Old and new straight-line detectors: Description and comparison

S. El Mejdani*, R. Egli, F. Dubeau

Laboratoire MOIVRE, Département d’informatique, Faculté des sciences, Université de Sherbrooke, 2500, boul. de l’Université, Sherbrooke, Qué., Canada J1K2R1

Received 28 September 2005; received in revised form 23 August 2007; accepted 13 November 2007

Abstract

Straight-line detection is important in several fields such as robotics, remote sensing, and imagery. The objective of this paper is to present several methods, old and new, used for straight-line detection. We begin by reviewing the standard Hough transform (SHT), then three new methods are suggested: the revisited Hough transform (RHT), the parallel-axis transform (PAT), and the circle transform (CT). These transforms utilize a point-line duality to detect straight lines in an image. The RHT and the PAT should be faster than the SHT and the CT because they use line segments whereas the SHT uses sinusoids and CT uses circles. Moreover, the PAT, RHT, and CT use additions and multiplications whereas the SHT uses trigonometric functions (sine and cosine) for calculation. To compare the methods we analyze the distribution of the frequencies in the accumulators and observe the effect on the detection of false local maxima. We also compare the robustness to noise of the four transforms. Finally, an example with a real image is given.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Straight-line detection; Point-line duality; Hough transform; Parallel-axis transform; Circle transform

1. Introduction

The goal of this paper is to propose and analyze old and new methods for detecting straight lines in images, using a point-line duality approach. After a review of the standard Hough transform (SHT), three new methods are proposed. These new methods are: the revisited Hough transform (RHT), which uses the Cartesian equation of a line; the parallel-axis transform (PAT), which is obtained from the parallel-axis coordinate system; and the circle transform (CT), which uses the normal equation of a line.

Point-line duality was introduced by Hough [1] in 1962 to detect straight lines in an image. It was extended by Rosenfeld [2] in 1969 and by Duda and Hart [3] in 1972. They developed the (standard) Hough transform, which has since been a traditional method for straight-line detection in images. It has been used in a number of applications: for road detection in satellite images by Geman and Jedynak [4], for robot localization by Hoppenot et al. [5], for robust bar-code reading by Muniz et al. [7], and for the detection of boat wakes by Rey et al. [8] and Magli et al. [9]. Since this method uses sinusoids, Tuytelaars et al. [10] have recently suggested an extension of the SHT, called the cascade Hough transform (CHT), which uses only straight lines. We also present an extension of the SHT which uses only straight lines and which is simpler than both the CHT and SHT. This method has recently been introduced simultaneously and independently in Refs. [11,12].

A point-line duality also appears in the parallel-axis coordinate system. A point in \( \mathbb{R}^n \) is represented by a poly-line in the parallel-axis coordinate system and conversely, a poly-line represents a point in \( \mathbb{R}^n \). The parallel-axis coordinate system is used for the visualization of multidimensional information. Sets of data in several dimensions are visualized by drawing a poly-line for each point through regularly spaced parallel axes. This technique introduced by Inselberg [13] and Inselberg and Dimsdale [14,15] has been used for several applications: in robotics by Cohan and Yang [16], in aerospace by Helly [17], in management by Desai and Walers [18], and in aerial navigation by Inselberg [19]. We use the parallel-axis coordinate system to develop a new method of straight-line detection as we did for the RHT [11].
In Section 2, we present and review how the SHT is used to detect the points aligned on a straight line in an image. We give the definition and properties of the SHT and describe its implementation. Section 3 is devoted to the RHT. This transform uses the Cartesian equation of a line directly. We present its definition and properties and also describe its implementation. In Section 4, we present a new method called PAT.

First we introduce the parallel-axis frame of reference proposed by Inselberg [13]. Then we give the definition of the PAT, which uses a parallel-axis representation of the plane and the point-line duality to define a new algorithm for straight-line detection in an image. We present the properties of this transform and describe its implementation. The CT is presented in Section 5. We give the definition of the CT, which uses a point-line duality to define a new algorithm for straight-line detection in an image. It is in fact the Hough transform but with a different parameter space [20]. We also present the properties of this new transform and discuss its implementation. In Section 6 we compare the four methods under four different aspects. We consider the transformation of the white image and the corresponding distributions of the votes in the accumulators. Then we illustrate the robustness to the occurrence of artifacts by detecting parallel lines. The analysis of the robustness of the methods to noise follows. Finally, we apply the methods on a real image. Conclusions are given in Section 7.

2. Standard Hough transform

The SHT was introduced by Hough [1] and extended by Rosenfeld [2] and by Duda and Hart [3]. We briefly review this method and point out its underlying point-line duality.

2.1. Definition

The SHT is based on the equation

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

If \( \theta \) and \( \rho \) are fixed, Eq. (1) is the normal equation of a line, where \((\cos \theta, \sin \theta)\) is its normal vector and \(|\rho|\) its distance from the origin. On the other hand, if \( x \) and \( y \) are fixed, Eq. (1) is the equation of a sinusoid. Indeed, if we set \((x, y) = r(\cos \phi, \sin \phi)\), then Eq. (1) becomes

\[ \rho = r \cos(\theta - \phi). \tag{2} \]

From Eq. (1) we define two applications.

For the first application, with \( P(x_0, y_0) \in \mathbb{R}^2 \) we associate the set \( \mathcal{S}(x_0, y_0) \in 2^\mathbb{R}^2 \) defined by

\[ \mathcal{S}(x_0, y_0) = \{ (\theta, \rho) \in \mathbb{R}^2 | x_0 \cos \theta + y_0 \sin \theta = \rho \}. \]

For any \((\theta, \rho) \in \mathcal{S}(x_0, y_0)\), Eq. (1) is a line passing through \( P(x_0, y_0) \). Conversely, for any line passing through \( P(x_0, y_0) \) there exists \((\theta, \rho) \in \mathcal{S}(x_0, y_0)\) such that Eq. (1) is the equation of that line. Hence, \( \mathcal{S}(x_0, y_0) \) in fact represents the set of all lines in \( \mathbb{R}^2 \) passing through \( P(x_0, y_0) \) (Fig. 1).

