

Planar Object Detection under Scaled Orthographic Projection

Julián Ramos Cózar, Nicolás Guil Mata, and Emilio López Zapata

Dept. of Computer Architecture, University of Málaga,
{julian,nico,ezapata}@ac.uma.es

Abstract. In this work a new method to detect objects under scaled orthographic projections is shown. It also calculates the parameters of the transformations the object has suffered. The method is based on the use of the Generalized Hough Transform (GHT) that compares a template with a projected image. The computational requirements of the algorithm are reduced by restricting the transformation to the template edge points and using invariant information during the comparison process. This information is obtained from a precomputed table of the template that is directly transformed and compared with the image table. Moreover, a multiresolution design of the algorithm speeds-up the parameters calculation.

1 Introduction

Planar object recognition and three-dimensional (3D) pose estimation are the most important tasks in computer vision. Some applications related to these methods are object manipulation, autonomous vehicle driving, etc.

Perspective projection is one of the most suitable models for real camera image formation. However, it introduces non-linear relations. Simpler camera models can be used [1] under more relaxed situations which produce a negligible error. When the size of the object is relatively small (in comparison to its depth) and placed near the camera optical axis, the orthoperspective projection (or, in general, the subgroup of affine transformations [2]) is a good approximation to the perspective projection [3]. Here, the scaled orthographic projection will be studied. This projection only incorporates the distance and the foreshortening effects, but it eliminates the non-linear dependencies.

There are several methods to find the transformations between a planar template and an image where this template is included and viewed from an arbitrary position [4]. Selection of the most suitable method will be based on considerations such as the kind of shapes we are dealing with and the ability to cope with practical problems such as occlusion, noise, etc.

A simple approach to object detection is to find, for every possible orientation and position, the template transformation that produces a better matching with the image shape. However, the search space can become overwhelmingly large. An efficient evaluate and subdivide search algorithm is carried out in [5] in order

to find the affine transformation that brings the larger number of model features close to image features. A variant of the Hausdorff distance is used as a similarity measure between the two point sets.

Search space can be reduced if invariant features are used. Geometric invariants are shape descriptors that remain unchanged under geometric transformations such as changing the viewpoint. Algebraic invariants are obtained using the classical results derived from perspective geometry of algebraic curves. The fundamental invariant in projective geometry is the cross-ratio (CR). Guo Lei [6] uses the CR to recognize 3D views of different polygons. B.S. Song et al. present in [7] a target recognition method based in CR that selects stable points for complex scenes. More general algebraic invariants can be derived from configurations of conics [8], points [9] and lines [10]. Besides the algebraic invariants, Rothwell et al. use a canonical frame, invariant to perspective transformations of the plane, to implement index functions that select models from a model database, as part of a recognition system [11].

Differential invariants can also be applied. In [12], affine invariants requiring one point correspondence and second order derivative, or requiring two points correspondences and first order derivative, are used to determine whether one curve may be a perspective projection of another one.

Other methods based on Fourier Descriptors [13] are aimed to exploit viewpoint invariants, but they are not fully invariant in perspective transformations. Furthermore, they are not robust when occlusion appears.

Several methods have been proposed to detect the 3D pose of a planar object based on the Hough transform (hashing methods). Using the HT we can take advantage of its useful properties, like relative insensitivity to noise, and robustness to occlusions. However, it needs high computational and storage requirements.

In this paper we undertake a new approach to planar object detection based on the Generalized Hough Transform (GHT) that compares template and image information in order to calculate the transformations between the template and the corresponding image shape. Computational and storage requirements are greatly reduced by using invariant template and image information derived from the gradients and positions associated with the edge points. Invariant information is stored in template and image tables which are compared during the detection process. Template (or model) table information will be directly modified during algorithm application without needing further computationally expensive shape processing.

The rest of the paper is organized as follows. Next section introduces the mathematical expressions involved in an orthoperspective transformation. Section 3 presents the new method for planar shape detection and the expressions that allows us to speed-up the generation and comparison of the tables. In section 4, several real experiments have been carried out in order to test the algorithm's behaviour. Finally, in section 5 several related works are analyzed more deeply.

