Short Papers

Statistical Hough Transform

Rozenn Dahyot

Abstract—The Standard Hough Transform is a popular method in image processing and is traditionally estimated using histograms. Densities modeled with histograms in high dimensional space and/or with few observations, can be very sparse and highly demanding in memory. In this paper, we propose first to extend the formulation to continuous kernel estimates. Second, when dependencies in between variables are well taken into account, the estimated density is also robust to noise and insensitive to the choice of the origin of the spatial coordinates. Finally, our new statistical framework is unsupervised (all needed parameters are automatically estimated) and flexible (priors can easily be attached to the observations). We show experimentally that our new modeling encodes better the alignment content of images.

Index Terms—Hough transform, Radon transform, kernel probability density function, uncertainty, line detection.

1 Introduction

CONSIDERING a set of points in a 2D plane, the Hough transform maps each point of coordinates (x, y) to all the variables (ρ, θ) in the Hough space with the relation:

$$\rho = x \ \cos\theta + y \ \sin\theta. \tag{1}$$

If a set of observations $\mathcal{S}_{xy} = \{(x_i, y_i)\}_{i=1\cdots N}$ is aligned on one straight line with coefficients $(\hat{\theta}, \hat{\rho})$, then the family of curves $\{C_i(\theta, \rho) : \ \rho = x_i \ \cos\theta + y_i \ \sin\theta, \forall i \in [1; N]\}$ intersects in the Hough space at $(\hat{\theta}, \hat{\rho})$. This property is used to robustly perform the estimation of $(\hat{\theta}, \hat{\rho})$ by incrementing a discrete two-dimensional histogram defined on the space of variables (θ, ρ) for each point of the curves $\{C_i\}$. The highest bin of this histogram allows then to estimate the parameters $(\hat{\theta}, \hat{\rho})$ of the line. This technique has been proposed to recover lines in images more than four decades ago [1] and refined (as expressed here) in the early 1970s [2]. Many works have since been proposed both to generalize the Hough transform to more complex shapes than straight lines and to improve its computational efficiency [3]. The Hough transform has recently been proven to be a statistically robust estimator for finding lines [4].

The Hough transform has, however, one main weakness: the probability density function $p_{\theta\rho}(\theta, \rho)$ of the parameters (θ, ρ) in the Hough space is estimated using a discrete two-dimensional histogram [5]. Therefore, the trade-off in between the number of bins in the histogram and the number of available observations is crucial. Too many bins for too few observations would lead to a sparse representation of the density. Too few bins would also reduce the resolution in the Hough space, and therefore, limit the precision of the estimates. Hence, it is important to extract the most relevant information from all available observations to model the distribution of (θ, ρ). To overcome histogram limitations, the main contribution of this paper is to propose a statistical kernel modeling of the Hough transform so that the resulting estimate $\hat{p}_{\theta\rho}(\theta, \rho)$ is continuous and includes as much information as available. The variables x, y, θ, and ρ are modeled as continuous random variables from which some observations may (or may not) be available. In particular, we define the following sets of observations:

- $\mathcal{S}_{xy} = \{(x_i, y_i)\}_{i=1\cdots N}$ is the set of observations of the spatial random variables (x, y). This set is used in the Standard Hough Transform.
- $\mathcal{S}_{\theta xy} = \{(\theta_i, x_i, y_i)\}_{i=1\cdots N}$ is the set of the location with an observation of the angle θ. Indeed, when considering images, the angle of the gradient can locally be computed and used as an observation of θ.
- $\mathcal{S}_{\theta\rho} = \{(\theta_i, \rho_i)\}_{i=1\cdots N}$: knowing θ_i, x_i, and y_i, the measure ρ_i can be computed using (1) and also used as an observation.

In addition to the observations, we attached a prior p_i to each observation i. Our statistical framework completely generalizes the Standard Hough Transform and shows the clear links between the Hough and the Radon transforms. The second main contribution of this paper is to take advantage of the relation (1) between the random variables. This allows us to propose three different estimates of $p_{\theta\rho}(\theta, \rho)$ for each set of observations ($\mathcal{S}_{xy}$, $\mathcal{S}_{\theta xy}$, and $\mathcal{S}_{\theta\rho}$). This new framework is presented in Section 3.

One drawback of kernel modeling is that it requires the estimation of bandwidths. We propose in Section 4 a method to set automatically those bandwidths. We show in Section 5 how accurate and also how resistant to noise our estimates are. Many articles have been published on the Hough transform in the last 50 years or so and it has been applied to many different applications, such as image and video [6] processing, astronomy [7], or geoscience [8]. We start with a nonexhaustive review in Section 2 on local appearance-based features and the Hough Transform.

2 Context

2.1 Distribution of Local Appearance-Based Measures

Schmid et al. [9] have defined local descriptors of the intensity surface of images to detect interesting points (e.g., corners) and match images. Using similar local appearance-based features, Schiele and Crowley [10] have proposed to model their distributions using multidimensional histograms, and detection and recognition of objects can then be performed by comparing histograms. More recently, due to the increasing computational power of computers, kernel modeling [11] succeeded histograms for the modeling of the distribution of local descriptors and the Mean-shift procedure, used for finding modes of kernel densities, has found many applications in computer vision [12].

2.2 Local Hough Features

We note $(x, y) \to I(x, y)$ a surface image and its first order derivatives $I_x(x, y)$ and $I_y(x, y)$. For each pixel i at location (x_i, y_i), the following three local appearance-based measures are computed in [13]:

$$\begin{cases} \|\nabla I_i\| = \|\nabla I(x_i, y_i)\| = \sqrt{I_x^2(x_i, y_i) + I_y^2(x_i, y_i)} \\ \theta_i = \theta(x_i, y_i) = \arctan\left(\dfrac{I_y(x_i, y_i)}{I_x(x_i, y_i)}\right) \\ \rho_i = \rho(x_i, y_i) = x_i \cdot \dfrac{I_x(x_i, y_i)}{\|\nabla I(x_i, y_i)\|} + y_i \cdot \dfrac{I_y(x_i, y_i)}{\|\nabla I(x_i, y_i)\|}. \end{cases} \tag{2}$$

• *The author is with the Department of Statistics, School of Computer Science and Statistics, Trinity College Dublin, Room 128, Lloyd Institute, Dublin 2, Ireland. E-mail: Rozenn.Dahyot@tcd.ie.*

Manuscript received 25 Mar. 2008; revised 14 Oct. 2008; accepted 20 Nov. 2008; published online 26 Nov. 2008.
Recommended for acceptance by M. Lindenbaum.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2008-03-0162.
Digital Object Identifier no. 10.1109/TPAMI.2008.288.

