10

Computer-based diagnosis of pigmented skin lesions

Hitoshi Iyatomi
Hosei University, Faculty of Science and Engineering
Japan

1. Introduction – Malignant melanoma and dermoscopy

The incidence of malignant melanoma has increased gradually in most parts of the world. There is a report that the incidence of melanoma is now approaching 50 cases per 100,000 population in Australia (Stolz et al., 2002) and another report describes that a total of 62,480 incidence and 8,420 deaths are estimated in United States in 2008 (Jemal et al., 2008). Although advanced malignant melanoma is often incurable, early-stage melanoma can be cured in many cases, particularly before the metastasis phase. For example, patients with a melanoma less than or equal to 0.75 mm in thickness have a good prognosis and their five-year survival rate is greater than 93% (Meyskens et al., 1998). Therefore, early detection is crucial for the reduction of melanoma-related deaths. On the other hand, however, it is often difficult to distinguish between early-stage melanoma and Clark nevus, one of melanocytic pigmented skin lesion, with the naked eye, especially when small lesions are involved.

Dermoscopy, a non-invasive skin imaging technique, was introduced to improve accuracy in the diagnosis of melanoma (Soyer et al., 1987). It uses optical magnification and either liquid immersion and low angle-of-incidence lighting or cross-polarized lighting to make the contact area translucent, making subsurface structures more easily visible when compared to conventional macroscopic (clinical) images (Tanaka, 2006). Fig 1 shows examples of (a) a clinical image of early stage of melanoma and (b) a dermoscopy image of the same lesion. The dermoscopy image has no defused reflection on the skin surface and shows the internal structures clearly. In this case, an experienced dermatologist found regression structures (tint color areas) in the dermoscopy image and concluded that this lesion should be malignant.

Several diagnostic schemes based on dermoscopy have been proposed and tested in clinical practice including the ABCD rule (Stolz et al., 1994), Menzies' scoring method (Menzies et al., 1996), the 7-point checklist (Argenziano et al., 1998), the modified ABC-point list (Blum et al., 2003), and the 3-point checklist (Soyer et al., 2004). A systematic review covering Medline entries from 1983 to 1997 revealed that dermoscopy had 10-27% higher sensitivity (Mayer et al., 1997). However, dermoscopic diagnosis is often subjective and is therefore associated with poor reproducibility and low accuracy especially in the hands of inexperienced dermatologists. Despite the use of dermoscopy, the accuracy of expert dermatologists in diagnosing melanoma is estimated to be about 75-84% (Argenziano et al., 2003).

In order to overcome the above problems, automated and semi-automated procedures for classification of dermoscopy images and related techniques have been investigated since the late 1990s. This chapter introduces the recent advancement of those investigations with Internet based melanoma screening system developed by authors.
The following of the chapter is organized as follows: section 2 describes the diagnostic scheme of melanomas and outlines computer-based melanoma diagnosis in methodology and past studies; section 3 introduces an outline of our web-based melanoma screening system and its architectonics; section 4 explains Asian specific melanomas found in acral volar regions and their automated method for diagnosis; section 5 describes the remaining issues needing to be addressed in this field and the conclusion is given in section 6.

(a) clinical image (b) dermoscopy image

Fig. 1. Sample of (a) clinical image and (b) dermoscopy image of same lesion (malignant melanoma)

2. Diagnosis of melanoma

This section firstly introduces the diagnosis scheme for melanomas, namely how dermatologists diagnose melanomas, and then introduces the computer-based diagnosis in terms of methodological outline and past studies.

2.1 Diagnosis scheme for melanomas

This subsection introduces well-known and commonly-used diagnosis scheme for dermoscopy images, the ABCD rule (Stolz et al., 1994) and the 7-point checklist (Argenziano et al., 1998) for further understanding.

2.1.1 ABCD rule

This is one of the most well-known semi-quantitative diagnosis schemes. It quantifies the asymmetry (A), border sharpness (B), color variation (C) and the number of differential structures (D) present in a lesion. Table 1 summarizes these definitions and their relative weights. A describes the degree of asymmetry of the tumor. Assuming a pair of orthogonal symmetry axes intersecting at the centroid of the tumor, A can be 0 (symmetry along both axes), 1 (symmetry along one axis), or 2 (no symmetry). B represents the number of border octants with sharp transitions. C indicates the number of significant colors present in the tumor, of which six are considered to be significant: white, red, light-brown, dark-brown, blue-gray, and black. Finally, D represents the number of differential structures (pigment network, structureless or homogeneous areas, streaks, dots, and globules) present in the tumor. Using the ABCD rule, the total dermoscopy score (TDS) is calculated as follows:

$$TDS = (A \times 1.3) + (B \times 0.1) + (C \times 0.5) + (D \times 0.5). \quad (1)$$

TDS below 4.75 indicates benignity, whereas TDS above 5.45 indicates malignancy. A score between these limits corresponds to a suspicious case that requires clinical follow-up.
Fig.2 shows the sample dermoscopy image. Dermatologists will find no symmetry border (A=2), more than half of the tumor border has sharp color transition (B=5), four colors (white, light-brown, dark-brown, blue-gray) in the tumor area (C=4) and five defined structures (pigment network, structureless or homogeneous areas, streaks, dots, and globules) (D=5).With these criteria, TDS becomes 7.6 (>5.45) and they can conclude this tumor is malignant.

Criterion	Description	score	weight
Asymmetry	Number of asymmetry axes	0-2	× 1.3
Border	Number of border octants with sharp transition	0-8	× 0.1
Color	Number of significant colors from: white, red, light-brown, dark-brown, bule-gray, and black	1-6	× 0.5
Differential structures	Number of differential structures from: pigment network, structureless, streaks, dots, and globules	0-5	× 0.5

Table 1. Brief summary of ABCD rule (Stolz et al., 1994)

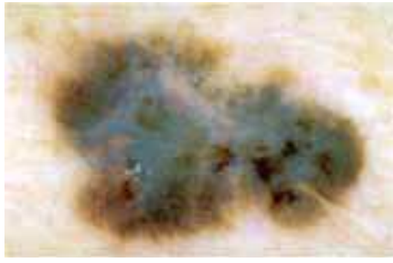


Fig. 2. Sample of dermoscopy image (malignant melanoma)

2.1.2 7-point checklist

This is another well-known diagnostic method that requires the identification of seven dermoscopic structures that are shown in Table 2. The score for a lesion is determined as the weighted sum of the structures present in it. Using the 7-point checklist, the total score (TS) is calculated as follows:

$$TS = (\#_{major} \times 2) + (\#_{minor}). \quad (2)$$

Here $\#_{major}$ and $\#_{minor}$ are the number of dermoscopic structures (see Table 2) present in the image. If TS is greater than or equal to 3, then the lesion is considered to be malignant.
In Fig. 2, dermatologists will find "blue-whitish veil", "irregular streaks", "irregular pigmentation", "irregular dots/globules", and "regression structures". Accordingly, TS becomes 6 (≥3) and they consider this tumor to be malignant. Note again that the scores in the above examples may vary among physicians. See (Argenziano et al., 2003) in detail.

