DEVELOPMENT OF A DERMATOLOGICAL WORKSTATION: PRELIMINARY RESULTS ON LESION SEGMENTATION IN CIE L*A*B* COLOR SPACE

Y. Vander Haeghen[1,2], J.M. Naeyaert[2] and I. Lemahieu[1]

[1]Dept. Engineering, ELIS-MEDISIP, University Gent,
St.-Pietersnieuwstraat 41, 9000 Gent, Belgium,
Yves.VanderHaeghen@rug.ac.be
[2]Dept. Dermatology, University Hospital, De Pintelaan 185,
9000 Gent, Belgium

ABSTRACT

In the past we have proposed a pseudo-colorimetric imaging system for dermatological uses. The system exhibits moderate accuracy, but high precision or reproducibility. Acquired images are stored in the well-defined gamma-corrected sRGB color space. This color space has a known relationship to the perceptually uniform CIE $L^*a^*b^*$ color space, which allows the development of segmentation methods based on perceptual color differences. This should give segmentation results which are more in agreement with a human observer than if performed in uncalibrated RGB or other tristimulus color spaces. To demonstrate this a very simple semi-automatic segmentation method is proposed and tested, and it is hoped that more sophisticated methods, e.g. watersheds and clustering, may equally benefit from a similar colorimetric approach.

The basic premise of the proposed segmentation method is that a skin lesion is surrounded by a band of 'normal' skin. The colors of this band are determined when the user draws a rough outer contour of the lesion, and are used to shrink the contour if the local color is similar to the local 'normal' skin color. As in general there is no consensus between dermatologists about the lesion contour, testing the method was primarily concerned with the reproducibility of the segmentation method versus that of a manually drawn contour. This reproducibility was assessed using two possibly diagnostically relevant parameters computed from the contours: the area in pixels, and the average $L^*a^*b^*$ color inside the contour. The test was conducted using a set of 30 distinct images, each used twice for a total of 60 segmented images per observer. Each segmented image was also scored by the observer. Four observers took part in the tests, and general satisfaction with the method was good, although the presence of hairs still posed some problems. Reproducibility was not really improved compared to manual segmentation, as long as one and the same observer was concerned. However, with multiple observers working on the same images the results of the algorithm improved, mainly due to the poor reproducibility of manual contours between observers. Lastly, the method is very fast and needs only minor interaction to operate properly.

Keywords: *Dermatology, Color segmentation, Colorimetry, $CIEL^*a^*b^*$*

1. INTRODUCTION

In dermatology, color and color difference as well as shape often convey important diagnostic information. Although visual inspection by the dermatologist goes a long way, quantitative color and shape measurements might be helpful when investigating pigmented lesions, especially skin cancer where early diagnosis is crucial but quite difficult. In order to perform such measurements a reproducible method for segmenting the lesion under study must first be developed.

Images of skin lesions show surprising variability in shape and color. Together with the presence of hairs this makes the robust detection of a lesion border very difficult without some human intervention. Several papers have been published on this subject, e.g. [1, 2, 3, 4, 5, 6], claiming good segmentation results. These publications present many different methods: radial search starting from a central point, principal component transform followed by histogram based techniques, adaptive thresholding, fuzzy C-means clustering, watershed segmentation followed by a region classifier, etc. .... Some of the presented methods also made use of the CIE $L^*a^*b^*$ color space, but it was unclear how proper CIE $L^*a^*b^*$ color triplets could be obtained if the RGB primaries of the acquisition color space are un-

This work was supported by the Belgian FWO, ref. FW7801

known. Anyway, here we do not aspire to solve the segmentation problem of skin lesions, but rather to demonstrate the potential of a more 'colorimetric' approach to image processing using a very simple segmentation algorithm. Some of the obvious advantages of this approach are a wider applicability and the possibility of cross-validation of any developed segmentation method. Such an approach, however, assumes precise knowledge of the nature of the image data, e.g. the primaries and white point of the color space in which it is defined.

