14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

# MOTION ESTIMATION BY THE PSEUDO-WIGNER VILLE DISTRIBUTION AND THE HOUGH TRANSFORM

*Noemí Carranza, Filip Sroubek and Gabriel Cristóbal*

Instituto de Óptica (CSIC), Serrano 121, 28006 Madrid (SPAIN)
phone: + (34) 915616800, fax: + (34) 915645557, email: {noemi, filip, gabriel}@optica.csic.es,
http://www.iv.optica.csic.es

**ABSTRACT**

*A fundamental problem in processing sequences of images is the computation of the optical flow, an approximation to the real image motion. This paper presents a new algorithm for optical flow estimation based on a spatiotemporal-frequency approach, specifically on the computation of the Wigner-Ville distribution and the Hough Transform. Experimental results are compared with an implementation of the variational technique for local and global motion estimation, where it is shown that the results are accurate and robust to noise degradations.*

## 1. INTRODUCTION

Motion estimation is a key problem for the analysis of image sequences. By means of motion estimation, 3D scene properties and motion parameters can be obtained from a dynamic scene. It is also useful for performing motion segmentation; for computing the focus of expansion and time-to-collision; for performing motion-compensation image encoding; stereo disparity; for measuring blood flow and heart-wall motion in medical imagery, etc [1].

From the information available from a sequence of images, it is only possible to derive an estimate of the motion field, which is called optical flow. Although optical flow is generally not equivalent to the true motion field, it is quite similar in most of the cases.

Numerous theoretical and practical studies on the optical flow estimation from image sequences and on the useful information it contains have been performed. These methods can be classified into three main categories:

- Differential techniques: these approaches compute image velocity from spatiotemporal derivatives of image intensities [1].
- Correlation techniques: these approaches define the displacement as an approximation to velocity as a shift that yields the best fit between contiguous time-varying image regions [1].
- Frequency-based methods: these techniques are based on the use of frequency and phase information by means of tools like the Fourier transform to estimate the optical flow [2].

Each technique has several advantages and disadvantages. Numerical differentiation is sometimes impractical because of small temporal support or poor signal-to-noise ratio. In these cases, it is natural to consider correlation techniques. Among the advantages brought by the frequency-based methods, it is found that motion-sensitive mechanisms operating on spatiotemporally oriented energy in Fourier space can estimate motion in image signals for which matching approaches would fail [1]. One example can be the motion of random dot patterns.

The differential method has a major drawback in the estimation of the first and second derivatives of the pixel intensity, mainly in the case of noisy images. To improve noise robustness, a common strategy is to use regularization based on variational integrals. Frequency-based methods are in general more robust to noise [3].

Among the different techniques for computing optical flow using frequency-based methods, the spatiotemporal-frequency approach (STF) has been proposed, which gives a simultaneous representation of a signal in space and spatial frequency variables [4]. One implementation of the STF approach employs the Wigner-Ville distribution (WVD) as an underlying STF image representation.

In this paper, an algorithm for computation the optical flow based on the computation of the Wigner-Ville distribution and the Hough Transform is presented and compared with an implementation based on variational techniques. The paper is organized as follows. Section 2 briefly reviews the problem of computing optical flow and explains both methods. Section 3 provides an explanation of the new algorithm proposed. Section 4 evaluates both methods and Section 5 concludes the paper.

## 2. A BRIEF REVIEW OF MOTION ESTIMATION TECHNIQUES

### 2.1 Differential Techniques

The initial hypothesis for measuring image motion is that the intensity structures of local time-varying image regions are approximately constant under motion for at least a short duration [5]. Let $i(\bar{x},t)$ denote the intensity function, where $\bar{x}=(x,y)$ represents the pixel position and $t$ is the time. If the intensity remains constant, then

$$i(\bar{x},t) = i(\bar{x}+\delta\bar{x}, t+\delta t) \tag{1}$$

where $\delta\bar{x}$ is the displacement of the local image region at $(\bar{x},t)$ after time $\delta t$. Expanding the left-hand side of this equation in a Taylor series yields

$$\nabla i \cdot \bar{v} + i_t = 0 \tag{2}$$

where $\nabla i = (i_x, i_y)$ is the spatial intensity gradient, $i_t$ is the derivate of the intensity with respect to time, and $\bar{v} = (v_x, v_y)$ is the image velocity.