Major criterion	weight
1. Atypical pigment network	× 2
2. Blue-whitish veil	× 2
3. Atypical vascular pattern	× 2
Minor criterion	weight
4. Irregular streaks	× 1
5. Irregular pigmentation	× 1
6. Irregular dots / globules	× 1
7. Regression structures	× 1

Table 2. Brief summary of 7-point checklist (Argenziano et al., 1998)

2.2 Computer-based diagnosis of melanomas

Several groups have developed automated analysis procedures to overcome the above mentioned problems – difficulty, subjectivity and low reproducibility of diagnosis, and reported high levels of diagnostic accuracy. The pioneering study of fully automated diagnosis of melanoma was conducted by Green et al. (Green et al., 1991). In their study tumor images were captured using a CCD color video camera.

Table 3 shows the list of recent studies in this topic. The diagnosis process of most automated diagnosis methods can be divided into three steps:

(1) Determination of tumor area from dermoscopy image,

(2) Extraction of image features from image,

(3) Building of the classification model and evaluation.

In the following section, outline of these steps is described with our Internet-based system as an example.

Reference	**Classifier**	**#Images**	**SE(%)**	**SP(%)**	**comments**
Gunster et al., 2001	k-NN	5363	73.0	89.0	(not dermoscopy images)
Elbaum et al., 2001	Linear	246	100	85.0	
Rubegni et al., 2002	ANN	550	94.3	93.8	
Hoffmann et al.,2003	Logistic	2218	-	-	AUC*=0.844
Blum et al., 2003	ANN	837	82.3	86.9	
Oka et al., 2004	Linear	247	87.0	93.1	**Internet-based**
Burroni et al., 2004	Linear	174	71.1	72.1	Only melanoma in situ
Seidenari et al., 2005	Linear	459	87.5	85.7	AUC*=0.933
Menzies et al., 2005	Logistic	2420	91.0	65.0	
Celebi et al., 2007b	SVM	564	93.3	92.3	
Iyatomi et al., 2008b	ANN	1258	85.9	86.0	**Internet-based** AUC=0.928

* area under the ROC (receiver operating characteristics) curve

Table 3. Recent studies in automated diagnosis for melanomas

3. Internet-based melanoma screening system

The software-based approaches introduced in the section 2, however, have several problems or limitations for practical use. For example, results of these studies are not comparable because of the different image sets used in each one. In addition, these studies were designed to develop a screening system for new patients using standalone systems and therefore they have not been opened to the public.

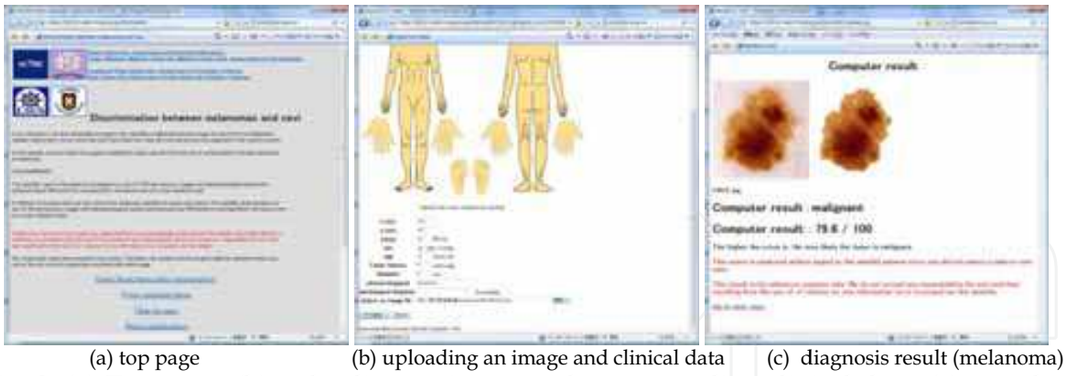
In 2004, authors developed the first Internet-based melanoma screening system and opened it for public use with the intention to solve abovementioned issues (Oka et al., 2004). The URL of the site has changed and it is now http://dermoscopy.k.hosei.ac.jp. Here we show the current top page of the site in Fig.3(a). When one uploads a dermoscopy image, inputs photographed body region of the tumor and the associated clinical data (Fig. 3(b)), the system extracts the tumor area, calculates the tumor characteristics and reports a diagnosis based on the output of built linear or artificial neural network classifier (Fig. 3(c)). Collecting many dermoscopy images for building a classifier is the most important issue to ensure system accuracy and generality. However, this is not a trivial task because obtaining the diagnosis information, dermatologists usually need histopathological tests or long term clinical follow-up. To address this issue, our system is designed to store the uploaded dermoscopy images into our database and waiting a final diagnosis by pathological examination etc. as a feedback from the users, if available.

Since we made this system open to the public, we have identified several issues that would make the system more practical. We have thus focused on the following topics: (1) expansion of the image database for building a classifier, (2) development of a more accurate tumor area extraction algorithm, (3) extraction of more discriminative diagnostic features, (4) development of an effective classification model, and (5) reduction of the system response time.

The latest version of our system (Iyatomi et al., 2008b) features a sophisticated tumor-area extraction algorithm that attains superior extraction performance to conventional methods (Iyatomi et al., 2006) and linear and back-propagation artificial neural network classifiers. The system has a capability to accept the usual melanocytic pigmented lesions (e.g. Clark nevi, Spitz nevi, dermal nevi, blue nevi, melanomas etc. - research target of most conventional studies as listed in Table 1) and it can also accept acral volar skin lesions that are found specifically in palm and sole area of non-white people. Acral lesions have completely different appearances and therefore a specific classification model is required to analyze these lesions (details are described in section 4). Our system automatically selects the appropriate diagnostic classifier based on the location of lesions provided by the user and yields the final diagnosis results in the form of a malignancy score between 0 and 100 within 3-10 seconds (see Fig.3(c)).

