USING AN EXACT PERFORMANCE OF HOUGH TRANSFORM FOR IMAGE TEXT SEGMENTATION

Dimo T. Dimov

Institute of Information Technologies, Bulgarian Academy of Sciences
E-mail: dtdim@iinf.bas.bg

ABSTRACT

The (r,q)-interpretation of Hough transform (HT) is treated herein considering its facilities to localize straight lines or cut-offs in a given image. The definition of "*exact*" HT is introduced for the image and for the given choice of the respective HT-domain size (r_{size}*q_{size}). The idea of "*HT inner noise*" is also defined using a mean square estimate for the HT-deviations from the exact HT. Hence, a constructive method is proposed for a set of algorithms realising digital HT of an arbitrary small inner noise. The experiment of algorithms and a comparative analysis of results are briefly represented. The proposed method and algorithms are applied for pre-processing of hand-written text images. Besides the localisation of text rows and words, the approach proposed is especially effective for evaluation of the averaged letter slope of handwriting in the case of critical resolution of 2÷3 pixels per line and/or per distance among lines.

1. INTRODUCTION

The recognition process in computer analysis of documents is usually simplified by preliminary image decomposition to simple domains of minimal area simultaneously being mostly filled with graphics. These domains are the true objects for traditional recognition. In text recognition, similar domains are intuitively defined for the text rows, isolated words and/or phrases. These domains usually possess a "longitudinal stretch" and therefore can be effectively localised by "projective" techniques of HT-type. This work was provoked by unexpected troubles when using HT for preliminary processing of hand-written text images.

2. SETTING OF THE PROBLEM

Hough [1] has introduced the transform using the fact of analytic geometry that every line defined in the 2/D Euclidean space $E^2(x,y)$ by equation $L(x,y,a,b) = 0$ of two independent parameters (a and b), can be represented as a point (a,b) in the $E^2(a,b)$ dual space of parameters, and that this representation is an isomorphism. Instead of this canonical linear form, Duda and Hart [3] have proposed the usage of the *normal equation* form, where the parameters are respectively r, which is the distance from the origin Oxy and q – the normal slope angle to the abscissa Ox. Their approach leads to the both significant advantages:

- The values of the parameters r and q are limited that is a substantial plus in computer modelling.
- HT can be also considered as parallel projection by analogy with the Radon transform, which is well known from computer tomography [4].

The present report deals with the (r,q)-interpretation of HT used for textual image pre-processing [5]. Information for other interpretations, generalisations and applications of the classical HT-technique can be found, for instance in the never out-of-date survey of Illingworth and Kittler [2].

The idea is well known. If we have a sufficient number of projections for the textual image along several directions, preferably through equal intervals Dq of the slope angle q_s to the abscissa Ox, $q_s = -p/2 + sDq$, $s = 0,1,\ldots\lceil p/Dq \rceil$, then the projection that is perpendicular to the text row axis could accumulate the "most-outstanding" maxima corresponding to these rows and the minima corresponding to the inter-row spaces. On the other hand, one (r,q)-HT applied to the total image could give us in principle all the projections $h = h_q(r) = h(r,q_s)$, $r \in [r_{min}, r_{max}]$, $q_s \in [-p/2, p/2]$. So, it is possible to find the projection of the "most outstanding" extrema using classical techniques of functional analysis.

A substantial facilitation here is that the configuration of considered extrema is arranged along the line $q = a - p/2$, i.e. in parallel to the Or axis of the HT space, where α is the average row slant demanded. Hence, the search of extrema for the bivariate HT accumulation function $h = h(r,q)$ can be replaced with the search of extrema for a unidimensional function $x = x(q)$ being defined as a mean square estimate for the behaviour of h. For example, we have the estimate:

