On the Merits of the Gaussian Mixture as a Model for Oriented Edgel Distributions

P. Assheton and A. Hunter

May 7, 2008

Abstract

The aim of this report is to establish the credibility of the Gaussian Mixture Model (GMM) as a model for the distributions of oriented edgels of rigid and biological objects in noisy images. This is tackled in two stages: first, the response of the Sobel filter to noisy pixels is analysed to show that the result holds for smooth rigid objects. Secondly, arguments are presented to support the proposition that the model can also effectively capture the added uncertainty introduced by natural shape variation, as found in images of biological objects. The result has particular application in the extension of the Generalized Hough Transform (GHT) to deformable shapes; in particular it offers a tailored and manipulable alternative to the non-parametric kernel density estimate used by Ecabert and Thiran in [3].

1 Introduction

In [3], Ecabert and Thiran propose a simple and efficient augmentation of the Generalized Hough Transform (GHT) [1] for the detection of mildly deformable shapes. Where GHT searches for instances matching a single template shape, Ecabert and Thiran learn a non-parametric probability distribution from a dataset of *similar* example shapes. In this vein, Ecabert and Thiran's approach replaces the *list* of possible centroid offsets in each row of the GHT R-table with a nonparametric representation for the *distribution* of such offsets. This allows for mildly deformable object shapes to be modelled, with a gradation in significance assignable to different centroid hypotheses.

In the following we argue that the nonparametric probability density estimate in such a model may be replaced by a semi-parametric Gaussian Mixture Model (GMM), bringing with it the various inferential advantages associated with a parameterised model. In section 2 we will use results derived in the appendix concerning the response of the Sobel filter to noisy images to prove our case for smooth rigid objects in noisy images. We subsequently argue in section 3 that the sort of shape variation found in biological shapes may also be well incorporated into the same model. In section 4 we will briefly outline work that is underway to exploit this result.

2 Gaussian Mixtures for Rigid Shapes

Given that there is i.i.d. Gaussian additive noise at each pixel, and that the true angle of the edge is θ, we show in appendix 4 that the angle φ retrieved by the sobel filter is also subject to approximately Gaussian uncertainty,

Figure 1: Edgels at similar orientations form short curve segments and sufficiently short curves approximate straight line segments. (a) An arbitrary shape. (b) Edgels extracted using Sobel filter. (c) Those edgels whose orientations lie between 0.6 and 1.1 radians from vertical.

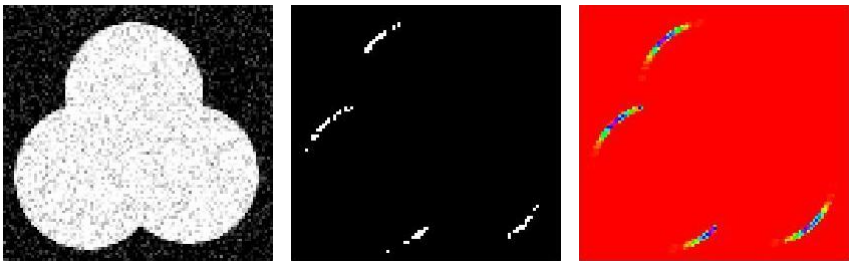

Figure 2: (a) Fig 1 with added noise, $\sigma = .25$, (b) edgels with orientations $\varphi \in (0.6, 1.1)$ and (c) the sum of 1000 samples as in (b) shows that the probabilistic structure is of a bivariate Gaussian bent to fit the curve.

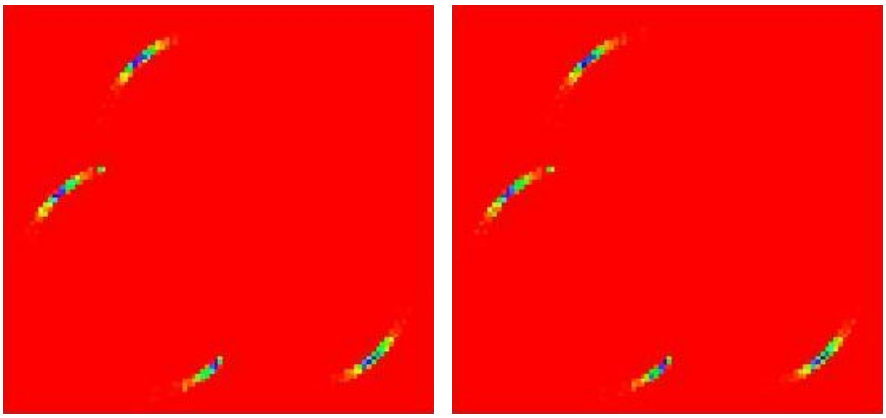

Figure 3: Spatial distribution of edgels falling within a given angular bin, generated as in fig 2c. The angular bins are (a) $\varphi \in (0.8, 0.9)$ and (b) $\varphi \in (0.849, 0.851)$. The images are almost identical.

