



**HAL**  
open science

## Segmentation of Erythema from Skin Photographs for Assisted Diagnosis in Allergology

Elodie Roullot, Jean-Eric Autegarden, Patrick Devriendt, Francisque Leynadier

► **To cite this version:**

Elodie Roullot, Jean-Eric Autegarden, Patrick Devriendt, Francisque Leynadier. Segmentation of Erythema from Skin Photographs for Assisted Diagnosis in Allergology. International Conference on Advances in Pattern Recognition, Aug 2005, Bath, United Kingdom. pp.754-763. hal-00756086

**HAL Id: hal-00756086**

**<https://hal.science/hal-00756086>**

Submitted on 22 Nov 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Segmentation of Erythema from Skin Photographs for Assisted Diagnosis in Allergology

Elodie Roullot<sup>1</sup>, Jean-Eric Autegarden<sup>2</sup>, Patrick Devriendt<sup>1</sup>, Francisque Leynadier<sup>2</sup>

<sup>1</sup> ESME Sudria, Department Signal & Telecoms, 38 rue Molière, 94200 IVRY-SUR-SEINE, France. [roullot@esme.fr](mailto:roullot@esme.fr)

<sup>2</sup> Hôpital Tenon, Service d'allergologie, 4 rue de la Chine, 75020 PARIS, France.

**Abstract.** More than 2 people out of 10 suffer from allergies, which can take various forms, from eczema to anaphylactic reactions with possible lethal consequences. Diagnosis is achieved through so-called "prick-tests" or IDR (intra-dermo-reaction): the injection of a small quantity of substances suspected to cause the allergic manifestation induces an erythema, the size of which is a useful indicator for the diagnosis. The manual surface measurement is time-consuming and inaccurate. This article presents a method for the semi-automatic measurement of the erythema from a photograph of the skin, taken in such conditions that lighting problems are minimized. The method is based on region growing and takes advantage of the most significant color spaces; the Lab space appears to be the best suited. It was tested on nearly 100 images, taken by various operators, on patients with various skin pigmentations; it gave promising results and proved to be robust.

### 1 Introduction

Allergies are classified as the fourth most important public health problem in the world. Numerous epidemiologic studies show that 20 to 25% of people suffer from allergic troubles. Allergic manifestations take various forms, among which eczema, rhinitis, asthma, or even Quincke oedema or anaphylactic reaction, with possible lethal consequences.

In order to prevent from the most severe anaphylactic manifestations, it is necessary to have an acute knowledge of the substances that are responsible for the allergic reaction, called the "allergens". This diagnosis is achieved by the allergologist through cutaneous tests: "prick-tests" or IDR (intra dermo reaction). In prick-tests, a drop of solution containing the allergen to be tested is dropped on the forearm of the patient, which is then punctured in order to make the allergen penetrate into the skin ; in IDR's, the allergen is injected into the skin with a syringe.

An allergic reaction locally induces a dilatation of blood vessels and the release of chemical substances due to the "axon reflex"[1], leading after 15 to 20 minutes to the apparition of an erythema, characterized by a local reddening of the skin and an in-

crease in its temperature; at the center of the erythema a papula is formed, looking like a mosquito bite. The surface of the erythema is an indicator for the severity of the allergy.

The mean diameter of each erythema is measured with a scale; each value, corresponding to a given allergen, is compared to 2 reference values: the positive and negative standards. The positive standard indicates the standard degree of reaction of the patient and is achieved by the injection of codein phosphate which involves for everybody a positive reaction. The negative standard is achieved by the injection of physiological serum which never involves a positive reaction (only the puncture is involved in an eventual change in the skin appearance). The reaction induced by a given allergen is considered to be positive if it is equivalent to or larger than the positive standard; if it is comparable to the negative standard, the reaction is considered to be negative.

This technique appears very fastidious especially due to the huge number of allergens to be tested (up to 20); moreover, the measures are corrupted by errors due to the inter- and intra-operators variability, thus making a follow-up of the patients more difficult. Moreover, the diameter of the erythema does not give a good estimation of the surface, the measure of which would be preferred by the allergologists.