2. THE IMAGING SYSTEM

To address the many problems of imaging systems in providing reproducible visual records and allowing precise quantitative color and geometric measurements, we have recently proposed a pseudo-colorimetric imaging system [7, 8, 9]. It consists of a 3-chip CCD camera controlled through the serial port of the PC, a zoom lens, a 150 Watt halogen light source connected to a continuous annular diffuser with an optical fiber, and a frame grabber, see fig. 1. The field of view of the CCD camera is 1.6 cm by 1.2 cm. With an image containing 760 by 570 pixels the resolution is thus 47.5 pixels/mm. A blue filter changes the color temperature of the light source from 2800 K to around 6500 K, making it a rough approximation for the CIE D65 illuminant. The camera output uses the PAL analog RGB format, which is digitized using 8-bit precision per color channel. The camera gamma-correction circuitry is on in order to minimize quantization errors, and this operation is inversed in software on the host PC.

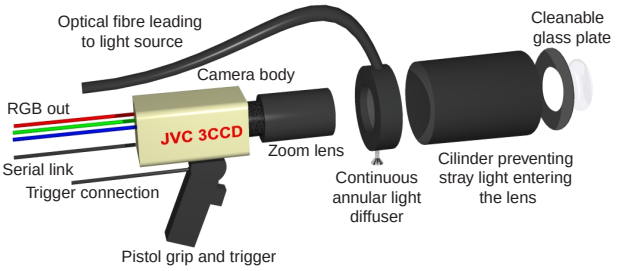

Figure 1: Schematical drawing of the camera of the imaging system. All components are drawn approximately to scale.

The imaging system exhibits moderate accuracy ($\langle \Delta E^*_{ab} \rangle = 6.2$ and $\Delta E^*_{ab} < 13.3$), but high precision or reproducibility ($\langle \Delta E^*_{ab} \rangle = 0.3$ and $\Delta E^*_{ab} < 1.2$). Resulting images are defined in the standard $sR^*G^*B^*$ color space [10], which has a D65 white point. This allows immediate realistic viewing on any modern CRT based computer monitor set to a color temperature of 6500K, as well as meaningful exchange of images, e.g. for telemedicine purposes. The $sRGB$ space has a known relationship to the CIE XYZ and CIE $L^*a^*b^*$ color spaces, so that perceptual color differences ΔE^*_{ab} can be computed. The relatively restricted gamut of $sRGB$ shouldn't be a problem for skin imaging as very vivid colors will rarely occur.

3. SEGMENTATION

The idea of the proposed semi-automatic segmentation method is to use the dermatologist for what he is very good at, namely the detection of the rough lesion border, but to relieve him of the laborious process of determining a detailed lesion contour. Although this may introduce some unwanted user-dependency, it is believed the benefits will outweigh this.

3.1. The contour

The contour consists of a set of vertices spaced by a predetermined distance called the vertex resolution. The vertices can move radially inward or outward based on image properties in one or more areas around the vertex. If during such a move the distance between two vertices becomes less than half the vertex resolution that vertex is removed. If, on the other hand, the distance between two vertices becomes larger than 2 times the vertex resolution, a new vertex is inserted in the middle, see fig. (2). This resample scheme is based on the one in [11].

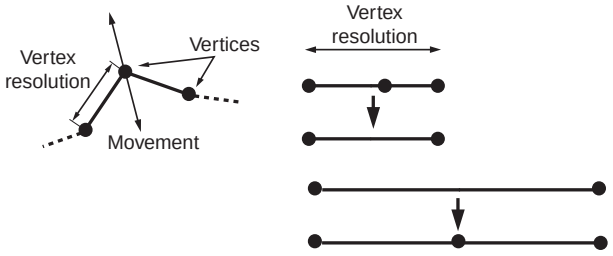

Figure 2: Contour consisting of vertices, and resample scheme.