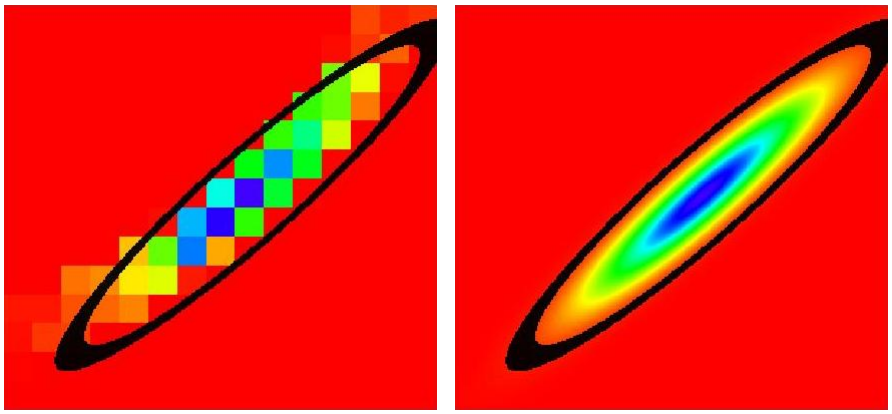

Figure 4: (a) Closer view of Gaussian-like shape from fig 3a, compared to (b) bivariate Gaussian with same statistics.

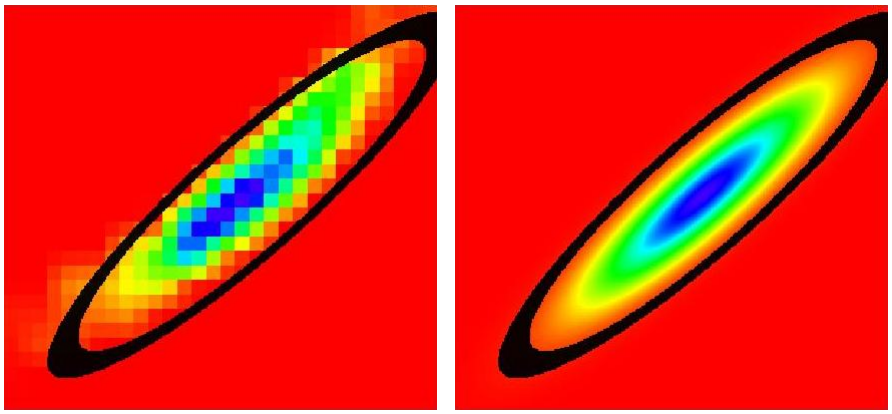

Figure 5: (a) Fig 4a bilinearly interpolated to double resolution and (b) bivariate Gaussian with same statistics.

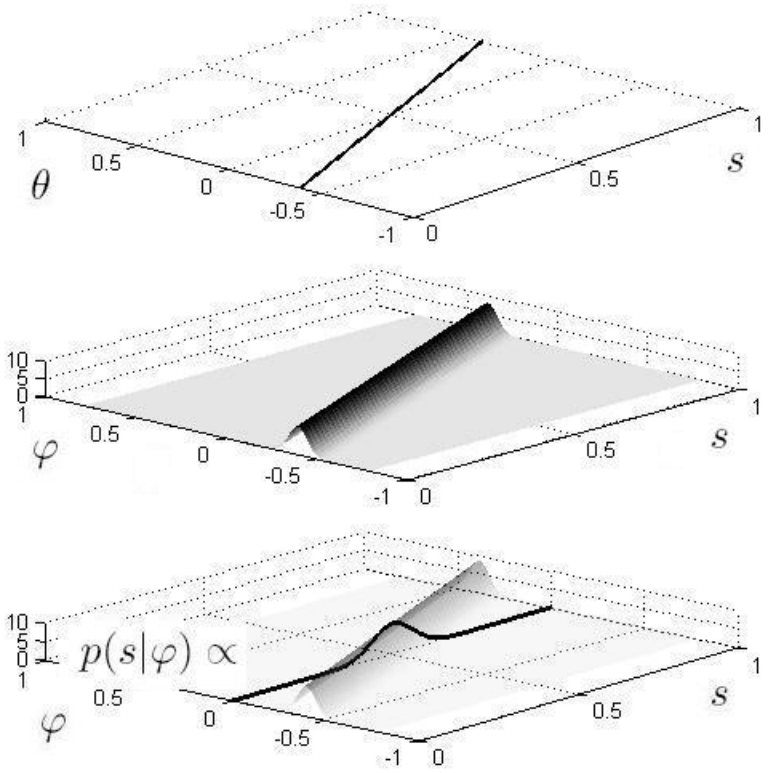

Figure 6: From true edge orientation to perceived edge orientation: (a) edge orientation $\theta(s)$ along the length of a segment of constant curvature is linear in the distance s along the line, (b) probability $p(\varphi|s)$ of 'perceiving' an edge of a given orientation at location s along the edge is Gaussian and (c) overlaid onto b, the likelihood of location along the line given a perceived edge orientation is proportional to another Gaussian.

$$p(\varphi|\theta) \approx \mathcal{N}(\varphi; \theta, \sigma_\theta^2) \tag{1}$$

where $\sigma_\theta \approx .476\sigma/k$, for pixel noise variance σ^2 and edge contrast (i.e. gradient magnitude) k.

Then, assuming the contrast along a given edge is roughly constant, and that pixel noise is uniform across the image, the variance σ_θ^2 in the retrieved edge angle for a given edge is also constant. We shall assume this to be the case. We will also assume that curves of natural objects are typically sufficiently smooth that within a local neighbourhood they may be considered to have constant curvature κ.

For this local neighbourhood, then, the relationship $\theta(s)$ between arc-length index s, and orientation θ of the curve edge, is linear

$$\theta(s) = \kappa s + \theta_0 \tag{2}$$

as plotted in fig 6a for a curve of length 1, $s \in (0, 1)$, with curvature $\kappa = \frac{\pi}{4}$ and initial orientation $\theta_0 = -\frac{\pi}{8}$.

We may combine (2) with (1) to obtain an expression for the probability of retrieving the angle φ using the sobel filter, at any point s along the line:

$$p(\varphi|s) = p(\varphi|\theta(s)) = \mathcal{N}(\varphi; \kappa s + \theta_0, \sigma_\theta^2) \tag{3}$$

which is plotted in fig 6b.

