Published in IET Computer Vision
Received on 9th February 2009
Revised on 30th June 2009
doi: 10.1049/iet-cvi.2009.0014

ISSN 1751-9632

Automatic identification of landmarks in digital images

S. Palaniswamy[1,2] *N.A. Thacker*[2] *C.P. Klingenberg*[1]

[1]*Faculty of Life Sciences, The University of Manchester, Manchester, UK*
[2]*Imaging Science and Biomedical Engineering, The University of Manchester, Manchester, UK*
E-mail: Sasirekha.Palaniswamy@manchester.ac.uk

Abstract: The authors present an automated system for feature recognition in digital images. Morphometric landmarks are points that can be defined in all specimens and located precisely. They are widely used in shape analysis and a typical shape analysis study involves several hundred digital images. Presently, the extraction of landmarks is usually done manually and the process of identifying the landmarks is an important and labour-intensive part of any such analysis. This process is time-consuming, and quite often the research questions are dependent on the duration of obtaining these data. The authors show that a single training image with its landmark co-ordinates is enough to independently estimate the landmarks of any individual within a particular data set. The reliability and accuracy of the method can be further enhanced by using multiple training images. The precision, repeatability and robustness of the algorithm have been evaluated. It is shown in this study that the method is sufficiently accurate to replace the manual identification of landmarks. The generic nature and intrinsic capability of the feature recognition process enables this method to be easily incorporated into other recognition tasks.

1 Background

Morphological landmarks are points that can be located precisely and establish an unambiguous one-to-one correspondence among all the specimens and are widely used in shape analysis [1]. Points like the tip of the nose or the outer corner of the left eye are possible landmark points of the human face. Analyses of shape investigate the arrangement of landmark points relative to each other. A substantial body of statistical methods is available for the analysis of configurations of landmark points [2].

This framework of shape analysis by landmarks is increasingly used in many biological and medical applications and is widely applied in many other fields. The anatomical landmarks can be easily collected in two dimensions from digital images and this approach is quite useful as the landmarks can be collected from non-model organisms or even fossils. The X and Y co-ordinates of these landmarks are usually obtained by manually digitising the location of these co-ordinates appropriately based on their anatomical context. The configuration of landmarks have helped identify the possible source of re-infesting specimens and encounter the epidemiologically challenging vectors of Chagas disease [3]. The potential of using geometrical morphometric techniques as an invaluable tool for recognising taxonomic data is being explored [4]. Other scientific applications include investigating the study of size and shape to examine the effects of experimental treatments, genotype or other factors directly in the anatomical aspect. The use of landmarks has been adapted to specific biological contexts such as genetics [5–7], geographic differentiation [8] and the study of morphological integration [9, 10].

Developing an automated system for locating landmarks in digital images of Drosophila wings is largely significant, as it is an excellent model for the study of genetics of development and evolution of morphological form [11, 12]. They contain a wealth of interesting biological information and its simple, flat, two-dimensional structures enables convenient handling. Specialised algorithms and plug-ins have been used to semi-automate the process and to enhance the speed of the manual digitisation. There have been previous

attempts to automate the process of landmark location on the Drosophila wing. The method presented in [12] is a semi-automated process where the operator initiates the process by marking two landmark positions and the system fits a series of spline curves to the margin and the veins of the wing, and defines the landmarks as the intersections of the splines. The drawback of such a system is that the landmarks are not always at the exact location of the intersection of veins and the splines may not exactly match the veins (because of the 'stiffness' of the spline interpolation) and the system has problems identifying wings of species with highly melanised spots at the intersections.

Another such system to locate the landmarks on digital images of bee wings is being developed at the Paris National Museum of Natural History [13]. It applies the techniques of mathematical morphology and skeletonisation to obtain the landmarks. However, using these techniques are not guaranteed to be robust. The method also requires human intervention in terms of loading the data and identifying the initial set of points to start the process and the pre-processing step includes certain parameters that have to be set by the operator.

2 Introduction

In this paper, we focus on the automatic extraction of morphometric landmarks in digital images of Drosophila (see Fig. 1). Morphometric landmarks are points that can be defined in all specimens and located precisely. They are widely used in shape analysis and a typical shape analysis study involves several hundred digital images. Presently, the extraction of landmarks is usually done manually and the process of identifying the landmarks is an important and labour-intensive part of any such analysis. This process is time-consuming, and quite often the research questions are dependent on the duration of obtaining these data. To address these problems, many researchers have focused on using specialised algorithms and semi-automated methods to enhance the speed and reliability of the digitisation process [12]. Although these methods increase the efficiency of human effort, the requirement for an observer has not been eliminated as the methods require the initial set of landmark co-ordinates to be identified manually. 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 as an important goal.

Automating the process of landmark extraction has numerous advantages over manual digitisation, as it will not only diminish the labour needed for shape analysis, but it also will eliminate the source of error (mistakes made during digitising and subtle differences between observers). Furthermore, automatic extraction of features from images can potentially change the way in which landmarks are chosen for morphometric studies. Whereas the traditional approach requires landmarks to be chosen a priori, based on outside knowledge of the study system, the approach using automated image analysis raises the possibility to identify and extract features from the total information contained in the images that are maximally informative in the context of a particular research project.

3 Methods

The data set consists of 856 grey-scale images of the Drosophila wings (as shown in Fig. 1*a*). Data acquisition is carried out by mounting the fly wings in rows on a microscopic slide and flattening with a cover slip. The digital images are obtained using an appropriate digital camera mounted on the microscope and attached to a computer. A calibration image is generally obtained along with each of the data set to standardise the difference in magnifications between different data sets. Two independent sets of images with five images in each set and their landmark co-ordinates are used as training (reference) images.

The analysis system is constructed from four key stages: a feature-based detection of the fly wing structure,

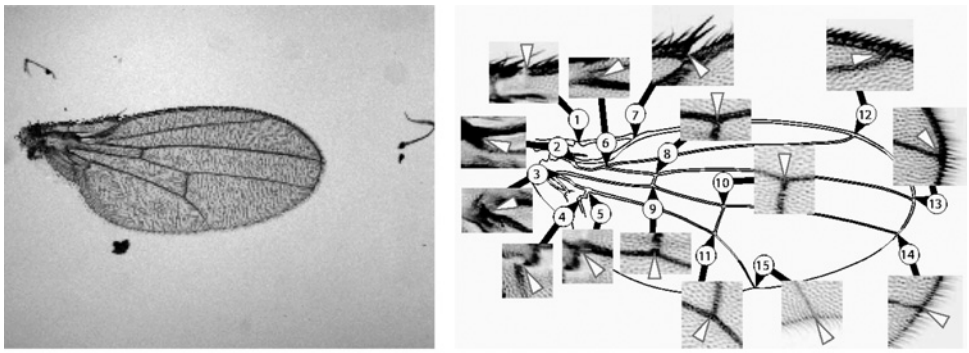


Figure 1 *Drosophila wing and morphometric landmarks*

a Image of a Drosophila wing
b Schematic representation showing the 15 morphometric landmarks

IET Comput. Vis., 2010, Vol. 4, Iss. 4, pp. 247–260
doi: 10.1049/iet-cvi.2009.0014

recording the compact invariant shape descriptors using the pairwise geometric histogram (PGH) representation, global estimation of the pose using the probabilistic Hough transform (PHT) and finally a correlation-based refinement of individual features.