This equation is called the optical flow constraint (OFCE). Unfortunately, one scalar equation is not enough for finding both components of the velocity field. It gives only the component in the direction of the gradient, that is, the normal flow. This problem is usually called the aperture problem [6].

### 2.2 Variational techniques

Several ideas have been proposed to overcome the aperture problem, for instance, second-order derivative constraints, weighted least squares approach, parametric models of velocity or regularization of the velocity field [6]. The later approach is the one we have implemented and evaluated here because actually it is considered one of the best techniques proposed in the literature [7].

The main idea is to formulate a minimization problem of the form

$$\inf_{v}\left(A(\bar{v}) + S(\bar{v})\right) \tag{3}$$

where $A(\bar{v}) = \int_{\Omega} (\bar{v} \cdot \nabla i + i_t)^2 dx$ is called the *fidelity* term and $S(\bar{v}) = \alpha \sum_{j=1}^{2} \int_{\Omega} \left|\nabla \bar{v}_j\right|^2 dx$ represents the *smoothing* term, for every $\bar{v}(x, y) \in \Omega$. Nevertheless, this smoothing term disregard the detection of discontinuities, which is a very important cue for sequence analysis. In order to preserve discontinuities, different approaches have been considered for computing discontinuous optical flow fields by modifying the smoothing term $S(\bar{v})$. The regularization term changes into

$$\sum_{j=1}^{2} \int_{\Omega} \Phi\left(\left|\nabla \bar{v}_j\right|\right) dx \tag{4}$$

where the function $\Phi$ is at most linear and allows noise removal and edge preservation. This regularization term has been extended also in the temporal dimension in order to regularize also the temporal information between frames. For $\Phi$, we have used the total variation, which is in this case $\Phi(s) = s$.

### 2.3 Frequency Based Methods

The motion estimation in the frequency domain is based on one fundamental property (Fourier shift theorem), which can be derived by analyzing a video sequence through a three dimensional Fourier transform. We assume again that we can represent an image sequence through a function $i_0(x, y)$ such as

$$i(x, y, t) = i_0(x - v_x t, y - v_y t) \tag{5}$$

where the main assumption here is that moving objects must move with a uniform velocity vector ($v_x$, $v_y$) and must have a constant illumination. Now, by calculating the spatial and temporal Fourier transform of the sequence $i(x, y, t)$, we obtain

$$I(f_x, f_y, f_t) = I_0(f_x, f_y)\delta(v_x f_x + v_y f_y + f_t) \tag{6}$$

where $I_0$ represents spatial Fourier transform of $i_0$ and $\delta$ is the Dirac delta function.

Thus, $I(f_x, f_y, f_t)$ is nonzero only in a plane, which is called *motion plane* [8]. This plane passes through the frequency origin. Its equation is given by

$$v_x f_x + v_y f_y + f_t = 0 \tag{7}$$

Estimating parameters of this plane leads to estimate velocity vector components ($v_x$, $v_y$) of the moving object. This equation (7) can also be viewed as the Fourier transform of the OFCE

$$\frac{\partial i(x, y, t)}{\partial t} + v_x \frac{\partial i(x, y, t)}{\partial x} + v_y \frac{\partial i(x, y, t)}{\partial y} = 0 \tag{8}$$

## 3. A NEW ALGORITHM FOR OPTICAL FLOW ESTIMATION BY WVD AND HT

### 3.1 The WVD of an image sequence

The major motivation for considering the use of STF image representation approach as a basis for computing optical flow comes from the literature on mammalian vision. In particular, some investigations have demonstrated that many neurons in various cortical areas of the brain behave as spatiotemporal-frequency bandpass filters [9]. In the field of non-stationary signal analysis, the WVD has been used for the representation of speech and image. Jacobson and Wechsler [4,10] were the first to suggest the use of the WVD for the optical flow estimation.

The Wigner Distribution was introduced by Wigner as a phase space representation in Quantum Mechanics, and it gives a simultaneous representation of a signal in space and spatial frequency variables [11]. Later, in the area of signal processing, Ville derived the same distribution that Wigner proposed several years before [12]. The WVD can be considered as a particular occurrence of a complex spectrogram in which the shifting window function is the function itself [13].