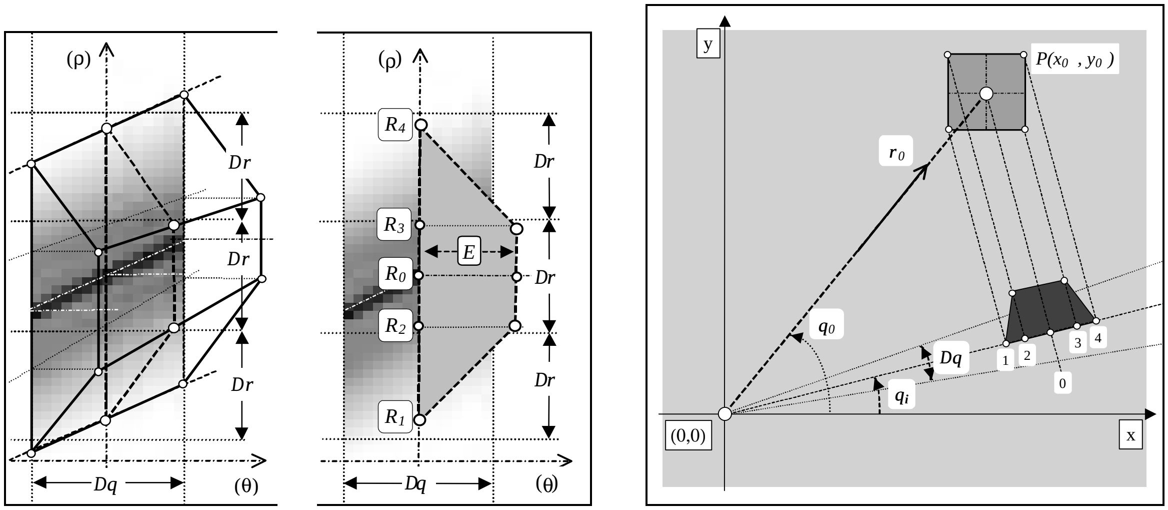
$$x^2(q) = \mathop{\mathrm{E}}_{r}\left[\left(h_q(r) - \mathop{\mathrm{E}}_{r}\left[h_q(r)\right]\right)^2\right], q \in [-p/2, p/2] \qquad (1)$$

that is the current variance for the sections $h = h_q(r) = h(r,q)$, $r \in [r_{min}, r_{max}]$ with q as a parameter, or another estimate:

$$x^2(q) = \mathop{\mathrm{E}}_{r}\left[\mathrm{grad}^2\left(h_q(r)\right)\right] = \mathop{\mathrm{E}}_{r}\left[\left(¶h_q(r)/¶r\right)^2\right] \qquad (1a)$$

where E[h(.)] is the mean value of h(.) in both the cases.

A similar strategy can be also used to evaluate the average font slope of text rows, words and/or phrases.



(a) **(b)** **(c)**

Fig.1. Geometric interpretation of the method in the strip $[q_i - Dq/2,\ q_i + Dq/2)$**:**

a) Trapezium-like cosine-shaped hexahedron $\mathrm{TC6}(x_0, y_0, q_i)$, $q_i = q_{\min} \div (Dq) q_{\max}$ **;**

b) Symmetrical trapezium $\mathrm{ST}(x_0, y_0, q)$, $q \in [q_i - Dq/2,\ q_i + Dq/2)$**;**

c) The accumulated projection for the pixel $P(x_0, y_0)$ **as a symmetrical trapezium** $\mathrm{ST}(x_0, y_0, q)$**.**

We shall intuitively consider as ***basic algorithm*** the (r,q)-HT of discrete performance that is usually referenced for the above-mentioned purposes [2, 4]:

⟨s0⟩ Form the accumulator $H(r,q)$ with zero elements.

⟨s1⟩ **For** each pixel $P(x,y)$ of the image $F(x,y)$ **do:**

if $F(x,y) \neq 0$ **then**

{ $r_0 = (x^2 + y^2)^{1/2}$, $q_0 = \mathrm{arctg}(y/x)$;

for $q = -p/2$ to $p/2$ step Dq **do:**

{ $r = r_0 \cos(q_0 - q)$; $H(r,q)$ += $F(x,y)$; } **}**

The available computer resources concerning both memory and processing speed naturally restrict the described approach because the HT processing speed can be evaluated in the multiplicity of $O(x_{\text{size}}.y_{\text{size}}.Q_{\text{size}})$. A popular rule to avoid computational problems here is to reduce the discrete intervals Dr and Dq for the HT-space, and in this way to improve the preciseness of HT localising facilities. So,

$$Dr = r_{\text{size}}/R_{\text{size}}, \quad Dq = q_{\text{size}}/Q_{\text{size}}, \tag{2}$$

where r_{size} usually amounts to the longest distance between two image pixels, while q_{size} is often set as one-half of revolution for more compact performance:

$$r_{\text{size}} \leq \mathbf{diag}(x_{\text{size}}, y_{\text{size}}) = (x_{\text{size}}^2 + y_{\text{size}}^2)^{1/2},\ q_{\text{size}} = p\ . \tag{2a}$$

