Analysis of Feature Invariance and Discrimination for Hand Images: Fourier Descriptors versus Moment Invariants

Andre L. C. Barczak, Andrew Gilman, Napoleon H. Reyes and Teo Susnjak
IIMS
Massey University
Albany Campus, Auckland, New Zealand
Email: a.gilman@massey.ac.nz, a.l.barczak@massey.ac.nz

***Abstract*—This papers addresses the issue of feature extraction for the case of gesture recognition. Two important properties are covered in the experiments, invariance to certain transformations and discrimination power. The study of these two characteristics was carried out using Moment Invariants and Fourier Descriptors, specifically for ASL (American Sign Language) images, and for that purpose a new dataset for the ASL images was created. Three types of images - greyscale, binary and contours - were used with the Moment Invariants. Two different Fourier Descriptors were compared, namely complex coordinates and central distances. The results showed trends for the invariance and discrimination powers for both feature sets. The paper shows how to gather information that allow researchers to choose between two competing feature extraction methods before attempting to train a classifier.**

I. Introduction

Gesture recognition is a challenging area in computer vision. There are many methods used to recognise hand posture for simple static gesture sets, e.g. [1] and [2]. Methods that use traditional feature extraction methods, such as Haar-like features ([3] and [4]) or various shape-based features [5], allow researchers to separate the feature extraction from the classification process. Usually artificial neural networks and related methods do not allow for such empirical analysis [6].

For the particular problem at hand, this paper investigates two of the well-known feature extraction algorithms, namely Fourier Descriptors (FD) and geometric Moment Invariants (MI). The primary objective of this work is to establish the strengths and limitations between the two algorithms and possibly devise simple criteria for choosing which approach is suitable for any given scenario. In the literature, moment invariants have been noted to be scaling, translation and rotation invariant but are susceptible to numerical instabilities due to the objects' symmetry and presence of noise [7]. On the other hand, Fourier descriptors have been noted to suffer from inaccuracies due to transformations that amplify small contour deviations. In addition, good object segmentation is a necessary precursor to generating profiles, which in turn produce the Fourier descriptors. A complete comparison between the algorithms should not only look at the relative invariance to translation, scaling and rotation, but also on how the features are distributed when considering the different classes defined for the set of images. In our work, we used the ASL (American Sign Language) gestures, a popular benchmarking set of gestures.

This work also examines how the details of the implementation can influence the performance of the algorithms. For example, Moment Invariants can be extracted in a number of ways. One can segment the object first, then compute the Moment Invariants from the greyscale image directly. Alternatively, Moment Invariants can be calculated from a binarised image, or from an object's contour image. The calculation of moment invariants is also prone to the limited precision of the system's representation of variables (e.g. single versus double precision). Yet another known limitation for Moment Invariants is the symmetry of objects [7].

Fourier descriptors can also be computed in a number of different ways. For example, choosing different shape signatures and contour sub-sampling methods can influence the accuracy of the features subject to the usual geometric transformations. Again, there are sources of inaccuracies that make the theoretical true invariance fail under certain circumstances. In practice, the features have to be extracted by a limited sample of points, which leads to a lot of variability even when contours are taken from transformed images that have been filtered to try to minimise the noise.

The paper is organised as follows. The next section briefly discusses related work, and justifies the methods used in this work. A brief section discusses the image database used for all feature extraction methods. The section after that shows how the Moment Invariants were computed, and the results presented for rotation invariance, scaling invariance, class invariance and discrimination powers. The same sequence of analysis is presented for the Fourier Descriptors. Finally, in the last section the results are discussed and a conclusion is presented.

II. Related Work

Recently, a paper comparing the same features was published and concluded that Fourier descriptors outperform Hu's Moment Invariants [8]. However, the analysis was carried out based on the final results of classification, without enough details on how the features were extracted and how the classifier was trained. We should improve the analysis taking into consideration the fact that we can separate the feature extraction from the classification. Moreover, Hu's moment invariant set has been known to contain redundancies since Flusser's publication in 2000 [9], therefore Flusser's set should have been used instead. Finally, implementation aspects were not discussed in enough details.

Balslev [10] analysed the noise tolerance of Moment Invariants, and concluded that for images with sizes bigger than about 29 pixels, the tolerance to rotation and scaling was reasonably good for moments up to the 4^{th} order. The noise tends to be more pronounced when the images are small, and the combination of rotation and scaling transformations can yield even larger noises.