2 Scaled Orthographic Projection

The necessary transformations to project a planar object into the image plane are presented here, where f is the focal distance of the camera lens, d is the distance between the focal point and the intersection of the object plane with the z -axis (the optical axis), and \mathbf{n} being the normal vector to the object plane. The projection of this vector onto $x-y$ and $x-z$ planes allows us to determine the pan, τ , and the tilt, δ , angles, respectively. Thus, τ is the angle between the projection of \mathbf{n} onto the $x-y$ plane and the x -axis. On the other hand, δ is the angle between the projection of \mathbf{n} onto the $x-z$ plane and the z -axis.

The relationship between an object point (x_i, y_i, z_i) and its corresponding image point (u_i, v_i, f) can be expressed using the orthographic transformation:

$$\begin{aligned} u_i &= f \cdot \frac{x_i \cdot \cos\delta \cdot \cos\tau - y_i \cdot \sin\tau}{d} \\ v_i &= f \cdot \frac{x_i \cdot \cos\delta \cdot \sin\tau + y_i \cdot \cos\tau}{d} \end{aligned} \quad (1)$$

3 Planar Object Detection

The detection process must find the occurrence of a template in an image. Six different groups of parameters indicate the transformation to be applied to the template to generate the object in the image. 1) *Scaling* of the template in the object plane represented by matrix S . 2) *Displacement* along the object plane, d_x and d_y , represented by matrix D . 3) *Rotation* in the object plane, β , represented by matrix R_β . 4) *Tilt* angle around object plane x -axis, δ , represented by matrix R_δ . 5) *Pan* angle around object plane z -axis, τ , represented by matrix R_τ . 6) *Scaling* of the scaled orthographic projection, represented by matrix S_o .

After application of the previous transformations to the template, the new coordinates of the template in the image plane are given by:

$$(u_i, v_i)^t = S_o \cdot R_\tau \cdot R_\delta \cdot R_\beta \cdot D \cdot S \cdot (x_i, y_i)^t \quad (2)$$

Note that the only transformation that introduces a distortion in the image object is the application of the tilt angle. The rest of transformations only change either the size (scaling), the orientation or the position.

3.1 Invariant Transformation

Recent work with the GHT [14] shows a new method to detect bidimensional shapes that uses invariant information to displacement and scaling.

The edge points of the image are characterized by the parameters $\langle x, y, \theta \rangle$ where x and y are the coordinates of the points in a two-dimensional space and

θ is the angle of the gradient vector associated with this edge point. An angle, ξ , called *difference angle* is also defined. Its value indicates the positive difference between the angles of the edge point gradient vectors that will be paired.

From this description we can derive a transformation from the original image that generates new invariant information for the displacement and the scaling, based on paired points. Thus, let p_i and p_j be two edge points, $\langle x_i, y_i, \theta_i \rangle$, $\langle x_j, y_j, \theta_j \rangle$ their associated information, and ξ the difference angle to generate the pairing. Then, the transformation \mathcal{T} can be expressed as follows:

$$\mathcal{T}(p_i, p_j) = \begin{cases} (\theta_i, \alpha_{ij}) & : \theta_j - \theta_i = \xi \\ \emptyset & : elsewhere \end{cases} \tag{3}$$

where

$$\alpha_{ij} = \left(\arctan \frac{y_i - y_j}{x_i - x_j} \ \angle \ \theta_i \right) \tag{4}$$

that is, α_{ij} is the positive angle formed by the line that joins p_i and p_j and the gradient vector angle of the point p_i .

The information generated by the application of the previous transformation is stored in a table in order to improve the detection process speed. The multivalued characteristics of the previous transformation will be apparent during table building. Next, we show the contents of this table:

Orientation table (\mathcal{OT}). The information generated by the \mathcal{T} transformation is contained by this bidimensional table. The α_{ij} and θ_j values are stored in rows and columns, respectively. When a pairing with α_{ij} and θ_j value is calculated, the $\mathcal{OT}[\alpha_{ij}][\theta_j]$ position is incremented. Because different pairings might coincide with the same α_i and θ_{ij} values, the content of $\mathcal{OT}[\alpha_{ij}][\theta_j]$ will indicate how many of them have these values. The information stored in this table is invariant to scale and displacement. Note that a rotation τ of an image in a plane causes a rotation of τ columns of its \mathcal{OT} .