For the second application, with any \( Q(\theta_0, \rho_0) \in \mathbb{R}^2 \) we associate the set \( \mathcal{D}(\theta_0, \rho_0) \in 2^\mathbb{R}^2 \) defined by

\[ \mathcal{D}(\theta_0, \rho_0) = \{ (x, y) \in \mathbb{R}^2 | x \cos \theta_0 + y \sin \theta_0 = \rho_0 \}. \]

For any \((x, y) \in \mathcal{D}(\theta_0, \rho_0)\), Eq. (1) is a sinusoid passing through \( Q(\theta_0, \rho_0) \). Conversely, for any sinusoid passing through \( Q(\theta_0, \rho_0) \) there exists \((x, y) \in \mathcal{D}(\theta_0, \rho_0)\) such that Eq. (1) is the equation of that sinusoid. Hence, \( \mathcal{D}(\theta_0, \rho_0) \) in fact represents the set of all sinusoids passing through \((\theta_0, \rho_0)\) (Fig. 2).

The two applications from \( \mathbb{R}^2 \) to \( 2^\mathbb{R}^2 \) given by

\[ P(x_0, y_0) \rightarrow \mathcal{S}(x_0, y_0) \tag{3} \]

and

\[ Q(\theta_0, \rho_0) \rightarrow \mathcal{D}(\theta_0, \rho_0) \tag{4} \]

determine or define the SHT.

2.2. Properties of the SHT

An image \( \mathcal{I} \) is described by the Cartesian coordinates of its points. We denote by \( \mathcal{H} \) the parameter space of coordinates \((\theta, \rho)\) associated with the lines. According to its definition, the SHT has the following properties:

**Property 1.** A point in an image \( \mathcal{I} \) corresponds to a sinusoid in the parameter space \( \mathcal{H} \).

**Property 2.** A point in the parameter space \( \mathcal{H} \) corresponds to a line in the image \( \mathcal{I} \).
Property 3. Points on a given line in an image $\mathcal{I}$ define sinusoids in the parameter space $\mathcal{H}$ which intersect at the same point.

Property 4. Points on a given sinusoid in the parameter space $\mathcal{H}$ correspond to lines in the image $\mathcal{I}$ which intersect at the same point.

Property 5. Using the preceding notation, we have:

- if $(\theta, \rho) \in \mathcal{H}(x_0, y_0)$ then $(\theta + l\pi, (-1)^l \rho) \in \mathcal{H}(x_0, y_0)$ for all $l \in \mathbb{Z}$;
- $\mathcal{H}(\theta, \rho) = \mathcal{H}(\theta + l\pi, (-1)^l \rho)$ for all $l \in \mathbb{Z}$.

From the last property, we use only the parameter space for $\theta$ in an interval of width $\pi$. Indeed, if sinusoids intersect at the point $(\theta, \rho)$, they intersect also at the points $(\theta + l\pi, (-1)^l \rho)$ for all $l \in \mathbb{Z}$.

2.3. Implementation

We consider gray-level images $\mathcal{I}$ having $(N \times N)$ pixels. The SHT algorithm is implemented in three stages.

Stage 1. Pre-treatment of the image $\mathcal{I}$ (binary image).

Stage 2. Transformation of edge points of the image $\mathcal{I}$ into sinusoids in the parameter space $\mathcal{H}$. With each edge point pixel $P(x_i, y_j)$ in the image $\mathcal{I}$ we associate the sinusoid of equation $\rho(\theta) = x_i \cos \theta + y_j \sin \theta$ and we increment the vote of all cells corresponding to the points on this sinusoid.

Stage 3. Detection of the points in the parameter space $\mathcal{H}$ through which the sinusoids pass most frequently (a search for local maxima called peaks).

3. Revisited Hough transform

We propose here a new version of the Hough transform based on the Cartesian equation of a line (see Refs. [11,12]). A modification of the SHT was proposed by Tuytelaars et al. [10], but our approach is simpler.

3.1. Definition

Let us consider the equation

$$y = mx + b.$$  

If $m$ and $b$ are fixed, Eq. (5) is the Cartesian equation of a line of slope $m$ and y-intercept $b$. On the other hand, we can also fix $x$ and $y$ and again, we get the equation of a line with respect to $(m, b)$.

In Eq. (5), the slope $m$ can be large, even infinite in the vertical line case. It is this observation which made that the Hough transform was not used with Eq. (5). But let us observe that when the absolute value of the slope $m$ is greater than 1, $|m| \geq 1$, we can rewrite the equation for the line in the form

$$x = \bar{m}y + \bar{b}$$  

with $|\bar{m}| \leq 1$. Thus, any line can be written according to Eqs. (5) and/or (6). If $m = 0$ in Eq. (5), the line is horizontal and if $\bar{m} = 0$ in Eq. (6) the line is vertical.

We will denote these lines by

$$D(m, b) = \{(x, y) \in \mathbb{R}^2 | y = mx + b\}$$

and

$$\bar{D}(\bar{m}, \bar{b}) = \{(x, y) \in \mathbb{R}^2 | x = \bar{m}y + \bar{b}\}.$$  

A point $P$ of Cartesian coordinates $(x_0, y_0)$ can be identified with the intersection of all lines passing through it. We consider two cases.

Case 1: For each line Eq. (5) with $|m| \leq 1$ passing through $P(x_0, y_0)$ we have $y_0 = mx_0 + b$.

