Tina Memo No. 2008-006
Presented at VIE 2008, Xi An, China. (prize paper)

Automated Landmark Extraction in Digital Images - Performance Evaluation.

Sasirekha Palaniswamy, Neil.A.Thacker, Christian Peter Klingenberg.

Last updated
21 / 7 / 2008

WWW.TINA-VISION.NET

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

Automated Landmark Extraction in Digital Images - Performance Evaluation

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, UK

Abstract

In this study, we present an automated system for feature recognition in digital images. Such a system is very important for morphometric analysis as it will replace the time consuming and labour intensive process of manual identification. 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. The performance of the system, its reliability and robustness are evaluated in comparison to the expert manual performance.

1 Introduction

The framework of shape analysis by landmarks is increasingly used in many biological and medical applications and widely applied in many other fields. For instance, the configuration of landmarks have helped identify the possible source of re-infesting specimens and encounter the epidemiologically challenging vectors of Chagas disease [7]. The potential of using geometrical morphometric techniques as an invaluable tool for recognising taxonomic data is being explored [2]. 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 [16, 12]; geographic differentiation [8], and the study of morphological integration [11].

In this paper, we focus on the automatic extraction of morphometric landmarks (fig:1) 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 [3, 6]. A typical shape analysis study involves several hundred digital images and extracting landmarks manually is time consuming. 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 [9]. Although these methods increase the efficency of human effort, the need for the manual observer has not been eliminated as the methods require 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.

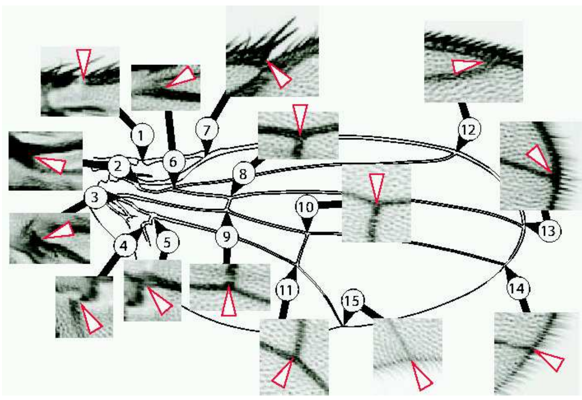


Figure 1: A set of 15 morphometric landmarks in *Drosophila*.

2 Methods

The dataset consists of 856 gray scale images of the *Drosophila* wings. Two independent sets of reference (training) images with 5 images in each set is used. 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. Before extracting features in the image it is important to preprocess the image with an appropriate enhancement technique. In this study, a Difference of Gaussian (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, 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.

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. It is shown in [14] 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 [4], but has to be modified to meet the needs of detecting ridge features in this study. The enhanced features and a 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)

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. These geometric properties have the characteristics that are invariant to in-plane rotation, translation and line extension. 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. By comparing such distributions (PGH) for the model and the scene data 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. This can be achieved by using the Bhattacharya similarity metric [13]. In addition, when compared in this way, the PGH shape matching scheme is robust to the loss of data due to fragmentation noise and occlusion. It is also characterised by its invariance to parts of the same linear feature, thereby making the method suitable for the detection of extended lines via clustering. Appropriate weighting of the entries in the PGH ensures sudden changes are not introduced in the representation by gradual changes in input data and 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 [15].

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 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. 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. Furthermore, the constraints imposed by pairs of scene lines and the associated Gaussian error model, results in the a pose estimation that is equivalent to computing a robust least-squares fit between the model and the scene data. Although, the probabilistic Hough transform is used to locate models using the positions, orientations and scales hypothesised by the scene line labels, the orientation and the scale of the model are determined using separate one parameter Hough transforms using data consistent with hypothesised locations. The orientation is determined by taking into account of the difference in orientation between the matching model and scene line [1].

2.4 Template Matching

The Hough prediction of the location of the landmark co-ordinates is based on the global shape of the features. To obtain sub-pixel accurate landmark location demanded by the Morphometric studies, 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. The template matching stage 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.

