

PRIP-TR-94

May 22, 2005

Segmentation and Surveying of Cutaneous Hemangiomas

Sebastian Zambanini

Abstract

This report presents an automatic method for the surveying of cutaneous hemangiomas by means of a fully automatic hemangioma segmentation and a ruler visible in the images to be examined. The segmentation of regions belonging to the hemangioma is accomplished with a pre-segmentation based on a perceptron followed by a postprocessing step where regions with untypical properties are rejected. The spatial resolution of the images is determined by computing the maximum euclidean distance between two marks of the ruler. Various experiments with the images gathered are reported, showing not only the satisfactory results on most of the images but also the problems arising on particular images.

1 Introduction

Cutaneous hemangiomas are benign tumors made up of newly-formed blood vessels in the skin. They occur in about 5-10 percent of one year olds and can grow very large if not treated early. Although they are not life-threatening, they may impinge on vital structures and interfere with breathing, vision, eating, or hearing and, especially if occurring in the face, can also cause psychological problems [2].

For the investigation of various methods of treatment it is necessary to determine the area which the hemangioma covers to find out, how much it has increased or decreased since the last examination. In order to make the evaluation of various methods easier, faster, more comfortable and in most cases more precise than a manual segmentation and surveying of the affected skin area, an automated method is highly desirable.

This technical report presents an automated method for surveying cutaneous hemangiomas that were photographed along with a ruler to determine the scale of the image. The problem can be divided in two main tasks:

1. Determining the scale of the images by means of the ruler visible in the image.
2. Segmentation of the skin area belonging to the hemangioma.

The images used in this project have been kindly provided by the *Department of Dermatology* at the *Vienna General Hospital* which is making a long-range study to compare different methods of treatment for cutaneous hemangiomas. The photos were taken using an analog photo camera and digitalized with a scanner. All images have a resolution of 512x768 pixels and a bit depth of 8 bits per color channel.

Because of the circumstance that nearly all patients of the study are infants who normally do not keep still when being photographed, and the bad illumination in the examination room, the quality of the images is partially quite poor. For instance sometimes parts of the hemangioma are located in dark areas (shadows), because they are not sufficiently illuminated by the flash of the camera, or the sharpness of the whole image is very low.

In order to increase reliability there are always taken two or more images of the hemangioma, i.e. there are at least two images of a specific hemangioma at a specific date. All images are labeled with a five-digit code to uniquely identify them.

Fig. 1 shows five different images of hemangiomas describing also their miscellaneous appearances. Fig. 1(a) shows a hemangioma with the usual strawberry-red color and a rough surface. Fig. 1(b) shows a more nodule-shaped hemangioma with a flat surface. In Fig. 1(c) there can be seen a hemangioma with regressing, slight reddish parts. Fig. 1(d) and (e) are good examples for images with poor quality. In Fig. 1(d) parts of the hemangioma are situated in an inadequately illuminated region and Fig. 1(e) is generally of low sharpness.

The report is organized as follows: Section 2 explains in detail the algorithm used for determining the scale of the images. In Section 3.1 first various segmentation algorithms are evaluated for our purpose, followed by a presentation of the complete segmentation process used for this project, divided into *Preprocessing* (3.2), *Classification* (3.3) and *Postprocessing* (3.4). In Section 4 several experiments performed on the data gathered with the algorithm proposed are presented and discussed. A Conclusion is finally given in Section 5.

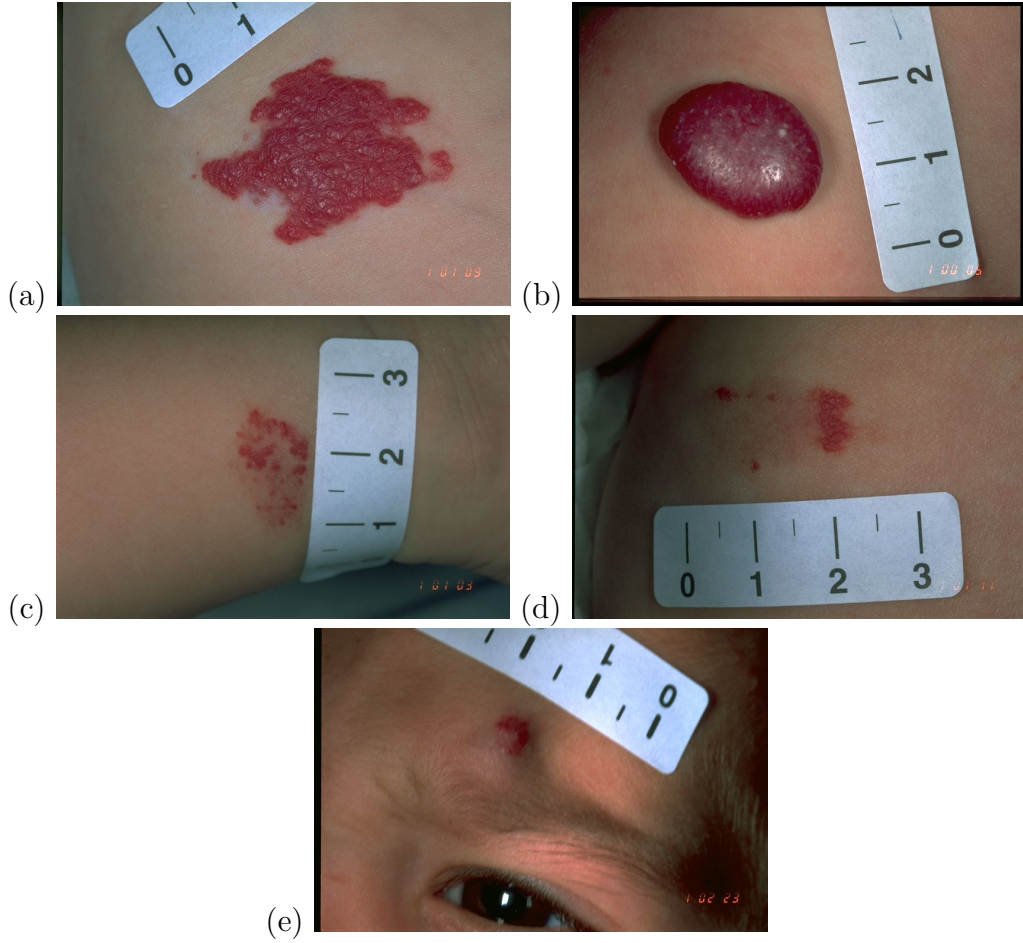


Figure 1: Five images of hemangiomas

2 Computing the Scale of the Images

All images of the study show a ruler beside the hemangioma. The ruler has 4 bold lines in 1 cm distance steps. Therefore the task of the algorithm is to compute the euclidean distance between two lines to get the spatial resolution of the images. The area of the hemangioma is then simply calculated by multiplying the number of hemangioma pixels with the area of one pixel. We neglect the error due to the fact that normally both hemangioma and ruler are not situated on a planar surface parallel to the image plane. The main steps of the algorithm are:

1. Segment the ruler
2. Calculate the ruler orientation and rotate it in a horizontal position
3. Get the number of pixels between two lines by scanlining through the ruler

1. Segmenting the ruler: First we have to segment the ruler. Since all rulers are white and differ extremely from the rest of the image, this can simply be done by global thresholding with the H and V channel of the HSV color model. Sometimes small regions not belonging to the ruler can also remain by this operation, so we take only the largest region in the computed mask [Fig. 2(a) and 2(b)] .

2. Rotation in horizontal position: In the next step we determine the orientation of the region describing the ruler. For this purpose we calculate the angle between the x-axis and the major axis of the ellipse that has the same second-moments as the region (the ellipse with the same orientation as the ruler). Once we have this angle, we can rotate the ruler in a horizontal position [Fig. 2(c)].

3. Scanlining: For robustness we use three scanlines to determine the number of pixels between two marks (a scanline is a vector containing the pixel values of a specific line (y-value) of an image from left to right). The first scanline is on the midpoint between the top and bottom y-value of the ruler. The two others are 10 pixels above and below, respectively [Fig. 2(d)]. To determine the scale we take the maximum number of pixels between two marks in all of the three scanlines. With this method we use the part of the ruler which lies most normal to the camera and has the smallest curvature and therefore has to be the most precise measurement. Occasionally it can happen that one or more marks are not recognized in the scanline (e.g. when the ruler has a strong curvature), hence too large distances with more than 200 pixels between two marks are rejected (in all 122 images appropriated for us no greater distance than 155 pixels could be found). If n denotes the number of pixels coming to one cm, the area of one pixel in the image is $(1/n)^2$. The example given in Fig. 2 results in a maximum distance of 138 pixels between two marks, i.e. one pixel of the image has an area of $\sim 0.0000525 \text{ cm}^2$.

3 Segmentation of the Hemangiomas

The task of the segmentation algorithm is to find the regions in an image belonging to the hemangioma. In the next Section we first give an overview about several existing methods proposed in publications dealing with the segmentation of skin lesions and explain the problems with them for our purpose and why we finally decided to use a perceptron for segmentation. After that our segmentation method is described in detail, divided into a preprocessing step (Section 3.2), a classification step (Section 3.3) and a postprocessing step (Section 3.4).

3.1 Evaluation of Existing Skin Lesion Segmentation Methods for Hemangioma Segmentation

Thresholding: A thresholding operation is often used for the segmentation of skin lesions [3], [4]. However, finding reliable parameter values for thresholding operations without user interaction working well with all of our images can not be done. This is mainly

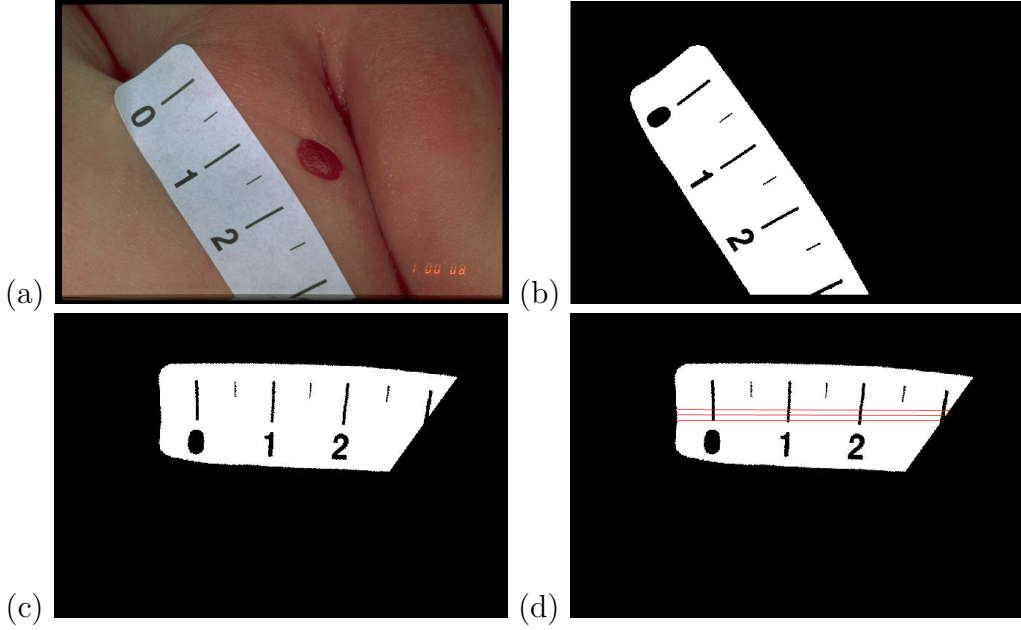


Figure 2: Example for determining the scale of an image

caused by the low contrast between skin and hemangioma regions, since hemangiomas mainly appear in bright red [13].