The proposed method extracts the ridges (linear features such as wing veins) using the noise characteristics of the image and by approximating the grey-level profile using a difference of Gaussian (DoG) filter. The ridges obtained are approximated by line segments and the geometric relationships between them are encoded in PGHs, an approximation to the probabilistic density function for the geometric co-occurrences in the data. Shape correspondence is determined by comparing and matching the pairwise histograms of the scene and model data. A PHT, (robust likelihood) is used to determine the hypothesised landmark location. Sub-pixel estimation of the landmark location is performed by template matching, that is, correlating a small region around the Hough estimated landmark location.

3.1 Image preprocessing and feature extraction

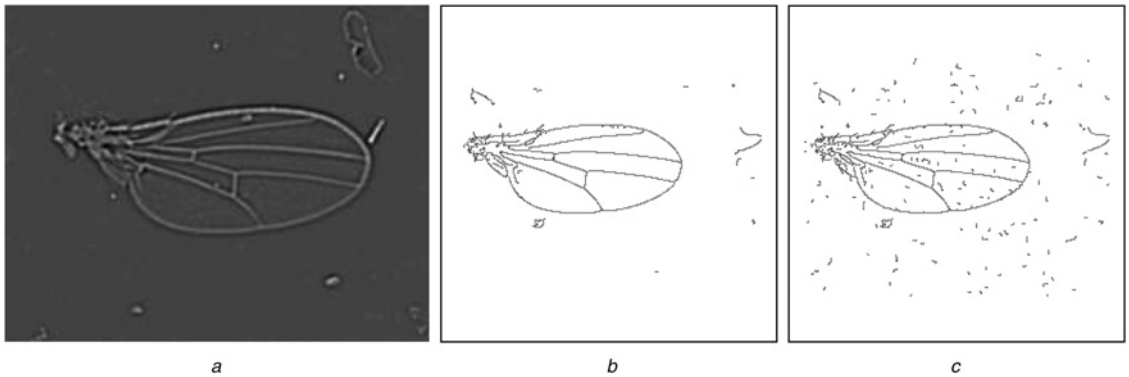
The feature extraction stage involves extracting essential information from the digital images and retaining only those features that we are interested in. Before extracting features in the image, it is important to preprocess the image with an appropriate enhancement technique. In this study, a DoG filter is employed for a variety of reasons. It is important to choose a filter with a shape profile similar to the features that we are looking for so that it can provide a maximal response when centred over the features. This is consistent with the concept of a matched filter. In this case, the shape of the DoG filter along with its radial weighting allows us to reduce the significance of the image data away from the centre and define a zero response for uniform background regions. Also, the width of the DoG filter can be optimised to detect features at a particular scale. It is essential for the filter to accommodate changes in the scale of the features up to a certain factor and this can be achieved by taking the difference of two Gaussian filters with convolution widths that differ by that factor. The radially symmetrical shape of the DoG filter allows the enhancement of the features at all orientations equally. In addition, applying a linear operator such as a DoG filter does not introduce any arbitrary spatial dependency on the noise process in the output image. Moreover, the DoG filter applies noise filtering, eliminating unwanted high-frequency image structure, such as bristles (see Fig. 2).

The feature enhancement stage and the thresholding of a response from the features in the feature detection process can be related to a conventional hypothesis test, as used for significance testing [14]. This framework potentially allows the construction of a feature detector for any form of local image structure and therefore provides very little constraint on the definition of what we would like to detect. Using the concept of error propagation, the uniformity of the noise in the image and its effect in the output image can be estimated. The first-order analysis (error propagation) provides an understanding of the effects of noise, and provides a necessary but not sufficient condition for algorithm stability. For any function, $f(y) = x$, a small change δ_x in x can cause a change δ_y in the function $f(y)$. This change in the function, δ_y, can be estimated by taking the partial derivatives of the function, $f(y)$, with respect to x and considering the small change δ_x as $\delta_y = (\mathrm{d}y/\mathrm{d}x)\delta_x$. Thereby, any change in the function, that depends on a variable x, can be calculated from the amount of change on the variable x itself.

For the specific case of convolution of an image, the output image can be represented as below

$$f_{(x,y)} = \sum_m \sum_n G_{mn} I_{(x+m,y+n)} \tag{1}$$

Any change σ in the input image I will cause a change δf in the output image f. Therefore the change in the output



a b c

Figure 2 *Result of the preprocessing stage and the feature extraction on Drosophila wing image*

a Preprocessed image (DoG filter)
b Feature extraction on original image
c Feature extraction on noisy image (≃10 times added noise)

image can be calculated from estimates of the variance in the input image.

For uniform independent random noise σ, this can be represented as

$$\delta f^2_{(xy)} = \sum_m \sum_n G^2_{mn}\sigma^2 \tag{2}$$

In the above equation, the effects of the parameters of the convolution kernel can be summarised by a constant, K

$$\sum_m \sum_n G^2_{mn} = K \tag{3}$$

Therefore the changes in the output image can be re-written as below

$$\delta f^2_{(xy)} = K\sigma^2 \tag{4}$$

It is evident that any change in the output image is directly proportional to the noise in the input image and spatially consistent. Such behaviour will result in a 'robust' response to image data, in that responses detected above fixed thresholds will behave equivalently everywhere in the image. The original Canny algorithm is ideal for detection of step edges [15], but has to be modified to meet the needs of detecting ridge features in this study [16]. The enhanced features and a corresponding orientation estimates are provided to a similar framework, which extracts extended structures of linked features using hysteresis thresholding (lower threshold approximated to 1.5 times σ noise and upper threshold set to 3.0 times σ noise) and non-maximal suppression. The shape of wing profiles combined with the need for some kind of radial weighting in the detection process can be interpreted as a likelihood estimate for the scale of local image structure (Appendix).

3.2 Geometric features

The shape of arbitrary image regions can be represented using feature vectors that are obtained by grouping features; deriving the geometrical relationship between these vectors can best represent the shape information in comparison to the individual features on their own. The geometric relationship between a set of feature vectors is unique with respect to their shape representation and provides a more reliable shape representation scheme compared to the primitive features such as edges. The geometric features characterised by the geometric properties (angle, ratio of the length of lines) between a set of feature vectors (string of edgels or lines) have the characteristics that are invariant to in-plane rotation, translation and line extension.

Geometric features such as the relative angle between the directional vectors of two line segments is a useful feature for shape representation [17]. The perpendicular distances from the endpoints of a line segment to the extension of another line segment, denoted by $d_{\min}$ and $d_{\max}$, is another geometrical feature that is invariant to rotation and translation. However, this feature is based on the absolute distances and therefore scale sensitive. This problem could be overcome by normalising the perpendicular distances with the length of the line segments. The combination of the relative angle and the perpendicular distances provides an useful set of geometric features that can be used in high-level shape representation (Fig. 3). The geometric features are robust to fragmentation (as shown in Fig. 3*b* and 4*b*) and this characteristic is crucial to achieve robust shape representation.