To sensibly combine the multiple estimated locations into a unique hypothesised landmark location, we use the least square matching scores generate a ranking. The landmark values obtained by the best three ranking reference images are compared to ensure that the values are consistent. If the separation between them is below certain distance threshold (determined from histograms of repeated localisation) , the mean of these will provide the final estimated landmark value of the automated system. Alternatively, the mean of the landmark values are obtained from the best two ranking reference images. By checking for the consistency of the results obtained by the best three scoring reference images in this way, we can reduce the possibilities of getting outliers. It improves the reliability of the results obtained and ensures an improvement in the accuracy of the final landmark estimation.

3 Evaluation

To test the repeatability and robustness of the automated system, the automated method is used to extract the landmarks on the above dataset but with two sets of 5 reference images. 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. Finally, an overall comparison is performed between manual and automatic mark-up locations in the context of a morphometric analysis using procrustes alignment.

4 Results

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. The accuracy of the results obtained by the automated system is evaluated by comparing the results to the manual results obtained by an expert. 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.

The plots (fig:3(a)-4(c)) 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.

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. The plots (fig:5-6) 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.

The plots (fig:7-8) show the comparison in performance obtained by increasing the number of reference images. This improves the accuracy of certain features (landmarks 6 & 12). However there is no evidence of significant improvement on most other landmarks. This could be due to the fact that some of the features for instance landmark 12 (see fig:1) can be highly variable. By increasing the number of reference images the chance for a suitable matching reference image increases, thereby improving its accuracy. It was noted that the presence of thick bristles in the surrounding region affects the accuracy of the landmark location. On the other hand, most of the 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.

The plots (fig:6-10) show the comparison of the width of the distribution of the Procrustes co-ordinates for the manual and the automated results. The figure (2) shows the Procrustes fit for the manual and automated results (offset slightly for easy comparison). It is clear that after 2% outlier removal, i.e., about 17 images, the width of the distribution is comparable to an expert manual performance. While manual mark-up does not produce these outliers, we consider such losses to be negligible in the context of such experiments.

5 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. The work provide 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, are intended to provide a compact representation which generalises well to unseen data. The ability to deal with scale change has also been identified [1], in a manner consistent with the SIFT approach [5]. Though the PGH representation, the arguments regarding its completeness [15] and techniques for statistical matching [13], predate this more popular work by half a decade. The simple procedure of regional template matching is shown to be extremely robust in sub-pixel localisation of the features.

In this paper, we have focused mainly on the evaluation of the performance of the automated system. The results from the automated system are compared to the human performance. The method is tested for its repeatability and robustness. The results are further analysed for an improvement in performance by providing more reference images. We can see that the combination of the landmarks obtained by 10 reference images performs better than the individual sets of 5 reference images, especially on complicated features (fig:7-8). Significant improvement on the accuracy of the most challenging landmarks such as landmark 12, is attributed to the high biological variability, with additional training images providing more appropriate data with which to compute a precise location.

6 Conclusion

In this study, we present an automated system for feature recognition in digital images. The performance of the system and its robustness is 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 (2).

A fully automated system to extract morphometric landmarks is useful for the analysis of biological variability [8] and several other contexts [2]. The automated method has potential advantages and makes large scale studies in the field of genetics and evolutionary morphometrics feasible [10, 16, 8]. 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*.

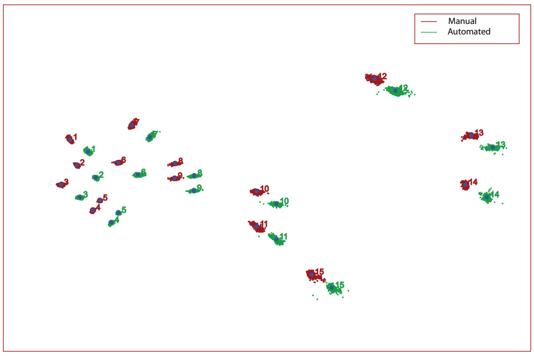


Figure 2: Procrustes Fit - Comparison of the width of distribution between manual (red) and automated (green) results.

References

[1] A.P. Ashbrook, N.A. Thacker, P.I. Rockett, and C.I. Brown. Robust recognition of scaled shapes using pairwise geometric histograms. *proc. BMVC 95*, pages 503–512, 1995.

[2] J.M. Becerra and A.G. Valdecasas. Landmark superimposition for taxonomic identification. *Biological Journal of the Linnean Society*, 81:267–274, 2004.

