Vision-based Detection of Stair-cases

Stephen Se
Department of Computer Science, University of British Columbia
Vancouver, B.C. Canada V6T 1Z4
and
Michael Brady
Department of Engineering Science, University of Oxford
Oxford OX1 3PJ, U.K.

Abstract

Stair-cases are useful environmental landmarks for navigation in mobility aids for the partially sighted. In this paper, a texture detection method using Gabor Filters is proposed to detect distant stair-cases. When close enough, stair-cases are then detected by looking for groups of concurrent lines, where convex and concave edges are partitioned using intensity variation information. Stair-case pose is estimated by a homography search approach. Using an *a priori* stair-case model, search criteria and constraints are established to find its vertical rotation and slope. These algorithms have been applied to both synthetic and real images with promising results.

1 Introduction

The problem we discuss here arose originally as part of the navigation function of a **T**echnological **A**id aimed at helping **P**artially **S**ighted (TAPS) which aims to provide a full mobility and navigation capability for partially sighted people. Obstacle detection [22, 16] and elevation changes detection such as kerbs [21] and steps are two of the most basic requirements of any TAPS. This paper concerns the visual detection of steps.

Stair-cases are useful environmental features that the partially sighted needs to be made aware of. A stair-case may be a location by which to orient themselves (for example you are passing the steps of St Paul's) or an important way point along the route that they wish to travel.

In this paper, we aim to detect the presence of a stair-case and to estimate its orientation within a 5m range. The gradient, i.e. how steep the stair-case is, is not essential but useful information for the user since it can forewarn the partially sighted person that the stairs are steep or is normal.

2 Texture Detection

Figure 1 shows an image sequence as a user approaches a stair-case from far away. All edge detectors to a greater or less extent, more or less explicitly, smooth the input data, for example, Sobel performed weighed averaging of the signal in the direction orthogonal to the direction that the gradient is sought. We have chosen to use Canny which uses Gaussian smoothing, as do a lot of other edge detectors, this serves to make the general point.

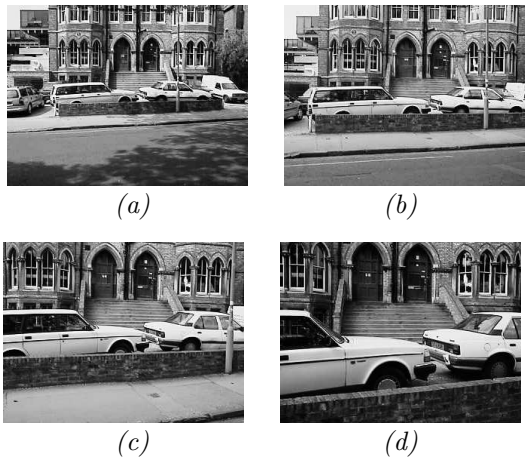

Figure 1: *An image sequence as a person approaches a stair-case from far away.*

During Canny edge detection, the image is smoothed first and this smoothing process blurs the edges. When the stair-case is far away, it may be just a few pixels wide for its tread and its riser, any smoothing will make them indistinguishable. Figure 2 shows the edge detection output for the image sequence, where edges in some cases cannot be resolved. Therefore, we will treat the stair-case pattern as a type of texture for detection.

2.1 Gabor Filter

Texture segmentation requires simultaneous measurements in both the spatial and the frequency domains. Filters with small bandwidth in the frequency domain allow us to make finer distinction among different textures. Meanwhile, accurate localisation of texture boundaries requires filters that are localised in spatial domain. However, the width of a filter in the spatial domain and its bandwidth in the frequency domain are inversely related.

Therefore, Gabor filters [8] are widely used in texture classification and segmentation as they

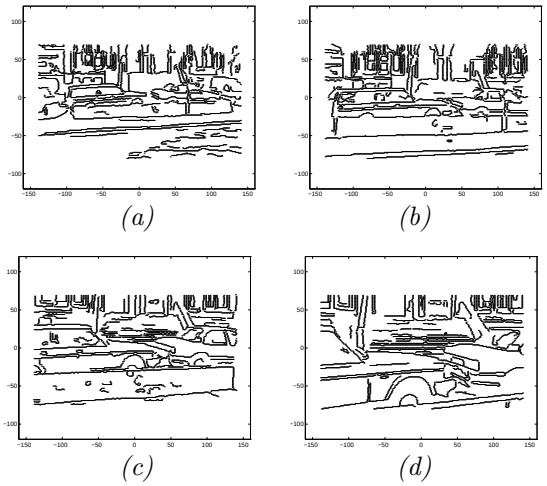

Figure 2: *Edge detection results for the image sequence. It can be seen that the edges cannot be resolved in some cases as the stair-case is too far away.*

have joint optimum resolution in both the spatial and the frequency domains [6]. They have also been shown to be good fits to the receptive fields of simple cells in the striate cortex [5] and hence their relationship to models for the early vision of mammals.

