

Computational vision systems for the detection of malignant melanoma

ILIAS MAGLOGIANNIS¹ and DIMITRIOS I. KOSMOPOULOS²

¹Department of Information and Communication Systems Engineering, University of Aegean, 83200 Karlovassi;

²National Centre for Scientific Research 'Demokritos,' Institute of Informatics and Telecommunications, 15310 Aghia Paraskevi, Greece

Received September 6, 2005; Accepted October 10, 2005

Abstract. In recent years, computational vision-based diagnostic systems for dermatology have demonstrated significant progress. We review these systems by first presenting the installation, visual features utilized for skin lesion classification and the methods for defining them. We also describe how to extract these features through digital image processing methods, i.e. segmentation, registration, border detection, color and texture processing, and present how to use the extracted features for skin lesion classification by employing artificial intelligence methods, i.e. discriminant analysis, neural networks, and support vector machines. Finally, we compare these techniques in discriminating malignant melanoma tumors versus dysplastic naevi lesions.

Introduction

Malignant melanoma is among the most frequent types of skin cancer and one of the most malignant tumors. Its incidence has increased faster than that of almost all other cancers and the annual incidence rates have increased at the rate of 3-7% in fair-skinned populations in recent decades (1). Advanced cutaneous melanoma is still incurable, but melanoma diagnosed in early stages can be cured without complications. However, the differentiation of early melanoma from other pigmented skin lesions is not trivial even for experienced dermatologists. The issue has attracted the interest of many researchers who have developed systems for the automated detection of malignant melanoma in skin lesions, which will be surveyed here.

The main design issues for a machine vision system for melanoma detection concern the image acquisition setup,

image processing and classification methodology. More specifically, the following questions need to be addressed:

1. How can we acquire good quality images?
2. How are the image features defined? i.e. What are we looking for?
3. How are these features detected in the image? (usually trivial for humans, but non-trivial for machines).
4. How many of the defined features should be used for optimal results? (feature selection).
5. Which classifiers are used and how is the 'importance' of each feature determined in classification?
6. How can we assess the performance of a classifier?

Materials and methods

Image acquisition. The first step in machine vision-based expert systems involves the acquisition of the tissue digital image, which answers question 1. The main techniques used for this purpose are epiluminescence microscopy (ELM or dermoscopy) and image acquisition using still or video cameras. By placing a thin layer of oil on a lesion, then pressing a special hand-held microscope against the oil field on the skin of the patient, ELM provides a more detailed inspection of the surface of pigmented skin lesions and renders the epidermis translucent, making many features visible. New techniques have been presented that use multispectral images in which the chosen wavelengths interact preferentially with constituents of the skin and are able to reveal the structure of the skin lesion (2).

The construction of systems with the ability to capture reliable and reproducible images of skin is rather challenging due to equipment and environmental constraints, such as image resolution, image noise, illumination, skin reflectivity and pose uncertainty. The use of commercially available photographic cameras is quite common in skin lesion inspection systems, particularly for telemedicine purposes (3). However, the poor resolution in small skin lesions, i.e. lesions with a diameter of <0.5 cm, and the variable illumination conditions are not easily handled, and high-resolution devices with low-distortion lenses must therefore be used. However, the requirement for constant image colors necessary for image reproducibility remain unsatisfied, as it requires real-time, automated color calibration of the camera, i.e. adjustments and corrections to operate within the dynamic range of the camera and always measure the same color regardless of

Correspondence to: Dr Ilias Maglogiannis, Department of Information and Communication Systems Engineering, University of Aegean, 83200 Karlovassi, Greece
E-mail: imaglo@aegean.gr

Key words: skin lesion, pattern analysis, melanoma, dermoscopy, discriminant analysis, neural networks, support vector machines