3.3 Pairwise geometric histograms

In order to support efficient shape matching, it is important that the edge features are represented in a compact and invariant manner. Geometric relationships between the underlying edge features, such as angle and perpendicular distance, provide an efficient and robust shape description. Encoding the set of geometric features by means of a histogram can be understood as recording the statistical variability in the shape in the form of probability density distributions.

The PGH is an approximation to the probabilistic density function

$$H_i(\theta, d) = P(\theta_i - \theta_j, d_{ij}|e_i) \tag{5}$$

for the geometric co-occurrences of an edgel e_j given e_i as a function of relative angle $\theta_i - \theta_j$ and perpendicular separation d_{ij}. This is a well-established method of shape representation [18, 19] based on robust recording of the distribution of pairwise geometric relationships between local shape features that can support recognition of identical geometric features in the test images using standard pattern recognition techniques as described in Section 3.5 [20–22]. There is considerable robustness to the loss of data due to fragmentation noise and occlusion [17, 18]. The method is also known to be complete, in that the original structure of the object can be reconstructed from the set of histograms describing a shape. This representation is invariant for portions of the

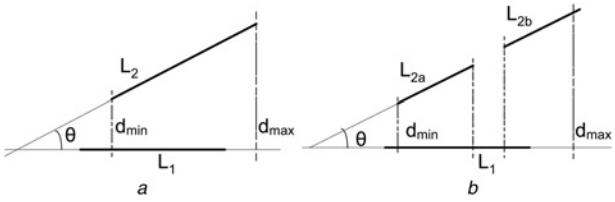


Figure 3 *Geometric features (angle and perpendicular distance)*

a Geometric features
b Geometric features (fragmented)
The features are robust to fragmentation indicating efficient shape representation

IET Comput. Vis., 2010, Vol. 4, Iss. 4, pp. 247–260
doi: 10.1049/iet-cvi.2009.0014

same linear feature so that it can be constructed by considering a linearisation of the edge map, thereby making the method suitable for the detection of extended lines via clustering.

The importance of a pair of line segments defining the representative shape can be encoded by entering the product of their lengths at the value of the entry. Appropriate weighting of the entries in the PGH ensures sudden changes are not introduced in the representation by gradual changes in input data; therefore matching performance degrades smoothly. Blurring the entries in the histograms is used as a way of encoding measurement uncertainty regarding the true position and orientation of the line segments. This assists in robustness of the algorithm by accommodating the subtle differences in location of the features due to the variability in the line segmentation process. The scale of binning and blurring specifies the allowable extent of differences when comparing similar shapes [17]. Since any arbitrary shape can be approximated by a set of straight lines, the PGH is applicable to even linearised smooth curves [17]. The invariance properties of PGH and its insensitive characteristics to noise makes it an efficient shape representation technique.

3.4 Histogram construction

During training normalised PGHs are generated for model features and during analysis PGHs are recorded for scene features. This can be related to recording the statistical variability in the shape in the form of probability density distributions. Measurement errors introduced by the approximation of edge strings by straight line segments has to be accounted for during the construction of the geometric histograms. This can be achieved by convolving the histograms with an approximate error function and it would provide a closer representation of the shape data it encompasses [17, 23].

The linearisation of edge pixels results in characterisation of a string of edge pixels into linear segments and the geometric relationships between each pair of these line segments can be encoded in the PGH. PGH is characterised as a frequency distribution scheme that records the geometric relationship (angle and perpendicular distances) between a pair of features (such as two line segments). This technique is a reliable and efficient descriptor for classifying shape/image features. Each of the line segments in the model image is taken in turn as the reference line and the perpendicular distance to all other features and the angle of orientation to all other features with respect to the reference line is recorded. The frequency of the geometric features is recorded as a two-dimensional histogram with an angle axis $(0 - 2\pi)$ and distance axis (range of perpendicular distances) to record the distribution of the geometric features within a shape. This method is considered to be an efficient shape representation scheme and it is translationally and rotationally invariant as the angle and the perpendicular distance between a pair of lines would be the same irrespective of their orientation and position. The approximation of the edge data is allowed up to a certain specified accuracy by means of blurring. The blurring of entry along each axis allows encoding the uncertainty regarding the true position and orientation of each line segment. The scale of binning and extent of blurring defines the extent of allowable differences when matching similar shapes (see Fig. 4).

The values in the histogram indicates the presence of features in the image at a particular geometric relationship. The importance of the features can be emphasised by taking the total entry made in the histogram to be equal to the product of the lengths of the reference line and the scene line in consideration. Also, since the length of the reference line is fixed, this can be considered to be a direct analogue to Poisson sampling. This enables the use of likelihood-based similarity measures to compare the histogram entries on the basis of a square root dot product (see Section 3.5) [17]. The success of the PGH lies in the matching strategy adopted to compare the measure of

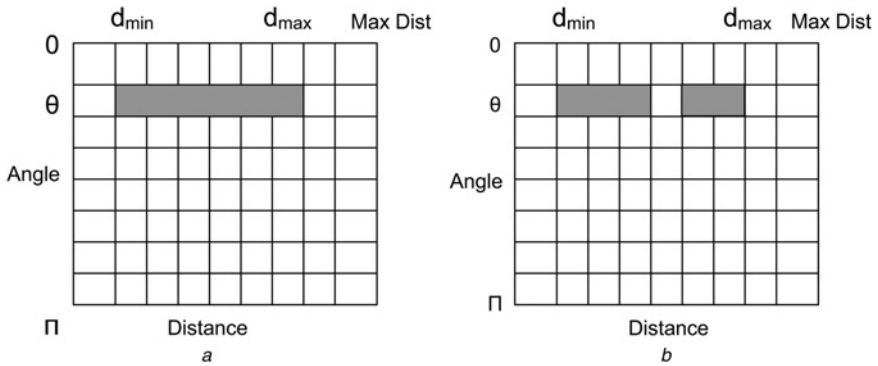


Figure 4 *Pairwise geometric histogram – robustness to fragmentation*

a PGH entry
b PGH entry (fragmented features)
The fragmented entries are identical to the non-fragmented features

similarity between the scene and model features. The measure of histogram similarity between the scene and the model features can be achieved by comparing their probability distributions.

3.5 Histogram matching

The histogram matching enables robust classification of shape features by finding similarity between the scene and the known model features. During training, PGH are constructed for model features and during analysis, PGH are constructed for scene features. Shape representation comprises many geometric histograms, each representing a single model feature and shape recognition is done by identifying the correspondences between the model and scene features (as shown in Fig. 5). The degree to which a scene feature matches a particular model feature can be determined by comparing their probability distribution (histograms similarity). This can be interpreted as the statistical likelihood measure.