All these reasons make an automatic measure of the surface very challenging; the expected improvements are: less time spent for measuring, less intra- and inter-operators variations in the measurement, more accuracy thanks to the surface measurement instead of the diameter measurement, and the storage possibility, making a follow-up of the patients easier. However, the clinical motivation of this work is not to provide a fully automated diagnosis system, but rather to provide assistance to the allergologist to make his diagnosis. Thus, it should be focused on the segmentation problem but it should not be dealt with an automatic diagnosis whether the patient is allergic or not.

Several research works emerged during the last years, dealing with the automation of erythema detection. They make use either of spectroscopy, or of echography, or of visible imaging. Several teams including Kopola et al. [2], Dawson et al. [3], Ferguson-Pell et al. [4], Hajizadeh-Saffar [5] or Diffey [6] suggest calculating an "erythema index" from reflectance measures in different carefully chosen spectral bands, by means of a spectroscope. Several methods among these were evaluated by Riordan et al. [7]. Some of them allow detecting erythemas even in highly pigmented skins; however, all of them only provide an index (generally an erythema index and a melanin index) that characterizes the global change in color, whereas routinely the allergologist is interested in the size of the erythema.

Seidenari [8] et al., for their part, use echography to quantify erythemas induced by patch-tests; once again, this provides values representing the intensity of the reaction whereas the allergologist is interested in the surface.

The surface can only be obtained automatically from images of the skin; although the problem of segmentation is dealt with very often in the context of melanoma detection for example, this problem has been addressed very rarely in the context of erythema detection. Unfortunately, the methods used for segmenting melanomas can not be transposed easily for erythema detection, since melanoma images are much more

contrasted than erythema images. Nischik et al. addressed the problem of erythema segmentation from visible light images and proposed a method for segmenting the erythema, based on the CIELab color space and on the comparison between two images acquired before and after prick-testing [1]. This method presents a major drawback: it implies a motion analysis in order to suppress artifacts due to the unavoidable motion between both images, since the reaction can be read 15 to 20 minutes after prick-testing.

We therefore propose an original approach to segment the erythema and provide a surface measurement of it, using only one image taken after reaction.

In Section 2 we briefly describe the image capture system and the image database; then we address in Section 3 a colorimetric study in order to extract the best color space to use. In Section 4 we describe the method, and the results are presented and discussed in Section 5.

## 2 Image Database

As said before, the detection is performed from digital photographs of the arm and forearm. These digital photographs are taken using a standard color CCD camera with a spatial resolution of 752 x 582 pixels. The camera is connected to a personal computer via a digitalization card (Ellips Riowin) allowing real-time visualization of the acquired image and eventual tuning of some acquisition parameters, as well as saving a color image in bitmap format (BMP).

The illumination system is composed of two neon tubes placed behind unpolished and translucent Plexiglas plates, in order to provide an homogeneous illumination and thus to avoid reflection artifacts.

90 images of the arm or forearm were acquired at [anonymous] hospital ([town], [country]) from January 2003 to June 2004 on both male and female adult patients who agreed to take part in these tests. Most patients had lightly pigmented skins (Caucasian for most of them, and Metis or Mongoloid), and only one patient had a very dark skin (Negroid). Ten randomly chosen images formed the learning database and the others were used for evaluation purposes.

## 3 Colorimetric Study

A first examination of the learning database revealed that the classical RGB (red, green, blue) color space is not the best suited for the segmentation of erythemas. In this section, we study several color spaces to extract the best suited for discriminating the reddened skin from the [normal] skin. The CMY color space is the complementary color space to the RGB one: it is based on subtractive synthesis (as in printers) while the RGB space is based on additive synthesis (as in monitors). HSV (hue, saturation, value) and HSL (hue, saturation, luminance) systems correspond to a psychovisual description of color, based on the human eye function. The LMS color space is also

based on the human eye: each stimulus L (long), M (middle) and S (short) represents the spectral sensitivity of the photoreceptive cells of the eye. The CIE Lab color space is derived from the XYZ standard color space, defined by the CIE in 1931 by a homographic transformation of the RGB space. L represents the luminance whereas a and b respectively represent the red-green chrominance and the blue chrominance.

