國立臺灣大學生農學院生物產業機電工程學系

學士專題

Quantitative Evaluation of the Floral Shape Variation in *Sinningia speciosa* Domestication

大岩桐花表現型的影像處理及特徵分析

翁立剛

指導教授：郭彥甫 教授

中華民國 101 年 12 月

臺灣大學生物產業機電工程學系

學士專題著作保證書

本人保證所撰之下述學士專題絕無涉及任何抄襲之情事。

專題名稱：

Quantitative Evaluation of the Floral Shape Variation in *Sinningia speciosa* Domestication

大岩桐花表現型的影像處理及特徵分析

學生姓名：　　翁立剛

學號：　B98611036

國立臺灣大學

生物產業機電工程學系

翁立剛君學士專題

Quantitative Evaluation of the Floral Shape Variation in *Sinningia speciosa* Domestication

大岩桐花表現型的影像處理及特徵分析

經考試合格特此證明

指導教授

中華民國　　　年　　　月

Abstract

Floral shape variation is of great interests to botanic scientists and evolutionary biologists. This study quantifies shape variation in Darwin’s Gloxinia (*Sinningia speciosa*) using image processing and geometric morphometric methods. Darwin’s Gloxinia has two shape forms – bilateral symmetric (zygomorphic) wild type and radially symmetric (actinomorphic) ornamental breeding (peloria). It is characterized by its easy crossing within its species; hence, is selected to be the study object in this research. In this work, the wild type Darwin’s Gloxinia was crossed with the peloric one. The face view and side view images of the second generation crossed flowers were taken by the regular digital camera. Image processing algorithms were applied to segment the flowers from their background, and to acquire landmarks, i.e., the characteristic points of the flowers. Generalized Procrustes analysis was applied to the landmarks to define the flower shapes excluding their sizes, rotation, or transformation information. The variation in floral shapes was then investigated by principal component analysis. It was found that the first three principal components capture most structure variations in side-view and face-view of the corolla. It is shown that the shape variation of Darwin’s Gloxinia can be adequately captured and quantified by our approach.

Keyword: Floral shape variation, Geometric morphometrics, Principal component analysis, Generalized Procrustes analysis

Acknowledgement

I would like to express my appreciation to my thesis advisor, Professor Yan-Fu Kuo, for his guidance and supports. His profession in evolutionary biology, machine learning and writing skill improved my research and presentation skills a lot.

I would like to express my appreciation to my advisor, Professor Ta-Te Lin, for his guidance and supports. His expertise in image processing and development environment improved my programing skills and presentation skills a lot.

I would like to thank Hao-Chun Hsu and Tzu-Keui Lee. Without their contribution to this research, this paper will go on so smoothly. Mr. Hsu’s profession in evolutionary biology and cultivation techniques taught me many knowledge of plants. Mr. Lee’s profession in machine learning taught me to understand the principles and applications in geometrics analysis methods.

Tables of Contents

List of Figures

List of Tables

Introduction

This study aimed to the analysis on flower shape variation in Darwin's Gloxinia (*Sinningia speciosa*). Darwin's Gloxinia is a species with high degree of diversity in appearances. Its flower is trumpet-shaped with narrow tube and flared petals (see Fig. 1). In the study, an actinomorphic, (radially symmetric, see Fig. 1(a)), cultivar was crossed with a zygomorphic, (bilaterally symmetric, see Fig. 1(b)), one (Hsu et al., 2009). The derived second generation (F2) population with different appearance is used as subject to discuss floral shape variation.

The shape of flower is usually represented by a set of characteristic coordinate points, also referred to as landmarks (Adams et al., 2004; Klingenberg, 2010), along flower contour. These landmarks are in two forms – primary and secondary landmarks. In this study, the primary landmarks are defined as the intersection points of adjacent lobes; the secondary landmarks are defined as the points uniformly distributed on the flower petal contour between two adjacent primary landmarks. By the landmarks, the shape variation can be quantified statistically.

Image processing algorithms were applied on flower images for landmark identification in preventing of the damages to samples. GrabCut algorithm was implemented to segment the flower images from their background images. Suzuki85 algorithm was implemented to retrieve the contour line from foreground images. Landmarks were then identified after the steps above. The conversion ratio between pixel and centimeter in the image must be calculated from the ruler ticks in the background to estimate the flower specimen size due to the non-fixed focus photographing.

