A Scalable, High-Precision, and Low-Noise Detector of Shift-Invariant Image Locations

Georgii Khachaturov *

Departamento de Sistemas, Universidad Autónoma Metropolitana, Av. San Pablo 180, Col. Reynosa Tamaulipas, C.P. 02200, D.F., Mexico

Abstract

A scalable, high-precision, and low-noise detector of shift invariant locations in grayscale images is presented. It leads to a wide range of novel image-to-'data structures' processing algorithms. Experiments with a single algorithm of this range prove that (*i*) the output structures convey great amount of semantically relevant information about the original image; (*ii*) this information can be successfully extracted and used in subsequent applications.

Keywords:

Feature detection

Shift invariance

Multi-scale processing

Image-to-'data structures' processing

* Telephone: **+52 55 5318 9532, ext. 139;** Fax: **+52 55 5394 4534**
E-mail address**: xgeorge@correo.azc.uam.mx**

1. Introduction

Shift invariance is a property opposite to *edges* – an edge corresponds to the location of the strongest change of image along some direction; shift invariance corresponds to a location with a direction of zero change. These two properties are dual each other in the following sense.

A functional of 'shift resistance' can be easily introduced. Such a functional of two arguments – image location and a direction in image plane – would compute the magnitude of spatial derivative along the input direction for a fragment at the input location. Edge detection can be carried out by search for local maxima of this functional; detection of shift invariance can be achieved by search for those directions for which shift resistance is equal to zero. For a fixed value of the first argument, the direction of the strongest change of the functional is orthogonal to the direction of zero change.

Well-known edge detectors such as those by Sobel, Roberts, Canny, and some others, see survey by (Mlsna and Rodriguez, 2005), can be derived from particular instances of such a shift resistance functional.

These detectors use convolution masks to signal an edge. The number and dimension of the masks depend on the detector chosen, but regardless of the choice, a dual set of masks can be derived to signal zeros of shift-resistance functional, instead of local maxima. The dual set of masks can be easily found using the orthogonality mentioned in the previous paragraph.

The author is unaware of any study that develops this idea consistently. Probably, no such research was undertaken because its practical value seems rather questionable and many drawbacks are obvious from the beginning. Indeed, the precision of output direction would be rather low whereas the probability of wrong detection be rather high; such a low quality of detected features introduces serious obstacles for subsequent applications, for instance, that deal with grouping features into larger objects.

This paper presents a scalable low-level detector based on a quite different idea than the above duality principle. The detector input is a point of image frame represented in R^2 (i.e., input components are not integer, but real); a positive response includes a high precision (real) estimate of the direction; it has a low level of wrongly detected shift-invariant locations; it is robust with respect to additive noise.

The detector was presented originally by (Khachaturov, 1995) as the Three Frequencies Method (the 3FM). Later, the 3FM was combined with a post-processing filter to suppress too "trivial" shift invariant locations – those that the 3FM generates at any point of an image represented by a linear function of two arguments, (Khachaturov and Moncayo-Muños, 2004).

Presence of a directional vector in output is a common property for a detector of shift invariance and for an edge detector. This similarity may provoke some expectations about resemblance of other properties or criteria for these kinds of detectors. For example, recall the principles declared by (Canny, 1986) for an optimal edge detector: good detection, good localization, and minimal response. One can try to impose these criteria for optimality of a detector of shift invariance. However, this idea proves to be futile.

Indeed, for a linear function of two arguments, *any* location in the domain of definition satisfies the property of shift invariance. That is, in contrast to an edge detector that generates 1D output, a shift invariance detector may generate positive responses in 2D domain. Hence, it is impossible to apply at least one of Canny's principles – minimal response – to a detector of shift invariance.

A natural doubt arises: Do the shift invariant locations have any practical meaning, or not?

(Khachaturov and Moncayo-Muños, 2004) show an example of processing where shift invariant locations are applied to a well known computer vision problem. It is achieved by involving the 3FM in an intermediate processing that leads to a kind of robust features. Then these features are used as input data for an instance of the correspondence problem that is successfully solved. [This processing performs exhaustive multi-scale filtering of an input image by the 3FM. For a fixed scale, after a positive response of the detector, the filtering is interrupted temporary to trace the curve that would contain just detected shift-invariant element. After the filtering for a scale is over, traced curves are formally merged into larger objects. For two curves to be adjoined so, they must have at least one pair of shift invariant elements close in position and orientation. Then, such 'larger objects' of different scales are agglomerated into objects of even more complex structure by a similar rule (a couple of 'larger objects' constructed for adjacent scales must be merged if they have at least one pair of close shift invariant elements). These complex objects are then used to construct a synthetic image in the way described below in Section 3. Finally, the robust features are constructed at local minima of the synthetic image.]

The present article refines mathematical fundamentals of the 3FM, describes the post-processing filter, the idea of which is suggested in (Khachaturov and Moncayo-Muños, 2004). As well, it presents quite new elements: (*i*) an analysis of the computational complexity, (*ii*) a comparison of the detector of shift-invariance versus Canny's edge detector, (*iii*) and quantitative experiments with simulated images that prove specific advantages of the 3FM.