All these values (R, G, B; C, M, Y; H, S, V; H, S, L; L, M, S; X, Y, Z; L, a, b) were computed for each image of the learning database. An SNR study revealed that the chrominance  $a$  provides the best discrimination of erythematous skin versus normal skin. These results confirm those obtained by Fullerton et al. [9] and Weatherall et al. [10]; they recommend the use of the Lab color space, which takes into account the non linearity of color perception by the human eye.

## 4 Color-based Segmentation

### 4.1 Choice of the Method

As said before, the segmentation of erythema is a difficult task, because of the colorimetric heterogeneity of the skin (both the normal and the reddened skin present a marmor aspect), as well as because of the low contrast between erythema and normal skin. This problem belongs to the family of problems dealing with the segmentation of low SNR images. For such images, classical segmentation methods such as contour detection approaches are not well suited, that is why we concentrate our attention on constrained region segmentation approaches. Among these methods, structural ones are well suited to the problem since they allow constraining more or less the pattern of the extracted region [11].

Moreover, each image acquired represents the full arm or forearm, and can thus contain several erythemas corresponding to different allergens. It does not appear judicious to develop a system that would be able to recognize automatically to each kind of allergen each area of the arm corresponds, since it will induce many technical constraints for the operators performing the test, whereas it is not very constraining for him/her to have a minimum level of interaction with the system. On the contrary, the medical specialists generally prefer keeping a given degree of control on the automatic diagnosis process rather than obtaining results from a fully automatic system in which they have low confidence. Thus, it appears as a reasonable compromise to require the allergologist to click on the corresponding erythema when prompted with the name of an allergen, for example, as follows:

- ? launch image acquisition
- ? for each allergen:
  - ? if an erythema is present:
    - ? mouse click within the reaction area
    - ? detection is thus performed automatically from the given seed
    - ? view, validate and store the results, both quantitative (image) and qualitative (surface)
  - ? else, store the result (absence of reaction) and skip to next allergen.

Taking into account both the considerations about the segmentation problem and the integration of the algorithm within a future computer-aided diagnosis software, a method based on region growing with manual input of the seeds appears consistent. In the next subsection we describe our algorithm based on pixel aggregation, with a strong constraint on the connectivity of the region and a weak constraint on its pattern.

#### 4.2 Segmentation Algorithm

The algorithm is based on region growing [11,12] where the initial regions have a unitary size and are chosen manually by mouse clicking. Aggregation of new pixels is iterative and depends on two criteria: adjacency and colorimetric similarity. Each iteration begins with the constitution of a candidate set, based on adjacency; all candidates satisfying the colorimetric criteria are aggregated, and the new region is then used for the following iteration. The algorithm stops when no more candidates are eligible for aggregation.

The different steps of the algorithm are detailed below:

- ? Initialization: manual choice of  $N$  seeds chosen within the erythema, forming the  $N$  initial regions  $R_i$ .
- ? For each region  $i$ :
  - ? selection of the candidate set  $C_i$ , i.e. the set of pixels adjacent to the current region, obtained with a morphological dilatation:

$$C_i = (R_i \oplus E_s) \cap R_i \quad (1)$$

where  $\oplus$  represents the morphological dilatation with a  $3 \times 3$  square structural element  $E_s$ .

- ? among the candidates  $c_i^k$ , selection of those satisfying the colorimetric similarity criterion:

$$a(c_i^k) > T(a(R_i)) \quad (2)$$

where  $a$  represents the chrominance  $[a]$  in the Lab color space,  $T$  is an automatic threshold computed from statistics in the region  $R_i$ .

- ? if no candidate was selected at the previous step, algorithm ends.
- ? else, creation of the new region  $R_{i+1}$  as the union of region  $R_i$  and selected pixels  $c_i^k$ .

#### 4.3 Automatic Threshold Tuning

A statistical study of the chrominance  $[a]$  was performed on the 10 images of the learning database: a coarse thresholding helped select the arm and suppress the background, and then the histogram of the chrominance  $[a]$  was computed. This study revealed that the mean and variance of the chrominance varied sensibly from an image to another. Therefore the threshold was chosen as a function of the mean and variance of the chrominance  $[a]$ . For the 10 images of the learning database, the erythema sur-

face was computed for various thresholds and compared to the reference surface (which is described further in Section 5). The threshold  $T$  varies as follows:

$$T = \bar{a}(I) + \lambda \sigma_a(I) \quad (3)$$

where  $\bar{a}$  is the mean chrominance over the image  $I$ ,  $\sigma_a$  is the variance of  $\bar{a}$  over the image  $I$ , and the parameter  $\lambda$  varies from 0.6 to 1.4. The value of  $\lambda$  giving the exact diameter is interpolated; over the learning database, the mean  $\lambda$  equals 0.97 meaning that the best threshold is  $\bar{a}(I) + 0.97\sigma_a(I)$ .