We assume that each point on the curve is as likely to appear unoccluded and undisturbed by noise as every other point. Then, the prior distribution $p(s)$ is uniform. Furthermore, we assume an object's contour to be a closed curve, so we may expect that the *uniform* distribution $p(s)$ has support at all values for which $p(\varphi|s)$ is significant. Then, by Bayes' Theorem,

$$p(s|\varphi) \propto p(\varphi|s) \tag{4}$$

This in itself does not prove that $p(s|\varphi)$ is a Gaussian *for given* φ. However, the fact that $p(\varphi|s)$ is, for given s, Gaussian with constant variance and a *mean that is linear in* s does, as is illustrated in fig 6c. We prove this syntactically by swapping the independent variable with the mean of the Gaussian (3), as is permitted by the symmetry of the Gaussian, followed by a change of coordinates, so that s becomes the independent variable:

$$\begin{aligned} p(\varphi|s) = \mathcal{N}(\varphi; \kappa s + \theta_0, \sigma_\theta^2) = \mathcal{N}(\kappa s + \theta_0; \varphi, \sigma_\theta^2) = \\ \mathcal{N}(\kappa s; \varphi - \theta_0, \sigma_\theta^2) \propto \mathcal{N}\left(s; \frac{\varphi - \theta_0}{\kappa}, \left(\frac{\sigma_\theta}{\kappa}\right)^2\right) \end{aligned} \tag{5}$$

Since the expression at the right hand side sums to unity over s and is proportional to $p(\varphi|s)$ (by (5)), then by (4) this expression gives us $p(s|\varphi)$

$$p(s|\varphi) = \mathcal{N}\left(s; \frac{\varphi - \theta_0}{\kappa}, \frac{\sigma_\theta^2}{\kappa^2}\right) \tag{6}$$

Of course, while curvature might be approximately constant across a small stretch of the curve, it will in general (for non-circular objects) not be so along the whole curve. In fact, the pattern of curvature changes in a distinctive shape is, by definition, likely to be quite individual to that shape. This is the reason that we use a separate model for each angular bin in our algorithm, so that any arbitrary shape may be modelled. If our angular divisions are sufficiently fine, we may approximate that the distribution of edgels along a given curve segment,

$$p(s|\varphi \in (\varphi_{min}, \varphi_{max})) \propto \int_{\varphi_{min}}^{\varphi_{max}} p(\varphi|s) d\varphi \tag{7}$$

for each angular bin $(\varphi_{min}, \varphi_{max}]$ also as Gaussian.

The justification for this approximation is fairly trivial, but it may be formalised as follows. Since we are using quite narrow bins for φ, we approximate the function $p(\varphi|s)$ as linear between φ_{min} and φ_{max} for any given bin[1]. This then gives

$$p(s|\varphi \in (\varphi_{min}, \varphi_{max})) \propto p(\bar{\varphi}|s) \tag{8}$$

where

$$\bar{\varphi} = \frac{\varphi_{max} - \varphi_{min}}{2} \tag{9}$$

so that

$$p(s|\varphi \in (\varphi_{min}, \varphi_{max})) = p(s|\varphi)|_{\varphi=\bar{\varphi}} \tag{10}$$

which is simply the Gaussian $p(s|\varphi)$, with φ as the central value of the interval $(\varphi_{min}, \varphi_{max})$.

2.0.1 An Example

This is illustrated in fig 2. Fig 2a shows the effect of adding Gaussian noise with standard deviation $\sigma = .15$ to the clean image in fig 1a. Edgels were extracted from this image by thresholding the gradient magnitudes from the sobel filter. These gradients were also used to calculate the orientation of each identified edgel. Those edgels of a given orientation $\varphi \approx .85$ (specifically, those for which $\varphi \in (0.6, 1.1)$) are shown in fig 2b. In order to illustrate the distribution of these edgels, we generated 1000 different noisy images, each as per fig 2a. We then extracted from each the edgels for which $\varphi \in (0.6, 1.1)$, as in fig 2b. Finally, we summed up these edge-images, to yield fig 2c. This result verifies our calculation that the distribution of edgels should take the form of a Gaussian wrapped along the curve around each point where the curve hits $\theta = .85$.

2.1 Limiting Distribution is Gaussian Mixture Model

The interval $\varphi \in (0.6, 1.1)$ is quite large, and was used for visualization purposes. If we could reduce the variance along the curve (and, hence the length) of these distributions sufficiently, the curves

[1]This approximation is helped by the fact that, as the maximum entropy distribution, the Gaussian is the smoothest possible (least curved) probability distribution.

would start to take on linear form, as is demonstrated in fig 1. Since this is a large interval, the linear assumption underlying the approximation of (7) with (8) is quite weak, and it is possible to reduce the variance of (7) by reducing the width of the angular bins.