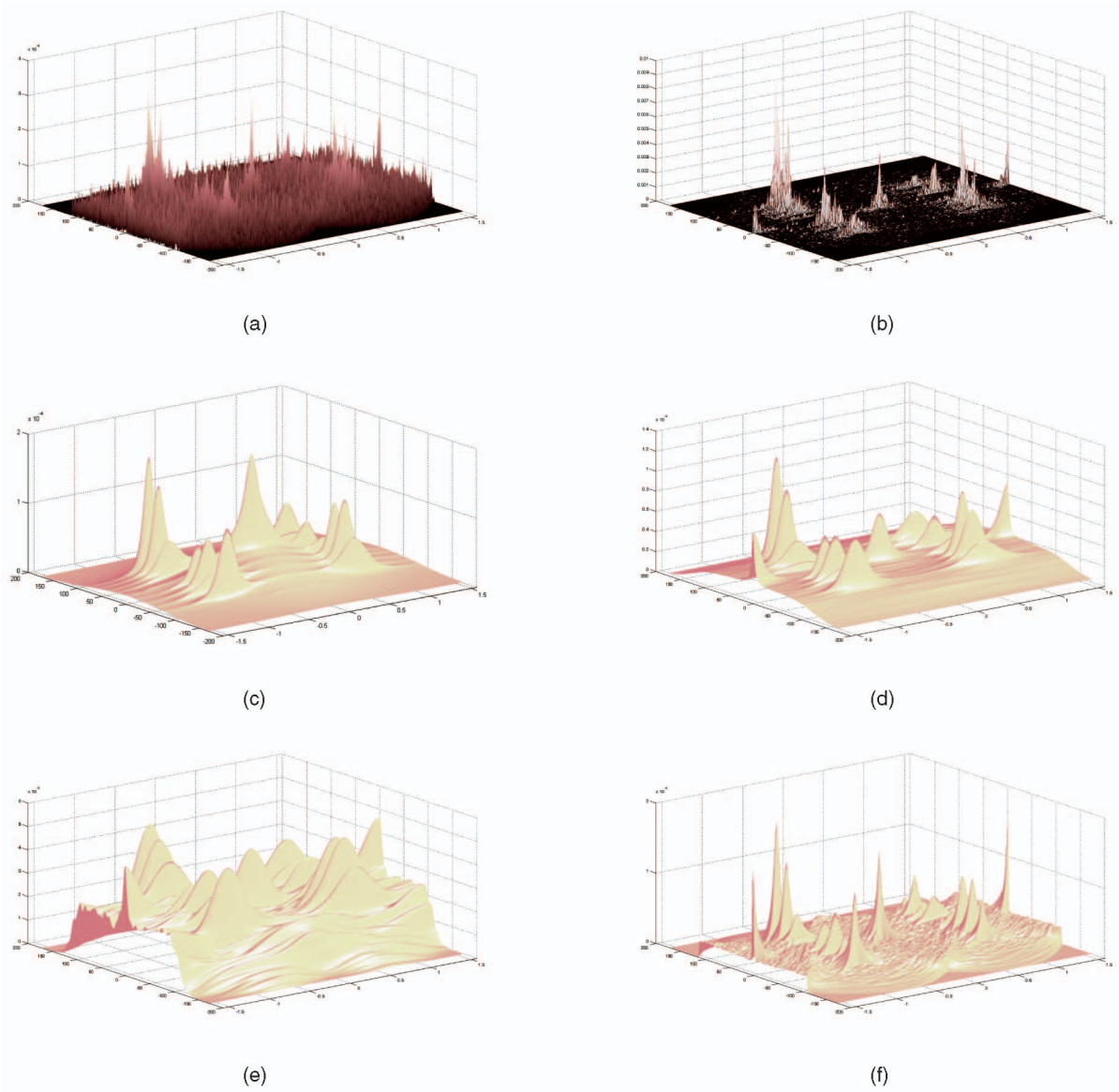

Fig. 1. Statistical Hough transform on the image *diamond* in Fig. 3b with additional Gaussian noise $\epsilon \sim \mathcal{N}(0, 20^2)$. (a) Histogram $h(\theta, \rho)$ (P1). (b) Weighted histogram $h_w(\theta, \rho)$ [14] (P1). (c) $\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{\theta\rho})$ (SVB$\theta\rho$-P1). (d) $\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{\theta\rho})$ (VB$\theta\rho$-P1). (e) $\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{\theta\rho})$ (FB$\theta\rho$-P1). (f) $\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{\theta xy})$ (VBθ-P1).

No edge segmentation is performed in [13]: The observations of all three measures $\{(\theta_i, \rho_i, \|\nabla I_i\|)\}_{i=1,\cdots,N}$ are computed for all pixels in the image and the joint distribution $p_{\theta\rho\|\nabla I\|}(\theta, \rho, \|\nabla I\|)$ is estimated by a three-dimensional histogram. The estimated density is then used to detect appearing, disappearing, or changing objects in a sequence [13].

Integrating this 3D histogram w.r.t. the magnitude of the gradient gives an estimate (2D histogram) of the density function $p_{\theta\rho}(\theta, \rho)$ that is noisy (all pixels in the image have been used). An example of such histogram computed for the image in Fig. 3b is shown in Fig. 1a.

In another approach proposed by Ji and Haralick [14], the image $I(x, y)$ is locally approximated by a plane:

$$I(x, y) = \alpha y + \beta x + \gamma + \epsilon, \tag{3}$$

where ϵ is a centered Gaussian noise. An estimate of the Hough parameters is then computed by

$$\begin{cases} \tan\theta_i = \dfrac{\hat{\alpha}}{\hat{\beta}} \\ \rho_i = x_i \cos\theta_i + y_i \sin\theta_i, \end{cases} \tag{4}$$

where $\hat{\alpha}$ and $\hat{\beta}$ are locally estimated using least squares on the neighborhood of (x_i, y_i). The spatial derivatives of $I(x, y)$ modeled in (3) give also estimates for α and β [14]. The standard error σ_{θ_i} made in computing the angle θ_i can also be quantified by [14], [15]:

$$\sigma^2_{\theta_i} = \frac{\sigma^2}{\|\nabla I_i\|^2}, \tag{5}$$

where σ^2 is the variance of the noise on the derivatives I_x and I_y, which can be known in advance or estimated from the observations of the magnitude of the gradient [16], [17], [18]. The variance in (5) confirms the intuition that the uncertainty of the computed orientation θ_i increases as the gradient magnitude decreases. A similar result has also been found in [19].

For the observation ρ_i, an estimate of its variance can also be computed for pixel i:

$$\sigma_{\rho_i}^2 = \cos^2\theta_i \ \sigma_{x_i}^2 + \sin^2\theta_i \ \sigma_{y_i}^2 + \sigma_{\theta_i}^2 (y_i \cos\theta_i - x_i \sin\theta_i)^2, \quad (6)$$

where $\sigma_{x_i}^2$ and $\sigma_{y_i}^2$ are the variances of the spatial coordinates (x_i, y_i). As noted in [14], the origin of the spatial coordinates is better chosen in the center of the image in order to limit the error done on the feature ρ_i since its variance (c.f., (6)) depends on the location (x_i, y_i).

For each observation (θ_i, ρ_i) for the pixel i at location (x_i, y_i), we can compute their uncertainty $(\sigma_{\theta_i}, \sigma_{\rho_i})$ using (5) and (6), once the spatial uncertainty σ_{x_i} and σ_{y_i} are set (see Section 4.2). A weighted histogram can then be computed using $\frac{1}{2\pi\sigma_{\theta i}\sigma_{\rho i}}$ as weights [14]. It is shown in Fig. 1b. The localization of the peaks are much easier to detect than in Fig. 1a, but it is still a discontinuous representation of the density $p_{\theta\rho}(\theta, \rho)$.