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

[4] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern analysis and Machine Intelligence*, 8(6):679–698, 1986.

[5] D.G.Lowe. Object recognition from local scale-invariant features. *Int. Conf. Comp. Vision (ICCV)*, pages 1150–1157, 1999.

[6] I.L. Dryden and K.V. Mardia. *Statistical Shape Analysis*. Chichester, John Wiley and Sons, 1998.

[7] J.P. Dujardin, C.B. Beard, and R. Ryckman. The relevance of wing geometry in entomological surveillance of triatominae, vectors of chagas disease. *Infection, Genetics and Evolution*, 7:161–167, 2007.

[8] A.S. Gilchrist, R.B.R. Azevedo, L. Partridge, and P. O'Higgins. Adaptation and constraint in the evolution of drosophila melanogaster wing shape. *Evolution & Development*, 2:114–124, 2000.

[9] D. Houle, J.G. Mezey, P. Galpern, and A. Carter. Automated measurement of drosophila wings. *BMC Evolutionary Biology*, 3(25):1471–2148, 2003.

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

[11] C.P. Klingenberg, L.J. Leamy, and J.M. Cheverud. Integration and modularity of quantitative trait locus effects on geometric shape in the mouse mandible. *Genetics*, 166:1909–1921, 2004.

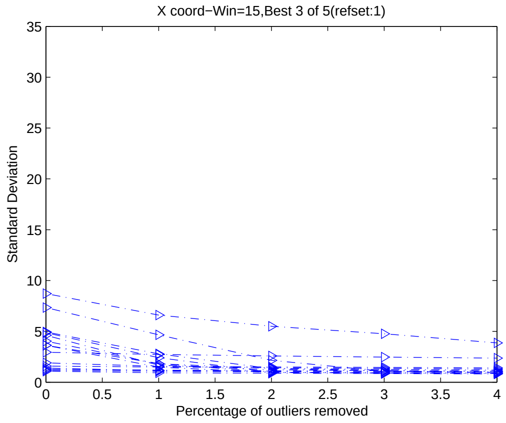
[12] C.P. Klingenberg, L.J. Leamy, E.J. Routman, and J.M. Cheverud. Genetic architecture of mandible shape in mice: effects of quantitative trait loci analyzed by geometric morphometrics. *Genetics*, 157:785–802, 2001.

[13] P.I.Rockett. N.A.Thacker, F.J.Aherne. The bhattacharyya metric as an absolute similarity measure for frequency coded data. *Kybernetika*, 34(4):363–368, 1997.

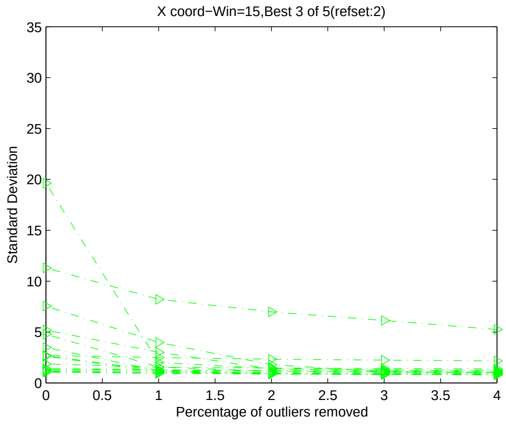
[14] S. Palaniswamy, N.A. Thacker, and C.P. Klingenberg. A statistical approach to feature detection in digital images. *Leeds Annual Statistical Research Workshop*, pages 146–149, 2006.

[15] N.A. Thacker, P.A. Riocreux, and R.B. Yates. Assessing the completeness properties of pairwise geometric histograms. *Image and Vision Computing*, 13(5):423–429, 1995.

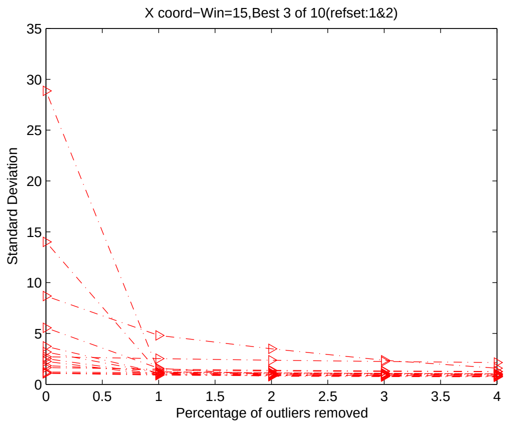
[16] E. Zimmerman, A. Palsson, and G. Gibson. Quantitative trait loci affecting components of wing shape in drosophila melanogaster. *Genetics*, 155:671–683, 2000.