Thus, $(m, b)$ is such that

$$b = (-x_0)m + y_0$$  

and corresponds to the line segment

$$L(-x_0, y_0) = \{(m, b) \in \mathbb{R}^2 | |m| \leq 1 \text{ and } b = (-x_0)m + y_0\}.$$  

Conversely, for any $(m, b) \in L(-x_0, y_0)$, Eq. (5) is a line passing through $P(x_0, y_0)$. Hence, we have a correspondence defined by

$$P(x_0, y_0) \rightarrow L(-x_0, y_0)$$  

which associates with a point $P(x_0, y_0)$ all lines of slope $|m| \leq 1$ passing through it. This correspondence is illustrated in Fig. 3.

Case 2: In the same way, for any line Eq. (6) passing through $P(x_0, y_0)$ we have

$$x_0 = \bar{m}y_0 + \bar{b}.$$  

Thus, $(\bar{m}, \bar{b})$ is such that

$$\bar{b} = (-y_0)\bar{m} + x_0.$$
which corresponds to the line segment
\[ \bar{L}(\bar{x}_0, \bar{x}_0) = \{(\bar{m}, \bar{b}) \in \mathbb{R}^2 \mid |\bar{m}| \leq 1 \text{ and } \bar{b} = (-\bar{x}_0)\bar{m} + \bar{x}_0 \}. \]

Conversely, for any \((\bar{m}, \bar{b}) \in \bar{L}(\bar{x}_0, \bar{x}_0)\), Eq. (6) is a line passing through \(P(x_0, y_0)\). Hence, we have a correspondence defined by
\[ P(x_0, y_0) \rightarrow \bar{L}(\bar{x}_0, \bar{x}_0) \] (10)
which associates with a point \(P(x_0, y_0)\) all lines of slope \(|m| \geq 1\) passing through it. This correspondence is illustrated in Fig. 4.

To establish an inverse correspondence we proceed in a similar way by considering two cases.

Case 1: Let a point \(Q(m_0, b_0) \in [-1, 1] \times \mathbb{R}\). We identify this point \(Q(m_0, b_0)\) to the set of all lines passing through it. For each such line there exists \((x, y) \in \mathbb{R}^2\) such that the equation of the line is
\[ b = (-x)m + y. \]
As \(Q(m_0, b_0)\) is on this line, one must have
\[ b_0 = (-x)m_0 + y, \]
thus
\[ y = m_0x + b_0. \] (11)

Hence, we have an inverse transform, as illustrated in Fig. 5, from \([-1, 1] \times \mathbb{R}\) to \(2\mathbb{R}^2\) which associates with any point \(Q(m_0, b_0)\) the line \(D(m_0, b_0)\) of Eq. (11). We denote this correspondence by
\[ Q(m_0, b_0) \rightarrow D(m_0, b_0). \] (12)

Case 2: Let us consider a point \(\bar{Q}(\bar{m}_0, \bar{b}_0) \in [-1, 1] \times \mathbb{R}\). We identify this point \(\bar{Q}(\bar{m}_0, \bar{b}_0)\) with the set of all lines passing through it. For each such line there exists \((x, y) \in \mathbb{R}^2\) such that the Cartesian equation of the line is
\[ \bar{b} = (-y)\bar{m} + x. \]
As \(\bar{Q}(\bar{m}_0, \bar{b}_0)\) must be on each line, we must have
\[ \bar{b}_0 = (-y)\bar{m}_0 + x, \]
thus
\[ x = \bar{m}_0y + \bar{b}_0. \] (13)

We have established another inverse transform, illustrated in Fig. 6, from \([-1, 1] \times \mathbb{R}\) to \(2\mathbb{R}^2\) which associates with any point \(\bar{Q}(\bar{m}_0, \bar{b}_0)\) the line \(D(\bar{m}_0, \bar{b}_0)\) of Eq. (13). We denote this correspondence by
\[ \bar{Q}(\bar{m}_0, \bar{b}_0) \rightarrow \bar{D}(\bar{m}_0, \bar{b}_0). \] (14)

Relations (8), (10), (12), and (14) are used to define the RHT.

3.2. Properties of the RHT

Let us define by \(\mathbb{R}^H\) and \(\mathbb{R}^H\) the accumulators in the two bands or parameter spaces for the RHT. According to its definition, this transform has the following properties.

Property 1. A point in the image \(\mathcal{I}\) corresponds to two line segments in the parameter spaces \(\mathbb{R}^H\) and \(\mathbb{R}^H\).

Property 2. A point in either of the parameter spaces \(\mathbb{R}^H\) or \(\mathbb{R}^H\) corresponds to a line in the image \(\mathcal{I}\).

Property 3. Points on a given line in \(\mathcal{I}\) define line segments in the parameter spaces \(\mathbb{R}^H\) or \(\mathbb{R}^H\) which intersect at the same point.

Property 4. Points on a given line segment in the parameter spaces \(\mathbb{R}^H\) or \(\mathbb{R}^H\) correspond to lines in the image \(\mathcal{I}\) which intersect at the same point.

3.3. Implementation

We consider gray-level images \(\mathcal{I}\) having \((N \times N)\) pixels. The RHT algorithm is implemented in three stages.
Stage 1. Pre-treatment of the image $I$ (binary image).

Stage 2. Transformation of the edge points of the image $I$ into line segments in the parameter spaces $RH$ and $RH$.

We fix the origin at the center of the image $I$ and with each edge point pixel $P(x_i, y_j)$ in $I$ we associate two line segments:

(a) one in $RH$ of equation

$$b = (-x_i)m + y_j \quad (m \in [-1, 1]),$$

and we increment the vote of all cells corresponding to the points on this segment (Fig. 7);

(b) one in $RH$ of equation

$$\tilde{b} = (-y_j)\tilde{m} + x_i \quad (\tilde{m} \in [-1, 1]),$$

and we increment the vote of all cells corresponding to the points on this segment (Fig. 8).

Stage 3. Detection of the points in $RH$ and $RH$ through which the line segments pass most frequently (a search for local maxima called peaks).

4. Parallel-axis transform

Based on the system of parallel coordinates and its point-line duality, we present a new method for line detection in images [11].

