THE HOUGH TRANSFORM'S IMPLICIT BAYESIAN FOUNDATION

Neil Toronto, Bryan S. Morse, Dan Ventura, Kevin Seppi

Brigham Young University
Department of Computer Science
Provo, UT, USA 84602

ABSTRACT

This paper shows that the basic Hough transform is implicitly a Bayesian process—that it computes an unnormalized posterior distribution over the parameters of a single shape given feature points. The proof motivates a purely Bayesian approach to the problem of finding parameterized shapes in digital images. A proof-of-concept implementation that finds multiple shapes of four parameters is presented. Extensions to the basic model that are made more obvious by the presented reformulation are discussed.

***Index Terms*—** Image processing, Machine vision, Hough transforms, Bayes procedures

1. INTRODUCTION

The Hough transform is a standard industry workhorse for finding parameterized shapes in digital images [1, 2]. (See [3] for a comprehensive survey.) It is image-global and robust to noise and occlusion. However, discretizing the parameter space results in two fundamental problems. First, it creates a trade-off between precision and sensitivity to noise, making selection of discretization level critical to correct convergence [4]. Second, the time and space required scales exponentially with additional parameters, limiting it to very simple parameterizations. Much of the research in the Hough transform concentrates on sidestepping this second problem for specific shapes or applications.

A continuous-parameter formulation overcomes both of these problems but leaves the problem of searching the parameter space. A continuous-parameter *Bayesian* formulation immediately suggests a search algorithm (Markov chain Monte Carlo sampling), suggests useful extensions to the basic algorithm, and provides a good framework for analyzing some of the Hough transform's properties. In this paper, we present a simple continuous-parameter Bayesian formulation and prove that it is equivalent to the Hough transform. We then build a model of multiple circle detection with deformations (a four-parameter problem) and use it to find blood platelets to demonstrate the viability of the Bayesian approach. We conclude with a discussion of extensions that the new formulation makes obvious.

2. RELATED WORK

Since the introduction of the Hough transform into computer vision [2], most research into the Hough transform has focused on algorithms and implementation [3]. Some new algorithms that result from injecting Bayesian methods into the Hough transform have been quite successful [5, 6, 7, 8]. Our work differs from these primarily in that we prove that the Hough transform is already a Bayesian method.

Very little Hough transform research has focused on understanding its behavior [9, 10, 11]. Most that have done so consider only the straight line fitting case and note similarities to well-known M-estimators.

Our work is closely related to Bayesian deformable models [12] and Bayesian dynamic contours [13]. The addition of the Hough transform to the ever-growing suite of Bayesian techniques allows it to be combined with many of them in well-grounded, principled ways.

3. PROOF OF EQUIVALENCE

3.1. Definitions

The objective is to find a parameterizable shape in an $m \times n$ digital image.

A *parameterization* of a shape is an assignment of variables that defines the shape. A parameterization is always defined in the context of a function that produces the shape or an equation that defines the shape. (Note that this is general enough to include the generalized Hough transform [14].)

A *feature set* F is the input to the Hough transform. We will assume that features are found by running an edge detector over the image and placing the x, y coordinates of the detected edges into F. In this paper, $k = |F|$. Without loss of generality, this proof can be extended to deal with any kind of features rather than just the coordinates of detected edges—for example, most Hough transforms incorporate some notion of edge orientation.

Generally, the Hough transform proceeds as follows:

1. Collect image features into F,
2. Create an accumulator of discrete parameterizations,
3. For each feature, increment the accumulator cell of every parameterization that could have produced it,
4. Low-pass filter the accumulator to help mitigate discretization issues, and
5. Find local maxima in the accumulator.

The local maxima found in the last step are regarded as candidates for parameterizations of shapes likely to have produced the feature set F.

The accumulator includes only parameterizations that are judged to be feasible—for example, for circles of radius $r = 12$, the accumulator's domain might be $x_c \in [-12, m+12]$, $y_c \in [-12, n+12]$. This will be referred to as the parameterization's *feasible domain*.

email: ntoronto@cs.byu.edu, morse@cs.byu.edu, ventura@cs.byu.edu, kseppi@cs.byu.edu

A rather more mathematical formulation of the Hough transform is approximately equivalent. Let $\mathbf{s}$ be a parameterization. Construct a discrete implicit *voting function* v:

$$v(\mathbf{s}, x_i, y_i) = \begin{cases} 1 & \text{if } (x_i, y_i) \text{ lies on shape } \mathbf{s} \\ 0 & \text{otherwise} \end{cases}$$

The Hough transform H of the image is

$$H(\mathbf{s}) = \left[\sum_{i=1}^{k} v(\mathbf{s}, x_i, y_i)\right] * K$$

where $\cdots * K$ represents the low-pass filter: convolution with some blurring kernel K. Note that this formulation does not restrict parameterizations to a discrete set of values: any parameterization that can be plotted is allowed.

