SIFT descriptor to set landmarks in biological images

Van Linh LE[1,3], Marie BEURTON-AIMAR[1], Adrien KRAHENBUHL[1], Nicolas PARISEY[2]

[1]Laboratoire Bordelais de Recherche en Informatique
351, cours de la Libration, F-33405 Talence cedex, France

[2]Institute for genetics, the environment and plant protection (IGEPP), INRA 1349
35653, Le Rheu, France

[3]Faculty of Information Technology, Dalat university
1 Phu Dong Thien Vuong, Dalat, Lamdong, Vietnam

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

Résumé – Les auteurs publiant au GRETSI et utilisant le traitement de texte LaTeX 2ε trouveront ci-dessous quelques indications destinées à leur faciliter la tâche. Le fichier `gretsien.tex` qui contient le présent document respecte les contraintes fixées ; recopiez le, par exemple sous le nom `monarticle.tex`, et placez votre texte aux endroits appropriés.

Abstract – Image analysis is a large field of computer vision and it has applied in practice with many application in the different majors such as medicine, machine vision, biology, ... In the past of biology techniques, microscopic techniques were applied to study the structure details, but the current research requires more accuracy in the area, perimeter, localization of the object in the image by using the suitable features. In this content of this paper, we will focus on the problem of setting the landmarks on biological images which are used in many biological studies. The landmarks are indicated by applying a combination of SIFT and some techniques. Besides, we also show a difference way to use SIFT descriptor with another size of the sample region. The efficiency of the method is evaluated on two set of images: left and right mandibles of beetle. The complete work is implemented in MAELab and freely available as a library.

1 Introduction

Morphometry analysis is an important field of image analysis in biology. It is used to characterize the shape variations of the organism. From obtained information, the biologists can evaluate the evolution of the organism or detect the difference between the organisms. Depending on the requirements of the application, the output of analysis process can be the shape, the color, the pattern,... or the landmarks (points of interest). The landmarks are the 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 have been manually identified. The manual identification is time-consuming and that could vary depend on the identifying operator.

In this paper, we focus on the method to automatically determine the landmarks in 2D biological images (see Fig. 1), specifically beetle images. The proposed method consists segmentation, registration, and descriptors comparison techniques. Besides, this method is also shown as a difference ways to apply the SIFT descriptor for keypoint detection. The work is evaluated on the dataset of beetles from Brittany lands with

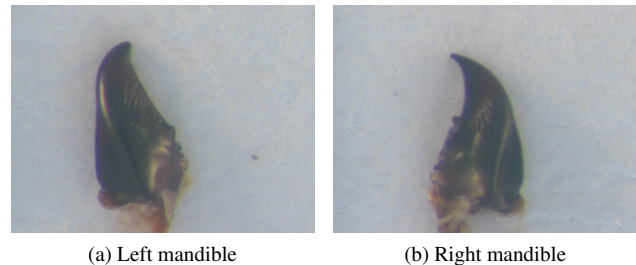(a) Left mandible   (b) Right mandible

FIGURE 1 – The pictures of beetle mandibles took by biologists.

a collection of 293 beetles. For each beetle, a set of landmarks has been manually located by the biologists. In our study, we used that dataset as ground truth to evaluate the position of the automatically estimated landmarks.

2 Landmark descriptor

For each beetle, the morphometric landmarks have been manually identified on mandibles images by the biologists : 16 and 18 landmarks for each left and right mandible, respectively. The consideration problem is how to detect automatically the landmarks on the mandible image to replace the manual

ones. In the whole of the process, a source and a target image are used. The landmarks will be estimated on the target image by using the manual landmarks of the source image. Note, the source image is chosen randomly from the set of images.

In this section, firstly, we have a summary about the general utilization of SIFT. Then, we will discuss detail about our method with a difference way of SIFT to estimate the landmarks.

2.1 SIFT method

SIFT method has been proposed by Lowe [7, 8]. It is used to extract distinctive invariant features from the images. Besides, the features from SIFT can use to match between different views of source and target images such as rotation, scale, noise. SIFT method includes four stages : (1) scale-space extrema detection, (2) keypoint localization, (3) orientation assignment, and (4) keypoint descriptor.