3.2. The segmentation algorithm

The basic premise of the proposed method is that a skin lesion is surrounded by a band of 'normal' skin. To start the algorithm the user draws a rough outer contour of the lesion, making sure the vertices mark representative areas of normal skin. Next, the average color is computed around each vertex to obtain a local measure of the normal skin color. Hereafter, the contour is resampled based on the vertex resolution. Notice that if during resampling a vertex is added its local skin color will become the average of the local skin colors of the two neighboring vertices. This is opposed to first resampling the contour and then computing the local normal skin color, which denies the

user the choice of some of the reference areas for normal skin and poses problems, e.g. when a skin lesion touches the border of the image. These local normal skin colors are then used to conditionally shrink the vertices radially inwards, based on their color difference with the current local color average. As long as this difference is within a predefined shrinking tolerance in ΔE^*_{ab} units, shrinking goes on, until all vertices come to a standstill. Recapitulating, we get:

1. Draw a rough outer contour of the skin lesion, marking the representative areas of normal skin;
2. Compute and store the average color around each vertex as a measure of the local normal skin color;
3. Resample the contour according to the vertex resolution;
4. Compute and store the color difference in ΔE^*_{ab} units between the average color around each vertex and the corresponding local normal skin color;
5. Move the vertex radially inwards if the vertex color difference is below a certain threshold called the shrinking tolerance;
6. Stop if no vertex has moved, else repeat from step 3.

The only two parameters of this algorithm are the vertex resolution which is set to 12 pixels, and the shrinking tolerance which is set to 8 ΔE^*_{ab} units. The averaging areas around the vertices are squares the size of the vertex resolution. More complex shapes, e.g. based on the local shape of the contour, could easily be implemented but would conceivably be slower.

3.3. Testing the segmentation

As in general there is no consensus between dermatologists about a lesion contour, i.e. there is no golden standard, testing the segmentation algorithm is primarily concerned with its reproducibility versus that of a manually drawn contour.

In order to test the segmentation method a test set of 30 images of skin lesions is put to 4 dermatologists, called observers from now on. Each image is presented twice in a random sequence, and is segmented both manually and using the proposed semi-automatic segmentation method. The difficulty of manually segmenting and the quality of the semi-automatic contour are both scored, see table 1. Two features which conceivably play a role when diagnosing the lesion are computed from both the manual and the semi-automatic contours: the area inside the contour in pixels, and the average $CIEL^*a^*b^*$ value inside the contour. For segmentation method $i = 1(\text{manual}), 2(\text{semi-automatic})$, measurement $l = 1, 2$, image $j = 1, \ldots, 30$ and observer $k = 1, \ldots, 4$ these features are respectively notated as A_{ijkl} and $(L^*_{ijkl}, a^*_{ijkl}, b^*_{ijkl})$.

Score	Difficulty	Contour
1	very hard, uncertain where the border is	bad, completely wrong
2	hard, contains parts where border is uncertain	poor, many errors
3	average, diffuse but defined	average, some errors
4	easy, partly crisp and partly diffuse but defined	good, a few errors
5	very easy, crisp border	perfect

Table 1: Subjective scores for difficulty of segmentation and contour quality of the result of the semi-automatic segmentation.

In order to asses the reproducibility of a method i for the contour area A, the standard deviation s_{ijk} could be used, computed over both measurements l, and this for each image j and observer k:

$$\begin{aligned} \langle A_{ijk} \rangle =& \frac{\sum_{l=1}^{2} A_{ijkl}}{2}, \\ s_{ijk} =& \sqrt{\frac{\sum_{l=1}^{2} (A_{ijkl} - \langle A_{ijk} \rangle)^2}{2}}. \end{aligned} \tag{1}$$

However, due to the great variety in lesion areas it can be argued that this makes comparison of the reproducibility measure *between* images impossible, and we therefore propose to divide the standard deviation by the average area to obtain a better measure of reproducibility r_{ijk} for method i, image j and observer k:

$$r_{ijk} = \frac{s_{ijk}}{\langle A_{ijk} \rangle}. \tag{2}$$

Note that the sample standard deviation was divided by 2 and not by 1 as usual for an unbiased estimator for the population standard deviation.