The Bhattacharya similarity metric [17] can be used to compare the distributions (PGH) for the model and the scene data. By this, it is possible to establish the shape correspondence between them in a way which directly identifies binary relationships, thereby avoiding the combinatorial problems generally encountered with interpretation trees. The similarity metric provides a quantitative estimation of the likeness between the model feature (reference image) and the scene feature (test image). The Bhattacharrya measure B_{ij} computes the degree of match between them and is expressed in the form of a dot product correlation of the histograms of lines H_i and H_j.

$$B_{ij} = \sum_{\theta}^{2\pi} \sum_{d}^{d_{\max}} \sqrt{H_i(\theta, d) H_j(\theta, d)} \tag{6}$$

The Bhattacharya metric for the histogram matching is statistically valid and can be related, via the χ^2 variable, to a maximum likelihood similarity metric and can be derived as an approximation to Fisher's exact test as a method for comparing two distributions. The similarity metric must account for the errors (slight variations in the PGHs of identical features due to noise and blurring in the entries to account for subtle variations) and degrade slowly. The hypothesised matches can then be used as input to pose an estimation algorithm such as the generalised Hough transforms. Scale independent recognition can be achieved by representing an object at a range of scales [24]. However, this property of the PGH representation was not applied in this study.

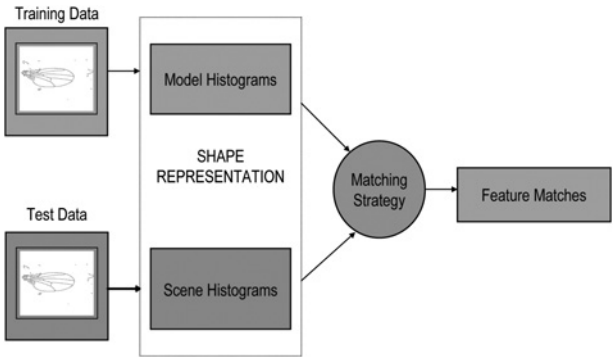


Figure 5 *Schematic representation of histogram matching strategy*

3.6 Probabilistic Hough transform

The PHT can be used to determine the presence and location of relevant shapes of the model in the scene image by finding spatially consistent groups of features. A reference frame for each of the stored model shapes is defined with reference to an arbitrary position and all the features are measured in relation to this frame. The relative position of the classified scene lines can be estimated using the relative position of the model feature in the reference frame. However, this method is subjective to variations such as line fragmentation. This could be overcome by improved constraints imposed by a pair of scene lines (tuple Hough transform). The hypothesised location of the model is established based on the conditional probability that any pair of scene lines will be measured at a given position.

A PHT (robust maximum likelihood) is used to make an estimate of the global position and orientation of each wing in the test data. The hypothesised pose estimation is modelled by a standard error propagation model to account for the variability of the line segmentation process and the likely occurrences of measurement errors. Entries in the 2D location histogram are made according to the localisation covariance, propagated from the errors on the constraint lines. This takes proper account of errors, resulting in improved robustness and more accurate determination of model position, orientation and scale in comparison to the more conventional form of this algorithm. The entries in the Hough arrays are constructed from pair of lines (n, m). The equivalent probabilistic form for the Hough transform $H(x, y)$ used to find the position of a model in a scene is given by the expression

$$\begin{aligned} H(x, y) &= \sum_{n}^{N} \sum_{m}^{N} \log(p(x, y|n) p(x, y|m)) \\ &= \sum_{n}^{N} \log(p(x, y|n)) \sum_{m}^{N} \log(p(x, y|m)) \end{aligned} \tag{7}$$

so that the Hough entry can be considered as the square of the robust log likelihood $L(x, y)$ for the localisation of the object

$$H(x, y) = L(x, y)^2 = \left(\sum_{n}^{N} \log(p(x, y|n)) \right)^2 \tag{8}$$

doi: 10.1049/iet-cvi.2009.0014

Each input feature in the PHT is a pair of scene line labels that hypothesises the position, orientation and scale of a model in the scene. The entry is made only when the pair of lines are in reasonable agreement about the position, orientation and scale of the model. However, this hypothesised location is prone to error due to its dependency on the reliability of the extracted line segments. The Gaussian error model is applied to the position at the end of the line segment and by applying standard error propagation, the appropriate entry into the PHT (Hough space) can be determined. Furthermore, the constraints imposed by pairs of scene lines and the associated Gaussian error model results in the pose estimation that is equivalent to computing a robust least-squares fit between the model and the scene data.

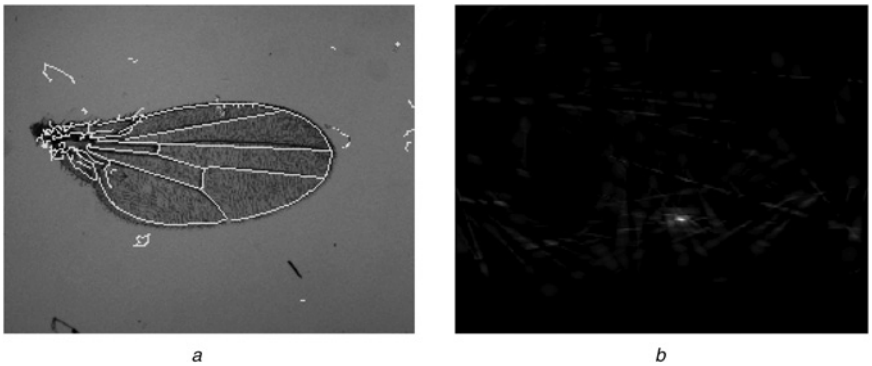
During array construction $H_{nm} = \log(p(x, y|n)p(x, y|m))$ is estimated from a 2D Gaussian distribution centred at the position of the model hypothesised by the m, nth pair of scene line labels with variance propagated from the individual line location errors. This tuple-based construction is more reliable in cases of line fragmentation and it helps remove background noise from the Hough array and has some computational advantages. The variability of the line segmentation process and the uniform error on the scale estimates are independent and are adjusted to give a quantitative estimate of the hypothesised location of a pre-defined reference point from pair of scene lines. Training from example data involves recording the perpendicular distance, 'd', from each model line to the reference point. During pose estimation, extending each pair of scene lines at the appropriate perpendicular distance will intersect at the hypothesised reference point (see Fig. 6). Error at the point of intersection can again be estimated by standard error propagation. The PHT is expected to perform with improved robustness and provides more accurate determination of the model position, orientation and scale in comparison with the generalised Hough transform as it takes into consideration of the errors.

3.7 Determination of orientation and scale

The PHT is used to locate models using the positions, orientations and scales hypothesised by the scene line labels. However, the orientations and scales of the models are not determined explicitly. This can be determined separately using 1-parameter Hough transforms using data consistent with hypothesised locations. For each model position determined, a 1-parameter orientation Hough transform and 1-parameter scale Hough transform can be constructed from entries selected on the basis of consistency between the scene lines and model position. The orientation is determined by taking into account of the difference in orientation between the scene line and model line to which it matched [20, 24]. Comparing the perpendicular distance from the scene line to the model position to this same distance in the model itself would yield the scale. Peaks in these Hough transforms would give the orientation and scale of the model at that position in the scene.