2.3 Recent Works on the Hough Transform

Many works have been published on the Hough Transform since its first publication [1]. Recently, Aggarwal and Karl proposed to robustly detect lines in noisy environment in the Hough space by adding prior modeling on the variables (θ, ρ) [20]. Several probabilistic Hough transforms, related to the RANSAC approach [21], have also been proposed [22], [23]. Of particular interest for this paper is the Meanshift clustering approach in the Hough domain proposed by Bandera et al. [23], where a continuous kernel modeling of $p_{\theta\rho}(\theta, \rho)$ with variable bandwidth is introduced. However, their modeling is deduced from a very different approach and their resulting process requires many parameters to be manually tuned.

In [24], a probabilistic interpretation of the Hough transform is proposed: The histogram $H(\theta, \rho)$ of the variables (θ, ρ) in the standard Hough transform is interpreted as $\log(p(\theta, \rho|\mathcal{S}_{xy})) \propto \sum_{i=1}^{N} \log p(\theta, \rho|(x_i, y_i))$. In [6], the standard Hough transform is extended and related to a robust M-estimator function [25].

Usually histograms are interpreted as being direct estimates of a probability density function [11] and not their logarithm. We propose in Section 3 another statistical interpretation of the Hough transform.

2.4 Reliability, Confidence, and Uncertainty of Local Features

The notion of uncertainty of a measurement is the amount by which an observed value differs from its true value. For instance, Steele and Jaynes [26] studied the uncertainty of the spatial localization of a corner detector. The notion of *repeatability* of a detector [9], defined by its robustness at detecting image features independently from perturbations in the imaging conditions, is closely related to uncertainty. A measurement with high uncertainty under a particular perturbation is indeed unlikely to be repeatable.

Another way to understand uncertainty is in considering that an observation is a random sample from an unknown distribution. For instance, lets assume a Gaussian distribution of the angle θ centered on its estimate θ_i and with standard deviation as defined in (5). Now, if we consider two instances of the same scene $d(x, y)$, $I_1(x, y) = d(x, y) + \epsilon^{(1)}(x, y)$ and $I_2(x, y) = d(x, y) + \epsilon^{(2)}(x, y)$, with Gaussian noise $\epsilon \sim \mathcal{N}(0, \sigma^2)$, then the measurements of the angle θ in flat areas (i.e., when the gradient is close to zero) are going to be randomly sampled from a uniform distribution on the interval $[-\pi/2; \pi/2]$. Therefore, it is unlikely that the measurements of θ on flat areas are corresponding in both images at the same position (x, y). The measure of the angle is not a reliable measurement to match images when it has been computed on uniform regions. On the contrary, when there is a contour ($d_x \neq 0$ or $d_y \neq 0$), the variance of the measure θ is small. As a consequence, the observed angle is sampled from a narrow Gaussian distribution and it should accurately repeat itself from one instance of an image I_1 to another I_2.

One traditional way to deal with this uncertainty on the features is to throw away unreliable observations before inferring decision. For instance, in the Standard Hough Transform, only pixels on the edges are used (i.e., pixels i with high gradient magnitude, and therefore, low uncertainty on the angle θ_i). Edge selection can be risky as relevant information can be lost in the process. In the next sections, we propose a new formalism that does not require edge segmentation as a preliminary step.

3 Statistical Hough Transform

From the sets of observations $\mathcal{S}_{\theta\rho}$, $\mathcal{S}_{\theta xy}$, and $\mathcal{S}_{xy}$, we propose three estimates of the probability density function $p_{\theta\rho}(\theta, \rho)$ using kernel modeling [11].

3.1 Kernel Density Modeling of $\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{\theta\rho})$

Using the set of observations $\mathcal{S}_{\theta\rho}$, we model the distribution $p_{\theta\rho}(\theta, \rho|\mathcal{S}_{\theta\rho})$ using kernels by

$$\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{\theta\rho}) = \sum_{i=1}^{N} \frac{1}{h_{\theta_i}} k_\theta\left(\frac{\theta - \theta_i}{h_{\theta_i}}\right) \cdot \frac{1}{h_{\rho_i}} k_\rho\left(\frac{\rho - \rho_i}{h_{\rho_i}}\right) p_i, \quad (7)$$

where h_{θ_i} and h_{ρ_i} are the variable bandwidths. Their estimations are explained in Section 4. The kernels $k_\theta(\cdot)$ and $k_\rho(\cdot)$ have been chosen Gaussians so that (7) gives a continuous and smooth estimate of the density $p_{\theta\rho}(\theta, \rho)$.

3.2 Kernel Density Modeling of $\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{\theta xy})$

In the previous section, we have used a subset of the observations leaving apart the location $\{(x_i, y_i)\}_{i=1\cdots N}$. We want now to take into account those observations to model first the density $p_{\theta\rho xy}(\theta, \rho, x, y)$. Using the Bayes formula, we can write

$$p_{\theta\rho xy}(\theta, \rho, x, y) = p_{\rho|\theta xy}(\rho|\theta, x, y) \ p_{\theta xy}(\theta, x, y). \quad (8)$$

As noticed by Bonci et al. [15], when x, y, θ are known, the variable ρ is deterministic, by definition, in (1). Therefore, we propose to model the conditional probability as follows:

$$p_{\rho|\theta xy}(\rho|\theta, x, y) = \delta(\rho - x\cos\theta - y\sin\theta), \quad (9)$$

where $\delta(\cdot)$ is the Dirac distribution. As a consequence, only $p_{\theta xy}(\theta, x, y)$ is to estimate using kernels with the set of observatons $\mathcal{S}_{\theta xy}$:

$$\begin{aligned}
\hat{p}_{\theta xy}(\theta, x, y|\mathcal{S}_{\theta xy}) &= \sum_{i=1}^{N} \hat{p}_{\theta xy}(\theta, x, y|\theta_i, x_i, y_i) \ p_i \\
&= \sum_{i=1}^{N} \hat{p}_\theta(\theta|\theta_i) \ \hat{p}_x(x|x_i) \ \hat{p}_y(y|y_i) \ p_i \\
&= \sum_{i=1}^{N} \frac{1}{h_{x_i}} k_x\left(\frac{x - x_i}{h_{x_i}}\right) \frac{1}{h_{y_i}} k_y\left(\frac{y - y_i}{h_{y_i}}\right) \\
&\quad \times \frac{1}{h_{\theta_i}} k_\theta\left(\frac{\theta - \theta_i}{h_{\theta_i}}\right) p_i.
\end{aligned} \quad (10)$$

Note that we have assumed the variables θ, x, and y independent given their observations (θ_i, x_i, y_i). By integration with respect to the spatial variables (x, y), an estimate of the Hough transform $p_{\theta\rho}(\theta, \rho)$ can be computed:

$$\hat{p}(\theta, \rho|\mathcal{S}_{\theta xy}) = \sum_{i=1}^{N} \frac{1}{h_{\theta_i}} k_\theta\left(\frac{\theta - \theta_i}{h_{\theta_i}}\right) R_i(\theta, \rho) \ p_i, \quad (11)$$

where $R_i(\theta,\rho)$ is the Radon transform of the spatial kernels:

$$R_i(\theta,\rho) = \int\int \delta(\rho - x\cos\theta - y\sin\theta)\ \frac{1}{h_{x_i}} k_x\left(\frac{x-x_i}{h_{x_i}}\right) \times \frac{1}{h_{y_i}} k_y\left(\frac{y-y_i}{h_{y_i}}\right) dx\ dy. \tag{12}$$