The above definition of $H(\mathbf{s})$ amounts to feature points casting "fuzzy" votes for parameterizations. A more general form allows these fuzzy votes to be cast directly:

$$H(\mathbf{s}) = \sum_{i=1}^{k} v_d(\mathbf{s}, x_i, y_i) \tag{1}$$

where v_d is a *voting distribution* with range $[0, 1]$.

3.2. The Bayesian Network

The Bayesian network used in this proof is shown in Fig. 1. It is easiest to understand as a generative model of single shapes: a shape $\mathbf{s}$ (which is the same $\mathbf{s}$ as in $H(\mathbf{s})$) with a uniform distribution $P(\mathbf{s})$ over its parameters in their feasible domain generates k edge features x_i, y_i, each with the same probability density $P(x_i, y_i|\mathbf{s})$.

Even though it is extremely simple, we use a Bayesian network at this point because extensions to the basic model can become quite complex.

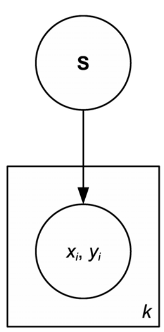


Fig. 1. A Bayesian network for finding a single parameterized shape.

3.3. Theorem

For all types of parameterizeable shapes, and for all voting distributions $v_d(\mathbf{s}, x_i, y_i)$, there exists a probability distribution $P(x_i, y_i|\mathbf{s})$ and a constant c such that, in the feasible domain of $\mathbf{s}$, $H(\mathbf{s}) = c + \log P(\mathbf{s}|x_1, y_1, ..., x_k, y_k)$.

In other words, the Hough transform computes an unnormalized log probability distribution over shape parameters given edge features. In particular, the global maximum of the Hough transform is the most probable single shape given the edge features.

Proof. We first derive $P(\mathbf{s}|x_1, y_1, ..., x_k, y_k)$. Starting from the network's joint distribution,

$$P(\mathbf{s}, x_1, y_1, ..., x_k, y_k) = P(\mathbf{s}) \prod_{i=1}^{k} P(x_i, y_i|\mathbf{s})$$

$$P(\mathbf{s}|x_1, y_1, ..., x_k, y_k) = \frac{P(\mathbf{s}) \prod_{i=1}^{k} P(x_i, y_i|\mathbf{s})}{P(x_1, y_1, ..., x_k, y_k)}$$

Because $P(x_1, y_1, ..., x_k, y_k)$ is constant and $P(\mathbf{s})$ is uniform over the feasible domain, they may be represented by a single constant. Let $\alpha = P(\mathbf{s})/P(x_1, y_1, ..., x_k, y_k)$. Then

$$P(\mathbf{s}|x_1, y_1, ..., x_k, y_k) = \alpha \prod_{i=1}^{k} P(x_i, y_i|\mathbf{s}) \tag{2}$$

and

$$\log P(\mathbf{s}|x_1, y_1, ..., x_k, y_k) = \log \alpha + \sum_{i=1}^{k} \log P(x_i, y_i|\mathbf{s}) \tag{3}$$

Let

$$P(x_i, y_i|\mathbf{s}) = \frac{1}{\beta} \exp(v_d(\mathbf{s}, x_i, y_i)) \tag{4}$$

where

$$\beta = \int_{x=0}^{m} \int_{y=0}^{n} \exp(v_d(\mathbf{s}, x, y))\, dy\, dx \tag{5}$$

The constant β normalizes v_d so that $P(x_i, y_i|\mathbf{s})$ is a valid probability distribution. Then

$$v_d(\mathbf{s}, x_i, y_i) - \log \beta = \log P(x_i, y_i|\mathbf{s}) \tag{6}$$

thus,

$$v_d(\mathbf{s}, x_i, y_i) = \log P(x_i, y_i|\mathbf{s}) + \log \beta \tag{7}$$

Substituting v_d as given by Equation (7) into Equation (1):

$$\begin{aligned} H(\mathbf{s}) &= \sum_{i=1}^{k} [\log P(x_i, y_i|\mathbf{s}) + \log \beta] \\ &= k \log \beta + \sum_{i=1}^{k} \log P(x_i, y_i|\mathbf{s}) \end{aligned}$$