4.1. Definitions

The PAT is based on two representations of the plane $\Pi$. The first representation uses the standard Cartesian coordinate system $XOY$ and the second representation is related to the parallel-axis system introduced by Inselberg [13].

We start with the plane $\Pi$, defined by the Cartesian coordinate system $XOY$, and a point $P(x_0, y_0)$. For the parallel-axis representation of the plane $\Pi$, let us place the $X$-axis on a line $D_x$ and the $Y$-axis on a line $D_y$, $D_x$ and $D_y$ being parallel, separated by a distance $d$ and perpendicular to a third line $D$. The parallel-axis system is denoted by $DD_xD_y$. With the point $P(x_0, y_0)$ in $XOY$ we associate a line passing through the points $A(0, x_0)$ and $B(d, y_0)$ in $DD_xD_y$, as illustrated in Fig. 9. We thus have a correspondence of the plane $\Pi$ with values in $2^\Pi$, defined by

$$P(x_0, y_0) \rightarrow \mathcal{L}(x_0, y_0) = \left\{ (\zeta, \eta) \in \mathbb{R}^2 | \eta = \left(\frac{y_0 - x_0}{d}\right)\zeta + x_0 \right\}. \quad (15)$$

Now let us consider the line expressed by the equation

$$D(m, b) = \{(x, y) \in \mathbb{R}^2 | y = mx + b\}$$

in the Cartesian coordinate system. With each point $P(x, y)$ on this line we associate the line $\mathcal{L}(x, y)$ in $DD_xD_y$. We have

$$\eta = \left(\frac{y - x}{d}\right)\zeta + x, \quad (16)$$

and we can write

$$\eta = \frac{mx + b - x}{d}\zeta + x = \frac{b}{d}\zeta + \left(1 - \frac{1 - m}{d}\right)x.$$

We observe that the point $Q\left(\frac{d}{1-m}, \frac{b}{1-m}\right)$ is common to all lines of Eq. (16). It is the point of intersection in $DD_xD_y$ of the
Fig. 10. Line-point correspondence.

Fig. 11. Point-line correspondence.

Finally, we consider a line in $D_x D_y$ given by the intersection with the $D_x$ and $D_y$ axes at $(0, x_0)$ and $(d, y_0)$. We thus have

$$\eta = \left( \frac{y_0 - x_0}{d} \right) \zeta + x_0. \tag{18}$$

To each point $Q(\zeta, \eta)$ on this line corresponds a line expressed in $XOY$ by the equation

$$y = \left( 1 - \frac{d}{\zeta} \right) x + \frac{d \eta}{\zeta},$$

Using relation (18) between $\zeta$ and $\eta$ we obtain the lines

$$y = \left( 1 - \frac{d}{\zeta} \right) (x - x_0) + y_0,$$

and we observe that the point $P(x_0, y_0)$ is common to all these lines. Thus, we have the correspondence depicted in Fig. 12, defined by

$${\mathcal L}(x_0, y_0) \rightarrow P(x_0, y_0). \tag{19}$$

The point in $D_x D_y$ associated with the line of equation $D(m, b)$ in $XOY$ has coordinates $(\frac{d}{1-m}, \frac{b}{1-m})$. It follows that $0 < \frac{d}{1-m} \leq d$ if and only if $m \leq 0$. Fig. 13 gives the value of the coordinate $\frac{d}{1-m}$ of the representative point for the line $D(m, b)$. Thus, each line of slope $m \leq 0$ has a representative point between the parallel axes $D_x$ and $D_y$.

We also observe a symmetry for this representation of the points between $D_x$ and $D_y$. For a slope $m \leq 0$, let $m = \tan(\theta)$ and $\tilde{m} = \tan(\frac{\pi}{2} + \theta) = \cot(\theta) = \frac{1}{m}$. We thus have

$$d - \frac{d}{1-m} = d - \frac{d}{1 - \tan(\theta)}$$

$$= d \left( 1 - \frac{1}{1 - \tan(\theta)} \right)$$

$$= \frac{d}{1 - \cot(\theta)}$$

$$= \frac{d}{1 - m}$$

Lines $\mathcal{L}(x, y)$ associated with the points $P(x, y)$ on $D(m, b)$. Associating this point with the line leads us to a line-point correspondence from $\mathcal{L}$ to $\Pi$, as shown in Fig. 10, given by

$$D(m, b) \rightarrow Q \left( \frac{d}{1-m}, \frac{b}{1-m} \right). \tag{17}$$

Any line can be associated with a point in this way, except for lines of slope $m = 1$ which give a set of parallel straight lines in $D_x D_y$.

Conversely, let us take a point $Q(\zeta_0, \eta_0)$ in $D_x D_y$. For any line passing through $Q(\zeta_0, \eta_0)$ and intersecting the $D_x$ and $D_y$ axes at $(0, x)$ and $(d, y)$, the relation between the values of $x$ and $y$ is given by the following equation:

$$y = \left( 1 - \frac{d}{\zeta_0} \right) x + \frac{d \eta_0}{\zeta_0},$$

which represents a line in the $XOY$ coordinate system, with slope

$$m = 1 - \frac{d}{\zeta_0} \left( \zeta_0 = \frac{d}{1-m} \right),$$

and $y$-intercept

$$b = d \frac{\eta_0}{\zeta_0} \left( \eta_0 = \frac{b}{1-m} \right).$$

We thus have the correspondence depicted in Fig. 11, defined by

$$Q(\zeta_0, \eta_0) \rightarrow D \left( 1 - \frac{d}{\zeta_0}, \frac{d \eta_0}{\zeta_0} \right).$$
Fig. 13. Horizontal coordinate $\frac{d}{1-m}$ of the representative point for the line $D(m, b)$.

Fig. 14. Symmetry of the representation. (a) Symmetry compared to line of slope $m = -1$. (b) Symmetry compared to $\frac{1}{2}$.

and similarly

$$b - \frac{b}{1 - m} = \frac{b}{1 - \tilde{m}}.$$

These values are illustrated in Fig. 14.