Yang et al. [5] published a complete survey of different shape extraction techniques. This can be used as a reference for future comparisons of Moment Invariants and Fourier Descriptors with other shape features.

Rodtook and Makhanov [11] analysed rotation invariance of different Moment Invariants (Zernike, Fourier-Mellin and Wavelet moments) for silhouettes of airplanes and musical instruments. They concluded that the transformational noise in digital images can significantly affect the invariance properties of moments, even if theoretically they are invariant to scaling and rotation.

One needs to study the effects of the transformational noise in invariant features before using them as a discriminative feature in a classification system. Moreover, the effects will be influenced by the way in which the features are extracted.

It seems that there is a gap in the literature related to the comparison of these two sets of features. The analysis should be made independently, without the influence of classifiers. Also, the nature of the images or objects studied often influences the properties, especially invariance to certain transformations. In this work, the scope was limited to the study of the invariance characteristics of Moment invariants and Fourier Descriptors for ASL gestures. The invariance was constrained to scaling and rotation. Real hand images were used in the comparison. For completion, slightly different methods were implemented to compute the features.

III. The image database

Images of ASL postures were taken using a 640x480 pixels camera, with controlled illumination with a green background. The segmentation is straightforward because the hand occupies a large portion of the image and the subjects used a wrist cover. A total of 2425 images were used, composed of 36 gestures (from A to Z, and from 0 to 9). The image collection used several repetitions with 5 different illuminations and 5 different subjects. The images were then transformed using a well known tool called Imagemagick. After scaling the images between 0.5 to 2.0 (10 % intervals), and rotating them between -45 to 45 degrees (with 5 degrees intervals), the total number of images is 86530.

Firstly, we analysed the features from the invariance point of view. The images changed by the geometric transformations should yield the same features in theory. In practise, due to the discrete nature of the problem, it is expected that the features will not be truly invariant.

Secondly, we looked at how the features cluster in the hyperspace given the various classes of gestures. This is a good indication of how strongly the features correlate with the particular classes (gestures). In order to analyse the discrimination powers without the use of classifiers, we computed the Euclidean Distance (ED) between the classes' centroids. This simple measure can give us an indication of how the classes are distributed in the hyperspace.

IV. Moment Invariants

For completion, Moment Invariants were computed from three types of images, contour, binary and grey-scale (figure 1). The same contour images were also used for the Fourier descriptors, so the accuracy can be compared fairly.

Fig. 1. Image samples: grey-scale, binary and contour.

The Moment Invariants implementation used Flusser's set [9]. Flusser used a complex number approach to determine a minimum independent set for moments of a certain order. Using his set of moments, we computed moments up to the 4^{th} order, which yields 11 independent Moment Invariants.

We implemented the Moment Invariants extraction using SAT's, a method described earlier in [12]. Only 5 of the original Hu's moments [13] are used , as moments ϕ_2 and ϕ_3 are dependent [9]. The other 6 Moment Invariants are:

$$\psi_6 = 2\{\eta_{11}[(\eta_{30}+\eta_{12})^2-(\eta_{03}+\eta_{21})^2] - (\eta_{20}-\eta_{02})(\eta_{30}+\eta_{12})(\eta_{03}+\eta_{21})\} \quad (1)$$

$$\psi_7 = \eta_{40}+\eta_{04}+2\eta_{22} \quad (2)$$

$$\psi_8 = (\eta_{40}-\eta_{04})[(\eta_{30}+\eta_{12})^2-(\eta_{03}+\eta_{21})^2] + 4(\eta_{31}+\eta_{13})(\eta_{30}+\eta_{12})(\eta_{03}+\eta_{21}) \quad (3)$$

$$\psi_9 = 2(\eta_{40}-\eta_{04})(\eta_{30}+\eta_{12})(\eta_{21}+\eta_{03}) - 2(\eta_{31}+\eta_{13})[(\eta_{30}+\eta_{12})^2-(\eta_{21}+\eta_{03})^2] \quad (4)$$

$$
\begin{aligned}
\psi_{10} = (\eta_{40} - 6\eta_{22} + \eta_{04}) \\
\{[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]^2 - \\
4(\eta_{30} + \eta_{12})^2(\eta_{03} + \eta_{21})^2\} \\
+ 16(\eta_{31} - \eta_{13})(\eta_{30} + \eta_{12})(\eta_{03} \\
+ \eta_{21})[(\eta_{30} + \eta_{12})^2 - (\eta_{03} + \eta_{21})^2]
\end{aligned} \quad (5)
$$

