A VIEW-GEOMETRIC APPROACH TO VIEW AND OCCLUSION INVARIANT SHAPE RECOGNITION AND RETRIEVAL

Alper Yilmaz[a] and Gabor Barsai[b*]

[a] The Ohio State University, Dept. of Geod. Science,Photogrammetric Computer Vision Laboratory
The Ohio State University -yilmaz.15@osu.edu
[b] Gotmaps? Inc., 2981 Wicklow Rd., Columbus, OH - gabor@gotmaps.biz

KEYWORDS: image interpretation, 2-D feature extraction, computer vision, feature recognition, machine vision

ABSTRACT:

In this paper, we propose a recognition technique for geographic features which are represented as closed contours. Our algorithm relies on the planar projective geometry between the contours and exploits the properties of the Fourier Transform. One of the contributions of the proposed method is that it can recognize features acquired from any viewing direction and even partially occluded. Another contribution is that this method is independent of the starting point of the contour and the digitizing direction. In addition, this method does not require conjugate, or matching points that are traditionally required in projective geometry. The experimental results on an in-house developed geographic database and the Brown University shape database show robust recognition performance.

1 INTRODUCTION

The measurement of shape similarity between two objects is an essential task in many areas, including object recognition, classification, mobile mapping, surveillance, trajectory calculation, event analysis and retrieval (Zhang and Lu, 2001) (Schenk, 2001) (Noor et al., 2006); and it has received a lot of attention since the earliest pictures were taken. For example, by World War I, the opposing sides routinely took aerial photos of each others' positions and identified landmarks for intelligence gathering. The abundance and easy access today to digital images makes matching especially relevant among the listed tasks. Ideally, recognition of objects should be projection, scale, translation and rotation invariant, just as they are in human vision. This, however, is a very complex problem, since numerous times an object is occluded and many objects rarely appear the same twice, due to different camera/observer positions, variable lighting or object motion. According to Meyer (Meyer, 1993), the ultimate goal in this regard is to investigate automatic object recognition in unconstrained environments by means of outlines of the objects, which we will refer to as the contours. In this paper, we study the problem of matching 2-D contours. One of the reasons for the popularity of contour-based analysis techniques is that edge detection constitutes an important aspect of shape recognition by the human visual system (van Otterloo, 1991) (Schenk, 2001). Rui (Rui et al., 1998), Zahn and Roskies (Zahn and Roskies, 1972), Zhang and Fiume (Zhang and Fiume, 2002), Wallace and Wintz (Wallace and Wintz, 1980) use Fourier Descriptors to match contours. Other methods used to recognize shapes are moment based and structure based approaches. The advantages of moments (easy to calculate) are outweighed by their disadvantages (not intuitive) (Teague, 1980) (Zhang and Lu, 2001). In particular, it is difficult to correlate high-order moments with one of these shape features (DeValois and DeValois, 1980). The representation of curves/contours using FDs gives a continuous function. Using FDs, a better reconstruction of the curve/contour can be created than by just using moments. Using only moments, reconstruction of the curve is difficult, if not impossible. Belonged (Belongie et al., 2002) develop a measure called shape context for comparison. Shapiro (Shapiro, 1979) writes about the structure of shape in and early work, about how shapes can be defined, and compares different structure methods. Using structural information is, however, not efficient when compared to contours, and structural methods, especially those using graph-like representations, usually lead to variants of the computation-intensive graph isomorphism algorithm (Shapiro, 1979) (van Otterloo, 1991) (Zhang and Lu, 2001). For an extended introduction of these techniques, please refer to any of the several survey papers by DeValois and DeValois (DeValois and DeValois, 1980), van Otterloo (van Otterloo, 1991), Loncaric (Loncaric, 1998) and Veltkamp (Veltkamp, 2001). Many of these matching methods rely on simple transformations, such as translation, rotation and scaling. Recognition under more general transformations, such as affine and projective transform, however, has not been fully examined, due to the complex nature of these transforms. A projective transform is also known as a homography. In a homography, a ratio of ratios or cross ratio of lengths on a line is the only projective invariant. The main motivation behind this work is that 2-D homography may overcome the problem of noise sensitivity and boundary variations. The Fourier transform, or the Fourier descriptors (FDs), of the contour are used to represent the curve parametrically. We propose to use the homography transform along with the FDs to match contours, applying the digitized coordinates of the contours. We use FDs, since ideally, shape representation should be invariant to scale, rotation translation and starting point, robust to noise, errors, efficient in computing the representative terms and efficient for use in matching (Rui et al., 1998), and many of these task are handled effectively by the Fourier transform. An important contribution of this paper is the elimination of the requirement of corresponding points. The paper by Belongie (Belongie et al., 2002) is probably closest in spirit of this paper, although with a different approach. For this study, several images of countries, lakes and other features were digitized from maps, satellite images, silhouettes to obtain the contours.