Then, we can use the \mathcal{OT} table to compare invariant information of the template and the image under scaled orthographic projection. Before starting the generation of the \mathcal{OT} tables, the template must be transformed as follows:

$$(u_i, v_i)^t = R_\delta \cdot R_\beta \cdot (x_i, y_i)^t \tag{5}$$

In this way, the projected template will be similar to the image shape except for a different scaling, displacement and orientation. Then the template and image \mathcal{OT} s will be different by a rotation of τ columns. This displacement can be calculated applying a matching process for both tables [14].

In general situation, where the values for the (δ, β) angles are unknown, a \mathcal{OT} table for the template needs to be generated for each (δ, β) value. Each template table is compared with the \mathcal{OT} table of the image and a value for τ is calculated, see Fig. 3.1. The maximum voting will indicate the correct solution.

The table generation and comparison for each δ and β values may require a high computational complexity. However, several improvements have been carried out to obtain a good performance.

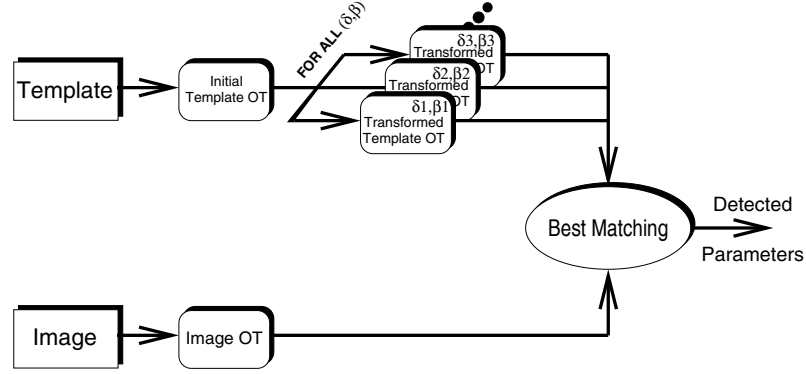


Fig. 1. Block-Diagram of the system

3.2 Table Generation and Comparison

The new OT table generation is accomplished by only using the edge points of the template shape. Normally, the number of edge points in the template is lower than in the image, where several different shapes can appear. This makes the computational complexity of our method to be lower than other algorithms using the image shape [15].

On the other hand, the OT table construction is based on both the calculation of the gradient vectors of the edge points (θ) and the value for the difference angles (α). This calculation can take a high computational time. In order to reduce this time, we have studied the modifications in an angle value that arise under a scaled orthographic projection. This allows to generate a OT table for a concrete δ and β value by directly transforming the original OT template table.

Modification of the Gradient and Difference Angles

Lemma 1. Let θ_i be the gradient angle of the template edge point (x_i, y_i) . Then, the transformed gradient vector, θ_{Ti} , after application of expression (5) is :

$$\theta_{Ti} = atan(\tan(\beta + \theta_i)/\cos\delta) \tag{6}$$

An important consequence of the previous lemma is that if (x_i, y_i) and (x_j, y_j) are two paired points in the template using a pairing angle of ξ , the transformed points after applying expression (5) are also paired.

Lemma 2. Let α_i be the difference angle value between two paired points using a pairing angle of ξ . Then, the transformed difference angle, α_{Ti} , after application of expression (5) is:

$$\alpha_{Ti} = atan(\tan(\beta + \theta_i + \alpha_i)/\cos\delta) - \theta_{Ti} \tag{7}$$

The demonstrations of the previous lemmas are straightforward, so we have omitted them due to space limitations. The two previous lemmas allow the creation of the new \mathcal{OT} tables for each δ and β angles without needing the calculation neither the projected edge points nor the new difference angle for the paired points. The \mathcal{OT} tables are sparse, so only the positions with values different from zero are stored. This reduces the required storage and allows the implementation of a more efficient transformation and comparison process.