In the first stage, a difference of Gaussian (DoG) [3] (Eq. 1) function is applied to identify the interest points on all scale and orientation of image. The keypoints are taken as the maximal and minimal of the result of DoG function at mutliple scale.

$$D(x, y, \sigma) = (G(x, y, k_i\sigma) - G(x, y, k_j\sigma)) * I(x, y) \quad (1)$$

Where :
— $D(x, y, \sigma)$ is the DoG image
— $I(x, y)$ is the original image
— $G(x, y, k\sigma)$ is the Gaussian blur at scale $k\sigma$.

The scale-space extrema detection produces too many kepoint candidates, some of them are unstable. In the second stage of SIFT, the key point candidates are localized and refined by suppress the key points which have the low contrast or are poorly localized along an edge.

Then, the orientation and gradient magnitude of key points are calculated by applyiing the following equations (Eq. 2) :

$$m(x, y) = \sqrt{v_x^2 + v_y^2}$$
$$\theta(x, y) = tan^{-1}(v_y/v_x) \quad (2)$$

Where :
— $v_x = I(x+1, y) - I(x-1, y)$
— $v_y = I(x, y+1) - I(x, y-1)$
— $I(x, y)$ is the gray value at position (x, y) in the image,
— $m(x, y)$ is the gradient magnitude of the pixel at position (x, y),
— $\theta(x, y)$ is the orientation of the pixel at position (x, y).

At the end, the descriptor is computed for each key point based on the orientation and gradient magnitude to ensure the invariance to image location, scale, and rotation. First, an orientation histogram with 8 bins is created on 4×4 pixel neighborhoods. This histogram is computed from the magnitude and orientation values of each pixel in the 4×4 region. The descriptor around a keypoint is a histogram of a 16×16 region around it which is combined from the 4 histograms of the 4×4 subregions. Then, all the values of the histogram are presented as a vector. To reduce the effects of illumination changes,

the vector is normalized and applied a threshold of 0.2 and the vector is again normalized.

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 correspondence keypoints between the source and target image. The problem is carried from the choosing the best points from the large set of the candidates. This problem is solved in our method by limit the search space before applied the SIFT, plus changing the size of the grid to calculate the SIFT descriptor.
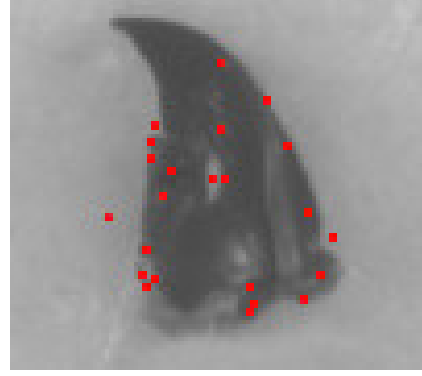

FIGURE 2 – SIFT keypoints in a right mandible

2.2 SIFT using into landmark area

As mentioned in section 2.1, we have tried to use original of SIFT to detect the landmarks. But the result shows a lot of candidates for the estimated landmarks. Additional, the position of the detected points are very far when we compare with the ground truth ones (manual landmarks). To hope for a better result, we use the SIFT method with another way, we do not consider all points of the image as usually ; instead, we register two images first. Then, the area (called patch) around each source landmark is detected then a larger patch is extracted in the target image at the same position. The SIFT descriptor is calculated as the original way but the size of the sample is 9×9 (instead 16×16). This size has been obtained 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 L2-distance with following equation (Eq. 3) :

$$L(D1, D2) = \sum_{i=0}^{n} \sqrt{(D1_i - D2_i)^2} \quad (3)$$

Where :
— n is the number of directions
— $D1$ and $D2$ are two descriptors of size n,
— $D1_i$ and $D2_i$ are the ith descriptor values.