Since x_{size} and y_{size} are given image dimensions (in pixels) then only R_{size} and Q_{size}, i.e. the HT accumulator dimensions remain for choice. But the enlargement of R_{size} and Q_{size} over some limits will roughen the HT approximations to such an extent that the "*inner noise*" level of the approximated HT comparatively to the "*exact*" HT (i.e. for given x_{size}, y_{size} and chosen $R_{\text{size}}, Q_{\text{size}}$) will make it out of any sense.

This work offers a constructive method for algorithms of approximated HT, which inner noise, i.e. HT-approximation error (see Def.1 and 2 below), is arbitrary close to zero.

3. DESCRIPTION OF THE METHOD PROPOSED

We introduce the idea of the *exact* HT for the input image $f = f(x,y)$ in viewpoint of the analytical description of HT as a Radon transform [4]:

$$h(r,q) = \iint_{RoI} f(x,y)\, d\big(x\cos(q) + y\sin(q) - r\big)\, dx\, dy \qquad (r,q) \in \boldsymbol{RoHT} \tag{3}$$

where ***RoI*** is the definition domain of the image $f = f(x,y)$, $(x,y) \in \boldsymbol{RoI}$; ***RoHT*** is the definition domain for HT of the image, i.e. for $h = h(r,q)$; and $d(.)$ is the Dirac's function.

Definition 1: The ***exact HT*** $h = h(r,q)$, $(r,q) \in \boldsymbol{RoHT}$ for a discrete image $f = f(x,y)$, $x = x_{\min} \div (Dx) x_{\max}$, $y = y_{\min} \div (Dy) y_{\max}$ is defined by the sum

$$h(r,q) = \sum_{y_0 = y_{\min}}^{(\Delta y)\, y_{\max}} \sum_{x_0 = x_{\min}}^{(\Delta x)\, x_{\max}} \tilde{h}\big(P(x_0, y_0), r, q\big),\ (r,q) \in \boldsymbol{RoHT}$$

of partial HT-s: $\tilde{h}\,(P(x_0, y_0), r, q) =$

$$= f(x_0, y_0) \iint_{P(x_0, y_0)} d\big(x\cos(q) + y\sin(q) - r\big)\, dx\, dy\ ,$$

$(r,q) \in \boldsymbol{RoHT}$, for each "real" pixel $P(x_0, y_0) \subset \boldsymbol{RoI}$, with dimensions Dx and Dy and centre (x_0, y_0).

Here and in what follows, the expressions of the type $x = x_{\min} \div (Dx) x_{\max}$ mean $x = x_{\min} + sDx$, $s = 0,1,\ldots \lceil x_{\max}/Dx \rceil$, where $\lceil . \rceil$ is the integer part of (.).

In this way, we consider the HT for a given pixel $P(x_0, y_0)$ of the discrete image as a volume bounded by the plain $h = 0$ and the enveloping surface $h = \tilde{h}\,((x_0, y_0), r, q)$. We interpret this volume as a ***generalized*** $\tilde{\mathbf{C}}$**osinusoid** $(P(x_0, y_0), r, q)$ that consists of a continuum of conventional Cosinusoids each one corresponding to a perfect point in the "real" pixel area $P(x_0, y_0) \subset$ ***RoI***:

$$P(x_0, y_0) \Leftrightarrow \begin{cases} \widetilde{\text{C}}\text{osinusoid}(P(x_0, y_0), r, q) \\ \equiv \bigcup_{(x,y)\in P(x_0,y_0)} \text{Cosinusoid}((x, y), r, q) \end{cases} \quad (4)$$