The paper is organized as follows: Section 2 describes projective geometry for a background on homography. Section 3 outlines the methodology used: converting the x and y coordinates into periodic functions for use by the Fourier transform

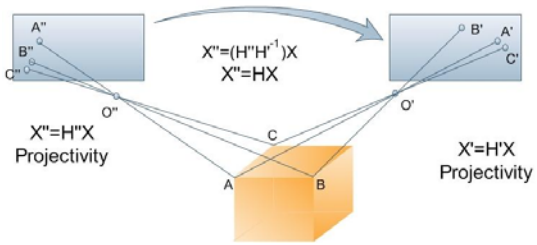


Figure 1: Planar homography between two views of a target. Planar homography requires corresponding points across two views.

and the Power Spectrum, then using the values of the Power Spectrum as the basis for the least squares adjustment to find the projectivity between the two Power Spectrums. Section 4 shows the robust recognition results from the condition number and variance.

2 PROJECTIVE GEOMETRY

Methods, like similarity, affine and projective transforms, which exploit projective geometry require point correspondences across different views, such that any given point in one view corresponds to one and only one point in the other, and vice versa. We establish a relationship between the x, y coordinates of two contours of the same object on different sources with the planar homography matrix. If a planar object is imaged from multiple viewing positions, the result is a projective image-to-image homography (Hartley and Zisserman, 2000). Under planar homography, points in one image are transformed to the points in the other image as:

$$x' = Hx, \quad (1)$$

where x' = (x', y', 1) and x = (x, y, 1) are corresponding points across images in homogeneous coordinates and H is the 3 × 3 matrix known as the planar homography matrix:

$$H=\begin{pmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{pmatrix} \quad (2)$$

For geometric features imaged from a distant viewpoint, the projective homography reduces to affine homography:

$$H=\begin{pmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ 0 & 0 & 1 \end{pmatrix} \quad (3)$$

This approximation is considered adequate, since the projective portion is minimized compared to the affine (Shapiro and Zisserman, 1995).

3 METHODOLOGY

While establishing correspondences across images is feasible and has been a common practice among researchers, finding corresponding points between two contours is not easy and at times it is impossible. Due to this observation, a common tradition has been to represent the contours in parametric form prior to any processing. Parametric contour representation can be generated by methods including but not limited to: curvature/spline based, polar coordinate (geometric) based, trigonometric based, wavelet transform based, Fourier descriptor (FD) based. In contrast to others, FD and wavelet based parametric contour representations provide continuous functions. Compared to the FD based methods, wavelet based methods, however, involve intensive computation, and it is usually not clear which basis would be a better choice to represent the contour. (Zhang and Lu, 2001). Let two periodic functions x(t) and y(t) describe the contour, such that we treat x and y coordinates as independent dimensions (Zahn and Roskies, 1972). We will be using periodic functions, since it is analogous to shifting the starting point on a contour. In order to objectively compare these functions to other x and y coordinates, we set a standard reference to the length of the contour, and select its period as 2 Π. Let l be the arc length from an arbitrary starting point, to a point p, and let L be the entire length of the closed curve. In this form the arc-length is converted to its angular representation by:

$$\varphi = 2\Pi \frac{l}{L} \quad (4)$$

This suggests that L is the period of the x and y functions of a contour. Given a digital image, the contour of the geometric shape constitutes a dense set of discrete points. Expressing the parametric form of the discrete point set can be obtained by taking the Fourier transform of x and y independently:

$$F(x) = \sum_{n=-\infty}^{n=\infty} f(x)e^{-i\omega(n)t}, \quad (5)$$

$$F(y) = \sum_{n=-\infty}^{n=\infty} f(y)e^{-i\omega(n)t}, \quad (6)$$

where F is the Fourier transform of the f(x). For finite terms, n = (−N/2)...(N/2 − 1), where N is the number of points (even number).

