Probabilistic Models for 2D Active Shape Recognition using Fourier Descriptors and Mutual Information

Natasha Govender
MIAS (CSIR)
South Africa
ngovender@csir.co.za

Jonathan Warrell
MIAS (CSIR)
South Africa
jwarrell@csir.co.za

Philip Torr
University of Oxford
United Kingdom
philip.torr@eng.ox.ac.uk

Fred Nicolls
UCT
South Africa
fred.nicolls@uct.ac.za

I. ABSTRACT

Shape recognition is essential for robots to perform tasks in both human and industrial environments. Many algorithms have been developed for shape recognition with varying results. However, few of the proposed methods actively look for additional information to improve the initial shape recognition results. We propose an initial system which performs shape recognition using the euclidean distances of Fourier descriptors. To improve upon these results we build multinomial and Gaussian probabilistic models using the extracted Fourier descriptors and show how actively looking for cues using mutual information can improve the overall results. These probabilistic models achieves excellent results while significantly improving on the initial system.

II. INTRODUCTION

The use of robots in industrial and household environments is steadily on the increase. A huge part of robots functioning in these environments is recognising objects. Textural and feature-based approaches are often not appropriate for these types of applications because parts may contain little or no distinctive features other than boundary shape. Environments may not have consistent lighting conditions which can adversely affect these approaches. We use Fourier descriptors to extract boundary information to perform shape recognition. Polar co-ordinates are selected as our shape signature for the descriptors.

Two shape recognition systems are proposed in this paper. The first system uses the euclidean distance between the descriptors to determine the class of each shape. We chose the toy problem of a child's shape puzzle because the shapes were relatively arbitrary and certain shapes were also similar. This was selected to determine the robustness of the system. The second system aims to improve on the shape recognition system that uses just euclidean distance. For this system close-up images of the shapes were captured as input to the system. Here we propose using the Fourier descriptors extracted in a probabilistic manner. Multinomial and Gaussian distributions are built using the Fourier descriptors. We then include an active vision component in the form of mutual information. When the system is determining the correct object sequence, mutual information provides the system with the ability to select the position in the sequence which it is most unsure about.

In our experiments we show that using the probabilistic models with mutual information outperforms both using just the Fourier descriptors as well as the probabilistic models without mutual information.

The structure of the paper is as follows: The next section describes related work. Section IV elaborates on the problem and Section V discusses the Fourier descriptors. Section VI describes the polar co-ordinates and the results for the first system. A complete description of the probability models is presented in Sections VII. Section VIII presents further experimental results and discussion. The conclusion is discussed in Sections IX.

III. RELATED WORK

Various shape representation methods, or shape descriptors, exist in the literature. These methods can be classified into two categories: region based versus contour based. In region based techniques, all the pixels within a shape are taken into account to obtain the shape representation [19],[18]. Contour based shape representation exploits shape boundary information.

Fourier descriptors are contour based and capture global shape features in the first few low frequency terms, while higher frequency terms capture finer features of the shape. Wavelet descriptors can also be used to model shape and have an advantage over Fourier descriptors in that they maintain the ability to localise a specific artifact in the frequency and spatial domains [20]. However, wavelet descriptors are impractical for higher dimensional feature matching [22]. The Fourier Descriptor method can also be easily normalized.

Fourier descriptors are a widely used, all purpose shape description and recognition technique. They have been used in a variety of fields over the years, including commerce, medical, space exploration, and technical sectors. In the field of computer vision, Fourier descriptors have been used for human silhouette recognition [12] for surveillance systems, content based image retrieval [24],[13], shape analysis [23],[14], character recognition [15],[16] and shape classification [11]. In these methods, different shape signatures have been exploited to obtain the Fourier descriptors. These include central distance, complex coordinates, curvature function, and

Fig. 1. The board with the shapes removed

Fig. 2. The shapes that need to be recognised

cumulative angles [7]. Most systems use complex coordinates to model the shape boundary [21],[12] but we use polar coordinates because in our experiments this method produces more accurate results.

