

# SIFT descriptor to set landmarks on biological images

Van Linh LE<sup>1,3</sup>, Marie BEURTON-AIMAR<sup>1</sup>, Adrien KRAHENBUHL<sup>1</sup>, Nicolas PARISEY<sup>2</sup>

<sup>1</sup>Laboratoire Bordelais de Recherche en Informatique (LaBRI)  
351, cours de la Libération, F-33405 Talence cedex, France

<sup>2</sup>Institute for genetics, the environment and plant protection (IGEPP), INRA 1349  
35653, Le Rheu, France

<sup>3</sup>Faculty of Information Technology, Dalat university  
1 Phu Dong Thien Vuong, Dalat, Lamdong, Vietnam

van-linh.le, beurton, adrien.krahenbuhl@labri.fr, nparisey@rennes.inra.fr

**Résumé** – L’analyse d’images est une des étapes majeures du traitement d’images numériques, on l’applique aussi bien en imagerie médicale ou biologique, en vision par ordinateur ... Classiquement, l’image est très présente en biologie, les différentes techniques d’acquisition (photos depuis l’imagerie microscopique pour étudier les structures anatomiques jusqu’au scanner, tomographie ou IRM pour étudier les tissus).

**Abstract** – Image analysis is a large field in image processing and it has been applied in practice with many applications in the different majors such as medicine, machine vision, biology, ... In biology, images are widely used from a long time, to study molecule structures or behaviors, tissue characteristics and in general to measure and to classify anatomical details. Classification of biological samples can be obtained from studying morphological features but at this time, setting morphological markers is done manually. In this work, we have focused on the problem to replace manual operations by automatic procedures to set landmarks which are points of interests in biological images. This paper presents how we have designed a specific way to use the SIFT descriptor to improve the results that we have obtained before to achieve this task. The two main characteristics of our method are the reduction of the patch area to compute the source descriptor and the definition of a restrictive search area of the target. The efficiency of the method is evaluated on two sets of images: left and right mandibles of beetles belonging to a study of the national institute of agriculture (INRA). The complete method is implemented in a framework called MAELab and freely available as a C++ library.

## 1 Introduction

Morphometry analysis is an important field of image analysis in biology. It is used to characterize the shape variations of the organisms. From obtained information, the biologists can evaluate the evolution of an organism or detect differences between several ones. Depending on the requirements of the application, the output of analysis process can be measures of shape, color ... or identification of pattern or landmark (points of interest) positions. Landmarks are points along an image outline that store a lot of important information about the shape of the image. The morphometric landmarks are precise points defined by the biologists. They are used in many biological studies and included into the classification tasks. Until now, the morphometric landmarks are mainly manually identified. The manual identification is time-consuming and could vary a lot depending on the operator.

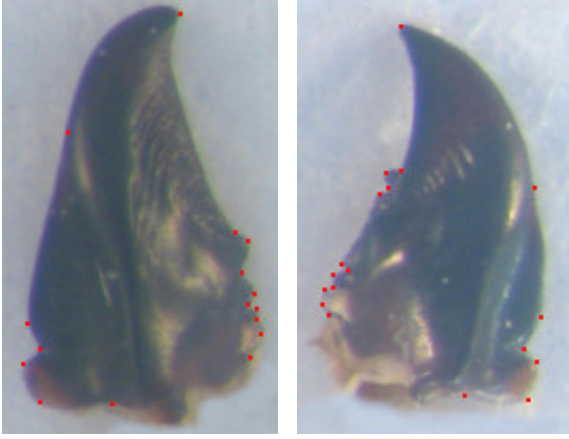
In this paper, we focus on a specific part of our framework which consists to introduce the computing of SIFT descriptors. This information about landmark areas allows to reduce the size of the source patch and the target search space at the same time. These works are evaluated on a dataset corresponding to a collection of 293 beetles from Brittany lands. The first step of these works concerns the left and right mandibles. For

each beetle, a set of landmarks has been manually positioned by the biologists (see Fig. 1). In our study, we used this dataset as ground truth to evaluate the position of the automatically estimated landmarks. Our framework also contains steps of segmentation, registration, and descriptors comparison. Each part will be discussed on the experiments results.

## 2 Landmark descriptor

For each beetle, the morphometric landmarks have been manually set on mandibles images by the biologists : 16 and 18 landmarks for each left and right mandible, respectively. The considered problem is the automatic detection of these landmarks on a mandible image to replace the manual one. In the whole of the process, the landmarks will be estimated on a target image by using the manual landmarks of a source image. The source image is chosen randomly from the set of all images.

In this section, firstly, we have a summary about the general usage of SIFT. Then, we will discuss about the details of our method of landmark estimation.