For the average $L^*a^*b^*$ color of a contour we propose following formulas to obtain a measure of reproducibility for method i, image j and observer k:

$$\begin{aligned} (\langle L^*_{ijk} \rangle, \langle a^*_{ijk} \rangle, \langle b^*_{ijk} \rangle) =& \frac{\sum_{l=1}^{2} (L^*_{ijkl}, a^*_{ijkl}, b^*_{ijkl})}{2}, \\ \Delta E^*_{ab,ijkl} =& \Big((L^*_{ijkl} - \langle L^*_{ijk} \rangle)^2 + (a^*_{ijkl} - \langle a^*_{ijk} \rangle)^2 \\ & + (b^*_{ijkl} - \langle b^*_{ijk} \rangle)^2 \Big)^{1/2}, \\ \langle \Delta E^*_{ab,ijk} \rangle =& \sum_{l=1}^{2} \frac{\Delta E^*_{ab,ijkl}}{2}. \end{aligned} \tag{3}$$

Eqs. (2) and (3) can be used to compare the reproducibility of both methods *for one observer performing the segmentation on one image*. These can then be compared over the set of observers and images, called cases from here on, in order to remove some of the variability introduced by those two factors outside our experimental control. It is assumed that these reproducibility measures are also representative for one observer segmenting and measuring *different* images of the *same* skin lesion, e.g. when performing a lesion follow-up over time. Within this context we can speak of an intra-observer reproducibility. In reality it will not always be the same person who segments and measures a skin lesion, and so we also need to determine the total reproducibility of the manual and semi-automatic segmentation, which includes a part due to inter-observer variations. For this eqs. (2) and (3) are extended by computing the feature averages per segmentation method i and image j, over all the measurements l and observers k:

$$\begin{aligned}\langle A_{ij}\rangle =& \frac{\sum_{k=1}^{4}\sum_{l=1}^{2} A_{ijkl}}{8},\\ s_{ij} =& \sqrt{\frac{\sum_{k=1}^{4}\sum_{l=1}^{2}(A_{ijkl}-\langle A_{ij}\rangle)^2}{8}},\\ r_{ij} =& \frac{s_{ij}}{\langle A_{ij}\rangle},\end{aligned} \tag{4}$$

and

$$\begin{aligned}(\langle L^*_{ij}\rangle, \langle a^*_{ij}\rangle, \langle b^*_{ij}\rangle) =& \frac{\sum_{k=1}^{4}\sum_{l=1}^{2}(L^*_{ijkl}, a^*_{ijkl}, b^*_{ijkl})}{8},\\ \Delta E^*_{ab,ijkl} =& \Big((L^*_{ijkl}-\langle L^*_{ij}\rangle)^2 + (a^*_{ijkl}-\langle a^*_{ij}\rangle)^2\\ &+ (b^*_{ijkl}-\langle b^*_{ij}\rangle)^2\Big)^{1/2},\\ \langle \Delta E^*_{ab,ij}\rangle =& \sum_{k=1}^{4}\sum_{l=1}^{2}\frac{\Delta E^*_{ab,ijkl}}{8}.\end{aligned} \tag{5}$$

4. RESULTS AND DISCUSSION

Although testing the segmentation algorithm was primarily concerned with the reproducibility of the segmentation method versus that of a manually drawn contour, the histograms of the scored values can be seen in fig. (3). These are very subjective measures, but the low number of poorly segmented images versus the number of difficult to segment images is nevertheless a clue as to the good performance of the algorithm. See fig. 4 for an example of a though lesion.