3.3 Pose and Position Estimation

A matching algorithm has to compare, by columns, the transformed template \mathcal{OT} table – for each possible tilt, pan value – and the \mathcal{OT} of the image for different shifting values in order to calculate the pose (tilt, pan and rotation angles). The scale and the displacement values are obtained, after eliminating the previous transformations in the contour of the tested image, in the same way as in [14].

4 Experimental Results

As we showed in the previous sections, the detection process is based on the use of the gradient and difference angles associated with the edge points. Then, the accuracy of our method will be limited by the accuracy of the angle detection. Good initial values for the angles and edge points are obtained by applying a Canny operator.

We have used different images in order to test the behaviour of the whole detection process. All the examples have been executed in a SGI Workstation with a R-10000 processor at 225 MHz. The range of the tilt, rotation and pan angles for the experiments are $(0^\circ, 60^\circ)$, $(-30^\circ, 30^\circ)$, and $(0^\circ, 360^\circ)$, respectively. Images are 8 bit greyscale with half PAL size (384x288 pixels). Model images are acquired in a fronto-parallel view. There is no need for any camera calibration procedure. The pairing angles have been chosen taking into account the template's shape [14], in a way that a significant subset of the contour points of the template are paired. For this purpose the gradient's histogram of the shape must be studied.

In addition, the algorithm uses a multipass approach to accelerate the program execution. Starting with a coarse estimation of the pose angles, successive steps focus in a narrower interval around the solution estimated in the previous step. Finally, the scale and displacement values are calculated.

The accuracy of the detection process is checked by using the likelihood function stated in [16]. First, this function measures the similarity, attending to edge points localization and gradient angle values, between transformed template and image contours. A visual checking is also carried out by superimposing the projected template shape on the image one.

First, we have applied the detection to B&W model images that have been projected using a warping process (Fig. 4). In this manner, we can know the

Table 1. Parameters of the warping deformation for the test images

Template	T_{POINTS}	I_{POINTS}	(δ, β, τ)	Scale	Likelihood (%)
Moon	801	602	(45,10,-20)	0.90	87.37
Indalo	1231	1312	(39,13,29)	1.15	83.58
Clover	568	599	(50,19,-29)	1.20	82.39
Key	622	531	(34,-28,13)	0.92	92.43

exact tilt, rotation, pan and scale values which have been used and check the accuracy of the detection. This allows us to study the theoretical behaviour of our programs. The template properties and experiment configurations are showed in Table 1. The first and second columns indicate the number of points detected by the Canny operator in both the template and transformed template. The following two columns show the parameters of the projection of the template by a warping process. Finally, in the last column, the value of the maximum theoretical likelihood is expressed. This value is calculated comparing the projected template with the warped image for the optimal pairing angles using the exact transformation parameters in pose and scale (the displacement has to be estimated). A 100% likelihood is not reached due to the discrete nature of the image transformation.

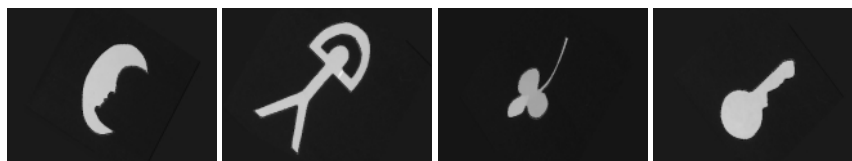
**Fig. 2.** Warped images

Table 2 summarizes the parameters used in program execution and the results obtained for the detection proposes. (ξ_1, ξ_2) are the used pairing angles. When two pairing angles are used, the achieved accuracy is better in general. The (δ, β, τ) , scale and the likelihood for the detected parameters are depicted in the subsequent columns. Finally, the computational time for the multipass and sequential algorithm is shown in the last two columns of Table 2. An important speed-up is achieved for the multipass approach.

Finally, several real images, obtained using a CCD camera, are presented in Fig. 3 in order to show the behaviour of the proposed algorithms in practical situations. The images in the left column are the templates used and the images in the right column shows a visual checking of the solution, by superimposing the projected template on the image.