We have seen that lines of slope $m \leq 0$ have representative points in the system of parallel axes $D_x, D_y$ between the vertical lines $D_x$ and $D_y$ (Fig. 15). To obtain the lines of slope $m \geq 0$ in the same way, we simply change the axes so that the slope of the line becomes $\tilde{m} \leq 0$. It is sufficient to set

$$\tilde{y} = y \quad \text{and} \quad \tilde{x} = -x$$

to obtain

$$y = mx + b,$$

$$y = (-m)(-x) + b,$$

$$\tilde{y} = (-m)\tilde{x} + b$$

and by setting $\tilde{m} = -m$ and $\tilde{b} = b$ we obtain

$$\tilde{y} = \tilde{m}\tilde{x} + \tilde{b}.$$
Hence, when \( m \geq 0 \) we have \( \tilde{m} \leq 0 \). Lines of slope \( m \geq 0 \) thus have representative points in the system of parallel axes \( D_x D_y \) between the vertical lines \( D_x \) and \( D_y \) (Fig. 16).

To establish the PAT, we need the point, in \( D_x D_y \) or \( D_x D_y \), associated with a line in \( XOY \) to fall between the parallel lines \( D_x \) and \( D_y \), or \( D_x \) and \( D_y \). We distinguish between two cases.

**Case 1:** For a point \( P(x_0, y_0) \) and lines of slope \( -\infty \leq m \leq 0 \) passing through it, we consider the following correspondence (Fig. 17):

\[
P(x_0, y_0) \rightarrow \mathcal{L}(x_0, y_0)
\]

\[
= \left\{ (\zeta, \eta) \in \mathbb{R}^2 \mid \eta = \zeta \left( \frac{y_0 - x_0}{d} \right) + x_0 \right\}. \quad (20)
\]

**Case 2:** For a point \( P(x_0, y_0) \) and lines of slope \( -\infty \leq \tilde{m} \leq 0 \) (i.e. \( 0 \leq m \leq +\infty \)) passing through it, we consider the following correspondence (Fig. 18):

\[
P(x_0, y_0) \rightarrow \mathcal{L}(-x_0, y_0) = \left\{ (\zeta, \eta) \in \mathbb{R}^2 \mid \eta = \zeta \left( \frac{y_0 + x_0}{d} \right) - x_0 \right\}.
\]

\[
P(\tilde{x}_0, \tilde{y}_0) \rightarrow \mathcal{L}(\tilde{x}_0, \tilde{y}_0) = \left\{ (\zeta, \eta) \in \mathbb{R}^2 \mid \eta = \zeta \left( \frac{\tilde{y}_0 - \tilde{x}_0}{d} \right) + \tilde{x}_0 \right\}. \quad (21)
\]

Conversely, we have the following two correspondences.

**Case 1:** We have the correspondence illustrated in Fig. 19 which associates a line with any point \( Q(\zeta_0, \eta_0) \) between the parallel lines \( D_x \) and \( D_y \) (0 \( \leq \zeta_0 \leq d \)), as follows:

\[
Q(\zeta_0, \eta_0) \rightarrow D \left(1 - \frac{d}{\zeta_0}, d \frac{\eta_0}{\zeta_0}\right)
\]

\[
= \left\{ (x, y) \in \mathbb{R}^2 \mid y = \left(1 - \frac{d}{\zeta_0}\right)x + d \frac{\eta_0}{\zeta_0} \right\}. \quad (22)
\]

**Case 2:** Finally, we have the correspondence illustrated in Fig. 20 which associates a line with any point \( \tilde{Q}(\zeta_0, \tilde{\eta}_0) \) between the parallel lines \( D_x \) and \( D_y \) (0 \( \leq \zeta_0 \leq d \)), as follows:

\[
\tilde{Q}(\zeta_0, \tilde{\eta}_0) \rightarrow \tilde{D} \left(1 - \frac{d}{\zeta_0}, \tilde{\eta}_0 \right)
\]

\[
= \left\{ (x, y) \in \mathbb{R}^2 \mid y = \left(1 - \frac{d}{\zeta_0}\right)x + \tilde{\eta}_0 \right\}. \quad (23)
\]

Relations (20)–(23) are used to define the PAT.
4.2. Properties of the PAT

An image $I$ will be described using the Cartesian coordinates of its points. We denote by $\mathcal{P}I$ and $\mathcal{O}I$ the parameter spaces of the PAT. According to its definition, the PAT has the following properties.

Property 1. A point in the observation space $I$ corresponds to a line segment in each of the parameter spaces $\mathcal{P}I$ and $\mathcal{O}I$ (Figs. 17 and 18).

Property 2. Points aligned in the observation space give intersecting line segments in parameter spaces $\mathcal{P}I$ or $\mathcal{O}I$ (Figs. 17 and 18).

Property 3. A point in parameter space $\mathcal{P}I$ or $\mathcal{O}I$ corresponds to a line in the observation space $I$ (Figs. 19 and 20).

Property 4. Points belonging to the same line segment in the parameter spaces $\mathcal{P}I$ and $\mathcal{O}I$ give lines which meet at the same point in the observation space $I$ (Figs. 19 and 20).

4.3. Implementation

In our implementation, discrete steps are used for the observation and parameter spaces. The gray-level image $I$ is of dimension $N \times N$ pixels. The parameter spaces $\mathcal{P}I$ and $\mathcal{O}I$ both have $N_d \times N_{xy}$ cells (or pixels). We use $\mathcal{P}I$ and $\mathcal{O}I$ to record the number of times that a cell is traversed by the segments traced by the PAT. There are three stages in the implementation of the PAT algorithm.

Stage 1. Pre-treatment of the image $I$ (binary image).

