Statistical Hough Transform

Rozenn Dahyot

***Abstract*— Local descriptors of the image surface and their distributions are attributes used in many computer vision applications. Using additional information measuring their uncertainty, a more accurate estimate of their probabilistic distribution can be modelled. This concept is illustrated here by proposing two kernel based representation of the distribution of the Hough variables that is more robust than the standard Hough transform.**

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

I. 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 $\mathcal{S} = \{(x_i, y_i)\}_{i\in\mathcal{I}}$ is aligned on one straight line with coefficient $(\hat{\theta}, \hat{\rho})$, then the family of curves $\{C_i(\theta, \rho) : \ \rho = x_i \ \cos\theta + y_i \ \sin\theta, \forall i \in \mathcal{I}\}$ 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 2-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 us 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 seventies [2]. Many works have since been proposed to generalize the Hough transform to more complex shapes than straight lines, and also, to improve its computational efficiency [3]. The Hough transform has also 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 $\hat{p}_{\theta\rho}(\theta, \rho)$ of the parameter (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. It is therefore important to extract the most relevant information from all available observations to model the distribution of (θ, ρ).

The focus of this article is to extend some recent developments on the Hough transform [6], [7] that propose to encapsulate more prior information on the variables (ρ, θ). This time, the available set of observations includes not only the spatial positions but also the spatial derivatives of the image. Using the gradients of the image, both the magnitudes and directions, allows us to estimate statistics of interest about the Hough variables θ and ρ. After a short review in section II that sets the context of this study, following [6], [7], we propose to infer some statistics of interest of the Hough variables in paragraph III. To overcome the discrete modelling using histograms, an estimate of the probability density function of the Hough variables using kernel mixtures with a variable bandwidth [8], is proposed in section V. The resulting density is continuous, and, thanks to the prior information extracted from the spatial derivatives of the image, the pre-segmentation of the edges of the image is shown to be an unnecessary preliminary step for selecting observations. The modelling of the probability density function $p(\theta, \rho)$ benefits then from all available pixels in the image and not only from the edges. The Standard Hough Transform is also presented as a particular case of this new approach named Statistical Hough Transform. The Hough transform has been used in many applications and for illustration here, some results of line detection on real images from snooker and tennis broadcasts [9], [10], and road scene footage [11] are shown.

II. Context

A. Local appearance based measures

Schmid et al. [12] have defined local descriptors of the intensity surface of images to detect interesting points (e.g. corners) and to match images. Using similar local appearance based features, Schiele et al. [13] have proposed to model their distributions using multi-dimensional histograms, and detection and recognition of objects can then be performed by comparing histograms. More recently, thanks to the increasing computational power of computers, kernel modelling [14] succeeded histograms for the modelling 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 [15].

In [11], three local appearance based measures are defined:

$$\begin{cases} \|\nabla I(x,y)\| = \sqrt{I_x^2(x,y) + I_y^2(x,y)} \\ \theta(x,y) = \arctan\left(\frac{I_y(x,y)}{I_x(x,y)}\right) \\ \rho(x,y) = x \cdot \frac{I_x(x,y)}{\|\nabla I(x,y)\|} + y \cdot \frac{I_y(x,y)}{\|\nabla I(x,y)\|} \end{cases} \tag{2}$$

where $\|\nabla I(x, y)\|$ is the magnitude of the gradient, ρ is the alignment, and θ is the angle of the gradient at the position (x, y). $(\theta(x, y), \rho(x, y))$ gives locally an estimate of the Hough parameters. No edge segmentation is performed in [11]: the angle and alignment are computed everywhere in the image, and their distribution (joint with the magnitude of the gradient) is used to detect appearing, disappearing or changing objects in a sequence [11]. The distribution $p(\theta, \rho, \|\nabla I\|)$ is estimated by a 3-dimensional histogram computed using the observations $\{(\theta_i, \rho_i, \|\nabla I\|_i)\}$ measured for each pixel i [11]. Integrating this 3-D histogram w.r.t. the magnitude of the gradient gives an estimate (2-D histogram) of the density function $p(\theta, \rho)$ that is noisy since all pixels in the image have been used (e.g. see figure 1(c)). Further analysis is performed in [16] to understand in particular

the behaviour of the angle θ. Its variance is computed locally and it is shown to be inversely proportional to the magnitude of the gradient. When an edge occurs at the position (x, y), the magnitude of the gradient is high, and the measure $\theta(x, y)$ is an observation with a low variance (or low error). On the contrary when no edge exists at the position (x, y), the magnitude of the gradient is low, and the measure $\theta(x, y)$ is an observation with a high variance (or high error). This approach is described and extended in paragraph III.

B. Reliability, confidence and uncertainty of local features

As just illustrated with the local measure of the angle, image content and noise can alter the quality of the computed or estimated 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 et al. [17] studied the uncertainty of the spatial localisation of a corner detector. The notion of *repeatability* of a detector [12], 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. In section III-B, the only perturbation considered is a Gaussian noise on the image. The reliability or certainty of the measure of the angle of the gradient and its alignment under this perturbation, is then assessed.

C. Recent works on the Hough Transform

Many works have been published on the Hough Transform since its first publication [1]. Recently, Aggarwal et al. proposed to robustly detect lines in noisy environment in the Hough space by adding prior modelling on the variables (θ, ρ) [18]. Several probabilistic Hough transforms, related to the RANSAC approach [19], have also been proposed [20], [21]. Of particular interest for this article is the Meanshift clustering approach in the Hough domain proposed by Bandera et al. [21], where a continuous kernel based modelling of $p(\theta, \rho)$ with variable bandwidth is introduced. However their modelling is deduced from a very different approach and their resulting process requieres many parameters to be manually tuned.

This short state of the art is focusing mainly on the works of Ji and Haralick [6], and Bonci and Karl [7], pursuing the same purpose as [20], [21], [18] of robustly estimating the lines.

Ji et al. [6] has proposed to locally approximate an image by a plane:

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

Consequently, an estimate of the Hough parameters is given by:

$$\begin{cases} \hat{\theta}(x, y) = \frac{\hat{\alpha}}{\hat{\beta}} \\ \\ \hat{\rho}(x, y) = x \cos \hat{\theta}(x, y) + y \sin \hat{\theta}(x, y) \end{cases} \tag{4}$$

where $\hat{\alpha}$ and $\hat{\beta}$ are locally estimated using least squares in equation (3). Estimates of the variances σ_θ^2, σ_ρ^2 and covariance $\sigma_{\theta\rho}$ are also computed [6], [7]. Considering all available estimated parameters $\{(\theta_i, \rho_i, \sigma_{\theta i}^2, \sigma_{\rho i}^2, \sigma_{\theta\rho i})\}$, a 2-dimensional weighted histogram estimates the distribution $\hat{p}(\theta, \rho)$ using the inverse of the variances as weights.

