

# Automatic Image-Based Assessment of Lesion Development during Hemangioma Follow-Up Examinations

Sebastian Zambanini <sup>a,\*</sup>, Robert Sablatnig <sup>a</sup>, Harald Maier <sup>b</sup>,  
Georg Langs <sup>c,d</sup>

<sup>a</sup>*Computer Vision Lab, Institute of Computer-Aided Automation, Vienna University of Technology, Favoritenstr. 9/183, A-1040 Vienna, Austria*

<sup>b</sup>*Department of Dermatology, Medical University of Vienna, Waehringer Guertel 18-20, A-1090 Vienna, Austria*

<sup>c</sup>*Computational Image Analysis and Radiology Lab, Department of Radiology, Medical University of Vienna, Waehringer Guertel 18-20, A-1090 Vienna, Austria*

<sup>d</sup>*Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, 32, Vassarstreet, MA 02139 Cambridge, USA*

---

## Summary

*Objective:* This paper presents an automatic method for the quantification of the development of cutaneous hemangiomas in digital images. Two measurements on digital images acquired during follow-up examinations are performed: (1) the skin area affected by the lesion is measured and (2) the change of the hemangioma during follow-up examinations called regression is determined. Current manual measurements exhibit inter- and intra-reader variation, which impedes precision and comparisons across clinical studies. The proposed automatic method aims at a more accurate and objective evaluation of the course of disease than the current clinical practice of manual measurement.

*Methods and Material:* The proposed method classifies individual pixels and calculates the area based on a ruler attached to the skin. For the regression detection follow-up images are registered automatically based on local gradient histograms. The method was evaluated on 90 individual images and a set of 4 follow-up series consisting of 3-4 examinations.

*Results:* The absolute average error of the individual area measurements lies at  $0.0775 \text{ cm}^2$  corresponding to a variation coefficient of 8.82 %. The measurement of the regression area provides an absolute average error of  $0.1134 \text{ cm}^2$  and a variation coefficient of 7.40 %.

*Conclusions:* The results indicate that the proposed method provides an accurate and objective evaluation of the course of cutaneous hemangiomas. This is relevant

for the monitoring of individual therapy and for clinical trials.

## 1 Introduction

### 1.1 Objective

Cutaneous hemangiomas are benign tumors consisting of newly-formed blood vessels in the skin. They are the most common benign vascular tumors during infancy with a frequency of about 10 % [1]. Due to their potency for rapid proliferation they can threaten vital structures by tumor compression or tumor obstruction and can even impair vital functions such as breathing, vision, hearing, ingestion or excretion [2]. Safe and effective treatment at the earliest possible time can stop further proliferation, induce regression and prevent complications [3].

During treatment and during clinical trials that compare different treatment strategies, the accurate monitoring of the progress of hemangioma growth and regression is essential. More accurate measurements of the development provide better feedback during therapy and allow for more significant results given the limited number of patients included in clinical trials.

In Fig. 1 an example of a base-line and two follow-up examinations are depicted. At the first examination (Fig. 1(a)), the bigger part of the hemangioma shows a typical reddish saturated color caused by the excess of blood vessels and almost no regression has started yet, except for a small region indicated by a so-called *graying* in the center of the hemangioma. At the second examination 8 weeks later (Fig. 1(b)), the regression covers almost the entire hemangioma. 13 weeks after base-line (Fig. 1(c)), the entire hemangioma is regressing, and healthy skin regions start to appear.

In current clinical practice, the hemangioma area is estimated by a manual measurement of the height and width of the lesion during a clinical examination.

---

\* Corresponding author. Tel.: +43 1 58801 18366; Fax: +43 1 58801 18392  
*Email addresses:* [zamba@caa.tuwien.ac.at](mailto:zamba@caa.tuwien.ac.at) (Sebastian Zambanini),  
[sab@caa.tuwien.ac.at](mailto:sab@caa.tuwien.ac.at) (Robert Sablatnig), [harald.maier@meduniwien.ac.at](mailto:harald.maier@meduniwien.ac.at)  
(Harald Maier), [langs@csail.mit.edu](mailto:langs@csail.mit.edu) (Georg Langs).  
*URL:* <http://www.caa.tuwien.ac.at/cvl/people/zamba/> (Sebastian Zambanini).

The degree of regression is reported by a visual estimation of the fraction of the hemangioma showing a graying of the typical reddish hemangioma color. Evidently, this way of assessment is inaccurate since only the diameters and not the - typically irregular - shape of hemangioma are considered in its surveying. Furthermore, the assessment of regression degree is a subjective measurement since it depends on the rating of an individual dermatologist.

In this paper we propose a fully automatic method for the quantitative assessment of cutaneous hemangioma during therapy. Based on digital images that are currently obtained solely for reporting reasons, two measurements are performed:

- (1) The precise measurement of the area affected by the hemangioma.
- (2) The detection and quantification of regressing (healing) regions during follow-up examinations.

The aim of the work is to provide dermatologists with accurate and precise measurements of hemangioma size and temporal development. Currently, visual scoring is performed by specialists using a small number of reference values [3] that are related to but do not measure the actual extent of lesions. In the proposed approach, the actual area covered by the hemangioma is measured and compared across a temporal sequence of follow-up examinations. A registration of the hemangioma allows for a point by point comparison at different acquisition times, instead of global scores that decrease the local specificity of the assessment.

The benefit of this assessment is twofold: First, it can be used in clinical studies evaluating the efficacy of different treatments for hemangiomas. In this context, the significance of the results is increased by a more accurate and objective evaluation. And second, the assessment can be used during treatment of individual patients to evaluate the efficacy of treatments with increased precision and the ability to perform local comparisons of the lesion appearance.

The method works based on digital images with minimal constraints with regard to the image acquisition (Fig. 1). The hemangioma is detected by an automatic image segmentation based on a classifier trained in a supervised manner. The scale is retrieved from the reference ruler visible in the images. This procedure surveys the clearly reddish, unregressed regions of the hemangioma and can be applied to single images. The second part, the detection and surveying of regressing regions describes a change over time. It therefore includes a comparison of the current hemangioma appearance to the base-line examination. The two images are registered robustly using feature points and local gradient histograms. This makes a comparison of hemangiomas possible, even if images are acquired from different viewpoints, which typically is the case.

## 1.2 Related Work

To our knowledge, no research has been published on the image-based assessment of cutaneous hemangiomas. Nonetheless, various other types of skin lesions have been investigated in this context. In particular, the automatic image-based classification of melanoma in benign and malignant lesions has been studied extensively starting as early as 1987 [4]. An overview of work published until the year 2000 is given in [5]. Recent approaches are summarized in [6]. Classification of melanoma is thereby typically based on the *ABCD* rule (**A**symmetry, **B**order irregularity, **C**olor variegation and **D**iameter) used by dermatologists [7].

Another skin lesion type covered in the literature is psoriasis [8]. In psoriasis assessment, a psoriasis area and severity index [9] is used to evaluate the degree of disease. A variable thresholding technique for area computation as part of this assessment method is proposed by Roening et al. [10]. The automatic change detection in follow-up psoriasis images is addressed in Maletti and Ersbøll who propose a method for change detection in both registered [11] and unregistered images [12], each one by applying the multivariate alteration detection transform on the data. In [13] two other methods for change detection are explored: simple image subtraction and principal component analysis. A method for area computation using active contour models is presented in [14] and applied on images of leg ulcers.

However, the described methods cannot be applied directly to cutaneous hemangiomas, because the single measurement of affected skin area cannot be used as a complete assessment method for this lesion type, since the regression process starts with color changes (graying) in certain hemangioma regions and not with their complete disappearance. The change detection scheme proposed by Maletti and Ersbøll [11,12] shows suitable results on psoriasis images but is designed for the detection of textural changes, whereas changes in hemangioma images are indicated by subtle color changes during follow-up.

Although experiments reported in this paper were only conducted on hemangioma images, the method is applicable to other lesion types as well. Both the segmentation of the affected skin area and the detection of regression regions are based on a pixel-wise classification by a previously trained classifier, thus the method could be adapted to other lesion types if an adequate amount of annotated training data is available. The registration and localized analysis of skin appearance development can be transferred to other diseases in a straightforward way.

The remainder of the paper is organized as follows: the method is presented in Section 2, divided into the steps of preprocessing (Section 2.1), hemangioma segmentation (Section 2.2), follow-up registration (Section 2.3), regression detection

(Section 2.4) and scale computation (Section 2.5). Experimental evaluation results for both area measurement and follow-up regression detection are reported in Section 3 and discussed in Section 4. A conclusion is given in Section 5.