Stage 2. Transformation of the edge point pixels of $I$ into segments in the parameter spaces $\mathcal{P}I$ and $\mathcal{O}I$. With each edge point pixel $P(x_i, y_j)$ in the observation space (image $I$) we associate two line segments:

(a) one in the parameter space $\mathcal{P}I$ joining the points $M(0, x_i)$ and $N(d, y_j)$ (Fig. 17), and we increment the vote of all cells corresponding to points on this segment;

(b) one in the parameter space $\mathcal{O}I$ joining the points $M(0, N-x_i)$ and $N(d, y_j)$ (Fig. 18), and we increment the vote of all cells corresponding to points on this segment.

Stage 3. Detection of the points having the most votes in $\mathcal{P}I$ and $\mathcal{O}I$, which determine pixel alignments in the image $I$ (a search for local maxima called peaks).

5. Circle transform

Like the SHT, the CT uses the normal equation of a line, but a different parameter space formed of circles. This idea was presented in Ref. [20].

5.1. Definition

The CT is based on the normal equation of a line

$$x \cos \theta + y \sin \theta = \rho.$$  \hspace{1cm} (24)

To completely determine a line $\mathcal{D}(\theta, \rho)$ of Eq. (24) it is enough to know $\theta$ and $\rho$ or essentially the point $Q(\rho \cos \theta, \rho \sin \theta)$. For a point $P(x_0, y_0)$ on Eq. (24), the circle of radius $R = \frac{1}{2} ||OP|| = \frac{1}{2}\sqrt{x_0^2 + y_0^2}$ centered in $P_c\left(\frac{x_0}{2}, \frac{y_0}{2}\right)$ passes through the origin $O$, $P$, and $Q$. In fact, for any line Eq. (24) passing through $P(x_0, y_0)$, its corresponding point $Q(\rho \cos \theta, \rho \sin \theta)$ is on this circle. We thus have the following correspondence:

$$P(x_0, y_0) \rightarrow \mathcal{C}(P_c; R)$$ \hspace{1cm} (25)

which associates with a point $P(x_0, y_0)$ the circle $\mathcal{C}(P_c; R)$ corresponding to all the lines passing through $P(x_0, y_0)$ (Fig. 21).

Conversely, for any point $Q(\rho \cos \theta, \rho \sin \theta)$ in the plane, let us consider all the circles passing through $Q$ and the origin $O$ (Fig. 22). Let the center of such a circle be $P_c(x_c, y_c)$. Then the point $P(2x_c, 2y_c)$ is on the line of Eq. (24) because $P_c(x_c, y_c)$ is on the line

$$x \cos \theta + y \sin \theta = \rho$$ \hspace{1cm} (26)

We thus have the following correspondence:

$$Q(\rho \cos \theta, \rho \sin \theta) \rightarrow \mathcal{D}(\theta, \rho)$$

which associates with a point $Q(\rho \cos \theta, \rho \sin \theta)$ the line $\mathcal{D}(\theta, \rho)$ corresponding to all the circles passing through $Q$ and $O$.

The two relations (25) and (26) are used to define the CT.

5.2. Properties of the CT

An image $I$ is described using the Cartesian coordinates of its points. We denote by $\mathcal{P}$ the parameter space or the set of traces of the circles. According to its definition, the CT has the following properties:

Property 1. A point in an image $I$ corresponds to a circle in the parameter space $\mathcal{P}$ (Fig. 21).

Property 2. A point in the parameter space $\mathcal{P}$ corresponds to a line in the image $I$ (Fig. 22).
Property 3. Points on a given circle passing through the origin in the parameter space $\mathcal{P}$ correspond to lines in the image $\mathcal{I}$ which intersect at the same point (Fig. 21).

Property 4. Points on a given line in the image $\mathcal{I}$ correspond to circles in the parameter spaces $\mathcal{P}$ which pass through the origin and intercept at the same point (Fig. 22).

5.3. Implementation

In our implementation, discrete steps are used for the observation and parameter spaces. The gray-level discrete image $\mathcal{I}$ is of dimension $N \times N$ pixels. We denote by $\mathcal{P}_i (i = 1, 2)$ the discrete parameter spaces and the two accumulators needed, which is of resolution $N_x \times N_y$ cells. We use $\mathcal{P}_i (i = 1, 2)$ to record the number of times that a cell is traversed by the circles traced by the CT. There are three stages in the implementation of our CT algorithm.

Stage 1. Pre-treatment of the image $\mathcal{I}$ (binary image).

Stage 2. Transformation of the edge point pixels of image $\mathcal{I}$ into circles in the parameter spaces $\mathcal{P}_i (i = 1, 2)$. With each edge point pixel $P(x_i, y_j)$ in the image $\mathcal{I}$ we associate two circles:

(a) one circle $\mathcal{C}_1$ in the parameter space $\mathcal{P}_1$ whose diameter is the segment joining $P(x_i, y_j)$ and $O_1(-x, -y)$, where $x = \frac{N}{4}$;

(b) a second circle $\mathcal{C}_2$ in the parameter space $\mathcal{P}_2$ whose diameter is the segment joining $P(x_i, y_j)$ and $O_2(N + x, -y)$, where $x = \frac{N}{4}$.

The choice of the points $O_1$ and $O_2$ is based on the principle that if a line passes through the origin of the transform it will not be detected, so one needs to make the transform with two circles. The fact that the two “origins” are outside the image and at a distance $\sqrt{2}x$ ensures that we will detect all the lines in our image. We trace the two circles in parameter spaces $\mathcal{P}_i (i = 1, 2)$ (Fig. 23), and we increment the vote of all cells corresponding to the points of each circle.

Stage 3. Detection of the points having the most votes in $\mathcal{P}_i$ $(i = 1, 2)$ which determines pixel alignments in the image $\mathcal{I}$ (a search for local maxima called peaks).

6. Comparison

In theory all straight lines passing by the same number of pixels must have the same votes, if not it is an artifact. We test robustness, for the four methods to this artifact, for a totally white image. Also we test robustness to this artifact for images with high number of parallel straight lines (which approximate the white image). Robustness to the noise for the four methods is tested in the end of this section.