As the interval $\Delta\varphi = \varphi_{max} - \varphi_{min}$ is decreased, two things happen: first, the linear assumption behind (8) becomes increasingly strong so that the distributions on the curve segments become even more Gaussian-like. Secondly, these 'Gaussian' edge segments become shorter and so take on a more linear form. Let us assume that pixel noise is not sufficiently large to prevent this occurring. Then, bearing in mind that these curve segments also have width (since they represent a *ramp* in intensity values), the result is the following: *as the bin-width approaches zero, the distribution of edgels from a given line, whose orientation is within a given bin, approaches the bivariate Gaussian distribution. Taking into account the multiplicity of these segments, the probability distribution of the edgels from a particular object falling within a particular angular bin approaches the bivariate Gaussian mixture model.*

2.1.1 Further Illustration

This is illustrated in fig 3. Fig 3a shows the spatial distribution of edgels with a tighter orientation interval, $\varphi \in (0.8, 0.9)$. In fig 3b, we present the same distribution for an interval that is one twentieth the size, $\varphi \in (0.849, 0.851)$. The distributions are almost identical. This illustrates that by the time the interval is 0.1 radians wide, the distribution has almost converged, and no further benefit may be obtained by modelling to finer precision. This, in turn, supports our assumption that (7) may be approximated by (8), since the former converges upon the latter as angular bin width approaches zero. This is also shows that the inferiority of the discretized representation of edge orientation, compared with its true continuous value, is negligible.

In fig 4, we present a close up of one of the Gaussian-like shapes in the distribution, along with a bivariate Gaussian with the same statistics. The approximation is quite good. However, due to the small pixel-width of an edge, the picture is highly pixellated. In fig 5 we use bilinear interpolation to double the resolution of fig 4, and the match to the Gaussian distribution again appears to be a good one.

Since we expect to find our edgels for each angular bin distributed according to a GMM, we may model the entire set of edgels in the object by a set of GMMs–one GMM per angular bin. This is the approach that we take.

2.1.2 Comparison to Variable Generalized Hough Transform

The material in this section draws close comparison to Morimoto and Suenaga's Variable Generalized Hough Transform [9], for which they also calculate the appropriate spread in the voting kernel, given uncertainty in the retrieved edge-orientations. However, Morimoto and Suenaga make no assumptions as to the method of edge retrieval nor of the nature of the noise. To avoid such assumptions they assume instead that the orientation likelihood is uniformly distributed between two bounds. In contrast, we have assumed here that the Sobel filter is used for edge extraction and that noise is manifested as i.i.d. Gaussian noise in each pixel. This neatly leads to a justification for our use of the mixture of Gaussian Mixture Models in place of the R-table.

3 Gaussian Mixture Models for Natural Shapes

Having established in section 2 that an appropriate model for rigid shapes in GHT is the Gaussian Mixture Model, we now argue that the same model can also capture small amounts of natural variation, such as that found between different leaves of a given species. We will also argue that such variations are the same as small local variations within more complex objects such as humans. As such, we claim that the Gaussian Mixture Model in this form provides the ideal model for a prototype object shape, as used in shape clustering.

3.0.3 Gaussians in Biology

In order to produce tight voting distributions, the contours on all training exemplars should be fairly similar (if this is not the case, as for example with articulated objects, more sophisticated techniques must be employed, as explored later in this thesis). Then we assume that each oriented edge section exists in all data examples, at a similar location in each, though displaced by a small random offset, as defined by natural variation. It is common to assume that there are a large number of modes of such natural variation, each corresponding to a different gene or environmental factor. Suppose we assume these each to have the effect of an independent linear translation on each edge. Then, central limit theorem dictates that any one-dimensional slice of the distribution of their combination approaches Gaussian; again we arrive at the bivariate GMM as our model, subsequent to natural variation.

Indeed, the idea that the cumulative, additive effect of multiple randomly selected and independent genes through the central limit theorem should yield a Gaussian distribution for such biological variation is quite a common one. Since Quetelet introduced the notion of the Gaussian distributed 'average man' in the 19th century [12], the Gaussian has been used to model such biological modes of variation as human height [4] and chest circumference [11] and monkey hair length [7].

More recently, however, it has been argued that genes cannot be considered to contribute independently and additively to a given parameter, but rather cooperatively and acceleratively [2]. The conclusion is typically that such parameters should be modelled as log-normal, including fruit size [6] and weight [13], and flower size within a given species [10].

Regardless, if our model is to be any use in its context then the voting distribution must fit fairly tightly to the boundary; then, the standard deviation of such measurements as leaf length must be small in comparison to their mean, in which case the log-normal distribution approximates well the Gaussian.

3.1 The Variation of *Oriented* Edgels

At this point we must admit that the above argument, based on findings in univariate measurements in biology, cannot account for such two-dimensional behaviour in our edge-segments as rotation (or indeed, any tranformation other than translation). However, our model is not of specific edge segments, but of sections of edge of a given orientation. As such, when the underlying edge undergoes rotation, the section that is at a given orientation "slides" along the edge. In fact, upon closer inspection, we shall see that the effects of rotation may indeed be approximated as linear transformation in the edge segments, so that our argument still holds.

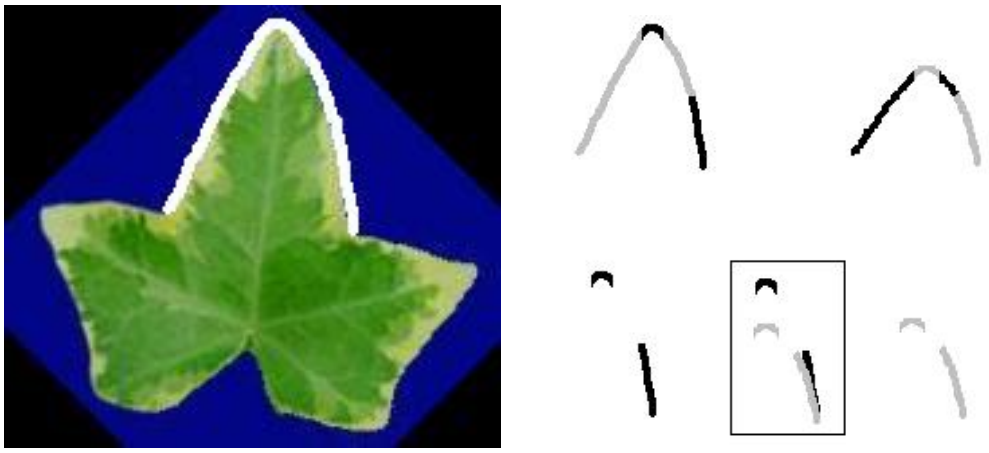

Figure 7: (a) An ivy leaf with one lobe outlined and (b) the effects on the edge segments of shortening the outlined lobe consists in a translation for edgels at the tip and rotation for edgels at the base.