Using equations (5) and (6) and after some manipulations, which includes dividing each side with $[e^{-i\omega(1)t}, e^{-i\omega(2)t}, ..., e^{-i\omega(n)t}]$, it is easy to show that the homography transform in equation 1 becomes

$$F(X0) = H \quad F(X) \quad (7)$$
$$(3,n) \quad (3,3) \quad (3,n)$$

where X = [x y scale]

Let Gk be the k^{th} Fourier coefficient. Then the coefficients in eq. 7 can be expressed by:

$$G0k = H \quad Gk. \quad (8)$$
$$(3,n) \quad (3,3) \quad (3,n)$$

Considering the case of a different starting point on the contour, according to Fourier theory a shift in the starting point of a function is the same as multiplying the coefficients by a rotation matrix (Schenk, 2001)

$$G0k = Gk\, e^{i2\Pi kn0/N} \quad (9)$$

where n0 is the shift in starting point (number of points), k is the harmonic coefficient, N is the total number of points on the contour. In this formulation, the phase shift is different for every k harmonic coefficient. This observation presents several problems, the most major being that shifting the starting point of an object is directly proportional to the harmonic number. In order to eliminate this effect, we use the Power Spectrum (PS) of the Fourier transform. Since the change of the starting point corresponds to a rotation matrix,

$$R=\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \quad (10)$$

Multiplying this matrix with its transpose would result in an identity matrix (since a rotation matrix is orthogonal). For complex numbers, this means multiplying the number with its complex conjugate:

$$PS = \sum_n \; ||f(x)e^{-i\omega(n)t}||^2 = F\,\overline{F} \quad (11)$$

where F is the Fourier transform of f(x), and $\overline{F}$ is its complex conjugate. The Power Spectrum of an object is the sum of the square of the magnitude of the x, y Fourier descriptors. In equation (8), if Gkx and Gky is the k^{th} complex Fourier descriptor for x and y, and G0kx and G0ky is the kth complex Fourier descriptor for x0 and y0, then the power spectrum is $(||Gkx||)^2$ + $(||Gky||)^2$ and $(||G'kx||)2$ + $(||G'ky||)^2$ for the k^{th} harmonic. This value is constant and independent of rotation or starting point, for any k. The result of this operation is a function where the only variable is the magnitude of each harmonic.

3.1 Matching with the Power Spectrum

We hypothesize that projectivity between two contours is sufficient for the existence of projectivity between their respective PSs. Hence, checking for the existence of projectivity between PSs suggests that two contours are projectively equivalent. Introducing the power spectrum into the harmonics in equation (7) and developing these equations establish a set of equations as in the following:

$$\begin{aligned} &||\mathbf{G0kx}||^2 + ||\mathbf{G0ky}||^2 = \mathbf{(h1)}^2||\mathbf{(Gkx)}||^2 + \\ &\mathbf{(h2)}^2||\mathbf{(Gkx)}||^2 + 2||\mathbf{(Gkx)}||\,||\mathbf{(Gky)}||\,\mathbf{h1h2} + \\ &\mathbf{(h3)}^2||\mathbf{(Gky)}||^2 + \mathbf{(h4)}^2||\mathbf{(Gky)}||^2 + \\ &2||\mathbf{(Gky)}||\,||\mathbf{(Gkx)}||\,\mathbf{h3h4.} \quad (12) \end{aligned}$$