Intensity image thresholding: A more interesting method for our purpose is given in [14]. Here an image showing a melanoma is first transformed into an intensity image where the intensity at a pixel shows the visual difference of that pixel to the background (i.e. the skin). Then a threshold value is determined from the average intensity of the $p\%$ highest gradient pixels in the obtained intensity image to find approximate lesion boundaries. Finally the lesion boundary is refined using edge information in the image. The problem of this algorithm for our task is choosing an eligible value for p . For an image containing a large hemangioma this value should be taken larger than for an image containing a small lesion. For a set of images with small variance of lesion sizes a failure made here can be compensated in the refining step. But hemangioma sizes vary strongly from one image to another. In Fig. 3 this problem is demonstrated with an image containing a small spot-shaped hemangioma [Fig. 3(a)] and an image containing a comparatively large hemangioma [Fig. 3(b)]. For the small hemangioma image a percentile $p=0.01\%$ looks adequate [Fig. 3(c)]. Segmenting the other image with this percentile detects the borders too far inside the hemangioma [Fig. 3(d)]. For that image a percentile $p=25\%$ is optimal [Fig. 3(e)]. However, this value is too large for the image containing the small hemangioma [Fig. 3(f)]. Nevertheless, the intensity image describing the visual difference between skin and hemangioma is a useful feature for our purpose and is used slightly modified as a feature for the classification step (see Section 3.3.3).

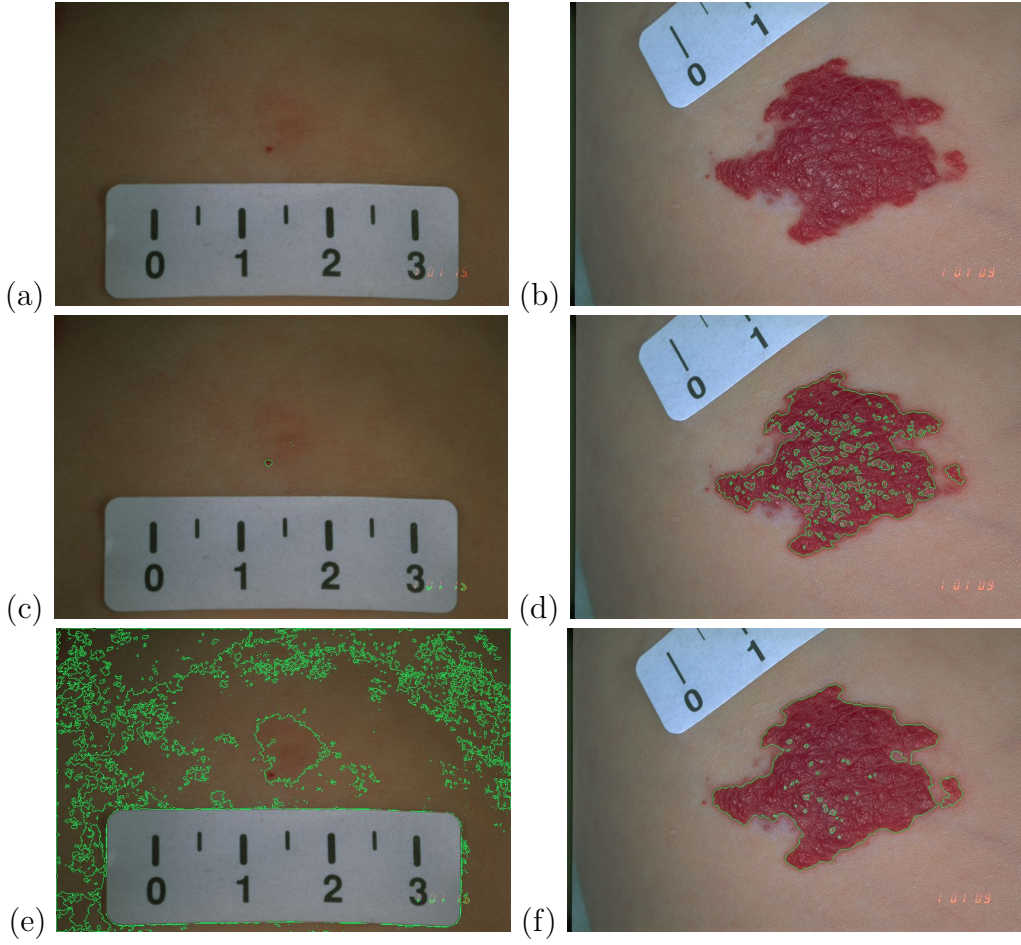


Figure 3: Results of intensity image thresholding of two images with different values for p .

Region-based methods: Further possibilities for segmenting skin lesions are region-based segmentation algorithms. With region-based algorithms not only the color values of pixels but also their spatial relationships are considered. Split-and-merge is a widely used region-based segmentation algorithm [7]. It first splits an image into homogeneous sub-regions and then merges together regions with similar average color values. Round et al's work on segmentation of skin lesions is basically an application of the split-and-merge algorithm [10]. Schmid and Fischer presented a region-based approach working with two-dimensional histogram analysis and fuzzy c-means clustering technique [12]. For our purpose we tested the region-based method explained in [5] on our images. The algorithm proposed in this work first splits the image into smaller regions until all the regions meet the homogeneity criteria set by a threshold. Second, the small split regions are grouped by DBSCAN clustering algorithm to form the final regions. The results of applying this method on two of our images are shown in Fig. 4.

The method was rejected due to the reason that after the clustering process it is very difficult to decide which clusters belong to the hemangioma and which do not. Although

the clustering of Fig. 4(a) induces a large cluster containing the major part of the hemangioma [Fig. 4(b)], note that the black regions are detected to be non skin before, see Section 3.2.2), there are several clusters near the border of the hemangioma which cannot be classified easily. A solution for that problem would be to change the parameters of the algorithm to merge together regions having more different color values. This would give us a single cluster containing the whole hemangioma. But although the image of Fig. 4(c) was clustered with the same parameters like the image of Fig. 4(a), the result [Fig. 4(d)] shows clusters, where skin and parts of the hemangioma were merged together. It was impossible to find a set of parameters working well with most of the images.

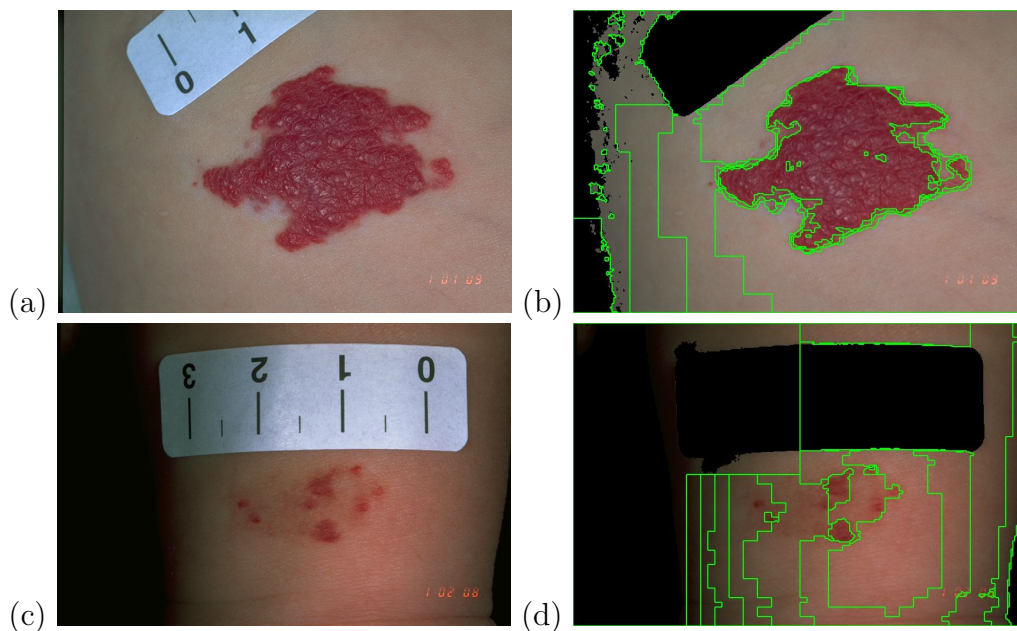


Figure 4: Result of DBSCAN clustering of two images.

As a conclusion the problem of all color segmentation algorithms is to determine which of the segmented regions are hemangioma regions. Making use of specific a priori knowledge about the general appearance of hemangiomas is difficult because of several reasons:

1. Hemangiomas can consist of several unconnected regions, so more than one region has to be classified as hemangioma then
2. Hemangiomas do not have a specific shape (with the exception that very longish regions can be classified as non hemangioma, see Section 3.4.2)
3. The hemangioma size in the images differs to a high extent

Our Methodology

Due to the difficulties mentioned previously we decided to use a classifier for segmenting the images that classifies each pixel in the image as hemangioma or non hemangioma on

the basis of its color values. Although postprocessing is still necessary (see Section 3.4), we get a basic segmentation by this method and can reject regions that are classified erroneously as belonging to the hemangioma later on.

Choice of the Classifier

In the following we had to determine the type of classifier suitable for our purpose. To get satisfactory classification results the data to be classified (i.e. the color values) have to show some form of clusters, where a cluster describes data values belonging to the same class. To test that we first normalized the RGB color values of 10 images in such a way that the skin has nearly the same color values in all images (see Section 3.2.3). By plotting their RGB values in 3D space [Fig. 5(a), the side views are shown in Fig. 5(b)-(d)] it can be seen that the color values of the hemangioma (red) are mostly separated from the other values (blue).