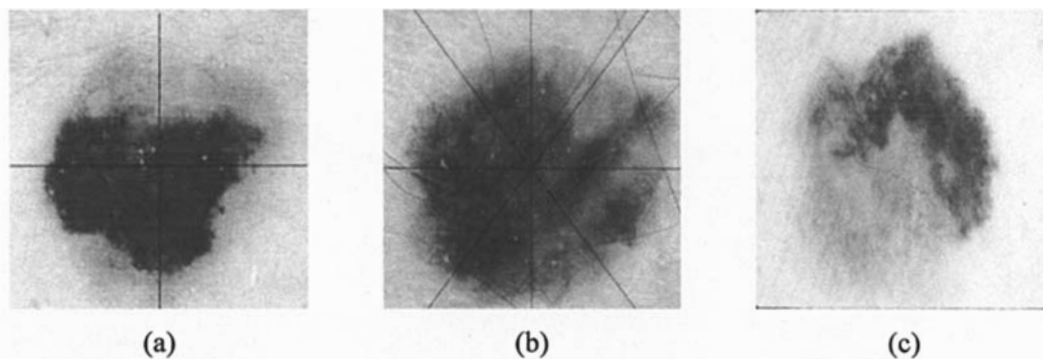


Figure 1. Asymmetry border color. Features include (a) asymmetry test, (b) border test, and (c) color variegation (7).

lighting conditions. The problem can be addressed by using video cameras that are parametrizable online and controlled through software at the price of higher complexity and costs (4,5).

Definition of features for detection of malignant melanoma. The features, i.e. the visual cues, used for melanoma detection were examined, providing answers to question 2. Similar to the traditional diagnosis procedure, the computer-based systems look for features and combine them to characterize the lesion as a malignant melanoma or dysplastic nevus. The employed features must be measurable and of high sensitivity, i.e. high correlation of the feature with malignant melanoma and high probability of true positive response. Further, the features should have high specificity, i.e. high probability of true negative response. Although both factors are considered important in the typical classification paradigm (a trade-off expressed by maximizing the area under the receiver-operating characteristic curve), the suppression of false negatives (i.e. increase of true positives) in the case of malignant melanoma is more important.

In the conventional procedure, the following diagnostic methods are mainly used (6): i) ABCD rule of dermoscopy ii) pattern analysis; iii) Menzies method; and iv) 7-point checklist. The features used for these methods are presented in this study.

The ABCD rule investigates the asymmetry (A), border (B), color (C) (Fig. 1), and differential structures (D) (Fig. 2) of the lesion and defines the basis for diagnosis by a dermatologist. More specifically: i) Asymmetry. The lesion is bisected by two axes that are positioned to produce the lowest asymmetry possible in terms of borders, colors, and dermoscopic structures; ii) Border. The lesion is divided into 8 pie-piece segments, then examined for the presence of sharp, abrupt cut-off pigment patterns at the periphery of the lesion or a gradual, indistinct cut-off; iii) Color. The number of colors present is determined and may include light brown, dark brown, black, red (red vascular areas are scored), white (if whiter than the surrounding skin), and slate blue; and iv) Differential structures. The number of structural components present is determined, i.e. pigment network, dots (scored if ≥ 3 are present), globules (scored if ≥ 2 are present), structureless areas (counted if area is $>10\%$ of lesion), and streaks (scored if ≥ 3 are present).

The pattern analysis method seeks to identify specific patterns that may be global (reticular, globular, cobblestone,

homogeneous, starburst, parallel, multicomponent, or non-specific) or local (pigment network, dots/globules, streaks, blue-whitish veil, regression structures, hypopigmentation, blotches, or vascular structures).

The Menzies method looks for negative (symmetry of pattern and presence of a single color) and positive features [blue-white veil, multiple brown dots, pseudopods, radial streaming, scar-like depigmentation, peripheral black dots/globules, multiple (5-6) colors, multiple blue/gray dots, and broadened network].

The 7-point checklist seeks to identify an atypical pigment network, blue-whitish veil, atypical vascular pattern, irregular streaks, irregular dots/globules, irregular blotches, regression structures.