3.3 Standard Hough Transform $\hat{p}_{\theta\rho}(\theta,\rho|\mathcal{S}_{xy})$

Let's assume now that the only available observations are the set of positions $\mathcal{S}_{xy}$. No prior information is available on the variable θ, therefore, its kernel can be replaced in (11) by the uniform distribution such that

$$\frac{k_\theta(\theta-\theta_i)}{h_{\theta_i}} = \frac{1}{\pi}.$$

Consequently, expression (10) becomes:

$$\hat{p}_{\theta\rho}(\theta,\rho|\mathcal{S}_{xy}) = \frac{1}{\pi}\sum_i R_i(\theta,\rho)\ p_i = \frac{1}{\pi} R(\theta,\rho), \tag{13}$$

where $R(\theta,\rho)$ is the Radon transform of the kernel estimate of the spatial coordinates:

$$R(\theta,\rho) = \int\int \delta(\rho - x\cos\theta - y\sin\theta)\ \hat{p}_{xy}(x,y|\mathcal{S}_{xy})\ dx\ dy \tag{14}$$

with

$$\hat{p}_{xy}(x,y|\mathcal{S}_{xy}) = \sum_{i=1}^{N} \frac{1}{h_{x_i}} k_x\left(\frac{x-x_i}{h_{x_i}}\right) \frac{1}{h_{y_i}} k_y\left(\frac{y-y_i}{h_{y_i}}\right)\ p_i. \tag{15}$$

Equation (13) gives a kernel estimate for the standard Hough transform. It is easy to compute using the Radon transform and its performance is illustrated in Section 5.1.

3.4 Priors $\{p_i\}_{i=1\cdots N}$

We consider the following priors:

1. All observations are equiprobable $p_i = \frac{1}{N},\ \forall i$. This setting is noted as P1 in the rest of the paper.
2. Among all of the pixels, only the ones on detected edges are used in the Standard Hough Transform. These binary priors (P2) are equiprobable for all the selected pixels and zero for the others.

Other priors could be used, but all obey the following constraints:

$$\left(\sum_{i=1}^{N} p_i = 1\right) \quad \wedge \quad (\forall i,\ 0 \le p_i \le 1).$$

4 Kernels and Bandwidths

It is usually acknowledged that the choice of the kernel $k(\cdot)$ has a limited impact on the estimate of the p.d.f. [11]. We have chosen the kernels k_θ and k_ρ as Gaussians and, in the next section, we discuss the choice of the spatial kernels k_x and k_y.

On the contrary, the choice of the bandwidths does impact on the estimate of the p.d.f. and in particular on its number of modes and bumps. Sections 4.2 and 4.3 propose and explain how we can set the bandwidths in the case of digital images.

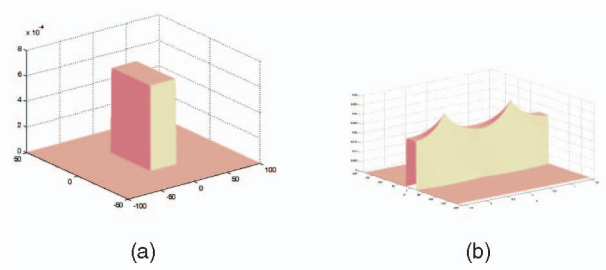

Fig. 2. (a) Spatial uniform kernel with (b) its Radon transform $R_i(\theta,\rho)$.

4.1 Spatial Kernels

Various kernels k_x and k_y can be used such as:

1. The Dirac kernels are defined as:

$$\begin{cases} \dfrac{k_x(x-x_i)}{h_{x_i}} = \delta(x-x_i) \\ \dfrac{k_y(y-y_i)}{h_{y_i}} = \delta(y-y_i). \end{cases} \tag{16}$$

In this case, no bandwidths are needed. The corresponding kernel $R_i(\theta,\rho)$ in the Hough space (cf., (11) and (12)) is then also a Dirac:

$$R_i(\theta,\rho) = \delta(\rho - x_i\cos\theta - y_i\sin\theta). \tag{17}$$

2. The Gaussian kernels are defined as:

$$\begin{cases} \dfrac{k_x(x-x_i)}{h_{x_i}} = \mathcal{N}\left(x_i, h_{x_i}^2\right) \\ \dfrac{k_y(y-y_i)}{h_{y_i}} = \mathcal{N}\left(y_i, h_{y_i}^2\right). \end{cases} \tag{18}$$

The corresponding kernel $R_i(\theta,\rho)$ in the Hough space is then

$$R_i(\theta,\rho) = \frac{1}{\sqrt{2\pi\left(h_{y_i}^2\sin^2\theta + h_{x_i}^2\cos^2\theta\right)}} \times \exp\left(\frac{-(\rho - (x_i\cos\theta + y_i\sin\theta))^2}{2\left(h_{y_i}^2\sin^2\theta + h_{x_i}^2\cos^2\theta\right)}\right). \tag{19}$$

3. A standard assumption to model the image grid is to choose uniform kernels [27]

$$\begin{cases} \dfrac{k_x(x-x_i)}{h_{x_i}} = \begin{cases} \dfrac{1}{h_{x_i}} & \text{for } \dfrac{|\mathrm{x}-\mathrm{x_i}|}{\mathrm{h_{x_i}}} < \dfrac{1}{2} \\ 0 & \text{otherwise} \end{cases} \\ \dfrac{k_y(y-y_i)}{h_{y_i}} = \begin{cases} \dfrac{1}{h_{y_i}} & \text{for } \dfrac{|\mathrm{y}-\mathrm{y_i}|}{\mathrm{h_{y_i}}} < \dfrac{1}{2} \\ 0 & \text{otherwise.} \end{cases} \end{cases} \tag{20}$$

In this case, we represented $R_i(\theta,\rho)$ in Fig. 2: The two peaks in $R_i(\theta,\rho)$ correspond to the diagonals of the spatial uniform kernel. So, from one observation (x_i,y_i), the kernel in the Hough space is then bimodal in the case of uniform spatial kernels, and the two directions of the diagonals have higher probabilities than the other directions.

In this work, we have chosen the spatial kernels as Gaussian so that no specific direction θ is favored. The Dirac kernel is then just

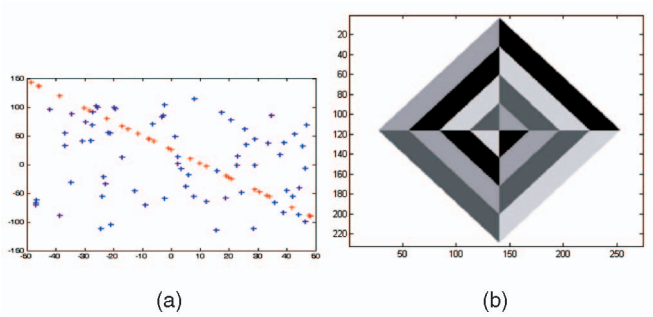

Fig. 3. Observations: a cloud of points, $\{(x_i, y_i)\}_{i=1,\cdots 100}$, and an image. (a) Sparse cloud of points. (b) Image *diamond* [15].

a special case of the Gaussian kernel when the bandwidth goes toward 0.