III. Statistics of local Hough features

A. Prior hypotheses

Assuming a noisy image $I(\mathbf{x})$ defined as:

$$I(\mathbf{x}) = d(\mathbf{x}) + b(\mathbf{x}) \tag{5}$$

with $d(\mathbf{x})$ the deterministic clean signal, $b(\mathbf{x})$ the noise and $\mathbf{x} = (x, y)$ is the position on the image surface. We assume that the distribution of the two derivatives of the image are normal:

$$\begin{cases} I_x(x, y) \sim \mathcal{N}(d_x(x, y), \sigma^2) \\ I_y(x, y) \sim \mathcal{N}(d_y(x, y), \sigma^2) \end{cases} \tag{6}$$

The position $\mathbf{x} = (x, y)$ is also modelled as a random variable that has the distribution $x \sim \mathcal{N}(\mu_x, \sigma_x^2)$ and $y \sim \mathcal{N}(\mu_y, \sigma_y^2)$. In the following paragraph, we derive statistical properties of the angle of the gradient θ and the alignment ρ that are defined with respect to I_x, I_y, x and y.

B. Statistical moments of θ

Let's define the ratio variable $z = \frac{I_y}{I_x}$. Assuming $\frac{b_x}{d_x} << 1$, the variable z can be approximated by:

$$z \quad = \frac{I_y}{I_x} \approx \frac{d_y + b_y}{d_x} \tag{7}$$

The mean and the variance of z can then be inferred:

$$\begin{cases} \mathbb{E}[z] = \frac{d_y}{d_x} \\ \\ \mathbb{V}[z] = \frac{\sigma^2}{d_x^2} \end{cases} \tag{8}$$

Applying the Delta method (c.f. appendix I), the statistics of the angle $\theta = \arctan z$ are:

$$\begin{cases} \mathbb{E}[\theta] = \mu_\theta \approx \arctan\left(\frac{d_y}{d_x}\right) \\ \\ \mathbb{V}[\theta] = \sigma_\theta^2 \approx \frac{\sigma^2}{d_x^2 + d_y^2} \end{cases} \tag{9}$$

In the case where I_x is small and $I_y >> b_y$, by defining $\theta = \operatorname{arccotan} \frac{I_x}{I_y}$, it is easy to show that the same result holds for the variance of θ in equation (9):

$$\begin{cases} \mathbb{E}[\theta] = \mu_\theta \approx \operatorname{arccotan}\left(\frac{d_x}{d_y}\right) \\ \\ \mathbb{V}[\theta] = \sigma_\theta^2 \approx \frac{\sigma^2}{d_x^2 + d_y^2} \end{cases} \tag{10}$$

Finally in the case when both variables (I_x, I_y) are centred on $d_x = d_y = 0$, the resulting distribution of θ is uniform on the interval $[-\pi/2; \pi/2]$. A short proof is given in appendix II. It means that when the region of the image is flat, the measurement of θ can take any value in the interval $[-\pi/2; \pi/2]$. It also means that if we observed two instances of the same scene $d(x, y)$ ($I_1(x, y) = d(x, y) + b_1(x, y)$ and $I_2(x, y) = d(x, y) + b_2(x, y)$), then the measurements of θ in flat areas are going to be randomly sampled from a uniform distribution. Therefore it is unlikely that the measurements 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 and therefore should accurately repeat itself from one instance of an image I_1 to another I_2.

C. Statistical moments of ρ

Using the relation in between ρ and θ (cf. equation (1)) and the properties listed in appendix I, the alignment is a function of three independent variables $\rho = \rho(x, y, \theta)$ and has the following statistics:

$$\mathbb{E}[\rho] = \mu_\rho \approx \mu_x \cdot \cos(\mu_\theta) + \mu_y \cdot \sin(\mu_\theta) \tag{11}$$

and the variance is:

$$\sigma_\rho^2 \approx (\cos\mu_\theta)^2\ \sigma_x^2 + (\sin\mu_\theta)^2\ \sigma_y^2 + \sigma_\theta^2(\mu_y\cos\mu_\theta - \mu_x\sin\mu_\theta)^2 \tag{12}$$

D. Remarks

The expression of the variances of θ and ρ are conformed to the ones found by Ji and Haralick [6]. A similar result for the variance of θ has also been found in [22]. This confirms the intuition that the uncertainty of the estimated orientation increases as the gradient magnitude decreases. Also, as noticed in [6], the origin of the spatial coordinates is better chosen in the centre of the image to limit the error done on the feature ρ, since its variance (c.f. equation (12)) depends on the location (μ_x, μ_y).

IV. Estimation of the statistics

A. Estimation of the variance of the noise

Lets consider the magnitude of the gradient of the image as a random variable:

$$\|\nabla I\| = \sqrt{I_x^2 + I_y^2} \tag{13}$$

The distribution of $\|\nabla I\|$ is a Rayleigh distribution when $d_x = d_y = 0$ (i.e. when the gradient is computed on a locally constant image surface). Assuming that a large proportion of the image has flat regions ($d_x = d_y = 0$) then the estimation of the standard deviation σ of the noise on the derivative can be robustly performed by locating the maximum of the distribution of the magnitude $\|\nabla I\|$ computed on the image [23], [24].

B. Estimation of $\mu_\theta, \sigma_\theta, \mu_\rho, \sigma_\rho$

Knowing one observation $(x_i, y_i, I_{x_i}, I_{y_i})$, assuming a Gaussian distribution for the variable θ and using the result in equation (9), the distribution of θ can be modelled by:

$$\theta \sim \mathcal{N}(\theta_i, \hat{\sigma}_{\theta_i}^2) \quad \text{with} \begin{cases} \theta_i = \arctan\frac{I_{y_i}}{I_{x_i}} \\ \hat{\sigma}_{\theta_i}^2 = \frac{\hat{\sigma}^2}{I_{x_i}^2 + I_{y_i}^2} \end{cases} \tag{14}$$

The derivatives (I_{x_i}, I_{y_i}) serve as an estimate of (d_{x_i}, d_{y_i}) since $\mathbb{E}[I_{x_i}] = d_{x_i}$ and $\mathbb{E}[I_{y_i}] = d_{y_i}$ (c.f. equation (6) with $I_{x_i} = I_x(x_i, y_i)$ and $I_{y_i} = I_y(x_i, y_i)$). Similarly, the variable ρ can be modelled by:

$$\rho \sim \mathcal{N}(\rho_i, \sigma_{\rho_i}^2) \quad \text{with}$$

$$\begin{cases} \rho_i = x_i\cos\theta_i + y_i\sin\theta_i \\ \hat{\sigma}_{\rho_i}^2 = (\cos\theta_i)^2\ \sigma_{x_i}^2 + (\sin\theta_i)^2\ \sigma_{y_i}^2 \\ \qquad + \hat{\sigma}_{\theta_i}^2(y_i\cos\theta_i - x_i\sin\theta_i)^2 \end{cases} \tag{15}$$

where the position (x_i, y_i) serves as an estimate for (μ_{x_i}, μ_{y_i}). The standard deviations $(\sigma_{x_i}, \sigma_{y_i})$ are manually set and chosen equal to 1 in the experiments section VII, to account for the digital image grid resolution.