Researchers that wish to automatically identify malignant melanoma exploit the available computational capabilities by searching for many of the above and additional features. The main features used for skin lesion image analysis and their calculation method are summarized below, answering question 3:

i) Asymmetry features. Asymmetry is examined with respect to a point on one or more axes. The asymmetry index is computed by first finding the principal axes of inertia of the tumor shape in the image by overlapping the two halves of the tumor along the principal axes of inertia and dividing the non-overlapping area differences of the two halves by the total area of the tumor.

ii) Border features. The most popular border features are the greatest diameter, area, border irregularity, thinness ratio (30), circularity index (32) and the variance of the distance of the border lesion points from the centroid location (32). Apart from regarding the border as a contour, emphasis is also placed on features that quantify the transition from the lesion to the skin. Such features are the minimum, maximum, average and variance responses of the gradient operator, applied on the intensity image along the lesion border.

iii) Color features. Typical color images consist of the three-color channels RGB (red, green, blue). The color features are based on measurements on these color channels or other color channels such as CMY (cyan, magenta, yellow), HSV (hue, saturation, value), YUV (Y-luminance, U-V chrominance components) or various combinations of them, linear or not. Color variegation may be calculated by measuring minimum, maximum, average and standard deviations of the selected channel values, as well as chromatic differences inside the lesion (7).

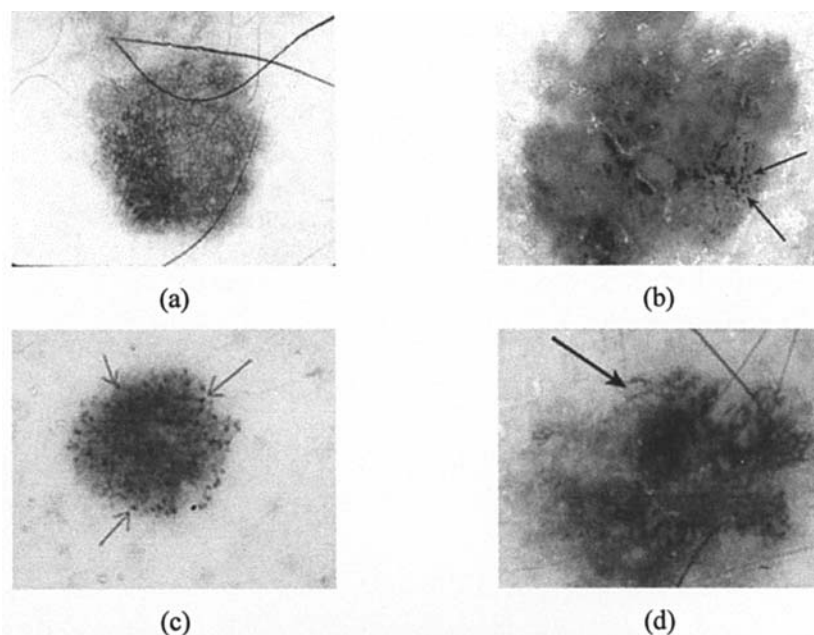


Figure 2. Differential structures. (a) Pigmented network, (b) dots, (c) brown globules, and (d) branched streaks (7).

iv) Differential structures. Differential structures, as described in the ABCD method, and most patterns used by pattern analysis, the Menzies method and the 7-point checklist are rarely used for automated skin lesion classification due to their complexity.

v) Skin lesion kinetics. Efforts have been made to elucidate the kinetics of skin lesions (24,25). The ratio of variances (RV) has been defined as (26):

$$RV = \frac{SD_{B^2}}{SD_{B^2} + SD_{I^2} + SD_{A^2}}$$

SD_{B^2} (standard deviation between days) is between day variance of the color variable computed using the mean values at each day of all wound sites and subjects. SD_{I^2} (standard deviation intra day) is the intra day variance of the color variable estimated from the computations at each day of all wound sites and subjects. SD_{A^2} (standard deviation analytical) is the variance of the color variable computed using normal skin sites of all subjects and times.