3.8 Template matching

The Hough scheme computes an estimate of landmark position based on the global shape of the features. To obtain sub-pixel accurate landmark location for use in the morphometric studies to determine shape variations, the estimated landmark position has to be refined. A template-based correlation matching is applied to achieve this. Local regions surrounding the landmark position of the template image and the Hough predicted landmark location on the scene image (rotated to same orientation as the template image) are correlated across a grid of spatial locations to obtain the landmark position based on the local image evidence. The parameter for the correlation window size is represented as δ. The regions are preprocessed with a DoG filter before the correlation matching is performed to ensure that high-frequency signals that are not part of the image structures are removed. Template matching is performed on the DoG image of the scene $D(I)$ and model (example



a b

Figure 6 *Probabilistic Hough transform located landmark*

a Flywing image with model overlaid
b Peak in the Hough transform
The bright spot in *b* indicates the corresponding Hough transform peak for landmark 15 on the flywing

mark-up) $D(M)$ data, over a small region δ around the Hough estimate for the feature in the scene data (see Fig. 7). The use of the DoG images also eliminates any image illumination offset and the matching is performed as a dot-product correlation in order to eliminate the effects of illumination scaling

$$L_{h_x\, h_y} = \frac{\sum_x^R \sum_y^R D(M(x, y))D(I(x + h_x, y + h_y))}{\sqrt{\sum_x^R \sum_y^R D(I(x + h_x, y + h_y))^2}} \quad (9)$$

where R is the region size.

This is directly equivalent to performing a least-squares comparison of the image regions with one free grey-level scale parameter. In this study, the denominator is presumed to be constant to save time on the computationally expensive calculations. To save processing time during alignment, the scene data are rotated to match the model data using the Hough estimate, which is assumed to be sufficiently accurate for final location of the landmark. The best possible match is identified and that location is transformed back onto the scene image. In addition, the least-squares values that provide the best matching score for the given template and scene region are output. These values enable to check the best matching templates (which can be selected for the final estimation of landmark position) to compare the consistency in their accuracy and to combine those landmark positions for improved localisation. This step not only supports quality control during the data analysis, but also allows one to check on the adequacy of the number of reference images and to eliminate any residual problems in alignment estimation, such as poor rotation estimates. The use of local template to detect salient feature in a small region characterises the similarity between the features that we are looking for.

The template matching stage enhances the precision and sub-pixel localisation of the landmark position obtained by the Hough transform pose estimation algorithm. However, it has to be noted that the sub-pixel localisation depends on several factors such as the orientation of the feature, the scale of the feature and the degree of blurring. The time taken for the template matching is directly proportional with the correlation window size and the search region. However, as it is performed over a confined space around the Hough predicted landmark location, it drastically reduces the computational time, in comparison to using it as a generalised first-stage approach.

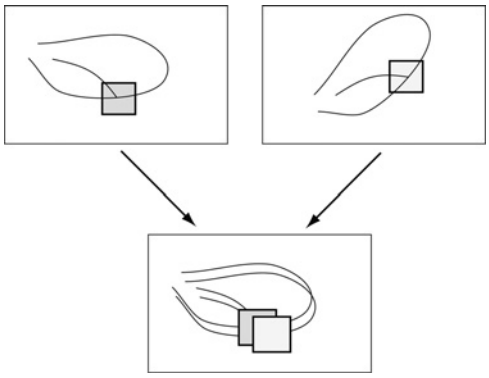


Figure 7 *Template matching illustration – small region around the known landmark on the training image and the Hough predicted landmark on the test image is correlated to refine the location of the landmark*

4 Evaluation of the accuracy of the automated system

A data set with 856 images (1280 × 1022 pixels) of the wings of Drosophila melanogaster is used to evaluate the accuracy, reliability and robustness of the automated system. The automated landmark identification system is implemented on this data set and the 15 landmarks (both x and y co-ordinates) are extracted for each of the images. As discussed in [16], multiple reference images are used to extract the landmark co-ordinates. To sensibly combine the estimated landmark locations by multiple training images into a unique hypothesised landmark location, we use the least-square matching scores to generate ranking of the best matching training images during the landmark extraction process.

To test the repeatability and robustness of the automated system, the landmarks are automatically extracted on the above data set but with two independent sets of five reference images in each. The coherent statistical accuracy in the estimation of the landmarks by two independent sets of reference images and an improved performance in their combination confirms the robustness of the algorithm. Furthermore, the landmarks obtained from both sets are combined to check the hypothesis that increasing the number of reference images would improve the performance of the automated system. This is expected to provide us with highly accurate estimates of the landmark position, due to the fact that in this scenario the best three reference images are obtained from both sets of reference images. The landmarks obtained by the best three ranking reference images are compared with one another to ensure that the landmarks obtained by each of the best three ranking reference images are consistent.

The above step ensures the exclusion of any outliers at this stage. This is achieved by calculating the distance between the landmarks obtained by the best three matching reference images. If the variation in their estimated landmark positions is below a certain threshold, that is, 10 pixels (which can be determined from histograms of repeated localisation), then the mean of the three values are computed to estimate the final landmark position. Alternatively, the mean of the landmark values obtained by the best two ranking reference images is used to estimate the final landmark position. This stage ensures that outliers

doi: 10.1049/iet-cvi.2009.0014

if any exist do not deteriorate the precision of the landmarks obtained by the automated system. It helps to improve the reliability of the results obtained and ensures an improvement in the accuracy of the final landmark estimation. The accuracy of the landmark location estimated by the automated system can be evaluated by comparing them to those of manually digitised landmark co-ordinates that are available (previously obtained for other morphometric studies) for this data set. The method is also shown to be stable to noise, well beyond the level normally present in this data set [25].

5 Results

The accuracy of the automated system can be determined by comparing the differences (in pixels) between the landmarks located by the automated system and the manual digitisation (see Fig. 8). It can be seen from Table 1, that the landmarks obtained by the automated system are located within few pixels accuracy in comparison to those manually digitised landmark locations. The results indicate that 90.64% of the landmarks in the data set (consisting of 896 images of size 1280 × 1022 pixels) are located within two pixels accuracy and 98.36% of the landmarks are located within an accuracy of five pixels from the manually digitised landmarks.

The analysis is carried out independently for each of the landmark co-ordinates as the feature extraction process on each of the landmark is independent of one another. The standard deviation for each of the landmark co-ordinates have been estimated and the results indicate that the landmarks can be located with high reliability. Figs. 9*a*–*f* show 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.

The effect of having a larger window size (δ) during the correlation matching has been examined by comparing the results obtained using the window size of $\delta = 15$ pixels and $\delta = 30$ pixels. Figs. 10*a* and *b* indicate that there has not been significant variation in their results. However, having a larger correlation window significantly impacts the speed of the algorithm, which is important when processing large numbers of images.

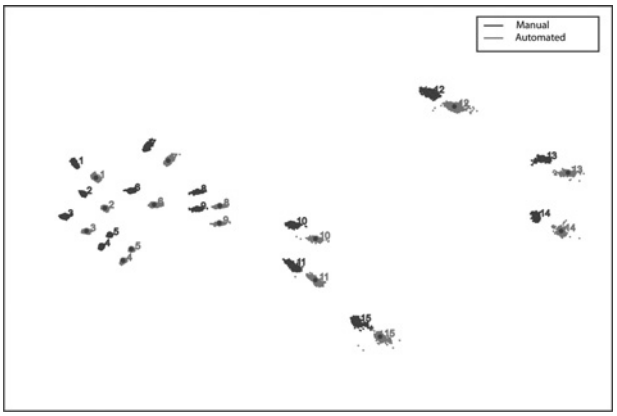


Figure 8 *Procrustes fit – comparison of the width of distribution between manual (dark grey) and automated (light grey) results*

Table 1 Accuracy of landmarks estimated by the automated system

Pixel range	No. of data points	Cumulative accuracy, %
0.5	18 815	69.99627976
1	2316	78.61235119
2	3232	90.63616071
3	1382	95.77752976
4	491	97.60416667
5	204	98.36309524
10	275	99.38616071
15	45	99.55357143
20	16	99.61309524
25	32	99.73214286
50	67	99.98139881
100	3	99.99255952
more	2	100
	26 880	

The results indicate that about 70% of the landmarks are located within sub-pixel accuracy in comparison to the landmarks located by manual digitisation. About 98.4% of the landmarks are located within five pixels from the landmarks estimated manually.