V. Statistical Hough Transform

From each observation $(I_{x_i}, I_{y_i}, x_i, y_i)$, we can compute the values (θ_i, ρ_i) and their variances as presented in the previous section. We will define the following sets of observations:

- $\mathcal{S}_{\theta\rho} = \{(\theta_i, \rho_i)\}_{i=1\cdots N}$ with their respective standard deviations $\{(\hat{\sigma}_{\theta_i}, \hat{\sigma}_{\rho_i})\}_{i=1\cdots N}$,
- $\mathcal{S}_{\theta xy} = \{(\theta_i, x_i, y_i)\}_{i=1\cdots N}$ with their respective standard deviations $\{(\hat{\sigma}_{\theta_i}, \hat{\sigma}_{x_i}, \hat{\sigma}_{y_i})\}_{i=1\cdots N}$.
- $\mathcal{S}_{xy} = \{(x_i, y_i)\}_{i=1\cdots N}$ with their respective standard deviations $\{(\hat{\sigma}_{x_i}, \hat{\sigma}_{y_i})\}_{i=1\cdots N}$.

We can next estimate $p_{\theta\rho}(\theta, \rho)$ using kernels with those different sets of observations.

A. Kernel density modelling of $p_{\theta\rho}(\theta, \rho|\mathcal{S}_{\theta\rho})$

Using the set of observations $\mathcal{S}_{\theta\rho}$ with their estimated variances $\{(\hat{\sigma}_{\theta_i}, \hat{\sigma}_{\rho_i})\}_{i=1\cdots N}$ as variable bandwidths, we can model the distribution using kernels by:

$$\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{\theta\rho}) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_{\theta_i}}k_\theta\left(\frac{\theta - \theta_i}{\hat{\sigma}_{\theta_i}}\right)\cdot\frac{1}{\hat{\sigma}_{\rho_i}}k_\rho\left(\frac{\rho - \rho_i}{\hat{\sigma}_{\rho_i}}\right) \tag{16}$$

where N is the number of observations available. The kernels $k_\theta(\cdot)$ and $k_\rho(\cdot)$ are naturally chosen Gaussians. Equation (16) gives a continuous and smooth estimate of the density $p_{\theta\rho}(\theta, \rho)$. However, using all available observations, we can estimate a more precise probability density function $p_{\theta\rho}(\theta, \rho)$ as shown in the following paragraphs.

B. Kernel density modelling of $\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{\theta xy})$

In the previous paragraph, 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:

$$\hat{p}_{\theta\rho xy}(\theta, \rho, x, y|\mathcal{S}_{\theta xy}) = \hat{p}_{\rho|\theta xy}(\rho|\theta, x, y, \mathcal{S}_{\theta xy})\cdot\hat{p}_{\theta xy}(\theta, x, y|\mathcal{S}_{\theta xy}) \tag{17}$$

As noticed by Bonci et al. [7], when x, y, θ are known, the variable ρ is deterministic by definition in equation (1), therefore equation (17) becomes:

$$\hat{p}_{\theta\rho xy}(\theta, \rho, x, y|\mathcal{S}_{\theta xy}) = \delta(\rho - x\cos\theta - y\sin\theta)\cdot\hat{p}_{\theta xy}(\theta, x, y|\mathcal{S}_{\theta xy}) \tag{18}$$

where $\delta(\cdot)$ is the dirac distribution. As a consequence, only $\hat{p}_{\theta xy}(\theta, x, y)$ is to model using kernels:

$$\hat{p}_{\theta\rho xy}(\theta, \rho, x, y|\mathcal{S}_{\theta xy}) = \delta(\rho - x\cos\theta - y\sin\theta)\cdot\frac{1}{N}\cdot \sum_i\frac{1}{\hat{\sigma}_{x_i}\hat{\sigma}_{y_i}\hat{\sigma}_{\theta_i}}k_x\left(\frac{x - x_i}{\hat{\sigma}_{x_i}}\right)\cdot k_y\left(\frac{y - y_i}{\hat{\sigma}_{y_i}}\right)\cdot k_\theta\left(\frac{\theta - \theta_i}{\hat{\sigma}_{\theta_i}}\right) \tag{19}$$

By integrating with respect to the variables $\mathbf{x} = (x, y)$, an estimate of the Hough transform $p_{\theta\rho}(\theta, \rho)$ can be computed. This is done in the next paragraph considering two types of kernels for the spatial coordinates.

1) Modelling spatial variables x and y with dirac kernels: We model the distribution of both spatial variables x and y with dirac kernels, so for instance for x (similarly for y), we have:

$$\frac{1}{\hat{\sigma}_{x_i}} k_x \left(\frac{x - x_i}{\hat{\sigma}_{x_i}} \right) = \delta(x - x_i) \tag{20}$$

then the integral equation (19) w.r.t. (x, y) can be solved:

$$\hat{p}_{\theta\rho}(\theta, \rho | \mathcal{S}_{\theta xy}) = \frac{1}{N} \sum_i \frac{1}{\sigma_{\theta_i}} k_\theta \left(\frac{\theta - \theta_i}{\hat{\sigma}_{\theta_i}} \right) \times \delta(\rho - x_i \cos\theta - y_i \sin\theta) \tag{21}$$

However, this result does not account for the uncertainty on the spatial coordinates (i.e. the assumption of dirac kernels for the spatial coordinates does not leave room for uncertainty on their measurements).

2) Modelling spatial variables x and y with Gaussian kernels: We assume Gaussian kernels for the spatial coordinates, for instance for x:

$$\frac{1}{\hat{\sigma}_{x_i}} k_x \left(\frac{x - x_i}{\hat{\sigma}_{x_i}} \right) = \frac{1}{\sqrt{2\pi}\hat{\sigma}_{x_i}} \exp - \frac{1}{2} \left(\frac{x - x_i}{\hat{\sigma}_{x_i}} \right)^2 \tag{22}$$

The integration of equation (19) w.r.t. (x, y) is also feasible but more complex:

$$\hat{p}_{\theta\rho}(\theta, \rho | \mathcal{S}_{\theta xy}) = \frac{1}{N} \sum_i \frac{1}{\hat{\sigma}_{\theta_i}} k_\theta \left(\frac{\theta - \theta_i}{\hat{\sigma}_{\theta_i}} \right) I_i(\theta, \rho) \tag{23}$$

with the function $I_i(\theta, \rho)$ defined in appendix III.

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

Lets 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 by the uniform distribution such that:

$$\frac{k_\theta(\theta - \theta_i)}{\hat{\sigma}_{\theta_i}} = \frac{1}{\pi}$$

So assuming dirac kernels for the spatial coordinates, equation (21) becomes as expected:

$$\hat{p}_{\theta\rho}(\theta, \rho | \mathcal{S}_{xy}) = \frac{1}{N\pi} \sum_i \delta(\rho - x_i \cos\theta - y_i \sin\theta) \tag{24}$$

We recognise in equation (24) the mathematical expression of the process described in the introduction. Indeed each distribution $\delta(\rho - x_i \cos\theta - y_i \sin\theta)$ modelled the probability of (θ, ρ) to belong to the curve $\{C_i(\theta, \rho) : \ \rho = x_i \ \cos\theta + y_i \ \sin\theta\}$.

Assuming Gaussian kernels for the spatial coordinates, equation (23) then becomes:

$$\hat{p}_{\theta\rho}(\theta, \rho | \mathcal{S}_{xy}) = \frac{1}{N\pi} \sum_i I_i(\theta, \rho) \tag{25}$$

with the function $I_i(\theta, \rho)$ defined in appendix III.