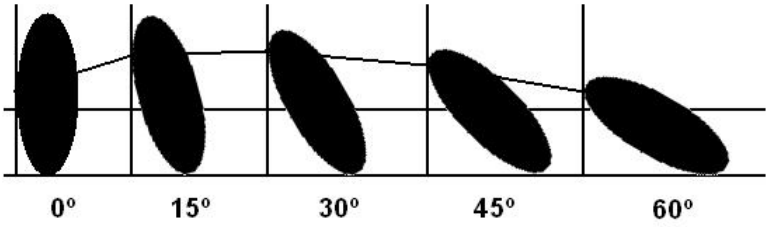

Figure 8: The motion of oriented edge segments compared with the motion of the shape. The edge segment that is vertical in the left hand image moves consistently downwards as the ellipse is rotated. But the edge segment that is perceived as vertical in each frame first moves up, and then down.

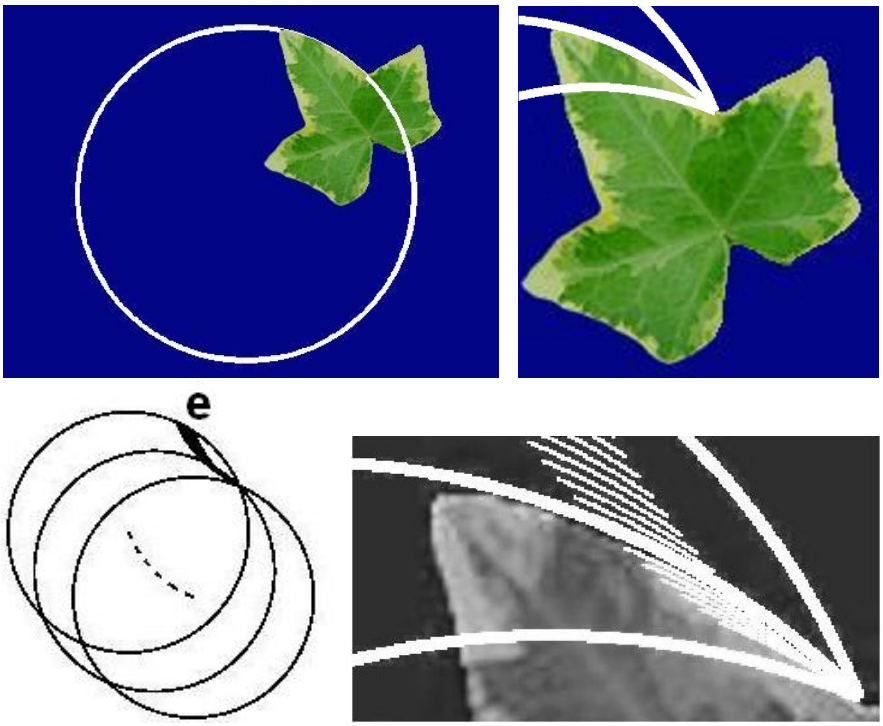

Figure 9: (a) Locally constant curvature assumption approximates the local curve with its ocsulating circle; (b) three possible degrees of rotation of this curve around a given pivot; (c) the interpretation of these rotations in terms of the osculating circle; the dotted line indicates the motion of the osculating circle centre, which moves parallel to the translation of the oriented edge segment e; and (d) a closeup of the motion of oriented edge segment e superimposed on the leaf that it models; the motion of e as translation is quite clear.

3.1.1 Translation and Rotation

We shall suppose that a natural mode of deformation may be modelled, with respect to its effect on a given curve segment, as a combination of a translation of the segment along a given line and a rotation of the curve segment relative to some pivot. For example, consider the typical variation in a lobe of an ivy leaf, as outlined in fig 7a. It is quite common for the angle of the whole lobe to vary slightly, in the extreme leading to 'wonky' looking leaves. This, then, represents rotation of the entire curve with respect to some pivot. An alternative mode of variation occurs because some lobes are shorter than those on other leaves. In this case, the variation must be considered as different on different segments of the lobe's contour. The tip, for example, will experience a translation towards the base of the lobe, while the sides near the base will experience just a slight rotation. These two motions are illustrated in fig 7b. Sections of the curve between the two will experience a linear mixture of rotation and translation, according to their location on the curve.

Since these form a smooth continuum, we approximate by assuming that the transform applied by a given mode of variation along a small section of the curve is constant, and comprising these two components (rotation and translation). The small sections we consider are those identified through binning edges by orientation, as discussed in section 2.

3.1.2 The Effect of Rotation on Oriented Edge Segments

The motion of these sections under rotation of their local curve requires some consideration, since rotation leads to a change in orientation of edges. This, in turn, means that the section of edge identified as falling into a given bin slides along the curve as it is rotated. An example is shown in fig 8. Here, we concentrate on the edge section, at the left hand side of the ellipse, that is vertical. For simplicity of demonstration, we focus on the vertical coordinate y of this edge section. The motion of the *section that is vertical before rotation* is monotonically downward, as the shape is rotated. However, the *segment, in each frame, that is vertical* first moves upwards, and then downwards, as demonstrated by the line y.