4.2 Bandwidths

For the variables θ and ρ, we naturally set the bandwidths $h_{\theta_i} = \sigma_{\theta_i}$ and $h_{\rho_i} = \sigma_{\rho_i}$ (see (5) and (6)). The densities $\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{\theta\rho})$ and $\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{\theta xy})$ for the image 3b, computed with these variable bandwidth estimates, are respectively, represented in Figs. 1d and 1f.

The spatial bandwidths (h_{x_i}, h_{y_i}) are estimated in a similar fashion as the variable bandwidth for θ. (h_{x_i}, h_{y_i}) reflect the uncertainty attached to the observations (x_i, y_i) of the variables (x, y). Because of the digitalization process (or quantification), the observations of the variables (x, y) have a precision ± 1, $\forall i$ (assuming a regular image grid resolution). Therefore, we set

$$h_{x_i} = h_{y_i} = 1, \ \forall i. \tag{21}$$

When no knowledge is available on the uncertainty of the observations, several unsupervised methods to set the variable bandwidths are proposed in [11]. For instance, the nearest neighbor approach can be used [11]:

$$h_{x_i} = h_{y_i} = \min_{(j\in[1;N])\wedge(j\neq i)} \{\|\mathbf{x}_i - \mathbf{x}_j\|\}, \tag{22}$$

with $\mathbf{x}_i = (x_i, y_i)$ and $\mathbf{x}_j = (x_j, y_j)$. Note that this approach would also give $h_{x_i} = h_{y_i} = 1$, $\forall i$ when the data are collected from a regular image grid. However, (22) is general and could be applied to other data sets when no knowledge about the uncertainty of the spatial observations is available.

4.3 Justification of the Choice of Bandwidths

We want to estimate $p_{\theta\rho}(\theta, \rho)$ such that the estimate is not sensitive to Gaussian noise on the intensity of the image. Let's assume that we have a noise free image $d(x, y)$ and that we can observe this image several times with different instances of the noise $\epsilon^{(n)}(x, y) \sim \mathcal{N}(0, \sigma^2)$, $\forall (x, y)$:

$$I^{(n)}(x, y) = d(x, y) + \epsilon^{(n)}(x, y),$$

then the extracted observations at pixel i are $\forall n$:

$$\begin{cases} x_i^{(n)} = x_i \\ y_i^{(n)} = y_i \\ \theta_i^{(n)} \sim \mathcal{N}(\mathbb{E}[\theta_i], \mathbb{W}[\theta_i]) \\ \rho_i^{(n)} \sim \mathcal{N}(\mathbb{E}[\rho_i], \mathbb{W}[\rho_i]). \end{cases}$$

In other words, the noise in the intensity of the image does not affect the measurement of the location (x_i, y_i) of the pixel i. However, it does affect the measurements of (θ, ρ) at pixel i. When we have only one instance of the image, then the expectation of (θ, ρ) is computed using (2) with the sole observation available at pixel i:

$$\begin{cases} \mathbb{E}[\theta_i] = \theta_i = \theta(x_i, y_i) \\ \mathbb{E}[\rho_i] = \rho_i = \rho(x_i, y_i). \end{cases} \tag{23}$$

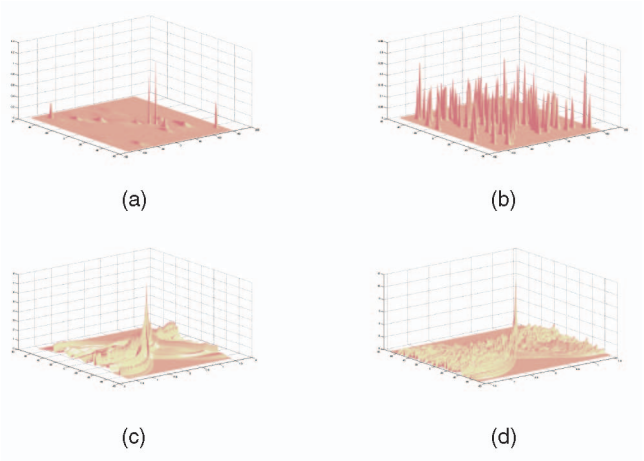

Fig. 4. Spatial and Hough densities. (a) $\hat{p}_{xy}(x, y|\mathcal{S}_{xy})$ FB; (b) $\hat{p}_{xy}(x, y|\mathcal{S}_{xy})$ VB; (c) $\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{xy})$ VB; and (d) $\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{xy})$ FB.

The variances $\mathbb{W}[\theta_i]$ and $\mathbb{W}[\rho_i]$ are estimated as proposed in [14], [15] (cf., (5) and (6)) and used to set the bandwidths at pixel i.

To verify our model, $n = 100$ instances of Fig. 3b were generated with different instance of the noise. It means that, for each pixel i, we can compute 100 times the values (θ_i, ρ_i) from which we compute their average values $(\mathbb{E}[\theta_i], \mathbb{E}[\rho_i])$ and variances $(\mathbb{W}[\theta_i], \mathbb{W}[\rho_i])$ using standard formula of statistics. Using these estimates in (7) gives the probability density function represented in Fig. 1c. Peaks in the simulated distribution Fig. 1c are slightly higher and narrower than in the estimated distribution Fig. 1d. It shows that our variable bandwidths of θ and ρ are slightly overestimated in Fig. 1d, but, nevertheless, our variable bandwidths give a good estimate.

Note that one major difference appears at the peaks at $-\pi/2$ and $\pi/2$ in Fig. 1c. Since the angle is modulo π, its occurrences swap randomly in between $-\pi/2$ and $\pi/2$ on the horizontal edge that splits the diamond into two. Consequently, the simulated variance for this line is very large, which flattens the peaks located at $\theta = -\pi/2$ and $\theta = \pi/2$. For comparison, we also computed the p.d.f. $\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{\theta\rho})$ with a fixed bandwidth h_{SJ} (Sheather-Jones plug-in [28]) and it is represented in Fig. 1e. One can note that this fixed bandwidth kernel estimate of $p_{\theta\rho}(\theta, \rho)$ does not show sharp peaks as for our variable bandwidth one (Fig. 1d).

4.4 Remark

We chose the kernels k_θ and k_ρ at pixel i as being Gaussian distributions. Consequently, the only parameters needed in our modeling are the expectation and the variance. Note that more complex kernels for θ and ρ can be estimated locally using steerable filters [29]. These would be more accurate in particular at corners.