D. Remarks

We have considered two hypotheses for the distribution of the spatial coordinates, dirac and Gaussian. Note that the dirac distribution can be seen as the Gaussian one with a variance null. Depending on the set of observations, we have shown that a continuous kernel modelling of the joint probability density function of the variables (θ, ρ) can be proposed. Our statistical framework also encapsulates nicely the Standard Hough Transform.

We are mainly interested here in the assessment of the estimates $\hat{p}_{\theta\rho}(\theta, \rho | \mathcal{S}_{\theta\rho})$ defined equation (16) and $\hat{p}_{\theta\rho}(\theta, \rho | \mathcal{S}_{\theta xy})$ defined equation (23). One major difference can already be spotted: the variance of ρ as defined in equation (15) and used as variable bandwidth in $\hat{p}_{\theta\rho}(\theta, \rho | \mathcal{S}_{\theta\rho})$, is depending on the spatial coordinates (x, y). This variance is then going to vary according to the distance of the observation from the origin, and points that are located far from the origin will have less impact on the overall distribution. This has already been noticed in section III-D and has motivated the choice of the origin of the coordinate system at the centre of the image to reduce the uncertainty on the variable ρ.

However the variance of ρ in $\hat{p}_{\theta\rho}(\theta, \rho | \mathcal{S}_{\theta xy})$, defined in equation (23), is not anymore dependent on the coordinates (x, y). As can be seen in the definition of the kernel $I_i(\theta, \rho)$ in equation (35), the variance of ρ is now equal to $\sigma_{y_i}^2 \sin^2\theta + \sigma_{x_i}^2 \cos^2\theta$ and using the hypothesis $\sigma_{y_i} = \sigma_{x_i} = \sigma_{\mathbf{x}}, \ \forall i$, it can then be simplified to $\sigma_{\mathbf{x}}^2$, the uncertainty on the spatial coordinates. The modelling of the distribution $\hat{p}_{\theta\rho}(\theta, \rho | \mathcal{S}_{\theta xy})$, defined in equation (23), is therefore not dependent on the choice of the origin of the coordinate system. This is a major difference with the $\hat{p}_{\theta\rho}(\theta, \rho | \mathcal{S}_{\theta\rho})$ defined in equation (16).

VI. Estimation of the modes

The modes of $\hat{p}_{\theta\rho}(\theta, \rho)$ indicate straight edges present in the picture. A natural way of finding the modes of the density function $\hat{p}(\theta, \rho)$ would be to apply a Meanshift search with the set of observations [8], [21]. Another approach used here, is to compute a discrete representation of $\hat{p}(\theta, \rho)$ 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. Those have been chosen $\delta_\theta = \pi/180$ and $\delta_\rho = 1$ in the experiments section VII. 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

$$l_\rho = \frac{w}{2} \cos\left(\arctan\frac{h}{w}\right) + \frac{h}{2} \sin\left(\arctan\frac{h}{w}\right) \tag{26}$$

The process to detect lines in images is summarized as follows:

1) Compute the derivatives of the image I_x, I_y. Here, we choose Gaussian filters with variance σ_g^2. Influence of the parameter σ_g will be shown in the next section.
2) Estimate the variance of the noise σ^2 from the distribution of the magnitude of the gradient.
3) Compute $\theta_i, \rho_i, \sigma_{\theta_i}, \sigma_{\rho_i}$ for each observation i in the image.
4) Compute the continuous densities $\hat{p}_{\theta\rho}(\theta, \rho | \mathcal{S}_{\theta\rho})$, as defined in equation (16), and $\hat{p}_{\theta\rho}(\theta, \rho | \mathcal{S}_{\theta xy})$, as defined in equation (23), on a discrete grid. One important characteristic of the angle is that it is defined modulo π and there is a discontinuity at the limit of the discrete space $-\pi/2$ and $\pi/2$. Indeed a horizontal edge can have points with either Hough coordinates $(-\pi/2, \rho)$ and $(\pi/2, -\rho)$. More generally, the point (θ, ρ) is equivalent to $(\theta + \pi, -\rho)$ and $(\theta - \pi, -\rho)$. Even if $\theta + \pi$ and $\theta - \pi$ are outside the discrete space, it is important to take into account the influence of their Gaussian kernel in the estimated mixture $\hat{p}_{\theta\rho}(\theta, \rho)$. In practice, this is achieved by duplicating all observations (θ_i, ρ_i) with $(\theta_i + \pi, -\rho_i)$ and $(\theta_i - \pi, -\rho_i)$ increasing artificially the cardinal of the set of observations from N to $3N$.

5) Detect the local maxima of the surfaces $\hat{p}_{\theta\rho}(\theta,\rho|\mathcal{S}_{\theta\rho})$ and $\hat{p}_{\theta\rho}(\theta,\rho|\mathcal{S}_{\theta xy})$, rank them in descending order and select the most significative ones to infer the corresponding lines.

VII. Experimental results

In the following experiments, we have set a priori the variance of the spatial coordinates $\sigma_{x_i}^2 = \sigma_{y_i}^2 = 1\ \forall i$. This choice is motivated by the natural grid of pixels in digital images. Also no particular pixels are selected in the image to compute the distributions, i.e. no contour segmentation has been performed to select observations: all the pixels in the images are used.

A. Synthetic image

The image diamond used in [7] is used here to compare the two distributions: $\hat{p}_{\theta\rho}(\theta,\rho|\mathcal{S}_{\theta\rho})$ is represented in figure 1(e) and $\hat{p}_{\theta\rho}(\theta,\rho|\mathcal{S}_{\theta xy})$ is shown in figure 1(f). Those two can be compared with a standard 2-dimensional histogram 1(c) and a weighted histogram 1(d) using $\frac{1}{2\pi\sigma_{\theta i}\sigma_{\rho i}}$ as weights as in [6], both computed on the observations $\mathcal{S}_{\theta\rho}$. The discrete sampling of the Hough space is the same for all the distribution. The continuous distributions are as expected, far smoother than the histograms and, $\hat{p}_{\theta\rho}(\theta,\rho|\mathcal{S}_{\theta xy})$ gives higher and sharper peaks than $\hat{p}_{\theta\rho}(\theta,\rho|\mathcal{S}_{\theta\rho})$. This can also be noticed in figure 5 where $\hat{p}_{\theta\rho}(\theta,\rho|\mathcal{S}_{\theta xy})$ gives finer information than $\hat{p}_{\theta\rho}(\theta,\rho|\mathcal{S}_{\theta\rho})$. Also despite the fact that all the pixels in the image have been used to compute these distributions, all the peaks corresponding to the straight edges are well detectable. Figure 1(b) shows the lines estimated using the maxima of the statistical Hough transform shown in figure 1(e).