Definition 2: The ***inner noise*** $d = d(r,q)$ of a given HT performance $H = H(r,q)$, $r = r_{\min}\div(Dr)r_{\max}$, $q = q_{\min}\div$ $\div(Dq)q_{\max}$, on an image $f = f(x,y)$, $x = x_{\min}\div(Dx)x_{\max}$, $y = y_{\min}\div(Dy)y_{\max}$ is defined by the absolute deviation from the exact HT on the same image that is:

$$d(r,q) = \left| H(r,q) - \bar{h}(r,q) \right|,$$

where $\bar{h}(r_0,q_0) = \iint_{\Pi(r_0,q_0)} h(r,q)\,\mathrm{d}r\,\mathrm{d}q$ is the integral value of the exact HT for a given ***HT-pixel*** $P(r_0,q_0) \equiv$ $\{(r,q) |\ r \in [r_0 - \Delta r/2, r_0 + \Delta r/2), q \in [q_0 - \Delta q/2, q_0 + \Delta q/2)\}$, $P(r_0,q_0) \subset$ ***RoHT***.

Obviously, the inner noise of HT can be evaluated by the similar approach used for the evaluation of the HT itself, see (1) and/or (1a). The estimate (1) is smoother than (1a). However, (1a) has been preferred here because of its greater sensibility to noise.

The proposed method can be resumed as follows:

- The HT of an input image is considered decomposed by image pixels, and a "*Generalised Cosinusoid*" (GC) of HT-space is set up in correspondence to each pixel.
- Each vertical cross-section of a given GC represents a "*Symmetric Trapezium*" (ST) for a given angle q. The ST position and dimension are easily computed in HT-space.
- Each GC is considered decomposed by strips of $\pm Dq/2$ for each discrete value of q. Such a GC part can be approximated by a "*Trapezium-like Cosine-shaped Hexahedron*" (TC6). Hence, the entire GC can be represented through a sequence of TC6-s, each of them being easily computed in an arbitrary precision by their vertical sections, which are ST-s.
- Each GC for an image pixel and consequently the HT of the entire image can be computed by sequential processing of the correspondent TC6-s, respectively their ST-sections. Many of the ST-s has equal dimensions despite their different positions in the HT-space, and this is of general importance for computation.

4. ALGORITHMIC ASPECTS

Several ways are possible to approximate the TC6's, in which the (r,q)-HT can be decomposed by the method.

Basic construction "B": It inherits the basic algorithm; see par.2. Each TC6 is computed on parts in the innermost loop.

New construction "N": The embedded order of loops is changed; the first is the loop on Q and the next is the double loop (on Y and X). The reason is that the sections $\text{ST}(x_0, y_0, q)$, $x_0 = XDx$, $y_0 = YDy$, along a direction $q = Q_i/Dq$, are identical but translated accordingly to the position of each image pixel $P(x_0, y_0) \subset$ ***RoI***. Therefore it is possible to compute the sections $\text{ST}_k = \text{ST}(x_0, y_0, q)$, $q = Dq\,(Q_i - 1/2 + k/K)$ in advance for all $k = 1,\dots K$ and $Q_i = Q_{\min},\dots Q_{\max}$; while the approximations of the corresponding $\text{TC6}(x_0, y_0, q)$ remain unchanged in the inner double loop (on Y and X).