The robustness of the parameter  $\lambda$  can also be evaluated from these tests: a variation of 10% of the standard deviation in the threshold induces a variation of about 5% in the diameter. We can therefore affirm that the threshold is quite robust.

## 5 Results and Discussion

### 5.1 Qualitative Study

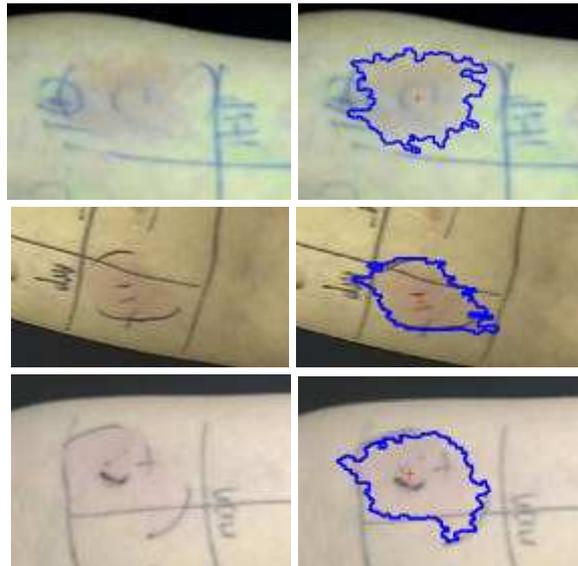
The algorithm was tested on the 90 erythema images of the database. Among the 90 images, 89 concern patients with lightly pigmented skins (Caucasian, Mongoloid, Metis) and one patient has a highly pigmented skin (Negroid). Some results are shown on Figure 1 for the clear skins and on Figure 2 for the dark skin. On these images, blue pencil marks can be noticed, corresponding to the marks used by the physician; note that these marks do not appear in the  $\bar{a}$  chrominance image and therefore do not disturb the algorithm (this would not be the case if a red pencil was used, as it was the case for some images that were therefore useless). When the system will be used instead of the manual measurement, these marks will not be needed anymore and will thus disappear. At this step of development, they are useful to compare the contour provided by the algorithm with the diameter estimated by the physician.

On Figure 1, it can be noticed that the lighting conditions differ considerably from one image to another; this does not seem to affect the algorithm, and confirms its robustness. Concerning the dark skin on Figure 2, the erythema of the unique negroid patient taking part in the tests could be detected in a satisfactory way.

In 5 images, the algorithm was not able to detect any contour. Possible reasons for these failures are: the low contrast of the image, due either to the image capture or to the skin reactivity of the patient; artifacts on the patient skin; defaults in the lighting systems providing an image with reflection artifacts that disturb the segmentation process.

The detection of an erythema lasts a few seconds, so that for one patient tested for 10 allergens, the total time including the mouse clicking, the automatic detection and storage of the results amounts to far less than one minute: less than the time for the patient to get dressed after the measurements. Routinely, the operators needs to draw the limits of the erythema before measuring (when measuring the diameter, the operator has to flatten the skin, which makes the skin reddening disappear), then measure, and finally report the value in a paper or electronic report; all these steps require sev-

eral minutes and are much more time-consuming than the mouse-clicking and detection. Moreover, as said before, the measures are corrupted by an important variability between different operators and even between different measures taken by the same operators. The automatic detection provides a reproducible measure and proved to be nearly independent of the manual input of the seed (if the seed is chosen within the erythema and not on the border).



**Fig. 1.** Examples of results obtained on various skin types among Caucasian, Mongoloid, and Métis patients. Left, the regions of interest in the original images; Right, the results superimposed on the original regions of interest. Blue pencil marks can be noted, which correspond to the separations between the allergens being tested, and to the limits used by the physician for the diameter measurement.