B. Comparison of $\hat{p}_{\theta\rho}(\theta,\rho|\mathcal{S}_{\theta\rho})$ with simulation

To verify our model, 100 instances of the image 1(a) were generated with different instance of the noise on the derivatives ($\sigma = 10$). It means that for each pixel i, we can compute 100 times the values of the angle and the alignment from which we compute their average values $(\mu_{\theta_i}, \mu_{\rho_i})$ and variances $(\sigma_{\theta_i}, \sigma_{\rho_i})$ using standard formula of statistics. Using these estimates in equation (16) gives the probability density function represented in figure 2. This is to be compared with $\hat{p}_{\theta\rho}(\theta,\rho|\mathcal{S}_{\theta\rho})$ in figure 1(e) estimated with only one instance of the image. Peaks in the simulated distribution 2 are slightly higher and narrower than in the estimated distribution 1(e). It shows that the computed variances of θ and ρ are slightly over-estimated. One major difference appears at the peaks at $-\pi/2$ and $\pi/2$. 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 in two. Consequently the simulated variance for this line is very large which flattens the peaks located at $\theta = -\pi/2$ and $\theta = \pi/2$.

C. Resistance to noise

Figure 3 presents the statistical Hough transform $\hat{p}_{\theta\rho}(\theta,\rho|\mathcal{S}_{\theta\rho})$ computed on the image 1(a) with different size of Gaussian filters. σ_g, the standard deviation of the filter, controls the size of the template used to compute the spatial derivative I_x and I_y. It is related to the size of the neighbourhood used by Ji et al. in [6] for their local least square estimation of $I_x(x,y)$ and $I_y(x,y)$. The larger this neighbourhood, the more numerous the observations used to perform the estimation that becomes more accurate and more reliable. The Gaussian filter offers the advantage over the least square estimation of weighting the neighbourhood according to the their relative position from the centre (x,y) of the window. If the neighbourhood (or σ_g) is too small, the statistical Hough transform is more sensitive to noise. Since the uncertainty increases with the noise, the distribution gets closer to a uniform distribution, drowning the relevant peaks.

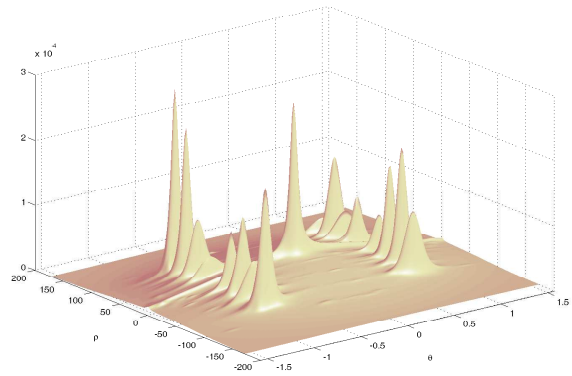


Fig. 2. Simulated statistical Hough transform (read paragraph VII-B).

Figure 4 shows the influence of both the noise level and the choice of the filter for the detection of the peaks. Very good resistance to noise can be observed with $\sigma_g = 1$.

The Hough transform is used in many applications such as sport video analysis. Figure 5 shows the result of line detection performed using our method on real images. No pre-selection of the observations has been performed before the computation of the statistical Hough transform. More natural scenes such as road scene images [11] show more texture that generate spurious peaks in the Hough space. However the first maxima are very informative on the geometry of the road and also for detecting the vanishing point. All the important straight edges are well detected in images of the tennis court and the snooker table, and even the very close parallels are distinguishable on the snooker table. One can also notice that, as expected, the estimate $\hat{p}_{\theta\rho}(\theta,\rho|\mathcal{S}_{\theta xy})$ gives finer details than $\hat{p}_{\theta\rho}(\theta,\rho|\mathcal{S}_{\theta\rho})$.

D. Computation

The computation of the Statistical Hough Transform is more intensive than the standard one, mainly because there is no prior selection of the observations (i.e. segmentation of the edges) and consequently the number of observations is far more important. Current implementation to compute $\hat{p}_{\theta\rho}(\theta,\rho|\mathcal{S}_{\theta\rho})$ in Matlab needs approximately 70s for a 290×290 (equivalent to $290 \times 290 \times 3$ kernels in the density functions) image on a computer with the following spec: CPU Intel Pentium(R) M processor 1.73GHz and 2GB memory. The matlab function is giving the possibility to split the image into several part to compute the probability density function into several lower memory demanding steps. This method is in fact very suitable for speed optimisation with parallel computing on multicore architecture.

VIII. Conclusion and future work

In this paper, it has been shown that using local appearance-based measures of angle and alignment computed with their variances, two probability density functions of the Hough parameters

(θ, ρ) can be modelled using kernels with variable bandwidths. Those distributions are continuous and smooth by nature. The detection of the maxima is easy and robust to noise. This new estimates require more time for computation but eliminate the prior need of edge segmentation.

The density $\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{\theta\rho})$ gives accurate results for detection of lines in images. However this can be improved by using $\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{\theta xy})$ that has the advantage of not being sensitive to the choice of the origin of the spatial coordinate system. Gaussian kernels have been used in this paper, however the proposed framework can easily be extended to other standard ones (e.g. Epanechnikov or Triangular kernels).

Appendix I
The Delta Method

Let x and y be two random variables linked by the relation $y = f(x)$ with f a smooth function that accepts a Taylor expansion. Knowing the first moments of x, mean μ_x and variance σ_x^2, the Delta method defines the moments of y [25]:

$$\begin{cases} \mathbb{E}[y = f(x)] = \mu_y \approx f(\mu_x) \\ \mathbb{V}[y = f(x)] = \sigma_y^2 \approx (f'(\mu_x))^2 \ \sigma_x^2 \end{cases} \tag{27}$$

It is generalisable to $y = f(x_1, x_2, \cdots, x_n)$ with $\{x_i\}_{i=1,\cdots,n}$, n independent random variables:

$$\begin{cases} \mathbb{E}[y = f(x_1, \cdots, x_n)] = \mu_y \approx f(\mu_{x_1}, \cdots, \mu_{x_n}) \\ \mathbb{V}[y = f(x_1, \cdots, x_n)] = \sigma_y^2 \approx \sum_{i=1}^{n} (\frac{\partial f}{\partial x_i}(\mu_{x_1}, \cdots, \mu_{x_n}))^2 \ \sigma_{x_i}^2 \end{cases} \tag{28}$$

Appendix II
Uniform regions $d_x = d_y = 0$

Considering the variable $z = \frac{I_y}{I_x}$, by definition [16], [26] , the probability density function of z is :

$$p_z(z) = \int\int_{D_z} p_{I_x,I_y}(I_x, I_y) \ dI_x dI_y \tag{29}$$

where $D_z = \{(I_x, I_y)|\frac{I_y}{I_x} \leq z\}$ and $p_{I_x,I_y}(.,.)$ is the joint probability density function of (I_x, I_y). By changing variable, equation (29) becomes:

$$p_z(z) = \int |I_x| \ p_{I_x,I_y}(I_x, zI_x) dI_x \tag{30}$$

Assuming the derivatives I_x and I_y independent, then we have $p_{I_x,I_y}(I_x, I_y) = p_{I_x} \cdot (I_x) \ p_{I_y}(I_y)$ with $p_{I_x}(I_x) = \mathcal{N}(0, \sigma^2)$ and $p_{I_y}(I_y) = \mathcal{N}(0, \sigma^2)$. So

$$\begin{aligned} p_z(z) &= \tfrac{1}{2\pi\sigma^2} \int |I_x| \ \exp^{-\frac{(1+z^2)I_x^2}{2\sigma^2}} \ dI_x \\ &= \tfrac{1}{\pi(1+z^2)} \end{aligned} \tag{31}$$