The multi-channel filtering approach uses a set of spatial filters with frequency- and orientation-selective properties. These Gabor filters of different frequencies and orientation are convolved with a textured image and there will be good response when the Gabor filter of the right frequency and orientation is used. In general, if there are different types of texture on the image, using this set of Gabor filters, and then combining with some pattern classifier, one can segment the different types of texture.

Jain and Farrokhnia [11] presented one such technique for texture segmentation using a bank of even-symmetric Gabor filters, with justification from psychophysical grounds [14]. The filtering operations using the filter set can be interpreted as computing the wavelet transform of the input image at selected frequencies. Computational savings can be realised by using subsampled filters and this only results in a modest degradation in segmentation accuracy [19].

There is also the filter-design approach which focuses on designing one or a few filters for a particular application. Instead of employing *ad hoc* banks of fixed parameters Gabor filters, Weldon and Higgins [26] tuned the parameters for specific texture segmentation problems. This is a supervised method as representative texture samples need to be given.

Rician statistics of filtered textures at two different Gabor-filter envelope scales are used to generate probability density estimates for each filtered texture over various filter parameters [27]. The filter design is then established by selecting the filter associated with the minimum predicted error.

For our current work, we use the multi-channel filtering approach with a set of predetermined filter parameters. In the experiments considered here, all the stair-case edges are horizontal, therefore, there will be maximum energy in the vertical direction and minimum energy in the horizontal direction. Since we do not know the right frequency, we convolve the images with a series of Gabor filters of different frequencies horizontally and vertically.

The goal is therefore, to determine the frequency when there is good response vertically, but poor response horizontally. Once this is identified, we can locate the stair-case too.

2.1.1 Algorithm

The stages of the algorithm are as follows:

For each frequency (between some lower and upper bound) and each orientation (discretised into a few directions),

- create a Gabor filter of the required frequency
- rotate the filter to the required orientation
- convolve the image with this filter
- do thresholding to discard poor responses
- discard edges of length less than L to avoid noise
- thinnering the edges
- keep regions with at least k edges to find stair-case of at least k steps
- compute the area of each such region and keep regions of at least area A

Subsequently, we choose the frequency and orientation that gives the maximum area, when a region exists. Also, we need to verify that the filter with the same frequency but at the orthogonal direction gives poor response. Hence, we obtain the frequency and orientation of the stair-case region (if any) in the image, as well as its approximate position.

There is a potential problem: if f is the right frequency satisfying the above criteria, multiples of f will give good response too.

2.1.2 Results

Figure 3 shows the Gabor filters for the different frequencies that we use in our experiments, ranging from 8 to 36. For our case, since the stair-cases are horizontal, we will just show results of convolution with filters horizontally and vertically.

For a sequence of images as a user approaches a stair-case, we know in advance that the frequency will be decreasing as one gets closer. Therefore, we can use the frequency obtained from the previous step as an upper bound for the current step, this will help to avoid choosing the multiples of the right frequency.

Considering the image sequence in Figure 1, we apply this algorithm with a discretisation of 4 first and then refine with a discretisation of 2. We obtain Figures 4 to 7 where the original image is shown on the left and the best frequency has been selected. The middle one shows the vertical filter

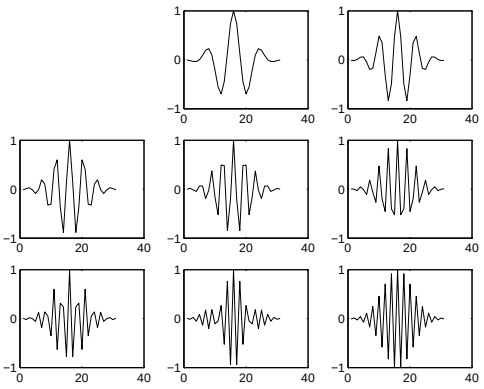

Figure 3: *Gabor filters at different frequencies. From 8 to 36 at increment of 4.*

Figure 4: *Image 1 from the sequence, frequency found is 26 (see text for details).*

result and the one on the right shows the horizontal filter result at the selected frequency. The stair-case has been localised and the frequencies for the sequence are found to be (26, 22, 18, 14) respectively.