Table 2. Results obtained for warped images

Image (points)	(ξ_1, ξ_2)	(δ, β, τ)	Scale	Like. (%)	Time _{Multi} (s.)	Time _{Seq} (s.)
Moon ₁	(180, -)	(45,10,-20)	0.90	87.37	0.20	3.73
Moon ₂	(180, 90)	(45,10,-20)	0.90	87.82	0.48	9.10
Indalo ₁	(180, -)	(39,10,-32)	1.15	71.80	0.51	9.72
Indalo ₂	(175, 125)	(39,13,-29)	1.15	83.58	0.96	18.91
Clover ₁	(180, -)	(50,18,-30)	1.20	78.68	0.14	2.74
Clover ₂	(45, 90)	(50,25,-27)	1.12	63.74	0.37	7.13
Key ₁	(180, -)	(34,-30,11)	0.92	94.63	0.12	2.36
Key ₂	(180, 150)	(33,-25,17)	0.92	88.37	0.30	5.63

Table 3. Practical situation

Image (points)	(ξ_1, ξ_2)	(δ, β, τ)	Like. (%)	Time (s.)
Pliers (1803)	(180, 135)	(20,-19,-40)	73.13	1.41
Jewish Harp (1483)	(180, 45)	(43,-24,8)	81.27	0.91
Cutter (1650)	(180, 135)	(20,-6,9)	73.46	0.47
Scissors (2561)	(45, -)	(12,35,4)	80.47	0.67

Table 3 summarizes the more important configuration values. Now, for real images, two pairing angles, except for *Scissors*, are necessary if a reasonable precision is looked for. The images were acquired from an unknown arbitrary point of view, so the detected pose in the second column is only for informative purposes. The values for the likelihood function reported by the algorithm show that good accuracy have been achieved, in comparison with the values reported for the warped images. Finally, the computational times are presented for the multipass algorithm.

5 Related Works

In this section, a brief review of another works that use the HT to estimate the 3D orientation and position of a planar patch is presented. Others papers, using the HT for detecting 2D transformations are not covered here [17,18].

The method described in [19] claims to be the first one in using the HT for planar object 3D pose estimation. It consists in mapping the classical 2D Hough space into another Hough space of the same dimensions using only geometrical considerations. The new space is related to the former through a 3D transformation that takes into account the orientation and location of an object. In this way, data and model are transformed in a common representation space, where a direct matching can be carried out by determining the pose of the actual object with respect to the known pose of the model. The method is based on the classical HT to detect straight lines, and hence it deals with rigid planar