TABLE 1
Robustness

$\frac{n_o}{N}\%$	70%	80%	90%
FB-1	99	98	86
FB-2	100	100	90
VB-1	89	83	20
VB-2	100	93	27

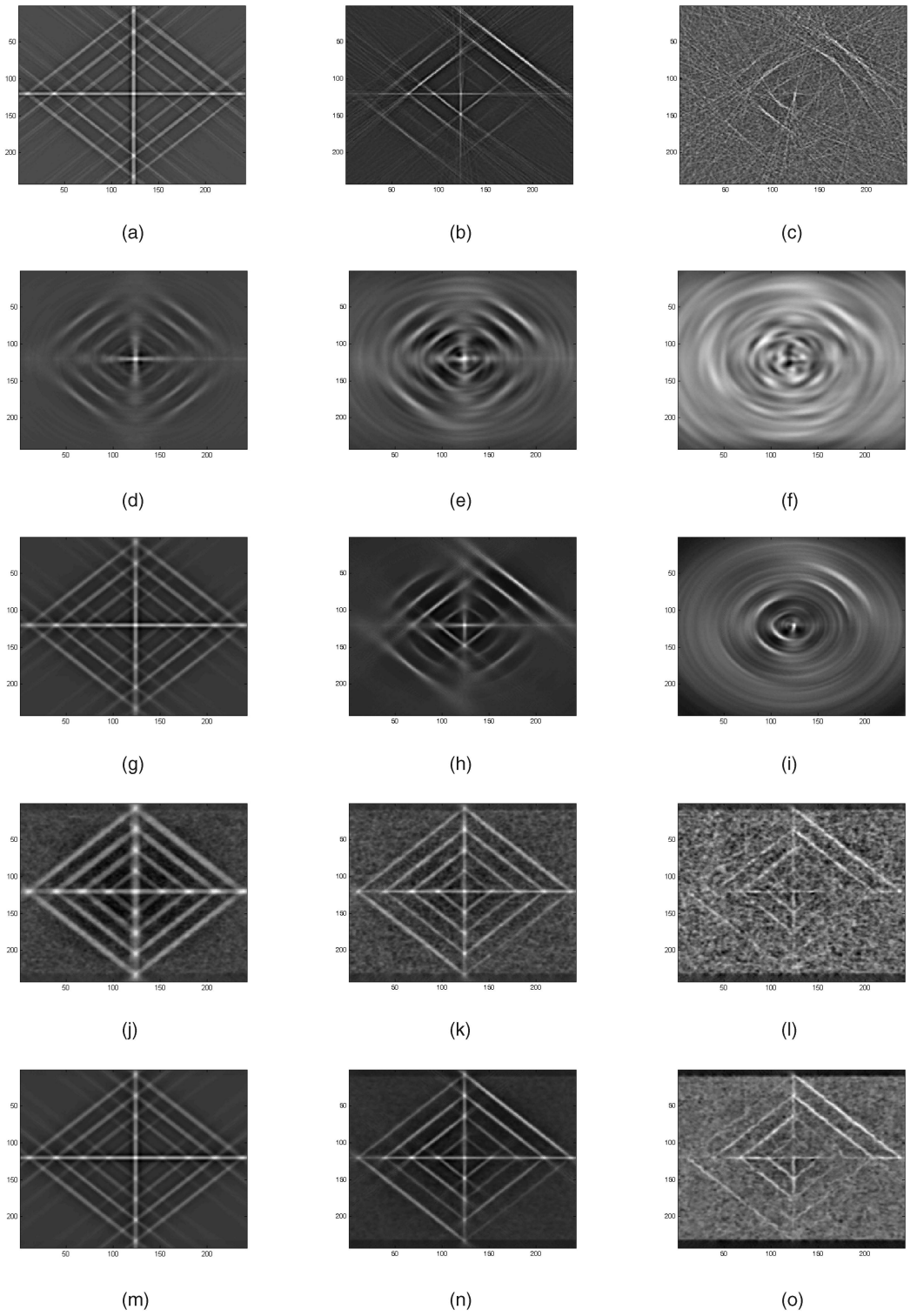

Fig. 5. Comparison of the inverse Radon transforms of the estimates of $p_{\theta\rho}(\theta\rho)$ computed for the image 3b with different level of noise. (a) $\mathcal{R}^*[h_w(\theta,\rho)]$ (P1-0). (b) $\mathcal{R}^*[h_w(\theta,\rho)]$ (P1-20). (c) $\mathcal{R}^*[h_w(\theta,\rho)]$ (P1-100). (d) $\mathcal{R}^*[\hat{p}_{\theta\rho}(\theta,\rho|\mathcal{S}_{\theta\rho})]$ (FB$\theta\rho$-P1-0). (e) $\mathcal{R}^*[\hat{p}_{\theta\rho}(\theta,\rho|\mathcal{S}_{\theta\rho})]$ (FB$\theta\rho$-P1-20). (f) $\mathcal{R}^*[\hat{p}_{\theta\rho}(\theta,\rho|\mathcal{S}_{\theta\rho})]$ (FB$\theta\rho$-P1-100). (g) $\mathcal{R}^*[\hat{p}_{\theta\rho}(\theta,\rho|\mathcal{S}_{\theta\rho})]$ (VB$\theta\rho$-P1-0). (h) $\mathcal{R}^*[\hat{p}_{\theta\rho}(\theta,\rho|\mathcal{S}_{\theta\rho})]$ (VB$\theta\rho$-P1-20). (i) $\mathcal{R}^*[\hat{p}_{\theta\rho}(\theta,\rho|\mathcal{S}_{\theta\rho})]$ (VB$\theta\rho$-P1-100). (j) $\mathcal{R}^*[\hat{p}_{\theta\rho}(\theta,\rho|\mathcal{S}_{\theta xy})]$ (FBθ-P1-0). (k) $\mathcal{R}^*[\hat{p}_{\theta\rho}(\theta,\rho|\mathcal{S}_{\theta xy})]$ (FBθ-P1-20). (l) $\mathcal{R}^*[\hat{p}_{\theta\rho}(\theta,\rho|\mathcal{S}_{\theta xy})]$ (FBθ-P1-100). (m) $\mathcal{R}^*[\hat{p}_{\theta\rho}(\theta,\rho|\mathcal{S}_{\theta xy})]$ (VBθ-P1-0). (n) $\mathcal{R}^*[\hat{p}_{\theta\rho}(\theta,\rho|\mathcal{S}_{\theta xy})]$ (VBθ-P1-20). (o) $\mathcal{R}^*[\hat{p}_{\theta\rho}(\theta,\rho|\mathcal{S}_{\theta xy})]$ (VBθ-P1-100).

5 Experimental Results

We compute a discrete representation of our estimates $\hat{p}_{\theta\rho}(\theta,\rho)$ on a fine grid $\theta \in (-\frac{\pi}{2} : \delta_\theta : \frac{\pi}{2})$ and $\rho \in [-l_\rho : \delta_\rho : +l_\rho]$. δ_θ represents the resolution of the discrete density on the axis θ and δ_ρ is the resolution in the direction ρ. These have been chosen as $\delta_\theta = \pi/180$ and $\delta_\rho = 1$ in the experiments Section 5. l_ρ is the maximum limit of ρ. Assuming an image of size $w \times h$ with the origin of the spatial coordinates in the middle of the image, then

$$l_\rho = \sqrt{\left(\frac{w}{2}\right)^2 + \left(\frac{h}{2}\right)^2}. \quad (24)$$

Once our estimates are computed, we want to assess how well they capture the alignment content of the observations. In the experiment described in Section 5.1, only one alignment occurs and we compare the accuracy of the detected maximum $(\hat{\theta}, \hat{\rho})$ in each of our p.d.f. estimates of $p_{\theta\rho}(\theta, \rho)$.