The way that we apply SIFT into our work is painted in Fig. 3. To detect the target landmarks, a registration is computed between the source and target images. Then, the patch P_m of source and P_s of the target are created with the size of P_m is

smaller than the size of P_s. For each pixel in P_s, a sub-patch P'_s is extracted with the same size of P_m. When the P'_s have a part outside of P_s, the pixels do not belong to the patch will be considered. Then, the distance $L(P_m, P'_s)$ is computed by using Eq. (3). This work is finished when all the pixel in patch P_s are considered. The position of estimated landmark corresponds to the position of the sub-patch P'_s with smallest distance $L(P_m, P'_s)$. Finally, the position of the estimated landmarks are set in the original position of the target image.
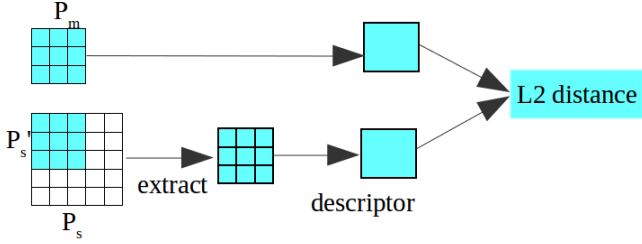

FIGURE 3 – Steps of SIFT descriptors comparisons between the patch P_m of the source image and the patches P'_s of the target image.

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

3 Experiments and results

The method is tested on two sets (left and right mandible) of beetles. After verifying the dataset, it remains 290 images of right mandibles and 286 left mandible images. In all valid images, a set of manual landmark is indicated by biologists : 18 for each right mandible, 16 for each left mandible.

As discussed, before applying the SIFT descriptor to estimate the landmarks on the target image. We have estimated the search space that contains the landmarks.

Firstly, the segmentation is applied on both of source and the target image. The Canny algorithm [2] was chosen to finish this step. To use the Canny algorithm, two thresholds value must be provided (T_{lower}, T_{upper}). As mentioned in [5], indicating 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 indicated by analysis the histogram of the image [6]. The ratio of two threshold values is chosen as $1 : 3$ to consider a wide range of the values. During Canny computing, the direction of the gradient of each pixel belonging to the contours is kept and will be used later. At the end of segmentation step, a simple algorithm is applied to remove the edge inside the main contours.

Then, the PCA iteration is computed to register two lists of contours points from the images [9, 4]. The lists of contour points are used as the input. The centroid point and principal axis are computed for each image contours. Two lists of contours points are registered by computing the translation and rotation parameter values. However in some case, the result of the segmentation step could exist the noise, it could affect to the registration step. To improve the registration, we have built up the PCA by iterating until stabilization (PCAI). Finally, the SIFT descriptor is finished to determine the estimated landmark on the target image.

We have run the method on all usable images. The results are given in difference : estimated landmarks are well positioned on some target images but not pleasing on others. As raised, the mandible images can be exist difference sizes because beetles have also difference sizes. We detected that our method is sensible to this parameter. To improve the result, we have inserted a pre-process step before the computing of the SIFT descriptor to estimate the scale between source and target image. The bounding box of contours in source and target are defined by checked the coordinate of the contour points. The scale between two images is the ratio of two bounding boxes.

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.
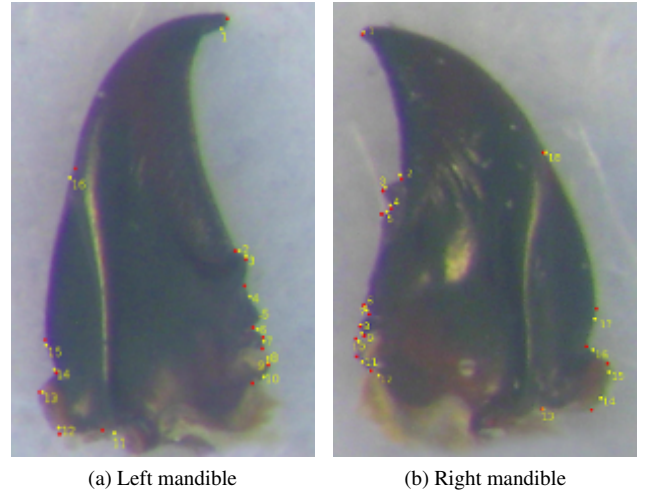

(a) Left mandible   (b) Right mandible

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

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

Beside the global results, we are also interested by the accuracy of each estimated landmark. The error measure is the same with the first one (distance between manual and corresponding estimated landmark) but the acceptance is done with standard deviation of the distances [1]. Figs 6 and 7 show the