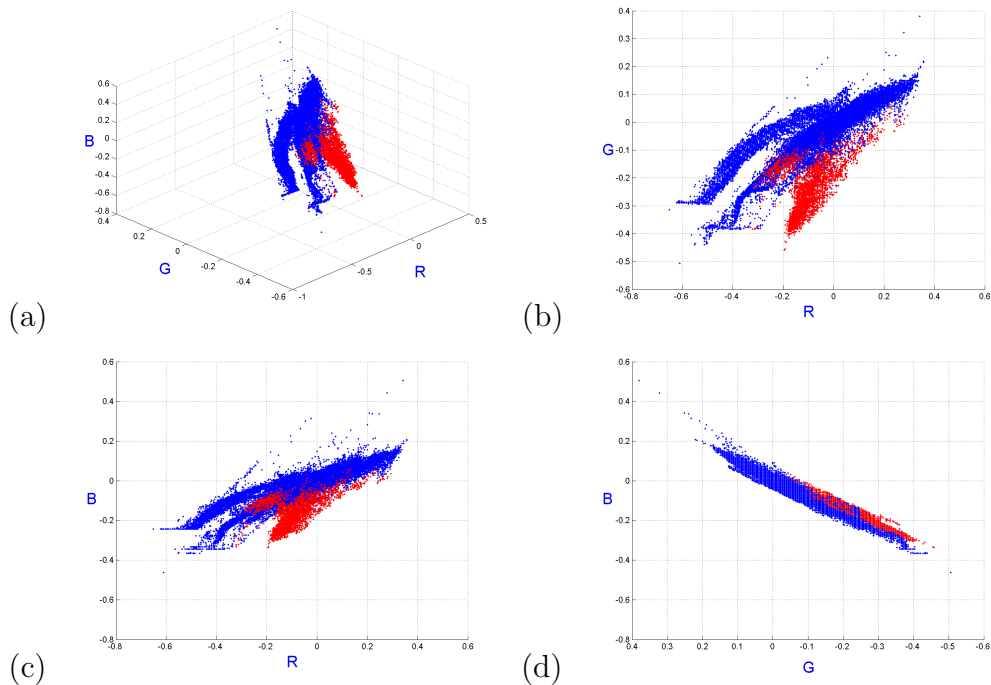


Figure 5: 3D Plot of normalized RGB color values in 10 images. The red points represent the color values of the hemangioma pixels.

Single-layer perceptron: The single-layer perceptron is a simple type of binary classifier having the advantage of a simple and fast classification [9]. Looking at it geometrically a single-layer perceptron classifies the data of Fig. 5 by putting a plane in the 3D space that divides the data set in two classes. Therefore an adequate basic segmentation result can be assumed on our images by applying a single-layer perceptron. As we see later this assumption is confirmed, especially if we take more appropriate color values from other

color models (see Section 3.3.2).

Neural network: Beside the single-layer perceptron various other classifiers exist. A neural network (multi-layer perceptron) is an improvement of the traditional single-layer perceptron with the ability of setting non-linear decision boundaries instead of only linear ones. In other words, a neural network can put a curved surface in 3D space as decision boundary. We have tested neural networks for segmenting our images with the conclusion, that the results are not better than by segmenting with a single-layer perceptron. Since a single-layer perceptron is simpler and faster in classification, we kept the perceptron.

k -Nearest Neighbor classifier: The k -Nearest Neighbor (k-NN) classifier determines the k nearest feature vectors of a reference set for every feature vector [1]. The specific feature value is mapped to the class having the majority in the k nearest feature vectors. Due to the complex feature distribution a k -NN classifier with $k = 5$ achieved better segmentation results than the single-layer perceptron on our images. The k-NN classifier achieved an average false positives rate of 9.2% and an average false negatives rate of 16.1% on 7 images, where the perceptron achieved an average false positives rate of 26.6% and an average false negatives rate of 16.9% on the same 7 images. Nevertheless, the k-NN classifier is not practical for our purpose because of its high computation time, since for our images with a resolution of 512x768 pixels in total 393216 feature vectors have to be classified for only one image.

Due to the similar results of the neural network and the bad computation time of the k -NN classifier we finally came to the decision to use the single-layer perceptron for classification.

3.2 Preprocessing

Before the classification step the images have to be preprocessed to improve the accuracy of the perceptron classifier and to reduce computation time. To remove noise a median filter is applied on the images. Furthermore, image regions containing no skin are masked out and the images are normalized in such a way that skin has nearly the same color values in all images.

3.2.1 Noise Removal

The task of the noise removal process is to reduce small structures in the images. Smoothing the images causes less color variations in the skin and hemangioma and makes thereby the segmentation process more accurate. With low-pass filtering an image is smoothed by replacing each pixel by a weighted sum of its neighbors. However, with median filtering, the value of an output pixel is determined by the median of the neighborhood pixels. The advantage of this method in contrast to low-pass filtering is that it preserves the edges of the image while reducing the noise. In this work we use a median filter with a window size

of 5x5, i.e. the value of an output pixel is determined by the median of all pixel included in 5x5 window surrounding the particular pixel. By testing this window size was decided to be the best compromise between reducing noise and preserving relevant information in the images.

3.2.2 Non Skin Masking

A simple test for masking out non skin regions is used after the median filtering to exclude regions that likely are not part of the skin or the hemangioma (e.g. the ruler or dark areas in the background). That step is necessary for a robust determination of the skin's color (see Section 3.2.3) and enhances the speed of the overall segmentation step as well, since not all pixels of the image have to be classified. Our method is based on a heuristic proposed in [6] but substantially simpler. We only check two criteria for each pixel (R , G and B are the red, green and blue color values of the RGB color model):

1. $R < G$
2. $R < B$

If one of the criteria is given, the particular pixel is marked as non skin. This test makes use of the fact that skin has usually a reddish color and therefore shows a greater red portion than green and blue portion. By applying this test on all 122 images we got a almost perfect result for 96 images (classification error less than 1%). The rest shows an average classification error of about 5%. The non skin masking never causes problematic results for the further steps, because the hemangioma and the bigger part of the skin is never masked out. Fig. 6 shows the non skin masking results of three particular images (non skin regions are marked blue). Non skin masking of the image in Fig. 6(a) induces a nearly perfect result [Fig. 6(b)]. Worse results are achieved in Fig. 6(c) and (e). The failure of Fig. 6(d) is mainly caused by a red cloth in the dark background of the image. In Fig. 6(e) some parts of the skin are erroneously masked out in a rather badly illuminated region of the image. Here the skin loses its reddish appearance.

3.2.3 Normalization with Skin Color

To achieve more accurate classification results a normalization with the skin color has to be done on the images. In other words, we have to determine the color value of the skin in an image and subtract these value from all pixels with the aim of having nearly the same color value of (0,0,0) for skin pixels in all images. The necessity of this step is shown in Table 1 where we have computed the classification result (false positives and false negatives rate) of the perceptron for 15 images without normalization and compared it with the results achieved with normalization. The same feature set as for the classification step was used (see Section 3.3.2) and all images were reduced to a resolution of 256x384 and randomly divided into a training set (30%) and a test set (70%). For normalization we have tested three different variants of computing the skin color:

1. **Manual Normalization:** the user chooses three 3x3 windows near the border of the hemangioma and the mean color value of all windows is chosen to be the skin's

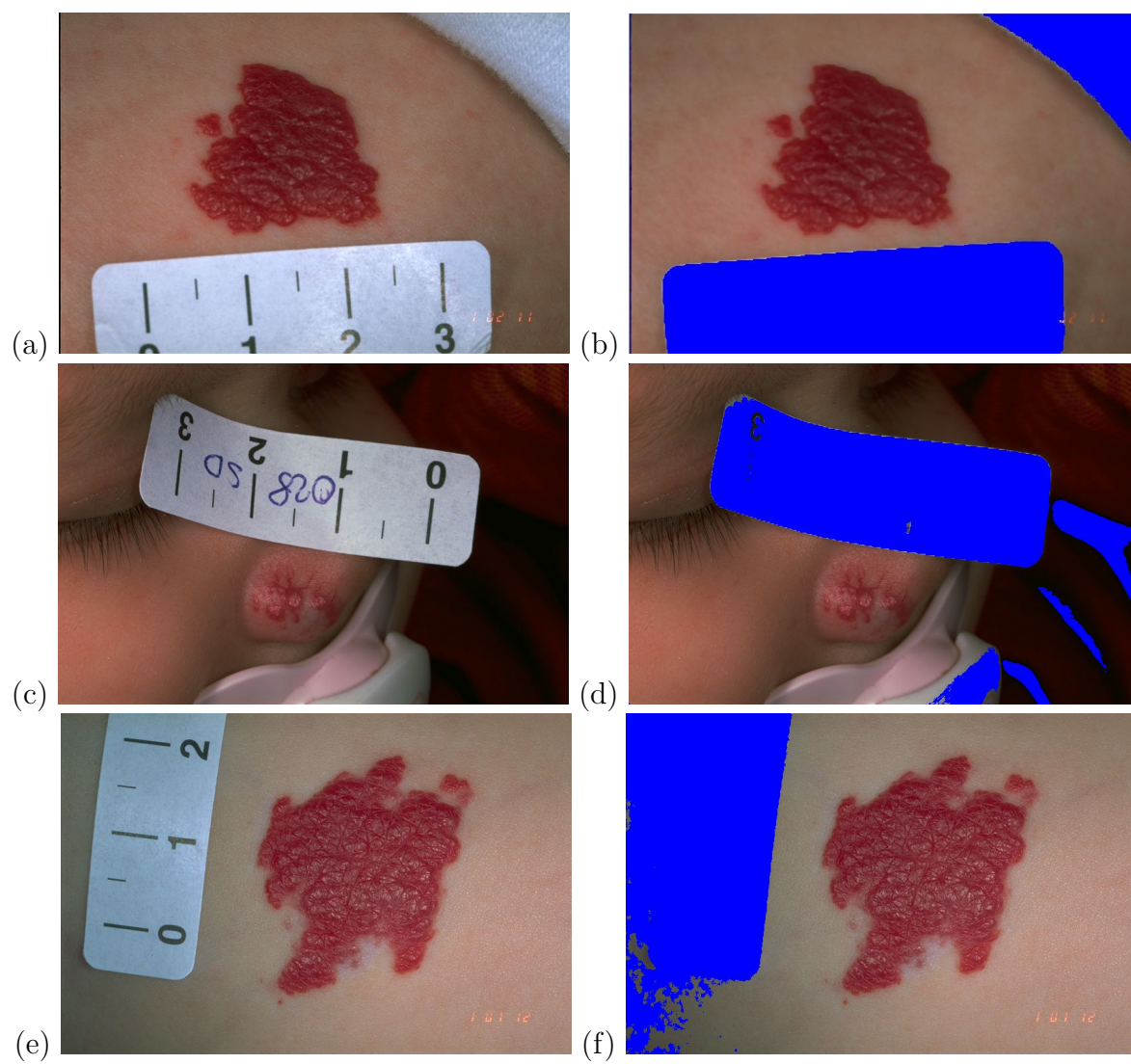


Figure 6: Non skin masking results of three images.