$$
\begin{aligned}
\psi_{11} = 4(\eta_{40} - 6\eta_{22} + \eta_{04})(\eta_{30}+ \\
\eta_{12})(\eta_{03} + \eta_{21})[(\eta_{30} + \eta_{12})^2 - (\eta_{03} + \eta_{21})^2] \\
- 4(\eta_{31} - \eta_{13})\{[(\eta_{30} + \eta_{12})^2 - (\eta_{03} + \eta_{21})^2]^2 \\
- 4(\eta_{30} + \eta_{12})^2(\eta_{03} + \eta_{21})^2\}
\end{aligned} \quad (6)
$$

Where η_{pq} are the normalised central moments. The equivalence of Hu's moments to Flusser's is as follows: $\psi_1 = \phi_1$, $\psi_2 = \phi_4$, $\psi_3 = \phi_5$, $\psi_4 = \phi_6$, $\psi_5 = \phi_7$.

A. Invariance to scaling and rotation in practice

Figure 2 shows the invariance to rotation for the 11 moment invariants. To create the figure, all the 2425 images were rotated from -45 to 45 degrees, with 5 degrees intervals. The moments were computed for each rotated image and the results were coalesced into mean and standard deviation. The relative standard deviation was computed for each one of the 2425 cases. The figure shows the mean of the relative standard deviation, a measure that shows how true the invariance is when considering resampling "noise" and implementation issues such as precision.

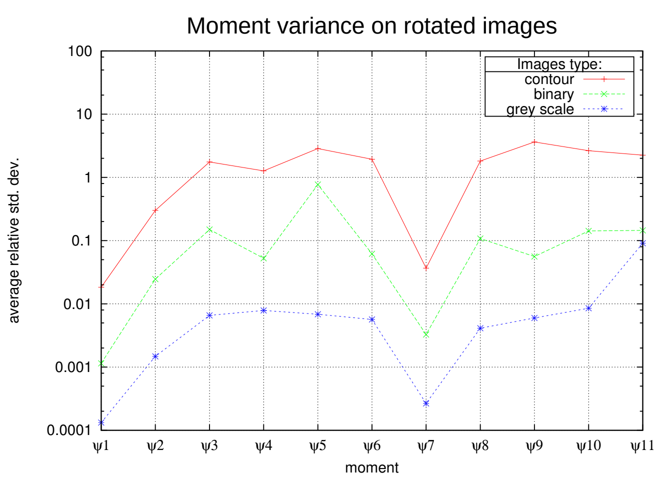

Fig. 2. Rotation invariance for the 11 moments. Mean Relative standard deviation for grey-scale, binary and contour images.

As expected, the best invariance is achieved when using the grey-scale images. The binary images present a slightly lower invariance properties. The contours, due to the more limited number of pixels that are different than zero, has the worst invariance to rotation.

Figure 3 shows the invariance to scaling for the 11 moment invariants. The figure was created in the same way as 2. All the 2425 images were scaled from 50% up to 200%, with 10% intervals. The mean of the relative standard deviation is shown.

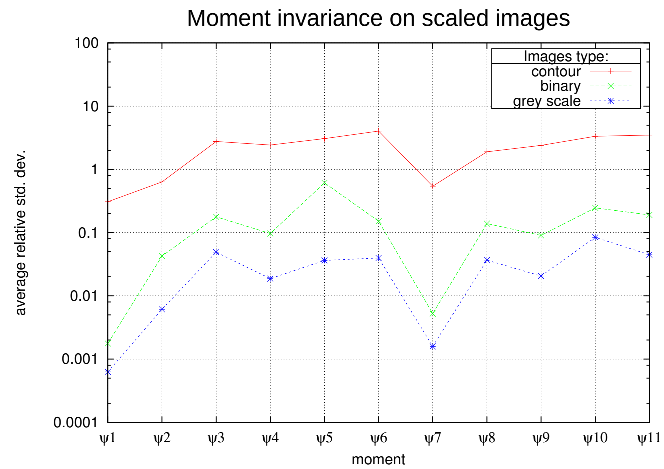

Fig. 3. Scaling invariance for the 11 moments. Mean Relative standard deviation for grey-scale, binary and contour images.