Feature selection. The success of image recognition depends on the correct selection of features used for the classification, which answers question 4. This is a typical optimization problem that may be resolved with heuristic strategies, greedy or genetic algorithms or other computational intelligence methods (9). The use of feature selection algorithms is motivated by the need for highly precise results, computational reasons and a peaking phenomenon often observed when classifiers are trained with a limited set of training samples. If the number of features is increased, the classification rate of the classifiers decreases after a peak (10,11).

Computational methods for skin lesion classification. We address and answer questions 5 and 6 by examining the most popular methods for skin lesion classification. The task mainly involves two phases after feature selection, learning and testing (7), which are analyzed in the following.

During the learning phase, typical feature values are extracted from a sequence of digital images representing classified skin lesions. The most classical recognition paradigm is statistical (12). Covariance matrices are computed for the discriminative measures, usually under the multivariate Gaussian assumption. Parametric discriminant functions are then determined, allowing classification of unknown lesions (discriminant analysis). The major problem of this approach is the need for large training samples.

Neural networks are networks of interconnected nodes composed of various stages that emulate some of the observed properties of biological nervous systems and draw on the analogies of adaptive biological learning. Learning occurs through training a large set of data where the training algorithm iteratively adjusts the connection weights (synapses), by minimizing a given error function (15,16). The weights of the features are thus automatically calculated. Popular choices for the error function in skin lesion image classification are the Euclidean distance or ratio deviation, defined as:

$$E = \sqrt{\sum_i (x_i - \bar{x})^2}, \quad E = \sum_i \left| 1 - \frac{x_i}{\bar{x}} \right|$$

where x_i is the i th sample and \bar{x} is the population mean (13).

The support vector machines (SVMs) is a popular algorithm for data classification into two classes (14,17,18). SVMs allow the expansion of information provided by a training dataset as a linear combination of a subset of the data in the training set (support vectors). These vectors locate a hypersurface that separates the input data with a good degree of generalization. The SVM algorithm is based on training, testing and performance evaluation, which are common steps in every learning procedure. Training involves the optimization of a convex cost function where there are no local minima to complicate the learning process. Testing is based on model evaluation using the support vectors to classify a test dataset. Performance evaluation is based on error rate determination, as test set data size tends to infinity.

Table I. Computer-based systems for the characterization of digital skin images.

Reference	Detection goal	Installation type	Visual features	Classification method	Success rates
(15,19)	Tumor, crust, hair, scale, shiny ulcer of skin lesions	Video RGB camera	Color (chromaticity) coordinates (more)	Neural networks	85-89% on average
(20)	Skin erythema	Video RGB camera	Color-CIE L*a*b*color space	Statistical	Monitoring indexes for follow ups
(21)	Burn scars	Video RGB camera	Image intensity, skin elasticity	Finite element analysis	Monitoring indexes for follow ups
(22)	Melanoma recognition	Video RGB camera	Color in RGB and HIS (more)	Statistical	5% deviation from manual diagnosis
(23)	Melanoma recognition	Tissue microscopy	Epidermal and dermal features (epidermis volume, thickness, dermal epidermal junction ratio, cellular and collagen densities)	Statistical	Difference was 5.33% in epidermal features and 2.76% for dermal features
(25)	Wound healing	Still CCD camera	Ratio of variances, in HIS and RGB	Healing indexes measuring, the wound area and the wound color	Monitoring indexes for follow-ups
(28)	Melanoma recognition	<i>In situ</i> , ultraviolet illumination	Auto fluorescence of skin tissues	Statistical	77% (81% manual diagnoses)
(29)	Melanoma recognition	Ultraviolet illumination	Imax/Imin, (fluorescence intensity)	Statistical	82.5% sensitivity, 78.6% specificity and 58.9% positive predictive value (average values of 14.3 for melanoma, 5.7 for naevi and 6.1 for other skin lesions)
(31)	Melanoma recognition	Epiluminescence microscopy (ELM)	RGB/HIS/border	Statistical (k-nearest-neighbor)	87% sensitivity and 92% specificity

Answering question 6, the performance of each classifier is tested using an ideally large set of manually classified images. A subset of them, e.g. 80% of the images, is used as a training set and the remaining 20% of samples is used for testing using the trained classifier. All possible combinations of the training and test images are explored to avoid bias in the solution.