(a) Results obtained using best 3 of 5 from Refset 1.

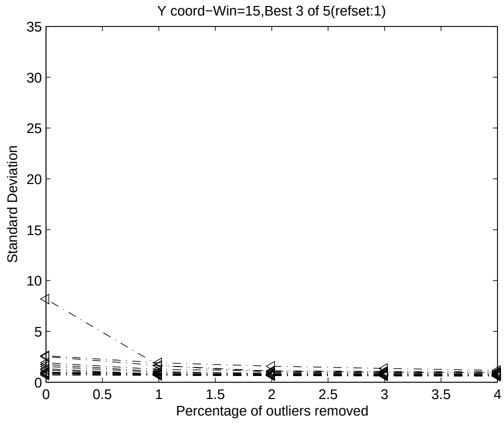


(b) Results obtained using best 3 of 5 from Refset 2.

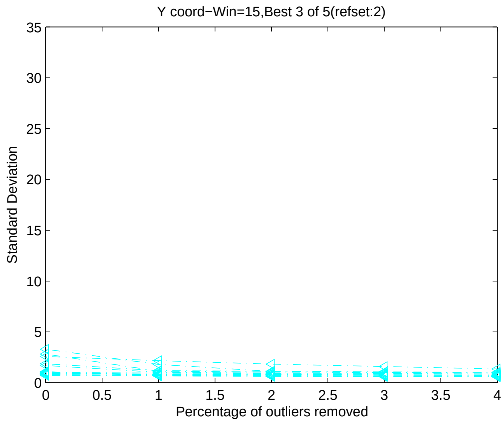


(c) Results obtained using best 3 of 10 from Refset 1&2.

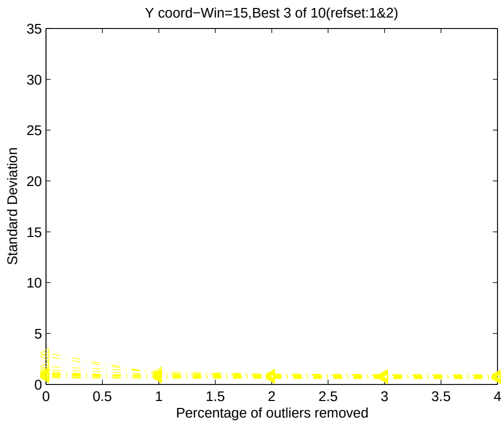
Figure 3: Standard deviation of the automated results from the manual results. The X co-ordinates for the 15 landmarks are plotted as a function of outlier removal to show the stability of the results.



(a) Results obtained using best 3 of 5 from Refset 1.



(b) Results obtained using best 3 of 5 from Refset 2.



(c) Results obtained using best 3 of 10 from Refset 1&2.

Figure 4: Standard deviation of the automated results from the manual results. The Y co-ordinates for the 15 landmarks are plotted as a function of outlier removal to show the stability of the results.

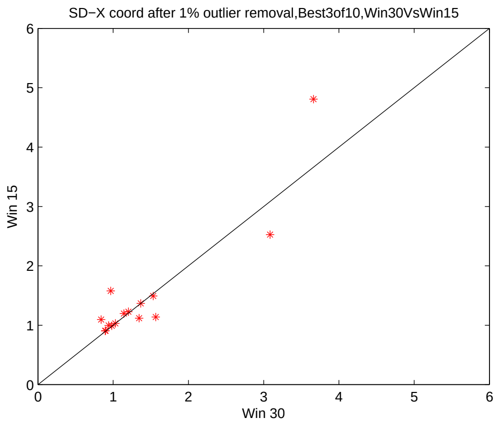


Figure 5: Comparison of the effect of template matching window size parameters (δ=15 & 30 pixels) on X co-ord standard deviation.