The whole p.d.f. $p_{\theta\rho}(\theta, \rho)$ encodes the probability of aligned edges in an image, and not only its maxima. We propose in Section 5.2 to visualize the corresponding (dual) density estimate $\hat{p}_{xy}(x, y)$ in the spatial domain computed by inverse Radon transform of our estimates of $p_{\theta\rho}(\theta, \rho)$. If our density estimates $\hat{p}_{\theta\rho}(\theta, \rho)$ encode the alignment content of the image well, then their Inverse Radon Transform will reflect it in the spatial domain by enhancing straight edges and discarding the rest (e.g., nonedges and nonstraight edges).

5.1 Estimate $\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{xy})$

N data points $\{(x_i, y_i)\}_{i=1,\cdots N=100}$ are randomly generated:

- $n_i = 30$ belongs to a straight line $(10 = x\cos(\pi/8) + y\sin(\pi/8)$ with $x \in [-50; 50])$,
- while $n_o = 70$ outlier points are uniformly distributed $(n_o + n_i = N)$.

All of these data points have equiprobable priors $p_i = \frac{1}{N}$ (P1), although the clouds of the observed points can also be interpreted as the result of an edge segmentation process in an image (P2).

An example of the data is shown in Fig. 3a. The corresponding spatial densities $\hat{p}_{xy}(x, y|\mathcal{S}_{xy})$ (see (15)) are computed with the variable bandwidth (VB) (Fig. 4a) (see (22)) and a fixed bandwidth (FB) in Fig. 4b with $h_{x_i} = h_{y_i} = 1,\ \forall i$. The corresponding densities $\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{xy})$ are in Figs. 4c and 4d, and despite 70 percent of outliers, the maximum is easily detected to give a robust estimate of the line. Out of 100 simulations, 99 maxima $(\hat{\theta}, \hat{\rho})$ of the density $\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{xy})$ computed with a fixed bandwidth (FB1) were in the vicinity of the true values $(\pi/8, 10)$ with a precision of 1 degree for the angle and 1 for ρ and 89 percent of the estimates with variable bandwidth (VB1). When relaxing the precision from 1 to 2 degrees and $\hat{\rho} = 10 \pm 2$, all of the 100 estimates $\{(\hat{\theta}, \hat{\rho})^{(n)}\}$ computed with the fixed (FB2) and variable bandwidths (VB2) are found.

This experience is repeated for higher percentages of outliers (see Table 1). When the proportion of outliers reaches 90 percent, the fixed bandwidth gives a more reliable estimation of the line than with the variable bandwidth. Indeed, more prior information on the uncertainty of the observations $\{(x_i, y_i)\}_{i=1,\cdots N=100}$ is known when setting $h_{x_i} = h_{y_i} = 1,\ \forall i$, and the estimated pdf is therefore better constrained.

5.2 Estimates $\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{\theta\rho})$ and $\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{\theta xy})$

The image *diamond* Fig. 3b has been used in our experiment with three levels of noise $\epsilon \sim \mathcal{N}(0, \sigma^2)$ with $\sigma = 0,\ 20,\ 100$. All of the experiments use spatial Gaussian kernels with bandwidths $h_{x_i} = h_{y_i} = 1,\ \forall i$, for a prior P1 (all pixels have equiprobable priors).

Fig. 5 shows the inverse Radon transform of our density estimates. FB$\theta\rho$ (respectively, FBθ) indicates that the Sheather and Jones plug-in bandwidth [28] is used to set the bandwidths for θ and ρ in computing $\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{\theta\rho})$ (respectively, $\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{\theta xy})$). The notation VB$\theta\rho$ (respectively, VBθ) stands for variable bandwidths use for both ρ and θ in computing $\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{\theta\rho})$ (respectively, $\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{\theta xy})$).

Comparing Figs. 5a, 5b, and 5c, we can note that the weighted histogram renders a good alignment content for a moderate noise. Several straight contours with low contrast seem to disappear quickly as noise increases. The weighted histogram is an estimate of $p_{\theta_\rho}(\theta\rho)$ that is fast to compute compared to the kernel-based estimates. In fact, the weighted histogram can be seen as a truncated estimate of $\hat{p}_{\theta\rho}(\theta\rho|\mathcal{S}_{\theta_\rho})$, where the tails of the Gaussian kernels are not taken into account.

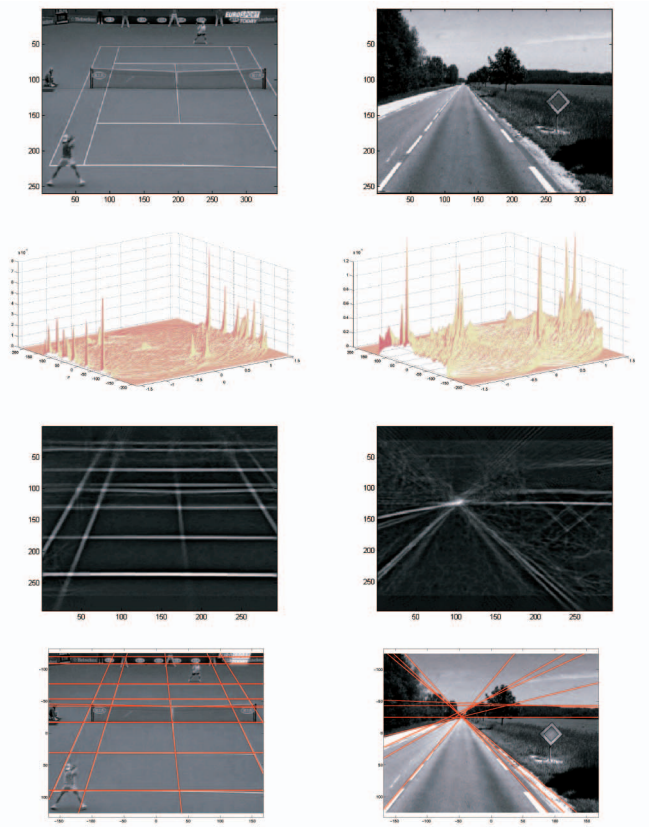

Fig. 6. Results on (top) real images, with their estimates $\hat{p}_{\theta\rho}(\theta\rho|\mathcal{S}_{\theta xy})$ and the inverse Radon transform, and (bottom) the detected lines.

Figs. 5d, 5e, and 5f compare the estimates $\hat{p}_{\theta\rho}(\theta\rho|\mathcal{S}_{\theta_\rho})$ computed with a fixed bandwidth. Even with no additional noise, the rendering of the straight contours is not very good as all pixels contribute with the same bandwidth. In contrast, Figs. 5g, 5h, and 5i show a better alignment content recovered using variable bandwidths as proposed in this paper. However, as the noise increases, the reconstruction gets very bad: the variable bandwidth computed for the variable ρ does depend on the noise level but also on the spatial position (x, y). As expected, as we go further away from the spatial origin (taken in the center of the image), the quality of the rendering of the straight contours deteriorates.