There have been a number of probabilistic based models for shape recognition proposed such as using Procrustean models[10], probability density functions [1], geometric features [5] and generative models [3] to name a few. None of these methods use Fourier descriptors as their input parameter to the shape recognition system. In addition none of these methods use active vision by incorporating mutual information to improve their initial results. Mutual information was introduced by [9] as a viewpoint selection mechanism for active vision, which has been subsequently used/proposed by [6][2][17]. As noted, this can be expensive to calculate, and requires the collection of extensive statistics at training time, although as [9] discussed, it provides the optimal strategy provided the underlying models are correct. Using mutual information also makes it easy to incorporate probabilistic assumptions to assist with active information selection. Our framework follows that of [9] and [4] in terms of the general Bayesian form of our updates and we use a sampling scheme to make the mutual information calculations tractable.

IV. The problem

A board containing cut out cartoons of different animals was used in the experiments. The shapes were removed from the board and placed on table. Figure 1 and Figure 2 display the board and the shapes used in the experiments. For each shape 20 close-up images were also captured. Information from these images are used in the probabilistic models.

V. Fourier Descriptors

These images are initially converted to grayscale and then into binary images. The method presented in [8] is used to detect and label the various objects boundaries. Each shape is then segmented from the image and stored. The same procedure is followed for all captured images.

The set of (x,y) boundary coordinates for each shape is converted to polar coordinates. The Fast Fourier Transform (FFT) is then taken for each set of values. The formula used is described in equation 1. Rotation and changes in the starting point only affect the phase of the descriptor. All the phase information can be removed by taking the absolute values of the descriptor elements. It has been shown that the low frequency components of the Fourier Transform are sufficient for shape recognition[12][24] and thus the entire transform does not need to be used. We found that using the first 15 Fourier co-efficients (excluding the very first component F(0)) provided sufficient discriminatory information to model a shape. F(0) is the lowest frequency term and is the only component in the Fourier descriptor that is dependent on the actual location of the shape. By ignoring the first component, it becomes translation invariant. F(0) tells us nothing about the shape; only mean position. The Fourier Descriptor is then normalized to remove any scaling effects. The FFT of the shape is described as:

$$\mathcal{F}(i) = \mathrm{FFT}\{\mathbf{r}\}_i, \tag{1}$$

where

$$r(s) = \sqrt{(x(s) - x_c)^2 + (y(s) - y_c)^2}, \tag{2}$$

x_c, y_c are the shape centroids, x(s), y(s) are the boundary coordinates of the s'th point, $\mathbf{r}$ is the vector of radii, and $\mathrm{FFT}\{.\}$ denotes the discrete fast fourier transform.

VI. Recognition using Polar co-ordinates

Each shape boundary is then matched to the shapes extracted from the board using euclidean distance. Since the energy in the Fourier components decreases sequentially, we artificially boost the contribution of each component. We have found that as the number of components increases, increasing the value used to boost the components works best when calculating the euclidean distance. The shape on the board

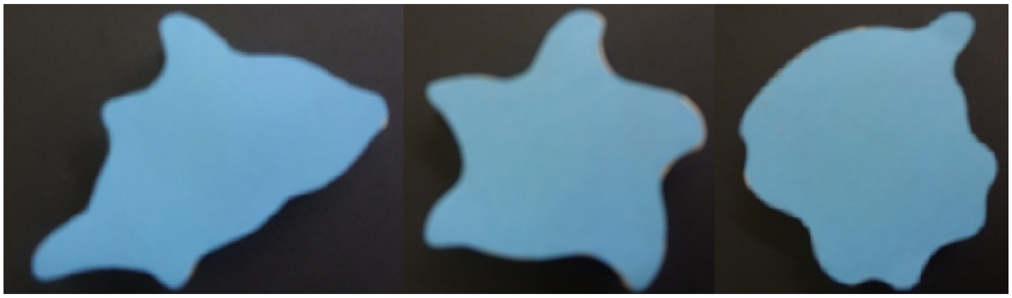

Fig. 3. Close-up images of the shapes

TABLE I
RECOGNITION RESULTS

Shape	Complex Coordinate Method	Polar Coordinate Method
Whale	yes	yes
Seal	no	yes
Fish	no	no
Crab	no	yes
Dolphin	yes	yes
Mussel	no	yes
Snail	no	yes
Octopus	no	yes
Star Fish	yes	yes
Tortoise	no	no

with the smallest euclidean distance to the shape on the table is considered to be the match.