## 2 Automatic Hemangioma Assessment

The overall workflow for the assessment of hemangioma development is illustrated in Fig. 2. Here four follow-up images of a specific hemangioma are given. The determination of the hemangioma area and the measurement of regions showing a regression is illustrated for image 4 in the series. In Fig. 2 the hemangioma region is indicated by a white area whereas the expanding regression during follow-up is indicated by a gray area. The following steps are conducted:

- (1) **Hemangioma Segmentation:** The image regions of Image 1 belonging to the hemangioma, called *hemangioma regions* are determined. In Fig. 2, this segmentation is indicated by the black border around the white area.
- (2) **Registration and Transformation:** The transformation aligning Image 1 with Image 4 defines a region of interest (ROI) corresponding to the expected location of the hemangioma in Image 4. This new region is again indicated by a black border. It is calculated based on the hemangioma segmentation in Image 1 and the transformation of Image 1 to Image 4.
- (3) **Regression Detection:** The regions exhibiting regression (the *regression regions*) are determined in the ROI, facilitated by a local comparison of color values between Image 4 and the transformed Image 1.
- (4) **Image Scale Computation:** The ruler visible in Image 4 is used to determine the image scale. By combining the information of image scale and regression/hemangioma regions, the current area of both the hemangioma and the regression regions is obtained.

### 2.1 Preprocessing

To account for the varying image acquisition conditions (e.g. pose of the patient, illumination, visible section of the body part, changing perspective) first a pre-processing is performed. To remove noise a 5x5 median filter is applied to the images. Furthermore, image regions containing no skin are detected and are excluded from further computation. The remaining image is normalized to obtain consistent skin color over the data set.

**Non-skin masking:** After the median filtering a simple test for masking out

non skin regions is used to exclude regions that likely are not part of the skin or the hemangioma (e.g. the ruler). This step is necessary for a robust skin color normalization. Our method is based on a heuristic proposed in [15] but substantially simpler. If the *red* ( $R$ ), *green* ( $G$ ), and *blue* ( $B$ ) channel values of each pixel fulfil  $R < G$  and  $R < B$  the pixel is marked as *non-skin*.

**Skin color normalization:** To account for changing illumination a normalization with regard to the skin color is performed. The most frequent color is determined by a histogram of the RGB color values present in the skin region, i.e. the most frequent value with intensity  $\mathbf{v}_0 = \langle R, G, B \rangle$  with  $\sqrt{R^2 + G^2 + B^2} > 120$  is used for normalization. That is, for each color value  $\mathbf{v}$  the normalized value is  $\mathbf{v}' = \mathbf{v} - \mathbf{v}_0$ . This procedure uses the large area of skin typically visible in the data for a color normalization and ensures a consistent color representation for the further processing of the images. Tests have shown that it is of equal accuracy as manual normalization.

## 2.2 Hemangioma Segmentation

The segmentation determines the regions in an image belonging to the hemangioma. Image segmentation algorithms usually subdivide images into regions which show a certain degree of coherence w.r.t. color or texture or use a priori knowledge about the shape of the object to be detected. Since hemangiomas can come in a variety of shapes and no assumption about their number and sizes for a given image can be made only the a priori knowledge of their color can be exploited. Commonly used color image segmentation algorithms usually deliver a set of connected regions where it is challenging to discern between hemangioma and non-hemangioma regions. Therefore, we use a classifier for segmenting the images that classifies each pixel in the image as hemangioma or non-hemangioma on the basis of features extracted from the pixel. A single-layer perceptron [16] is used for classification.

The single-layer perceptron uses four color features for classification. It is trained on an annotated training set of images and applied to the new images. This results in a label for each pixel, indicating its belonging or non-belonging to the hemangioma. In the following the features used for classification are described and a final postprocessing step eliminating highlight artifacts is explained.

**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 the color channels of the color spaces RGB, HSV and CIE 1976 L\*a\*b\* [17].  $G$  from RGB,  $H$  from HSV and  $a^*$  from L\*a\*b\* proved to be usable

for our purpose by achieving the best results with a perceptron classification of our images. Additionally we use a 4<sup>th</sup> feature *abdist*. As can be seen in Fig. 3, each of these features exhibits extensive differentiation between pixels belonging to the hemangioma and pixels belonging to the skin.

**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 [18]. In this paper the proposed method works on an intensity image describing the Euclidean distance between the skin and the lesion. If  $a_s, b_s$  denotes the  $a^*$  and  $b^*$  values of the skin (obtained from the normalization step) and  $a_p, b_p$  that from a particular pixel, its *abdist* is computed as  $\sqrt{(2a_s - 2a_p)^2 + (b_s - b_p)^2}$ . The difference of the  $a^*$  channel is multiplied by the factor 2, because the  $a^*$  value differs more between hemangioma and skin pixels than the  $b^*$  value.

**Treatment of highlights:** Highlights on the hemangioma are normally erroneously detected as healthy skin by the classifier. This is corrected by closing all holes occurring in the masked region. Since hemangiomas with large holes of normal skin could not be found in the data, they seem to be very rare and a possible error resulting from that operation can be neglected.

### 2.3 Registration of Follow-Up Hemangioma Images

The registration of follow-up images is a necessary prerequisite for the detection of regressions (see Section 2.4). It allows for the establishment of correspondences across the image series and thereby enables the propagation of the initial hemangioma shape to subsequent images. This makes a definition of a ROI for the regression detection possible, allows for shape and area comparisons, and finally for a direct comparison of corresponding positions on the hemangioma and according color values. This is essential for the detection of regression regions exhibiting subtle color differences.

#### 2.3.1 Robust Feature-Based Method for Registration of Follow-Up Hemangioma Images

We prefer a feature-based approach to an area-based one [19] since area-based methods are not well-adapted for the registration of changing structures like hemangiomas due to low similarities of corresponding image points in changing image regions of follow-up images. Feature-based methods are more suitable in this

case since they allow a registration also for missing correspondences by predicating the registration only on non-degraded image parts. Therefore, our approach is based on the detection and matching of distinctive interest points by means of local features. The resulting interest points are used for a robust estimation of the transformation between images by random sample consensus (RANSAC) [20]. To obtain reliable matches of interest points we use gradient histograms as used in the scale-invariant feature transform (SIFT) [21] for the description of interest point appearance and determination of correspondences between images. Under the assumption that hemangiomas are on a planar surface, the transformation between images can be simplified to a homography and estimated by the detected point correspondences. Note that the planarity assumption can be generalized to more complex surfaces. However, experiments indicate that it provides sufficient accuracy on our dataset. The use of RANSAC makes the registration more insensitive to the number of correct matches. Therefore, the risk of misregistration caused by a high amount of incorrect matches due to changing hemangioma structures is minimized.

### 2.3.2 Detection of Interest Points in Hemangioma Images

After segmenting the hemangioma in both images, first distinctive points are detected in the hemangioma region. For this, Canny edge detection [22] is performed on the green color component.

**Constraining interest points to the hemangioma region:** The analysis is constrained to the segmented hemangioma region, provided by the method described in Section 2.2. Only interest points inside or near the hemangioma region are accepted. This ensures that the planarity assumption for the homography estimation gives a good estimate of the actual surface, accounts for a better chance of reliable matches in the close vicinity of the hemangioma and reduces computational costs. Fig. 4(a) shows the segmentation of the hemangioma area (green border) and resulting rectangular ROI (blue frame).

**Interest point localization:** The most reliable interest points lie at the hemangioma border because the inner hemangioma parts change to a higher degree from one time to another. Therefore, in our method interest points are detected along edges in the green channel of the RGB-images. The green channel is chosen since for hemangiomas it shows the highest differences among the three RGB color channels between lesion regions and healthy skin (see Section 2.2 and Fig. 4(b)). Edge detection is accomplished by means of the Canny edge detector [22] and interest points are finally localized at edge pixels showing the highest gradient magnitude in a neighborhood of  $N$  pixels. By empirical tests a value of  $N = 3$  combined with low threshold of 0.1 and a high threshold of 0.2 (see [22]) has

proven to be adequate for obtaining a high number of distinctive interest points. In Fig. 4(d) the detected interest points are marked as black spots.

### 2.3.3 Matching of Interest Points

Once the interest points have been localized, a mathematical description of the interest points' local neighborhoods is provided by gradient histograms as used in SIFT descriptors [21] in order to match the interest points. The interest points are described by accumulating the orientations in a region around the interest point location. Gradient magnitudes and orientations are sampled in a  $16 \times 16$  array around the interest point location and weighted by a Gaussian window. The size of each array sample was empirically set to  $2.4 \times 2.4$  pixels, i.e. each descriptor's image patch has a size of  $38.4 \times 38.4$  pixels. By computing 8 orientation bins for every  $4 \times 4$  subarray of the overall  $16 \times 16$  array, an interest point is finally described by a 128-element feature vector.