A quick look at fig. 5 shows us that the intra-observer reproducibilities contain some outliers and is far from being normally distributed. However, with 120 cases and by virtue of the central limit theorem of statistics we may safely state that the average reproducibility measure *per case* is normally distributed. In that case we may use a t-test to test if the two

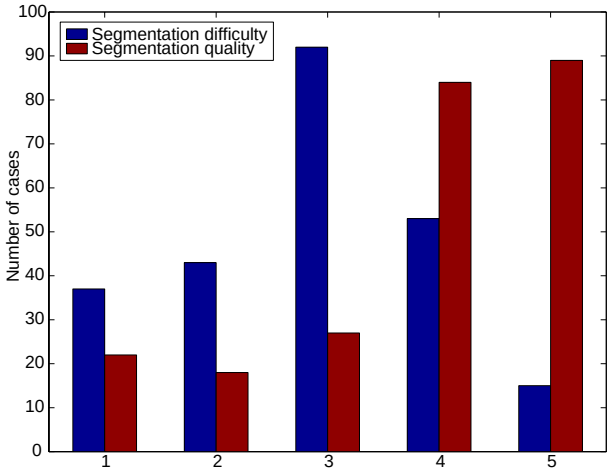

Figure 3: The subjective scores of the segmentation. See table 1 for an explanation of the labels.

methods have a different mean. An alternative which doesn't require normality of the underlying distributions is the non-parametric Wilcoxon matched pairs signed-ranks test [12], which is almost as powerful as the t-test. This test relies on the ability to rank the differences in reproducibility for each case according to their absolute value. If the assumption that the reproducibility measures can be compared across cases is not valid, we must fall back on the much less powerful matched pairs sign test [12], which does not require ranking of the differences in reproducibility over the cases.

The results of the t-test can be found in table 2, and as could easily be deduced from the histograms in fig. 5, no significant improvement was found for the semi-automatic segmentation method.

Feature	Avg. reprod.		Significant
	man.	semi-aut.	difference?
Area	0.2613	0.2455	No
Color	3.1774	3.0603	No

Table 2: Average intra-observer reproducibility for the manual and semi-automatic segmentation per case, and statistical significance of the t-test at the $p = 0.05$ level. Results are obtained by averaging over 120 cases (30 images and 4 observers).

Fig. 6 shows the histograms of the total reproducibilities r_{ij} and $\langle \Delta E^*_{ab,ij}\rangle$. There are some indications of differences in distributions, but the overall picture is still unclear due to the low number of samples. To be sure we will use the weakest statistical test, the matched pairs sign test, to demonstrate a significant improvement for the semi-automatic versus manual segmentation. This test poses minimal requirements on the data, and its results can be seen in table 3. This time a statistically significant improve-

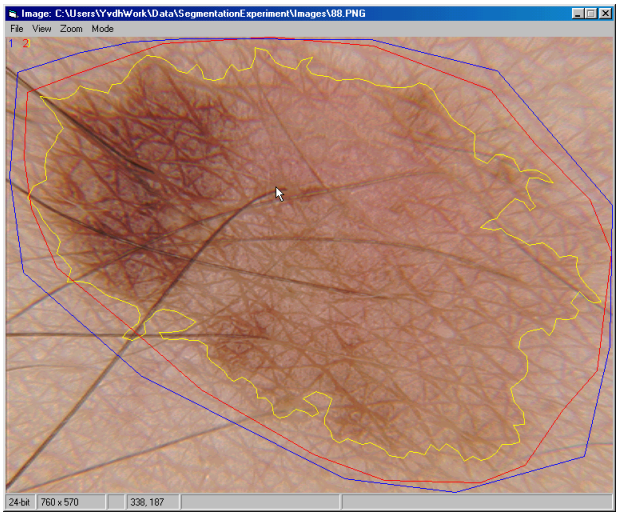

Figure 4: An example of the manual (red) and semi-automatic (yellow) segmentation of a difficult lesion (average difficulty score of 1.625). The rough start contour for the semi-automatic segmentation is in blue. Notice the problems with the dark hairs.

ment was found for both image features, although in absolute value the reproducibilities can hardly be called exceptional.

Feat.	Avg. reprod.		Nr. of improved images	Sign. diff. ?
	man.	semi-aut.		
Area	0.42	0.34	24/30	Yes
Color	7.6	6.3	26/30	Yes

Table 3: Average total reproducibility for the manual and semi-automatic segmentation per image, and statistical significance of the matched pairs sign-test at the $p = 0.05$ level. Results are obtained by averaging over 30 images.