Rearranging this equation and putting the unknowns into a vector form results in:

$$\begin{aligned} &(||\mathbf{G'kx}||^2 + ||\mathbf{G'ky}||^2) = \\ &(\,||\mathbf{Gkx}||^2\;\; ||\mathbf{Gky}||^2\;\; 2||\mathbf{Gkx}||\,||\mathbf{Gky}||\,) \\ &\begin{pmatrix} h1h1 + h2h2 \\ h3h3 + h4h4 \\ h1h2 + h3h4 \end{pmatrix} \quad (13) \end{aligned}$$

This is in the well known form of y = Ax, where y and A are the PS coefficients of the two contours, and x in this case is the unknown homography coefficients between the PSs. Since there are more equations than unknowns, we can use a least squares solution to calculate the homography coefficients. The matching between two contours can be expressed by evaluating the quality of the projective relation depicted in equation (13). Particularly, the quality of an equation system can be computed by the condition number of A or the empirical estimate of the residual variance.

Among these two measures, the condition number evaluates the residual error calculated from the least squares solution and the values of the PS. In order to calculate the condition number, let e be the error in y. Then the error in the solution $A^{-1}y$ is $A^{-1}e$.

The ratio of the relative error in the solution to the relative error in y is

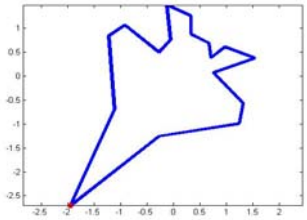
$$\kappa = (||A^{-1}e||\;||A^{-1}y||\;) / (||e||\;||y||). \quad (14)$$

A lower condition number suggests a better parameter estimation; hence a better matching between the contours. The variance, on the other hand, is the measure of statistical dispersion of the residual. In other words, it is a measure of how spread out a distribution is and how much variability there is in the distribution:

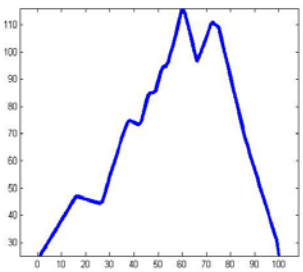
$$Var(X) = E((X - \mu)^2), \quad (15)$$

where μ is the average of the variables contained in X. In this paper, in order to compute the matching between two contours, we use a combination of both the condition number and the variance combination by weighting the variance with the inverse condition number:

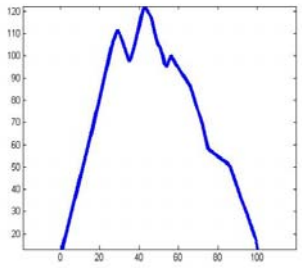
$$MatchingScore = 1/(\kappa\, Var(X)). \quad (16)$$



(a)



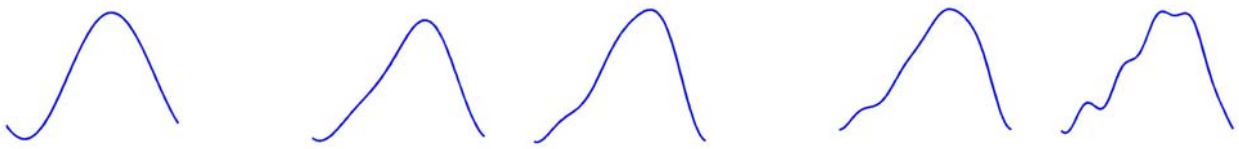
(b)



(c)

(a) original shape (F15), (b) x coordinate, (c) y coordinate

Figure 2: Analysis of an object by separating it into its x and y coordinates. Starting point is at nose of F15 plane, going clockwise.



(a) (b) (c) (d) (e)

Figure 3: Reconstruction of x coordinate of Figure 2 above using first six Fourier descriptors out of 100

.