A. Results

The shapes used in the experiments are in the form of animals which include a tortoise, whale, seal, dolphin, fish, crab and so on as seen in Figures 1-3. There are ten shapes in total. Many shape recognition systems use complex coordinates to model the shape boundary but we opted to use polar coordinates because in our experiments this method produces more accurate results. Table I shows the results obtained from both methods.

The complex coordinate system recognises four shapes correctly while the polar coordinate system correctly recognises eight out of the ten shapes. It incorrectly identifies the fish and the tortoise shapes. Figure 5 displays the first fifteen dimensions of the Fourier descriptors extracted for the fish shape from the board and the tortoise shape and the fish shape extracted from the cut out pieces. The system incorrectly recognises the fish shape as the tortoise as this produces the smallest euclidean distance. Looking at the Fourier descriptor components, we can see that the tortoise and fish descriptors are fairly similar (produces similar peaks) and could be easily confused. The fish shape actually has the second smallest euclidean distance. A similar situation occurs when trying to recognise the tortoise shape.

VII. PROBABILITY MODELS

A. Multinomial distribution

For the multinomial distribution we extracted the Fourier Descriptors from the dataset containing the close-up images of the shapes. The 20 close-up images for each shape were split into two sets containing 14 images for training and 6 images for testing. The training set was further split into two sets containing 7 images each. One was used for training and the other as a validation set. This was done to determine a quasi-ground truth histogram distribution which can be used for testing. The euclidean distance was calculated between every image in the training and validation set. This process was carried out 10 times. The minimum distance value was then determined which identified which object class the system thought each image belonged to. A distribution histogram for each image class was then calculated. A bias was placed at the correct class to provide the system with a reliable ground truth distribution.

Let N be the number of shapes. D the dimension of the Fourier transforms (in this case we used 15 descriptors). Let x represent as possible permutation x $\in \mathcal{P} \subset \{1..N\}^N$. Here, $\mathcal{P}$ denotes all permutations of N objects, hence is the subset of $\{1..N\}^N$ which contains no repetitions. Observation O for the close-up shapes takes the form $O = [O_1, O_2, ...O_N]$, where $O_n \in \mathcal{V} \subset (\mathbb{Z}^+)^N$ which are the counts in the histogram for test images in class n derived above. Let $\theta = [\theta_1, \theta_2...\theta_N]$ represent the parameters of the distributions for each of the object classes. For the multinomial model we use $\theta_n = \alpha_n$, where α_n is the multinomial mean vector set using the counts from the training images. For initialisation a noisy prior is selected for the board. This is done to incorporate the effects/noise that may occur due to illumination changes and the camera or lens used. The prior for the board can be represented by $\pi(x)$. The probability of a permutation given all observations is described as:

$$P(x|O) = \frac{P(O|x) \cdot \pi(x)}{P(O)} \propto \pi(x) \prod_n P'(O_n|\theta_{x_n}), \qquad (3)$$

where $P'(O_n|\theta_{x_n}) = \text{Mult}(O_n|\alpha_{x_n}) = (M!/\prod_m O_{nm}!) \prod_n \alpha_{x_n}^{O_{nm}}$ for the multinomial likelihood, where m ranges across the histogram bins, and M is the number of test images per class.

Bayes theorem can be used to update the probability after each new individual observation. This is given by:

$$\begin{aligned} P_0(x) &= \pi(x) \\ P_t(x|O_1..O_t) &\propto P'(O_{n(t)}|\theta_{x_{n(t)}})P_{t-1}(x|O_1,..O_{t-1}), \end{aligned} \qquad (4)$$