Geometric morphometrics analysis (GM) is a collection of coordinate points of landmarks that examine the shapes of flower quantitatively. General Procrustes analysis (GPA) was applied to eliminate variance irrelevant to shape, e.g., variance of translation, orientation, and scaling, from the landmark dataset. Principal component analysis (PCA) was used to capture the major trends of variation.

This objective of this study is to quantitatively investigate the floral shape variation among the F2 specimens of Darwin's Gloxinia. A user interface program was developed to semi-automatically identify and collect landmark coordinates from the floral images.

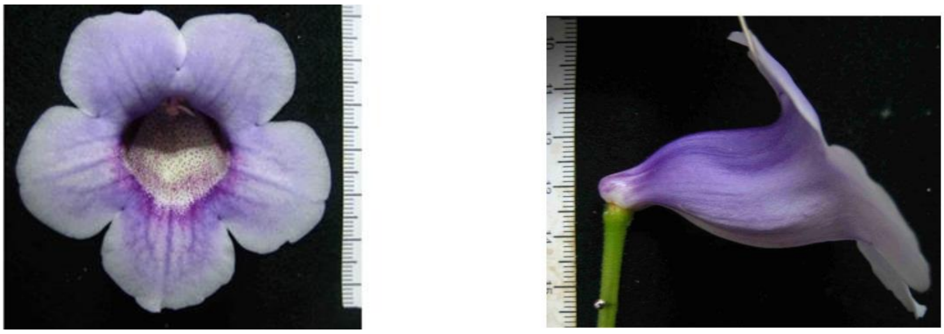
Materials and Methods

Flower samples

The flower samples for this study were developed by inbreeding two species of *Sinningia speciosa*. The first generation (F1) was bred through intercrossing two parents, accession 'Carangola' and cultivar 'Peridots Darth Vaders' (see Fig. 1). The second generation (F2) was developed and had segregated by selfing of one F1 plant.

Fig. 1. (a) Accession 'Carangola' face view; b, side view) and cultivar 'Peridots Darth Vaders' (c, face view; d, side view)

The flowers of F2 inbred strain were selected as experimental samples (see Fig. 2). The sample for shape comparison must be homologous. Therefore, the F2 plants with less or more than five petals are excluded. The face view flowers of an inflorescence were subjected to image acquisition to prevent including abnormal floral patterns (Rudall and Bateman, 2003). A total of 73 F2 plants was selected, and 2 flowers were sampled from each individual plant.



(a) **(b)**

Figure 2. The images of experimental samples, F2 plants, in the (a) face view and (b) side view

Image acquisition

The face and side view images of each flower sample in full bloom were captured with Canon® SD1000 digital camera. The face and side view images were photographed confronting the plane of unfolded petal lobes (see Fig. 1(a), 1(c)) and the dorsiventral plane **(ref)** of the flowers (see Fig. 1(b), 1(d)), respectively.

Graphic user interface

The graphic user interface (GUI) was implemented with a program written with Qt Creator (Nokia®) and OpenCV (Intel®), and developed for landmark identification. The GUI (see Fig. 3) is for user through mouse clicks to select the ROI, primary landmarks, display the pre-processed, processed image and information of coordinates of landmarks.

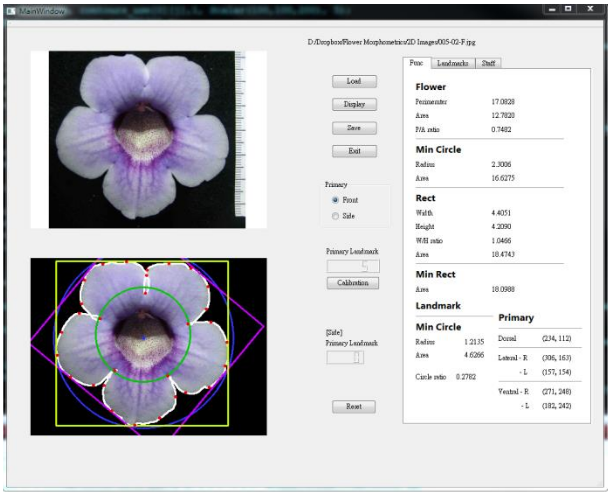


Figure 3. The image processing GUI

Floral landmark identification

Image processing algorithms were applied on flower samples for landmark identification. Here the landmarks of a flower are a set of characteristic coordinate points on flower contour that are used to describe the flower shape. The landmarks are in two forms – primary (homologous) and secondary landmarks. In this study, the primary landmarks are defined as the intersection points of adjacent lobes; the secondary landmarks are defined as the points uniformly distributed on the flower petal contour between two adjacent primary landmarks. Figure 4 shows the flowchart of the landmark identification.