The WVD distribution of a moving sequence is a 6-dimensional function defined as

$$\begin{aligned} &W_i(x, y, t, w_x, w_y, w_t) = \\ &\iiint R_i(x, y, t, \alpha, \beta, \tau) e^{-j(\alpha w_x + \beta w_y + \tau w_t)} d\alpha d\beta d\tau \end{aligned} \tag{9}$$

where

$$\begin{aligned} &R_i(x, y, t, \alpha, \beta, \tau) = \\ &i(x + \alpha, y + \beta, t + \tau) i^*(x - \alpha, y - \beta, t - \tau) \end{aligned} \tag{10}$$

and where * denotes complex conjugation.

Again, for the case where a time-varying image $i(x, y, t)$ is uniformly translating at some constant velocity, the WVD of this image is

$$W_i(x, y, t, w_x, w_y, w_t) = \delta(v_x w_x + v_y w_y + w_t) W_i(x - v_x t, y - v_y t, w_x, w_y) \quad (11)$$

From (11), the WVD of a linearly translating image with velocity ($v_x$,$v_y$) is everywhere zero except in the plane defined by

$$\{(x, y, t, w_x, w_y, w_t) : v_x w_x + v_y w_y + w_t = 0\} \quad (12)$$

Equivalently, for an arbitrary pixel at $x, y, t$, each local STF spectrum of the WVD is zero everywhere except on the plane defined in (12). For this reason, if a procedure for estimating the velocity associated with a given STF spectrum is found, we will obtain a space and time varying optical flow function. In the next sections we will describe how the Hough Transform can be used for computing the optical flow.

### 3.2 Hough Transform

A recurring problem in digital image processing is the detection of straight lines in digitized images. The Hough Transform (HT) and least squares (LS) algorithm are two well known pose estimation methods. The LS method is based on the minimization of an objective function: the sum of squared distances between image features and the model. The HT detects an object and estimates the pose parameters by computing the largest subset of image features fitting a rigid template. Its strong points are the ability to discard features belonging to other objects and the robustness against incomplete data and noise [14]. These characteristics are very important in our problem due to the fact of the presence of cross-terms of the Wigner-Ville distribution, so we can use a more efficient method for determining the motion plane.

The parameterization specifies a straight line by the angle θ of its normal and its perpendicular distance ρ from the origin. If we restrict θ to the interval [0, π], then the normal parameters for a line are unique. With this restriction, every line in the *x-y* plane corresponds to a unique point in the θ-ρ plane [15].

### 3.3 Algorithm proposed

As seen in the previous section, a line can be completely characterized in the Hough plane, as well as a line and also a plane. This feature has been used to determine the velocity by means of the HT applied to the STF spectrum.

Our first implementation was based on the use of the HT on the whole spectrum in order to find the plane. This way, each pixel of the spectrum with a nonzero value was represented in the Hough plane. However, those pixels of the WVD belonging to cross-terms influenced on the final result, and sometimes can produce incorrect solutions. For this reason, we have used another approach based on the HT computation for each of the frames of the spectrum in order to detect a line on each of them and in this way discarding the information from cross-terms pixels. Furthermore, this implementation is computationally less demanding.

At the end, our problem can be reduced to find one straight line in each temporal frame. Furthermore, we already know that the spectrum contains the frequency origin, and we can observe that the lines in different temporal frames are parallel. Taking into account these facts and by means of applying the HT the plane will be detected.

We will illustrate our algorithm with an example. Given a sequence composed of a circular object moving with a uniform horizontal velocity, performing the WVD we will find a plane which is represented in each temporal-frequency frame by a line (see Fig. 1).

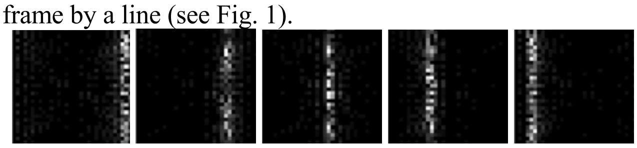

Figure 1 – Several frames of the WVD of the sequence

With our method, we perform the HT for each of the frame of the spectrum. The result of the HT applied to one frame is shown in Fig. 2, where the position of the line can be estimated through the angle and distance to the centre of the plane.