For non-acral lesions, the system achieved 85.9 % SE, 86.0% SP and 0.928 area under the receiver operating characteristics (ROC) curve (AUC) using a leave-one-out cross-validation test on a set of 1258 dermoscopy images (1060 nevi and 198 melanomas) and for acral volar lesions, 93.3% SE, 91.1% SP and 0.991 AUC on a set of 199 dermoscopy images (169 nevi and 30 melanomas). Fig. 4 shows the ROC curve for our latest screening system for (1) non-acral and (2) acral lesions.

In this section, key components of our web-based system, namely determination of tumor area, extraction and selection of important image features, building classifiers, and their performances are described. In the following section (section 4), introduction of acral volar skin lesions and their automated diagnosis is explained.



(a) top page (b) uploading an image and clinical data (c) diagnosis result (melanoma)

Fig. 3. (a) Top page, (b) uploading an image and corresponding clinical data, (c) sample of result page

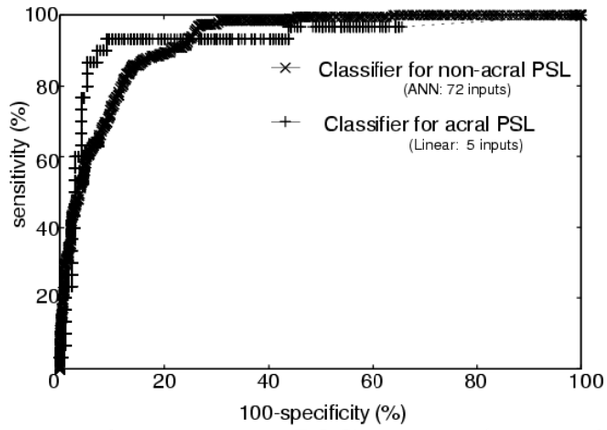


Fig. 4. Receiver operating characteristics (ROC) curves for our latest Internet-based melanoma screening system (classifier for non-acral lesions and acral lesions)

3.1 Tumor area extraction from surrounding skin

Diagnostic accuracy highly depends on the accurate extraction of the tumor area. Since the late 90s, numerous solutions that address this issue have been reported (Celebi et al., 2009). A notable problem with these studies is that the computer-extracted regions were often smaller than the dermatologist-drawn ones resulting in the area immediately surrounding the tumor, an important feature in the diagnosis of melanoma, being excluded from the subsequent analysis (Grana et al., 2003). Therefore, there is need for developing a more accurate tumor area extraction algorithm that produces results similar to those determined by the dermatologists.

We developed "dermatologist-like" tumor area extraction algorithm (Iyatomi et al., 2006) introduces a region-growing approach that aims to bring the automatic extraction results closer to those determined by expert dermatologists. To the best of author's knowldge, this algorithm was developed based on the largest number of manual extraction results by expert dermatologists at present and quantitatively evaluated performance showed that is had almost equivalent extraction performance to that by expert dermatologists. With those reasons, this algorithm is now used on our web-based screening system.

In this section, a brief summary of our method is introduced and explained. For more detailed information on this topic, please see survey paper (Celebi et al., 2009).
The "dermatologist-like" tumor area extraction algorithm is developed based on a total of 319 dermoscopy images from EDRA-CDROM (Argenziano et al., 2000) (244 melanocytic nevi and 75 melanomas) and their manual extraction results of tumor area by five expert dermatologists. The algorithm consists of four phases:
(1) initial tumor area decision,
(2) regionalization,
(3) tumor area selection, and
(4) region-growing.
In the following subsections, we introduce the method briefly and show some examples. For more details, please refer to the original article (Iyatomi et al., 2006).

3.1.1 Initial tumor area decision phase

This method uses two filtering operations before the selection of a threshold. First, the image was processed with a Gaussian filter to eliminate the sensor noise. Then, the Laplacian filter was applied to the image and the pixels in the top 20% of the Laplacian value were selected. Only these selected pixels were used to calculate a threshold. The threshold was determined by maximizing the inter-group variance (Otsu, 1988) with blue channel of image and the darker area was taken as a tentative tumor area.

3.1.2 Regionalization phase

Because many isolated small regions were created in the previous phase, these needed to be merged in order to obtain a continuous or a small set of tumor areas. First, a unique region number was assigned to each connected region. Second, a region smaller than predefined certain size (ratio to the image size) was combined with the adjacent larger region that shares the longest boundary. This phase makes it possible to manipulate the image as an assembly of regions.

3.1.3 Tumor area selection phase

Tumor areas were experimentally determined by selecting appropriate areas from the segmented regions using experimentally decided rules. The main objective of this phase is to eliminate undesired surrounding shadow areas that are sometimes produced by narrow shooting area of the dermoscopy. The regions that fulfilled these specific conditions were selected as the tumor region.

3.1.4 Region-growing phase

The extracted tumor area was expanded along the pre-defined border by a region-growing algorithm in order to bring it closer to the area selected by dermatologists. This method traverses the border of the initial tumor using a window of S×S pixels. When the color properties of the inner V_{in} and outer V_{out} regions of the tumor are similar, all of the neighborhood pixels are considered as part of the tumor area. This procedure is performed on each and every border pixel. This modification makes the tumor size larger and the

border of the tumor is redefined. This procedure is repeated iteratively until the size of tumor becomes stable.

3.1.5 Evaluation of tumor area extraction

We used a total of 319 dermoscopy images and evaluated the algorithm from a clinical perspective using manually determined borders from five expert dermatologists. Five dermatologists, with an average of 11 years of experience manually determined the borders of all tumors using a tablet computer. Even though these were expert dermatologist, their manual extraction results of the tumor area showed more than minor differences from each other (standard deviation of the extracted area is 8.9% of the tumor size in average) and therefore the determination of the standard tumor area (STA) for each image was necessary to be done in advance. We compared the extraction results from 5/5 medical doctor (5/5 MD) area (the region that is selected by all five dermatologists) to 1/5 MD area (the region that is selected by at least one dermatologist) and evaluated the standard deviation (SD) of the selected area. We concluded that the area extracted by two or more dermatologists (2/5 MD area) could be taken as the standard tumor area (STA).

Fig. 5 shows examples of tumor extraction results. From left to right: (a) dermoscopy image, (b) extraction result by conventional thresholding method, (c) extraction result by our "dermatologist-like" method, and (d) manual extraction result by five expert dermatologists. In manual the extraction results, the black area represents the area selected by all five dermatologists and the gray one is that selected by at least one dermatologist.