The first descriptor is the average, and will be left out. (a) Second descriptor only (b) Sum of second and third (c) Sum of second, third and fourth (d) Sum of second, third, fourth and fifth (e) Sum of second, third, fourth, fifth and sixth

For least squares matching, the number of terms has to be the same on both sides of the equation. Since the number of points on each feature may not be the same, we use the lower number of 3 coefficients between the compared shapes for each matching. Using less coefficients gives us the opportunity to compare the same first coefficients with each other, without changing the shape itself, it is also more efficient, since the coefficients need to be calculated once. The other solution would be to delete some of the points in the shape with the higher number of points, but this would change the contour itself. The results are comparable to Belongie (Belongie et al., 2002), with the added benefit of having a simpler algorithm than shape contexts, although our method only considered the outline contours, not contours within contours.

4 RESULTS

We tested the method with both geographic and non-geographic features. The geographic features are manually extracted from maps and images. The non-geographic images, on the other hand, are obtained from the silhouette database from the Brown University (Database, 2007). The contours used in this study are represented by a set of sampled points. Some contours were digitized with intentional errors, others were sampled (at regular intervals) from an output from an edge-detector. There is nothing unique about the detected edges, such that they are not intersection points or break points. The number of sampled points was also varied. The results are generated by comparing each feature against all the features used in the experiments. The similarity between the features are generated using equation (16) which exploits the empirical estimate of variance and the condition number of the equation system generated from the homography transform between the Fourier descriptors of the contours. We tabulated the matching recognition performance of the proposed method in the form of a confusion matrix which is shown in Figure 5. Ideally, the regions marked by red outlines, which correspond to the clusters of objects, should have highest similarity and the other regions in the matrix should have no similarity. Representing a high match by white and no match by black color codes, the performance of the method provides shades of gray which shows robust matching performance. An affine projection of F15 gives close to 0 error when compared to the original, as expected, due to round-off. It even provides robust matching for different projections, like with the two instance of the Mexico. Similar performance results are observed when the features are occluded as shown for two instances of the Lake Superior and the occluded hand silhouettes. We have observed similar performances for three instances of the Mexico map, where the occluded version very well matches with the two other instances. An interesting observation is that, the match score tells us that F15 and F16 have some similarity which can be considered true since both are silhouettes of planes. We should note that for a human observer, all the geographic features have some similarity, such as most maps used in the experiments have small peninsulas visible at one end of the feature. This observation, however, is not valid for the Staten Island, which has a more elliptical shape and has a smoother outline compared to the rest.

5 CONCLUSION

This paper provides a novel approach to matching objects represented in the form of a silhouette. Compared to many other recognition method in the literature, our method allows extracted silhouettes and their outlines to contain noise and occlusions. Additionally, the method resolves projective deformations to the objects which occur due to perspective viewing effects. The proposed approach exploits the projective geometry, which results in a robust and computationally simple procedure. An important contribution in this regard is the elimination of the point correspondences, having the same staring points on the silhouette outline and the direction of digitization of the outline. Experimental results show the robustness of the proposed method.

REFERENCES

Belongie, S., Malik, J. and Puzicha, J., 2002. Shape matching and object recognition using shape contexts. IEEE Trans. on Pattern Analysis and Machine Vision 24(24), pp. 509–522.

Database,2007.http://www.lems.brown.edu/vision/researchareas/siid/index.html. Brown University.

DeValois, R. L. and DeValois, K., 1980. Spatial vision. Ann. Rev. Psychol.

Hartley, R. and Zisserman, A., 2000. Multiple View Geometry. University Press, Cambridge.

Loncaric, S., 1998. A survey of shape analysis techniques. Pattern Recognition 31(8), pp. 983–1001.

Meyer, Y., 1993. Wavelets, algorithms and applications. Society for Industrial and Applied Mathematics.

Noor, H., Mirza, S., Sheikh, Y., Jain, A. and Shah, M., 2006.

Model generation for video based object recognition. ACM Multimedia 2006.

Rui, Y., She, A. and Huang, T. S., 1998. A modified fourier descriptor for shape matching in mars. Image Databases and Multimedia Search, Series on Software Engineering and Knowledge Engineering 8, pp. 165–180.

Schenk, A., 2001. Digital Photogrammetry Vol.1. Terrascience.

Shapiro, L. and Zisserman, A., 1995. 3d motion recovery via affine epipolar geometry. International Journal for Computer Vision 16, pp. 147–182.