$p_z(z)$ is then the Cauchy distribution. Since the angle is defined as $\theta = \arctan z$ and $p_\theta(\theta) = p_z(\tan\theta)|\tan'\theta|$, then

$$p_\theta(\theta) = \frac{1}{\pi}, \quad \forall\theta \in \left[-\frac{\pi}{2}; \frac{\pi}{2}\right] \tag{32}$$

Appendix III
Integral $I_i(\theta, \rho)$

The integral to solve is:

$$\begin{aligned} I_i(\theta, \rho) = \frac{1}{2\pi\hat{\sigma}_{x_i}\hat{\sigma}_{y_i}} \int\int_{-\infty}^{+\infty} \delta(\rho - x\cos\theta - y\sin\theta) \\ \exp - \frac{(x - x_i)^2}{2\hat{\sigma}_{x_i}^2} \exp - \frac{(y - y_i)^2}{2\hat{\sigma}_{y_i}^2} \ dxdy \end{aligned} \tag{33}$$

By integration on x we get:

$$\begin{aligned} I_i(\theta, \rho) = \frac{1}{2\pi\hat{\sigma}_{x_i}\hat{\sigma}_{y_i}\cos\theta} \\ \int_{-\infty}^{+\infty} \exp - \frac{(\frac{\rho - y\sin\theta}{\cos\theta} - x_i)^2}{2\hat{\sigma}_{x_i}^2} \exp - \frac{(y - y_i)^2}{2\hat{\sigma}_{y_i}^2} dy \end{aligned} \tag{34}$$

By integrating w.r.t. y:

$$\begin{aligned} I_i(\theta, \rho) = \frac{1}{\sqrt{2\pi(\hat{\sigma}_{y_i}^2 \sin^2\theta + \hat{\sigma}_{x_i}^2 \cos^2\theta)}} \\ \exp\left(\frac{-(\rho - (x_i\cos\theta + y_i\sin\theta))^2}{2(\hat{\sigma}_{y_i}^2 \sin^2\theta + \hat{\sigma}_{x_i}^2 \cos^2\theta)}\right) \end{aligned} \tag{35}$$

Acknowledgment

Part of this work has been funded by the European Network of Excellence on *Multimedia Understanding through Semantics, Computation and Learning* MUSCLE FP6-5077-52, `www.muscle-noe.org`.

References

[1] P. Hough, "Methods of means for recognising 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," *Communications of the ACM*, vol. 15, pp. 11–15, January 1972.

[3] J.-Y. Goulermas and P. Liatsis, "Incorporating gradient estimations in a circle-finding probabilistic hough transform," *Pattern Analysis and Applications*, pp. 239–250, 1999.

[4] A. Goldenshluger and A. Zeevi, "The hough transform estimator," *The Annals of Statistics*, vol. 32, no. 5, October 2004.

[5] A. S. Aguado, E. Montiel, and M. S. Nixon, "Bias error analysis of the generalised hough transform," *Journal of Mathematical Imaging and Vision*, vol. 12, pp. 25–42, 2000.

[6] Q. Ji and R. M. Haralick, "Error propagation for the hough transform," *Pattern Recognition Letters*, vol. 22, pp. 813–823, 2001.

[7] A. Bonci, T. Leo, and S. Longhi, "A bayesian approach to the hough transform for line detection," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 35, no. 6, November 2005.

[8] D. Comaniciu, V. Ramesh, and P. Meer, "The variable bandwidth mean shift and data-driven scale selection," in *International Conference on Computer Vision*, Vancouver, Canada, July 2001, pp. 438–445.

[9] H. Denman, N. Rea, and A. Kokaram, "Content-based analysis for video from snooker broadcasts," *in Special Issue on Video Retrieval and Summarization, Journal of Computer Vision and Image Understanding*, vol. 92, pp. 141–306, 2003.

[10] A. Kokaram, N. Rea, R. Dahyot, M. Tekalp, P. Bouthemy, P.Gros, and I. Sezan, "Browsing sports video: trends in sports-related indexing and retrieval work," *IEEE Signal Processing Magazine*, vol. 23, no. 2, pp. 47–58, 2006.

[11] R. Dahyot, P. Charbonnier, and F. Heitz, "Unsupervised statistical change detection in camera-in-motion video," in *IEEE proceedings of the International Conference on Image Processing*, Thessaloniki, Greece, October 2001.

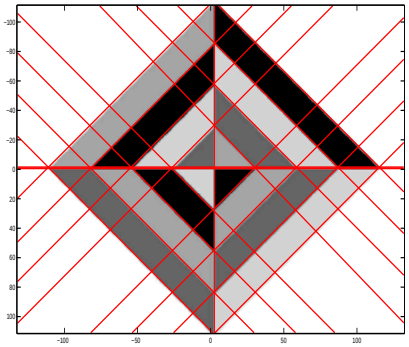
[12] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of interest point detectors," *International Journal of Computer Vision*, vol. 37, no. 2, pp. 151–172, 2000.

[13] B. Schiele and J. L. Crowley, "Recognition without correspondence using multidimensional receptive field histograms," *Journal International Journal of Computer Vision*, vol. 36, no. 1, pp. 31–50, January 2000.

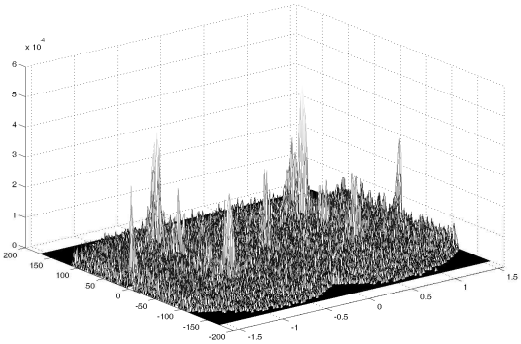
[14] C. Bishop, *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.
[15] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, May 2002.
[16] R. Dahyot, "Appearance based road scene video analysis for the management of the road network," Ph.D. dissertation, University of Strasbourg I, France, November 2001, published in french.
[17] R. M. Steele and C. Jaynes, "Feature uncertainty arising from covariant image noise," in *IEEE conference on Computer Vision and Pattern Recognition*, 2005, pp. 1063–1069.
[18] N. Aggarwal and W. C. Karl, "Line detection in image through regularized hough transform," *IEEE Transactions in Image Processing*, vol. 15, no. 3, March 2006.
[19] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
[20] D. Walsh and A. E. Raftery, "Accurate and efficient curve detection in images: the importance sampling hough transform," *Pattern Recognition Society*, vol. 35, pp. 1421–1431, 2002.
[21] 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.
[22] P. Meer and B. Georgescu, "Edge detection with embedded confidence," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 12, pp. 1351–1365, December 2001.
[23] R. Dahyot, N. Rea, A. Kokaram, and N. Kingsbury, "Inlier modeling for multimedia data analysis," in *IEEE International Workshop on MultiMedia Signal Processing*, Siena Italy, September 2004, pp. 482–485.
[24] 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.
[25] L. Wasserman, *All of Statistics - A concise course in statistical inference*. Springer, 2004.
[26] H. Stark and J. W. Woods, *Probability, Random Processes, and Estimation Theory For Engineers*, 2nd ed. Prentice Hall, 1994.