Although RANSAC is capable of handling a large portion of incorrect matches, a preselection of the matches with highest confidence can improve the stability. Therefore, matches are determined by means of lowest Euclidean distance of interest point descriptors. All matches are sorted in terms of the distance between the nearest and the second nearest neighbor in a descending order, and finally only the first  $n$  matches are accepted. Based on initial experiments,  $n$  is determined by  $n = 2\sqrt{m_1}$ , rounded to the nearest integer value, where  $m_1$  is the number of interest points detected in Image 1. Fig. 5(a) shows exemplarily the matching result between the image from Fig. 4 and the corresponding image from the following examination.

### 2.3.4 Robust Transformation Estimation Based on Interest Point Matches

The final step in the image registration procedure is the computation of the homography that maps the sensed image onto the reference image. The homography is computed from the pairs of matched interest points and assigns each point in the hemangioma of the sensed image to a point in the reference image. Part of the matches detected by comparing local appearance are incorrect or *outliers*, and have to be detected and discarded. A homography can be estimated by the normalized direct linear transform (DLT) which minimizes the algebraic error [23]. Since every match is equally considered for homography estimation by the DLT algorithm, it is not robust against outliers. To detect outliers the RANSAC scheme [20] is applied for the homography estimation. RANSAC is a robust model fitting method which is able to cope with a large portion of incorrect data samples.

In Fig. 5 the final results for two consecutive images from a follow-up series are shown. Fig. 5(b) shows the remaining inliers determined by RANSAC of the initial matches depicted in Fig. 5(a). In this example 31 of the 76 initial matches are classified as inliers. The final transformed sensed image can be seen in Fig. 5(c). The difference between the transformed image and the reference image is shown in Fig. 5(d).

#### 2.4 Detection of Regressing Hemangioma Regions

While at the beginning hemangiomas typically exhibit a saturated red color, during the course of therapy regressions appear as pale gray regions, a process often referred to as *graying* [2]. Typically, these regressions do not occur uniformly on the whole hemangioma, but start at certain regions and expand during time [2]. The detection and size estimation of such regions is an important parameter to assess the healing process.

The image of Fig. 6(a) was taken at the base-line examination and the entire hemangioma is a saturated reddish color, except for a small inner region. Image Fig. 6(b) was taken 6 weeks later and yet the hemangioma shows regressions in its inner regions. These regressions are marked with a white border in Fig. 6(c). The regression detection relies on the pixel-wise classification and a comparison to the base-line examination. The registration (Section 2.3) provides the necessary correspondence for a pixel-wise comparison of follow-up images. Differences between regressing and not regressing regions are less distinctive than for the segmentation of the hemangioma area (see Fig. 3 for a comparison). Nevertheless, the experiments in Section 3.5 show that the method achieves a similar accuracy.

The procedure of the proposed method for the detection of regressing hemangioma regions is depicted in Fig. 7: consider a series of four follow-up images (Image 1 to Image 4). Between each consecutive image a homography is computed by the registration method described in Section 2.3.1 ( $H_{12}$ ,  $H_{23}$  and  $H_{34}$ ). In the first step the region encompassing the hemangioma is determined by the segmentation method of Section 2.2 (white area). Next, the mask defining the hemangioma region is transformed with the composite homography  $H = H_{34}H_{23}H_{12}$  to define the ROI for the classification process. In the last step, in this hemangioma region all pixels are classified as *regressing* (gray area in Fig. 7) and *not regressing*, using the differences in color values between Image 1 and Image 4 as additional features whereas the mapping between pixel positions on the hemangioma is computed via the composite homography  $H$ . In this context, *regressing* includes (already healed) healthy skin as well. However, a separation between regressing and healthy skin can be made by the hemangioma segmentation algo-

rithm (see Section 2.2).

The images in Fig. 6 illustrate that the differentiation between non-regressing and regressing regions is far less pronounced than the difference between skin and hemangioma. To deal with varying illumination conditions during image acquisition all features are skin-normalized by the method presented in Section 2.1. For feature selection the method proposed by Krizek et al. [24] was used. Their algorithm removes irrelevant and redundant features by a weight modification of training samples similar to the AdaBoost algorithm [25]. As a result, the following four features were selected: (1) the green channel from the RGB color model, (2) the a\*-channel from the L\*a\*b\* color model, (3) abdist (see Section 2.2), and (4) the difference between the green channel image from the base-line examination and the green channel image from the current examination. In Fig. 8 the values of these chosen features are shown for the image of Fig. 6(c), normalized to the range 0 to 1 for presentation (black represents 0, white represents 1). A multi-layer perceptron [26] is used for the classification.

### 2.5 Computation of the Image Scale

All images show a ruler close to the hemangioma. The ruler has 4 bold lines in 1 cm distance steps (see Fig. 1). The task of the algorithm is to compute the Euclidean distance between two lines to obtain the scale of the images. First the ruler is segmented by global thresholding. Afterwards, three scanlines are defined at the midpoint between the top and bottom of the ruler and parallel but 10 pixels to both sides of the center line. Finally, the maximum distance measured between two marks is taken to compute the image scale.

## 3 Experiments

### 3.1 Data Description

The images used in this paper were acquired by the *Division of Special and Environmental Dermatology* at the *Medical University of Vienna*. The study was approved by the ethics committee of the Medical University of Vienna and the parents of all participating children signed informed consent. The photos were acquired using an analog camera and digitized with a scanner. All images have a resolution of 512x768 pixels and a bit depth of 8 bits per color channel. In total the data set contains images from 120 examinations. 24 images are part

of 6 different cases consisting of 3-5 follow-up examinations (1 consisting of 3, 4 consisting of 4 and 1 consisting of 5 follow-up examinations). Additionally, for 35 examinations, a second image depicting the same hemangioma and taken within a few minutes is available. This is used for an evaluation of the method’s precision by measuring the variation of achieved measurements on these image pairs (see Section 3.3 and 3.6).

### 3.2 Accuracy of the Hemangioma Segmentation Algorithm

**Setup:** Experiments were performed on the whole set of 120 images gathered during clinical examinations. Manually annotated ground truth data (standard of reference) was provided by a dermatologist. 30 randomly selected images were used for training and the remaining 90 images served as the test set. To evaluate the accuracy of the segmentation for every image the following error metrics were applied: (1) the error rate, (2) false positive rate, (3) false negative rate, (4) absolute area difference, and (5) border error, defines as  $Area(A \cup M) - Area(A \cap M) / Area(M)$ , where  $A$  and  $M$  are the regions obtained by the automatic segmentation and the manual segmentation, respectively. The border error [15] is the most significant error metric because it is independent of both the size of the hemangioma and the size of the image. Note that the manually annotation of the data can not be seen as absolute gold standard since the annotations of only one dermatologist were available and variation between specialists should be assumed. To assess this point, an evaluation of the precision of the automatic measurement of hemangioma size is reported in Section 3.3. It captures both variability due to repeated annotation and due to repeated image acquisition.

**Results:** The segmentation of all 90 images results in a 6.8% average segmentation error rate, with 5.5% false positives, 11.6% false negatives, and absolute area difference of  $0.0965cm^2$ , and a border error of 32.1%. Generally, the obtained average border error of 32.1 % is hardly influenced by few particular outliers with border errors of more than 100 %. However, the majority of the images (54 out of 90) could be segmented with a border error of less than 20 % and only 15 images yield an error of more than 50 %. In Fig. 9 some results are depicted. Fig. 9(a)-(c) belongs to the best segmentation results with border errors of 3.6 %, 5.7 % and 6.8 %, respectively. Fig. 9(d)-(f) belongs to the worst segmentation results with border errors of 247.7 %, 137.5 % and 141.2 %, respectively.

### 3.3 Precision of Automatic Measurement of Hemangioma Size

**Setup:** To assess the precision of the entire procedure of hemangioma size measurement the hemangioma area on 20 pairs of images, depicting the same hemangioma and taken within a few minutes, were measured. In the absence of error, both images should have precisely the same computed hemangioma area. The actual error is estimated by the absolute area difference and variation coefficient of both measurements. The variation coefficient of multiple measurements is defined as their standard deviation divided by their mean. To have an estimate of the minimal achievable error, given the differences in image acquisition, reference measurements with manual segmentation and scale computing were made on the same image pairs.

**Results:** The average difference of hemangioma area is  $0.0775 \text{ cm}^2$  while the average variation coefficient is 8.82 %. The reference measurements with manual segmentation and scale computing lead to an average difference of  $0.0394 \text{ cm}^2$  and an average variation coefficient of 4.53 %.