Figs. 11*a* and *b* show the comparison in performance obtained by increasing the number of reference images. This improves the accuracy of certain features (landmarks 6 and 12), (see Figs. 11*a* and *b*). This could be attributed to the fact that some of the features for instance landmark 12 (see Fig. 1*b*) has high biological variability. It was also noted that the presence of thick bristles in the surrounding region affects the accuracy of the landmark location. By providing additional training images the chance for a more appropriate matching reference image increases, thereby enabling an improvement in its accuracy. However, there is no evidence of significant improvement on most other landmarks, as on the other hand, these landmarks are simplistic features and an increase in the number of reference images does not provide additional information which contributes to a substantial improvement in its accuracy.

Finally, an overall comparison is performed between the landmark locations obtained using the automated method

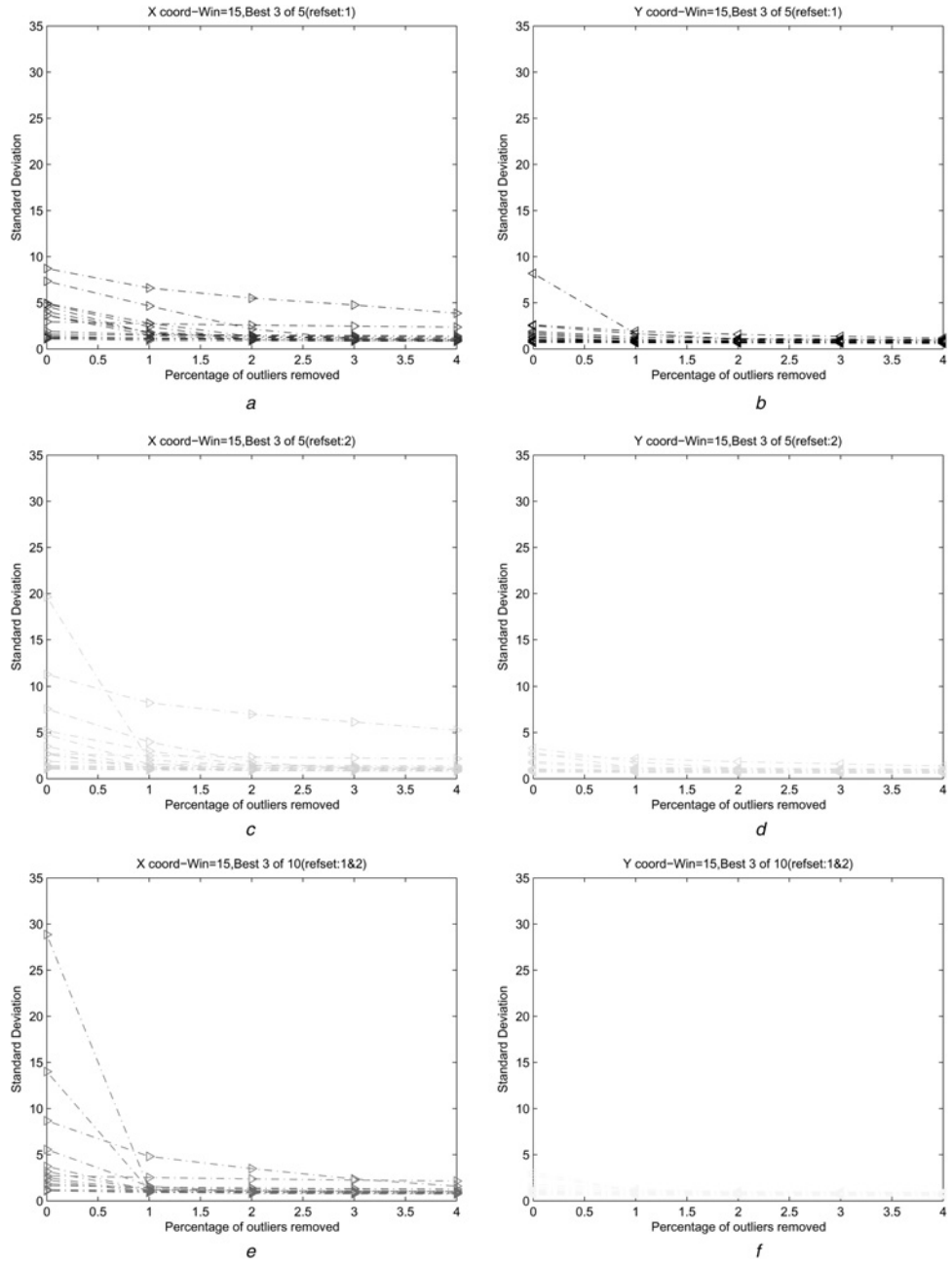


Figure 9 *Standard deviation of the automated results from the manual results*

a The plot shows the SD of *X* co-ordinates obtained using the best three out of five reference images from Reference set 1.
b The plot shows the SD of *Y* co-ordinates obtained using the best three out of five reference images from Reference set 1.
c The plot shows the SD of *X* co-ordinates obtained using the best three out of five reference images from Reference set 2.
d The plot shows the SD of *Y* co-ordinates obtained using the best three out of five reference images from Reference set 2.
e The plot shows the SD of *X* co-ordinates obtained using the best three out of ten reference images from both Reference set 1 and 2.
f The plot shows the SD of *Y* co-ordinates obtained using the best three out of ten reference images from both Reference set 1 and 2.
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.

and manual digitisation in the context of a morphometric analysis using Procrustes alignment (Procrustes analysis is carried out by superimposing configurations of landmarks in two or more specimens to achieve an overall best fit. The aligned co-ordinates constitute a new set of variables encoding the shape information [1, 2].). Figs. 12*a* and *b* show the comparison of the width of the distribution of the Procrustes co-ordinates for the manual and the automated results. The Procrustes fit for the manual and automated results is shown in Fig. 8 (they are offset slightly for easy comparison). It is clear that after 2% outlier removal, that is, about 17 images, the width of the distribution is comparable to an expert manual performance. Although manual digitisation does not produce these outliers, we consider such losses to be negligible in the context of such experiments.