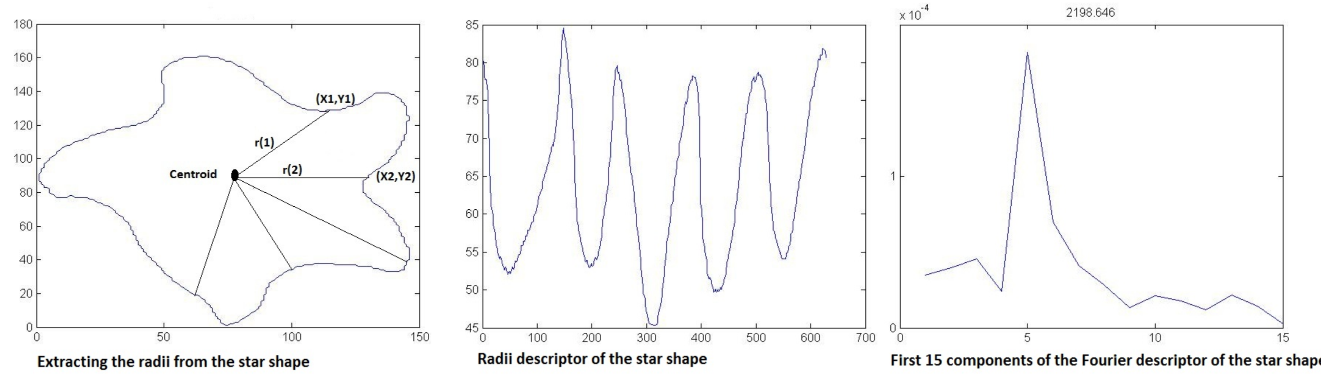

Fig. 4. Boundary of a shape depicting the Fourier descriptors converted to polar co-ordinates

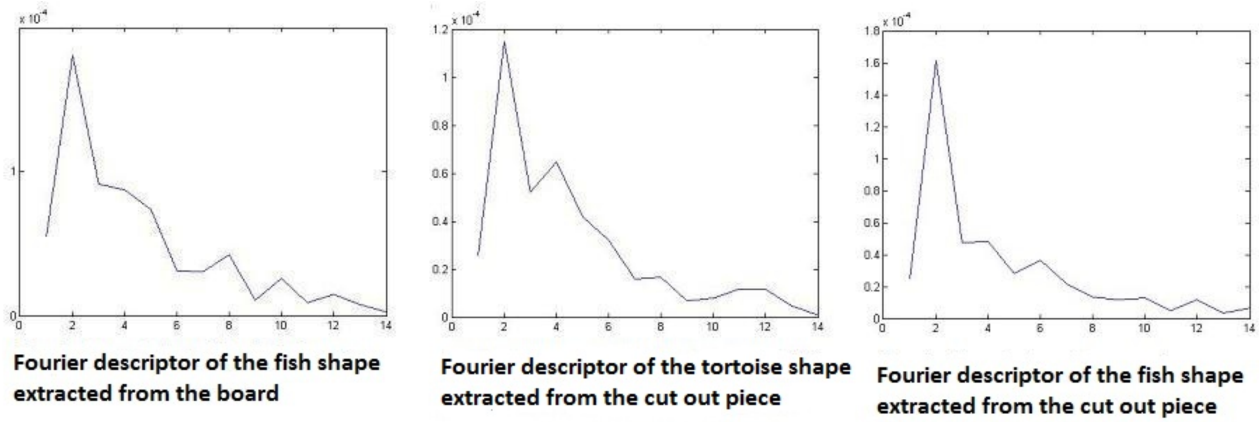

Fig. 5. Fourier Descriptors

where n(t) is the index of the observation seen at time t.

Mutual information(MI) assists in the selection of the position to look at since there can be no repetitions. Once the system is fairly certain of the position of a class in the permutation, mutual information can assist in deciding which position to look at next i.e. which is the most uncertain. Randomly selecting the next position to look at does not not take this information into account.The Mutual Information selection rule is as follows:

$$n(t+1) = \text{argmax}_{n \neq n(1)..n(t)} \, \text{MI}(O_n; x). \tag{5}$$

Mutual information values increases with uncertainty. In this equation we want to select the position in the permutation with the most uncertainty for a given observation.