	Test Set			Training Set		
	False Pos.	False Neg.	Sum	False Pos.	False Neg.	Sum
No Norm.	0.0%	99.8%	99.8%	0.0%	99.8%	99.8%
Manual Norm.	29.7%	14.8%	44.5%	29.5%	14.5%	44.0%
Median Norm.	30.2%	15.5%	45.7%	30.0%	15.2%	45.2%
Histogram Norm.	29.4%	14.7%	44.1%	29.2%	14.4%	43.6%

Table 1: Classification results after various normalization methods.

color (this is only for testing, since we want an automated procedure with no user interaction).

2. **Median Normalization:** the median value for each color channel in the image is chosen
3. **Histogram Normalization:** a 3D histogram of the RGB color channels is created and the maximum RGB value, which has a brightness greater than 120, is chosen. This threshold was decided to be the most appropriate by empirical tests.

Both Median and Histogram Normalization make use of the fact that after non skin masking the majority of the remaining pixels in an image represent skin.

As can be seen in Table 1, without normalization the perceptron is not able to find an useful decision boundary. The perceptron has set a decision boundary that classifies nearly all pixels as non hemangioma and therefore we get a false positives rate of 0.0% and a false negatives rate of 99.8% without normalization. With normalization we obtain more usable decision boundaries resulting in a total error rate (false positives rate plus false negatives rate) of 44.5% (Manual Normalization), 45.7% (Median Normalization) and 44.1% (Histogram Normalization) in the test set. Although median and histogram normalization accomplish nearly the same results, the subjective visual result on particular images is better with histogram normalization. It can also be seen that the manual normalization does not achieve better results than the histogram normalization (44.1% against 44.5% total error rate in the test set) which is an indication that the histogram normalization works accurately.

3.3 Classification of the Hemangiomas

As mentioned before, our segmentation method is based on a perceptron that classifies all pixels in the images by means of their color values. As we see later in Section 3.4, this first segmentation result has to be postprocessed to exclude non hemangioma regions with hemangioma-like color values.

In Section 3.3.1 first the general functioning of the single-layer perceptron is explained. In Section 3.3.2 and 3.3.3 the selection of adequate features for our images is discussed. Finally Section 3.3.4 describes the training phase of the perceptron.

3.3.1 The Single-Layer Perceptron

The single-layer perceptron is a simple binary classifier based on a mathematical model for the behavior of a single biological neuron.

Assume that we have a d -dimensional feature vector $x \in \mathbb{R}^d$ and two classes w_1, w_2 . Our goal is to find a mapping $g : \mathbb{R}^d \rightarrow \mathbb{R}$ with

$$g(x) > 0 \text{ if } x \in w_1 \quad (1)$$

$$g(x) < 0 \text{ if } x \in w_2 \quad (2)$$

g is often called the *discriminant function* and has the form

$$g(x) = \sum_{i=1}^d w_i x_i - \theta = \mathbf{w}^T \mathbf{x} - \theta \quad (3)$$

where

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_d \end{pmatrix}, \mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ \dots \\ w_d \end{pmatrix} \quad (4)$$

w is normally called the weight vector and θ the *bias*. If we call $o(x)$ the output of the perceptron, we have

$$o(\mathbf{x}) = g(\mathbf{w}^T \mathbf{x} - \theta) = \begin{cases} 1 & \text{if } \mathbf{w}^T \mathbf{x} \geq \theta \\ -1 & \text{if } \mathbf{w}^T \mathbf{x} < \theta \end{cases} \quad (5)$$

where $g()$ is the signum function:

$$\text{sgn}(a) = \begin{cases} +1 & \text{if } a \geq 0 \\ -1 & \text{if } a < 0 \end{cases} \quad (6)$$

It is common practice to "pull" the bias into the weight vector by introducing additional coordinates $x_0 = 1$ and $w_0 = -\theta$.

$${}^a\mathbf{x} = \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ \dots \\ x_d \end{pmatrix}, {}^a\mathbf{w} = \begin{pmatrix} -\theta \\ w_1 \\ w_2 \\ \dots \\ w_d \end{pmatrix} \quad (7)$$

Perceptron training

Let $S_{Tr} = \{\mathbf{X}, \mathbf{t}\}$ denote a set of N augmented input vectors $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_N) \in \mathbb{R}^{(d+1) \times N}$ and corresponding class labels $\mathbf{t} = (t_1, \dots, t_N), t_i \in \{1, -1\}$, called *training set*.

Goal: find an augmented weight vector \mathbf{w} such that

$$o(\mathbf{x}_i) = \text{sgn}(\mathbf{w}^T \mathbf{x}_i) = t_i, 1 \leq i \leq N \quad (8)$$

Perceptron Learning Rule: if a training vector \mathbf{x}_j with $t_j = 1$ is misclassified, add a multiple of $\mathbf{x}_j t_j$ to \mathbf{w} : $\mathbf{w}_{\text{new}} = \mathbf{w} + \gamma \mathbf{x}_j t_j$. Likewise if a training vector \mathbf{x}_j with $t_j = -1$ is misclassified, subtract a multiple of $\mathbf{x}_j t_j$ from \mathbf{w} : $\mathbf{w}_{\text{new}} = \mathbf{w} + \gamma \mathbf{x}_j t_j$. The factor γ is called *learning rate*.

3.3.2 Feature Selection

For classification we have to define a set of features showing a big difference between skin and hemangioma pixels. Possible features for the classification are at first all the three color channels of these three significant color spaces [8]:

1. **RGB:** In RGB color space, each color is represented by a three number triple. The components of this triple specify, respectively, the amount of red, green and blue in the color.
2. **HSV:** In the HSV color space, each color is again represented by a three number triple. The first component, *Hue*, describes the basic color in terms of its angular position on a "color wheel". The *Hue* is normally described in terms of degrees. The second component of the HSV triple is *Saturation*, which can be thought of as the "purity" of the color. The third component is *Value*, which is a measure how "bright" the color is.
3. **CIE 1976 L*a*b*:** The CIE 1976 L*a*b* color space was created by the *Commission Internationale de l'Eclairage* and has the special ability, that the color differences perceived by the human eye correspond to colorimetrically measured distances. The L^* -axis in this three-dimensional color space is known as the lightness, the other two coordinates a^* and b^* represent redness-greenness and yellowness-blueness, respectively.

Further we have created a 10th feature called *abdist* (see Section 3.3.3). To find an appropriate set of features we have ascertained the false negatives and false positives rate for every meaningful group of features on a test set of 15 normalized images (Table 2). The images were reduced to a resolution of 256x384 before and randomly divided in a training set (30%) and test set (70%).

Although the best result was obtained with all ten features (22% false positives and 17.9% false negatives on the test set), the set $\{G, H, a^*, abdist\}$ is the better choice. It has produced the lowest false negatives rate of all groups (14.7%), which is more important in the classification step. This is because of the fact that in the postprocessing step mainly the false positives rate will be improved. Therefore, we have decided to use the features

G , H , a^* and $abdist$ for classifying the pixels of an image. As can be seen in Fig. 7, each of these features has a rather big difference between pixels belonging to the hemangioma and pixels belonging to the skin.

Features	Test Set		Training Set	
	False Pos.	False Neg.	False Pos.	False Neg.
$L\ a^*\ b^*$	23.2 %	37.9 %	23.1 %	37.2 %
$abdist$	33.8 %	25.2 %	33.7 %	25.0 %
a^*	30.0 %	31.7 %	29.9 %	31.3 %
$G\ H\ a^*$	30.2 %	34.9 %	30.0 %	34.3 %
$H\ S\ a^*$	23.3 %	46.3 %	23.3 %	45.5 %
$V\ abdist$	28.0 %	27.0 %	27.7 %	26.6 %
$G\ H\ a^*\ abdist$	29.4 %	14.7 %	29.2 %	14.4 %
$R\ G\ H\ S\ V\ a^*$	24.8 %	41.5 %	24.8 %	41.0 %
$H\ S\ abdist$	21.3 %	23.9 %	21.3 %	23.6 %
all	22.0 %	17.9 %	21.9 %	17.6 %

Table 2: False positives and false negatives of a set of feature groups of 15 images trained with a perceptron.

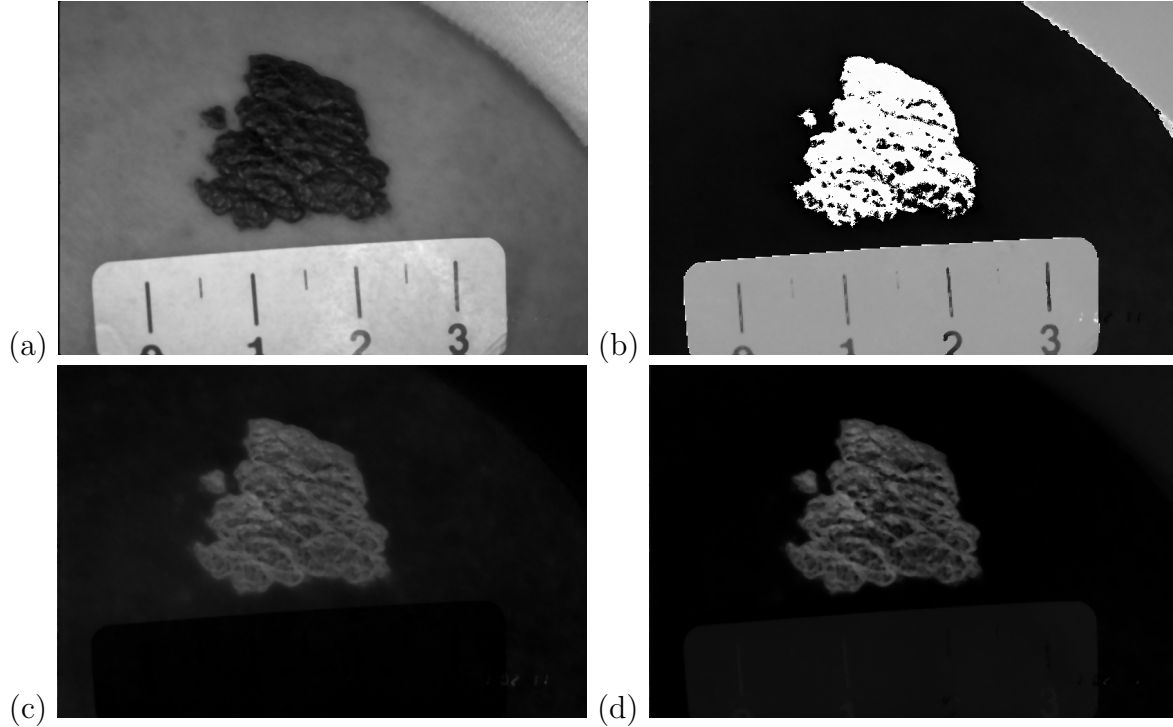


Figure 7: The intensity images of the four features (a) G , (b) H , (c) a and (d) $abdist$ of the image in Fig. 6(a).