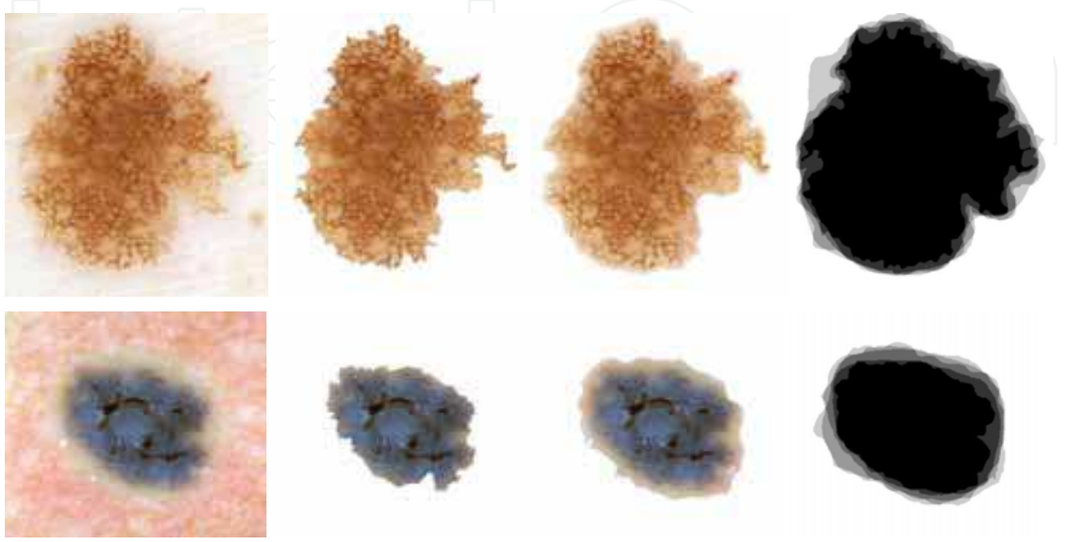
We used *precision* and *recall* criteria for performance evaluation. Their definitions are as follows:

$$precision = \frac{\text{correctly extracted area}}{\text{extracted area}}. \qquad recall = \frac{\text{correctly extracted area}}{\text{STA}}. \tag{3}$$

Note that "correctly extracted area" is the intersectional parts of the STA and the extracted area. The *precision* indicates "How accurate the extracted area was" and *recall* indicates "How well the tumor area was extracted". Those are ambivalent criteria and good extraction requires both precision and recall in high levels.

The summary of evaluation results for tumor area extraction is shown in Table 4. The conventional thresholding method showed excellent precision (99.5%) but low recall (87.6%), because this method tended to extract the inner area of the STA. This score indicated that the extracted area was smaller and almost all of the extracted area was in the tumor area. The characteristics of peripheral part of tumors are important for diagnosing melanoma so that inadequate extraction could have lost important information. Other computer-based methods using the clustering technique showed a similar trend; they had high precision but low recall when compared with the results of dermatologists. Given that the SD of the tumor areas manually extracted by five dermatologists was 8.9%, the precision of the proposed algorithm can be considered to be high enough and the extracted areas were almost equivalent to those determined by dermatologists. In addition, this algorithm provided better performance compared with that by non-medical individuals. With this result, we can consider we don't have to prepare manual interface for tumor area extraction when we widen the target audience of the system for non medically-trained individuals.

Celebi *et al.* compared recent seven tumor area extraction algorithms using a total of 90 dermoscopy images with manual extraction results by three expert dermatologists as a gold standard (Celebi et al., 2007a). In their evaluation, our dermatologist-like tumor area extraction algorithm achieved the lowest error in the benign category (mean $\pm$ SD = 10.66$\pm$5.13%) and the second lowest in the overall image set (11.44$\pm$6.40%).



(a) dermoscopy image | (b) conventional method | (c) "dermatologist-like" method | (d) five expert dermatologists

(upper) Clark nevus — precision = 100, recall = 79.8 | precision =98.3, recall =92.9

(lower) Melanoma — precision = 100, recall = 73.9 | precision =95.0, recall =96.1

Fig. 5 Comparison of tumor area extraction results: From left to right: dermoscopy image, extraction result by conventional thresholding method, extraction result by our "dermatologist-like" method, and manual extraction result by five expert dermatologists.

Methods	precision	recall
Conventional thresholding	99.5	87.6
Average of 10 non-medical individuals	97.0	90.2
Dermatologist-like	94.1	95.3

Table 4. Summary of tumor extraction performance

3.1.6 Importance of tumor area extraction from diagnostic performance

The effectiveness of our extraction method for the diagnostic accuracy was evaluated using a total of 319 dermoscopy images (same dataset). We extracted a tumor area by conventional and dermatologist-like methods, calculated a total of 64 image features (Oka et al., 2004) from each image, and then built a linear classifier using a incremental stepwise input selection method. The final diagnostic accuracy was evaluated by drawing the ROC curve of each classifier and the area under the each ROC curve (AUC) was evaluated. Our dermatologist like tumor extraction algorithm had improved diagnostic accuracy over the

other conventional methods. AUC increased from 0.795 to 0.875 with improvement of tumor extraction algorithm. When the diagnostic threshold was defined at a sensitivity of 80%, our extraction method showed approximately 20% better accuracy in specificity.

3.2 Feature extraction from the image

After the extraction of the tumor area, the tumor object is rotated to align its major axis with the Cartesian x-axis. We extract a total of 428 image related objective features (Iyatomi et al., 2008b). The extracted features can be roughly categorized into asymmetry, border, color and texture properties. In this section, a brief summary is described, please refer the original article for more details.

(a) Asymmetry features (80 features): We use 10 intensity thresholds values from 5 to 230 with a stepsize of 25. In the extracted tumor area, thresholding is performed and the areas whose intensity is lower than the threshold are determined. From each such area, we calculate 8 features: area ratio to original tumor size, circularity, differences of the center of gravity between original tumor, standard deviation of the distribution and skewness of the distribution.

(b) Border features (32 features): We divide the tumor area into eight equi-angle regions and in each region, we define an $S_B \times S_B$ window centered on the tumor border. In each window, a ratio of color intensity between inside and outside of the tumor and the gradient of color intensity is calculated on the blue and luminance channels, respectively. These are averaged over the 8 equi-angle regions. We calculate four features for eight different window sizes; 1/5, 1/10, 1/15, 1/20, 1/25, 1/30, 1/35 and 1/40 of the length of the major axis of the tumor object L.