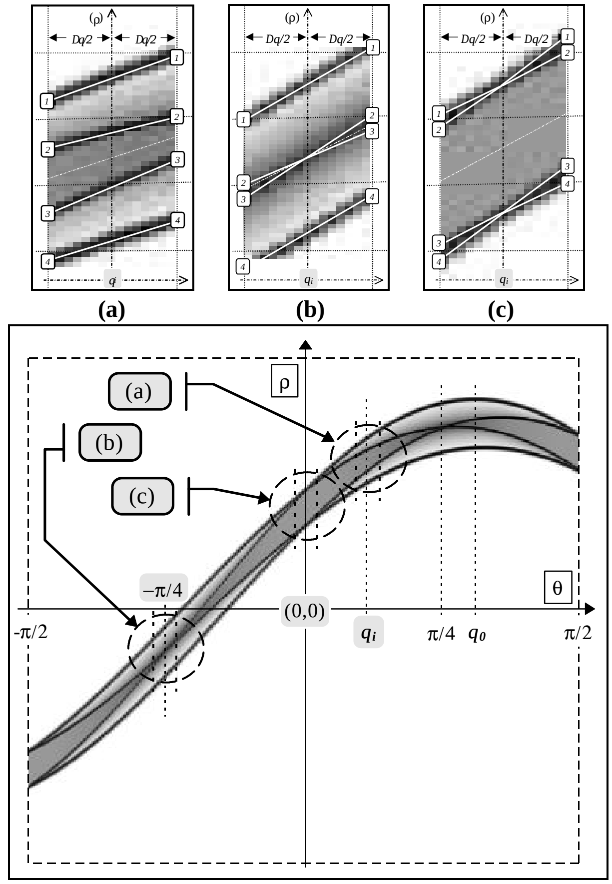


Fig.2. Three basic types of TC6 (*a view from above*)

Speedy construction "S": It resembles construction ***B***, differing in that all necessary trapezia (of number K for each direction Q) are computed in advance and then the control enters the main triple loop (Y, X, Q).

Several algorithms, which differ in the way of TC6 computation have been developed based on these three constructions, namely ***B1, B2, N1, N2, S1, S2, S3,*** besides ***B_***(the basic algorithm) and ***BF*** (an exact FP algorithm). Always herein, K that we call a "*constructive parameter*" is chosen $K = 3$.

Table 1. Results for $(X_{\text{size}}=Y_{\text{size}}=29)$ **square, by rising the HT-parameters: from** $(R_{\text{size}}=49, Q_{\text{size}}=36)$ **to** $(R_{\text{size}}=257, Q_{\text{size}}=178)$**, i.e. » 1:5**

	B_	***B1***	***B2***	***N1***	***N2***	***S1***	***S2***	***S3***	***BF***
hA_max	1485	19	21	119	119	119	412	837	24
	1334	30	21	41	41	41	76	60	20
dE1_mean	181,3	4,33	4,24	17,9	17,9	17,9	41,2	80,5	3,80
	76,4	1,58	1,58	1,83	18,0	1,80	2,54	3,23	0,28
dE1_sigma	78,4	1,02	1.02	10,5	10,5	10,4	19,9	48,1	1,29
	36,5	0,48	0,39	0,72	0,72	0,72	0,90	1,22	0,26
HT_time[s]	0.1	3.6	3.6	2.0	1.8	1.8	1.0	0.8	3.2
	0.6	28.9	26.8	28.2	24.1	25.7	12.9	7.7	26.0

5. EXPERIMENTS

The test image, we prefer due to easier precision of the HT-inner-noise evaluation, is a *square*, instead of a *disk* as usual [4]. The chosen square size is $X_{\text{size}} = Y_{\text{size}} = 29$.

Table 2. The recommended versions defined by a combination of two concurrent requirements.

Preferred Version	By lower values of Q_{size} and R_{size}	By higher values For Q_{size} and R_{size}
Processing speed	***S1 , S2 , S3 , BF***	***S2 , S3 , BF , S1***
No speed requirements	***BF , B1 , B2***	***BF , N1 , N2***