Inspection of some of the contours, e.g. as in fig. 4, showed that the main problem for proper segmentation was the presence of hairs. This could possibly be remedied by giving the user the opportunity to interactively modify the resulting contour if deemed necessary. This was tried, but turned out to be laborious due to the high contour resolution which meant that many points had to be moved. A possible solution is the use of so-called 'rational Gaussian' curves [13] or similar method, which allow easy interactive repositioning of parts of contours. The segmentation algorithm also still exhibits a reproducibility which may be too low for practical purposes, but this could possibly be improved by using information from the first segmented image of a lesion to guide segmentation of subsequent images of that same lesion.

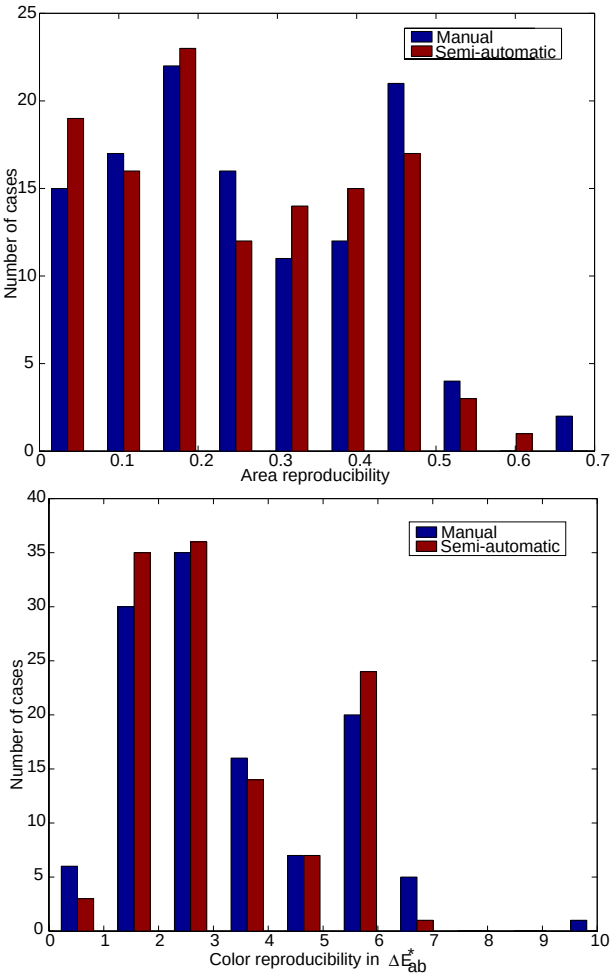

Figure 5: The histograms of the intra-observer reproducibility measures r_{ijk} (top) and $\langle \Delta E^*_{ab,ijk} \rangle$ (bottom) for all cases.

5. CONCLUSION

A simple segmentation algorithm using concepts from the domain of colorimetry was presented. Subjectively, the method gave satisfying results. Objectively it was demonstrated that it could be used to iron out some of the inter-observer variability when compared to manual segmentation. The algorithm is not meant to be a definitive solution to skin lesion segmentation, but rather demonstrates the importance of proper knowledge of the image data to be processed and the potential for a more colorimetric approach to image processing.