We can rewrite the above equation in terms of the conditional entropy as follows:

$$\text{MI}(O_n; x) = H(x) - H(x|O_n), \tag{6}$$

where H(.) represents the Shannon entropy and $H(.|.)$ represents the conditional entropy. We need to minimise the conditional entropy. This is described as:

$$n(t+1) = \text{argmin}_{n \neq n(1)..n(t)} \, H(x|O_n). \tag{7}$$

The conditional entropy can be written as:

$$\begin{aligned} H(x|O_n) \quad = \quad & - \sum_{O_n \in \mathcal{V}} P_t(O_n) [\sum_{x' \in \mathcal{P}} P(x'|O_n, O_{n(1)}, ...O_{n(t)}) \cdot \\ & \log(P(x'|O_n, O_{n(1)}, ...O_{n(t)}))]. \end{aligned} \tag{8}$$

To evaluate $P_t(O_n)$, we introduce mixing coefficients β

$$\beta_m = \sum_{(x|x_n = m)} P_t(x), \tag{9}$$

for m = 1..N, which weight the likelihoods for each class. This gives us

$$P_t(O_n) = \sum_m \beta_m \cdot P'(O_n|\theta_m). \tag{10}$$

To avoid exhaustively summing across $\mathcal{V}$ in equation 8, we can consider the conditional entropy as the expectation across $P_t(O_n)$ and approximately evaluate the sum by sampling from this distribution.

$$\begin{aligned} H(x|O_n) &= E_{o\sim P_t(O_n)}[H(x|o)] \\ &\approx \frac{1}{n}\sum_{o_i} H(x|o_i) \\ &= -\frac{1}{n}\sum_{o_i}[\sum_{x'\in\mathcal{P}} P(x'|o_i, O_{n(1)}, ...O_{n(t)}) \cdot \\ &\quad \log(P(x'|o_i, O_{n(1)}, ...O_{n(t)}))], \end{aligned} \tag{11}$$

where E denotes the expectation. o_i in equation 10 represents the samples drawn from the mixture distribution described in equation 10 where i ranges from 1 to n number of samples.

B. Gaussian Distribution

The image set was treated in the same manner as used in the multinomial distribution. The training images were used to learn a Gaussian distribution for each class, $\theta_n = (\mu_n, \sigma_n)$. For σ_n we used a diagonal covariance matrix. For each observation O_n we included all test images $O_n = [O_{n1}...O_{nM}]$, where M is the number of test images per class. The feature space is $\mathcal{V} = (\mathbb{R}^+)^{DM}$ since we have one D dimensional Fourier descriptor for each image. For the likelihood in equation 10 , we used joint likelihood of these observations.

$$P'(O_n|\theta) = P'(O_n|\mu, \sigma) = \prod_{i=1..M} \mathcal{N}(O_{ni}|\mu, \sigma), \tag{12}$$

where $\mathcal{N}$ represents the Gaussian distribution and O_{ni} is the i'th descriptor of observation n.

Since feature space is now continuous the summation in equation 8 changes to an integral. We can use the sampling technique to approximate this integral as in equation 11.

$$\begin{aligned} H(x|O_n) &= \int_{\mathcal{V}} H(x|o)P_t(o)do \\ &= E_{o\sim P_t(O_n)}[H(x|o)] \\ &\approx -\frac{1}{n}\sum_{o_i}[\sum_{x'\in\mathcal{P}} P(x'|o_i, O_{n(1)}, ...O_{n(t)}) \cdot \\ &\quad \log(P(x'|o_i, O_{n(1)}, ...O_{n(t)}))]. \end{aligned} \tag{13}$$

The sampling distribution used here is the same as described in equation 10, with $P'(O_n|\theta)$ as in equation 12.

VIII. Experiments

The sequence on the board was used as the ground truth. Noise was added to the initial models to take into account possible illumination changes and noise introduced by the camera. Each simulation was run 100 times with a different split of the training and the testing images each time. For both models, we want to identify the correct sequence. Once the system is fairly certain about the object at a specific position mutual information allows us to select the next position to look at which the system is most unsure about. In the random case this position is randomly selected. For the shape puzzle the initial model for board was very good so we introduced an artificial flipping method where two object positions would be flipped at random for 20% of the objects. The reason for doing this was we wanted to demonstrate the effectiveness of using mutual information when the initial guesses are not very accurate. We ran simulations with restricted numbers of objects ranging from 4 to 10, and display the results when only 7 and 10 objects are used. Since we found it was computationally expensive to go through all the possible combinations when using 9 or 10 objects, we introduced the sampling method as discussed to reduce this complexity.