It must be noticed that for classification the H values have to be rotated 180 degrees to cope with the circular fashion of the data. In the HSV color model the red color value lies

at 0° , i.e. similar red color values can have very different H values. Since hemangiomas are red, the red color values of H are the most relevant for classification and we have to bring these values in a linear range.

3.3.3 The Feature *abdist*

The feature *abdist* stands for the Euclidean distance between the skin and the hemangioma in the $L^*a^*b^*$ color space without consideration of the luminance L^* and intensification of the a^* component. This feature is adopted from [14] (see Section 3.1). In this paper the proposed method works on an intensity image describing the Euclidean distance between the skin and the lesion.

The advantage of the *CIE 1976 $L^*a^*b^*$* color space compared with other color spaces is that it defines color in such a way that the Euclidean distance between two colors is proportional to their visual difference. This property is very useful for our purpose, since the color distance in RGB space between hemangioma and skin regions is not as high as the perceived difference.

If a_s, b_s denotes the a^* and b^* values of the skin (obtained from the normalization step, see 3.2.3) and a_h, b_h that from the hemangioma, *abdist* is computed as follows:

$$abdist = \sqrt{(2a_s - 2a)^2 + (b_s - b)^2} \quad (9)$$

The difference of the a^* channel is multiplied with the factor 2, because the a^* value differs more between hemangioma and skin pixels than the b^* value. At last for contrast enhancement we apply a Gaussian function of the form $G(x) = 1/\sqrt{2\pi\sigma}(1 - \exp(-x^2/2\sigma^2))$ with $\sigma = 0.5$ on the intensity image. As can be seen in Fig. 8, this function increases higher values (hemangioma) and decreases lower values (skin) in the intensity image. For an example of *abdist* see Fig. 7(d).

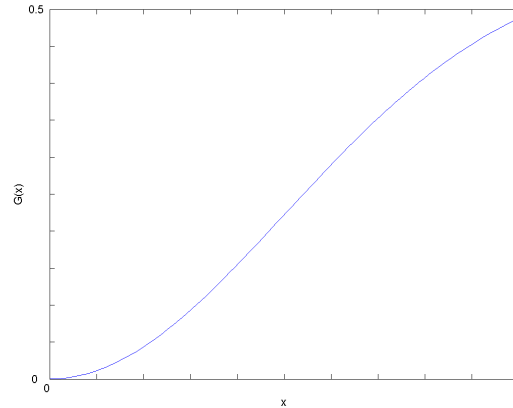


Figure 8: Gaussian function $G(x) = 1/\sqrt{2\pi\sigma}(1 - \exp(-x^2/2\sigma^2))$ with $\sigma = 0.5$

3.3.4 Training of the Single-Layer Perceptron

To train the perceptron we took a training set of 15 images representing a wide variety of different appearances of hemangiomas as shown in Fig.1. The ground truth, i.e. the class labeling of every pixel of the images, was manually determined by ourself by drawing the borders of the hemangiomas in a paint program. All features were brought in the range $[0, 1]$ and normalized with histogram normalization. Best results were achieved with a learning rate $\gamma = 0.0001$ and 50 training cycles. Thereafter no improvement of the classification error was recognized. On the training set the trained perceptron achieved a false positives rate of 30.9% and a false negatives rate of 14.2%.

3.4 Postprocessing

In the postprocessing step several operations are applied to the mask obtained from the classification. To improve the false negatives rate we have to find a way to detect highlights on the hemangioma that were erroneously classified as non hemangioma. Contrariwise we also have to reject regions, which have similar color values to those of the hemangioma but do not belong to it. These regions include mostly:

- The red five-digit code in the lower right corner of the image
- Parts of the human body that naturally have a similar color to hemangiomas (i.e. lips)
- Skin regions near the border between skin and dark background regions (that seems to be a special property of the film used in the study)
- Other red objects like earrings, baby-soothers etc.

In the following we describe all operations used to get the final mask.

3.4.1 Hole Closing

For solving the highlights problem mentioned above all holes occurring in the mask are closed. Since hemangiomas with large holes of normal skin could not be found in the 122 images appropriated for us, they seem to be very rare and a possible error resulted from that operation can be neglected. The effect of hole closing can be seen in Fig. 9: the highlights detected as non hemangioma by the classifier [Fig.9(a)] are added to the mask by hole closing [Fig.9(b)].

3.4.2 Rejecting Regions

Size-based: First of all regions with a size smaller than 40 pixels are deleted from the mask. In all images no hemangioma region with a size smaller than 60 pixel is contained, therefore this value is proven to be adequate to exclude artifacts.

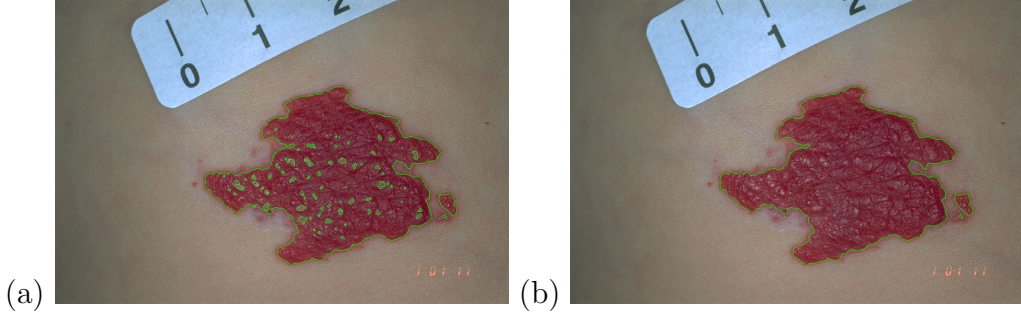


Figure 9: Result of hole closing on a particular image.

Position-based: To exclude regions belonging to the five-digit code of the photo, regions having their centroid in a 30x100 frame in the lower right corner of the image are also taken out. The hemangiomas are normally located vaguely in the middle of the image. Even if parts of it are overlapping with this region, still their centroid is close to the middle of the image.

Roundness-based: To detect the regions near the border between skin and dark background areas we take advantage of the fact that they mostly stretch over a thin, long line. To discern this regions from hemangioma regions we use *shape descriptors* [11], in particular the *roundness*. The *roundness* of a region is defined as:

$$roundness = \frac{4A}{\pi D_{max}^2} \quad (10)$$

where A is the area and D_{max} the length of the major axis of the ellipse that has the same second-moments as the region. Hemangiomas are rather compact and round, so we reject regions with a very small roundness (in this project we used a empirically determined threshold of 0.1).

Color-based: The last criterion we check for each region has to do with his *saturation* and *value* (from the HSV color model). The classifier often chooses by mistake dark and less saturated regions (in comparison to the other regions of the image) to be part of the hemangioma, because they have high values in the *abdist* feature. To exclude these regions we calculate the mean S and V values of the whole image after non skin masking and compare it with the mean values of the region. That is to say we calculate

$$\frac{Mean(S_{reg})}{Mean(S_{img})} + \frac{Mean(V_{reg})}{Mean(V_{img})} \quad (11)$$

whereas $Mean(S_{reg})$, $Mean(S_{img})$, $Mean(V_{reg})$ and $Mean(V_{img})$ are the mean values of the S and V values of the image and the region. This computed value is small for regions with a relative low saturation and/or brightness compared to the rest of the image.

Therefore, all regions having a lower value calculated with this formula than a specific threshold are deleted from the mask (by testing we found an appropriate threshold of 1.3).

Recapitulating the following steps are performed in postprocessing:

- Close holes
- Reject regions with:
 1. $Area < 40$
 2. $Roundness < 0.1$
 3. $Centroid\ in\ frame\ (440-470, 610-710)$
 4. $\frac{Mean(S_{reg})}{Mean(S_{img})} + \frac{Mean(V_{reg})}{Mean(V_{img})} < 1.3$

Fig. 10 shows the computed hemangioma borders before [Fig. 10(a)] and after [Fig. 10(b)] the postprocessing step of a particular image. The two large, longish regions are rejected because of their low roundness. The false positives rate of this segmentation is decreased from 90.3% to 5.5%, the false negatives rate from 17.7% to 2.2%.

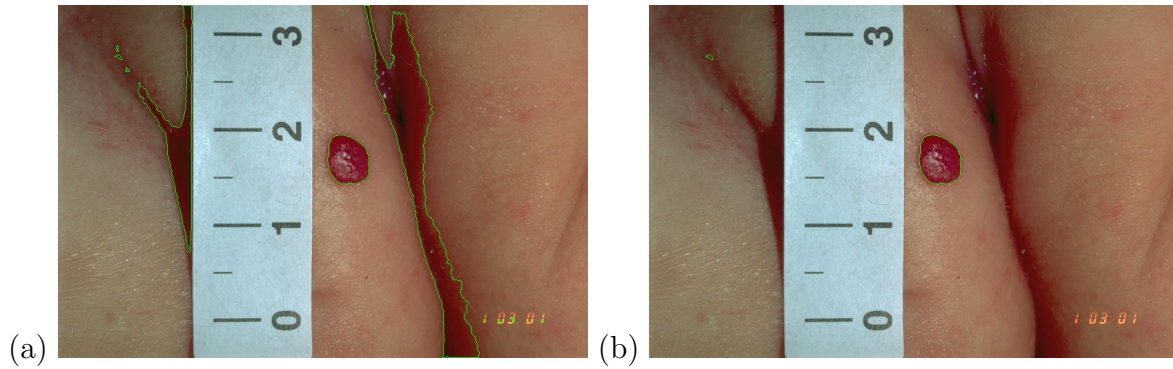


Figure 10: Example of the changes made by the postprocessing step.

