Tina Memo No. 2008-007
Presented at LASR 2008, Leeds.

Automatic Identification System for Morphometric Landmarks.

Sasirekha Palaniswamy, Neil A Thacker and Christian Peter Klingenberg.

Last updated
21 / 7 / 2008

Imaging Science and Biomedical Engineering Division,
Medical School, University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT.

Automatic Identification System for Morphometric Landmarks

Sasirekha Palaniswamy[1], Neil A Thacker[2] & Christian Peter Klingenberg[1]

[1]Faculty of Life Sciences, [2]Imaging Science and Biomedical Engineering,
The University of Manchester

1 Introduction

In this paper, we present the automatic extraction of morphometric landmarks (see fig:1(a)) in digital images of *Drosophila*. Morphometric landmarks are points that can be defined in all specimens and located precisely. They establish an unambiguous one-to-one correspondence among the specimens and are widely used in shape analysis (Bookstein 1991, Dryden and Mardia 1998). A typical shape analysis study involves several thousand digital images and extracting landmarks manually is time consuming. To address these problems, researchers have focused on using specialised algorithms and semi-automated methods to enhance the speed and reliability of the digitization process (Houle et al., 2003). Although these methods increase the efficiency of human effort, the need for the manual observer has not been eliminated as the methods require the operator to digitize the initial set of landmark co-ordinates. Also, the systematic variations between different individuals raises an issue relating to reliability and repeatability of this process of digitisation. Therefore, complete automation of the process has been identified an important goal.

2 Materials and Methods

The dataset (Cambridge data) consists of 856 gray scale images of the *Drosophila* wings. The dataset has two repeat images of each wing. The training data consists of two independent sets of reference (training) images with 5 images in each set and their landmark co-ordinates. The analysis system is constructed from four key stages: a feature based detection of the fly wing structure, recording the compact invariant shape descriptors using the pairwise geometric histogram (PGH) representation, global estimation of the pose using the probabilistic Hough transform and finally a correlation based refinement of individual features.

2.1 Feature Extraction

The feature extraction stage involves extracting essential information from the digital images and retaining only those features that we are interested in. The image is preprocessed with a Difference of Gaussian (DoG) filter. The conventional Canny edge detection algorithm is modified to detect the ridge (edge) features in this study. The enhanced features and corresponding orientation estimates are provided to a similar framework which extracts extended structures of linked features using hysteresis thresholding and non-maximal suppression.

2.2 Pairwise geometric histograms (PGH)

Geometric relationships between the underlying edge features, such as angle and perpendicular distance, provide an efficient and robust shape description. The invariant shape characteristics are encoded by a set of geometric features by means of a histogram. The shape correspondence between the model (training) and the scene histograms can be established using the Bhatttacharya similarity metric (Rockett et al., 1997).

2.3 Hough Transform

The Hough transform finds spatially consistent groups of features within a scene, identifying the presence of a shape at some position and orientation. The probabilistic Hough transform is used to locate models using the positions, orientations and scales hypothesised by the scene line labels. The orientation is determined using separate one parameter Hough transforms.

2.4 Template Matching

A template based correlation matching is applied to obtain sub-pixel accuracy of landmarks. It is equivalent to performing a least squares comparison of the image regions with one free grey level parameter. In addition the least squares values that provide the best matching score for the given template and scene region are output. These values not only support quality control during the data analysis, to check the best matching templates, but also enables combination of landmark positions for improved localisation.

The landmark extraction is carried out independently for each of the landmark co-ordinates using different reference images as the feature extraction process on each of the landmark is independent of one another. Using multiple reference images improves the reliability of the results obtained and ensures an improvement in the accuracy of the final landmark estimation. To sensibly combine the multiple estimated locations (using different reference images) into a unique hypothesised landmark location, we use the least square matching scores to generate ranking. The landmark values obtained by the top ranking reference images are compared to ensure that the values are consistent.

3 Results

The landmarks obtained by the automated system is evaluated to test the statistical accuracy, robustness and reliability of the automated system. The standard deviation between the landmark co-ordinates obtained by the automated system and manual digitization is estimated. The results indicate that the landmarks can be located with high reliability as 70% of the landmark co-ordinates in the dataset (consisting of 896 images of size 1280 x 1022 pixels) were located within 0.5 subpixel accuracy, 90.64% within 2 pixels and 98.36% of the landmarks were located within an accuracy of 5 pixels from the manually digitized landmarks (Palaniswamy 2008). The plots (fig:2(a)-2(b)) shows the results of the automated performance for the X and Y as a function of outlier removal. It is clear that about 1% outliers need to be removed to make the results consistent.