Shapiro, L. G., 1979. A structural model of shape. Dept. of Computer Science, Virginia Polytechnic Technical Report CS79003-R.

Teague, M. R., 1980. Image analysis via the general theory of moments. Optical Society of America 70, pp. 920–930.

van Otterloo, P. J., 1991. A Contour-oriented Approach to Shape Analysis. Prentice Hall.

Veltkamp, R. C., 2001. Shape matching: Similarity measures and algorithms. In: Proc. of the International Conference on Shape Modeling and Applications, Genova, Italy, pp. 188–197.

Wallace, T. P. and Wintz, P. A., 1980. An efficient 3-d aircraft recognition algorithm using normalized fourier descriptors.

Computer Graphics and Image Processing 13, pp. 99–126.

Zahn, C. T. and Roskies, R. Z., 1972. Fourier descriptors for plane closed curves. IEEE Transactions on Computers 21(3), pp. 296–281.

Zhang, D. and Lu, G., 2001. A Comparative Study on Shape Retrieval Using Fourier Descriptors with Different Shape Signatures. Proc. of International Conference on Intelligent Multimedia and Distance Education, Fargo, ND, USA.

Zhang, H. and Fiume, E., 2002. Shape matching of 3-d contours using normalized fourier descriptors. In: Proceedings of the Shape Modeling International, IEEE, (SMI.02)

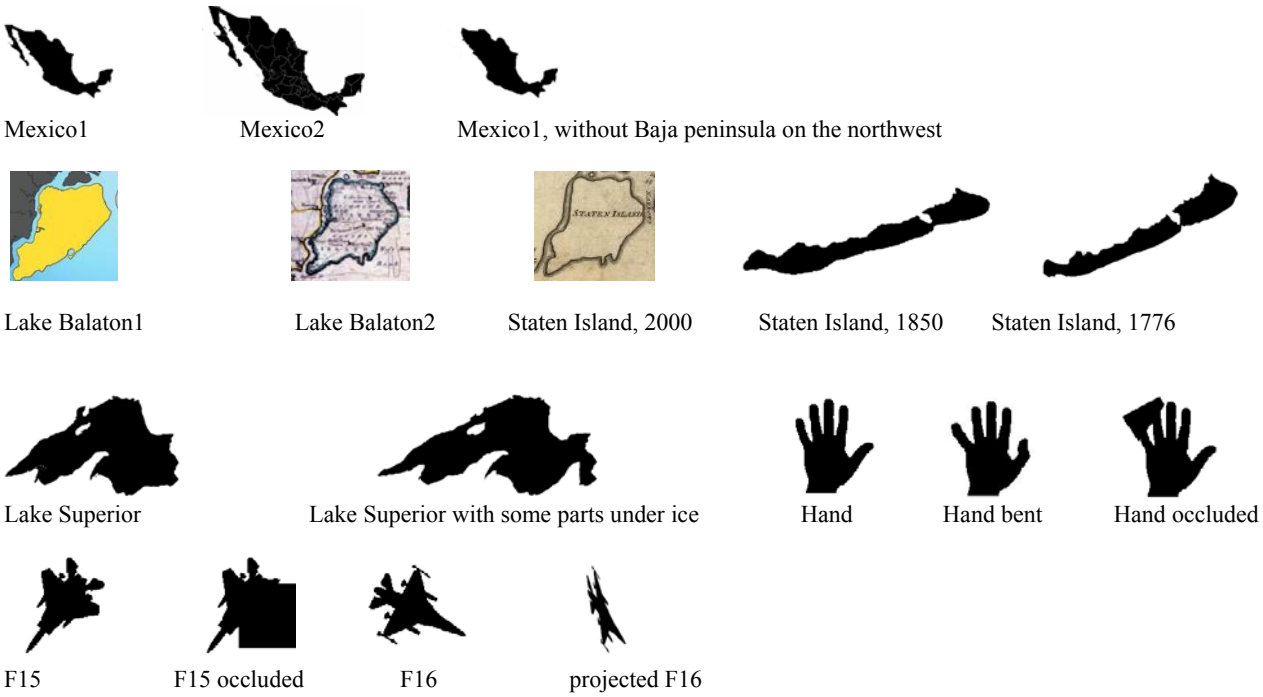


Figure 4: Images used to perform the experiments.The silhouette images are obtained from the Brown U. silhouette database, and the geographic features are from in-house generated database.