TABLE II
Timings for single position update

Method	# Objects	Random (s)	MI (s)
Multinomial	7	0.003	0.074
	10	0.155	33.916
Guassian	7	0.013	0.090
	10	0.173	33.651

The multinomial distribution is initialised using the class histogram calculated at the start.

The probability after looking at 7 and 10 view are 92% and 94% respectively. This provides better results than just using the euclidean distance of the Fourier descriptors as show in Table 1. Also as shown in Figure 6, MI information outperforms randomly selecting the next position to visit. The figure shows how the probability of the correct permutation changes with the number of views seen, and the percentage of correct objects when the permutation with the highest probability is chosen at each number of viewpoints.

In the Gaussian case, the class histogram is not used. Instead we use a covariance matrix to calculate the likelihoods as explained in Section VII-B. The probability after looking at 7 and 10 view are 99.2% and 99.8% respectively. This provides much better results than just using the euclidean distance of the Fourier descriptors and also outperforms the multinomial distribution.

In Table II the average timings are given for choosing the position to view next and updating the current distribution after making an observation. For the mutual information, we used 20 samples per position. The timings are similar for both the multinomial and Guassian distributions. As shown, the mutual information increases the time taken over random selection, although this could be reduced by using fewer samples while trading off accuracy.

IX. Conclusion

We have presented a system which extracts Fourier descriptors from ten different animal shapes to be used for shape recognition. Initially recognition was performed using the euclidean distance between the shapes. This resulted is an accuracy of 80% with the system confusing the fish and the tortoise shapes. We then set about using the Fourier descriptors from the shapes in probability models. We showed that using mutual information to actively select the next most uncertain position in the sequence provides better results than randomly selecting the next position. Both the models correctly identify all the objects. This paper has shown that using probability models for shape recognition, incorporating information about

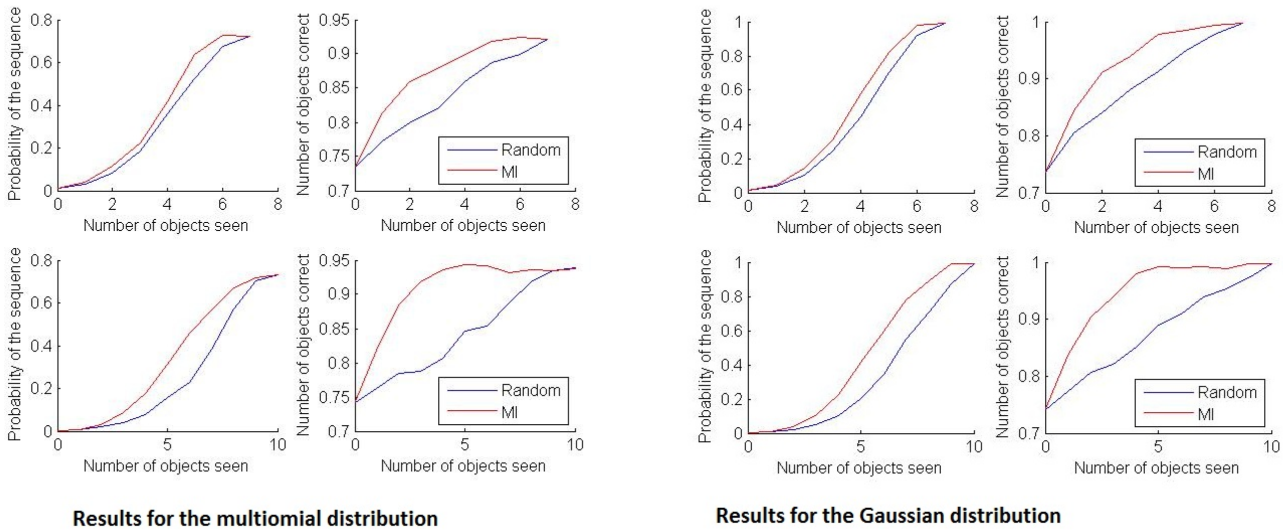

Fig. 6. Results for the multinomial and Guassian distributions using sequences of 7 and 10 objects averaged over 100 splits of data

the current state of the system (MI) and actively selecting which uncertain position to look at produces excellent results.