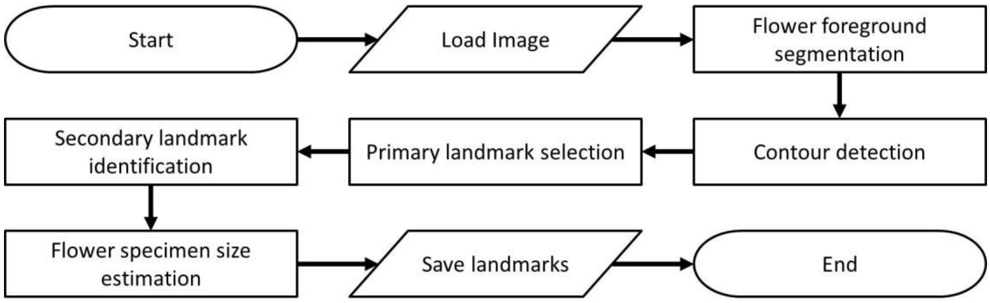


Figure 4. Flowchart of flower landmark identification

Flower foreground segmentation

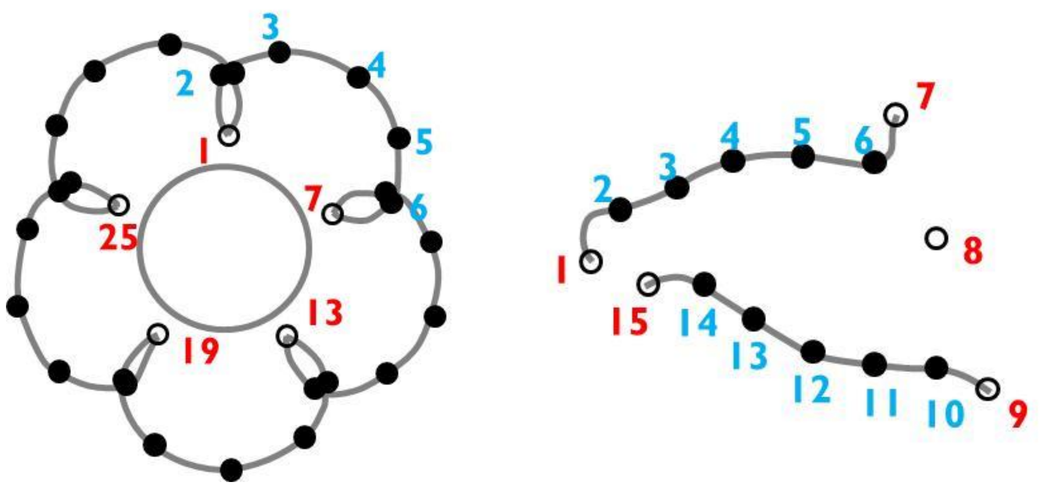
GrabCut algorithm (Rother *et al.*, 2004), an interactive tool, was applied for flower foreground segmentation. In the procedure, a region of interest (ROI) enclosing the flower object was first assigned by user. The ROI was designated to contain the foreground; the region outside ROI was considered as part of the background. The foreground image was then determined from the ROI based on the color and texture differences using statistical models. The background image could also be obtained by extracting the original image from the foreground image.

Contour detection

Suzuki85 algorithm (Suzuki *et al.*, 1985) was applied to identify the flower contour lines from the foreground images. The results were binary images that contained only pixels of the contour lines. Note that, in the face view images, the lobe contours usually overlap one the other by the intersection of two lobes (see Fig. 2). This makes the contour of the lobe in the back invisible and undetectable. To solve this problem, it was assumed that the overlapped lobe contour sections were symmetric. The mirror mapping of the front lobe contour was used to replace the invisible section of the back lobe contour.

Primary landmark selection

The primary landmarks were manually selected. This is because they are at the intersection of lobe, tube, or sepal, and are challenging to be accurately detected by image processing algorithms. The landmarks were assigned through mouse clicks in a user interface developed in the program. Five landmarks were chosen for both the face and side view images, respectively. In the face view images, the landmarks were assigned starting from dorsal lobe and proceeding in counterclockwise order, labeled as number 1, 7, 13, 19, and 25. In the side view images, the landmarks were assigned from the intersection point of sepal and tube, labeled as 1, 7, 8, 9, and 15. Figure 5 shows the primary landmarks and their assigned numbers.