doi: 10.1049/iet-cvi.2009.0014

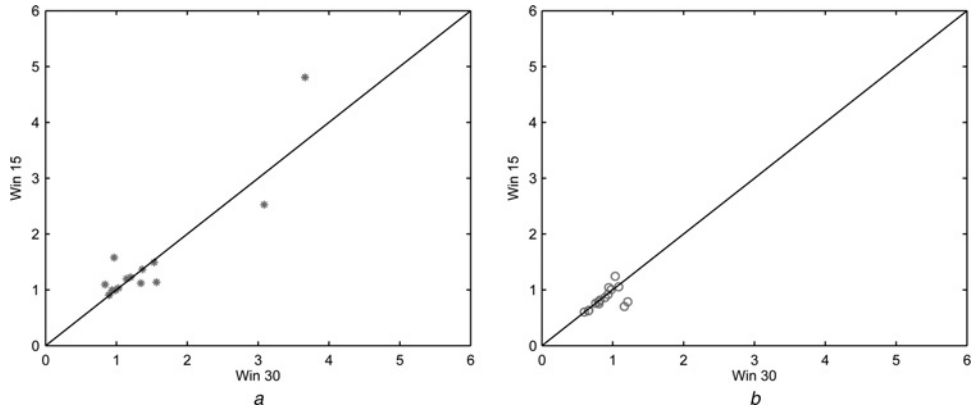


Figure 10 *Comparison of the effect of template matching window size parameters ($\delta = 15$ & $\delta = 30$ pixels) on X and Y co-ordinates standard deviation. Average of best three sets of landmarks out of ten is used after removing outliers (1%)*

a Standard deviation of *X* co-ordinates
b Standard deviation of *Y* co-ordinates

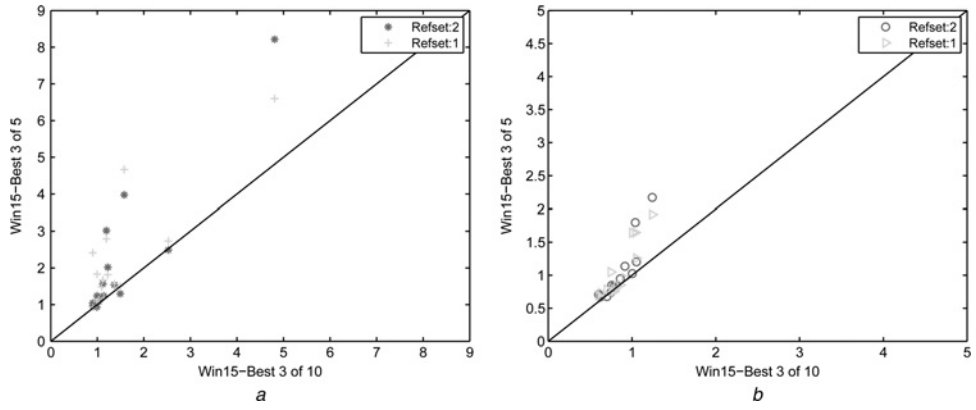


Figure 11 *Results obtained by the best three matching templates from a set of five and ten reference images are compared after removing outliers (1%)*

a Standard deviation of *X* co-ordinates
b Standard deviation of *Y* co-ordinates

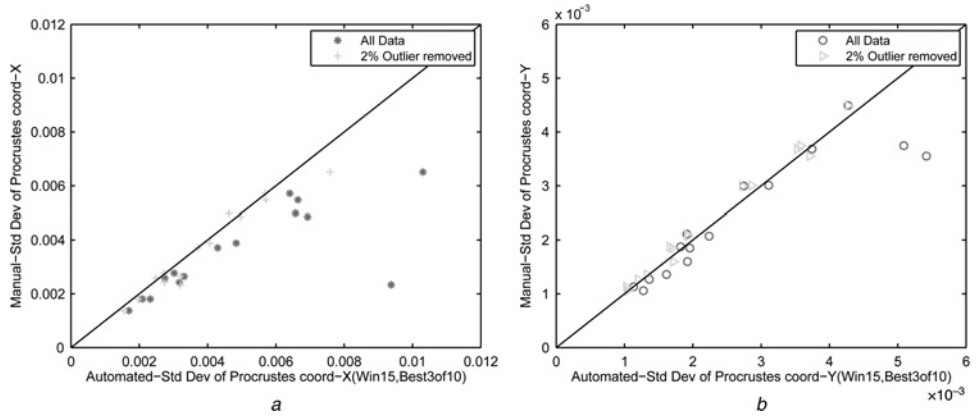


Figure 12 *Comparison of the standard deviation of the Procrustes (X and Y) co-ordinates obtained by manual and automated methods*

a Standard deviation of Procrustes *X* co-ordinates
b Standard deviation of Procrustes *Y* co-ordinates

6 Discussion

The automated method presented in this study is based on a stable feature extraction process, which is inherently robust and has a high degree of noise stability. This study provides confirmation that the PGH representation is a reliable shape descriptor. Such behaviour is a fundamental requirement for any object recognition algorithm. In combination with the Hough transform, the combined system has the ability to capture pose variation and offers reliability even in cases of line fragmentation and occlusion. These properties are present in this approach by design; specifically, attention to the treatment of shape data as measurement, combined with the requirements for image invariants and the use of appropriate statistical matching processes. In particular, the desired invariance to in-plane rotation, translation, illumination and aperture selection is intended to provide a compact representation, which generalises well to unseen data. The ability to deal with scale change has also been identified [20, 24], in a manner consistent with the Scale Invariant Feature Transform (SIFT) approach. The PGH representation and the arguments regarding its completeness [17] and techniques for statistical matching [17] predate this more popular work by a decade.

The simple procedure of regional template matching is shown to be extremely robust in sub-pixel localisation of the features. The performance of the template matching depends on how accurately the template of the model matches the scene features. The templates for each of the landmark regions could have been created synthetically using the accurate description of the scene information pertaining to that particular landmark. However, it is essential to take into account of the biological variability and the presence of bristles across the outline of the wing. It also involves the complexity in constructing the possible template models across the different landmarks. Therefore a deliberate choice of simplified process was to correlate the regions surrounding the respective landmarks in the model and the scene image. This process allows comparing regions of real images and avoids the issues surrounding real feature models.

The current system is capable of providing a set of 30 landmark co-ordinates (15 landmarks) on an image of size, 1280×1022 pixels in about 3 min on a SUN Sparc Ultra 5 workstation. The time taken for an expert to digitise using specialised algorithms is $\sim$40 s, although maintaining this speed across a large data set would be unrealistic.

7 Conclusions

In this study, we have presented an automated system for feature recognition in digital images. The performance of the system and its robustness have been evaluated by comparing it to the manual results obtained by an expert. The results achieved allows us to draw the conclusion that the automated method presented in this paper is a reliable system for extracting features such as morphometric landmarks on digital images. The landmarks obtained by the automated system are input into a morphometric software package using Procrustes co-ordinates for further scientific analysis. It is shown in this study that the method is sufficiently accurate to replace the manual digitisation process (8).

A fully automated system to extract morphometric landmarks is useful for the analysis of shape [1, 2], biological variability [8] and several other contexts [4]. The generic nature of object recognition and feature location incorporated in this automated system enables easy modification to locate features in a variety of other feature recognition tasks. The method is intrinsically robust to changes in shape and based firmly on the statistical interpretation of data analysis. The automated landmark identification system will benefit major research groups in the field of morphometrics and will easily be transferable to other relevant areas. The automation of shape analysis has major potential advantages regarding standardisation, as the landmarks can be located without any manual intervention and will make large-scale studies easily feasible [5, 8, 11]. 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.