### 3.4 Registration of Follow-Up Hemangioma Images

**Setup:** To assess the accuracy of the proposed registration method for follow-up images we applied our algorithm on four of the six different image series. Two series (consisting of 4 and 5 examinations, respectively) were excluded from this and the following experiment of regression detection (see Section 3.5) for two reasons: (1) image quality impedes a consistent manual annotation of regressing regions and (2) hemangioma changes are too high to obtain a registration through the automatic or manual definition of corresponding points (due to missing acquisitions and thereby long intervals between follow-up images). Therefore, each image series used consists of three to four images, resulting in a total of 11 registrations marked by two characteristics, patient (1,2,3,4) and position in the sequence of registrations (A,B,C): 2A shows the registration of the images from the first and the second examination, 2B indicates the registration of the images from the second and the third examination, and so on. The registration error is measured by manually defining 5 matching points for each follow-up image and measuring the average pixel distance achieved with the estimated homography. To measure the consistency of three circularly concatenated registrations, an additional error metric called *average reference point displacement* is reported which uses two images from an examination (Image  $A$  and Image  $A'$ ) and one from the subsequent examination (Image  $B$ ). The three homographies  $H_{A,B}$  (from image  $A$  to image  $B$ ),  $H_{B,A'}$  and  $H_{A',A}$  between the images are computed and the composite

homography  $H_{A,A} = H_{A',A}H_{B,A'}H_{A,B}$  is built. In the absence of error,  $H_{A,A}$  is the identity matrix and every point in Image  $A$  is not displaced by a transformation with it. Inevitably, there is an error which can be measured by the average displacement of points in Image  $A$ . Therefore, a set of reference points is uniformly distributed, separated by 10 pixel distances inside the hemangioma region of Image  $A$  and transformed by the composite homography  $H_{A,A}$ . The error is computed as the average Euclidean distance between reference points and their corresponding transformed points. This error is not stated for the registration 2C because in this case only one image from examination 1 is available. In our experiments best results were achieved by iterating 2000 times and allowing a maximum distance of 5 pixels for the inlier decision (in a typical image 5 pixels correspond to  $\sim 0.375mm$ ).

**Results:** The individual and average errors of the registrations are listed in Table 1. The average distance error of 5 test points and the average reference points displacement never exceeds 10 pixels ( $\sim 0.8mm$ ).

Fig. 10 shows the sensed image, the reference image and the difference image between the transformed sensed and the reference image for the follow-up images of patient 1. Since the appearance of a hemangioma changes over time, some regions show higher differences than others. For example, in Fig. 10(1C) differences occur at the regressing regions of the reference image.

### 3.5 Accuracy of Regression Detection on Follow-up Images

**Setup:** For the evaluation of the regression detection procedure the algorithm was tested on the four follow-up series described in Section 3.4. For every follow-up image the absolute area difference and relative border error, compared to manually annotated standard of reference, was measured. For an evaluation of the whole dataset a leave-one-out cross-validation scheme was applied: for every image of a series a new classifier was trained with the data from all series except the one containing the present image and tested on this image. As described in Section 2.4, a multi-layer perceptron classifier was used. The network consisted of 2 hidden layers with 20 and 10 hidden units, respectively. The network was trained by quasi-newton optimization [26] and training performance was measured by the mean squared error.

Since the overall goal of the proposed method is the automatic determination of the area/percentage of the hemangioma showing regression signs, these values are also reported for the same image series. They are listed separately in Table 2. As discussed in Section 2.4, the change of hemangioma color values is a discriminative feature for the classifier. To verify this, the experiment was repeated without using

difference features, i.e. only G, a\* and abdist were used.

**Results:** The results of the experiment are shown in Table 1 and Table 2. Fig. 11 shows the results obtained on the single images. Here the region of manual detection is marked by a green border whereas the automatically detected region is marked by a blue border. The largest error occurs in image (3B) with a border error of 60.2 % but this image exhibits a strong oversaturation caused by the inadequate image acquisition. For the test reporting the area/percentage of the regressing region, it can be seen that the fraction of the hemangioma showing a regression increases or remains constant in all follow-up series. The only exception is the previously mentioned poor-quality image (3B). For patient 1 the algorithm assesses already the whole hemangioma area of the first image (1A) as regressing (percentage = 100 %).

As a result of using no difference feature for the regression detection, all error values deteriorate compared to the values obtained by using registration and difference features. Border error increases from 27.1 % to 50.1 % and absolute area difference from  $0.1278 \text{ cm}^2$  to  $0.3486 \text{ cm}^2$ .

### *3.6 Precision of Regression Detection on Hemangioma Images taken at the same Time*

**Setup:** Since the manual determination of regressing regions is an uncertain and difficult task (even for dermatologists), this experiment deals with the comparison of surveyed regression regions in image pairs showing the same hemangioma, consistent with the experiment reported in Section 3.3. Again the absolute difference and variation coefficient between the two measurements are calculated. 2B was excluded from this test because no second image was available for that case.

**Results:** The values are listed in Table 3. Compared to the hemangioma segmentation results, the variation coefficient is even less (7.40 % to 8.82 %, see Section 3.3), while the absolute difference between measured regression on image pairs shows a comparable value ( $0.1134 \text{ cm}^2$  to  $0.0775 \text{ cm}^2$ ).

## 4 Discussion

As part of the hemangioma assessment procedure, the first task is to determine and measure the whole skin region affected (Section 3.2). By the design of the algorithm, errors in area computation arise due to three different reasons:

**Error in image scale computation:** Tests have shown that scale computation is accurate (average error rate of 1.55 % on 20 images), with very few exceptions where the ruler was inadequately placed on the skin or photographed. Therefore, this type of error is assumed to have the lowest impact on area measurement.

**Error in hemangioma segmentation:** Hemangioma segmentation works reliably and accurately on saturated hemangiomas with no or few regressions, as shown in Section 3.2 where 54 out of 90 test images could be segmented with a border error of less than 20 %. Note that the border error is a very sensitive measurement. In the work of Zhang et al. [27] dermatologists define a border error of less than 50 % as sufficient for melanoma segmentation and their results are comparable to our results on hemangioma images (see below). Nevertheless, our results also show that the algorithm can fail completely on regressing hemangiomas because of the lower contrast between skin and hemangioma regions. However, typically the first images during follow-up studies exhibit no or few regressions and can therefore be segmented accurately. The affected skin area in later examinations can then be determined by the proposed registration method.

**Variation in image acquisition:** For an optimal measurement both ruler and hemangioma have to be situated parallel to the view plane of the camera. Whereas this is naturally violated by the fact that hemangioma and skin are never perfectly planar, errors are also caused by an inadequate positioning of the camera. The tests made on the image pairs in Section 3.3 show that the variation coefficient achieved with the presented method (8.82 %) is close to the one achieved with a manual surveying (4.53 %). An error of  $\sim 4\%$  can be seen as the natural limitation of the method, i.e. the minimal achievable error rate.

**Registration of follow-up images:** The proposed method for image registration has to cope with the changing hemangioma appearances between examinations. This is achieved by a feature-based method using local gradient histograms to obtain reliable matches between consecutive images. The inevitable occurrence of incorrect matches caused by the varying image content is mastered by the use of a RANSAC scheme for robust homography estimation describing the transformation between the two images. Experimental results reported in Section 3.4 show that the majority of the images can be registered without considerable errors. The average reference point displacement of  $\sim 0.4mm$  accumulated by three subsequent registrations is acceptable with an average hemangioma diameter of  $\sim 14mm$ . Sources of error are (1) the assumption of hemangioma planarity due to the homography transformation model and (2) global or local misregistrations caused by strong changes in hemangioma appearance and thereby a high fraction of false matches, either solely in specific regions or on the whole hemangioma. However, the proposed method is designed to deal with changing hemangioma appearances and the experiments show that it is applicable on the given follow-up

data.

**Regression detection:** The tests on the detection of regressing regions (Section 3.5) verify that the registration is accurate enough to support the multi-layer perceptron in classification between regressing and not regressing regions by means of difference features. Nevertheless, an important aspect of regression detection is the uncertainty of the manual annotation and therefore the use of the border error as performance measure. However, by analyzing the determined absolute areas of the detected regression regions in follow-up series (see Table 2), an enlargement of regressing regions is reported. The uncertainty of manual annotation is overcome by a comparison of determined regression areas, measured on image pairs showing the same hemangioma at the same time. Because of the similar error values (variation coefficient of 7.40 % compared to 8.82 % of the hemangioma segmentation method), it is assumed that this error is again mainly caused by variations in the image acquisition procedure, i.e. different lighting conditions and camera viewpoints. Unfortunately, regression detection could be evaluated only on a limited set of four follow-up series. A larger dataset would provide support for the general conclusions made. Nonetheless, the results on these series are consistent when the image quality is high enough (consider the case 2B as an exception due to the oversaturation caused by the improper image acquisition) and therefore it is assumed that a larger dataset would not substantially decrease the accuracy of the proposed method.