Let $c = k \log \beta - \log \alpha$, so that $k \log \beta = c + \log \alpha$. Substituting into the above and from Equation (3) yields

$$\begin{aligned} H(\mathbf{s}) &= c + \log \alpha + \sum_{i=1}^{k} \log P(x_i, y_i|\mathbf{s}) \\ &= c + \log P(\mathbf{s}|x_1, y_1, ..., x_k, y_k) \end{aligned}$$

which completes the proof.

4. A SHORT ANALYSIS

A full analysis of the implications of this proof is beyond the scope of this paper; however, some aspects deserve mention.

The Hough transform assumes that each feature is generated by either a single shape or by noise. This is generally false. The Bayesian network makes it clear how to deal with multiple shapes: add more shape nodes (Fig. 2).

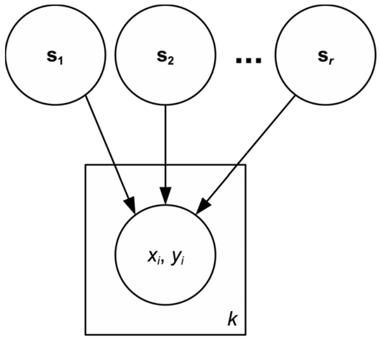


Fig. 2. A network for finding multiple shapes.

It follows from Equation (4) that the lowest possible value of $P(x_i, y_i|\mathbf{s})$ is $1/\beta$. Because of this, $P(x_i, y_i|\mathbf{s})$ can account for (equivalently, ignore) edges not produced by $\mathbf{s}$. While this explains why the Hough transform tends to be robust to noise and the presence of other shapes, notice that the proportion of allowable noise cannot be easily controlled in the Hough transform because β cannot generally be expressed in closed form. An explicit representation is in order.

More importantly, $1/\beta$ provides a statistical explanation for the Hough accumulator's complexity. A direct comparison can be drawn between the Hough transform's fat-tailed inference and estimating the center of a Cauchy density with known spread. (In the Hough transform's case, the spread of $P(x_i, y_i|\mathbf{s})$—or fuzziness of the voting function—is fixed by the convolution kernel.) The Cauchy's general insensitivity to outliers makes the likelihood very complex under fixed spread, so that in the worst case a mode exists near every sample [15]. The Hough transform's posterior density is complex and multimodal for the same reason, though it also may include ridges.

Some of this complexity is mitigated by making strong assumptions about the nature of the data. One popular method is including orientation in the feature points and accounting for it in the likelihood. (The proof of concept that follows does this and also includes level-set curvature.) More complexity is often mitigated using various techniques to post-process the accumulator. With global shape detection and parameterization expressed as a Bayesian process, much of this may be replaceable by better modeling.

5. PROOF OF CONCEPT

To demonstrate the viability of the Hough transform-based Bayesian approach, we develop a model based on the causal relationships shown in Fig. 2. The model is used to find blood platelets on a slide, which can be thought of as a four-parameter problem: find each platelet's center, radius, and amount of border deformation. A straightforward Hough transform would be unable to complete this with good accuracy in any reasonable amount of time.

Each $\mathbf{s}_j$ is a vector of independent random variables $\sigma_j^2, x_j, y_j, r_j$. Let

$$\sigma_j^2 \sim \mathrm{InvGamma}(\alpha, \beta) \qquad x_j \sim \mathrm{Gaussian}(m/2, m)$$
$$r_j \sim \mathrm{Gaussian}(13, 1) \qquad y_j \sim \mathrm{Gaussian}(n/2, n)$$

Recall that m and n are the width and height of the image. The new circle parameter σ^2 models slight deformations in the cell borders. (We also let it handle edge displacement due to discretization.) The inverse Gamma parameters α and β are chosen so that the mean and variance are 2^2 and 1, respectively.

For this test the number of circles, r, is 15. We chose a constant number to keep the proof of concept simple. (More robust ways to deal with an unknown number of shapes is discussed briefly later.) We chose 15 circles to demonstrate that the formulation allows circles with centers outside the image—anywhere the prior probability of x_j and y_j is nonzero.