(a) **(b)**

Figure 5. Numbers assigned to the primary and secondary landmarks in the (a) face view image and (b) side view image

Secondary landmark identification

The secondary landmarks were identified automatically by the procedure below. In a front view image, the complete flower contour was partitioned into lobe contour sections using the primary landmarks as the separation points. Each lobe section was then further divided into six equally long segments. The endpoints of the segments were defined as the secondary landmarks. Here a total of five secondary landmarks was chosen for each lobe because it had been shown that five points were adequate to describe the lobe shape **[Ref.]**. In a side view image, the up and down flower contour sections between primary landmark 1 and 7, and 9 and 15 were respectively divided into six equally long segments. The endpoints of the segments were defined as the secondary landmarks. Figure 5 shows the secondary landmarks and their assigned numbers. A total of 35 secondary landmarks, including 25 in the face view and 10 in side view images was collected for each flower.

Flower specimen size estimation

The images of the flower specimen were taken with a scale ruler placed vertically in the background. The ruler was isolated from the flower (see Fig. 2) and would be contained in the background images from the grabCut algorithm. The image was converted into binary for counting number of pixel. A histogram that gave the number of white pixels for each row was first generated for of a background image. The horizontal stripes in the histogram would correspond to the ruler ticks, each of which represented one millimeter in length. The conversion ratio of the ruler tick and image pixel was calculated by averaging the number of pixels between two neighboring stripes throughout the histogram. The flower specimen size was then estimated based on the ratio.

Morphometrics

Morphometrics was applied to the landmark coordinates from image processing for evaluation of floral shape variation. The floral shape is defined as the form that doesn't alter by translation, scaling, or rotation. Geberal Procrustes analysis (GPA) was performed to remove the irrelevant information from the coordinates of landmarks (see Fig. 6). In this procedure, the mean shape, average of the coordinate points of landmarks, was calculated. The centroid of each individual landmark coordinates was translated to the mean shape. The translated landmark coordinates was scaled and rotated to minimize the deviation between it and the mean shape. This was applied to all the individual landmark coordinates recursively until obtaining the minimum deviation. The final coordinates are called GPA landmarks.