8 Acknowledgment

This paper is an extended version of a paper delivered at the Visual Information Engineering Conference 2008 (29/7-1/8/2008).

9 References

[1] BOOKSTEIN F.L.: 'Morphometric tools for landmark data: geometry and biology' (Cambridge University Press, 1991)

[2] DRYDEN I.L., MARDIA K.V.: 'Statistical shape analysis' (John Wiley and Sons, Chichester, 1998)

[3] DUJARDIN J.P., BEARD C.B., RYCKMAN R.: 'The relevance of wing geometry in entomological surveillance of triatominae, vectors of chagas disease', *Infect. Genet. Evol.*, 2007, **7**, pp. 161–167

[4] BECERRA J.M., VALDECASAS A.G.: 'Landmark superimposition for taxonomic identification', *Biol. J. Linnean Soc.*, 2004, **81**, pp. 267–274

[5] ZIMMERMAN E., PALSSON A., GIBSON G.: 'Quantitative trait loci affecting components of wing shape in drosophila melanogaster', *Genetics*, 2000, **155**, pp. 671–683

[6] KLINGENBERG C.P., LEAMY L.J.: 'Quantitative genetics of geometric shape in the mouse mandible', *Evolution*, 2001, **55**, pp. 2342–2352

doi: 10.1049/iet-cvi.2009.0014

[7] KLINGENBERG C.P., LEAMY L.J., ROUTMAN E.J., CHEVERUD J.M.: 'Genetic architecture of mandible shape in mice: effects of quantitative trait loci analyzed by geometric morphometrics', *Genetics*, 2001, **157**, pp. 785–802

[8] GILCHRIST A.S., AZEVEDO R.V.R., PARTRIDGE L., O'HIGGINS P.: 'Adaptation and constraint in the evolution of drosophila melanogaster wing shape', *Evolution and Development*, 2000, **2**, (2), pp. 114–124

[9] KLINGENBERG C.P., BADYAEV A.V., SOWRY S.M., BECKWITH N.J.: 'Inferring developmental modularity from morphological integration: analysis of individual variation and asymmetry in bumblebee wings', *Amer. Natural.*, 2001, **157**, pp. 11–23

[10] KLINGENBERG C.P., LEAMY L.J., CHEVERUD J.M.: 'Integration and modularity of quantitative trait locus effects on geometric shape in the mouse mandible', *Genetics*, 2004, **166**, pp. 1909–1921

[11] KLINGENBERG C.P.: 'Morphometrics and the role of the phenotype in studies of the evolution of developmental mechanisms', *Gene*, 2002, **287**, pp. 3–10

[12] HOULE D., MEZEY J.G., GALPERN P., CARTER A.: 'Automated measurement of drosophila wings', *BMC Evol. Biol.*, 2003, **3**, (25), pp. 1471–2148

[13] THARAVY D.: Personal communication. Paris National Museum of Natural History, 2007

[14] PALANISWAMY S., THACKER N.A., KLINGENBERG C.P.: 'A statistical approach to feature detection in digital images'. Interdisciplinary Statistics Bioinformatics (LASR 2006), 2006, pp. 146–149

[15] CANNY J.: 'A computational approach to edge detection', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1986, **8**, (6), pp. 679–698

[16] PALANISWAMY S.: 'Automatic identification of landmarks in digital images of *Drosophila*: analysis of left-right asymmetry in *Drosophila* wings'. PhD thesis, The University of Manchester, UK, 2008

[17] THACKER N.A., RIOCREUX P.A., YATES R.B.: 'Assessing the completeness properties of pairwise geometric histograms', *Image Vis. Comput.*, 1995, **13**, (5), pp. 423–429

[18] EVANS A.C., THACKER N.A., MAYHEW J.E.W.: 'Pairwise representations of shape'. Proc. Int. Conf. on Computer Vision and Pattern Recognition, 1992

[19] EVANS A.C., THACKER N.A., MAYHEW J.E.W.: 'The use of geometric histograms for model based object recognition'. Proc. Fourth British Machine Vision Conf., 1993, pp. 429–438

[20] ASHBROOK A.P., THACKER N.A., ROCKETT P.I., BROWN C.I.: 'Robust recognition of scaled shapes using pairwise geometric histograms'. Proc. British Machine Vision Conf., 1995, pp. 503–512

[21] ASHBROOK A.P.: 'Pairwise geometric histograms for object recognition: developments and analysis'. PhD thesis, The University of Sheffield, UK, 1998

[22] EVANS A.C.: 'Geometric feature distributions for shape representation and recognition'. PhD thesis, The University of Sheffield, UK, 1994

[23] RIOCREUX P.A., THACKER N.A., YATES R.B.: 'An analysis of pairwise geometric histograms for view-based object recognition'. Proc. British Machine Vision Conf., 1994

[24] ASHBROOK A.P., ROCKETT P.I., THACKER N.A.: 'Multiple shape recognition using pairwise geometric histogram based algortihms'. Proc. IEE Image Processing, 1995

[25] PALANISWAMY S., THACKER N.A., KLINGENBERG C.P.: 'Automatic identification of morphometric landmarks in digital images'. Proc. British Machine Vision Conf., 2007, vol. 2, pp. 870–879

10 Appendix

It is a general notion that features in an image I can be enhanced, ready for threshold detection, using a convolution with a kernel shaped similar to the feature t

$$f(x, y) = \sum_{ij} t(i, j)I(x + i, y + j)$$

For an image with independent random Gaussian noise of width σ we can write the likelihood for a template t' describing an image region as

$$L = \sum_{ij} \frac{(\alpha(x, y)t'(i, j) - I(x + i, y + j))^2}{\sigma^2 + \mathrm{var}(t')}$$

where $\mathrm{var}(t')$ is the variance (i.e. the accuracy) of the assumed template, and α is a scale factor defining the strength of the linear contribution of t' to the image. Assuming that $\mathrm{var}(t')$ is independent of α, the likelihood estimate of α is

$$\alpha(x, y) = \sum_{ij} \frac{t'(i, j)I(x + i, y + j)}{\sigma^2 + \mathrm{var}(t')}$$

This justifies using the template

$$t = \frac{t'(i, j)}{\sigma^2 + \mathrm{var}(t')}$$

Exact knowledge of the template within a defined region, and ignorance elsewhere leads directly to $t = t'$. However, it is also sensible to assume that our knowledge of the expected

template reduces away from the centre (i.e. a radial weighting) such as

$$t = \frac{t'(r, \theta)}{\sigma^2 + w(r)}$$

For example $w(r)$ may be set to give an overall Gaussian weighting

$$t = t'G(r) \quad \text{with} \quad w(r) = 1/G(r) - \sigma^2$$

Notice, that from a matching perspective this does not give a unique interpretation for any particular t', as there are an infinite number of possible weightings we can use in the construction of a specific t. In addition, for a given level of image 'signal' the best response will be given to an image looking like t and not t'. In some applications more specificity to a specific shape may be achieved by locating features on the likelihood L, rather than using α.

IET Comput. Vis., 2010, Vol. 4, Iss. 4, pp. 247–260

doi: 10.1049/iet-cvi.2009.0014