Results

Systems in literature. The development of automated systems for melanoma classification preoccupies many biomedical

laboratories (22,23,27,29,31), which will be examined further. It is also interesting to include studies in our survey that discuss with the general problem of skin lesion image characterization, as they face similar problems. The lesions include tumor, crust, hair, scale, shiny and ulcer (15,19), erythema (20), burn scars (21) and wounds (24,25), among others.

The most common installation type appears to be the video camera, due to the control features it provides (15,19,20,21,22). The still camera is of use in some installations (24,25), while infra-red or ultraviolet illumination (*in situ* or *in vivo*) using the appropriate cameras is a popular

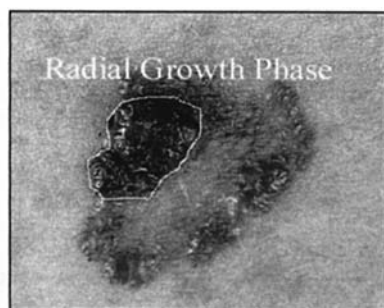


Figure 3. The RGP phase of melanoma is the circled area.

choice (27-29). Microscopy or epiluminescence microscopy installations have also been applied (23,31).

The most common features used for automated lesion characterization are those associated with color in various color spaces (RGB, HIS, CIELab), e.g. color values in (15,19,20). Some combine features in more than one color space for better results, e.g. HIS and RGB (22-25,31). Intensity characteristics (21) and ratios of maximum to minimum intensity value (29) have also been used. Asymmetry and border features as defined in Materials and methods are fairly common (31), while features based on differential structures are rare.

The most common classification methods are statistical (20,22,23,28,29). More advanced techniques such as neural networks have been applied (15,19), as well as the k-nearest neighborhood classification scheme (31).

Success rates for the methods presented in the literature indicate that efforts towards the automated classification of lesions and melanoma in particular may provide good results. These rates and other system features are summarized in Table I. We should note that the results are not comparable but rather indicative, mainly due to the fact that different

images from different cases are used. The classification success rates are also not applicable to the methods calculating healing indexes.

Classification results. We attempted to quantify the success rates of classification methods described above for the distinction of malignant melanoma from dysplastic nevus. Three groups of data were considered. The first group (VGP; vertical growth phase) consists of cases of malignant melanoma, with measurements taken on the entire extent of the lesion. The second group (RGP; radial growth phase) also refers to the malignant melanomas, but measurements are restricted to the dark area of the melanoma. The third group (DSP; dysplastic) comprises cases of dysplastic nevus (Fig. 3). Separate analyses were carried out, one between VGP and DSP, and the other between RGP and DSP. Both comparisons were made with linear discriminant analysis, by fitting a neural network model and utilizing the SVM algorithm. A training dataset of 34 cases at the Department of Plastic Surgery and Dermatology in Athens General Hospital were collected within a period of 6 months. The total number of lesions captured was 14 melanomas and 20 dysplastic naevi. Although the set is small, it provided some insight about classifier performance and how the number of features influences the results. The mean thickness of melanoma lesions was measured after biopsy at approximately 1.5 mm penetration through the skin.

The sensitivity and specificity rates using discriminant analysis, neural networks and support vector machines are presented in Tables II and III for VGP-DSP and RGP-DSP classification, respectively. We used the cross-validation or 'leave-one-out' estimator of the rate of correct classifications, which is obtained by monitoring how each observation is classified according to a function, then recalculated after omitting that observation from the analysis.

Table III. Sensitivity and specificity indexes of the RGP-DSP classification.