However, assuming that we can model the curve as of locally constant curvature (an assumption we have already made in section 2), the following insight allows a simple explanation of the pattern of motion of these edges. By assuming that the curvature is constant, we approximate our curve with its osculating circle, as illustrated in fig 9a. We note that the edges of a circle that are at a given orientation have a very simple relationship with the centre of the circle: they are always at the same distance and angle (and hence the same spatial offset) from the centre and, as such, when the circle is subject to rigid motion, the two will possess the same motion vector. So we may understand the motion of our oriented edge segments as the motion of the centroid of the osculating circle. In the case of rotation around a pivot other than the circle centre, this implies translation of the oriented edge in a circular motion. *Note that the oriented edge does not itself rotate, as its orientation is fixed*; rather, it is dragged at its set orientation around a circular path. This is illustrated in figs. 9c and 9d for the edge rotations overlaid on fig 9b; the motion of the osculating circle centre (shown as a dashed line in fig 9c) is parallel to that of the oriented edge segment e.

The phase and centroid of the circular motion of the oriented edge may differ from those of the rotation of the underlying curve; but the angular motion around its circular path is the same. We assume that in most natural non-articulated objects (such as ivy leaves) the angle by which typical deformation may rotate an edge segment is likely to be small enough that the arced translation in the oriented edge may be approximated as linear.

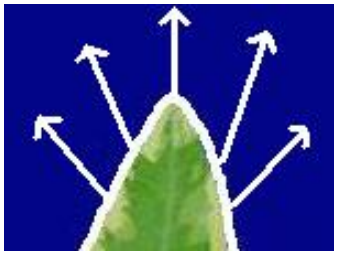

Figure 10: The mode of natural variation illustrated in fig. 7 may be modelled as linear translations of the oriented edge segments.

3.1.3 Incorporating Translation

At the start of this section we noted that the typical deformations in natural objects can be seen, locally to a given curve segment, as a combination of a rotation and a translation. We showed that the rotation element could be approximated in many cases as linear translation. This may then be additively combined with the translation element experienced by the segment, to yield a linear trajectory for each oriented curve segment, as illustrated in fig 10.

3.2 Entropy: An Alternative Justification

A rather nice alternative justification for using the Gaussian distribution is presented by Jaynes [8] in his famous book on probability. We must ask ourselves, how much of the structure present in the dataset represents the distribution of the population, and how much represents the individual characteristics of the sample members? The answer is that we do not know. We can expect, however, that we are able to extract a covariance matrix and a mean from the data that well represent those of the population. If this is the only information of which we can be certain, then the Gaussian distribution, being the maximum entropy distribution for given (co)variances, is the best representation of the information that we have available and, as such, represents precisely the level of certainty that we have in the likely locations of edgels in a novel image[2].

The maximum entropy property of the Gaussian has a potential further advantage. Peaks in a smooth vote surface may be identified with a lower sampling frequency (this is one of the major motivations for the use of the distance transform to blur the edge image in Chamfer matching techniques). Since the Gaussian is the highest entropy probability distribution, this also makes it the smoothest possible vote kernel.

3.3 Prototypes for Articulated Natural Objects

We make one final note in this section that, for those objects whose typical variation is too great to be efficiently modelled in this way, we may use a GMM as described to model a subset of the configurations typical to that object. An example might be the limbs of a walking pedestrian; the representation of the "modes of variation" as translations and rotations still applies, so that breaking the total variation up into a number of tighter models, each model becomes small enough to

[2]We note, however, that this is based on Bayesian reasoning about "certainty" rather than frequency and, as such, may be unpopular with some mathematicians.

fit our representation. Consider the boundaries of the region over which each GMM has significant probability mass. Taking the union across the set of GMMs, those boundaries over which continuous movement passes must overlap between the different models. Then it is no problem if we do not expect the configurations of a moving leg to be Gaussian distributed; it is only necessary to assume Gaussian variation at the boundaries of the *union* of the set of densities–which are likely to be those boundaries representing small biological variations, rather than those influenced by motion.

4 Conclusions and Future Work

In the above we have presented arguments for the use of the Gaussian Mixture Model in describing the distributions of oriented edgels detected by the Sobel filter in noisy images. We first studied smooth curves of rigid objects, before extending our analysis to incorporate the variation between biological specimens of the same species.

This result is currently being explored in two particular applications. In the first, we replace the non-parametric distributions employed by Ecabert and Thiran with GMMs (as described in section 1). Investigations are taking place into different ways of exploiting the semi-parametric nature of this alternative distribution. Particular areas of interest are: training based on a small dataset, optimisation of the alignment and scaling of the data exemplars as an integral part of EM training, and adjustment for the increase in spurious peaks generated by hypotheses of small-scaled instances.

In the second application, we consider the use of density networks to define a wider range of object deformation. This becomes possible through the results of this paper, since a density network can be used to define the parameters to a GMM, but cannot effectively define a nonparametric distribution of the form used by Ecabert and Thiran.

References

[1] D.H. Ballard. Generalizing the hough transform to detect arbitrary shapes. *Pattern Recognition*, 13:111–122, 1981.

[2] E.M. East. Inheritance of flower size in crosses between species of nicotiana. *Botanical Gazette*, 55(3):177–188, March 1913.