4 Experiments

In this section we report experiments with the data gathered and our algorithm applied on it to evaluate the proposed method. First of all we test the precision of the scale computing algorithm solely (Section 4.1). Then the improvement of the postprocessing step is shown by comparing the results of 29 images before and after postprocessing (Section 4.2). Furthermore, we analyze the differences between images of the same hemangioma photographed at the same date in terms of their computed area to get an estimation of the precision of the method (Section 4.3). For evaluating time series of the images another test is made with two series of hemangioma images photographed in monthly intervals (Section 4.4). The performance of our classifier is evaluated by means of *leave-one-out*

testing (Section 4.5). At last the possibility of performing a second iteration of the segmentation step with new normalization values is evaluated (Section 4.6).

Like for the data used in the training phase (Section 3.3.4), the ground truth of the images is annotated manually.

4.1 Precision of the Scale Computing Algorithm

For this experiment we computed the euclidean distance between two marks of the ruler of 20 images with the algorithm proposed in Section 2 and compared it with manually determined values (Table 3). These values are achieved by manually choosing the longest euclidean distance between two marks of the ruler visible in the image. Nevertheless, these values do not represent the real spatial resolution of the image since we do not take into account the curvature and position of the ruler just as in our algorithm. In addition the manually determined euclidean distance is not exact too, a failure of 1-2 pixels is possible. As can be seen in Table 3 the algorithm works well for most of the images (average error rate 1.55%).

Image	Computed Euclidean Distance	Manually Determined Euclidean Distance	Error
1	130 px	129 px	0.78 %
2	127 px	124 px	2.42 %
3	132 px	129 px	2.33 %
4	136 px	134 px	1.49 %
5	140 px	138 px	1.45 %
6	135 px	133 px	1.50 %
7	139 px	135 px	2.96 %
8	139 px	139 px	0.00 %
9	130 px	130 px	0.00 %
10	133 px	131 px	1.53 %
11	134 px	141 px	4.96 %
12	130 px	129 px	0.78 %
13	145 px	145 px	0.00 %
14	141 px	151 px	6.62 %
15	98 px	100 px	2.00 %
16	140 px	141 px	0.71 %
17	146 px	146 px	0.00 %
18	130 px	132 px	1.52 %
19	141 px	141 px	0.00 %
20	146 px	146 px	0.00 %
Average			1.55 %

Table 3: Comparison between the computed and the manually determined euclidean distance between two marks of the ruler in 20 images.

Problems arise only with images 11 and 14. These are shown in Fig. 11. In image 11 [Fig. 11(a)] there can be seen a white cloth that is segmented together with the ruler.

Therefore, in the computed, rotated ruler mask the three scanlines (red) are too far below and erroneously the distance between the numbers instead of the marks is taken for scale computing [Fig. 11(b)]. The same problem occurs in image 14 [Fig. 11(c) and 11(d)] but this time caused of the fact that only about the half of the ruler is visible in the image. Generally the algorithm works well on images with completely visible rulers that are surrounded only by skin.

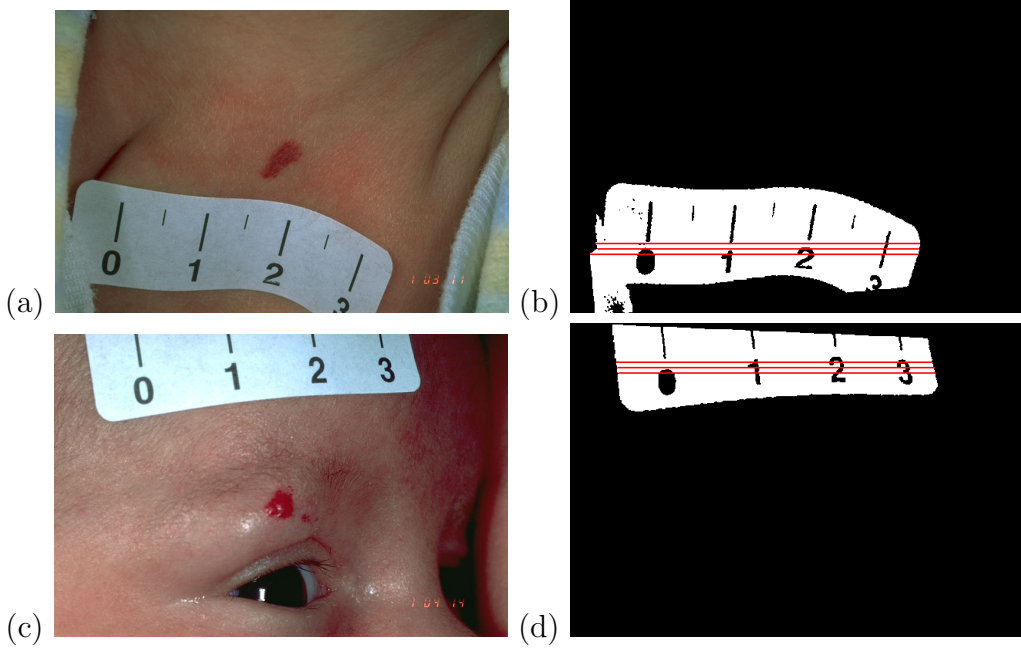


Figure 11: Image 11 and 14 of Table 3 with corresponding computed ruler masks. The three scanlines are marked red.

4.2 Results on Images and Improvement by Postprocessing

We have tested the segmentation of 29 single images to get an approximation of the overall error rate of the segmentation algorithm proposed. In addition we have tested the effect of the postprocessing step by comparing the false positives and false negatives rate of the same 29 images before and after postprocessing (Table 4). On average by postprocessing we get primarily an improved false positives rate (from 39.0% to 15.5%) and also the false negatives rate is decreased (from 17.4% to 13.1%). It can also be seen in the table that no image's error rate is increased significantly by the postprocessing step (merely images 1, 3, 22 and 25 show a negligible deterioration of the false negatives rate).

In Fig. 12 we show images of both good and bad segmentation results. Fig. 12(a)-(e) correspond to images 6, 8, 13, 23 and 28 in Table 4, respectively, and belong to the best segmentation results. Fig. 12(f)-(j) correspond to images 7, 14, 15, 19 and 21 in Table 4, respectively, and belong to the worst segmentation results.

	Before Postprocessing		After Postprocessing	
Image	False Positives	False Negatives	False Positives	False Negatives
1	17.7 %	13.3 %	15.9 %	14.3 %
2	0.3 %	12.2 %	0.3 %	10.4 %
3	21.5 %	36.3 %	1.3 %	36.4 %
4	6.3 %	16.1 %	0.8 %	7.9 %
5	41.6 %	27.7 %	9.9 %	27.2 %
6	26.2 %	7.7 %	2.9 %	5.9 %
7	67.6 %	3.2 %	51.4 %	3.2 %
8	1.1 %	8.4 %	0.8 %	2.4 %
9	0.5 %	48.0 %	0.2 %	15.8 %
10	32.3 %	4.0 %	7.1 %	3.9 %
11	3.1 %	12.6 %	2.7 %	10.5 %
12	90.0 %	18.0 %	4.2 %	2.2 %
13	1.3 %	9.1 %	0.6 %	5.3 %
14	89.4 %	46.9 %	11.7 %	46.6 %
15	76.2 %	3.8 %	63.8 %	1.8 %
16	9.2 %	40.8 %	0.0 %	5.8 %
17	75.4 %	9.2 %	6.4 %	9.2 %
18	67.9 %	6.7 %	58.9 %	6.7 %
19	29.0 %	14.8 %	24.1 %	14.8 %
20	72.3 %	19.7 %	19.2 %	18.3 %
21	96.2 %	16.1 %	96.2 %	14.2 %
22	26.1 %	8.2 %	6.9 %	8.3 %
23	3.6 %	5.1 %	3.6 %	5.1 %
24	97.2 %	17.2 %	5.2 %	17.2 %
25	52.4 %	5.1 %	43.3 %	5.7 %
26	0.0 %	10.0 %	0.0 %	9.3 %
27	24.0 %	46.9 %	9.4 %	45.7 %
28	49.0 %	6.9 %	1.7 %	4.6 %
29	54.3 %	30.2 %	1.8 %	20.8 %
Average	39.0 %	17.4 %	15.5 %	13.1 %

Table 4: False positives and false negatives rate of 29 images before and after the post-processing step.