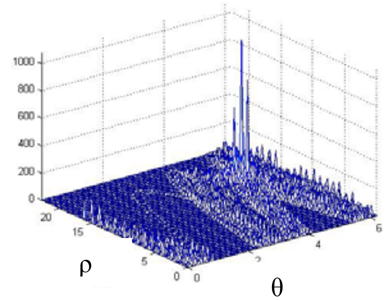


Figure 2 – HT of one frame of the WVD

Taking the maximum value of this HT, the information of the position for each line is provided. As every straight line found in one frame is parallel to the others, a peak for each line can be found in the Hough plane, which will have the same angle θ, and increasing ρ. This fact can be seen in Fig. 3, where we have summed up all the maximum peaks for all the frames.

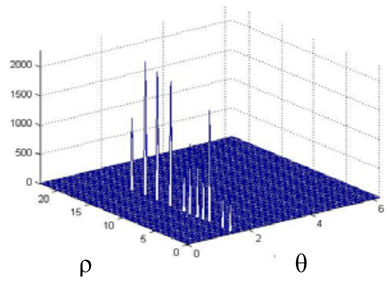


Figure 3 – HT of all the frames of the WVD

Actually, the information provided by the angle of one of the peaks would be enough for estimating the direction of the velocity. But most of the times, ideal conditions are not met (we must remember, for example, the presence of cross-terms induced by the WVD), and in some cases, by considering one of the frames we can end up with an erroneous solution, due to noise or other external factors. We propose to use the redundant information of all the frames and the property shown in the Fig. 3, which is that all the peaks form a straight line and by applying the HT to the summation of all the peaks we are able to eliminate the information of the erroneous peaks.

In order to estimate the magnitude of the velocity, the values of the different $\boldsymbol{\rho}$ obtained have been used(i.e. the distance from the lines to the origin), so as to estimate the slope of the plane.

As seen in the Fig. 4, the values of $\boldsymbol{\rho}$ should increase uniformly. Therefore, the HT can be used to evaluate the slope of this line and also for eliminating outliers.

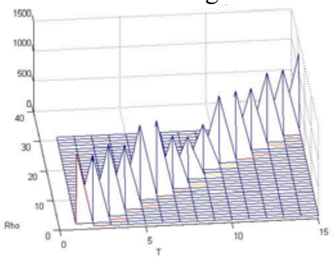


Figure 4 – Values of ρ of the HT for the different frames of the WVD

## 4. RESULTS

The new methodology proposed was applied to synthetic images of a moving circular object with constant intensity, which can be described as:

$$\begin{cases} x(t) = \rho \cdot \cos(\theta) + x_0 + t \cdot v_x \\ y(t) = \rho \cdot \sin(\theta) + y_0 + t \cdot v_y \end{cases} \tag{15}$$

where $\rho$ represents the radius of the circle with a constant gray level, θ varies from 0 to 2π, $[x_0, y_0]$ is the initial point, $t$ corresponds to the variation of the position with the time and finally, $v_x$ and $v_y$ are the velocity components of the circle. Fig. 5 shows three images of the sequence, for $v_x$=1 and $v_y$=-1.

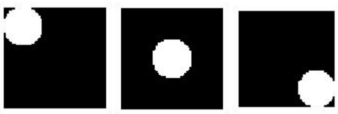

Figure 5 – Images of the synthetic sequence for $v_x = 1$ and $v_y = -1$

In order to estimate the global motion, we have obtained a smoothed 3D frequency spectrum by means of a Hanning filter which has been previously introduced [3].

The method was applied to distinct values of ρ, $[x_0, y_0]$ and $[v_x, v_y]$. Some results are shown on Table 1.

| *Actual translation* | | *Estimated translation* | |
|---|---|---|---|
| $V_x$ | $V_y$ | $v_x$ | $v_y$ |
| 1 | 0 | 1.0000 | -9.4941e-016 |
| 1 | -1 | 0.9717 | -0.9748 |
| -0.7 | 0.5 | -0.6773 | 0.5203 |
| 0.7 | -1 | 0.6935 | -0.9434 |
| -1.2 | -0.7 | -1.1973 | -0.6790 |

Table 1 – Translations in pixels/frame for several examples.

For these simple sequences, when we consider a moving object with a uniform velocity and we calculate a global motion, an accurate information about the optical flow can be obtained by means of the method based on WVD-HT.