[3] O. Ecabert and J.-P. Thiran. Adaptive hough transform for the detection of natural shapes under weak affine transformations. *Pattern Recognition Letters*, 25(12):1411–1419, 2004.

[4] Cristina Esteban, Laura Aud, Antonio Carrascosa, Mnica Fernndez-Cancio, Annalisa Prez-Arroyo, Angels Ulied, Pilar Andaluz, Rosa Arjona, Marian Albisu, Mara Clemente, Miquel Gussiny, and Diego Yeste. Human growth hormone (gh1) gene polymorphism map in a normal-statured adult population. *Clinical Endocrinology*, 66(2):258–268, February 2007.

[5] R.C. Geary. The frequency distribution of the quotient of two normal variates. *Journal of the Royal Statistical Society*, 93(3):442–446, 1930.

[6] B.H.A. Groth. The golden mean in the inheritance of size. *Science*, 39:581–584, 1914.

[7] Haruhisa Inagaki. A preliminary study on hair length in the japanese monkey (macaca fuscata fuscata). *Primates*, 26(3):334–338, July 1985.

[8] E.T. Jaynes. *Probability Theory: The Logic of Science*, chapter 7, The Central Gaussian, or Normal, Distribution, pages 701–732. Cambridge University Press, 2003.

[9] Masashi Morimoto and Yasuhito Suenaga. Variable generalized hough transform based on error analysis of curve gradient. *Systems and Computers in Japan*, 31(2):19–28, 2000.

[10] L. Powers. The nature of the interaction of genes affecting four quantitative characters in a cross between hordeum deficiens and h. vulgare. *Genetics*, 21:398–420, 1936.

[11] Adolphe Quetelet. Statement of the sizes of men in different counties of scotland, taken from the local militia. *Edinburgh Medical and Surgical Journal*, 13:260–264, 1817.

[12] Adolphe Quetelet. *A Treatise on Man and the Development of His Faculties*. Scholars Facsimilies & Reprint, 1842. ISBN 0-8201-1061-2.

[13] E.W. Sinnot. The relation of gene to character in quantitative inheritance. *Proceedings of the National Academy of Sciences*, 23:224–227, 1937.

Appendix: Modelling the Noise of the Sobel Filter

We demonstrate below that, assuming independent Gaussian noise at each pixel, up to a limited standard deviation, the edge orientations retrieved by the Sobel filter are subject to approximately Gaussian uncertainty. We model an edge as a discretized ramp in intensity values. We define the contrast k at a pixel as the magnitude of the gradient. If the true edge angle relative to vertical is θ the gradients in the x and y directions are given by $\delta_x = k \sin\theta$ and $\delta_y = k\cos\theta$.

The Sobel filter approximates four times the central difference across the pixel taken along a line of length 2 pixels, so to obtain an approximation to the gradient (δ_x, δ_y) and hence k we divide the output of our Sobel filter by $2 * 4 = 8$. We may then estimate θ using:

$$\theta = \operatorname{atan}\frac{\delta_x}{\delta_y} \tag{11}$$

Without loss of generality, assume that $\theta \in (-\pi/4, \pi/4)$. Also assume that the pixel noise is additive zero mean Gaussian, standard deviation $\sigma \leq 0.8k$ on the edge in question. In this case the differences calculated by the Sobel filter are uncorrelated, with $\sigma_{sobel} = \sigma\sqrt{12}$, as shown below. We can approximate the gradient using:

$$\Delta_x = \delta_x + \varepsilon_x \tag{12}$$

$$\Delta_y = \delta_y + \varepsilon_y \tag{13}$$

where Gaussian additive noise ε_x and ε_y has standard deviation $\sigma_x = \sigma_y = \sigma_{sobel}/8 = \sigma\sqrt{12}/8$.

The Sobel filter effects a linear transformation on the nine pixels within its kernel area. We vectorize this 3x3 matrix of pixels, and treat it as a multivariate Gaussian distribution $\mathbf{f} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2 I)$ where I is the 9x9 identity matrix and $\boldsymbol{\mu}$ represents the true values of the pixels' intensities, before noise. The Sobel filter is then represented by the matrix:

$$S = \begin{pmatrix} -1 & -2 & -1 & 0 & 0 & 0 & 1 & 2 & 1 \\ -1 & 0 & 1 & -2 & 0 & 2 & -1 & 0 & 1 \end{pmatrix} \tag{14}$$

so that

$$\mathbf{\Delta} = \begin{pmatrix} \Delta_x \\ \Delta_y \end{pmatrix} = S\mathbf{f} \tag{15}$$

Then, $\mathbf{\Delta}$ is drawn from a multivariate normal distribution $\mathbf{\Delta} \sim \mathcal{N}(S\boldsymbol{\mu}, \mathbf{\Sigma}_\Delta)$, with covariance matrix $\mathbf{\Sigma}_\Delta$:

$$\mathbf{\Sigma}_\Delta = S(\sigma^2 I)S^T = \sigma^2 \begin{pmatrix} 12 & 0 \\ 0 & 12 \end{pmatrix} \tag{16}$$

Hence, Δ_x and Δ_y are uncorrelated, each having standard deviation $\sigma_x = \sigma_y = \sigma\sqrt{12}$. Further, the ratio r of these follows a normal ratio distribution:

$$r = \frac{\Delta_x}{\Delta_y} = \frac{\delta_x + \varepsilon_x}{\delta_y + \varepsilon_y} \tag{17}$$