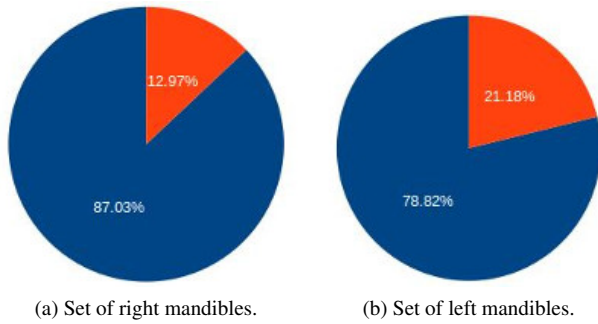(a) Set of right mandibles.   (b) Set of left mandibles.

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

proportion of well estimated landmarks on each dataset. With 18 landmarks of right mandible, the highest proportion is belong to the 1st landmark with **98.62%** ; the lowest proportion is **74.48%** for the 14th landmark. The remaining landmarks are also estimated with a high accuracy greater than **75%**. For left mandibles, the highest and lowest success rate are **93.01%** for the 1st landmark and **60.14%** for the 16th landmark. In this evaluation, we can see that the correct proportion on the 11th and 12th landmark of the left mandible and the 13th and 14th landmark of the right mandible are less than other landmarks. The reason is come from the noise of the contours at the base of mandible is higher than the top of the mandible.
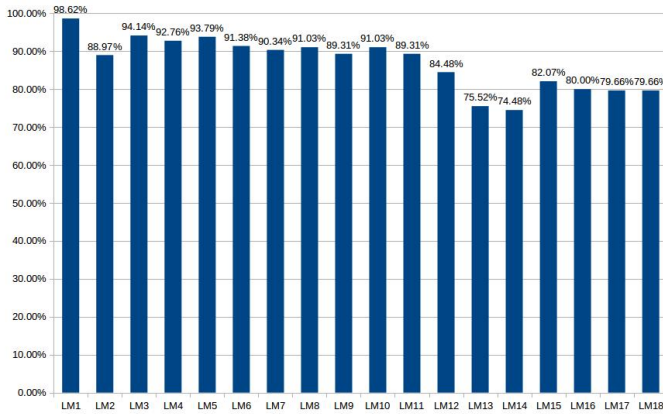

FIGURE 6 – The proportions of well estimated landmark for each model landmark of right mandibles.

4 Conclusion and discussion

The landmarks setting is the main way to analysis image in biology. In this paper, we have described a difference way to use the SIFT descriptor along with the segmentation and registration processes to estimate the landmark on the beetle. Firstly, each mandible has been segmented. Then, two contours are registered. Finally, SIFT descriptor is applied to find the best matching position of estimated landmarks. The results show that our method succeeds in locating all the landmarks of the
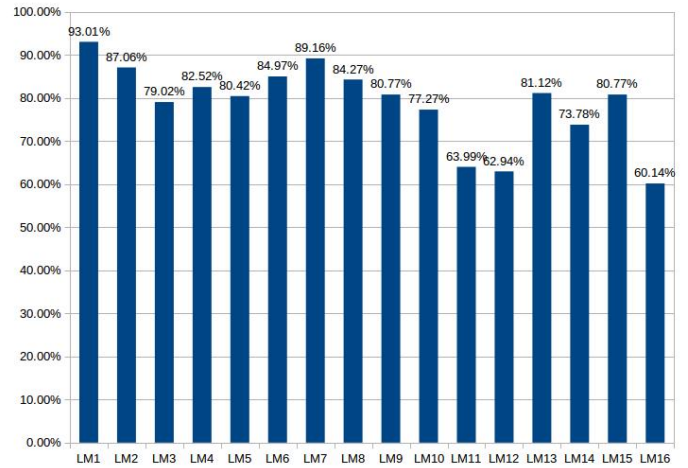

FIGURE 7 – The proportions of well estimated landmark for each model landmark of left mandibles.

image. The accuracy of the method is sufficient to be proposed to biologists as a replacement of the manual measures. Moreover, considering with previous work [6], the position of the estimated landmarks are improved a little bit. From now, the next stage consists the improving to increase the correct accuracy of the 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.