(a) Left mandible (b) Right mandible

FIGURE 1 – Example of beetle mandibles from the studied data set with manual landmarks.

## 2.1 The adapted SIFT method

The SIFT method has been proposed by D. Lowe [7, 8]. It is used to extract distinctive features from the images, potentially invariant to rotation, scale and noise. The SIFT features can be used to determine a matching between source and target images. The SIFT method is composed by four steps : (1) scale-space extrema detection, (2) keypoint localization, (3) orientation assignment, and (4) keypoint descriptor.

In the first step, a difference of Gaussian (DoG) [3] function is applied to identify the interest points at all scales and orientation of the input image. The keypoints are taken as the maximal and minimal of the result of DoG function computed at multiple scales. The scale-space extrema detection produces a lot of keypoint candidates, with some of them are unstable. In the second step of SIFT, the key point candidates are localized and refined by suppressing the ones which have the low contrast or are poorly localized along an edge.

Then, the orientation and gradient magnitude of key points are calculated by considered their 4-neighborhoods. Finally, the descriptor is computed for each key point, from the orientation and gradient magnitude. A descriptor is a region of  $16 \times 16$  pixels around the key point.

By applying the original SIFT into our problem, we have succeeded indicating the keypoints in the image (see Fig. 2). But we do not have the result when we try to extract keypoint correspondences between the source and target image. The problem is carried from choosing the best points from the large set of the candidates. To solve this problem, we have modified the method in order to limit the search space before applying the SIFT computing. We have also change the size of the grid to calculate the SIFT descriptor.

## 2.2 SIFT using into landmark area

As mentioned in section 2.1, we have tried to use the standard version of SIFT to detect the landmarks. But the result

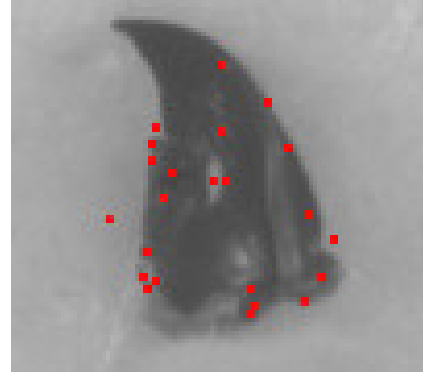


FIGURE 2 – SIFT keypoints in a right mandible.

contains a lot of candidates for the estimated landmarks. Moreover, when we compare the position of the detected points with the ground truth (manual landmarks), they are very far. Last, searching in whole image is very time consuming. To obtain better result, we propose to exploit the SIFT descriptors by another way.

Firstly, two images are segmented and two lists of contours points are registered. Then, the area (called patch) around each source landmark is defined and a larger patch is extracted in the target image at the same position. The SIFT descriptor is then calculated as usual. But we have reduced two first steps of SIFT and have changed the size of the sample regions for  $9 \times 9$  (instead of  $16 \times 16$ ). This size has been empirically determined after several tests. The SIFT descriptor for the patch is also an histogram containing the sum of pixel gradients for each consider direction. The comparison between two SIFT-descriptors is done by an L2-distance as in Eq. 1 :

$$L(D1, D2) = \sum_{i=0}^n \sqrt{(D1_i - D2_i)^2} \quad (1)$$

Where :

- $n$  is the number of directions
- $D1$  and  $D2$  are two descriptors of size  $n$ ,
- $D1_i$  and  $D2_i$  are the  $i^{th}$  descriptor values.

Our approach to use the SIFT descriptors into our work is presented in Fig. 3. To detect the target landmarks, a registration is computed between the source and target images. This step is described in the experiment results. Then, the patch  $P_s$  of the source and  $P_t$  of the target are created with the size of  $P_s$  smaller than the size of  $P_t$ . For each pixel in  $P_t$ , a sub-patch  $P'_t$  is extracted with the same size of  $P_s$ . When the  $P'_t$  have a part outside of  $P_t$ , the pixels outside the patch will be considered. Then, the distance  $L(P_s, P'_t)$  is computed following Eq. (1). The process ends when all the pixels in patch  $P_t$  are considered. The position of the estimated landmark corresponds to the position of the sub-patch  $P'_t$  giving the smallest distance  $L(P_s, P'_t)$ .

By experiments, a patch sample of  $9 \times 9$  pixels centered in each landmark on the source and the size of  $36 \times 36$  is kept for the patch on the target image.

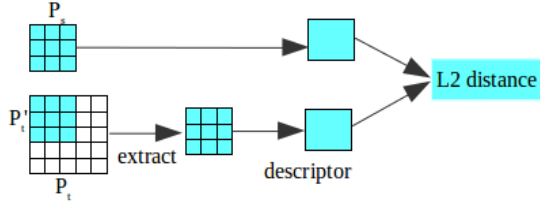


FIGURE 3 – Steps of SIFT descriptors comparisons between the patch  $P_s$  of the source image and the patches  $P'_t$  of the target image.