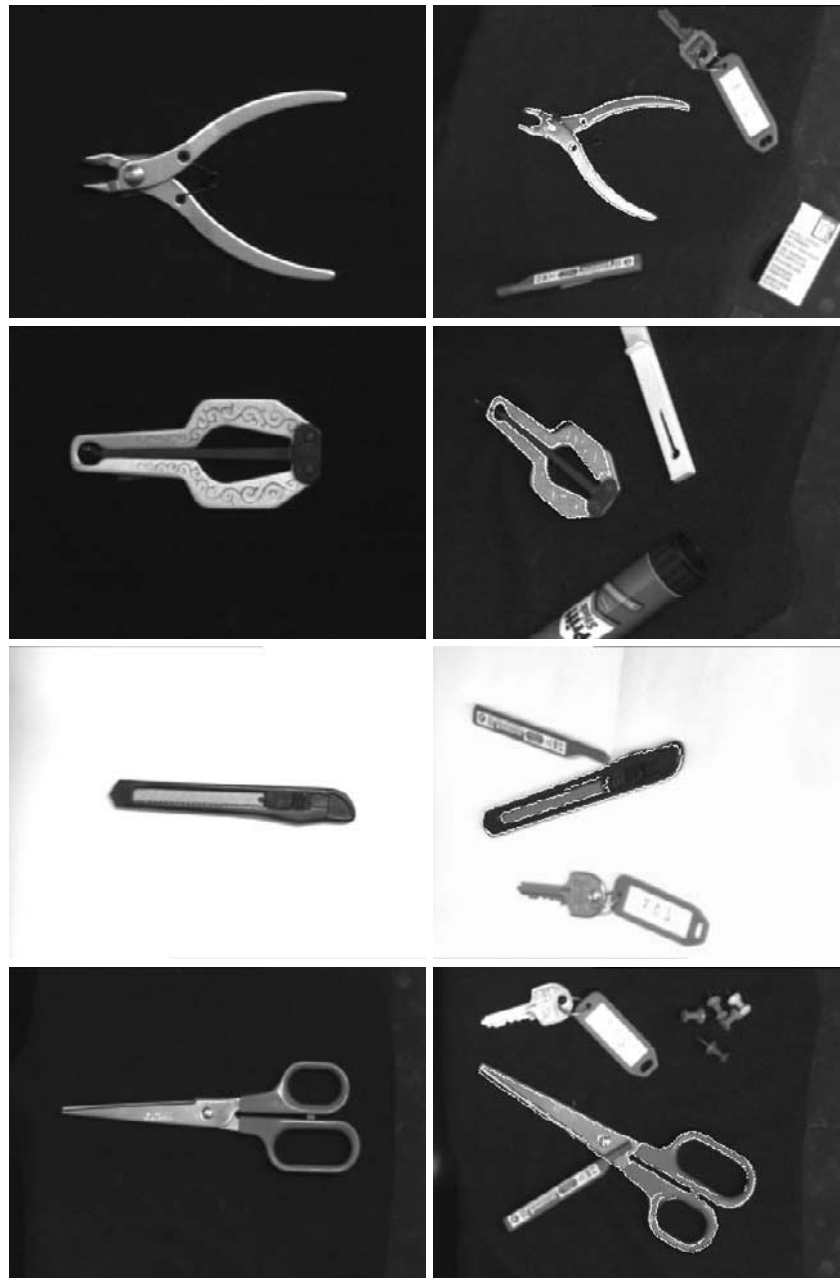


Fig. 3. Practical situation

objects bounded by straight segments. Planar shapes bounded by curvilinear edges, polyhedra and scenes containing several objects can also be tackled, in an indirect and more complex fashion, but they are not addressed in the work presented. The main limitation of this approach, as pointed by the authors, lies in the computational complexity of the matching process, which involves exploring a 6-dimensional space for pose estimation.

A new version of the GHT, called PTIGHT [20], is proposed for detecting a perspective transformed planar shape in a perspective image that is taken from an unknown viewpoint. In order to build a *perspective reference table* from a given template, all possible perspective shape images are derived by applying forward perspective transformation on the template for all viewing directions and positions. The overlapping of this information imposes two restrictions. First, the positions in the HT with votes greater than a threshold are considered potential candidates. This hypothesis must be verified by back-projecting the image shape over the template in order to find the best solution. The resolutions that can be achieved are poor, 5° for the tilt and pan angles in the reported results. Although the images tested there are not complex, the computational time is high. Also, a camera calibration is required in order to know the focal distance.

The algorithm presented in [15] detects planar objects that are rotated a tilt and pan angle, and distorted by a perspective projection before applying a chain of 2D transformations. Our method reduce the computational complexity of this approach by using a less complex perspective model and restricting the transformation to be applied to the template table which, usually, has a lower number of entries than the image table.

Finally, a technique dealing with parallel projected planar objects with rotational symmetry is presented by Yip [21]. The main contribution of the paper is that it provides a method to reduce the dimensionality of the HT space by breaking it down into several lower order ones. The problem is the shape form and projection model restrictions it imposes.

6 Conclusions

A new method to detect planar shapes under orthoperspective projection has been presented. The method copes with a situation in which the shape of the projected image presents a different displacement, orientation and scaling in relation to a template. The method is based on the GHT and generates invariant information, by using gradient information, that allows us to uncouple the parameter calculation and, in this way, reduce computational complexity. A new table generation is accomplished by only using the edge points of the template shape. Generally, the number of edge points in the template is lower than in the image, making the computational complexity of our method to be lower than algorithms using the image shape. Important improvements have been introduced to save gradient angle calculation of the projected points. Several examples, that show the accuracy of the algorithm with real images, have also been presented.