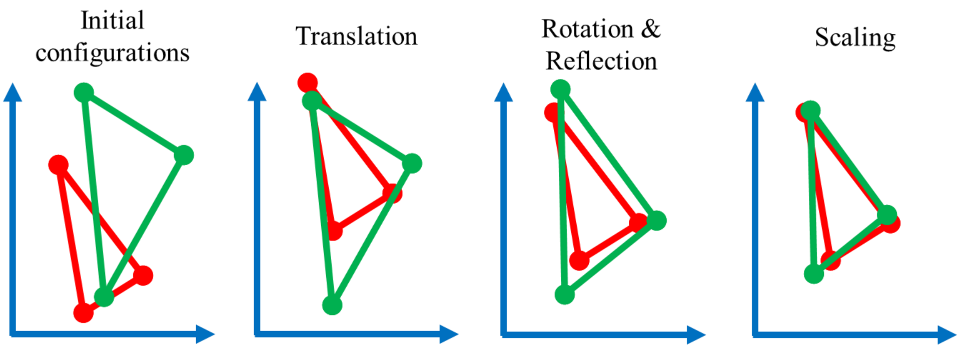


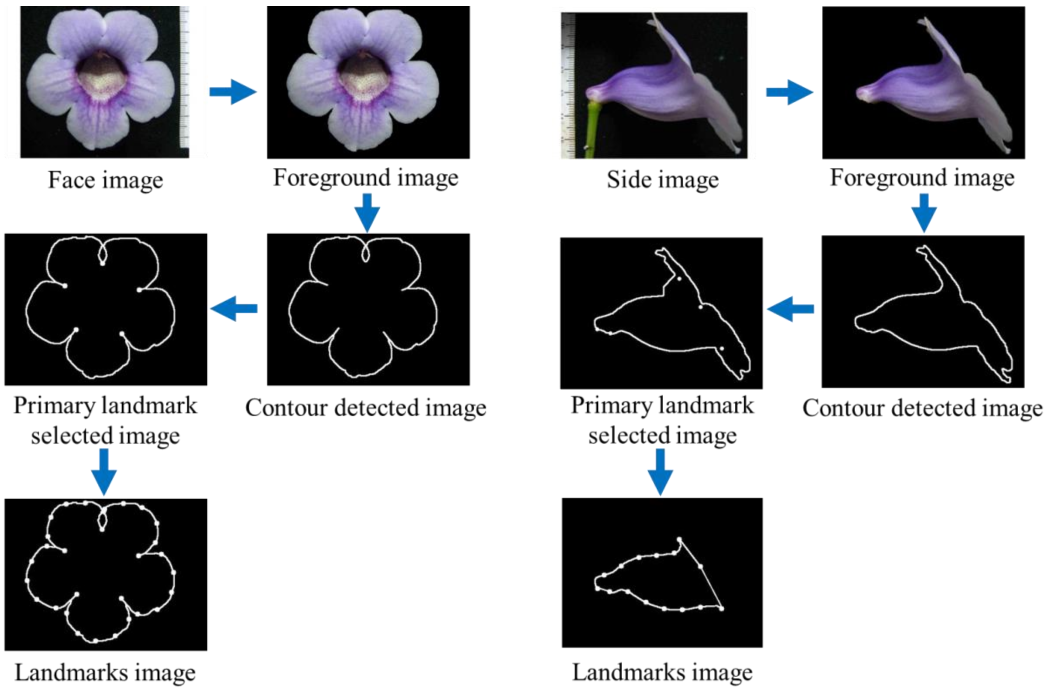
Figure 6. An overview of general Procrustes analysis

Principal component analysis

Principal component analysis (PCA) was applied to the GPA landmarks for dimension reduction. The GPA landmarks of a flower are highly correlated. Therefore, there exists a certain level of redundancy in the high-dimensional landmark vectors. Practically, the shape variation can be adequately represented with only a few significant variables. The PCA was performed to project the GPA landmarks into a set of orthogonal variables, namely principal components (PCs). The first few PCs account for most of the variation implicit in the landmarks and can well summarize the variance in shape with little loss of information. The analyses of variation were then performed with only the first few PCs.

Results and Discussion

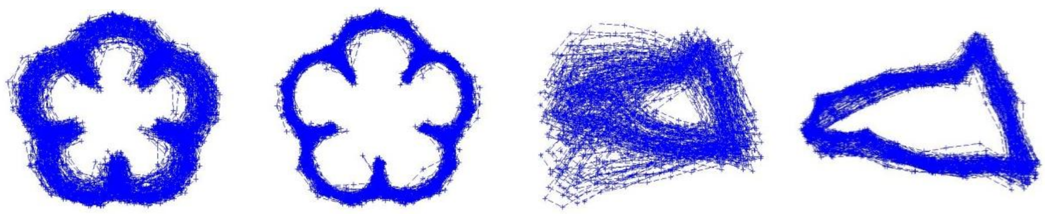
All the landmarks of flower were identified by the proposed procedure. Figure 7 shows the images at each stage. First, the original flower image was extracted. Then the flower contour was diagnosed. After that, the primary landmarks were selected manually in the GUI. The secondary landmarks were identified by the procedure.



(a) **(b)**

Figure 7. Landmark identification of the (a) face view and (b) side view image

The GPA was carried out to obtain shape information from the landmarks. Figure 8 shows the images before and after GPA in the front view and side view. It can be shown that the scaling, translation or rotation effects from the original images were minimized.



(a) **(b)** **(c)** **(d)**

Figure 8. The front view images (a)before and (b) after GPA, and side view

images (c) before and (d) after.

The principal components were acquired by the GPA landmarks calculated from PCA. The first (PC1), second (PC2) and third (PC3) principal components represent floral shape variation accounted for 19.2%, 16.0% and 15.8% in face view image and 54.5%, 13.5% and 9.4% in side view image, respectively (see Table 1). In the face view, the PC scores are too low to represent the floral shape variation well, though. the PC scores show a high percentage to the floral shape variation in the side view.

Figure 9 shows effect of first principal components in the face and side view. From the figure, it has shown that the first PC score is corresponded to the degree of overlap between flower lobes; in the side view, in the side view, the first principal component score obviously corresponds to the symmetry between dorsal and ventral of flower, also referred to the dorsi-ventral asymmetry.