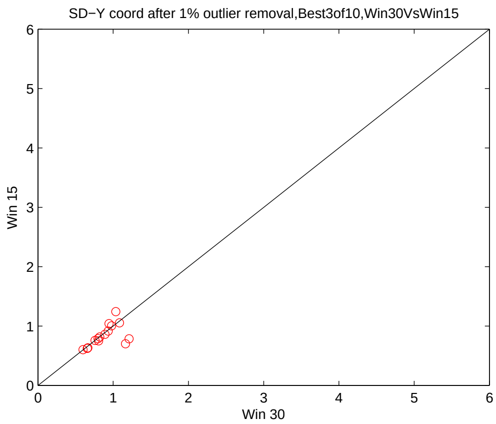


Figure 6: Comparison of the effect of template matching window size parameters (δ=15 & 30 pixels) on Y co-ord standard deviation.

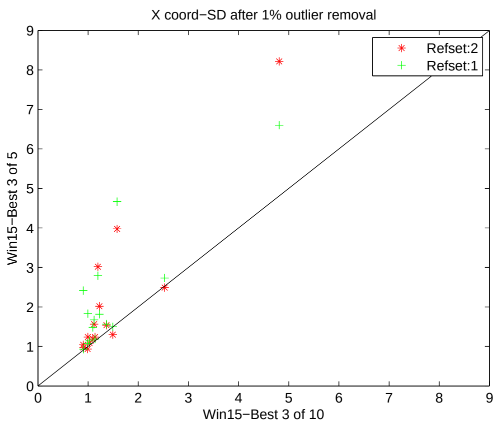


Figure 7: Results obtained by the best 3 matching templates from a set of 5 and 10 reference images are compared (X co-ord).

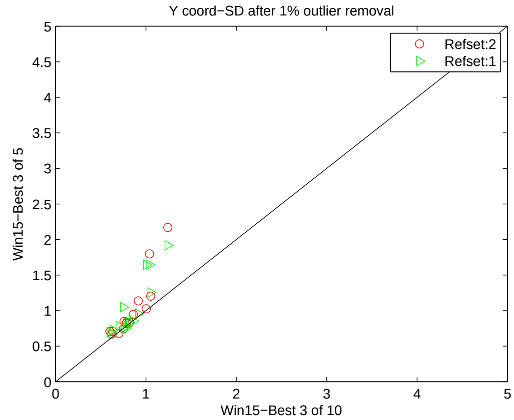


Figure 8: Results obtained by the best 3 matching templates from a set of 5 and 10 reference images are compared (Y co-ord).

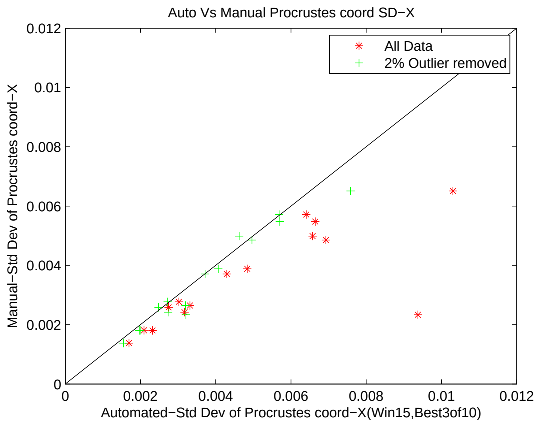


Figure 9: Comparison of the standard deviation of the Procrustes (X) co-ordinates obtained from manual and automated results.

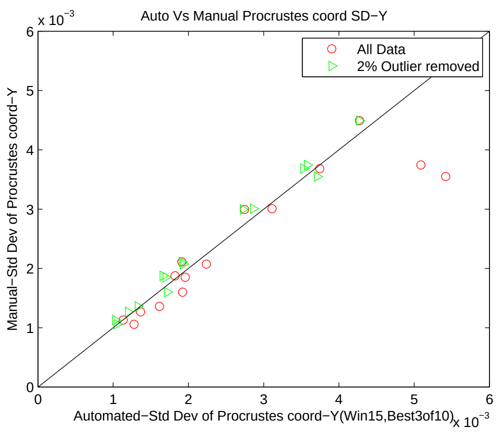


Figure 10: Comparison of the standard deviation of the Procrustes (Y) co-ordinates obtained from manual and automated results.