**Comparison with other methods:** As stated earlier, the image-based automatic assessment of cutaneous hemangiomas and the proposed scheme of detecting and measuring regressions by the use of registered follow-up images has not been covered before. Thus, no direct comparisons to other methods can be made. However, other skin lesion types were addressed in the past and there are some comparable results despite different input data, experimental setups and algorithm designs: in [27] border error was used as performance metric for melanoma segmentation and the authors state that a border error of less than 50% is considered as sufficient by dermatologists. They report a border error of less than 50 % in 55 out of 66 images. In our results for hemangioma segmentation 75 out of 90 segmentations have a border error of less than 50 % which corresponds to the same percentage of 83 %. For the area measurement of skin lesions, in [10] a method for psoriasis surveying was developed showing an error of less than 4 % to reference measurements in all cases. Although an exact comparison cannot be made because of nonexistent reference measurements, it must be noted that in the surveying of hemangioma/regression regions we obtain higher variation coefficients of 8.8 and 7.4 %, respectively. However, it is not clear to what extent the method reported in [10] is influenced by the image acquisition.

Based on the results we expect the method to outperform manual measurements

during clinical trials. The standard approach of height and width measurement is not sufficient to account for the highly irregular shape of the hemangioma. Additionally, the determination of the regressing regions during the healing process is more consistent, since inter- and intra-reader variability is avoided by the automatic classification of the skin region.

## 5 Conclusion

In this paper a method for the automatic image-based assessment of cutaneous hemangiomas is proposed. Cutaneous hemangiomas are a skin disease that causes an increased growth of blood vessels beneath the skin in local regions. The procedure measures the current area of the affected skin and detects and tracks regions exhibiting a regression during follow-up examinations. The accurate and precise measurement of the affected skin region, and in particular the development of the hemangioma during the follow-up examinations, is crucial during clinical trials and during treatment.

The current clinical standard for the assessment of the hemangioma's healing process is the manual measurement of its height and width and a coarse estimate of the percentage showing a regression by the physician. The area estimate does not account for the irregular shape, which becomes more relevant during regression, and consequently the regression estimation can exhibit considerable inter- and intra-reader variability. The proposed measurement method is not affected by inter- or intra-reader variation and makes the assessment of hemangioma development more accurate and objective. Thereby the quality of both clinical trials and long-term studies can be improved by a more consistent evaluation of the effect of therapies. Due to the follow-up registration it can quantify even small local changes of the hemangioma. The detection of regressions is made more consistent by decreasing the subjectivity of a manual regression quantification caused by differences of opinion between observers.

In clinical practice the procedure can be integrated into a framework that allows for an approval by the clinical expert, since, in addition to the labeling, the classifier provides a real-valued output that can be interpreted as the confidence in the result. Future work will focus on this integration.

## Acknowledgements

This research has been supported by the European Union (FP-SSP5-044450 COINS) and the Austrian National Bank Anniversary Fund (12537 COBAQUO, 13497 AORTAMOTION).

## References

- [1] B. Drolet, N. Esterly, I. Frieden, Hemangiomas in children, *New England Journal of Medicine* 341 (1999) 173–181.
- [2] K. Chiller, D. Passaro, I. Frieden, Hemangiomas of infancy: clinical characteristics, morphologic subtypes, and their relationship to race, ethnicity, and sex, *Archives of Dermatology* 138 (2002) 1567–1576.
- [3] H. Maier, R. Neumann, Treatment of strawberry marks with flashlamp-pumped pulsed dye laser in infancy, *Lancet* 347 (1996) 131–132.
- [4] N. Cascinelli, M. Ferrario, T. Tonelli, E. Leo, A possible new tool for clinical diagnosis of melanoma: The computer, *Journal of the American Academy of Dermatology* 16 (2) (1987) 361–367.
- [5] G. R. Day, R. H. Barbour, Automated melanoma diagnosis: Where are we at?, *Skin Research and Technology* 6 (2000) 1–5.
- [6] M. Celebi, H. Kingravi, B. Uddin, H. Iyatomi, Y. Aslandogan, W. Stoecker, R. Moss, A methodological approach to the classification of dermoscopy images., *Computerized Medical Imaging and Graphics* 31 (2007) 362–373.
- [7] F. Nachbar, W. Stolz, T. Merkle, A. Cognetta, T. Vogt, M. Landthaler, P. Bilek, O. Braun-Falco, G. Plewig, The ABCD rule of dermoscopy: High prospective value in the diagnosis of doubtful melanocytic skin lesions, *Journal of the American Academy of Dermatology* 30 (4) (1994) 551–559.
- [8] T. P. Habif, *Clinical Dermatology*, 2nd Edition, The C.V. Mosby Company, St. Louis, 1990.
- [9] P. C. M. van de Kerkhof, On the limitations of the psoriasis area and severity index (PASI), *British Journal of Dermatology* 126 (1991) 205.
- [10] J. Roening, R. Jaques, J. Kontinen, Area assessment of psoriatic lesions based on variable thresholding and subimage classification, in: *Proc. of Vision Interface'99*, 1999, pp. 303–311, <http://w3.gel.ulaval.ca/~parizeau/vi99/PDF-files/S10/paper48.pdf> (Accessed: 10 February 2010).

- [11] G. Maletti, B. Ersbøll, Change detection in registered psoriasis lesion image patterns, Tech. Rep. IMM-Technical Report-2003-10, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby (2003).
- [12] G. Maletti, B. Ersbøll, Texture alteration detection in bitemporal images of lesions with psoriasis, Tech. Rep. IMM-Technical Report-2003-8, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby (2003).
- [13] D. Gomez, C. Butakoff, B. Ersbøll, J. Carstensen, Automatic change detection and quantification of dermatological diseases with an application to psoriasis images, *Pattern Recognition Letters* 28 (9) (2007) 1012–1018.
- [14] T. Jones, P. Plassmann, An active contour model for measuring the area of leg ulcers, *IEEE Transactions on Medical Imaging* 19 (12) (2000) 1202–1210.
- [15] G. Hance, S. Umbaugh, R. Moss, W. Stoecker, Unsupervised color image segmentation with application to skin tumor borders, *IEEE Engineering in Medicine and Biology* 15 (1) (1996) 104–111.
- [16] F. Rosenblatt, The perceptron: A perceiving and recognizing automaton, Technical Report 85-460-1, Project PARA, Cornell Aeronautical Laboratory, Ithaca, New York (1957).
- [17] J. Kasson, W. Plouffe, An analysis of selected computer interchange color spaces, *ACM Transactions on Graphics* 11 (4) (1992) 373–405.
- [18] L. Xu, M. Jackowski, A. Goshtasby, D. Roseman, S. Bines, C. Yu, A. Dhawan, A. Huntley, Segmentation of skin cancer images, *Image and Vision Computing* 17 (1) (1999) 65–74.
- [19] B. Zitova, J. Flusser, Image registration methods: a survey, *Image and Vision Computing* 21 (11) (2003) 977–1000.
- [20] M. Fischler, R. Bolles, Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the ACM* 24 (6) (1981) 381–395.
- [21] D. G. Lowe, Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision* 60 (2) (2004) 91–110.
- [22] J. Canny, A computational approach to edge detection, *Pattern Analysis and Machine Intelligence* 8 (6) (1986) 679–698.
- [23] R. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd Edition, Cambridge University Press, Cambridge, 2004.
- [24] P. Krizek, J. Kittler, V. Hlavac, Feature selection based on the training set manipulation, in: Y. Tang, S. Wang, D. Yeung, H. Yan, G. Lorette (Eds.), *Proc. of the 18th International Conference on Pattern Recognition*, Vol. 2, IEEE Computer Society, Washington, DC, USA, 2006, pp. II: 658–661.

- [25] R. Schapire, Y. Singer, Improved boosting algorithms using confidence-rated predictions, *Machine Learning* 37 (3) (1999) 297–336.
- [26] C. M. Bishop, *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford, 1995.
- [27] Z. Zhang, W. Stoecker, R. Moss, Border detection on digitized skin tumor images, *IEEE Transactions on Medical Imaging* 19 (11) (2000) 1128–1143.

Table 1

Results of the proposed method for image registration and regression detection on 4 image series with 11 registrations in total. Initial matches depicts the number  $n$  of best keypoint matches taken for the registration. Inliers depicts the number of inliers after the use of random sample consensus (RANSAC). The average distance error of 5 test points and the average reference points displacement measures the error of the registration in pixel units. Absolute area difference and border error represent the absolute and relative error of the regression detection by comparison to the standard of reference.

Image	Initial matches $n$ (before RANSAC)	Inliers (after RANSAC)	Average distance error of 5 test points (pixels)	Average reference points displacement (pixels)	Absolute area difference ( $cm^2$ )	Border error (%)
1A	45	17	7.96	2.41	0.2508	39.9
1B	49	28	3.96	8.67	0.0576	20.7
1C	52	31	6.45	5.79	0.0310	16.3
2A	40	12	3.95	3.60	0.0712	20.0
2B	33	18	4.92	3.76	0.0835	60.2
2C	28	16	5.21	N/A	0.2403	38.6
3A	36	21	3.44	2.70	0.1996	38.6
3B	45	14	4.32	6.02	0.0071	17.4
4A	26	7	8.37	7.50	0.0315	15.9
4B	39	9	8.54	8.70	0.1126	17.6
4C	61	8	5.50	6.82	0.3204	18.2
<b>Average</b>	<b>41.27</b>	<b>16.45</b>	<b>5.69</b>	<b>5.60</b>	<b>0.1278</b>	<b>27.6</b>

Table 2

Absolute and relative determined areas of regressing hemangioma regions of the images from Table 1.