### 3 Experiments and results

The method is tested on two sets (left and right) of beetle mandibles. After verifying the dataset and suppressing of same unusable images (broken or containing a hidden mandible), it remains 290 images of right mandibles and 286 left mandible images. In all valid images, the position of a set of manual landmarks is indicated by biologists : 18 for each right mandible, 16 for each left mandible.

As discussed in previous section, before applying the SIFT descriptor to estimate the landmarks on the target image, we have to estimate the search space containing the landmarks.

Firstly, the segmentation is applied on both on source and target image. The Canny algorithm [2] was chosen to perform this step. To use the Canny algorithm, two thresholds value must be provided ( $T_{lower}$ ,  $T_{upper}$ ). As mentioned in [5], defining the threshold values is a difficult problem. If the threshold values are unsuitable, the contours could be far with ground truth (fewer contours or more noise). In our case, the lower threshold value is determined by analysis of the image histogram [6]. The ratio of two threshold values that we have chosen is 1 : 3 to consider a wide range of the values. During the Canny computing, the direction of the gradient of each pixel belonging to the contours is kept in order to be used later. At the end of the segmentation step, a simple algorithm is applied to remove edges inside the main contours.

Then, a Principal Component Analysis (PCA) is applied to register two lists of contours points from the source and the target [9, 4]. The lists of contour points are used as input. For each list of contours points, the centroid point and principal axis are computed. The centroid point coordinates are computed like the average coordinate of all contours points. The principal axis is a line connecting the centroid point to a contour point which has the minimum of average perpendicular distance to remaining contour points. Then, two lists of contours points are registered by computing the translation and rotation parameter values. The translation is computed as the distance between the centroid points of the source and the target. The rotation angle is the angle between the principal axes of two images. However in some case, the result of the segmentation step could contains noise, affecting to the registration step. To improve the registration, we have built up an iterative PCA until stabilization (PCAI). Finally, the SIFT descriptor allows to determine the estimated landmark on the target image.

We have run the method on all usable images. As results, the estimated landmarks are well positioned on the major part of targets but not in all. As precedingly raised, the mandible images can have different sizes due to the different sizes of beetles. We detected that our method is sensible to these differences. To improve the results, we have inserted a pre-process before the computing of the SIFT descriptor to estimate the difference of scale between the source and target images. The bounding box of the mandible contours in the source and target images is defined by checking the coordinate of the contour points. The scale between two images is defined as the ratio of the two bounding box sizes.

The Fig. 4 shows the final result for a right and a left mandible with manual and estimated landmarks. The estimated landmarks are quite near with the manual ones, as conformed by the following statistics.

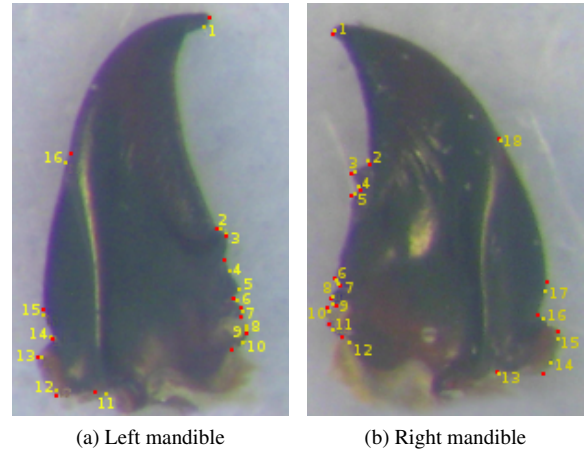


FIGURE 4 – The manual (in red) and estimated (in yellow) landmarks on a mandible.

First statistics are done on the mean accuracy of all landmarks on the target images. The error is computed as the distance between the manual and the corresponding estimated landmark on the target image with an accepted error from 1% to 2% of the bounding box's size (when we consider the scale of the image). According to this way, the results are shown in Fig. 5. The score of well-positioned landmarks is **87.03%** for the set of right mandibles and **78.82%** for left mandibles.

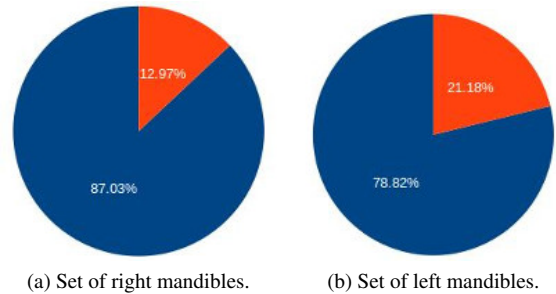


FIGURE 5 – The mean proportion of well and bad landmark locations of the two sets of left and right mandibles.