Rozenn Dahyot received a BSc in 1996 and MSc in 1998 in Physics and image processing from the University of Louis Pasteur (ULP) in Strasbourg France. In collaboration with the Laboratoire des Ponts et Chaussées, she received a PhD degree from ULP in 2001, working on robust detection and recognition of objects in road scene image databases, for the management of the road network. She then worked as a Marie Curie Research Fellow in Trinity College Dublin and in Cambridge university. Since 2005, she is a lecturer in the Computer Science department in Trinity College Dublin. Her research interests are in image, video and audio processing, object detection and recognition and statistical learning, amongst others.

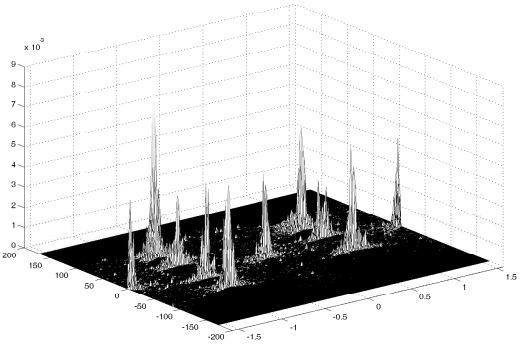
(a) Image *diamond*



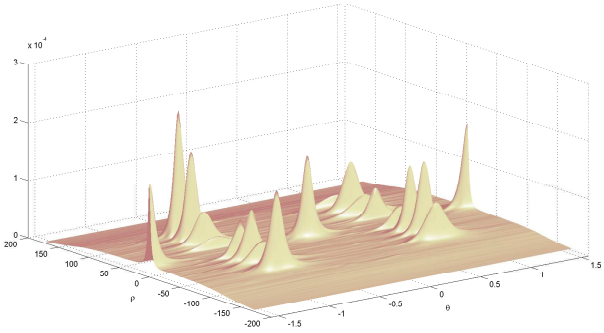
(b) Detected lines



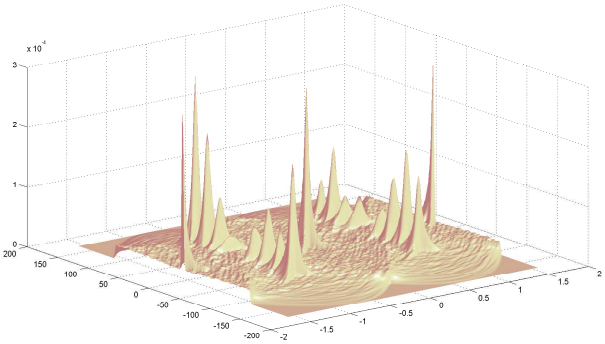
(c) Histogram



(d) Weighted histogram [6]



(e) Statistical hough transform $\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{\theta\rho})$



(f) Statistical hough transform $\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{\theta xy})$

Fig. 1. Different estimates of the probability density function $p_{\theta\rho}(\theta, \rho)$ on the image *diamond* 1(a) used in [7] with noise $\sigma = 10$.

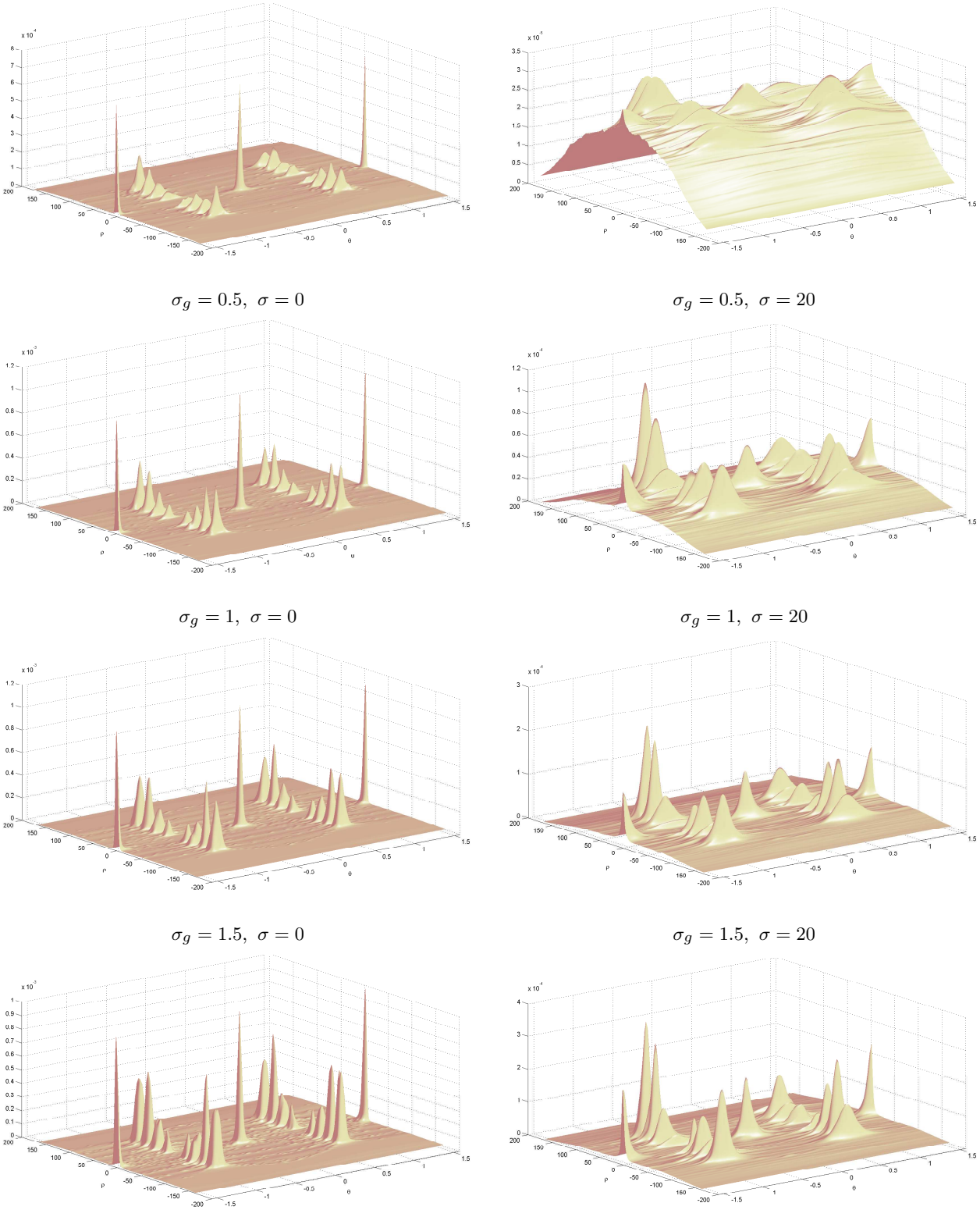


Fig. 3. Probability density function $\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{\theta\rho})$ computed on the image 1(a) with different noise level (σ) and different filters (σ_g).