Image	Regressing area ( $cm^2$ )	Total area of hemangioma ( $cm^2$ )	Percentage
1A	1.08	1.08	100.00
1B	1.38	1.38	100.00
1C	1.36	1.36	100.00
2A	0.78	1.03	75.34
2B	0.28	1.06	26.82
2C	1.09	1.11	97.84
3A	0.50	0.65	77.10
3B	0.77	0.77	99.76
4A	1.47	1.66	88.42
4B	2.11	2.20	96.10
4C	2.24	2.31	97.12

Table 3

Differences in automatically determined areas of regression between two images of the same hemangioma.

Image series	Absolute area difference ( $cm^2$ )	Variation coefficient (%)
1A	0.0906	6.18
1B	0.0854	4.56
1C	0.2570	14.75
2A	0.1165	9.19
2C	0.0986	6.63
3A	0.0467	7.83
3B	0.1272	13.02
4A	0.1597	7.14
4B	0.0120	0.39
4C	0.1402	4.33
Average	<b>0.1134</b>	<b>7.40</b>

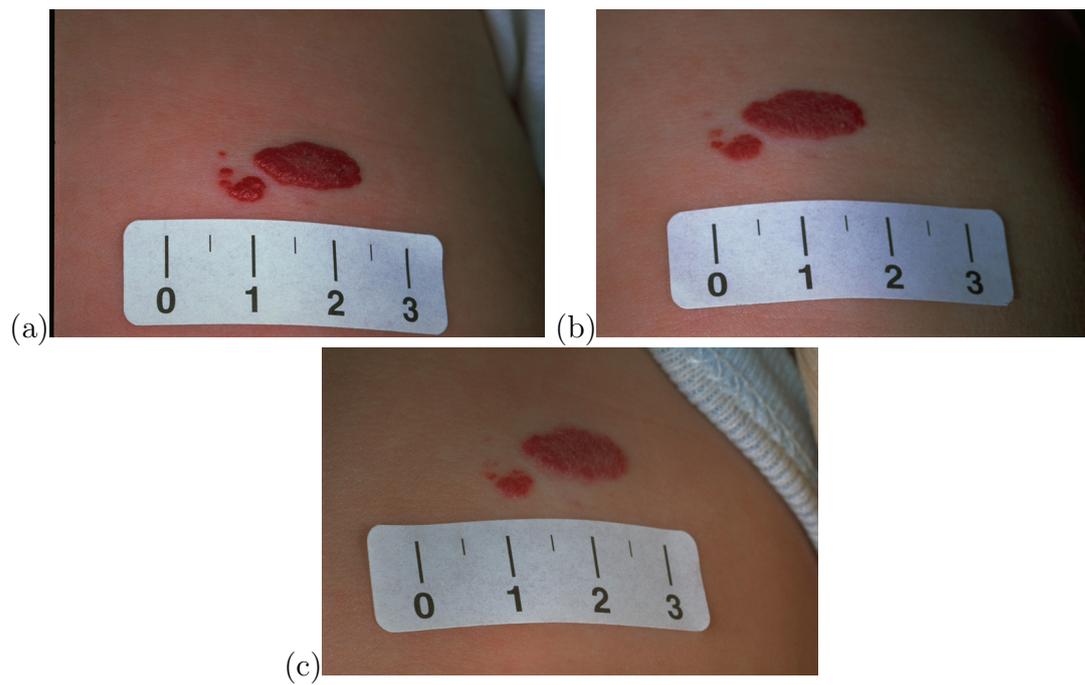


Fig. 1. A follow-up series of hemangioma images.

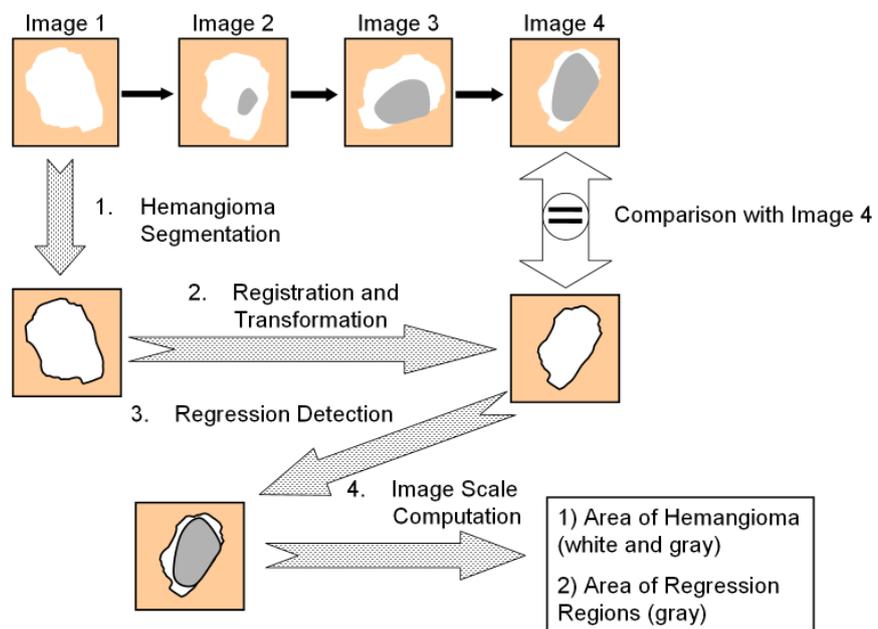


Fig. 2. Illustration of the workflow for the assessment of hemangioma development.

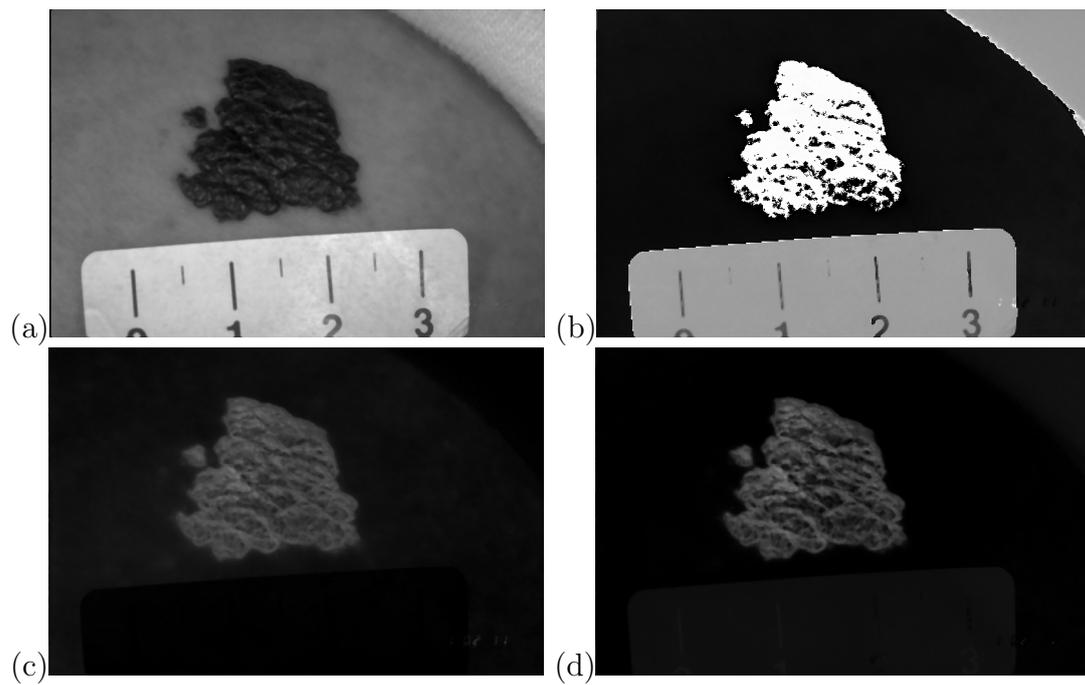


Fig. 3. The intensity images of the four features (a)  $G$ , (b)  $H$ , (c)  $a$  and (d)  $abdist$  of a particular hemangioma image.