(c) Color features (140 features): We calculated minimum, average, maximum, standard deviation and skewness value in the RGB and HSV color spaces, respectively (subtotal 30) for the whole tumor area, perimeter of the tumor area, differences between the tumor area and the surrounding normal skin, and that between peripheral and normal-skin (30×4=120). In addition, a total of 20 color related features are calculated; the number of colors in the tumor area and peripheral tumor area in the RGB and HSV color spaces quantized to 8^3 and 16^3 colors, respectively (subtotal 8), the average color of normal skin (R, G, B, H, S, V: subtotal 6), and average color differences between the peripheral tumor area and inside of the tumor area (R, G, B, H, S, V subtotal 6). Note that peripheral part of the tumor is defined as the region inside the border that has an area equal to 30% of the tumor area based on a consensus by several dermatologists.

(d) Texture features (176 features): We calculate 11 different sized co-occurrence matrices with distance value δ ranging from $L/2$ to $L/64$. Based on each co-occurrence matrix, energy, moment, entropy and correlation were calculated in four directions (0, 45, 90 and 135 degrees).

3.3 Feature selection and build a classifier

Feature selection is one of the most important steps for developing a robust classifier in any case. It is also well known that building a classifier with highly correlated parameters was adversely affected by so called "multi collinearity" and in such a case the system loses accuracy and generality.

In our research, we usually prepare two types of feature sets, (1) original image feature set and (2) orthogonal feature set. Using the original image feature set, the extracted image features are used directly as input candidates in the classifier and therefore we can clearly observe the relationship between image features and the target (e.g. diagnosis). However, using the original image features has the above mentioned potential risk. Note that the risk of multi collinearity is greatly reduced by appropriate input selection. On the other hand using the orthogonal feature set, finding the relationship between the image features and the target (e.g. diagnosis) becomes complicated, but this can show us the global trends with further investigation. To calculate the orthogonal image features, we extracted a total of 428 features per image and transformed them into the [0, 1] range using z-score normalization and then orthogonalized them using the principal component analysis (PCA).

The parameters used in melanoma classifiers are selected by an incremental stepwise method which determines the statistically most significant input parameters in a sequential manner. This method searches appropriate input parameters one after the other according to the statistical rule. This input selection method rejects statistically ignorable features during incremental selection and therefore, these highly correlated features were automatically excluded from the model. Note that using orthogonal feature sets frees from this problem.

The details of the feature selection is as follows:

(Step 0) Set the base parameter BP=null and number of the base parameter $\#_{BP}=0$.

(Step 1) Search one input parameter x* from all parameters $\boldsymbol{x}$ where regression model with x* yields best performance (lowest residual) among all. Set BP to x* and $\#_{BP}=1$.

(Step 2) Build linear regression models whose input elements are BP and x' without redundancy $\forall x' \in \mathbf{x}$, number of input is $\#_{BP}+1$ and select one input candidate x^ which has the highest partial correlation coefficient among x'.

(Step 3) Calculate the variance ratio (*F*-value) between the regression sum of squares and the residual sum of squares of the built regression model.

(Step 4) Perform statistical F-test (calculate p value) in order to verify that the model is reliable.

If $p<0.05$: BP $\leftarrow$ BP $+ x^{\wedge}$, $\#_{BP} \leftarrow \#_{BP}+1$ and return to (step 2). Else if $0.05 \leq p < 0.10$: discard x^ and return to (step 2) and find the next best candidate.Else if the developed model has a statistically negligible parameter x^ $(0.10 \leq p)$ among currently selected input, exclude x^ from BP, $\#_{BP} \leftarrow \#_{BP}-1$ and return to (step 2). Otherwise terminate the feature selection process.

Based on selected image features by above mentioned method, we built a back propagation artificial neural network (ANN) to classify dermoscopy images into benign or malignant.

Although ANNs have excellent learning and function approximation abilities, it is desirable to restrict the number of hidden-neurons and input nodes to a minimum in order to obtain a general classification model that performs well on future data (Reed et al., 1993).

In our network design, we had only one output node. This is because our aim was to classify the input as malignant or benign. All nevi such as Clark nevi, Reed nevi, blue nevi, and dermal nevi are equally considered as benign. Note that we assigned a training signal of 0.9 and 0.1 to melanoma and benign classes, respectively. If the output of the ANN exceeded the diagnostic threshold θ, we judged the input tumor as being malignant.
On a separate note, our system provides the screening results not only in the form of ``benign'' or ``malignant'', but also as a malignancy score between 0 and 100 based on the output of the ANN classifier. We assigned a malignancy score of 50 to the case where the output of the ANN was θ. For other values, we adjust this score of 0, 20, 80, and 100 according to the output of the ANN of 0, 0.2, 0.8 and 1.0, respectively using linear interpolation. This conversion is based on the assumption that the larger score of the classifier is, the more malignancy is. Although this assignment procedure is arbitrary, we believe the malignancy score can be useful in understanding the severity of the case.
We also built a linear classifier using the same method as a baseline for the classification performance comparison.

3.4 Performance evaluation

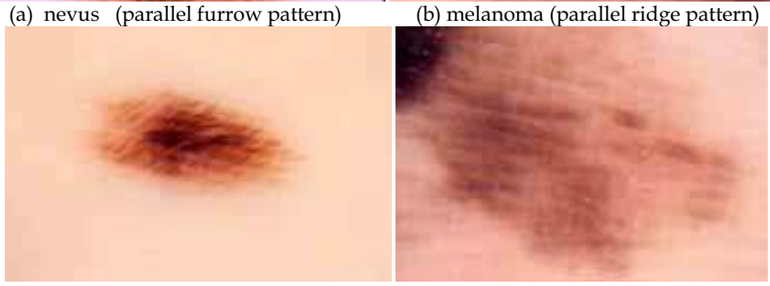
We used a total 1258 dermoscopy images with diagnosis (1060 cases of melanocytic nevi and 198 melanomas) from three European university hospitals (University of Naples, Graz, and Vienna) and one Japanese university hospital (Keio University). The diagnostic performance was evaluated by leave-one-out cross-validation test.
The incremental stepwise method selected 72 orthogonalized features from 428 principal components and all selected features were statistically significant ($p<0.05$). In this experiment, the basic back-propagation algorithm with constant training coefficients achieved the best classification performance among the tested training algorithms. The ANN classifier with 72 inputs - 6 hidden neurons achieved the best performance of 85.9% in SE, 86.0% in SP, and an AUC value of 0.928. Introducing a momentum term boosted the convergence rate at the expense of reduced diagnostic accuracy (Note that linear model with same inputs achieved 0.914 in AUC).
The classification performance is quite good considering that the diagnostic accuracy of expert dermatologists was 75-84% and that of histological tissue examination on difficult case sets was as low as 90% (Argenziano et al., 2003). In this study, we used ANN and linear models for classification. Using other models such as support vector machine classifier may improve performance, however importance of model selection is less than selecting efficient features in this task.
On the other hand, despite the good classification performance obtained, our system has several limitations regarding the acceptable tumor classes and the condition of the input images. At the present, the diagnostic capability of our system does not match that of expert dermatologists.The primary reason for this is the lack of a large and diverse dermoscopy image set.