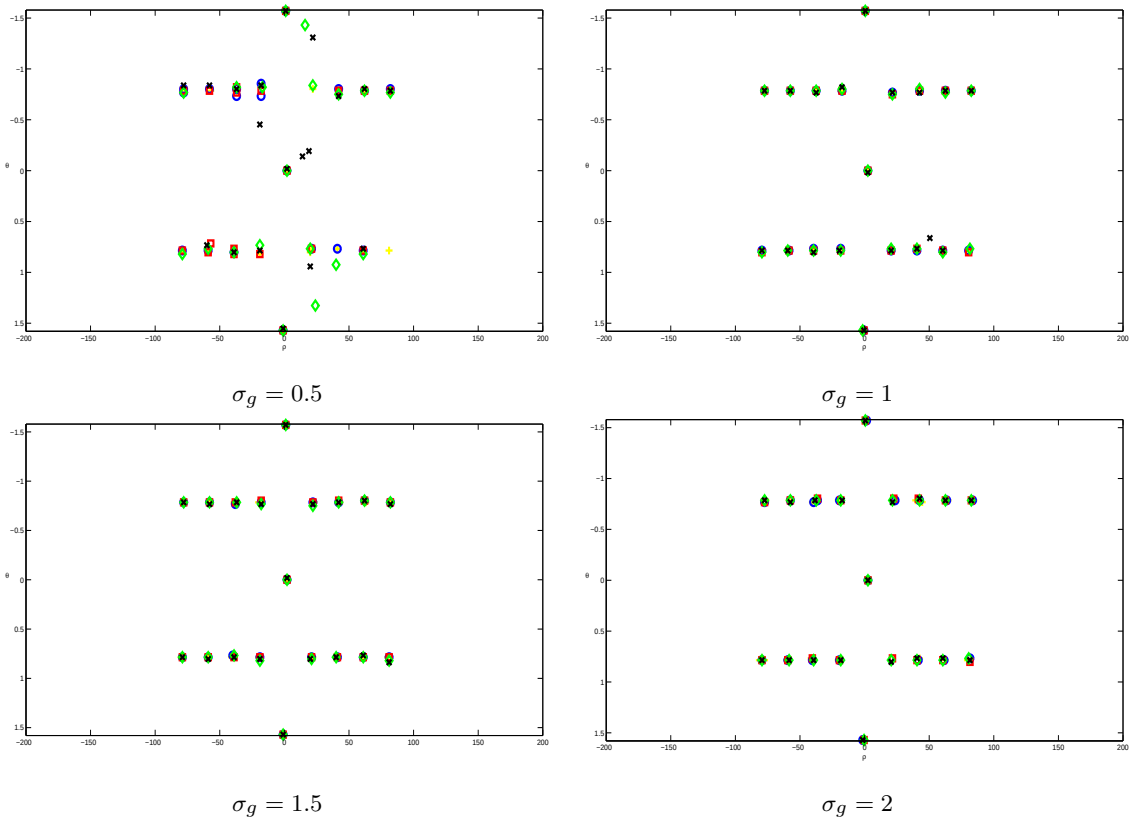


Fig. 4. Localisation of the first 19 maxima corresponding to the lines in image 1(a) for different filters (σ_g) and different noise level (σ).

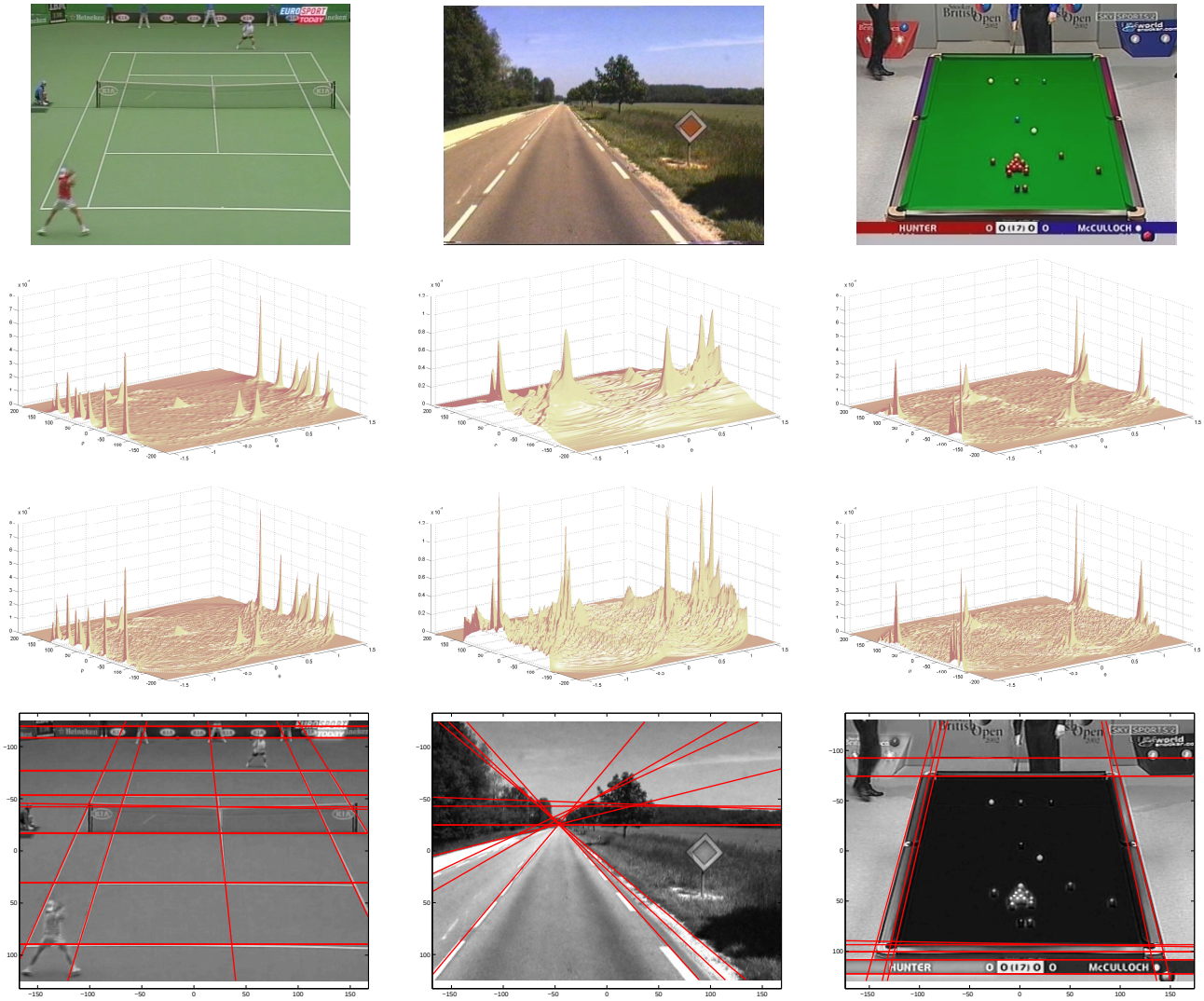


Fig. 5. Line detection. From top to bottom: real colour images, corresponding density $\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{\theta\rho})$, $\hat{p}_{\theta\rho}(\theta, \rho|\mathcal{S}_{\theta xy})$ and detected lines superimposed on the images.