Table 1 Contribution of principal components related to floral shape variation.

	Face view				Side view			
	PC1	PC2	PC3	Sum	PC1	PC2	PC3	Sum
Floral shape % of total	19.2	16.0	15.8	51.0	55.0	13.5	9.4	73.9

PC1 **-2STD** **Mean** **+2STD**

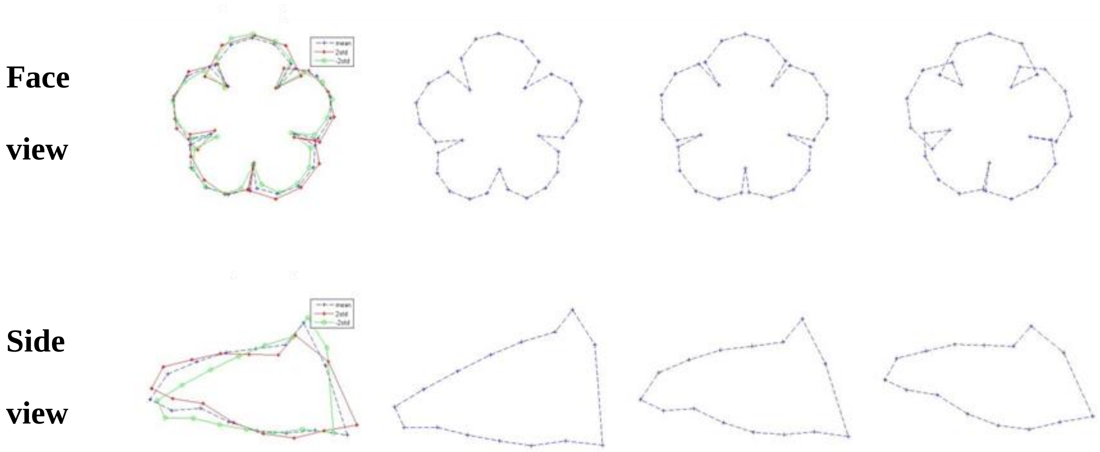


Figure 9. Effect of first (PC1) principal components in the face and side view

Conclusion

In this study, we have developed a program for landmark identification using image processing algorithms. The GM analysis successfully evaluated the floral shape variation. The indicative parameters were observed to represent the floral shape. In the face view, the PC1 is corresponded to the degree of overlap between flower lobes. In the side view, the PC1 is corresponded to the dorsi-ventral asymmetry.

References

Dalayap, R. M., M. A. J. Torres, and C. G. Demayo. 2011. Landmark and Outline Methods in Describing Petal, Sepal and Labellum Shapes of the Flower of Mokara Orchid Varieties. *International Journal of Agriculture and Biology* 13 (5):652-658.

Kawabata, S., M. Yokoo, and K. Nii. 2009. Quantitative analysis of corolla shapes and petal contours in single-flower cultivars of lisianthus. *Scientia Horticulturae* 121 (2):206-212.

Ueda, K., T. Baba, Y. Nakagawa, and K. Amano. 2005. Detection of Scale Intervals in Digital Images. Paper read at Data Engineering Workshops, 2005. 21st International Conference on, 05-08 April 2005.

Yoshioka, Y., M. Honjo, H. Iwata, S. Ninomiya, and R. Ohsawa. 2007. Pattern of geographical variation in petal shape in wild populations of Primula sieboldii E. Morren. *Plant Species Biology* 22 (2):87-93.

Yoshioka, Y., H. Iwata, R. Ohsawa, and S. Ninomiya. 2004. Analysis of petal shape variation of Primula sieboldii by elliptic Fourier descriptors and principal component analysis. *Annals of Botany* 94 (5):657-664.

花朵特徵非常多樣，就算同種間花朵的表現型都有許多不同，例如花朵的形狀 (Kawabata et al., 2009)、花冠筒的形狀、花朵外輪廓之生長標點 (Dalayap et al., 2011)、花冠筒傾斜程度、花朵對稱性等等特徵，為了能從特徵中得知不同表現型對花朵造成的影響，

目前許多型態類的研究主要對花瓣型態的變異進行研究 (Yoshioka et al., 2007; Dalayap et al., 2011; Kawabata et al., 2009; Yoshioka et al., 2004)，因此提出對其他特徵進行量化分析，希望藉由不同特徵的量化分析，找出新的花朵型態變異之量化方法。