**Fig. 2.** Example of result obtained on a skin of type Negroid. On this unique example it seems that the algorithm does not encounter any particular problem detecting the contour of the erythema, although the erythema is hardly visible with the naked eye.

## 5.2 Quantitative Evaluation

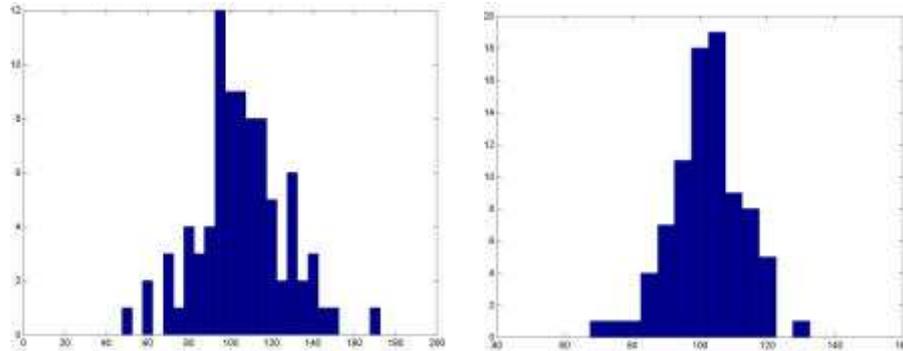
Two options can be considered for evaluating the algorithm; the first one consists in comparing the results to the clinical measurements performed manually, and the second one consists in comparing to a ground-truth surface that remains to be defined.

Actually, the measurements conducted routinely are diameter measurements performed by the nurses with a scale. These measurements suffer from a very important intra- and inter-operators variability, mainly due to two reasons. Firstly, the use of the scale implies a deformation of the natural curve of the arm and pushing on the skin implies color changes that can affect the measurement. Secondly, the erythema contour is not necessarily a circle; the nurses take as "diameter" measurement the biggest distance between the erythema borders. This is not so easy to appreciate with the human eye and is another factor for the variability of the measurement. Therefore, a quantitative evaluation based on these measurements does not appear to be judicious, and it appears necessary to perform evaluation by comparing to a "ground-truth" value.

To this end, for each image of the database an expert was asked to draw the erythema contour as precisely as possible; actually he pointed as many points on the contour as he needed and the corresponding polygon (with many vertices) was considered to be the ground-truth boundary. A "goodness" measure was defined as the ratio between the provided surface and the ground-truth surface. Figure 3 shows the histogram of this ratio computed over the database.

Figure 4(a) presents a "good" result of evaluation with the polygon and the detection result superimposed on the image. On this example, the evaluation metric (the surface ratio) is 95%; the area of the polygon is 6411 pixels<sup>2</sup> and that of the automatic contour is 6139 pixels<sup>2</sup>. A fictive diameter was computed as  $2\sqrt{area/\pi}$ , giving 90 pixels for the ground-truth, 88 pixels for the automatic detection. These values can be compared to the diameter measured manually (corresponding to the pencil parks) to 111 pixels. The error of 2 pixels is very low compared to the 21 pixels difference between the manual measure and the ground-truth. The difference between the manual measurement and the ground-truth provides a surface ratio of 152%, which corresponds to the worst cases of the automatic detection, as shown on Figure 3 (left).

Figure 4(b) and 4(c) shows two of the worse results obtained with the automatic detection. For the left image, the surface ratio is 76%; for the right image, the surface ratio is 155%. When comparing the results of Figure 4, it can be noticed that even in the worse cases, the contour is not so bad and is anyway at least as good as the one that could have been drawn manually. This illustrates the fact that surface ratios of 80% or 150% for example are considered as acceptable by the experts, and also contribute to the improvement of the measurements when compared to the manual ones.



**Fig. 3.** Left: Histogram of the ratio between the surface provided by the algorithm and the ground-truth surface. This surface comparison leads to ratios that go very fast far away from the ideal ratio of 100%; this is due to the fact that the variation of surface is quadratic. Right: the same histogram after correction of the quadratic effect due to the surface measurement. As a consequence the ratios are homogeneous with diameter ratios and are thus easier to understand intuitively.