While this distribution has a complex analytical form, Geary [5] proved that, provided that δ_y is unlikely to be negative (he suggests $\delta_y \geq 3\sigma_y$) and that ε_x and ε_y are uncorrelated,

$$t = \frac{\delta_y r - \delta_x}{\sqrt{\sigma_y^2 r + \sigma_x^2}} \tag{18}$$

has a standard normal distribution. From the definition of δ_y and the range limits of θ, we have that $\delta_y \geq 0.707k$. Then δ_y is always at least $0.707k/\sigma_y$ standard deviations above zero. In the worst case, $\sigma = k$; this gives 1.63 standard deviations, which corresponds to 95% confidence that Δ_y is above zero. We typically expect considerably lower image noise than this leading to higher confidence that $\delta_y > 0$. Then,

$$\tilde{\varphi} = \frac{1.1\sigma_x}{k} t + \theta \tag{19}$$

is also normally distributed, with mean θ and standard deviation $1.1\sigma_x/k$. Some algebraic manipulation shows that:

$$\tilde{\varphi} = 1.1 \frac{r\cos\theta - \sin\theta}{\sqrt{r^2 + 1}} \tag{20}$$

The angle θ may be estimated using $\varphi = \text{atan}(r)$. We show below that for likely values of r, $\varphi(r) \approx \tilde{\varphi}(r)$. Then for the significant part of its distribution φ is distributed as $\tilde{\varphi}$; approximately speaking, $\varphi \sim \mathcal{N}(\theta, (1.1\sigma_x/k)^2)$. Hence the distribution of φ for $\theta \in (-\frac{\pi}{4}, \frac{\pi}{4})$ is well approximated by a Gaussian, centred on θ, with variance depending only on image noise σ and edge contrast k:

$$p(\varphi|\theta) \approx \mathcal{N}(\varphi; \theta, \sigma_\theta^2) \tag{21}$$

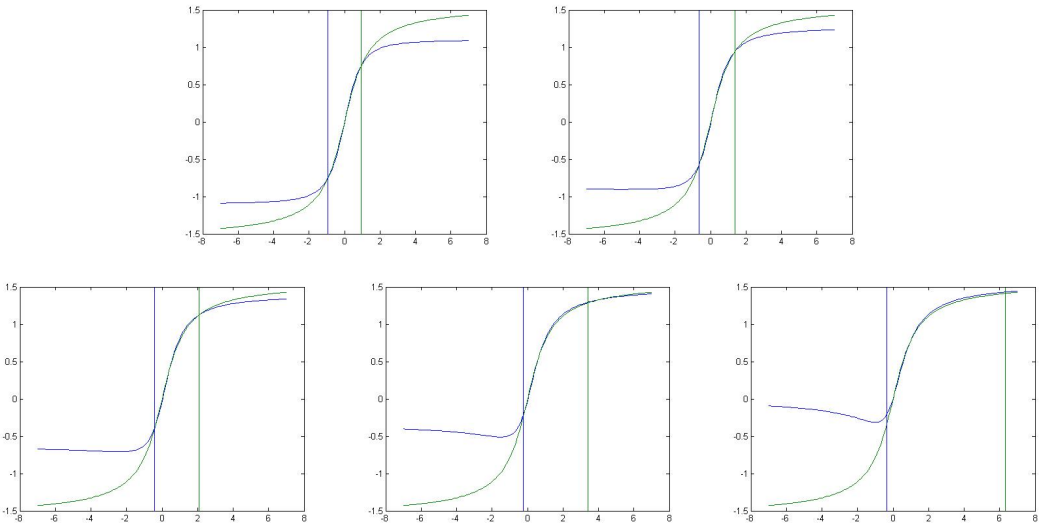

Figure 11: $\tilde{\varphi}(r)$ and $\varphi(r) = \text{atan}(r)$ for a range of values of θ. Bars represent 95% confidence intervals for r, for given θ, and where $\sigma = k$: $\theta = 0, \frac{\pi}{16}, \frac{2\pi}{16}, \frac{3\pi}{16}$ and $\frac{4\pi}{16}$. The graphs for negative θ are identical but mirror image.

where

$$\begin{aligned} \sigma_\theta &= 1.1\sigma_x/k \\ &= 0.476\sigma/k \end{aligned} \tag{22}$$

In fig 11, we show that, for likely values of r, $\varphi \approx \tilde{\varphi}$. Each graph plots atan$(r)$ alongside $\varphi(r)$ for a particular value of θ, as given by (18) and (19). The vertical bars show 95% confidence intervals for $\sigma = .8k$, generated using a million samples drawn from the ratio distribution of r; the approximation is excellent in this range.

The generalisation from $\theta \in (\frac{-\pi}{4}, \frac{\pi}{4})$ to any θ is relatively straightforward. We first extend the range to include $\theta \in (\frac{\pi}{4}, \frac{2\pi}{4})$, by swapping the δ_x and δ_y values. The effect this has is of turning the ratio r upside down to get r^{-1}. These r^{-1} values behave precisely as the r values for $\theta \in (\frac{-\pi}{4}, \frac{\pi}{4})$, so that atan$(r^{-1}) = (\frac{\pi}{2} - \theta)$ is also normally distributed as discussed. Due to the symmetry of the Gaussian function, this implies that θ is normally distributed with the same variance. Finally, generalising to angles in the other half plane may be achieved by noting that for each angle θ in that half plane, the angle $\theta^{opp} = \theta + \pi$ is in the opposite half plane and has the opposite gradient; that is, $\delta_x^{opp} = -\delta_x$ and $\delta_y^{opp} = -\delta_y$. Then the ratios for these two angles are both equal $r^{opp} = r$ and identically distributed. So the same mapping may be used for these angles to yield a normal distribution.