Let f_i represent a feature. In addition to edge position, it contains gradient direction and level-set curvature. Its distribution is

$$P(f_i|\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_r) \propto \max(\omega, c_d(\mathbf{s_1}, f_i), c_d(\mathbf{s_2}, f_i), ..., c_d(\mathbf{s_r}, f_i))$$

where ω is a small noise constant.

In the interest of conserving space, we only *describe* the function $c_d(\mathbf{s_j}, f_i)$, which yields the probability density of feature f_i given a single shape $\mathbf{s}_j$. It is the product of three probability densities: of the location of f_i given $\mathbf{s}_j$, of the orientation (gradient direction) of f_i given $\mathbf{s}_j$, and of the level-set curvature of f_i given $\mathbf{s}_j$. (The last is computed using the expected value of r from the curvature. Figs. 3(c) and 3(d) show that this reduces the ridges in $P(\mathbf{s}|f_1, ..., f_k)$ considerably.)

We use $\max$ to combine these probability densities because it represents the intuitive notion that a feature can be generated from at most one shape. It also tends to maximize the effect of "explaining away," which prevents multiple shape nodes from converging to the same parameters.

This model poses difficulties to correct inference, the greatest being that each of the 15 groups of 4 parameters is non-identifiable. (There are at least 15! modes.) Given that the purpose of this exercise is only to prove that a Bayesian formulation *can work*, we are satisfied by only reporting one of the posterior modes. We found that Markov chain Monte Carlo (MCMC) sampling tends to find these quickly and accurately. Further, assuming the circles infrequently intersect allows searching for one circle at a time, which is very good for efficiency. Fig. 3 shows a typical result after only 200 MCMC iterations per circle. Our Python code took just over two minutes, and a specialized algorithm in C would likely take single-digit seconds.

General time requirements are difficult to estimate, as they depend on the complexity of $P(\mathbf{s}_1, ..., \mathbf{s}_r|f_1, ..., f_k)$. Space requirements are linear in the number of feature points.

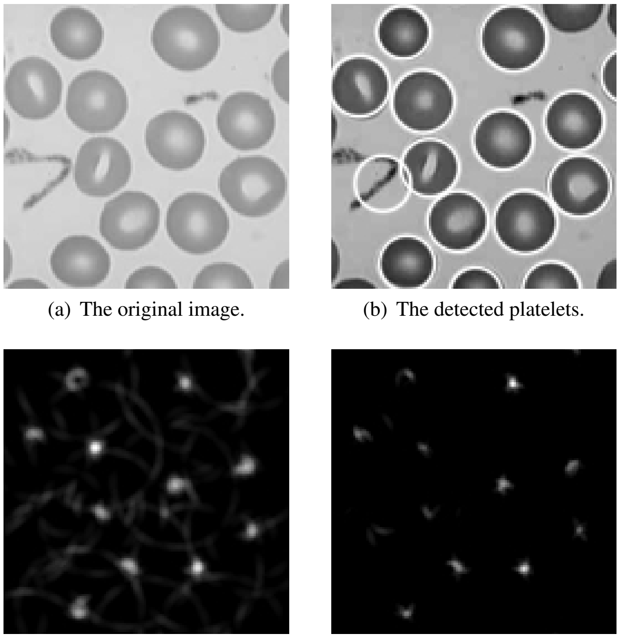
6. EXTENSIONS

Many of the extensions that the Bayesian formulation makes obvious are due to:

1. Prior belief about shape parameters can be formally expressed and integrated into the model, and
2. The model is capable of outputting a posterior distribution.

For example, it would be fairly easy to add a dynamic layer (attach a Kalman or particle filter) to the network to track shapes across video frames: the previous frame's posterior distribution over a shape would generate priors for the current frame. Geometric relationships can be modeled between components of compound shapes using priors as well.

Shape priors may come from other sources. For example, a face detection algorithm may output a distribution over the position (and possibly orientation) of a face, which the Bayesian model may use as



(a) The original image. (b) The detected platelets.

(c) A cross section of $\log P(\mathbf{s}|f_1, ..., f_k)$ with $r = 13$, $\sigma^2 = 1$. (d) The same cross section with level-set curvature incorporated.

Fig. 3. A four-parameter problem: finding blood platelets.

prior information in finding exact facial features. If the face detector is accurate enough, this would certainly increase speed and accuracy.