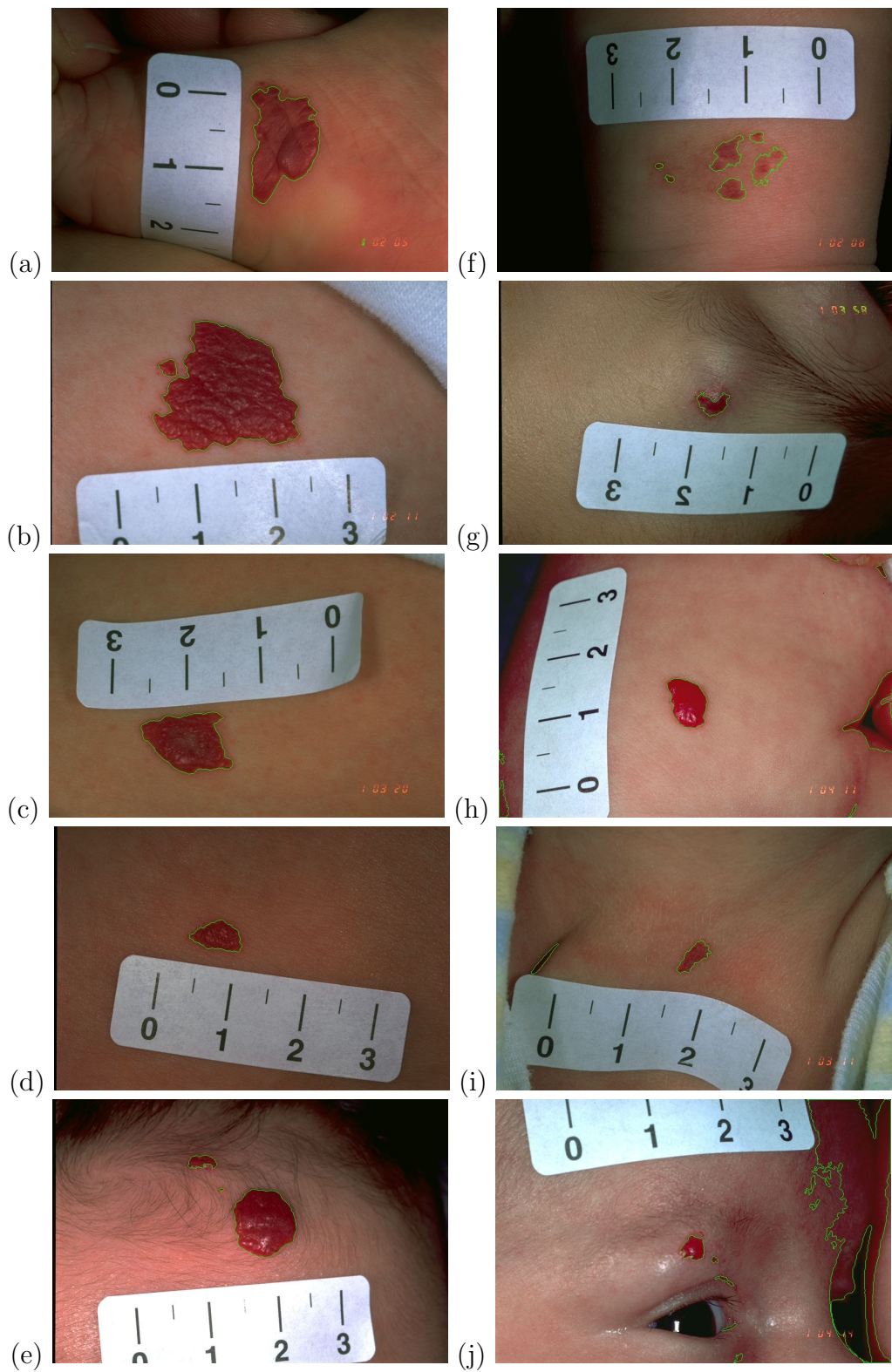


Figure 12: Computed segmentation borders (green) of 10 images with our algorithm.

4.3 Comparing Computed Areas of the Same Hemangioma

As mentioned above, all hemangiomas were photographed at least twice, which gave us the possibility to test the precision of the complete algorithm proposed, i.e. the scale computing algorithm combined with the segmentation algorithm. With a "perfect" algorithm both images should have precisely the same computed hemangioma area. Table 5 shows the computed areas of two different images displaying the same hemangioma at a specific time of 20 hemangiomas in total. The average error (difference in percent) of all 20 hemangiomas is 11.3%. This error is mainly caused by the circumstance that two photos of the same hemangioma are never photographed from exactly the same angle.

	Area Image 1	Area Image 2	Difference	%
	2.626 cm^2	2.397 cm^2	0.229 cm^2	8.7 %
	0.185 cm^2	0.219 cm^2	0.034 cm^2	15.5 %
	0.348 cm^2	0.363 cm^2	0.016 cm^2	4.3 %
	3.755 cm^2	3.600 cm^2	0.154 cm^2	4.1 %
	0.215 cm^2	0.222 cm^2	0.007 cm^2	3.2 %
	0.097 cm^2	0.119 cm^2	0.023 cm^2	18.9 %
	0.635 cm^2	0.592 cm^2	0.043 cm^2	6.8 %
	2.119 cm^2	1.889 cm^2	0.230 cm^2	10.9 %
	0.128 cm^2	0.130 cm^2	0.002 cm^2	1.8 %
	0.524 cm^2	0.569 cm^2	0.045 cm^2	7.9 %
	0.197 cm^2	0.176 cm^2	0.021 cm^2	10.8 %
	0.147 cm^2	0.123 cm^2	0.024 cm^2	16.5 %
	2.901 cm^2	2.475 cm^2	0.426 cm^2	14.7 %
	0.839 cm^2	0.735 cm^2	0.104 cm^2	12.4 %
	0.193 cm^2	0.115 cm^2	0.078 cm^2	40.5 %
	0.091 cm^2	0.106 cm^2	0.015 cm^2	14.1 %
	0.217 cm^2	0.204 cm^2	0.013 cm^2	6.0 %
	2.061 cm^2	2.057 cm^2	0.004 cm^2	0.2 %
	0.610 cm^2	0.659 cm^2	0.049 cm^2	7.4 %
	0.153 cm^2	0.120 cm^2	0.033 cm^2	21.3 %
Average			0.078 cm^2	11.3 %

Table 5: Difference in computed areas between two images of the same hemangioma.

4.4 Tests on Series of Hemangioma Images

In this experiment we observed the course of computed areas on two series. Both series contain three images of an hemangioma photographed in intervals of about one month to record their development. The results can be seen in Table 6. Series 1 shows a slight enlargement of the hemangioma in the first five weeks, but after 10 weeks the hemangioma has become very small. In series 2 the area of the hemangioma is quite constant. It seems, that the small differences between the three measurements are caused by error and the hemangioma has not changed his size. Fig. 13(a)-(c) shows the three segmentation results of series 1 in chronological order and Fig. 13(d)-(e) the segmentation results of series 2.

Series 1		Series 2	
Date	Area	Date	Area
2003-08-26	0.348 cm^2	2003-07-15	0.202 cm^2
2003-10-07	0.414 cm^2	2003-08-19	0.152 cm^2
2003-11-11	0.003 cm^2	2003-09-16	0.187 cm^2

Table 6: Results for computing the area of two series of hemangioma photographed in an interval of about a month.

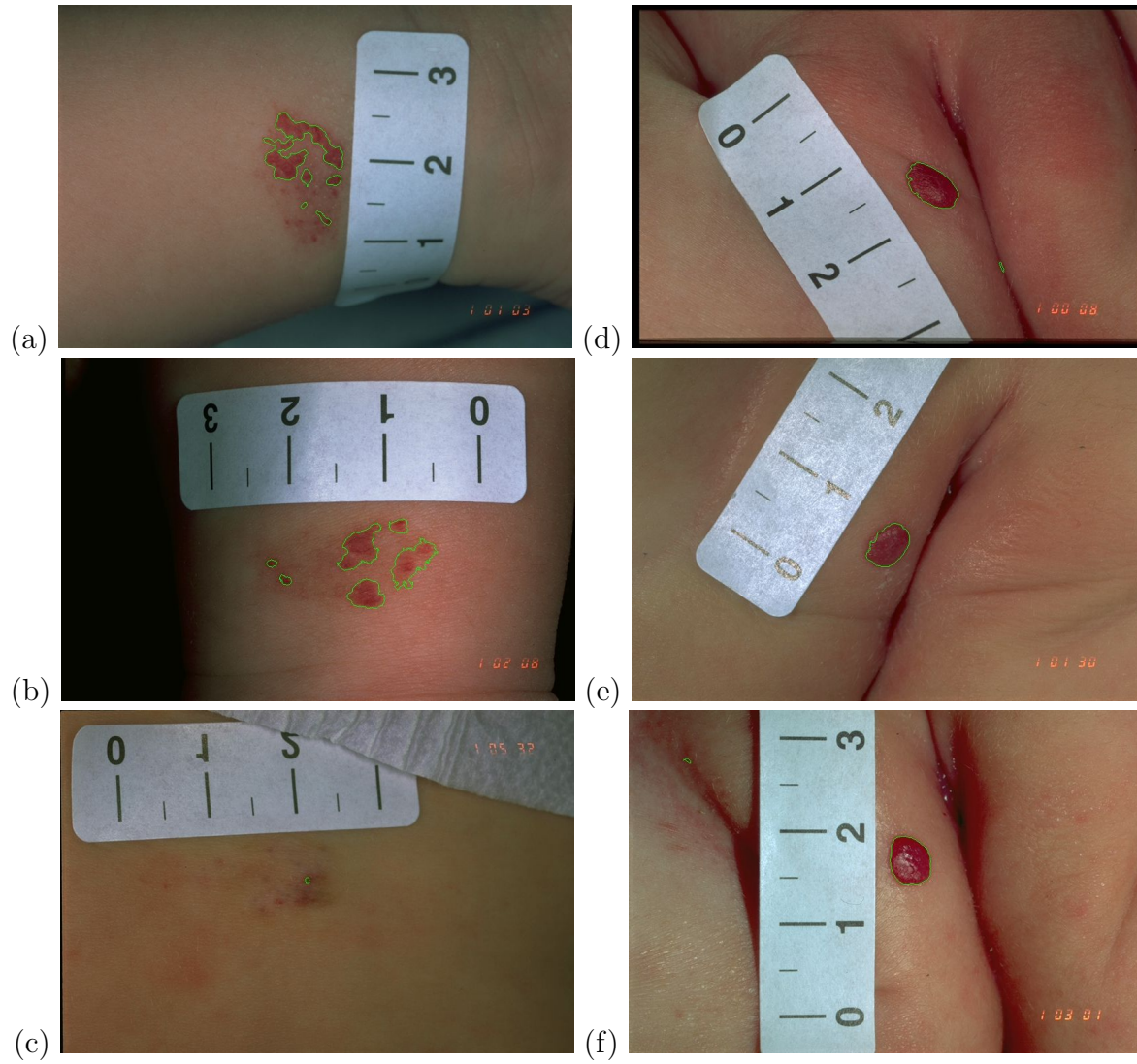


Figure 13: Segmentation results of two series photographed in an interval of about a month.

4.5 Leave-One-Out Test

The *leave-one-out* test (or *round robin* test) is an elegant and straightforward technique for estimating a classifier's error rate. The perceptron is trained with 28 images and tested on the 29th image. This is done 29 times, each time training a perceptron on all but one images and testing it on the single remaining image. Table 7 shows the false positives and false negatives rates obtained by this method for the 29 images that were already utilized for Table 4. For training the same parameters as in the training phase (Section 3.3.4) were taken. On the average a false positives rate of 36.1% and a false negatives rate of 21.9% is produced by this test.

Image	False Positives	False Negatives
1	11.9 %	19.3 %
2	0.1 %	14.8 %
3	18.6 %	45.1 %
4	8.4 %	10.6 %
5	42.6 %	41.8 %
6	20.9 %	14.3 %
7	61.0 %	7.0 %
8	0.5 %	12.6 %
9	0.2 %	55.1 %
10	24.9 %	4.6 %
11	0.9 %	18.5 %
12	89.7 %	14.2 %
13	0.5 %	8.6 %
14	90.0 %	44.0 %
15	74.2 %	3.4 %
16	4.7 %	45.3 %
17	72.7 %	25.0 %
18	64.1 %	8.5 %
19	23.4 %	21.2 %
20	76.9 %	20.6 %
21	96.2 %	20.1 %
22	27.2 %	9.3 %
23	2.3 %	8.9 %
24	97.4 %	16.8 %
25	25.8 %	15.5 %
26	0.1 %	14.3 %
27	13.5 %	58.5 %
28	37.5 %	10.1 %
29	59.3 %	47.5 %
Average	36.1 %	21.9 %

Table 7: False positives and false negatives rate of the *leave-one-out* test with a set of 29 images.