Figs. 5j, 5k, and 5l show the inverse Radon transform of the estimate $\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{\theta xy})$ computed with fixed bandwidth on the variable θ. The rendering reflects the alignment content of the image *diamond* well and does not deteriorate away from the spatial origin as for $\hat{p}_{\theta\rho}(\theta\rho|\mathcal{S}_{\theta_\rho})$. This is even improved when using variable bandwidths in Figs. 5m, 5n, and 5o.

Fig. 6 presents the inverse radon transform computed on the estimates $\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{\theta xy})$ of two real images. Only the alignment content remains while strong contours from objects with no straight edges do not. While computing the inverse Radon transform of the estimates of $p_{\theta\rho}(\theta, \rho)$ helps to visualize all the information encoded, many works on the Hough transform have only focused on detecting its modes (maxima) to recover the aligned segments (see Fig. 6). In [30], bumps are segmented in $p_{\theta\rho}(\theta, \rho)$ as an alternative to modes, to recover the spatial alignments.

6 Conclusion

We have proposed several kernel modelings to generalize the Hough Transform. This new approach is unsupervised since all the needed bandwidths are automatically estimated. In addition, by explicitly modeling dependencies between the variables, we proposed a kernel modeling that is not depending on the choice of the origin of the coordinates in the spatial domain. The resulting density $\hat{p}_{\theta\rho}(\theta\rho|\mathcal{S}_{\theta xy})$ encodes the alignment content of the images well and is robust to noise. Computing the Statistical Hough Transform on an image is more computationally expensive than the Standard Hough transform because all pixels of the image are used (instead of selecting only the edges) and, moreover, the tails of the kernels have to be taken into account in computing the estimates. Future work will investigate random and edge selection strategies to speed up the Statistical Hough Transform.

Acknowledgments

This work has been supported by the Enterprise Ireland Innovation Partnership IP-2006-412 and a Research Google Award.

References

[1] P. Hough, "Methods of Means for Recognizing Complex Patterns," US Patent 3 069 654, 1962.
[2] R.O. Duda and P.E. Hart, "Use of the Hough Transformation to Detect Lines and Curves in Pictures," *Comm. ACM,* vol. 15, pp. 11-15, Jan. 1972.
[3] J.-Y. Goulermas and P. Liatsis, "Incorporating Gradient Estimations in a Circle-Finding Probabilistic Hough Transform," *Pattern Analysis and Applications,* vol. 2, pp. 239-250, 1999.
[4] A. Goldenshluger and A. Zeevi, "The Hough Transform Estimator," *The Annals of Statistics,* vol. 32, no. 5, pp. 1908-1932, Oct. 2004.
[5] A.S. Aguado, E. Montiel, and M.S. Nixon, "Bias Error Analysis of the Generalized Hough Transform," *J. Math. Imaging and Vision,* vol. 12, pp. 25-42, 2000.
[6] M. Bober and J. Kittler, "Estimation of Complex Multimodal Motion: An Approach Based on Robust Statistics and Hough Transform," *Image and Vision Computing J.,* vol. 12, no. 10, pp. 661-668, Dec. 1994.
[7] P. Ballester, "Hough Transform and Astronomical Data Analysis," *Vistas in Astronomy,* vol. 40, no. 4, pp. 479-485, 1996.
[8] G.R.J. Cooper and D.R. Cowan, "The Detection of Circular Features in Irregularly Spaced Data," *Computers & Geosciences,* vol. 30, no. 1, pp. 101-105, Feb. 2004.
[9] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of Interest Point Detectors," *Int'l J. Computer Vision,* vol. 37, no. 2, pp. 151-172, 2000.
[10] B. Schiele and J.L. Crowley, "Recognition without Correspondence Using Multidimensional Receptive Field Histograms," *Int'l J. Computer Vision,* vol. 36, no. 1, pp. 31-50, Jan. 2000.
[11] B.W. Silverman, *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, 1986.
[12] D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach Toward Feature Space Analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 5, pp. 603-619, May 2002.
[13] R. Dahyot, P. Charbonnier, and F. Heitz, "Unsupervised Statistical Change Detection in Camera-in-Motion Video," *Proc. IEEE Int'l Conf. Image Processing,* Oct. 2001.
[14] Q. Ji and R.M. Haralick, "Error Propagation for the Hough Transform," *Pattern Recognition Letters,* vol. 22, pp. 813-823, 2001.
[15] A. Bonci, T. Leo, and S. Longhi, "A Bayesian Approach to the Hough Transform for Line Detection," *IEEE Trans. Systems, Man, and Cybernetics,* vol. 35, no. 6, pp. 945-955, Nov. 2005.
[16] G. Lai and R.D. Figueiredo, "A Novel Algorithm for Edge Detection from Direction-Derived Statistics," *Proc. IEEE Int'l Symp. Circuits and Systems,* vol. 5, pp. 37-40, May 2000.
[17] R. Dahyot, N. Rea, A. Kokaram, and N. Kingsbury, "Inlier Modeling for Multimedia Data Analysis," *Proc. IEEE Int'l Workshop Multimedia Signal Processing,* pp. 482-485, Sept. 2004.
[18] R. Dahyot and S. Wilson, "Robust Scale Estimation for the Generalized Gaussian Probability Density Function," *Advances in Methodology and Statistics (Metodološki zvezki),* vol. 3, no. 1, pp. 21-37, 2006.
[19] P. Meer and B. Georgescu, "Edge Detection with Embedded Confidence," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 12, pp. 1351-1365, Dec. 2001.
[20] N. Aggarwal and W.C. Karl, "Line Detection in Image through Regularized Hough Transform," *IEEE Trans. Image Processing,* vol. 15, no. 3, pp. 582-591, Mar. 2006.
[21] M.A. Fischler and R.C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Comm. ACM,* vol. 24, no. 6, pp. 381-395, 1981.
[22] D. Walsh and A.E. Raftery, "Accurate and Efficient Curve Detection in Images: The Importance Sampling Hough Transform," *Pattern Recognition,* vol. 35, pp. 1421-1431, 2002.
[23] A. Bandera, J.P.B.J.M. Pérez-Lorenzo, and F. Sandoval, "Mean Shift Based Clustering of Hough Domain for Fast Line Segment Detection," *Pattern Recognition Letters,* vol. 27, pp. 578-586, 2006.
[24] R.S. Stephens, "Probabilistic Approach to the Hough Transform," *Image and Vision Computing J.,* vol. 9, no. 1, pp. 66-71, Feb. 1991.
[25] P. Huber, *Robust Statistics.* John Wiley and Sons, 1981.
[26] R.M. Steele and C. Jaynes, "Feature Uncertainty Arising from Covariant Image Noise," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 1063-1069, 2005.
[27] J. Princen, J. Illingworth, and J. Kittler, "A Formal Definition of the Hough Transform: Properties and Relationships," *J. Math. Imaging and Vision,* vol. 1, no. 2, pp. 153-168, 1992.
[28] S.J. Sheather, "Density Estimation," *Statistical Science,* vol. 19, no. 4, pp. 588-597, 2004.
[29] W.T. Freeman, "Steerable Filters and Local Analysis of Image Structure," PhD dissertation, Massachusetts Inst. of Technology, 1992.
[30] R. Dahyot, "Bayesian Classification for the Statistical Hough Transform," *Proc. IEEE Int'l Conf. Pattern Recognition,* Dec. 2008.