In general, this model of shape detection can be integrated into almost any other Bayesian system.

The multiple-shapes model as depicted in Fig. 2 bears a strong resemblance to mixture models. In fact, shape detection can easily be regarded as a clustering problem: assigning features to generating distributions. This suggests that, for an unknown number of shapes, an infinite mixture model [16]—which allows a prior distribution on the number of shapes—may be appropriate.

Finally, if both presence and lack of edges is modeled, occlusions can also be modeled—and in fact, not only could the absence of an edge not reduce the probability of a parameterization, it could *increase* the probability if occluded by another shape.

7. CONCLUSION

We started with a proof that the Hough transform has an implicit Bayesian foundation. The short analysis that followed showed that

1. The Hough transform implicitly assumes that all features are produced by a single shape or noise.
2. The proportion of noise the Hough transform accepts cannot be controlled in general.
3. Fat-tailed shape densities explain the Hough accumulator's complexity.

We found that 1 and 2 can be overcome with a more explicit Bayesian reformulation.

We followed with a proof of concept, which was able to find four-parameter shapes very quickly and using much less space than the traditional Hough transform. We concluded with a discussion of possible extensions made obvious by the Bayesian reformulation: using a dynamic layer for shape tracking, using priors to express geometric relationships between shapes, using the output from other algorithms to generate priors, using an infinite mixture model to model an unknown number of shapes, and using lack of edge features in modeling occlusions.

8. REFERENCES

[1] P.V.C. Hough, "Method and means for recognizing complex patterns," U.S. patent number 3069654, 1962.

[2] Azriel Rosenfeld, "Picture processing by computer," *ACM Comput. Surv.*, vol. 1, no. 3, pp. 147–176, 1969.

[3] V. F. Leavers, "Which Hough transform?," *CVGIP: Image Understanding*, vol. 58, no. 2, pp. 250–264, 1993.

[4] Ming Zhang, "On the discretization of parameter domain in Hough transform," in *13th International Conference on Pattern Recognition*, 1996, vol. 2, pp. 527–531.

[5] B. Morse, D. Ashton, and M. Bastian, "A Bayesian exclusionary rule for Hough transforms," in *Proceedings International Conference on Computer Vision, Pattern Recognition, and Image Processing*, 1998.

[6] Qiang Ji and R.M. Haralick, "An optimal Bayesian Hough transform for line detection," in *IEEE International Conference on Image Processing*, 1999, vol. 2, pp. 691–695.

[7] F. Zana and J.C. Klein, "A multimodal registration algorithm of eye fundus images using vessels detection and Hough transform," in *IEEE Transactions on Medical Imaging*, 1999, vol. 18, pp. 419–428.

[8] A. Bonci, T. Leo, and S. Longhi, "A Bayesian approach to the Hough transform for line detection," in *IEEE Transactions on Systems, Man, and Cybernetics (Part A)*, 2005, vol. 35, pp. 945–955.

[9] N. Kiryati and A.M. Bruckstein, "What's in a set of points?," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1992, vol. 14, pp. 496–500.

[10] Menashe Soffer and Nahum Kiryati, "Guaranteed convergence of the Hough transform," *Computer Vision and Image Understanding: CVIU*, vol. 69, no. 2, pp. 119–134, 1994.

[11] Alexander Goldenshluger and Assaf Zeevi, "The Hough transform estimator," *Annals of Statistics*, vol. 32, pp. 1908–1932, 2004.

[12] S. Lakshmanan A.K. Jain, Yu Zhong, "Object matching using deformable templates," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1996, vol. 18, pp. 267–278.

[13] G. Storvik, "A Bayesian approach to dynamic contours through stochastic sampling and simulated annealing," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, October 1994, vol. 16, pp. 976–986.

[14] D. H. Ballard, "Generalizing the Hough transform to detect arbitrary shapes," pp. 714–725, 1987.

[15] Thomas S. Ferguson, "Maximum likelihood estimates of the parameters of the Cauchy distribution for samples of size 3 and 4," *Journal of the American Statistical Association*, vol. 73, no. 361, pp. 211–213, March 1978.

[16] C. E. Rasmussen, "The infinite Gaussian mixture model," in *Advances in information processing systems 12*, S.A. et al. Solla, Ed. 2000, pp. 554–560, MIT Press.