4. Diagnosis of Asian specific melanomas

In non-white populations, almost half of the melanomas are found in acral volar areas and nearly 30% of melanomas affect the sole of the foot (Saida et al., 2004). Saida *et al.* also reported that melanocytic nevi are also frequently found in their acral skin and

approximately 8% of Japanese have melanocytic nevi on their soles. They reported that about 90% of melanomas in this area have the parallel ridge pattern (ridge areas are pigmented) and 70% of melanocytic nevi have the parallel furrow pattern (furrow areas are pigmented). In fact, the appearance of these acral volar lesions is largely different from pigmented skin lesions found in other body areas and accordingly the development of a specially designed classifier is required for these lesions.
Fig. 6 shows sample dermoscopy images from acral volar areas. Expert dermatologists focus on parallel patterns and diagnose this lesion. However, automatic detection of the parallel ridge or parallel furrow patterns is often difficult to achieve due to the wide variety of dermoscopy images (e.g. fibrillar pattern, sometimes looks similar to parallel ridge pattern) and there has been no published methods on computerized classification of this diagnostic category. Recently authors found key features to recognize parallel patterns and developed a classification model for these lesions (Iyatomi et al., 2008a).
In this chapter, we introduce the methodology and results briefly and then discuss them.

(a) nevus (parallel furrow pattern) (b) melanoma (parallel ridge pattern)



(c) nevus (fibrillar pattern – looks like parallel ridge) (d) melanoma (parallel ridge pattern)

Fig. 6. Sample of acral volar pigmented skin lesions.

4.1 Strategy for diagnosis of acral volar lesions

A total of 213 acral volar dermoscopy images; 176 clinically equivocal nevi and 37 melanomas from four Japanese hospitals (Keio University, Toranomon, Shinshu University, and Inagi Hospitals) and two European university hospitals (University of Naples, Italy, University of Graz, Austria) as part of the EDRA-CDROM (Argenziano et al., 2000) were used in our study.
Identification of parallel ridge or parallel furrow patterns is an efficient clue for the diagnosis of acral volar lesions, however as described before, automatic detection of these patterns from dermoscopy image are often difficult. Therefore we did not extract these patterns (structures) directly but instead we constructed a parametric approach as we searched for non-acral lesions, namely by determining the tumor area extracting image features and classifying the image.

In our study,we developed an acral melanoma-nevus classifier and three detectors for typical patterns of acral volar lesions: parallel ridge pattern, parallel furrow pattern and fibrillar pattern. For melanoma-nevus classifier, the training signal of 1 or -1 was assigned to each melanoma and nevus case, respectively. Similarly, a training signal of 1 (positive) or -1 (negative) was assigned to each dermoscopic pattern. The dermoscopic patterns were identified by three experienced dermatologists and only those patterns of which at least two dermatologists agreed were considered. Note here that dermoscopic patterns were assessed independently of each other and therefore some cases received multiple or no assignments. As for a classification model, we used a linear model with the confirmation of whose enough performance for separating malignant tumors from others. The classification performance was evaluated by leave-one-out cross-validation.

4.2 Computer-based diagnosis of acral volar lesions

4.2.1 Determination of tumor area and details of material

The dermatologist-like tumor area extraction algorithm successfully extracted tumor area in 199 cases out of 213 cases (≈ 93.4%). In 14 cases (7 nevi and 7 melanomas), tumor area extraction process failed. This was due to the size of the tumor being larger than about 70% of the dermoscope field. Our algorithm is mainly for early melanomas which usually fit in the frame. Note that most of automated tumor area extraction algorithms meet this difficulty. Tumors in dermoscopy images have a wide variety of colors, shapes and sizes, and accordingly the pre-definition of the characteristics of tumor areas is difficult. Automated algorithms are designed to extract intended areas from the image for most cases with cost of mis-extraction of irregular cases.

Since larger lesions are relatively easy to diagnose, we deem that computer-based screening is not necessary. Also note that the false-extraction rate for melanomas was higher (19%) than that of nevi (4%) and therefore if extraction fails we can consider the lesion as potentially malignant in the first screening step.

Out of 169 nevi, parallel ridge, parallel furrow and fibrillar patterns were found in 5, 133 and 49 cases, respectively. A total of 11 cases of nevi had no specific patterns and, 28 nevi had both parallel furrow pattern and fibrillar pattern. One nevus had both a parallel ridge and a fibrillar pattern. In 30 melanomas, parallel ridge, parallel furrow and fibrillar patterns were found in 24, 2 and 1 cases, respectively. Five of the melanomas had no specific patterns and one of the melanomas had all three patterns.

4.2.2 Developed model

A total of 428 image features were transformed into orthogonal 198 principal components (PCs). From these PCs, we selected the effective ones for each classifier. Table 5 summarizes the number of selected PCs for each classification model (#PC), determination coefficient with adjustment of the degree of freedom R^2, standard deviation of mean estimated error E, the order number of the first 10 PCs lined by the selected sequence by stepwise input-selection method, and the classification performance in terms of SE, SP and AUC under leave-one-out cross-validation test. The SE and SP values shown are those that have the maximum product. The numbers in parentheses represent the performance when 14 unsuccessful extraction cases are considered as false-classification. Even though the number

of the test images was limited, good recognition and classification performance was achieved as well for acral volar pigmented skin lesions.