Stair-case pose is not important at this stage as the stair-case is still a long way away. When it is close enough for its edges to be resolvable, it will switch to the following approach which can then estimate its orientation and slope.

3 Stair-case Detection

Stair-cases can be characterised as a sequence of steps, which can be regarded as a group of consecutive kerb edges. Following the kerb detection approach [21], we start with Canny edge detection and Hough Transform line-fitting on the image.

Stair-case edges are parallel to each other in 3D space, we are not interested in spiral ones here. Therefore, when they are projected onto the image, these edges will intersect at a vanishing point (provided that they are not fronto-parallel to the image plane). Usually when a stair-case is seen

Figure 5: *Image 2 from the sequence, frequency found is 22.*

Figure 6: *Image 3 from the sequence, frequency found is 18.*

Figure 7: *Image 4 from the sequence, frequency found is 14.*

from a distance, the lines will be quite parallel to each other, and the vanishing point will be far from the image. It is therefore logical to search for concurrent lines when looking for a structure that originally consists of parallel lines.

3.1 Searching for Vanishing Point

There are two ways to find concurrent lines. One approach is to search for vanishing points using the Hough Transform [1, 18, 24, 13, 3]. After obtaining straight lines using the Hough Transform, we can apply another Hough Transform to find the intersection of these straight lines. The vanishing points are characterised as those points where most of the supporting line segment primitives intersect, so we can accumulate evidence provided by these line segments.

However, Collins and Weiss [4] considered vanishing point computation as a statistical estimation problem and observed that it is not reliable when not many lines are passing through that point. The accuracy level stays roughly the same as the number of lines drops from 100 down to 20, but degrades notably from 20 down to 5. In fact, any convergent group consisting of relatively small number of lines will be left undetected with this approach.

3.2 The Detection Algorithm

A different approach is proposed here where potential groups of candidate lines are generated and then tested for coincidence. This approach was employed by Utcke [25] for grouping and recognising zebra-crossing in cluttered images. Line clustering was also used in [15] to classify groups which share common vanishing points followed by vanishing point estimation.

Based on the projective property of structures with parallel lines, our algorithm [17] picks out groups of nearly parallel lines and checks for concurrency (hence finding the vanishing point) as hypotheses for stair-cases. Then it seeks further support from the other lines for these hypotheses, to determine the best hypothesis.

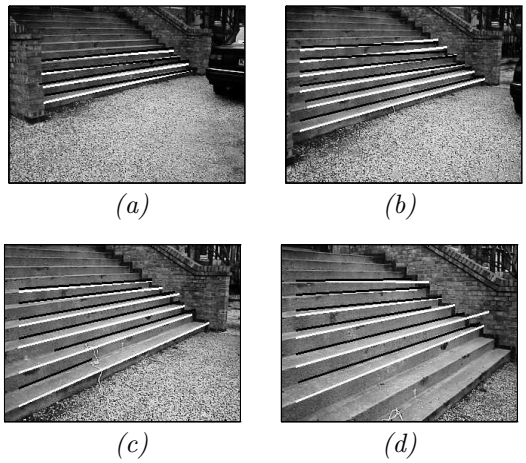

Figure 8: *Image sequence overlaid with the concave and convex edges found on the stair-case. Convex edges are marked in white while concave ones are marked in black.*

Moreover, since stair-case edges are usually long and close together, the algorithm discards short edges and edges that are far away from the rest which are likely to arise from features other than a stair-case. RANSAC [7] is employed to eliminate outliers and a least-squares procedure is then used to find the intersection of multiple lines [20].

3.3 Partition Stair-case Edges

There are various types of stair-cases, but we are most interested in the typical regular ones that have uniform tread and riser, i.e. constant slope. Using the technique described above, we obtain a hypothesis for some structure containing parallel lines. We would like to add further constraints to verify that the structure actually consists of two sets of equally-spaced parallel lines: the convex and the concave step edges. This is now a much stronger constraint for a regular stair-case compared to merely searching for structures with parallel lines.

Intensity variation of the stair-case is considered here as a cue on which to base the partition. The main idea is to detect concave and convex edges whenever there is a change of intensity from dark to light or from light to dark. Geometric constraints such as cross-ratios can be applied to refine the sets proposed by this method.