The same trend was observed with scaling, with the best invariance characteristics for the grey-scale images and the worst for the contour images. These results did not surprise because the Moment Invariants should be sensitive to noise. With a smaller number of points to smooth the noise, it was expected that with the contours the invariance would be worse than compared to grey-scale images. The geometric transformations may displace pixels in the contour images, making the relative distance between the pixels and the centre of mass of the contour to vary more wildly, and therefore yielding slightly different moments.

B. Invariance per class

We separated classes from figures 2 and 3 in order to analyse the effect of different images on the geometric invariance. It should be stressed that the next figures do not show the standard deviation per class, but rather the average of the relative standard deviation for *each image* of that class.

The best geometric invariance per class was for class '3' (figure 4). For class '3' the relative standard deviation was above 0.1 for ψ_{11} and below for all other ψ_n. The worst result for the invariance was for class '6' (figure 5). For class '6', only two of the moments had the relative standard deviation below the 0.1 line.

V. FOURIER DESCRIPTORS

Fourier descriptors are a form of harmonic analysis of the shape signature, which is a function derived from the shape's boundary coordinates. They are computed by taking the Fourier transform of the shape signature and normalising the resulting coefficients in a way that would result in a set of rotation-, translation- and scale-invariant descriptors.

Fig. 4. Invariance characteristics for class 3. In this plot all rotations and scaling were used for all original '3' images.

Fig. 5. Invariance characteristics for class 6. In this plot all rotations and scaling were used for all original '6' images.

This normalisation is signature dependent, for example some signatures (e.g. periodic cumulative angular function) are already invariant to the above transformation and do not require normalisation. A number of shape signatures have been described in the literature: complex coordinates, curvature function, cumulative angles, central distance, etc. The choice of the shape signature has an effect on the resulting Fourier descriptors and may have effect on their invariance, as well as discrimination power. Zhang and Lu [14] have shown that central distance outperforms other types of signatures; however, this was shown in the context of image retrieval. Hence, we utilised the central difference shape signature, as well as another popular signature, based on complex coordinates.

The central distance shape signature is computed from the set of object's contour's coordinates (x_n, y_n) by computing the distance between each coordinate and the object's centroid. The complex coordinates-based shape signature is computed by transforming the 2D coordinates of each contour point (x_n, y_n) to a 1D complex coordinate $(x_n + iy_n)$. The resulting shape signatures generated from our test images generally contained between 1500 and 3000 samples. These were resampled at 256 locations uniformly spread around the contour using linear interpolation. The discrete Fourier transform was then computed and the resulting coefficients were normalised in the following way. For the central distance shape signature, the signature itself is rotation and translation invariant; hence, the Fourier transform coefficient were normalised by the magnitude of the DC component to make the FDs scale invariant. Since the shape signature is a sequence of real numbers, applying the Fourier transform results in a set of conjugate symmetric coefficients. We found it sufficient to keep the 9 lowest frequencies. in the case of complex coordinates signature shape, the DC component was discarded to make it translation invariant, the remaining coefficients were normalised by the magnitude of the first positive coefficient to make it scale invariant, and the phase information was discarded to make it rotation invariant. As this shape signature is a sequence of complex numbers, we have kept coefficients $[-5, -4, -3, -2, -1, 3, 4, 5]$ as the nine Fourier descriptors.

Figure 6 shows the effect of rotation on FDs computed using both shape signatures. Both of these show similar degree of invariance to each other and a higher degree of invariance in comparison to the moment invariants computed from the contours. The Fourier Descriptors' amount of invariance is more similar to that of the binary moment invariants. Scaling of the images produces much the same results for the FDs. The amount of variability in these results may seem high in comparison to the theoretic invariance of the Fourier Descriptors; however, when compared to between-the-images variability within the same class, it appears not as significant. Figure 7 shows a plot of relative standard deviation computed over the original images (no scaling or rotation) of the same class and averaged over all 36 classes and a plot of relative standard deviation computed as before, but this time with both scaling and rotation for comparison.