Method	Total correct classification	Sensitivity	Specificity
Discriminant analysis using 4 features as the most significant for discrimination	33/34 or 97%	93%	100%
Discriminant analysis using 2 features as the most significant for discrimination	30/34 or 88%	86%	100%
Neural networks using 4 principal components as input	34/34 or 100%	100%	100%
Neural networks using 2 principal components as input	32/34 or 94%	86%	90%
SVM (first order polynomial kernels, 5 support vectors)	97%	93%	100%

Table II. Sensitivity and specificity indexes of the VGP-DSP classification.

Method	Total correct classification	Sensitivity	Specificity
Discriminant analysis using 4 features as the most significant for discrimination	33/34 or 97%	93%	100%
Discriminant analysis using 2 features as the most significant for discrimination	32/34 or 94%	86%	100%
Neural networks using 4 principal components as input	33/34 or 97%	93%	100%
Neural networks using 2 principal components as input	29/34 or 85%	79%	90%
SVM (Gaussian RBF kernel, sigma=4, 7 support vectors)	32/34 or 94%	86%	100%

Discussion

The most remarkable systems for the automated detection of malignant melanoma have been surveyed. These systems employ a variety of methods for image acquisition, feature definition and extraction, and lesion classification from features.

The most promising image acquisition techniques appear to be those that reveal the skin structure through selected spectral images. However, the issue of repeatability of measurements for follow-up studies has not been satisfactorily resolved.

Regarding features, it is clear that emphasis has been on the assessment of lesion size, shape, color, and texture. These statistical parameters were chosen primarily for computational convenience, and can be acquired with well-established analytic techniques at a manageable computational cost. However, they do not correspond to known biological phenomena or model human interpretation of dermoscopic imagery. On the contrary, the structural patterns considered essential for manual lesion categorization appear to have been neglected by the computational intelligence community, due to their complexity, although their exploitation could provide crucial information.

As far as the classification method is concerned, the SVM appears to perform better. However, it is the selected features that are critical for the performance of the classifier and training procedure, which must include the largest possible variety of cases.

The results presented by the research community are promising. It is now necessary to examine more patients in order to increase the number of cases, particularly during the classification phase. This will clarify the issue of selecting the most powerful variables for classification and may enable even better classification if the examination of differences in results between the two methods casts light on why misclassifications can arise.