We show in Figure 8 an image sequence in which a user walks towards a stair-case. Edges are detected using Canny edge detector and the resulting lines are fitted with the Hough Transform. Stair-case edges are then identified using the vanishing point constraint. Afterwards, they are partitioned using the intensity variation approach described above. Concave and convex edges are overlaid on the images with convex edges marked in white and concave ones marked in black. It can be seen that the partition obtained using intensity variation is stable, with most edges correctly classified.

4 Pose Estimation by Homography Search

We consider the recovery of vertical rotation and slope of stair-cases since these are important information for a partially sighted person, first in order to navigate themselves to arrive at the foot of the stairs and then to gauge the degree of caution necessary to mount the stairs.

We use a search approach which is similar to Witkin's search for tilt and slant from texture [28]. However, in the general shape from texture literature [10, 28, 12, 2, 9], isotropy of texture is assumed. In Witkin's case, a maximum likelihood estimator is derived to compute the tilt and slant which will give the best isotropy texture on back-projection.

For normal stair-cases (not for spiral ones), the orientation of each step is assumed to be the same, the texture is anisotropic and therefore an *a priori* model is required. The model we adopt here is a group of non-skewed parallel horizontal lines on the image when the stair-case is facing the camera.

The textured plane due to stair-case edges is a more constrained problem than a general textured surface, as there are only two rotational pose components: one around the vertical axis (vertical rotation θ); the other around the horizontal axis (slope ϕ).

Once we have obtained the vertical rotation θ and slope ϕ, we can compute the tilt τ and slant σ parameters, which are often used in the shape from texture literature, by the following equations:

$$\tau = \tan^{-1}(\frac{1}{\tan\phi \sin\theta})$$

$$\sigma = \cos^{-1}(\sin\phi \cos\theta)$$

Our aim is to transform the image to another view by a homography [20] so that the camera in the new view will be facing the stair-case head-on. We employ criteria based on our stair-case model while we search in a discretised space of (θ, ϕ).

4.1 Search Criteria

There are two components of our model-based search criteria: one for the vertical rotation and the other for the slope.

We know that if the prediction of vertical rotation is right, after the homography transformation, the camera will be facing the stair-case edges head-on and so horizontal edges are expected. Therefore, the criterion for the correct vertical rotation is based on how horizontal the transformed texture lines are, quantified by their slopes.

The criterion for the correct slope is based on how skewed the transformed textures lines are. The correct one will correspond to the case when the midpoints of all the texture lines lie on a vertical line. We can quantify this measure by computing the standard deviation for the u-coordinates of all the midpoints of the image lines. The slope of a stair-case cannot be down to 0° nor up to 90°, therefore, there exists a lower and an upper bound for ϕ.

For each discretised θ between -90° and 90° and ϕ between ϕ_{lower} and ϕ_{upper}, we transform

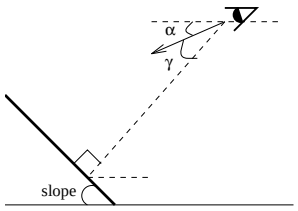

Figure 9: *The relationship between the stair-case slope, the camera tilt and the rotation angle required to achieve fronto-parallel status.*

the texture lines into a new view by homography and compute the product between the sum of absolute values of the slopes and the standard deviation of the u-coordinates of the midpoints, then we search for the lowest product over the two-dimensional space and hence obtain the corresponding θ and ϕ. To reduce the complexity of the search algorithm, we can employ a coarse-to-fine search strategy [20].

4.2 Equal Spacing Constraint

In most real images, we notice that it is usually not possible to see the full extent of the stair-case from one side to the other, either because the stair-case is so close that part of the stair-case is outside of the view, or it is partially occluded by some other objects, not least the stair rail.

Since the criterion of the search algorithm to determine the slope is by looking at how skewed the stair-case textured plane is, therefore occlusion of one side or both sides of the stair-case will prevent the algorithm from estimating the correct slope.

We have partitioned the stair-case edges into two groups: convex and concave, i.e. we have got two groups of equally-spaced lines in 3D.

After estimating the vertical rotation of the stair-case, we apply a plane-to-plane homography to the image to remove the vertical rotation effect, so that we are now facing the stair-case, but there is still the projective effect which makes each group of stair-case edges non-equally-spaced in the image. However, if we do a rotation transformation of an appropriate angle γ around the X-axis, we can obtain a fronto-parallel situation, where each group of stair-case edges are equally-spaced in the image as well.

Hence, we can use how well each group of stair-case edges are equally-spaced to determine the right rotation angle γ. At the specified discretisation, the algorithm searches for γ to rotate around the X-axis to get the best equally-spaced lines in the image.

From Figure 9, we can then estimate the slope ϕ of the stair-case with

$$\phi = 90^\circ - \alpha - \gamma \qquad (1)$$

where α is the tilt of the camera.

However, we need the slope value to define the homography to remove the vertical rotation effect. We have been using an arbitrary slope value for the vertical rotation homography search, as the slope does not affect how horizontal the texture lines are. But here the actual slope value is needed, otherwise the transformed texture lines will be distorted and the rotation γ found will be incorrect. But the goal of searching for the right rotation angle γ is to obtain the slope.

Therefore, we propose an iterative scheme using a starting value for the slope, say 45°. We apply the algorithm and obtain a value for γ, with Equation 1, we can then use a better slope for the next iteration and so on. This process is iterated three times.

5 Results

As a real application, it is crucial to know how reliable the estimation is, therefore error analysis are carried out. Details can be found in [20]. Now, we use the image sequence in Figure 8 where the measured slope is around 30° and the actual vertical rotation is around 45°. Firstly, the vertical rotation θ^* is estimated. Applying the iterative process, setting slope $\phi_0 = 45^\circ$ in the beginning:

$$\phi_0, \theta^* \longrightarrow \gamma_1$$

$$\phi_1, \theta^* \longrightarrow \gamma_2$$

$$\phi_2, \theta^* \longrightarrow \gamma_3$$

The iterative γ_i obtained at each stage is used to compute the slope ϕ_i (using Equation 1) for the next iteration, we obtain ϕ_3 at the end for the slope after three iterations.

Both the vertical rotation θ and the pitch angle γ are estimated at 0.1° discretisation, and then error analysis is carried out to compute their standard deviation. Standard deviation of 1 pixel is assumed for the original image points. The results are tabulated in Table 1. Since the stair-case is quite far away in step (a), the edges are close together and quite equally-spaced already, the estimation is therefore not good. Reasonably acceptable results are obtained for the others.

Step		(a)	(b)	(c)	(d)
θ		39.3	41.9	49.2	45.1
σ_θ		0.54	0.34	0.27	0.27
1^{st}	γ	2.8	44.2	54.3	53.6
	σ_γ	15.13	3.28	2.18	0.96
	slope	72.2	30.8	20.7	21.4
2^{nd}	γ	9.2	42.1	50.5	50.5
	σ_γ	2.60	3.41	1.97	3.05
	slope	65.8	32.9	24.5	24.5
3^{rd}	γ	7.7	42.4	51.3	51.1
	σ_γ	3.52	3.60	2.27	2.44
	slope	67.3	32.6	23.7	23.9

Table 1: θ and γ estimated at various steps for the image sequence in Figure 8 at 0.1° discretisation, with the corresponding standard deviation and the estimated slope.

6 Conclusion

In this paper, we apply a texture detection method using Gabor Filters to recognise distant stair-

cases. When they are close enough and their edges are resolvable, we look into stair-case detection by grouping lines and checking for concurrency, as the vanishing point is a projective property for structures with parallel lines. Intensity variation is used to partition the convex and concave edges afterwards. To recover the vertical rotation and the slope of the stair-case found, we employed homography with some search criteria as well as the equal-spacing constraint of the convex and concave groups.

Although the algorithms have been developed for TAPS, they are by no means limited to such applications. Stair-cases are important landmarks for outdoor mobile robots, for example in map-building applications or for navigation purposes.

These algorithms are working in theory, however in terms of speed, they are slow and far from real-time. Therefore, we will need to optimise them greatly before they can be used in practice. Moreover, further trials with different scenes are required to evaluate their robustness and performance, for instance using Receiver Operating Characteristic (ROC) curves [23]. Future work will also include planning a trajectory to approach the stair-cases found.

References

[1] S.T. Barnard. Interpreting perspective images. *Artificial Intelligence*, 21(4):435–462, November 1983.

[2] A. Blake and C. Marinos. Shape from texture: Estimation, isotropy and moments. *Artificial Intelligence*, 45(3):323–380, 1990.

[3] B. Brillault-O'Mahony. New method for vanishing point detection. *Computer Vision, Graphics, and Image Processing*, 54(2):289–300, September 1991.

[4] R.T. Collins and R.S. Weiss. Vanishing point calculation as a statistical inference on the unit sphere. In *Proceedings of the Third International Conference on Computer Vision*, pages 400–403, Osaka, Japan, December 1990.

[5] J.G. Daugman. Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Res.*, 20:847–856, 1980.

[6] J.G. Daugman. Uncertainty relation for resolution in space, spatial-frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am.*, 2(7):1160–1169, 1985.

[7] M.A. Fischler and R.C. Bolles. Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography. *Commun. Assoc. Comp. Mach.*, 24:381–395, 1981.

[8] D. Gabor. Theory of communication. *Journal of the Institute of Electrical Engineers*, 93:429–457, 1946.

[9] J. Garding. Shape from texture and contour by weak isotropy. *Artificial Intelligence*, 64(2):243–297, 1993.

[10] J.J. Gibson. *The Perception of the Visual World*. Houghton Mifflin, Boston, 1950.

[11] A.K. Jain and F. Farrokhnia. Unsupervised texture segmentation using gabor filters. *Pattern Recognition*, 24(12):1167–1186, 1991.

[12] K. Kanatani. Detection of surface orientation and motion from texture by a stereological technique. *Artificial Intelligence*, 23:213–237, 1984.

[13] M.J. Magee and J.K. Aggarwal. Determining vanishing points from perspective images. *Computer Vision, Graphics, and Image Processing*, 26:256–267, 1984.

[14] J. Malik and P. Perona. Preattentive texture discrimination with early vision mechanisms. *J. Opt. Soc. Am. A*, 7:923–932, 1990.

[15] G.F. McLean and D. Kotturi. Vanishing point detection by line clustering. *IEEE Trans. Pattern Analysis Mach. Intell. (PAMI)*, 17(11):1090–1095, November 1995.

[16] N. Molton, S. Se, J.M. Brady, D. Lee, and P. Probert. A stereo vision-based aid for the visually impaired. *Image and Vision Computing*, 16(4):251–263, 1998.

[17] N. Molton, S. Se, M. Brady, D. Lee, and P. Probert. Robotic sensing for the partially sighted. *Robotics and Autonomous Systems*, 26:185–201, 1999.

[18] L. Quan and R. Mohr. Determining perspective structures using hierarchical hough transform. *Pattern Recognition Letters*, 9:279–286, 1989.

[19] T. Randen and J.H. Husoy. Multichannel filtering for image texture segmentation. *Optical Engineering*, 33(8):2617–2625, August 1994.

[20] S. Se. *Computer Vision Aids for the Partially Sighted*. PhD thesis, Department of Engineering Science, University of Oxford, 1998.

[21] S. Se and M. Brady. Vision-based detection of kerbs and steps. In A.F. Clark, editor, *Proceedings of British Machine Vision Conference BMVC'97*, pages 410–419, Essex, September 1997.

[22] S. Se and M. Brady. Stereo vision-based obstacle detection for partially sighted people. In *Third Asian Conference on Computer Vision ACCV'98, Volume I*, pages 152–159, January 1998.

[23] J.A. Swets, W.P. Tanner, and T.G. Birdsall. Decision processes in perception. In R.N. Haber, editor, *Contemporary Theory and Research in Visual Perception*, pages 78–101. Holt Rinehart Winston, 1968.

[24] T. Tuytelaars, L. Van Gool, M. Proesmans, and T. Moons. The cascaded hough transform as an aid in aerial image interpretation. In *Proceedings of the Sixth International Conference on Computer Vision*, pages 67–72, Bombay, India, January 1998.

[25] S. Utcke. Grouping based on projective geometry constraints and uncertainty. In *Proceedings of the Sixth International Conference on Computer Vision*, pages 739–746, Bombay, India, January 1998.

[26] T.P. Weldon and W.E. Higgins. Design of multiple gabor filters for texture segmentation. In *Proceedings of IEEE International Conference on Acoustics, Speech, Signal Processing IV*, pages 2245–2248, Atlanta, GA, May 1996.

[27] T.P. Weldon, W.E. Higgins, and D.F. Dunn. Gabor filter design for multiple texture segmentation. *Optical Engineering*, 35(10):2852–2863, October 1996.

[28] A.P. Witkin. Recovering surface shape and orientation from texture. *Artificial Intelligence*, 17(1-3):17–45, August 1981.