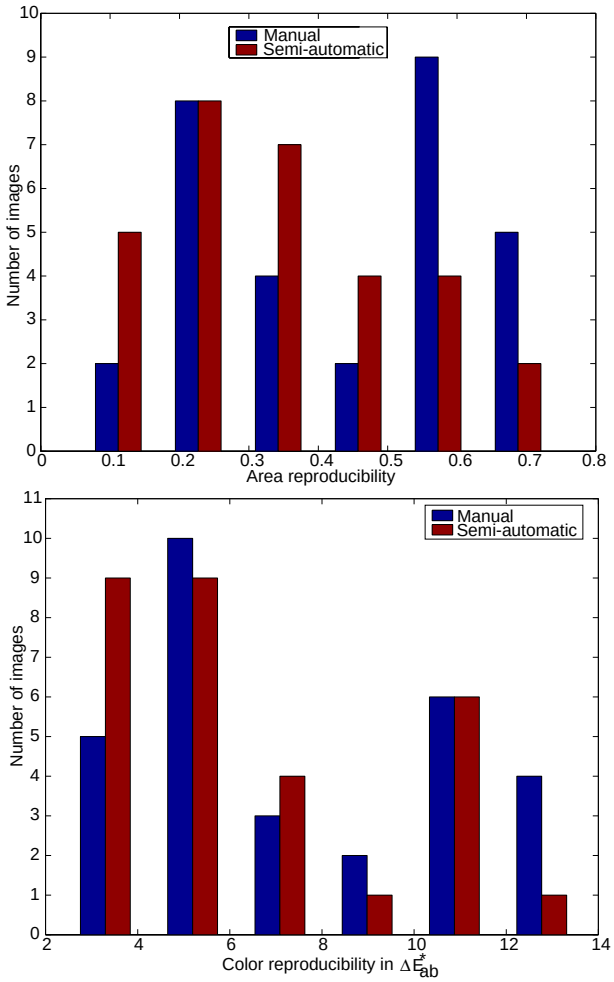

Figure 6: The histograms of the total reproducibility measures r_{ij} (top) and $\langle \Delta E^*_{ab,ij} \rangle$ (bottom) for all images.

6. REFERENCES

[1] J.E. Golston et al. Boundary detection in skin tumor images: an overall approach and a radial search algorithm. *Pattern Recognition*, 23(11):1235–1247, 1990.

[2] Scott E. Umbaugh. *Computer vision in medicine: Color metrics and image segmentation methods for skin cancer diagnosis.* Ph.D., University of Missouri - Rolla, 1990.

[3] Scott E. Umbaugh et al. An automatic color segmentation algorithm with application to identification of skin tumor borders. *Computerized Medical Imaging and Graphics*, 16(3):227–235, May-June 1992.

[4] Scott E. Umbaugh et al. Automatic color segmentation algorithms. *IEEE Engineering in Medicine and Biology*, pages 75–82, September 1993.

[5] Tim Lee et al. Describing the shape of melanocytic lesions. In *SPIE 99 Medical Imaging*. SPIE, 1999.

[6] Olivier Debeir et al. Pigmented skin lesion cartography by supervised image segmentation: pixel vs region classification. In *Proceedings In memoriam Pierre Devijver*, pages 102–107, 1999.

[7] Y. Vander Haeghen, J.M. Naeyaert, and I. Lemahieu. Consistent digital color image acquisition of the skin. In *20th Annual international conference of the IEEE Engineering in Medicine and Biology, Hong-Kong*, volume 20, pages 944–949. IEEE, 1998.

[8] Y. Vander Haeghen, J.M. Naeyaert, and I. Lemahieu. Towards consistent color image acquisition in dermatology. In *Internation congress on Imaging Science, Antwerp, Belgium*, pages 377–381. ICPS, 1998.

[9] Y. Vander Haeghen, J.M. Naeyaert, I. Lemahieu, and W. Philips. An imaging system with calibrated color image acquisition for use in dermatology. *IEEE Transcations on Medical Imaging*, 2000. To be published (Accepted with some minor revisions).

[10] IEC. Colour management in Multimedia systems - Part2: Colour Management, Part2.1: Default RGB colour space-sRGB. Technical report, International Electrotechnical Commision, 1998.

[11] Steven Lobregt and Max A. Viergever. A discrete dynamic contour model. *IEEE Transactions on Medical Imaging*, 14(1):12–24, march 1995.

[12] Amin D. Aczel. *Complete business statistics.* Irwin, 2nd edition, 1992.

[13] M Jackowski, A. Goshtasby, and M. Satter. Interactive tools for image segmentation. 1999.