0062(01)02782-3.