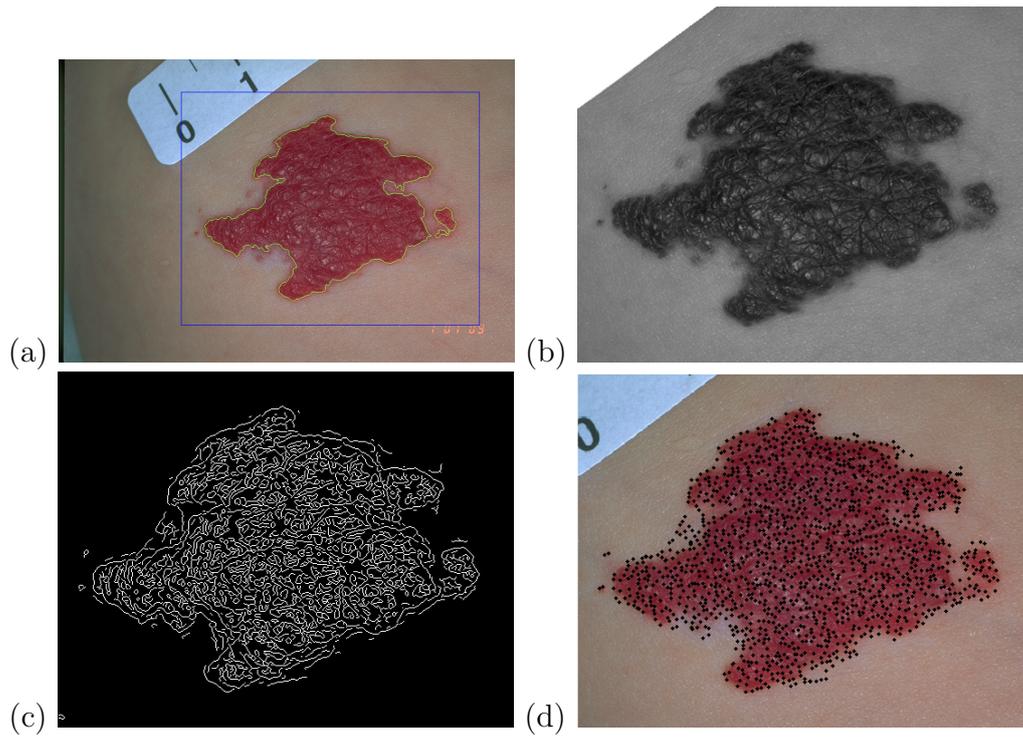


Fig. 4. Individual steps of the interest point localization procedure.

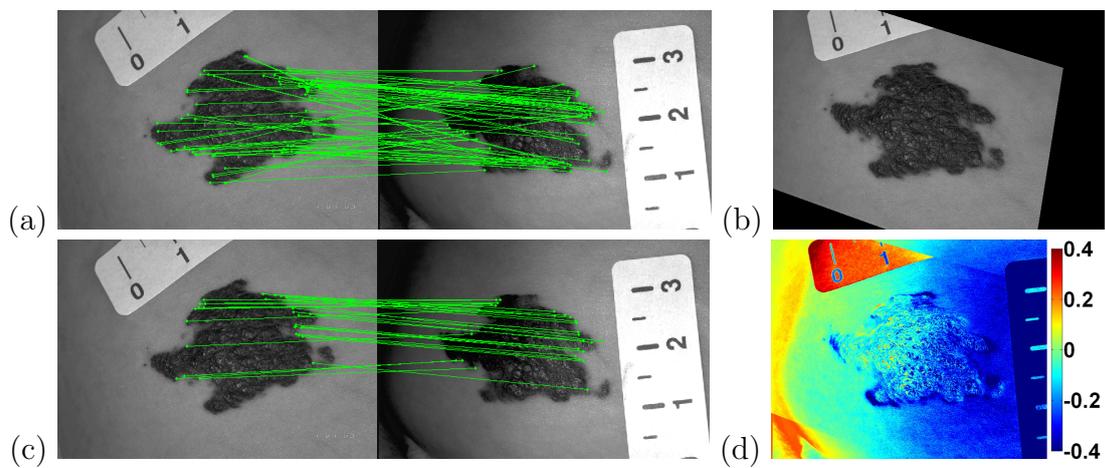


Fig. 5. (a) Initially detected matches between sensed image (left) and reference image (right), (b) detected inliers after removal of incorrect matches by RANSAC, (c) transformed sensed image and (d) image displaying the difference between the transformed sensed image and the reference image.



(a)

(b)

(c)

Fig. 6. (a),(b) Consecutive images of a hemangioma and (c) regressions marked with white border.

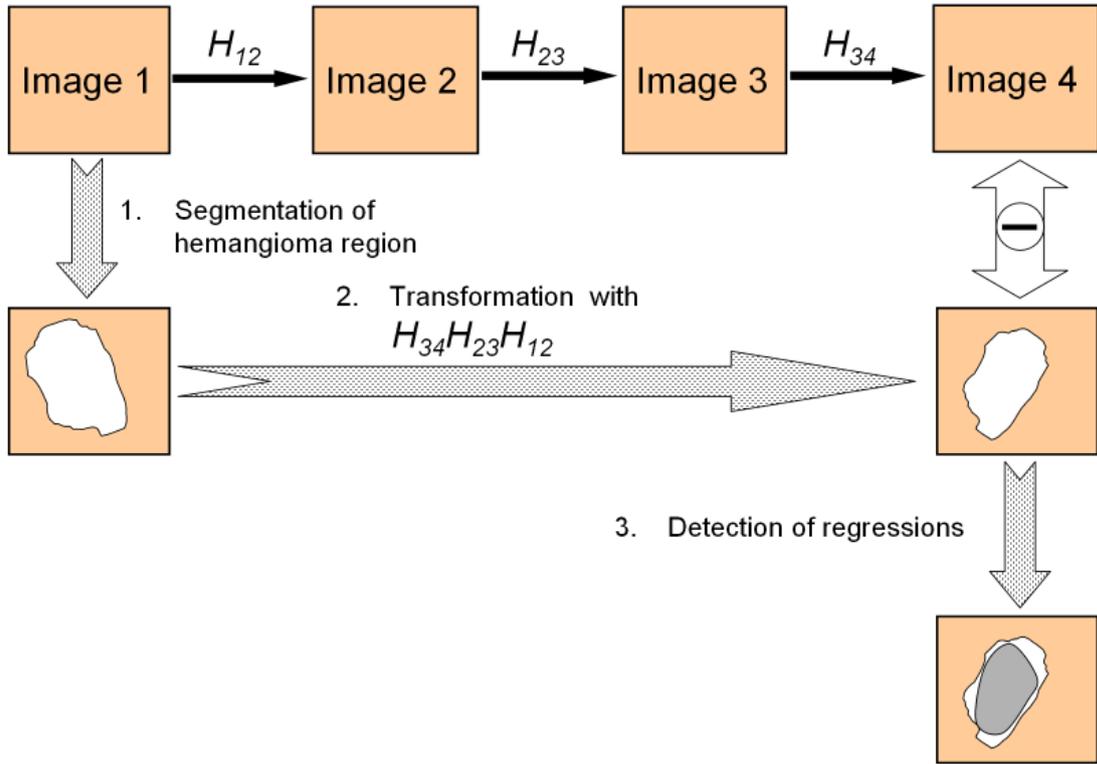


Fig. 7. Illustration of the regression detection procedure.

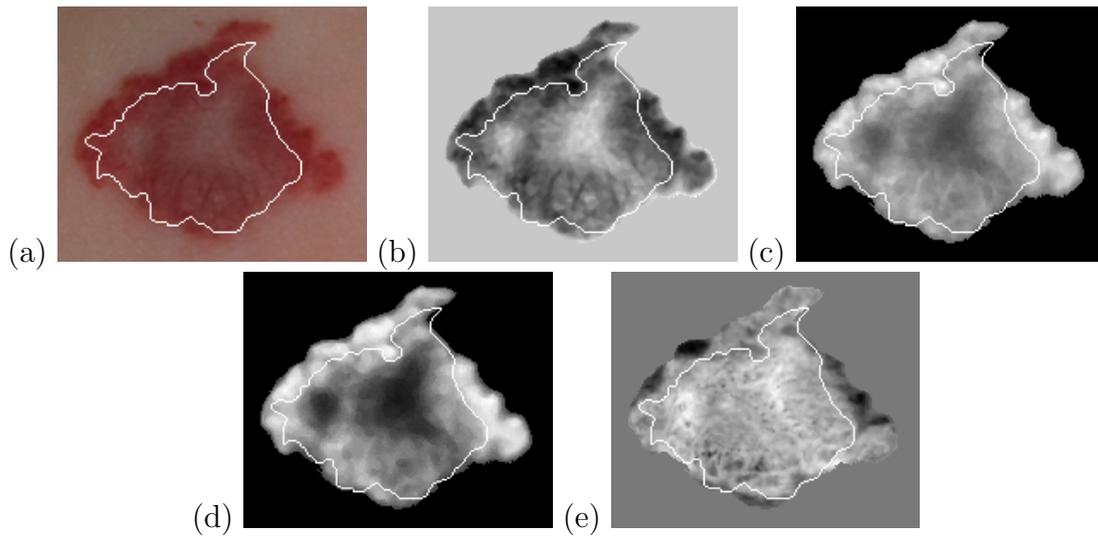


Fig. 8. (a) Original hemangioma image with regressing region marked by white border and values of features: (b)  $G$ , (c)  $a^*$ , (d)  $abdist$ , (e) difference of  $G$  between first follow-up and current image. Values are normalized to the range 0 to 1, thus 0 is represented by black and 1 is represented by white.