Beside the global results, we are also interested by the individual accuracy of each estimated landmark. The error measurement is the same with the first one (distance between manual landmark and the corresponding estimated landmark) but the acceptance is done with the standard deviation of the distances [1]. The Fig. 6 and 7 show the proportion of well estimated landmarks on each dataset. With 18 landmarks of right mandible, the highest proportion is obtained by the 1<sup>st</sup> landmark with **98.62%** ; the lowest proportion is **74.48%** for the 14<sup>th</sup> landmark. The remaining landmarks are also estimated with a high accuracy greater than **75%**. For left mandibles, the highest and lowest success rates are **93.01%** for the 1<sup>st</sup> landmark and **60.14%** for the 16<sup>th</sup> landmark. In this evaluation, we can see that the correct proportion on the 11<sup>th</sup> and 12<sup>th</sup> landmark of the left mandible and the 13<sup>th</sup> and 14<sup>th</sup> landmark of the right mandible are less than other landmarks. This is due to the noise of the contours at the base of mandibles higher than on the top part. The software MAELab proposes an implementation of the SIFT version for the automatic landmarks estimation. It is written in C++ and distributed as free library on the Github.

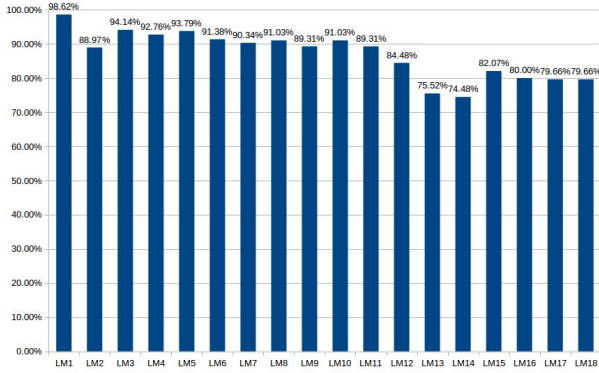


FIGURE 6 – The proportions of well estimated landmarks for each model landmark of right mandibles.

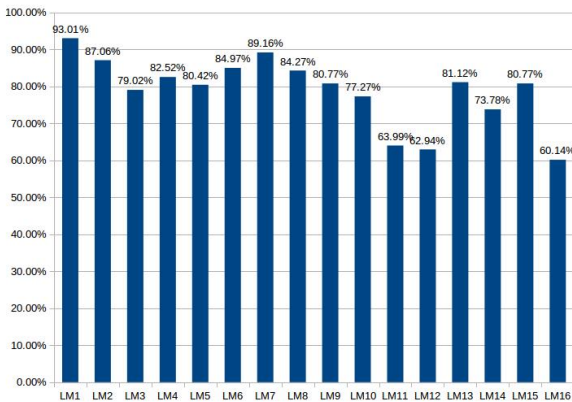


FIGURE 7 – The proportions of well estimated landmarks for each model landmark of left mandibles.

## 4 Conclusion and discussion

The landmark positioning is the main way for the image analysis in biology. In this paper, we presented a solution based on the SIFT descriptors for the segmentation and registration processes of the landmark estimation on beetle mandibles. Firstly, each mandible has been segmented and its contours are extracted. The SIFT descriptor is computed on each contour point to find the best matching position of estimated landmarks. The results show that this new method succeeds in locating all landmarks in a request image. The accuracy of the method is sufficient to be proposed to biologists as a replacement of the manual positioning. Moreover, considering the previous work [6], the precision of the position of the estimated landmarks has been improved. The next step consists to increase the number of well positioned landmarks in the bottom of the mandible.

## Références

- [1] J Martin Bland and Douglas G Altman. Statistics notes : measurement error. *Bmj*, 313(7059) :744, 1996.
- [2] John Canny. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6) :679–698, 1986.
- [3] Michael W Davidson and Mortimer Abramowitz. Molecular expressions microscopy primer : Digital image processing-difference of gaussians edge enhancement algorithm. *Olympus America Inc., and Florida State University*, 2006.
- [4] Ian Jolliffe. *Principal component analysis*. Wiley Online Library, 2002.
- [5] Dehua Li Jun Zeng. An adaptive canny edge detector using histogram concavity analysis. *International Journal of Digital Content Technology and its Applications*, 5, 2011.
- [6] L Lê Vãnh, M Beurton-Aimar, JP Salmon, A Marie, and N Parisey. Estimating landmarks on 2d images of beetle mandibles. *WSCG*, 2016.
- [7] David G Lowe. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. Ieee, 1999.
- [8] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2) :91–110, 2004.
- [9] Jonathon Shlens. A tutorial on principal component analysis. *arXiv preprint arXiv :1404.1100*, 2014.