4.6 Second Iteration of Classification

We tested the possibility to classify the hemangioma a second time to get better results. The idea was to obtain better normalization values, if we take the mean color value of the bounding box surrounding the regions segmented in the first iteration (without the color values of the already segmented regions). This should give us a good estimation of the skin color near the borders of the hemangioma. We have tested this method with the images 1-15 of Table 4. The results are shown in Table 8.

Image	False Positives	False Negatives
1	12.0 %	17.4 %
2	0.2 %	10.8 %
3	1.1 %	37.5 %
4	0.6 %	22.8 %
5	6.9 %	32.1 %
6	5.4 %	4.2 %
7	41.3 %	5.5 %
8	0.6 %	2.8 %
9	0.3 %	15.0 %
10	3.3 %	5.8 %
11	1.8 %	12.8 %
12	58.0 %	3.0 %
13	0.4 %	5.9 %
14	11.8 %	51.9 %
15	63.6 %	1.9 %
Average	13.8 %	15.7 %

Table 8: False Positives and false negatives rate achieved by a second iteration of the segmentation step. Significant changes compared to Table 4 are marked blue.

Due to the received results the idea was rejected. Both the average false positives and the false negatives rate is increased from 11.6% to 13.8% and from 12.9% to 15.7%. The results for particular images are similar to those of the first iteration, only image 7 evinces an improvement and images 4 and 11 are classified significantly worse. The different segmentation results of these three images are shown in Fig. 14. By normalizing image 7 [Fig.14(a)] with the new values we get a better segmentation result since the normalization values correspond more to the reddish skin near the hemangioma [Fig.14(b)]. The main problem of this procedure is that there are also included reddish color values at the hemangioma border in the bounding box and therefore the new values do not represent the skin color accurately in most cases. By this reason the second iteration segmentation of image 4 [(Fig.14(c))] does not detect the whole hemangioma [(Fig.14(d))]. Another problem is caused by falsely classified image regions causing an improper bounding box. This can be seen by the second segmentation iteration of image 12 [(Fig.14(e))]. The computed bounding box is too large because of a misclassified small region at the left side of the image. Therefore, by the second iteration new regions belonging to skin are added to the hemangioma segmentation [(Fig.14(d))].

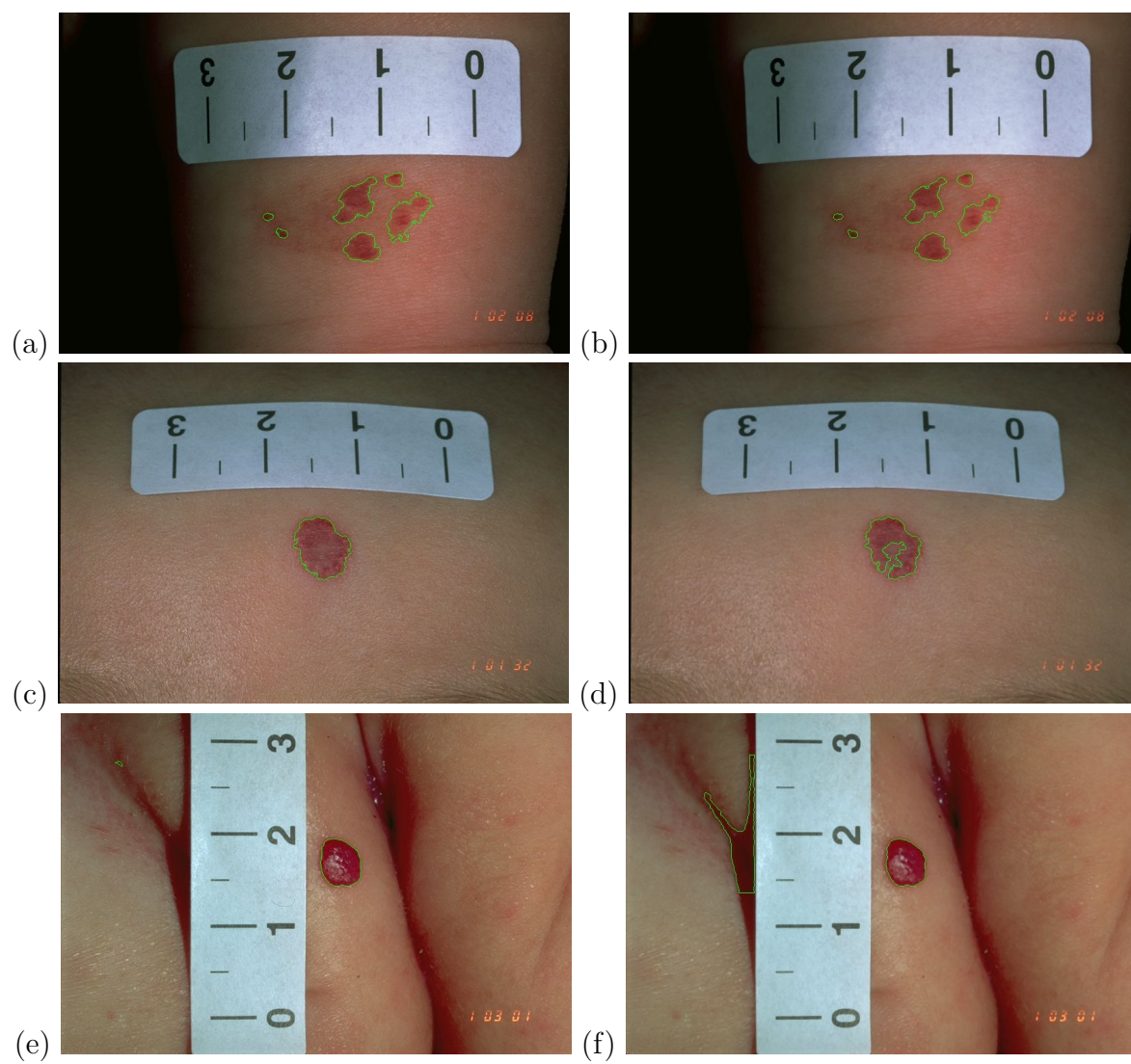


Figure 14: Three images segmented with a second iteration and without a second iteration.

5 Conclusion

For analyzing the course of evolution of cutaneous hemangiomas an automatic and fast method for determining their area in photographs is very useful. The algorithm proposed first computes the spatial resolution of an image by determining the euclidean distance between two marks of a ruler contained in the images. Then a hemangioma segmentation is accomplished by a single-layer perceptron classification by means of the pixel's color values. In a postprocessing step holes caused by highlights are closed and each segmented region is checked for various properties to exclude falsely classified regions.

Experimental results show that the segmentation works precisely on the majority of the images. Bad segmentation results are primarily caused by large regions with hemangioma-like color values that are not be rejected in the postprocessing step either. For these images a workaround could be to manually cut out the parts of the image containing these regions. Inaccuracies in computing the hemangioma area are mainly caused by the fact that often the images are not photographed from an optimal position and also do not contain an optimally placed ruler since the algorithm does not consider the spatial position of both the hemangioma and the ruler in the images. For an optimal result both hemangioma and ruler should be situated on a planar surface parallel to the image plane. Therefore, better results can be expected by photographing the hemangiomas from an angle of 90 degrees and placing the ruler as near as possible to the hemangioma. Generally an improvement of the algorithm's overall performance can be assumed by a better image acquisition process. This includes beside an optimal camera position and ruler placement an adequate illumination of the hemangiomas and the use of a digital instead of an analog photo camera.

Acknowledgements

The author would like to thank Dr. Harald Maier from the *Vienna General Hospital* for providing the test images and Prof. R. Sablatnig and G. Langs for their help and support.

References

- [1] T.M. Cover and P.E. Hart. Nearest neighbor pattern classification. *IT*, 13(1):21–27, January 1967. [8](#)
- [2] B.A. Drolet, N.B. Esterly, and I.J. Frieden. Hemangiomas in children. *New England Journal of Medicine*, 341:173–181, July 1999. [1](#)
- [3] H. Ganster, P. Pinz, R. Rohrer, E. Wildling, M. Binder, and H. Kittler. Automated melanoma recognition. *IEEE Transactions on Medical Imaging*, 20(3):233–239, March 2001. [3](#)
- [4] A. Green, N. Martin, J. Pfitzner, M. O'Rourke, and N. Knight. Computer image analysis in the diagnosis of melanoma. *J. of the American Academy of Dermatology*, 31(6):958–964, December 1994. [3](#)

- [5] Whenzao Guo and Y. Alp Aslandogan. Mining skin lesion images with spatial data mining methods. Technical Report CSE-2003-19, Department of Computer Science and Engineering, University of Texas at Arlington, July 2003. [5](#)
- [6] G.A. Hance, S.E. Umbaugh, R.H. Moss, and W.V. Stoecker. Unsupervised color image segmentation with application to skin tumor borders. *IEEE Engineering in Medicine and Biology*, 15(1):104–111, January/February 1996. [9](#)
- [7] S.L. Horowitz and T. Pavlidis. Picture segmentation by a tree traversal algorithm. *Journal of the ACM*, 23(2):368–388, April 1976. [5](#)
- [8] J.M. Kasson and W. Plouffe. An analysis of selected computer interchange color spaces. *ACM Transactions on Graphics*, 11(4):373–405, October 1992. [13](#)
- [9] F. Rosenblatt. The perceptron: A perceiving and recognizing automaton. Report 85-460-1, Project PARA, Cornell Aeronautical Laboratory, Ithaca, New York, January 1957. [7](#)
- [10] A.J. Round, A.W.G. Duller, and P.J. Fish. Colour segmentation for lesion classification. In *IEEE Proceedings of the Engineering in Medicine and Biology Society*, volume 2, pages 582–585, October/November 1997. [5](#)
- [11] J. C. Russ. *The Image Processing Handbook*. CRC Press, Boca Raton, FL, 2nd edition, 1995. [17](#)
- [12] P. Schmid and S. Fischer. Colour segmentation for the analysis of pigmented skin lesions. In *Proceedings of the Sixth International Conference on Image Processing and its Applications*, volume 2, pages 688–692, July 1997. [5](#)
- [13] F.A. Wirth and M.H. Lowitt. Diagnosis and treatment of cutaneous vascular lesions. *American Family Physician*, 57(4):765–773, February 1998. [4](#)
- [14] L. Xu, M. Jackowski, A. Goshtasby, D. Roseman, S. Bines, C. Yu, A. Dhawan, and A. Huntley. Segmentation of skin cancer images. *Image and Vision Computing*, 17(1):65–74, January 1999. [4](#), [15](#)