A further step on the analysis has been done, estimating the motion locally. This implementation should be used when the *a priori* information of assuming only one object in the sequence is unknown. Thus, a small window will be assigned to each pixel of the sequence, and the algorithm will be executed for each of the windows.

The main problem of this implementation will be choosing the optimum window size. If the window size is too small, the apperture problem will occur in some of the windows. On the other hand, a bigger size of window could lead to join several objects moving with different directions in the same window. In order to avoid this problem, a hierarchical implementation has been developed. The best estimation from different sizes of window will be chosen depending on the maximum value of the HT. An example with two hierarchical levels is shown in Fig. 8-a, where the image size is 128x128x25 pixels, and window sizes are 25 and 50 pixels. For these sizes of window, we can observe that the optical flow estimated in the regions near to the border of the circle is very accurate (compare with the results obtained with the variational method in Fig. 6-b). Only regions inside the circle provide uncertainty due to the apperture problem, and therefore values for optical flow are less accurate.

We have applied the variational method to the same simple sequence used before. We can observe the results for an arbitrary frame obtained for $v_x$=1 and $v_y$=0 in the Fig. 6-b, where it is shown the optical flow obtained after regularization. We can see that the values of optical flow are about 1 pixel per frame, but the optical flow doesn't always fit the right positions. The uncertainty of the optical flow obtained in regions inside the circle, which is due to the apperture problem, has been partially corrected by means of the regularization.

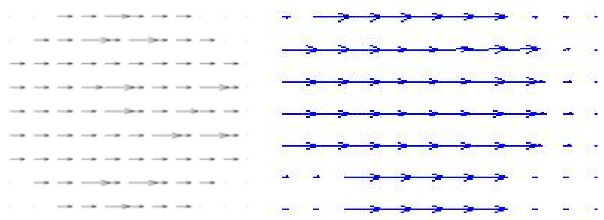

Figure 6 – a) Optical flow with local estimation using the method based on WVD-HT; b) Optical flow obtained using variational methods.

Differential methods provide an optical flow for each pixel of the sequence, so it can be a tough task to perform a direct comparison with the global motion estimation method based on WVD-HT, but we can make a comparison through the local estimation. Fig. 6-a and 6-b show that results obtained with our method are comparable with the variational technique and the motion field obtained with our method fits better the right positions.

Finally, to test the robustness against noise of the new algorithm proposed, several experiments have been conducted adding different types of noise to the sequence. Table 2 shows results of estimated translation for our sequence with added noise, when the actual translation is $V_x$ = 1 pixel/frame and $V_y$ = 0 pixel/frame. The Gaussian and Speckle added noise are characterized by its variance and the Salt & Pepper noise by its density.

| | *Gaussian* | | *Speckle* | | *Salt&Pepper* | |
|---|---|---|---|---|---|---|
| Variance/ Density | $V_X$ | $V_y$ | $V_X$ | $V_y$ | $V_X$ | $V_y$ |
| 0.05 | 1 | 0 | 1 | 0 | 1 | 0 |
| 0.1 | 1 | 0 | 1 | 0 | 1 | 0 |
| 0.2 | 1 | 0 | 1 | 0 | 1 | 0 |
| 0.4 | 1 | 0 | 1 | 0 | 1 | 0 |

Table 2 – Estimated translations provided by WVD-HT method in pixels/frame for actual translation $V_x$=1 and $V_y$=0 for different additive noise variances.

In table 3, results of estimated translation are shown when the actual translation is $V_x$ = 1 and $V_y$ = -1 pixel/frame respectively.

| | *Gaussian* | | *Speckle* | | *Salt&Pepper* | |
|---|---|---|---|---|---|---|
| Variance/ Density | $V_X$ | $V_y$ | $V_X$ | $V_y$ | $V_X$ | $V_y$ |
| 0.05 | 0.97 | -0.97 | 0.97 | -0.97 | 0.96 | -0.98 |
| 0.1 | 0.98 | -0.95 | 0.97 | -0.97 | 0.96 | -0.98 |
| 0.2 | 0.95 | -0.95 | 0.98 | -0.96 | 0.98 | -0.97 |
| 0.4 | 1 | 0 | 0.97 | -0.98 | 0 | -1 |

Table 3 – Estimated translations provided by WVD-HT method in pixels/frame for actual translation $V_x$=1 and $V_y$=-1 for different additive noise variances.