References

1. J. Y. Aloimonos. Perspective approximations. *Image and Vision Computing*, 18(3):179–192, August 1990. 479
2. Yu Cheng. Analysis of affine invariants as approximate perspective invariants. *Computer Vision and Image Understanding*, 63(2):197–207, March 1996. 479
3. S. C. Pei and L. G. Liou. Finding the motion, position and orientation of a planar patch in 3D space from scaled-orthographic projection. *Pattern Recognition*, 27(1):9–25, 1994. 479
4. Isaac Weiss. Review. Geometric invariants and object recognition. *International Journal of Computer Vision*, 10(3):207–231, June 1993. 479
5. W. J. Rucklidge. Efficiently locating objects using the Hausdorff distance. *IJCV*, 24(3):251–270, September 1997. 479
6. G. Lei. Recognition of planar objects in 3-D space from single perspective views using cross ratio. *RA*, 6:432–437, 1990. 480
7. B. S. Song, I. D. Yun, and S. U. Lee. A target recognition technique employing geometric invariants. *PR*, 33(3):413–425, March 2000. 480
8. C. A. Rothwell, A. Zisserman, C. I. Marinou, D. A. Forsyth, and J. L. Mundy. Relative motion and pose from arbitrary plane curves. *IVC*, 10:250–262, 1992. 480
9. D. Oberkampf, D. F. DeMenthon, and L. S. Davis. Iterative pose estimation using coplanar feature points. *CVIU*, 63(3):495–511, May 1996. 480
10. M. I. A. Lourakis, S. T. Halkidis, and S. C. Orphanoudakis. Matching disparate views of planar surfaces using projective invariants. *Image and Vision Computing*, 18(9):673–683, June 2000. 480
11. C. A. Rothwell, A. Zisserman, D. A. Forsyth, and J. L. Mundy. Planar object recognition using projective shape representation. *IJCV*, 16(1):57–99, September 1995. 480
12. R. J. Holt and A. N. Netravali. Using affine invariants on perspective projections of plane curves. *CVIU*, 61(1):112–121, January 1995. 480
13. Hannu Kauppinen, Tapio Seppänen, and Matti Pietikäinen. An experimental comparison of autoregressive and Fourier-based descriptors in 2D shape classification. *IEEE Transactions on Pattern Anal. and Machine Intell.*, 17(2):201–207, February 1995. 480
14. N. Guil, J. M. Gonzalez-Linares, and E. L. Zapata. Bidimensional shape detection using an invariant approach. *Pattern Recognition*, 32(6):1025–1038, 1999. 481, 482, 484
15. N. Guil, J. R. Cózar, and E. L. Zapata. Planar 3D object detection by using the generalized Hough transform. In *10th Intl. Conf. on Image Analysis and Processing*, pages 358–363, September 1999. 483, 488
16. A. K. Jain, Y. Zhong, and S. Lakshmanan. Object matching using deformable templates. *PAMI*, 18(3):267–278, March 1996. 484
17. S. C. Jeng and W. H. Tsai. Scale- and orientation-invariant generalized Hough transform: A new approach. *PR*, 24:1037–1051, 1991. 486
18. D. C. W. Pao, H. F. Li, and R. Jayakumar. Shapes recognition using the straight line Hough transform: Theory and generalization. *PAMI*, 14(11):1076–1089, November 1992. 486
19. Vittorio Murino and Gian Luca Foresti. 2D into 3D Hough-space mapping for planar object pose estimation. *Image and Vision Computing*, 15(6):435–444, 1997. 486

20. R. C. Lo and W. H. Tsai. Perspective-transformation-invariant generalized Hough transform for perspective planar shape detection and matching. *Pattern Recognition*, 30(3):383–396, March 1997. 488
21. Raimond K. K. Yip. A Hough transform technique for the detection of parallel projected rotational symmetry. *Pattern Recognition Letters*, 20(10):991–1004, November 1999. 488