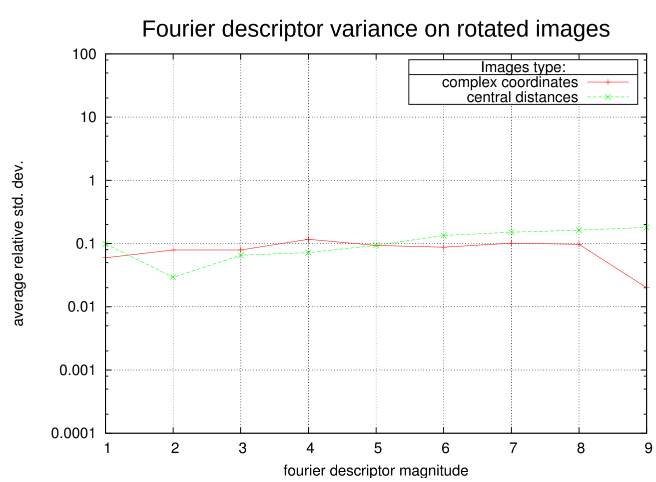

Fig. 6. Rotation invariance for the FDs.

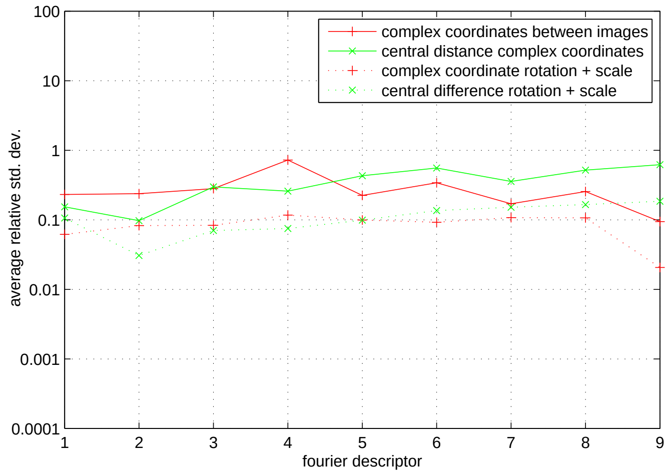

Fig. 7. This plot shows total scaling and rotation standard deviation for the FDs is much smaller than the average standard deviation within each class.

VI. Characteristics for FDs and MIs

Some gestures could be merged into a single class, either because they are too similar (e.g., '0' and 'O') or because they are slightly rotated in relation to each other (the cases of 'I' with 'J' and 'Z' with '1'). Other gestures have subtle differences that may show in one feature set, but not in others. E.g., gesture 'T' and 'E' differ only by the position of the thumb, and there may be very little difference between the image contours.

The Euclidean distances are plotted between the 36 classes in figure 9. Although this result cannot be in itself used to determine the discrimination power of a set of features, it can be used to gather data about how the various classes are distributed. If the distance is larger than the average, those two classes will be easily separable, even using a simple classifier based on thresholds. If the distance is small, nothing can be concluded, as those classes might still be separable when using a more sophisticated classification algorithm.

We noticed three important trends. Firstly, the fact that the feature set is invariant does not mean that it produces separable classes. It may be the case that the feature is robust to a certain transformation, but not discriminative enough for the defined classes. Secondly, the invariance to scaling and rotation for a single hand sample is so small, when compared to other hand samples of the same class, that for practical purposes one can consider that both, the Moment Invariants and the Fourier Descriptors are truly invariant. The exception would be the Moment Invariants computed from the contours of the images, as this had the largest relative standard deviation of all 5 feature sets. And thirdly, using different shape signatures for FDs may result in different discrimination powers. Building a classifier using multiple shape signatures may prove beneficial.

VII. Conclusion

This paper compared two feature extraction methods, Moment Invariants against Fourier Descriptors. Despite the fact that the theoretical invariance is not true for digital images and for the limited precision of the implementation, little effect was noticed in terms of the standard deviation for rotation and scaling of the same hand image sample.

Although the discrimination cannot be fully determined from the simple analysis carried out by computing the Euclidean distances, it gives some insights on how difficult (or how easy) it is to train a classifier between those classes. For the cases where the distances are larger, it might even be straightforward to find simple thresholds to separate the classes. For the cases where the distances are smaller, thresholds will not suffice. Originally we intended to show the discrimination power using a metric based on Mahalanobis distance; however, computation of the inverse of Moment Invariants' covariance matrix proved to be rather difficult with 'double' precision, because of severe ill-conditioning.