References

1. Marks R: Epidemiology of melanoma. *Clin Exp Dermatol* 25: 459-463, 2000.
2. Moncrieff M, Cotton S, Claridge E and Hall P: Spectrophotometric intracutaneous analysis – a new technique for imaging pigmented skin lesions. *Br J Dermatol* 146: 448-457, 2002.
3. Loane M, Gore H, Corbet and Steele K: Effect of camera performance on diagnostic accuracy. *J Telemed Telecare* 3: 83-88, 1997.
4. Maglogiannis I and Kosmopoulos D: A system for the acquisition of reproducible digital skin lesion images. In: *Technology and Healthcare*, IOS Press, pp425-441, 2003.
5. Gutenev A, Skladnev VN and Varvel D: Acquisition-time image quality control in digital dermatoscopy of skin lesions. *Computerized Med Imaging Graphics* 25: 495-499, 2001.
6. Giuseppe A: Dermoscopy of pigmented skin lesions: Results of a consensus meeting via the Internet. *J Br Dermatol* 48: 680-693, 2003.
7. Electronic available information at Skin Oncology Teaching Center: <http://www.dermoncology.com/>
8. Maglogiannis I, Caroni C, Pavlopoulos S and Karioti V: Utilizing artificial intelligence for the characterization of dermatological images. 4th International Conference: Neural Networks and Expert Systems in Medicine and Healthcare. *NNESMED* pp362-368, 2001.
9. Handels H, Ross TH, Kreusch J, Wolff HH and Pöppel SJ: Feature selection for optimized skin tumor recognition using genetic algorithms. *Artif Intell Med* 16: 283-297, 1999.
10. Jain AK: Advances in statistical pattern recognition. In: *Pattern Recognition, Theory and Applications*. Devijver PA and Kittler J (eds). Springer, Berlin, 1986.
11. Jain AK and Waller WG: On the optimal number of features in the classification of multivariate Gaussian data. *Patt Recog* 10: 365-374, 1978.
12. Duda RO and Hart PE: *Pattern classification and skin analysis*. John Wiley, New York, 1973.
13. Jain A and Dubes R: *Algorithms for Clustering Data*. Prentice Hall, Englewood Cliffs, NJ, 1988.
14. Burges C: A tutorial on support vector machines for pattern recognition. <http://www.kernel-machines.org/>.
15. Umbaugh S, Moss R and Stoecker W: Applying artificial intelligence to the identification of variegated coloring in skin tumors. *IEEE Eng Med Biol Magazine*, pp57-62, 1991.
16. Durg A, Stoecker W, Vookson J, Umbaugh S and Moss R: Identification of variegated coloring in skin tumors. *IEEE Eng Med Biol Magazine*, pp71-75, 1993.
17. Christianini N and Shawe-Taylor J: *An introduction to support vector machines*. Cambridge University Press, 2000.
18. Schölkopf B: Statistical learning and kernel methods. <http://research.microsoft.com/~bsc>.
19. Umbaugh S, Wei Y and Zuke M: Feature extraction in image analysis. *IEEE Eng Med Biol Magazine*, pp62-73, 1997.
20. Nischic M and Forster C: Analysis of skin erythema using true color images. *IEEE Trans Med Imaging* 16: 6, 1997.
21. Tsap L, Goldgof D, Sarkar S and Powers P: Vision-based technique for objective assesment of burn scars. *IEEE Trans Med Imaging* 17: 620-633, 1998.
22. Tomatis S, Bartol C, Tragni C, Farina B and Marchesini R: Image analysis in the RGB and HS color planes for a computer assisted diagnosis of cutaneous pigmented lesions. *Tumori* 84: 29-32, 1998.
23. Sanders J, Goldstein C, Leotta and Richards K: Image processing techniques for quantitative analysis of skin structures. *Computer Methods Programs Biomed* 59: 167-180, 1999.
24. Hansen G, Sparrow E, Kokate J, Leland K and Iaizzo P: Wound status evaluation using color image processing. *IEEE Trans Med Imaging* 116: 78-86, 1997.
25. Herbin M, Bon F, Venot A, Jeanlouis F, Dubertret M, Dubertret L and Strauch G: Assessment of healing kinetics through true color image processing. *IEEE Trans Med Imaging* 112: 39-43, 1993.
26. Chin S: The assessment of methods of measurements. *Stat Med* 9: 351-362, 1990.
27. Lohman W and Paul E: *In situ* detection of melanomas by fluorescence measurements. *Naturewissenschaften* 75: 201-202, 1988.
28. Bono A, Tomatis S and Bartoli C: The invisible colors of melanoma – a telespectrophotometric diagnostic approach on pigmented skin lesions. *Eur J Cancer* 32A: 727-729, 1996.
29. Chwirot B, Chwirot S, Redziski J and Michniewicz Z: Detection of melanomas by digital imaging of spectrally resolved ultraviolet light-induced autofluorescence of human skin. *Eur J Cancer* 34: 1730-1734, 1998.
30. Maglogiannis I and Zafiropoulos E: Utilizing support vector machines for the characterization of digital medical images. *BMC Med Informatics Decision Making* 4: 4, 2004.
31. Ganster H, Pinz P, Rohrer R, Wildling E, Binder M and Kittler H: Automated melanoma recognition. *IEEE Trans Med Imaging* 20: 233-239, 2001.
32. Bono A, Tomatis S, Bartoli C, Tragni G, Radaelli G, Maurichi A and Marchesini R: The ABCD system of melanoma detection: A spectrophotometric analysis of the asymmetry, border, color, and dimension. *Cancer* 85: 72-77, 1999.