Classifier type	#PC	R^2	E	Selected PCs (first 10)	SE(%)	SP(%)	AUC
Melanoma	45	.807	.315	**2,9,6,1,3,**15,91,40,20,98	100 (81.1†)	95.9 (92.1†)	0.993
Parallel ridge	40	.736	.363	**2,9,1,6,3,**59,20,88,77,33	93.1	97.7	0.985
Parallel furrow	35	.571	.614	**6,2**,145,15,**3**,98,70,24,59,179	90.4	85.9	0.931
Fibrillar	24	.434	.654	106,66,56,145,137,94,111,169,131,5	88.0	77.9	0.890

† When 14 unsuccessful extraction cases are treated as false-classification.
Table 5. Modeling result and classification performance for acral volar lesions

4.2.3 Important features for recognition of acral lesions

Since we used an orthogonalized image feature set in our analysis, we reached interesting results that compares to the clinical findings of a dermatologist.

For the melanoma-nevus classifier, many significant (small numbered) PCs were found in the first 10 selected features. The parallel ridge and parallel furrow detector were also composed of significant PCs, on the other hand, fibrillar pattern detector showed a different trend. The melanoma classifier and the parallel ridge detector have many common PCs. Particularly, the top five PCs for the two (2nd, 9th, 6th, 1st and 3rd PCs) were completely the same. Note that parameters chosen early in the stepwise feature selection were thought to be more important for the classification because the most significant parameters were statistically selected in each step. The common PCs are mainly related to asymmetry and structural properties rather than color. (See details in original manuscript (Iyatomi et al., 2008a)) The linear classifier using only these five components achieved 0.933 AUC, 93.3% SE, and 91.1% SP using a leave-one out cross validation test. Since the system with a smaller number of the inputs should have high generality in general and a linear model is the most simple architecture, we integrated this 5-input linear classifier on our server.

Dermatologists evaluate parallel patterns using the intensity distribution of the images and they consider the peripheral area of the lesion as important. We confirmed that our computer-based results also focus on similar characteristics as the dermatologists.

5. Open issues in this field

In order to improve system accuracy and generality, there is no doubt that the system should be developed with many samples as much as possible. The number of cases used in any of conventional studies is not enough for practical use at present. On the other hand, even if we can collect a large enough number of images and succeed at finding robust features for diagnosis, the accuracy of the diagnosis cannot reach 100%. In the current format, most of the conventional studies provide only the final diagnosis or diagnosis with limited information. It is desirable that the system provides the grounds for diagnostic results in accordance with quantitatively scored common clinical structures, such as those defined in the ABCD rule, the 7-point checklist, or others. However, since these dermoscopic structures are defined subjectively, their automated quantification is still difficult.

Recent studies on high-level dermoscopic feature extraction include (i) two studies on pigment network (Fleming et al.,1998) and globules (Caputo et al.,2002), (ii) four systematic studies on dots (Yoshino et al., 2004), blotches (Stoecker et al., 2005)(Pellacani et al.,2004), and blue-white areas (Celebi et al.,2008), and (iii) a recent study on parallel-ridge and parallel-furrow patterns (Iyatomi et al.,2008a). Although several researchers attempted to extract these features using image processing techniques, to the best of authors' knowledge, no general solution has been proposed, especially the evaluation of structural features such as pigment networks and streaks have remained an open issue.
We also find that when we widen the target user of an automated diagnostic or screening system from "dermatologist only" to physicians with other expertise or not-medically trained people, the system should have pre-processing schemes to exclude non melanocytic lesions such as basel cell carcinoma (BCC), seborrheic keratosis, and hemangioma. Identification of melanomas from those lesions is in not small cases easier than that from melanocytic lesion (e.g. Clark nevi) by expert dermatologists, but this is also important issue and almost no published results examine this topic.

6. Conclusion

In this chapter, recent investigations in computer-based diagnosis for melanoma are introduced with authors' Internet-based system as an example. Even though recent studies shows good classification accuracy, these systems still have several limitations regarding the acceptable tumor classes, the condition of the input images, etc. Note here again that the diagnostic capability of the present automated systems does not match that of an expert dermatologist. On the other hand, they would be efficient as a diagnosis support system with further improvements and they have the capability to find early stage hidden patients.

7. References

Argenziano, G.; Fabbrocini G, Carli P et al. (1998) Epiluminescence microscopy for the diagnosis of ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis, *Archives of Dermatology*, No. 134, pp. 1536-1570.

Argenziano, G.; Soyer HP, De Giorgi V et al. (2000). *Interactive atlas of dermoscopy CD*: EDRA Medical Publishing and New Media, Milan.

Argenziano, G.; Soyer HP, Chimenti S et al. (2003) Dermoscopy of pigmented skin lesions: Results of a consensus meeting via the Internet, *Journal of American Academy of Dermatology* , Vol. 48, No.5, pp. 679-693.

Blum, A.; Rassner G & Garbe C. (2003) Modified ABC-point list of dermoscopy: A simplified and highly accurate dermoscopic algorithm for the diagnosis of cutaneous melanocytic lesions, *Journal of the Americal Academy of Dermatology*, Vol. 48, No. 5, pp. 672-678.

Blum, A.; Luedtke H, Ellwanger U et al. (2004) Digital image analysis for diagnosis of cutaneous melanoma. Development of a highly effective computer algorithm based on analysis of 837 melanocytic lesions, *British Journal of Dermatology*, Vol. 151, pp. 1029-1038.

Burroni, M.; Sbano P, Cevenini G et al. (2005) Dysplastic naevus vs. in situ melanoma: digital dermoscopy analysis, *British Journal of Dermatology*, Vol. 152, pp. 679-684.

Caputo, B.; Panichelli V, Gigante GE. (2002) Toward a quantitative analysis of skin lesion images, *Studies in Health Technology and Informatics*, Vol. 90, pp. 509-513.

Celebi, ME.; Aslandogan YA, Stoecker WV et al. (2007a) Unsupervised border detection in dermoscopy images, *Skin Research and Technology*, Vol. 13, pp. 1-9.

Celebi, ME.; Kingravi HA, Uddin B et al. (2007b) A methodological approach to the classification of dermoscopy images, *Computerized Medical Imaging & Graphics*, Vol. 31, No. 6, pp. 362-373.

Celebi, ME.; Iyatomi H, Stoecker WV et al. (2008) Automatic Detection of Blue-White Veil and Related Structures in Dermoscopy Images, *Computerized Medical Imaging and Graphics*, Vol. 32, No. 8, pp. 670-677.

Celebi, ME.; Iyatomi H & Gerald S. (2009) Lesion border detection in dermoscopy images, *Computerized Medical Imaging and Graphics*, Vol. 33, No. 2, pp. 148-153.