Also, it shows which classes are likely to be merged. We notice that certain images would yield similar contours, because the correspondent ASL gesture only differs by the position of the thumb. The Fourier Descriptors or the Moment Invariants computed from the contours would not be able to differentiate such classes.

For future work we intend to carry out further discrimination analysis using Mahalanobis distance. We also intend to include other feature extraction methods, such as Zernike moments.

References

[1] M. Kolsch and M. Turk, "Fast 2d hand tracking with flocks of features and multi-cue integration," in *RealTimeHCI04*, 2004, pp. 158–164.

[2] S. Marcel, O. Bernier, J.-E. Viallet, and D. Collobert, "Hand gesture recognition using input/ouput hidden markov models," in *Proceedings of the 4th International Conference on Automatic Face and Gesture Recognition (AFGR)*, 2000.

[3] A. L. C. Barczak, F. Dadgostar, and M. J. Johnson, "Real-time hand tracking using the viola and jones method," in *SIP 2005*, Honolulu, HI, 2005, pp. 336–341.

[4] M. Kolsch and M. Turk, "Analysis of rotational robustness of hand detection with a viola-jones detector," in *ICPR04*, 2004, pp. III: 107–110.

[5] M. Yang, K. Kpalma, and J. Ronsin, *Pattern Recognition Techniques, Technology and Applications*. InTech, 2008, ch. A Survey of Shape Feature Extraction Techniques, pp. 43–90, iSBN 978-953-7619-24-4.

[6] S. Mitra and T. Acharya, "Gesture recognition: a survey," *IEEE Trans. on SMC part C*, vol. 37, pp. 311–324, 2007.

[7] J. Flusser, B. Zitova, and T. Suk, *Moments and Moment Invariants in Pattern Recognition*. Wiley Publishing, 2009.

[8] S. Conseil, S. Bourennane, and L. Martin, "Comparison of fourier descriptors and hu moments for hand posture recognition," in *European Signal Processing Conference (EUSIPCO)*, 2007.

[9] J. Flusser, "On the independence of rotation moment invariants," *Pattern Recognition*, vol. 33, pp. 1405–1410, 2000.

[10] I. Balslev, "Noise tolerance of moment invariants in pattern recognition," *Pattern Recogn. Lett.*, vol. 19, pp. 1183–1189, November 1998.

[11] S. Rodtook and S. Makhanov, "Numerical experiments on the accuracy of rotation moments invariants," *Image and Vision Computing*, vol. 23, no. 6, pp. 577–586, June 2005.

[12] A. L. C. Barczak and M. J. Johnson, "A new rapid feature extraction method for computer vision based on moments," in *International Conference in Image and Vision Computing NZ (IVCNZ 2006)*, Auckland, NZ, November 2006, pp. 395–400.

[13] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory*, vol. 8, pp. 179–187, 1962.

[14] D. Zhang and G. Lu, "A comparative study of fourier descriptors for shape representation and retrieval," in *Proceedings of the Fifth Asian Conference on Computer Vision (ACCV)*, January 2002, pp. 646–651.