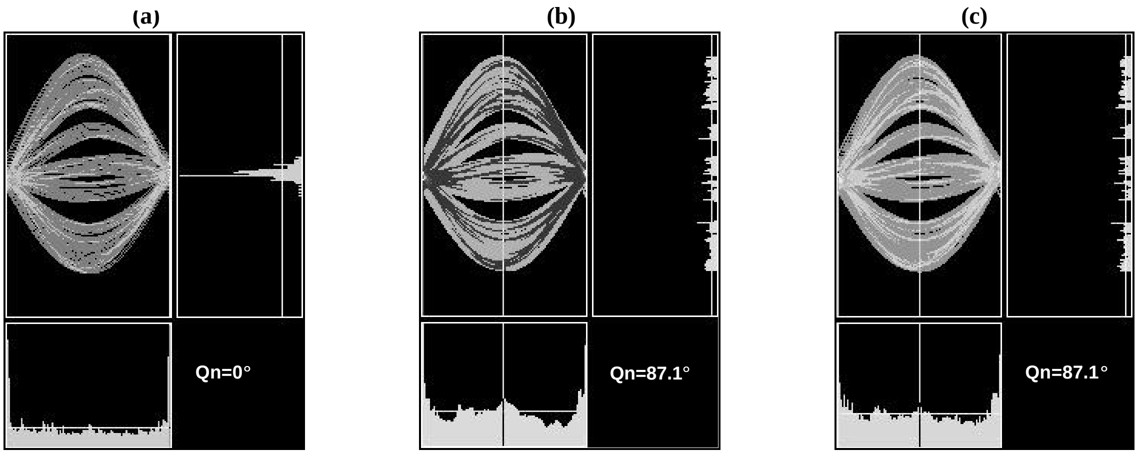


Fig.3. Evaluation of the letter slope q_n in a handwriting fragment (X_{size}=153, Y_{size}=31 ; R_{size}=183, Q_{size}=104):
(a) by the "rough" basic version *B_* (*it absents any indicative 2-d Max, the first Max gives a wrong* $q_n = 0°$);
(b) by the "precise" version *BF*, and (c) by the "speedy" version *S2* (*there is an indicative 2-d Max, i.e.* $q_n = 87{,}1°$).

5.1. Evaluation parameters

HT_time: the experimentally measured time of HT.
hA_max: the global maximum for picture $h = H[P_j][Q_i]$.
hE1_mean: the global average of $x \sim x(Q_i)$, see (1).
hE1_sigma: the mean square variation of $x \sim x(Q_i)$.

The HT inner noise (see Def.2) of the algorithmic versions here, can be evaluated by the similar estimate function *x*, see (1a). The corresponding parameters are denoted by the prefix "***d***" (instead of "***h***") in Table 1.

5.2. Comparative description of experiments

Table 1 compares lower towards (≈5 times) higher values for Q_{size} and R_{size}, for the chosen test image. So, we can order the versions by closeness to the exact HT as:

$$\boldsymbol{BF \prec B2 \prec B1 \prec (N1 \prec N2 \prec S1) \prec S2 \prec S3 \prec\prec B\_} \quad (5)$$

and by processing speed as:

$$\boldsymbol{B\_ \prec\prec S3 \prec S2 \prec (N2 \prec BF \prec S1 \prec N1) \prec B2 \prec B1} \quad (5a)$$

Table 2 shows a classification of the algorithmic versions by preferences expertly concluded herein.

5.3. Application in text image processing

Fig.3 illustrates the evaluation of the averaged font-slope q_n of a hand-written text. The shown phrase is isolated from a larger text image by the EX_HT, an experimental system for exact HT. The writing thickness here is 1÷2 pixels. The first maximum of the horizontal histograms corresponds to the phrase longitude, and the second one – to its letters' slope. The weak capability of usual algorithms and the efficiency of the algorithms proposed herein are illustrated.

6. CONCLUSION

More details about the method proposed can be found in [6]. The method has been successfully used for preliminary processing of textual images, and more precisely, for evaluation of the letter slope in images of handwriting. A similar approach of exact HT can be sought by an analogy for other interpretations of HT [2].

References:

[1] Hough P.V.C., Method and means for recognizing complex patterns, U.S. Patent 3-069-654, 1962.

[2] Illingworth J., J. Kittler, "A Survey of the Hough Transform," *Comp. Vision, Graphics, and Image Proc. J.*, Vol. 44, pp. 87-116, 1988.

[3] Duda R.O., P.E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Commun. ACM*, Vol. 15, pp. 11-15, 1972.

[4] Shapiro V.A., "From Radon to Hough Transform of Gray-Scale Images via Digital Halftoning," *Proc. of 8-th Scandinavian Conf. on Image Analysis*, pp. 665-672, 1993.

[5] Impedovo S., "Frontiers in Handwriting Recognition," *Fundamentals in HWR*, In S. Impedovo (Ed.), NATO ASI Series "F": "C&S Sci", Vol. 124, Springer-Verlag, Berlin, pp. 7-39, 1994.

[6] Dimov D.T., "Using Hough Transform for Image Segmentation of Textual Rows," *Working papers series of IIT-BAS*, ISSN-1310-652X, WP/49, pp. 1-26, 1998.