REFERENCES

[1] C.B Akgul, B.Sankur, Y. Yemez, and F.Schmitt. 3d model retrieval using probability density-based shape descriptors. In *IEEE Transactions of Pattern Analysis and Machine Intelligence*, volume 31, 2009.

[2] A.Singh A.Krause and C.Guestrin. Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, 2008.

[3] B.Krishnapuram, C.M. Bishop, and M.Szummer. Generative models and bayesian model comparison for shape recognition. *Ninth International Workshop on Frontiers of Handwriting Recognition*, 2004.

[4] H. Borotschnig, L. Paletta, M. Prantl, and A. Pinz. Active object recognition in parametric eigenspace. In *British Machine Vision Conference (BMVC)*, pages 629–638, 1998.

[5] D.Macrini, C.Whiten, R.Laganiere, and M.Greenspan. Probabilistic shape parsing for view-based object recognition. In *21st International Conference on Pattern Recognition*, 2012.

[6] E.Sommerlade and I.Reid. Information-theoretic active scene exploration. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[7] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Prentice Hall, 2002.

[8] R. M. Haralick and L. G. Shapiro. *Computer and Robot Vision*. Addision-Wesley Longman Publishing, 1992.

[9] J.Denzler and C.M. Brown. Information theoretic sensor data selection for active object recognition and state estimation. In *IEEE Transactions on PAMI*, 2002.

[10] J.M.Glover. Probabilistic procrustean models for shape recognition with an application to robotic grasping. Master's thesis, MIT, 2008.

[11] H. Kauppinen, T. Seppanen, and M. Pietikainen. An experimental comparison of autoregressive and fourier-based descriptors in 2d shape classification. In *IEEE Transaction on Pattern Analysis and Machine Intelligence*, volume 17, pages 201–207, 1995.

[12] R. D. D. Leon and L. E. Sucar. Human silhouette recognition with fourier descriptors. In *15th International Conference on Pattern Recognition*, pages 709–712, 2000.

[13] G. Lu and A. Sajjanahr. Region-based shape representation and similarity measure suitable for content-based image retrieval. In *IEEE Transactions on Pattern Analysis and Machine Learning*, pages 164–174, 1999.

[14] P. J. Van Otterloo. *A Contour oriented Approach to Shape Analysis*. Prentice Hall, 1991.

[15] E. Persoon and K. Fu. Shape discrimination using fourier descriptors. In *IEEE Transaction On Systems, Man and Cybernetics*, pages 170–179, 1977.

[16] T. W. Rauber. Two-dimensional shape description. Technical report, University Nova de Lisboa, Portugal, 1994.

[17] R.Rosales S.Yu, B.Krishnapuram and R.Rao. Active sensing. In *IEEE International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 639 – 646, 2009.

[18] G. Taubin. *Recognition and Positionning of Rigid Objects using Algebraic and Moment Invariants*. PhD thesis, Brown University, December 1990.

[19] C. H. Teh and R. T. Chin. On image analysis by the methods of moments. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 10, 1988.

[20] Q. M. Tieng and W. W. Boles. Recognition of 2d object contours using wavelet transform zero crossing representation. In *IEEE Transaction on Pattern Analysis and Machine Learning*, 1997.

[21] F. J. Janse van Rensburg, J. Treurnicht, and C. J. Fourie. The use of fourier descriptors for object recogntion in robotic assembly. In *5th CIRP International Seminar on Intelligent Computation in Manufacturing Engineering*, 2006.

[22] H. S. Yang, S. U. Lee, and K. M. Lee. Recognition of 2d contours using starting-point-independent wavelet coefficient matching. In *Journal of Visual Communication and Image Representation*, volume 9, pages 171–181, 1998.

[23] C. T. Zahn and R. Z. Roskies. Fourier descriptors for plane closed curves. In *IEEE Transaction on Computer*, volume 21, pages 269–281, 1972.

[24] D. Zhang and G. Lu. A comparative study on shape retrieval using fourier descriptors with different shape signatures. In *Victoria*, volume 14, pages 1–9, 2001.