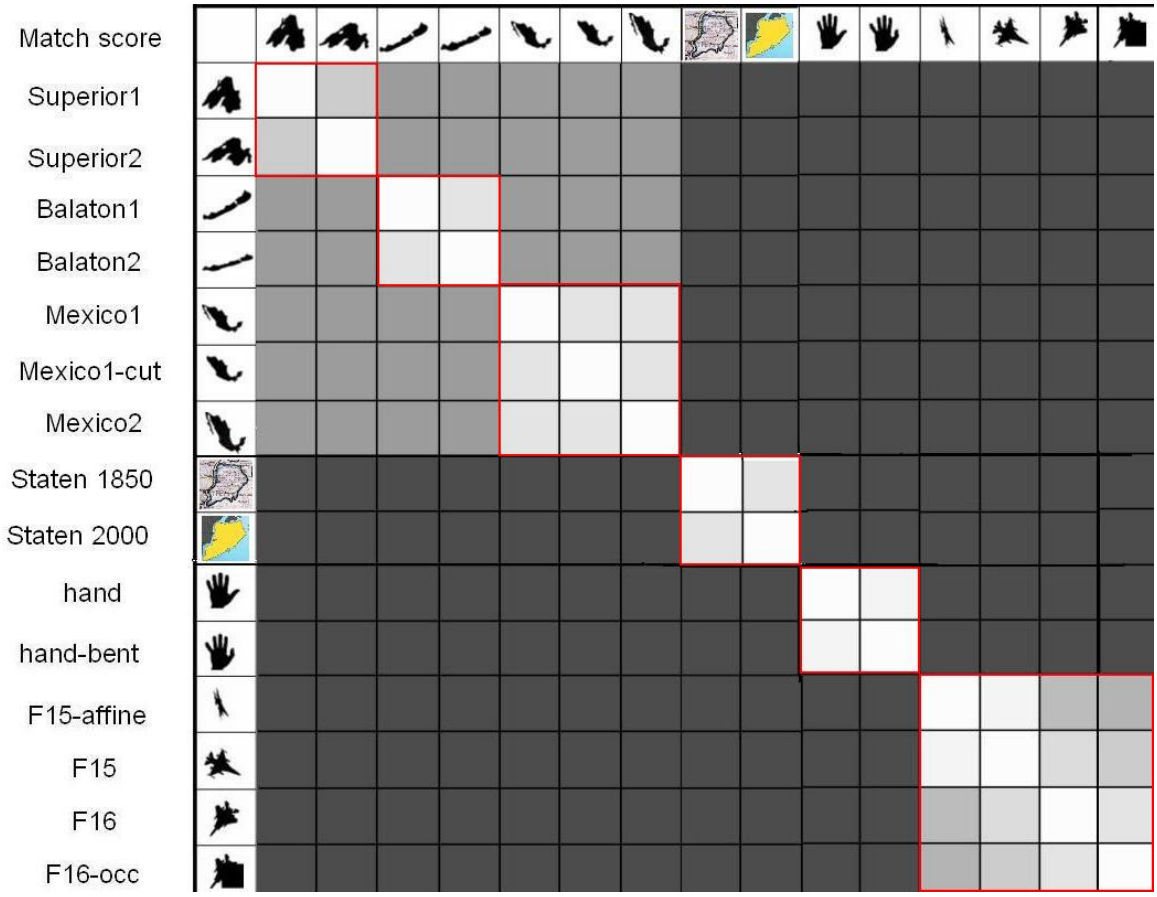


Figure 5: The matching scores for features displayed in the form of a confusion matrix where each feature is matched against all the other features. The light areas represent higher match scores and the red outlines represent the ground truth of the clusters in the datasets.