In all the subsections below, the accumulator size for each of the four transforms is $512 \times 256$ pixels. For Sections 6.1, 6.3, and 6.4 the resolution of image is $256 \times 256$, and for Section 6.2 the resolution of image is $512 \times 512$.

6.1. Comparison of distribution

In this section we compare the four transforms from the point of view of the distributions of votes in the accumulators and their consequences. The distribution of the votes in an accumulator indicates what pixels or what portion of the accumulator is really used and how.

To perform the comparison, we consider the mapping of a totally white image. Our totally white image is a square image (see Fig. 24). The mappings of the totally white image by the transforms appear in Fig. 25 for the SHT, Fig. 26 for the RHT, Fig. 27 for the PAT, and finally in Fig. 28 for the CT.

A simple implementation for the accumulator with a fast access to the pixel uses an array of $m \times n$ pixels. For this kind of implementation, Table 1 indicates that SHT uses only 54% of the space in the accumulator, followed by the RHT which uses 75% of the space, compared to the CT and the PAT which use 90% and 100% of the space, respectively. Hence, PAT has the...
advantage of using all the pixels without wasting memory. We could think about an implementation with sparse array using only the pixels which have a non-zero value, but this kind of implementation is much less effective in term of access time and more complex to implement.

Table 2 contains indications on the distribution of votes among the pixels which are used in the accumulator. More precisely it contains the minimal (non-zero) and maximal number of votes in a non-zero pixels of the accumulator (and hence the range of votes), the number of pixels with the maximal number of votes, and the means and standard deviations of the numbers of votes in a pixel.

Low maximal number of votes in non-zero pixels and low mean and standard deviation indicate that the votes are well distributed among the pixels. If we consider only a low maximal number of vote, PAT and RHT are better than SHT and CT. However, for low mean and standard deviation, we observe that CT is better than PAT which is better than RHT and SHT. In conclusion for this set of criteria we can conclude that PAT is better than RHT which is better than SHT. But it is not clear how well is CT compared to the other methods. It will happen that it is almost as good as PAT.

The maximal number of votes can be considered as an indicator of the presence or absence of artifacts when we use the method in a real image. Indeed, a high maximal number of votes indicate that image is apt to produce artifact rapidly in the
Table 2

Distribution of votes on the space used in the accumulator

<table>
<thead>
<tr>
<th>Transform</th>
<th>Minimal number of votes</th>
<th>Maximal number of votes</th>
<th>Number of pixels with max votes</th>
<th>Mean votes</th>
<th>Standard deviation of votes</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHT</td>
<td>1</td>
<td>1019</td>
<td>2</td>
<td>567.104</td>
<td>258.841</td>
</tr>
<tr>
<td>RHT</td>
<td>1</td>
<td>512</td>
<td>32,640</td>
<td>340.889</td>
<td>170.975</td>
</tr>
<tr>
<td>PAT</td>
<td>1</td>
<td>512</td>
<td>2</td>
<td>256</td>
<td>109.136</td>
</tr>
<tr>
<td>CT</td>
<td>1</td>
<td>1068</td>
<td>4</td>
<td>226.221</td>
<td>101.635</td>
</tr>
</tbody>
</table>

Fig. 29. Histogram of votes for the SHT.

Fig. 30. Histogram of votes for the RHT.

Fig. 31. Histogram of votes for the PAT.

Fig. 32. Histogram of votes for the CT.

For the SHT, the false peaks correspond to diagonal lines (see Fig. 34). The false peak for the RHT corresponds to horizontal and vertical lines which belong to the zone indicated in Fig. 35, and with the given resolution, gives a white image (see Fig. 36). For the PAT, the false peak corresponds to diagonal lines as indicated in Fig. 38. Finally, the false peak for the CT corresponds to horizontal and vertical lines as indicated in Fig. 40.

The histograms of the four methods are given in Figs. 29–32.

The presumed artifact for a method depends on the repartition of the votes in the accumulator, which depends on the voting method in itself and also on the discretization process which causes splitting and spreading [21].

6.2. Examples with vertical parallel lines

In this section we present examples to compare the four methods for parallel lines. As we increase the number of straight line the image like the white image and the possibility to get wrong votes or false peaks increases. We will detect sum artifacts associated with the methods.
For the image of 30 vertical lines in Fig. 41, the four methods (the SHT in Fig. 42, the RHT in Fig. 43, the PAT in Fig. 44, and the CT in Fig. 45), give similar results. But for the image of 50 vertical lines in Fig. 46, the SHT presents false peaks in Fig. 47 whereas the RHT in Fig. 48, the PAT in Fig. 49 and the CT in Fig. 50 do not. For the image of 80 vertical lines in Fig. 51, the RHT presents false peaks in Fig. 52, but the PAT in Fig. 53 and the CT in Fig. 54 do not show false peaks. For the image of 110 vertical lines in Fig. 55, both PAT and CT present false peaks, respectively, in Figs. 56 and 57.

When we increase the number of parallel lines, the image behaves like the white image. So the artifact detected by the
Fig. 39. Points having maximum votes in the CT accumulator. The points have been enlarged for better visibility.

Fig. 40. Lines having maximum votes, for the CT.

Fig. 41. Image containing 30 vertical lines.

Fig. 42. Lines detected by the SHT.

Fig. 43. Lines detected by the RHT.

Fig. 44. Lines detected by the PAT.
Fig. 45. Lines detected by the CT.

Fig. 46. Image containing 50 vertical lines.

Fig. 47. Lines detected by the SHT.

Fig. 48. Lines detected by the RHT.

Fig. 49. Lines detected by the PAT.

Fig. 50. Lines detected by the CT.
Fig. 51. Image containing 80 vertical lines.

Fig. 52. Lines detected by the RHT.

Fig. 53. Lines detected by the PAT.