In the case of the variational method, the added noise of each pixel has influence on the estimation of first and second derivatives of the pixel intensity. The optical flow obtained, obviously, gets worse as the noise increases.

## 5. CONCLUSIONS

In this paper, we have presented a frequency-based method for motion estimation based on the computation of the Wigner-Ville distribution together with the Hough Transform. Results have been shown, evaluated and compared with an implementation based on the variational method.

The estimation of global motion gives values of optical flow very close to the actual ones. In this case, we can say that the solution is more accurate and robust to noise than the one obtained with the variational method. Nevertheless, a more fair comparison between both methods has been performed with a local estimation of motion. Selecting an appropriate size of window and using a hierarchical approach, the proposed method gives similar results to the variational approach.

The advantages brought by the WVD-HT method are the robustness to noise and the accuracy in global estimation of motion. In both methods, WVD-HT's and variational's, one of the major drawbacks is the selection of the optimum parameters for the algorithms.

In the future, we are looking forward to applying the described method to other image processing problems, such as cardiac magnetic resonance motion estimation.

## 6. ACKNOWLEDGEMENTS

Financial support of this research was provided by the Grant Agency of the Czech Republic under the projects No. 102/04/0155, No. 202/05/0242 and by the following spanish grants: TEC2004-00834; TEC2005-24739-E; TEC2005-24046-E; 20045OE184 and PI040765. F. Sroubek was also supported by the Spanish States Secretary of Education and Universities fellowship.

## REFERENCES

[1] S.S. Beauchemin and J. L. Barron, "The Computation of Optical Flow", *ACM Computing Surveys*, vol. 27, No.3, pp. 433-467, September 1995.

[2] M.A. Arredondo, K. Lebart and D. Lane, "Optical flow using textures", *Pattern Recognition Letters*, vol. 25, pp. 449-457, 2004.

[3] W.I. Meyering, M.A. Gutierrez, S.S. Furuie, M.S. Rebelo, C.P. Melo, "Wigner-Ville Distribution applied to Cardiac Motion Estimation", *Computers in Cardiology 2000*, 2000, pp. 619-622.

[4] L. Jacobson and H. Wechsler, "Derivation of optical flow using a spatiotemporal-frequency approach", *Computer Vision, Graphics and Image Processing*, vol. 38, pp. 29-65, 1987.

[5] B. K. P. Horn and B. G. Schunck, "Determining Optical Flow", *Artificial Intell.* 17, 1981, pp. 185-204.

[6] G. Aubert, P. Kornprobst, *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, Springer-Verlag New York, LLC, 2001.

[7] G. Aubert, R. Deriche and P. Kornprobst, "Computing Optical Flow Via Variational Techniques", *SIAM Joutnal on Applied Mathematics*, vol. 60, N. 1, pp. 156-182, 1999.

[8] M. Pingault, D. Pellerin, "Motion estimation of transparent objects in the frequency domain", *Signal Processing*, vol. 84, pp. 709 – 719, April 2004.

[9] P. A. Laplante and A. D. Stoyenko, *Real-time imaging: theory, tehcniques and applications*, IEEE Press, 1996.

[10] L. Jacobson and H. Wechsler, "Joint spatial/spatial-frequency representation", *Signal Processing*, vol. 14, pp. 37-68, 1998.

[11] E. Wigner, "On the Quantum Correction for Thermodynamic Equilibrium", *Physical Review*, vol. 40, pp. 749-759, 1932.

[12] E. Ville, "Théorie et Applications de la Notion de Signal Analytique", *Cables et Transmision*, 2A, pp. 61-74, 1948.

[13] J. Hormigo and G. Cristóbal, "Image Segmentation Using the Wigner-Ville Distribution", *Advances in imaging and electron physics*, P.W. Hawkes (ed.), Academic Press, vol. 131, pp 65-80, 2003.

[14] J. S. Marques, "A probabilistic framework for the Hough Transform and Least Squares Pose Estimation", *EUSIPCO 1998*, Island of Rhodes, Greece, September 8-11. 1998.

[15] 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, No. 1, Jan. 1972.