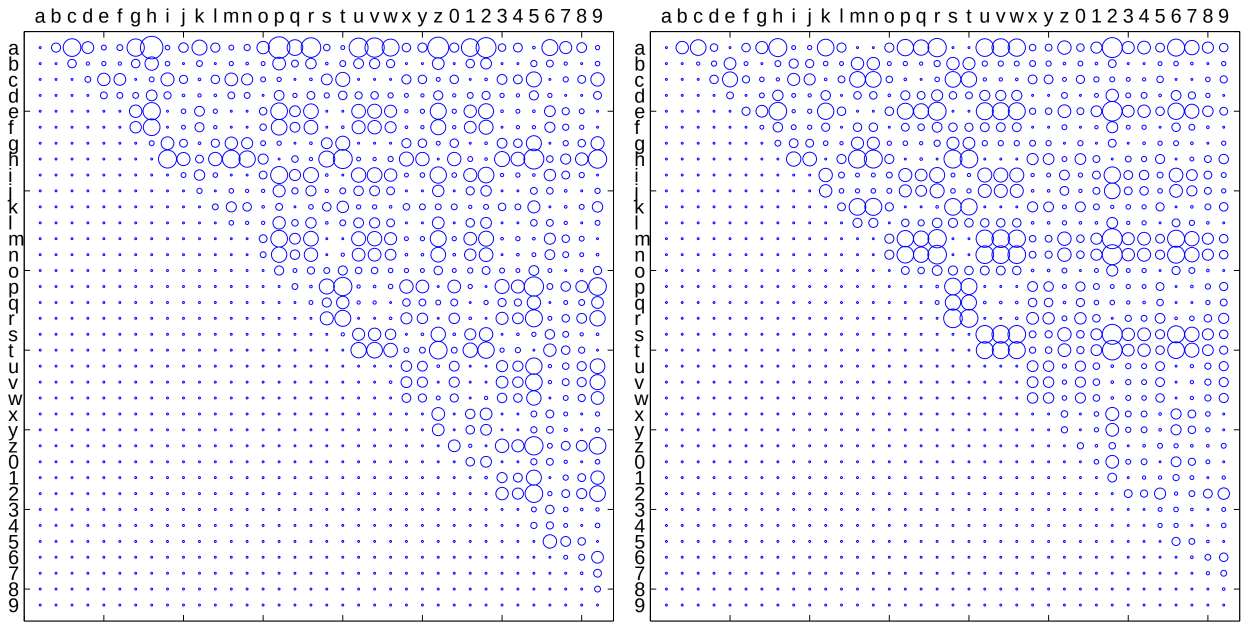

Fig. 8. Euclidean distances between classes for the Moment Invariants using grey (left) and binary (right) images. Circle diameters are proportional to the Euclidean distance. Because these are arbitrarily scaled, comparison is only valid within each figure.

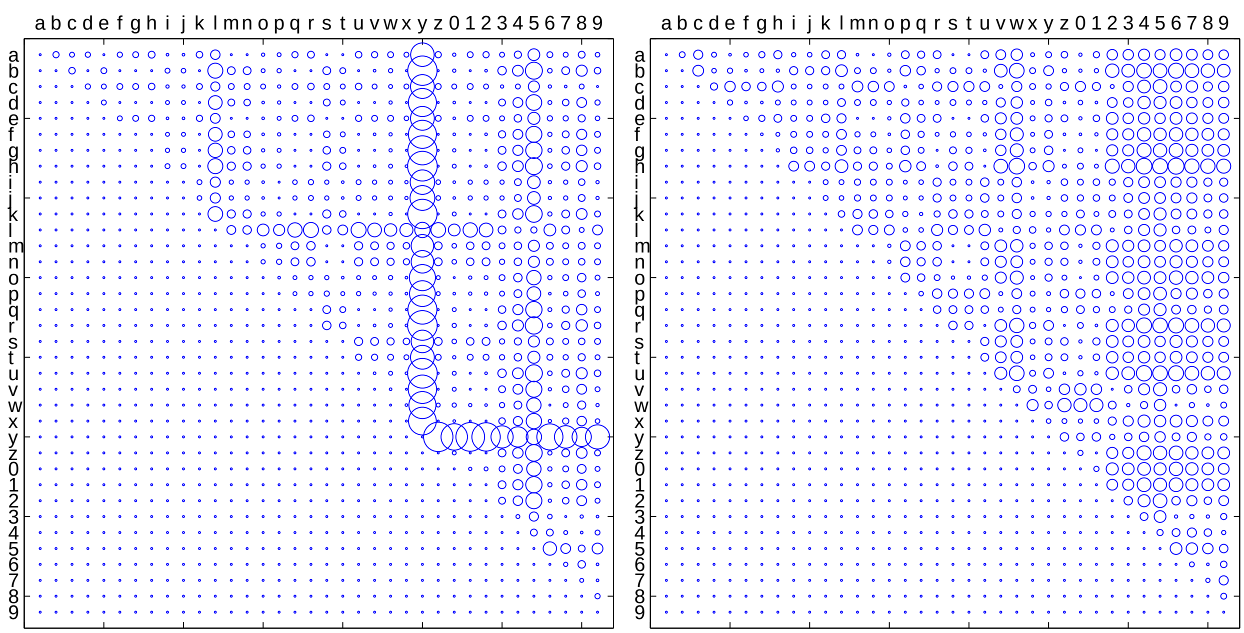

Fig. 9. Euclidean distances between classes for the Fourier Descriptors with complex coordinates (left) and with central distance (right). Circle diameters are proportional to the Euclidean distance. Because these are arbitrarily scaled, comparison is only valid within each figure.