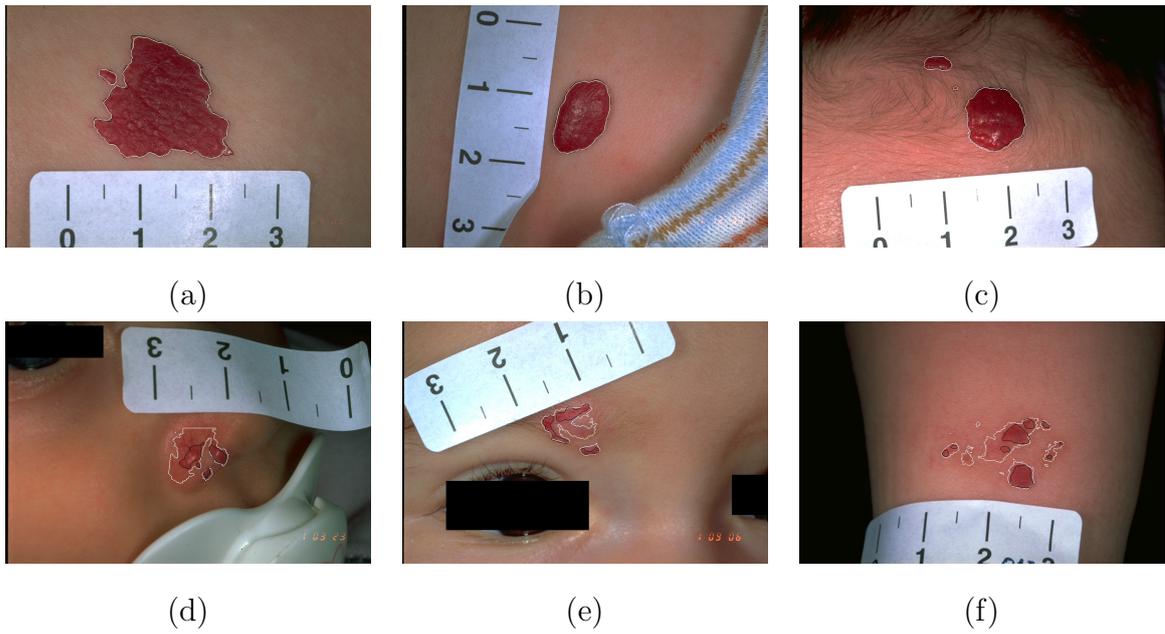


Fig. 9. Automatic segmentation (white) and standard of reference (black) of 6 images.

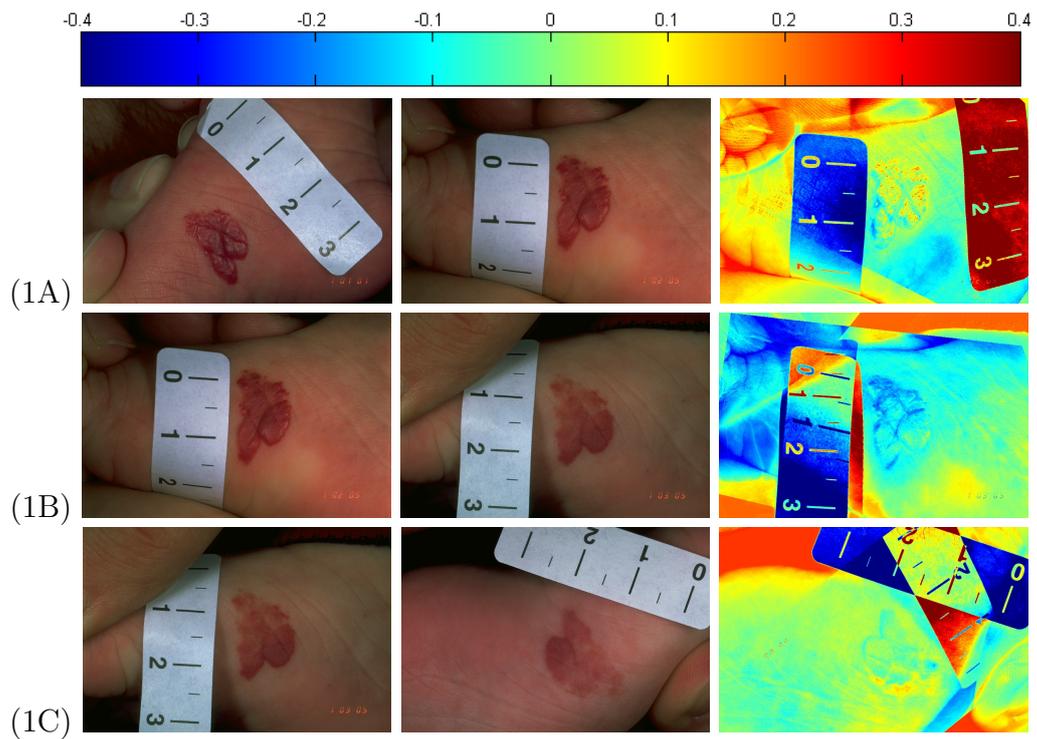


Fig. 10. Difference images (1A)-(1C) of patient 1 in Table 1 showing the same hemangioma at different times. The colorbar is shown at the top.

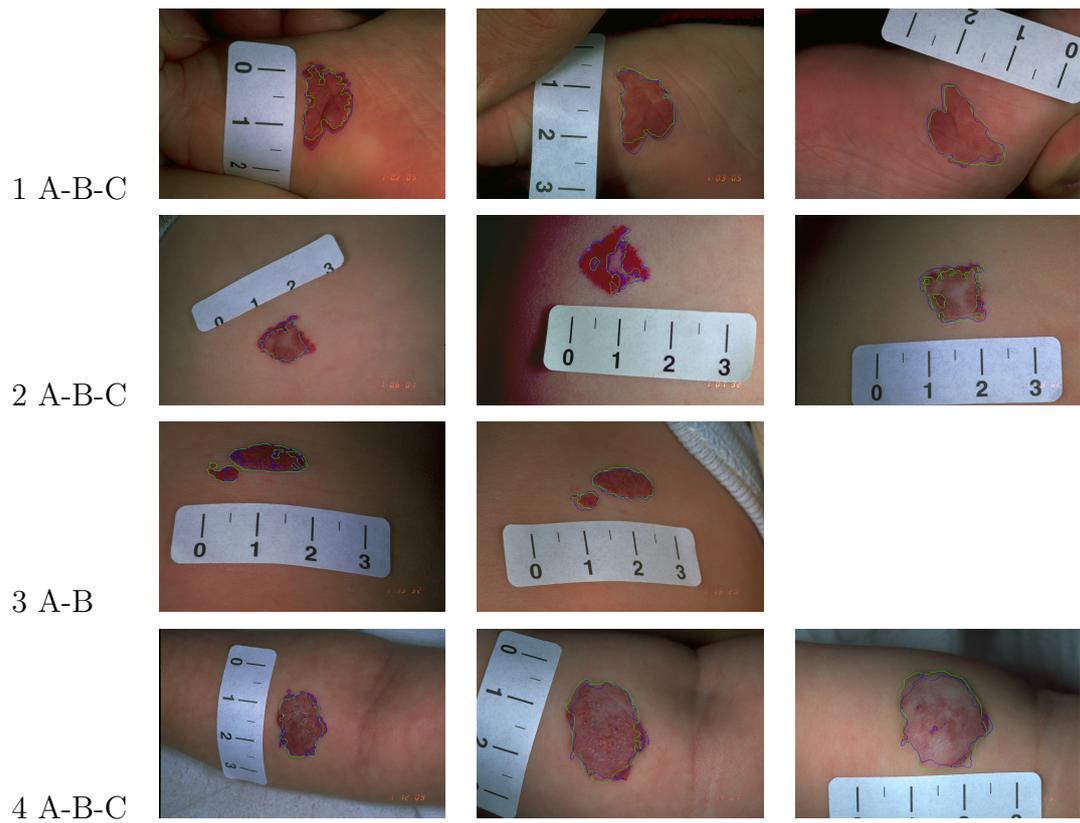


Fig. 11. Results of the automatic regression detection during follow-up examinations for 4 cases. Each row of the table represent the temporal development for one patient. (green: manual detection, blue: automatic detection).