Elbaum, M.; Kopf AW, Rabinovitz HS et al. (2001) Automatic differentiation of melanoma from melanocytic nevi with multispectral digital dermoscopy: a feasibility study, *Journal of American Academy of Dermatology* , Vol. 44, pp. 207-218.

Fleming, MG.; Steger C, Zhang J et al. (1998) Techniques for a structural analysis of dermatoscopic imagery, *Computerized Medical Imaging and Graphics*, Vol. 22, No. 5, pp. 375-389.

Ganster, H.; Pinz A, Rohrer R et al. (2001) Automated melanoma recognition, *IEEE Trans. on Medical Imaging*, Vol. 20, No. 3, pp. 233-239.

Grana, C.; Pellacani G, Cucchiara R et al. (2003) A new algorithm for border description of polarized light surface micriscopic images of pigmented skin lesions, *IEEE Trans. on Medical Imaging*, Vol 22, No. 8, pp. 959-964.

Green, A.; Martin N, McKenzle G et al. (1991) Computer image analysis of pigmented skin lesions, Melanoma Research, Vol. 1, pp. 231- 236.

Hoffmann, K.; Gambichler T, Rick A et al. (2003) Diagnostic and neural analysis of skin cancer (DANAOS). A multicentre stidy for collection and computer-aided analysis of data from pigmented skin lesions using digital dermoscopy. *British Journal of Dermatology*, Vol. 149, pp. 801-809.

Iyatomi, H.; Oka H, Saito M et al. (2006) Quantitative assessment of tumour area extraction from dermoscopy images and evaluation of the computer-based methods for automatic melanoma diagnostic system, *Melanoma Research*, Vol. 16, No. 2, pp. 183-190.

Iyatomi, H.; Oka H, Celebi ME et al. (2008a) Computer-Based Classification of Dermoscopy Images of Melanocytic Lesions on Acral Volar Skin, *Journal of Investigative Dermatology*, Vol. 128, pp. 2049-2054.

Iyatomi, H.; Oka H, Celebi ME et al.(2008b) An improved Internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm, *Computerized Medical Imaging and Graphics*, Vol. 32, No. 7, pp. 566-579.

Jemal, A.; Siegel R, Ward E et al. (2008) Cancer Statistics, *A Cancer Journal for Clinicians*, Vol. 58, No. 2, pp. 71-96.

Mayer, J.; (1997) Systematic review of the diagnostic accuracy of dermoscopy in detecting malignant melanoma, *Med. Journal of Australia*, Vol. 167, No. 4, pp. 206-210.

Menzies, SW.; Bischof L, Talbot H, et al. (2005) The performance of SolarScan - An automated dermoscopy image analysis instrument for the diagnosis of primary melanoma, *Archives of Dermatology*, Vol. 141, No. 11, pp. 1388-1396.

Meyskens, FL Jr.; Berdeaux DH, Parks B et al. (1998). Natural history and prognostic factors incluencing survival in patients with stage I disease, *Cancer*, Vol. 62, No. 6, pp. 1207-1214.

Oka, H.; Hashimoto M, Iyatomi H et al. (2004) Internet-based program for automatic discrimination of dermoscopic images between melanoma and Clark nevi, *British Journal of Dermatology*, Vol. 150, No. 5, p. 1041.

Otsu, N. (1998) An automatic threshold selection method based on discriminant and least square criteria, *Trans. of IEICE*, Vol. 63, pp. 349-356.

Pellacani, G.; Grana C, Cucchiara R et al. (2004) Automated extraction and description of dark areas in surface microscopy melanocytic lesion images, *Dermatology*, Vol. 208, No. 1, pp. 21-26.

Reed, R. (1993) Pruning algorithms - a survey, *IEEE Trans. on Neural Networks*, Vol. 4, No. 5, pp. 740-747.

Rubegn ,P.; Cevenini G, Burroni M et al. (2002) Automated diagnosis of pigmented skin lesions, *International Journal of Cancer* , Vol. 101, pp. 576-580.

Saida, T.; Miyazaki A, Oguchi S et al. (2004) Significance of dermoscopic patterns in detecting malignant melanoma on acral volar skin, Arch Dermatol, Vol. 140, pp. 1233-1238.

Seidenari, S; Pellacani G & Grana C. (2005) Pigment distribution in melanocytic lesion images: a digital parameter to be employed for computer-aided diagnosis, *Skin Research and Technology*, Vol. 11, pp. 236-241.

Stoecker, WV.; Gupta K, Stanley RJ et al. (2005) Detection of asymmetric blotches (asymmetric structureless areas) in dermoscopy images of malignant melanoma using relative color, *Skin Research and Technology*, Vol. 11, No. 3, pp. 179-184.

Stolz, W.; Riemann A, Cognetta AB et al. (1994) ABCD rule of dermatoscopy: a new practical method for early recognition of malignant melanoma, *European Journal of Dermatology*, Vol. 4, No. 7, pp. 521-527.

Stolz W.; Falco OB., Bliek P et al. (2002). *Color Atlas of Dermatoscopy -- 2nd enlarged and completely revised edition*, Berlin: Blackwell publishing, ISBN: 978-1-4051-0098-4, Berlin.

Soyer, HP.; Smolle J, Kerl H et al. (1987) Early diagnosis of malignant melanoma by surface microscopy, *Lancet*, No. 2, p. 803.

Soyer, HP.; Argenziano G, Zalaudek I et al. (2004) Three-Point Checklist of Dermoscopy: A New Screening Method for Early Detection of Melanoma, *Dermatology*, Vol. 208, pp. 27-31.

Tanaka M. (2006) Dermoscopy, *Journal of Dermatology*, Vol. 3, pp. 513-517.

Yoshino, S.; Tanaka T, Tanaka M et al. (2004) Application of morphology for detection of dots in tumor, *Procs. SICE Annual Conference*, Vol. 1, pp. 591-594.

New Developments in Biomedical Engineering
Edited by Domenico Campolo

ISBN 978-953-7619-57-2
Hard cover, 714 pages
Publisher InTech
Published online 01, January, 2010
Published in print edition January, 2010

Biomedical Engineering is a highly interdisciplinary and well established discipline spanning across engineering, medicine and biology. A single definition of Biomedical Engineering is hardly unanimously accepted but it is often easier to identify what activities are included in it. This volume collects works on recent advances in Biomedical Engineering and provides a bird-view on a very broad field, ranging from purely theoretical frameworks to clinical applications and from diagnosis to treatment.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen