Determining Image Orientation using the Hough and Fourier Transforms

Christopher Hollitt
School of Engineering and Computer Science
Victoria University of Wellington
Wellington, New Zealand
christopher.hollitt@ecs.vuw.ac.nz

Ahmed Sheikh Deeb
School of Engineering and Computer Science
Victoria University of Wellington
Wellington, New Zealand
ahmed.sheik.deeb@ecs.vuw.ac.nz

ABSTRACT

Many machine vision tasks require that there be an adequate estimate of the vertical direction. Many environments have a relative richness of vertical features, so extracting the favoured direction allows the vertical direction to be estimated. In this paper two different techniques are used to extract the orientation of image features, one based on Fourier transforms, the other on Hough transforms. The two techniques in combination are shown to provide promising estimates of the vertical in a variety of indoor environments.

Categories and Subject Descriptors

I.4.7 [**Image Processing and Computer Vision**]: Feature Measurement; I.5.4 [**Pattern Recognition**]: Applications—*Computer Vision*

Keywords

Orientation, Fourier Transform, Hough Transform

1. INTRODUCTION

Many machine vision tasks require that there be a reliable estimate of which way is up. Examples include image registration, prediction of ballistic target trajectories, maintenance of robot balance and robot navigation [8]. The need for a good estimate is particularly critical in robotic applications where a loss of orientation can result in damage to the robot or to surrounding objects.

Humans and other mammals use a combination of techniques to maintain pose estimates. It is known that acceleration measurements from the vestibular system, the proprioceptive sense and vision are involved in the process [9, 1]. This redundancy of approach allows an organism to maintain a reasonable estimate of its orientation even when one sensor modality is compromised by illness or the nature of an unusual environment.

The use of vision to establish orientation in animals appears to rely on the abundance of vertical features in many environments. This richness is perhaps not surprising for built environments, where vertical features are extremely common, however many natural environments also show a bias towards vertical features [2]. As a result, the early parts of the visual cortex of many species, including humans, have more neurons that respond to vertical features than other orientations, a feature described as the *oblique effect* [7, 6, 10]. This bias in the neuron population presumably reflects the importance of accurately sensing vertical features to establish the upward direction.

In this paper we describe the use of two techniques to visually estimate the roll orientation of a camera system. The first of the techniques uses the power spectral density of an input image to find the direction showing highest changes in image intensity. This is expected to correspond to the excess of vertical textures in the environment. In the second approach we use the Hough transform to find the directions of lines in the image. In built environments, long linear features are likely to be structural features, many of which are vertical. These two techniques are expected to show different dependence on environmental features, so we combine them to obtain a better estimate of camera roll.

IVCNZ '12 November 26 - 28 2012, Dunedin, New Zealand

2. METHOD

We explore two techniques to find the dominant direction of an image. A Fourier Transform based technique is used to find the direction with the greatest power in the spectrum and a Hough transform is used to find the orientations of lines in the input images. These two approaches target different aspects of the input image and so one would expect that using the two in concert will result in a more robust estimator of orientation.

2.1 Orientation from the Fourier Transform

The Fourier transform provides a convenient method to quantify the spatial variations in an image. An image that shows intensity variations with frequency ξ will result in a corresponding peak at ξ in the power spectrum of the image. The image location at which the intensity variation occurs is not important, as the power spectrum is invariant to displacement of the input image.

If an image contains many features in a particular direction, then it will tend to have high variation in image intensity perpendicular to that direction. Thus, its spectrum will show high power in a direction orthogonal to the dominant feature orientation. Consequently, when an image contains many upright (vertically oriented) features, then its Fourier transform will have large power in the horizontal direction.

Image spatial variations in directions away from the vertical will exhibit commensurate rotations away from the horizontal in their spectral power. We would therefore expect that the orientation in the power spectral density with the highest total intensity will be orthogonal to the upright direction as seen in the original image.

We wish to find the direction in which the input image shows the most spatial variation. To do this we sum the power spectral density along a variety of directions θ to find its angular dependence $g(\theta)$. The peak in $g(\theta)$ can indicates the direction in the PSD with maximum power and therefore can be used to infer the upwards direction in the original input.

We denote the Fourier transform of an input image $f(x, y) \in \mathbb{Z}^{n\times n}$ as $(\mathcal{F}f)(\xi_x, \xi_y) \in \mathbb{Z}^{n\times n}$. That is, we consider input images of size $n \times n$ pixels and transform them into a spectrum of the same size. The variables ξ_x and ξ_y are conjugate to x and y respectively. Equivalently we can consider the Fourier transform (and the power spectrum) to be functions of ξ_ρ and ξ_θ, which describe the spatial variations in the radial and angular directions. We can therefore write

$$g(\theta) := \sum_{\xi_\rho} \left|(\mathcal{F}f)(\xi_\rho, \xi_\theta)\right|^2 \tag{1}$$

To perform this calculation we first use a two-dimensional Fast Fourier Transform to find $(\mathcal{F}f)(\xi_x, \xi_y)$ and then square its absolute value elementwise to obtain the power spectral density. There are then two equivalent approaches for finding $g(\theta)$.

1. Convert the power spectrum to polar coordinates and then sum over ξ_ρ.

2. Rotate the power spectrum $(\mathcal{F}f)^2(\xi_x, \xi_y)$ by $-\theta_i$ and sum along the resulting ξ_x direction to find $g(\theta_i)$.

The second option is computationally cheaper when $g(\theta)$ need only be calculated over a subset of possible θ values. In many practical applications there would be a prior estimate of the upright direction and consequently it is reasonable to find g over only a restricted domain. We have therefore adopted the second approach for this work.

As a refinement to the above procedure, we consider that the rectangular domain of the power spectral density can lead to bias in the summation because features oriented near 45° from the input image axes can potentially accumulate more excessive power. This has the undesirable effect of biasing the technique to find higher power near 45°. To avoid this effect, we simply low pass filter the power spectral density to ensure that the upper frequency limit is the same in all directions.

In addition images often have considerable power at low spatial frequencies, due to illumination variations across a scene or changes from one type of surface to another. We expect that such changes will not prove particularly useful in our application, so we would prefer to concentrate on higher spatial frequencies. We therefore remove spatial frequencies below some lower limit $\xi_{\min}$ from the Fourier transform of the image.

The combination of these two effects allows us to construct a mask region $\mathcal{M}$ that is preserved for further processing.

$$\mathcal{M} := \left\{(\xi_x, \xi_y) : \xi_{\min} \le \sqrt{\xi_x^2 + \xi_y^2} \le n\right\} \tag{2}$$

We also define an indicator function $M(\xi_x, \xi_y)$ that can be used to remove spatial frequencies outside $\mathcal{M}$. This function is simply the point spread function of a brickwall bandpass spatial filter.

$$M_{\mathcal{F}}(\xi_x, \xi_y) = \begin{cases} 1, & (\xi_x, \xi_y) \in \mathcal{M} \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

Thus the total operation of the Fourier processing chain is as follows.

$$f(x, y) \Leftrightarrow (\mathcal{F}f)(\xi_x, \xi_y) \tag{4}$$

$$(\mathcal{F}f)_{\mathrm{m}}(\xi_x, \xi_y) = (\mathcal{F}f)(\xi_x, \xi_y) \times M_{\mathcal{F}}(\xi_x, \xi_y) \tag{5}$$

$$g(\theta) = \sum_{\xi_\rho} (\mathcal{F}f)^2_{\mathrm{m}}(\xi_\rho, \xi_\theta) \tag{6}$$

Finally we need to find the value of θ for which $g(\theta)$ is maximum. This corresponds to our Fourier-derived estimate of the upright direction.

$$\hat{\theta}_{\mathcal{F}} = \arg\max_{\theta} g(\theta) \tag{7}$$

To extract these values we find the location of the maximum value of $g(\theta)$ and extract the four points on either side of the peak. We then fit the nine resulting points to a quadratic using a linear least squares fit. $\hat{\theta}_{\mathcal{F}}$ and $\delta\hat{\theta}_{\mathcal{F}}$ are equated to the peak location and the location of the half-width half-maximum of the resulting quadratic. That is we fit the quadratic

$$g_{\text{fit}}(\theta) = a_2\theta^2 + a_1\theta + a_0 \tag{8}$$

around the peak of $g(\theta)$, and then find the estimates

$$\hat{\theta}_{\mathcal{F}} = -\frac{a_1}{2a_2} \tag{9}$$

$$\delta\hat{\theta}_{\mathcal{F}} = -\frac{\sqrt{a_1^2 - 4a_2\left(a_0 - \frac{g(\theta^\star)}{2}\right)}}{2a_2} \tag{10}$$

$$\text{where, } \theta^\star = \arg\max_{\theta} g(\theta) \tag{11}$$

The half-width half-maximum of the quadratic provides us a metric for estimate uncertainty $\delta\hat{\theta}_{\mathcal{F}}$. Figure 1 shows an illustration of the major steps in the Fourier Transform process.

2.2 Orientation from the Hough Transform

The Hough transform is a well known family of algorithms that can be used to parameterise geometric figures in an image [5, 4]. In this work we seek to locate the presence of linear features in the input image, with the expectation that they will be predominantly aligned in the vertical direction. Thus, we use the form of the Hough transform targeted at finding lines in an input image. This is the earliest and simplest form of the Hough transform [3] and has found wide application in a variety of machine vision applications.

The Hough transform algorithm accumulates votes into a parameter space (the Hough space), which describes the set of possible parameters for lines in the input image. The possible parameters are necessarily quantised so that votes can be accumulated into bins. Here we use arguably the most common parametisation of a line in Hough transforms and characterise them by their angle (θ) and normal distance from the origin (ρ).

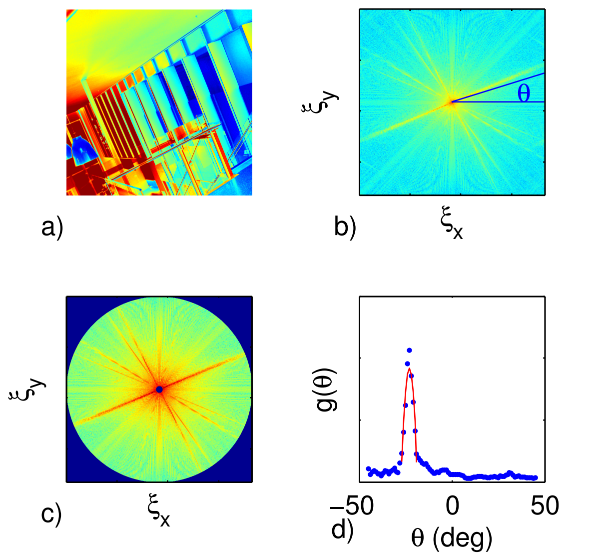

Figure 1: Illustration of the Fourier transform based technique to extract orientation. Subfigure a) shows the input image, b) shows the power spectral density of the image, c) is the same power spectral density after spatial filtering, and d) shows the variation in the power as a function of θ. The magnitude of $g(\theta)$ is arbitrary, so the scale has been omitted.

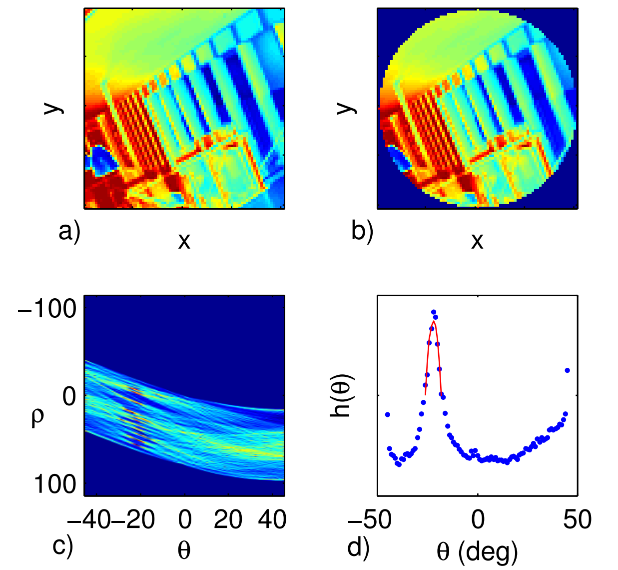

Figure 2: Illustration of the Hough transform based technique to extract orientation. Subfigure a) shows the input image, b) shows the masked image, c) is a portion of the Hough transform of the masked image and d) shows the variation in the sum of the square of the Hough transform as a function of θ. The magnitude of $h(\theta)$ is arbitrary, so the scale has been omitted.

During the Hough transform algorithm, each image pixel "votes" for all possible (ρ, θ) combinations that are consistent with that pixel. An image pixel $f(x_i, y_i)$ with non-zero intensity will cast a vote for all (ρ, θ) pairs that would result in a line through (x_i, y_i). In the simplest case, a binary input image is used and each vote has the same value.

Parameter tuples corresponding to lines in the image will accumulate many votes during the transform process. Because the transform has a high value at parameter tuples corresponding to image lines, a peak detection process yields the angle and position of image lines. In the current work we do not need full information about the lines, so a full peak detection process will not be required. We merely seek to find angles for which there appears to be an excess of linear features.

The Hough transform of a rectangular image is biased towards finding lines that cross into the corners of the image. This is because such lines can be longer that more nearly vertical or horizontal lines and can therefore accumulate more votes. To prevent this from upsetting our estimate we mask the input image so that lines in all directions can have a maximum length of n. To do this we form a mask function

$$M_{\mathcal{H}}(x, y) := \begin{cases} 1 & \sqrt{x^2 + y^2} \leq n \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

and find the masked image

$$f_{\mathrm{m}}(x, y) = f(x, y) M_{\mathcal{H}}(x, y) \tag{13}$$

We will consider masked images $f_{\mathrm{m}}(x, y) \in \mathbb{Z}^{n \times n}$, edge detect and threshold them and then find their Hough transforms $(\mathcal{H} f_{\mathrm{m}})(\rho, \theta) \in \mathbb{Z}^{n_\rho \times n_\theta}$. That is, we transform $n \times n$ pixel images into a parameter space with n_ρ values of normal distances to the origin and n_θ distinct angles.

In the application described here we are not interested in location of each line in the image. Instead we wish to extract the θ values for which there is an preponderance of lines. The voting process used to generate the Hough transform ensures that each θ value will contain the same number of votes, so summation of votes for a particular angle will not yield any information about the prevalence of lines in the image. However for angles which do correspond to lines in the input image, the votes will be concentrated at particular value (or values) of ρ. Consequently, any operation that reflects the clumpiness of the Hough transform as a function of θ can be expected to reveal the angles that contain disproportionate number of lines.

In this work we used the sum of the squares of the Hough transform along each angle. That is, we formed the function $h(\theta)$ according to

$$h(\theta) = \sum_\rho (\mathcal{H} f)^2(\rho, \theta)$$

To extract the estimate of roll angle we must find the peak in $h(\theta)$. We do this using the same quadratic fitting procedure outlined for the Fourier transform technique in section 2.1. This again results in an estimate of the roll angle and an estimate of our uncertainty in that angle, denoted $\hat{\theta}_{\mathcal{H}}$ and $\delta\hat{\theta}_{\mathcal{H}}$ respectively.

Figure 2 illustrates the operation of the Hough Transform based orientation estimation process.

2.3 Combination of the Transforms

The two techniques described above both produce an estimate of the roll angle between the input image and the vertical direction in the world. It can be expected that in some situations one or other of the approaches will struggle, due to the features of the particular environment being viewed. For example, one could imagine a scene without long linear features in which the Hough transform would not perform well, but the Fourier transform might. Similarly one could imagine scenarios that contain a lot of horizontal image variation that cause the Fourier method to fail, but in which the presence of a few strong vertical lines allow the Hough transform to perform correctly.

We therefore wish to combine the estimates derived from the two techniques to form a composite estimate. We will do this by combining weighting the two estimates according to our uncertainty in the locations of the peak values of $g(\theta)$ and $h(\theta)$. That is, we use the half-width half-maxima as proxies for the standard variation of the two measurements and combine them according to the usual conventions for normally distributed variables.

$$w_{\mathcal{F}} = \frac{(\delta\hat{\theta}_{\mathcal{H}})^2}{(\delta\hat{\theta}_{\mathcal{F}})^2 + (\delta\hat{\theta}_{\mathcal{H}})^2} \tag{14}$$

$$w_{\mathcal{H}} = \frac{(\delta\hat{\theta}_{\mathcal{F}})^2}{(\delta\hat{\theta}_{\mathcal{F}})^2 + (\delta\hat{\theta}_{\mathcal{H}})^2} \tag{15}$$

$$\hat{\theta} = w_{\mathcal{F}}\hat{\theta}_{\mathcal{F}} + w_{\mathcal{H}}\hat{\theta}_{\mathcal{H}} \tag{16}$$

$$\delta\hat{\theta}^2 = \frac{(\delta\hat{\theta}_{\mathcal{F}})^2(\delta\hat{\theta}_{\mathcal{H}})^2}{(\delta\hat{\theta}_{\mathcal{F}})^2 + (\delta\hat{\theta}_{\mathcal{H}})^2} \tag{17}$$

3. RESULTS

The algorithms described above were implemented using Matlab. A collection of images were taken inside the Alan MacDiarmid and Cotton buildings of Victoria University of Wellington using a consumer grade camera. The images were converted to greyscale, cropped to square images and then downsampled for use in the orientation estimation technique. No extrinsic ground truth was collected for these images, though it is straightforward for a human being to estimate the vertical direction from the images.

Figure 3 shows the performance of the algorithm when analysing a number of scenes. The images were downsampled to 729 pixels square for this test. The frequency mask used in the Fourier processing removed the used $\xi_{\min} = 27$ pixels. In practice the algorithm is not very sensitive to the mask size. The range of possible angles explored in each case was $\theta \in [-20..20]^\circ$ with a resolution of 1°.

As can be seen from figure 3, the algorithm appears to do a good job of estimating the vertical direction in a variety of scenes. In a collection of 39 indoor images we did not find any that caused an incorrect estimate.

A series of photographs of the same scene in different orientations are shown in figure 4. The inputs for this test were separate photographs, rather than rotated versions of a master photograph. Again, the algorithm successfully identifies the vertical direction in each image.

The performance of the algorithm was tested with images of a variety of resolutions. A variety of images were downsampled to a variety of resolutions and their performance measured. Most of the images showed no significant degradation of estimation performance decreasing resolution, though the error did increase as the resolution dropped. The image that produced the most dramatic low resolution behaviour was that shown in figure 3f. Figure 5 shows the estimated

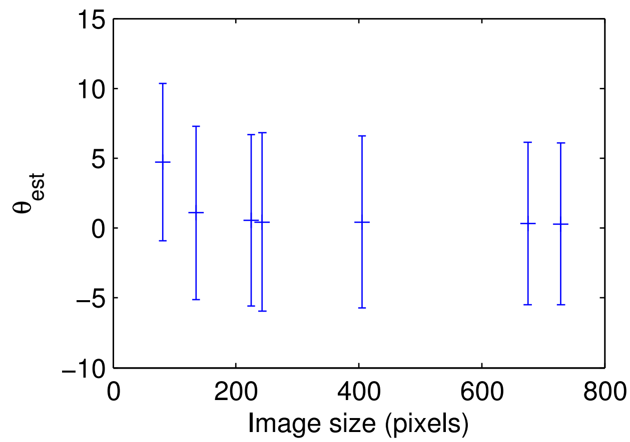

Figure 5: Variation in the estimated image orientation for the image in 3f as the image resolution is varied. The cross indicates the estimated roll angle and the error bars show the estimated error bound.

orientation angles obtained for each resolution for this image, along with the error bound on the estimate. Also shown is the variation in the error as a function of image resolution.

Even in this worst case, the performance of the estimate does not decay significantly with lower image resolution. Consequently it appears that considerable efficiency gains can be affected by working with low resolution input images. However, it is unclear whether this performance will persist in all environments. In particular one might expect that the performance of the Fourier based technique will be degraded if too much high frequency information is removed from the input images.

4. CONCLUSIONS AND FUTURE WORK

The combination of Fourier and Hough transform techniques is able to extract the dominant feature orientation from an image. Given that we expect that this direction will typically reflect the vertical direction in the world, this process can be used to infer the roll of an image with respect to its environment. The technique has been shown to work in a variety of indoor environments with different lighting conditions.

The next stage in the development of the algorithm requires careful characterisations of the error behaviour of each algorithm. At present the half-width half-maximum of the peaks of $g(\theta)$ and $h(\theta)$ are used as proxies for standard deviations of the two measurements. It is likely that these estimates are overly pessimistic, particularly in the case of the Hough transform technique, where figures 3 and 4 show much sharper peaks than reflected in the fitted quadratics. Comparison of the outputs of the two techniques with ground truth should allow for a better determination of the estimate variances and hence allow a less conservative estimate of roll angle.

This technique will then be combined with other orientation estimation techniques and deployed onto a robot for field testing in a variety of indoor and outdoor environments. In particular we will explore the roll angle estimation schemes in environments where one or other of the visual techniques breaks down due to a lack of suitable image features.

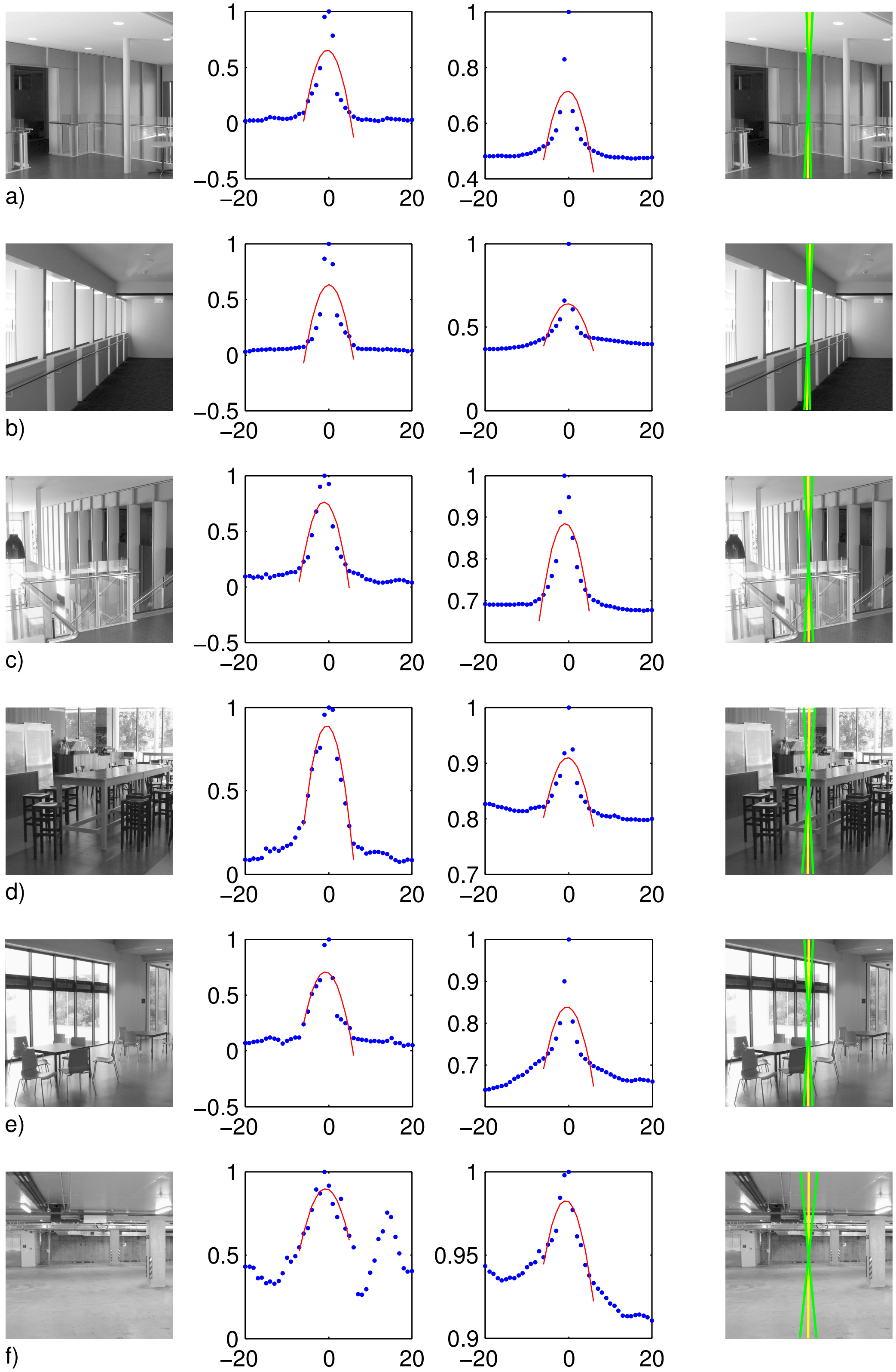

Figure 3: Demonstration of the orientation determining technique on number of different indoor scenes. For each scene, a) through f), the left column shows the input image, the second column is the estimate arising from the Fourier transform and the third column is the Hough transform technique. The x-axis in the middle two columns ranges over $\theta \in [-20..20]$, while the y-axis is normalised to the peak values of $g(\theta)$ and $h(\theta)$ respectively. The rightmost column is the initial scene with the estimated vertical direction superimposed through the centre of the image. The yellow line indicates the estimate of the vertical direction in the world, while the two flanking green lines indicate the error bound on the estimate.

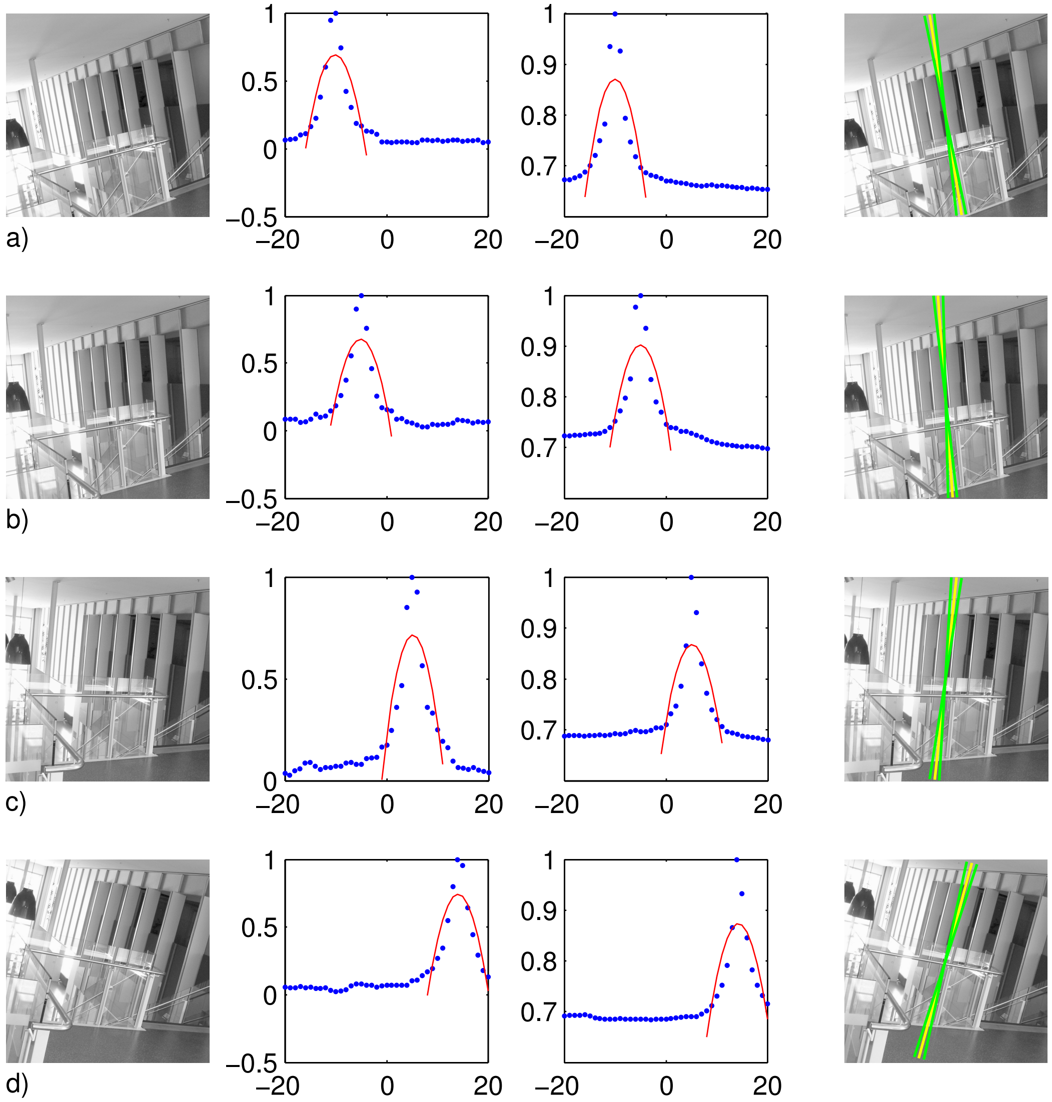

Figure 4: Demonstration of the orientation determining technique with the same scene captured in different orientations. The graphing conventions for each orientation, a) through d) are the same as those outlined for figure 3.

5. REFERENCES

[1] C. E. Bauby and A. D. Kuo. Active control of lateral balance in human walking. *Journal of Biomechanics*, 33(11):1433 – 1440, 2000.

[2] D. Coppola, H. Purves, A. McCoy, and D. Purves. The distribution of oriented contours in the real world. *Proceedings of the National Academy of Sciences*, 95(7):4002, 1998.

[3] R. O. Duda and P. E. Hart. Use of the Hough transformation to detect lines and curves in pictures. *Commun. ACM*, 15(1):11–15, 1972.

[4] E.R.Davies. *Machine Vision*. Morgan Kaufmann, 3rd edition, 2005.

[5] J. Illingworth and J. Kittler. A survey of the Hough transform. *Comput. Vision Graph. Image Process.*, 44(1):87–116, 1988.

[6] B. Li, M. Peterson, and R. Freeman. Oblique effect: A neural basis in the visual cortex. *Journal of Neurophysiology*, 90(1):204–217, 2003.

[7] R. J. W. Mansfield. Neural basis of orientation perception in primate vision. *Science*, 186(4169):1133–1135, 1974.

[8] S. Segvic and S. Ribaric. Determining the absolute orientation in a corridor using projective geometry and active vision. *Industrial Electronics, IEEE Transactions on*, 48(3):696 –710, jun 2001.

[9] D. Winter. Human balance and posture control during standing and walking. *Gait and Posture*, 3(4):193 – 214, 1995.

[10] E. Yacoub, N. Harel, and K. Uğurbil. High-field fMRI unveils orientation columns in humans. *Proceedings of the National Academy of Sciences*, 105(30):10607, 2008.