Fig. 54. Lines detected by the CT.

Fig. 55. Image containing 110 vertical lines.

Fig. 56. Lines detected by the PAT.
method is apt to occur and is the first line what happen here for SHT in Fig. 47, RHT in Fig. 52, PAT in Fig. 56 and CT in Fig. 57.

6.3. Example with noise

In this section we examine the robustness of the four methods to noise. We consider a synthetic image containing 11 straight-line segments and we add a Gaussian noise with parameters \((\mu = 0, \sigma)\). To obtain a binary image we use a threshold \(\tau = 12\). To detect the straight lines we use the algorithm described in Fig. 58. At each iteration we transform our image with (SHT, RHT, PAT, or CT) and detect the global maximum (peak) in the accumulator (in parameter space), we save it and we delete all points in the image (in image space) contributing to this maximum (this is the identify and remove idea proposed by Fiala [22]) and we set a reduced image. We repeat the process with the reduced image until \(n\) straight lines have been detected. For a given \(n\) we have exactly \(n\) straight lines detected by each method.

Note that the detection of peak can be enhanced by using filtering [23]. Here we used the standard butterfly filter [23, 24] for SHT, RHT, and PAT. No filtering process have been used with CT because no filter is adapted for this method.

Table 3 contains the result of this experiment (Figs. 59–93). It indicates the true lines (T) and false lines (F) detected by each method. We can conclude that PAT is slightly more robust to noise than RHT, which itself is more robust than SHT. CT is more sensitive for noise. The bad performance of CT can be caused by the absence of the filtering process.

6.4. A real image

In this section we apply the four methods to the real image in Fig. 94 in order to detect straight lines. We use the identify and remove algorithm described in the previous section. We present results for the identification of \(n = 11\) lines.

Fig. 95 is the edge-detected image of the real image of Fig. 94. The lines detected by the four transforms for \(n = 11\) are given in Figs. 96–99. In this example all the lines detected by the SHT, RHT, and PAT are significant except the two verticals detected by CT which do not represent any straight line in our edge image. These lines correspond to artifacts associated with the method; they are false peaks generated by the image. It is not surprising because CT is the most sensitive method to noise. Also we have two lines detected by SHT but not detected by the three other methods. These two lines pass at least by 20 pixels of less than the horizontal line in bottom of the edge image. We still attend the artifact cited in introduction of Section 6.
Fig. 59. $\sigma = 5$. 

Fig. 60. SHT. 

Fig. 61. RHT. 

Fig. 62. PAT. 

Fig. 63. CT. 

Fig. 64. $\sigma = 9$. 

Fig. 65. SHT. 

Fig. 66. RHT. 

Fig. 67. PAT. 

Fig. 68. CT.
Fig. 69. $\sigma = 10$.  

Fig. 70. SHT.  

Fig. 71. RHT.  

Fig. 72. PAT.  

Fig. 73. CT.  

Fig. 74. $\sigma = 11$.  

Fig. 75. SHT.  

Fig. 76. RHT.  

Fig. 77. PAT.  

Fig. 78. CT.
Fig. 79. $\sigma = 12$.

Fig. 80. SHT.

Fig. 81. RHT.

Fig. 82. PAT.

Fig. 83. CT.

Fig. 84. $\sigma = 14$.

Fig. 85. SHT.

Fig. 86. RHT.

Fig. 87. PAT.

Fig. 88. CT.
Fig. 89. $\sigma = 15$.

Fig. 90. SHT.

Fig. 91. RHT.

Fig. 92. PAT.

Fig. 93. CT.

Fig. 94. A real image.

Fig. 95. Edge image for the real image.

Fig. 96. 11 lines SHT.

Fig. 97. 11 lines RHT.

Fig. 98. 11 lines PAT.
Acknowledgments

7. Conclusion

Based on an extension of the point-line duality used by the (standard) Hough transform to detect straight lines in an image, we have analyzed three methods. Which methods are faster than the SHT because the PAT, RHT and CT use additions and multiplications whereas the SHT uses trigonometric functions (sine and cosine) for calculation. We compared the four methods for accumulator distribution of votes and conclude that the distribution of the votes is more accurate for CT and PAT than for SHT and RHT. Also we tested the robustness of these methods to the noise and it proved that the RHT and PAT are more robust to noise than the SHT and CT. Finally, we have presented the results of these methods on a real image for which we have better results with RHT and PAT than with SHT and CT.

All the four methods have a weakness compared to the artifact listed in Section 6. In future we have to improve these methods to design perfectly robust method to this artifact.

Acknowledgments

This work has been financially supported by NSERC (Natural Sciences and Engineering Research Council of Canada) via individual discovery grants to the last two authors.

References


About the Author—SAID EL MEJDANI received his B.Sc. degree in Mathematics in 1992 from University Ibn Toufail (Morocco) and his M.Sc. degree in Computer Science at University of Sherbrooke (Québec, Canada). He is currently a Ph.D. candidate in Computer Science at University of Sherbrooke.

About the Author—RICHARD EGLI is associate professor in the Department of Computer Sciences at University of Sherbrooke since 2000. He received his B.Sc. degree in Computer Science and his M.Sc. degree in Computer Sciences at University of Sherbrooke (Québec, Canada). He receives his Ph.D. in Computer Science from University of Montréal (Québec, Canada) in 2000. His research interests include computer graphics and digital image processing.

About the Author—FRANCOIS DUBEAU is professor in the Department of Mathematics at University of Sherbrooke since 1992. He taught at the Collège Militaire Royal de St-Jean (Québec, Canada) from 1982 to 1992. He received his Ph.D. degree in Mathematics from University of Montréal (Québec, Canada) in 1981, his M.Sc.A. degree in Industrial Engineering in 1973 and his B.Sc.A. degree in Engineering Physics in 1971 from École Polytechnique de Montréal (Québec, Canada). His research interests include applied mathematics, operational research, numerical analysis, and digital image processing.