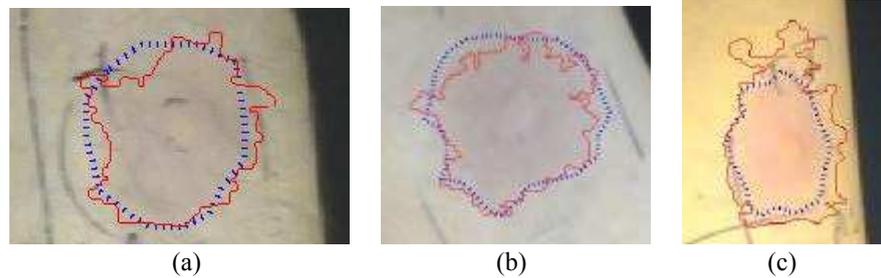
## 6 Conclusion

In this article, we addressed the problem of the automatic detection and quantification of reactions to allergic tests. This original problem was not addressed often in the literature, although it derives from an actual query from the allergologists.

In this study we focused on the segmentation problem; the automatic decision process was not addressed. Indeed the segmentation problem is the most difficult to solve because of the low contrast of erythema images and of the heterogeneity of the skin color. To this end a method was developed that combines color image analysis with a region growing algorithm; it provides a robust and reproducible contour detection from the manual input of a seed point within the erythema.

The method was applied on a database of nearly 100 images acquired during the consultations of allergologists at [anonymous] hospital. A rigorous evaluation allowed proving that the method has a satisfying detection rate and is quite robust. The physicians judged the software potentially helpful and very promising. In the future, it will be necessary to perform more tests especially on dark skins, since they were not well represented in the present database.

It is also planned to work on the acquisition and lighting systems: the system described in Section 2 is a provisory one and work is already being performed on a best suited system, such as those used by some dermatologists consisting of a digital camera together with a ring light.



**Fig. 4.** Comparison between the automatic contour (solid line), the ground-truth (dashed line) and the manual diameter (on image (a): dark pencil marks looking like parenthesis) in a "good" case (a) and in two "bad" cases ((b): the surface ratio is 76%; (c): the surface ratio is 155%).

## References

1. M. Nischik and C. Forster. Analysis of Skin Erythema Using True-Color Images. *IEEE Transactions on Medical Imaging*, 16(6):711-716, December 1997.
2. H. Kopola, A. Lahti, R. Myllyla, M. Hannuksela. Two-channel fiber optic skin erythema meter. *Optical Engineering*, 32:222-226, 1993.
3. J. B. Dawson, D. J. Barker et al. A theoretical and experimental study of light absorption and scattering by in-vivo skin. *Phys Med Biol* 25(4):695-709, 1980.
4. M. Ferguson-Pell and S. Hagsawa. An empirical technique to compensate for melanin when monitoring skin microcirculation using reflectance spectrophotometry. *Med Eng Physics* 17(2):104-110, 1995.
5. M. Hajizadeh-Saffar, J. W. Feather and J. B. Dawson. An investigation of factors affecting the accuracy of in vivo measurement of skin pigments by reflectance spectrophotometry. *Phys Med Biol* 35(9):1301-1315, 1990.
6. B. L. Diffey and P. M. Farr. Quantitative aspects of ultraviolet erythema. *Clin Phys Physiol Meas* 12(4):311-325, 1991.
7. B. Riordan, S. Sprigle and M. Linden. Testing the validity of erythema detection algorithms. *Journal of Rehabilitation Research and Development*, 38(1), January/February 2001.
8. S. Seidenari and B. Belletti. The Quantification of Patch Test Responses: A Comparison Between Echographic and Colorimetric Methods. *Acta Derm Venerol* 78:364-366, 1998.
9. A. Fullerton, T. Fischer et al. Guidelines for the Measurement of Skin Color and Erythema. *Contact Dermatitis* 35:1-10, 1996.
10. I. L. Weatherall and B. D. Coombs. Skin Color Measurements in Terms of CIELAB Color Space Values. *Journal of Investigative Dermatology* 99:468-473, 1992.
11. J.-P. Cocquerez and S. Philipp. *Analyse d'images : filtrage et segmentation*. Masson, 1995.
12. H. Maître. *Le traitement des images*. Lavoisier, 2003.