Finally, an overall comparison is performed between manual and automatic mark-up locations in the context of morphometric analysis using procrustes alignment to evaluate its suitability in the scientific studies involving larger datasets. The landmarks obtained are input into a Morphometric software package - *MorphoJ.* To eliminate the effects of scale, translational and rotation, Procrustes superimposition (Bookstein 1996) is applied. The results obtained by the automated system is further refined to eliminate $\approx 2\%$ of the outliers (≈ 17 images). The resulting shape variation after Procrustes superimposition is shown in figure: 1(b) along side the results obtained manually (slightly offset for easy comparison). The Procrustes ANOVA results of wing size and shape variation is presented in the (Tables: 1, 3 and 2, 4 respectively). The comparison of the results provide a direct evaluation of the scientific results obtained by the automated system. The results obtained by the automated and manual methods are consistent and indicates that all factors except 'side' are statistically significant for the size asymmetry (table: 1 and 3) and all factors are statistically significant for shape asymmetry (tables: 2 and 4).

Centroid Size - (Automated)					
Effect	*SS*	*MS*	*df*	*F*	*P (param.)*
Sex	14.120759	14.120759	1	554.150	<.0001
Individual	3.771311	0.025482	148	108.235	<.0001
Side	0.001701	0.001701	1	7.225	0.0080
Ind * Side	0.035079	0.000235	149	31.507	<.0001
Image Error	0.002219	0.000007	297		

Table 1: Asymmetry of Overall Wing Size. (Landmarks obtained by the automated system).

Shape, Procrustes ANOVA-Automated Results - (Auto).							
Effect	*SS*	*MS*	*df*	*Goodall F*	*P(param.)*	*Pillai tr*	*P(param.)*
Sex	0.03081822	0.00118532	26	34.086	<.0001	0.929	<.0001
Individual	0.13381055	0.00003477	3848	4.911	<.0001	15.209	<.0001
Side	0.00066362	0.00002552	26	3.605	<.0001	0.659	<.0001
Ind * Side	0.02743125	0.00000708	3874	10.665	<.0001	19.712	<.0001
Image Error	0.00512697	0.00000066	7722				

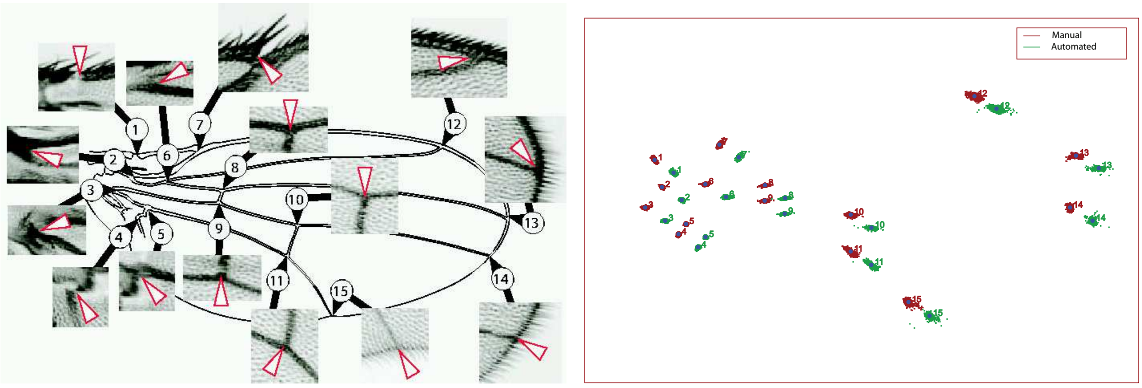
Table 2: Asymmetry of Shape. (Landmarks obtained by the automated system).

Centroid Size - (Manual).					
Effect	*SS*	*MS*	*df*	*F*	*P (param.)*
Sex	14.666568	14.666568	1	469.896	<.0001
Individual	4.775495	0.031212	153	198.792	<.0001
Side	0.000213	0.000213	1	1.353	0.2465
Ind * Side	0.024180	0.000157	154	61.56	<.0001
Image Error	0.000791	0.000003	310		

Table 3: Asymmetry of Overall Wing Size. (Landmarks obtained by manual digitization).

Shape, Procrustes ANOVA - (Manual).							
Effect	*SS*	*MS*	*df*	*Goodall F*	*P(param.)*	*Pillai tr*	*P(param.)*
Sex	0.02948371	0.00113399	26	31.569	<.0001	0.872	<.0001
Individual	0.14289389	0.00003592	3978	6.035	<.0001	16.191	<.0001
Side	0.00026496	0.00001019	26	1.712	0.0137	0.468	<.0001
Ind * Side	0.02383230	0.00000595	4004	20.060	<.0001	20.050	<.0001
Image Error	0.00239150	0.00000030	8060				

Table 4: Asymmetry of Shape. (Landmarks obtained by manual digitization).



(a) Schematic representation of the location of landmarks. (b) Procrustes Fit - Comparison of the width of distribution between manual (red) and automated (green) results.

Figure 1: Morphometric landmarks in *Drosophila*

4 Discussion

The coherent statistical accuracy in the estimation of the landmarks confirms the robustness of the algorithm. Although, there is a factor of two difference between the accuracy of the manual digitization and the automated system, the results indicate that the accuracy of the system can be considerably improved by using relevant matching reference images. The results achieved allows us to draw the conclusion that the automated method is a reliable system for extracting features such as morphometric landmarks on digital images. It is shown in this study that the method is sufficiently accurate to replace the manual digitization process (see figure: 1(b)). The automated method has potential advantages and makes large scale studies in the field of genetics and evolutionary morphometrics feasible. The intrinsic capability of the feature recognition process enables this method to be easily incorporated into other recognition tasks. The need for only a few reference images makes the system even more appealing. The algorithm will be made available as an open source package from *www.tina-vision.net and www.flywings.org.uk*.

References

Bookstein, F.L. (1991). *Morphometric tools for landmark data: Geometry and Biology.* Cambridge University Press, UK.

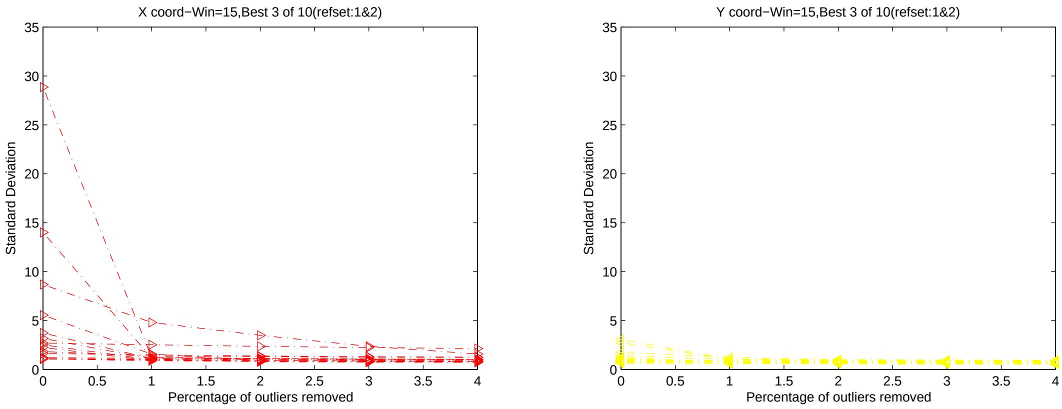
Dryden, I.L. and Mardia, K.V. (1998). *Statistical Shape Analysis.* Chichester, John Wiley and Sons.

Houle, D, Mezey, J, Galpern, P, and Carter, A. (2003). Automated measurement of Drosophila wings. *BMC*

Evolutionary Biology, **3(25)**.

Rockett, P.I, Thacker, N.A. and Aherne, F.J. (1997). The bhattacharyya metric as an absolute similarity measure for frequency coded data. *Kybernetika*, **34(4)**.

Palaniswamy, S. (2008). Automatic identification of landmarks in digital images. *Ph.D thesis (in progress), University of Manchester, UK.*



(a) X co-ord.

(b) Y co-ord.

Figure 2: Standard deviation of the automated results from the manual results. Best 3 of 10 matching reference images were used. The X and Y co-ordinates for the 15 landmarks are plotted as a function of outlier removal to show the stability of the results.