The author is unaware of prior works related to detection of shift invariance. So perhaps, the 3FM is the first detector of this kind, except the straightforward method mentioned above.

The rest of the paper is organized as follows: Section 2 introduces the algorithm of the 3FM and then its theoretical fundamentals; Section 3 describes the post-processing filter and a related experiment with a real image; Section 4 presents an analysis of numerical complexity that is applied in Section5; Section 5 compares the 3FM with edge detectors. In particular, while treating the numerical complexity, the 3FM is compared with Canny's detector; Section 6 summarizes experiments: an emphasis on the experiments with simulated images is done because they lead to quantitative estimates of precision, robustness, and noise; Section 7 contains a conclusion.

2. The 3FM

We start this section from a description of how the 3FM works. Then an explanation of why it works follows.

Being applied at point $\boldsymbol{v}\in R^2$ of the image frame, the 3FM constructs a set of rectangular windows centered at $\boldsymbol{v}$ and indexed by parameter $\alpha\in[0, \pi)$. Any window is obtained by rotation of a standard window at a standard (horizontal) orientation and α represents the corresponding rotation angle. The size of all windows is $Kh\times Lh$, where K and L are natural numbers and h is a real number that represents scaling parameter to control size of the windows.

While processing image content inside the windows, the 3FM computes values of three functions of α – $S_{-1}(\alpha)$, $S_0(\alpha)$, and $S_1(\alpha)$ – where for any option of index n, which can be –1, 0, or 1, $S_n(\alpha)$ is defined as follows:

$$S_n(\alpha)=(KL)^{-1}\,\Big|\sum_{\substack{k=1,\dots,K\\ l=1,\dots,L}} F_{k,l}(\alpha)\, e^{-\frac{2\pi i}{L}(1+\frac{n}{K})(kL+l)}\Big|. \qquad (1)$$

To define $F_{k,l}(\alpha)$ of equation (1), each window of the sequence is regarded as a $K\times L$-lattice formed by $h\times h$-square cells; any pair $\{k,l\}$ represents coordinates of a cell in the lattice; and finally, $F_{k,l}(\alpha)$ represents mean value of the image inside cell $\{k,l\}$ for that window of the set, which corresponds to rotation angle α.

Then, $\{S_n(\alpha)\}_{n=-1,0,1}$ are subjected to the following rule:

The 3FM generates a positive response at *v* iff there exists such α^* that

$$\{\ \alpha^* = \mathit{arg\ local\_max}\ S_0(\alpha), \text{ and simultaneously } S_{-1}(\alpha^*) = S_1(\alpha^*) = 0\}. \qquad (2)$$

The directional component of a positive response by the 3FM is denoted below by ***w***. Conventionally, ***w*** is identified with the direction of that side of the window rectangle which contains *K* cells. In more strict terms, let "*K*-side" stand for that side of the above standard window at horizontal position, which has length *Kh*. Then, for a positive response of the 3FM:

- The direction of local shift invariance of the image at location $\boldsymbol{v} \in \boldsymbol{R}^2$ is co-linear with the unit vector ***w*** obtained from the *K*-side by its normalization and rotation at angle α^*;
- The output of the 3FM is represented by the pair {***v, w***}.

In the rest of this section, we explain why the 3FM works.

First of all, let us justify the name of the 3FM.

The above 2D-table $\{F_{k,l}(\alpha)\}_{k=1,\dots,K;l=1,\dots,L}$ has an equivalent 1D-representation by row $\{f_t(\alpha)\}_{t=1,2,\dots,LK}$. This representation is as follows:

Let $(k,l) \leftrightarrow t$ be a one-to-one correspondence between nodes of $K\times L$-lattice and *KL*-row established by equation $t=kL+l$, so $f_t(\alpha)$ can be defined as $F_{k,l}(\alpha)$. That is, the two dimensional table of *F* is developed into a one-dimensional row of *f* as follows:

$$\begin{aligned}\{f_1(\alpha), f_2(\alpha), \dots, f_{LK}(\alpha)\} = \{\ & F_{1,1}(\alpha), F_{1,2}(\alpha), \dots, F_{1,L}(\alpha),\\ & F_{2,1}(\alpha), F_{2,2}(\alpha), \dots, F_{2,L}(\alpha),\\ & \dots\dots\dots\dots\dots\dots\dots\dots.\\ & F_{K,1}(\alpha), F_{K,2}(\alpha), \dots, F_{K,L}(\alpha)\}\ .\end{aligned}$$

For this representation, formula (1) is equivalent to $S_n(\alpha)=(KL)^{-1}|\sum_{t=1,2,\ldots,KL} f_t(\alpha)e^{-\frac{2\pi i}{L}(1+\frac{n}{K})t}|$.

It is clear that for $n= -1$, 0, and 1, the last expression coincides with three consecutive frequencies of discrete Fourier transform (DFT) of *f(t)* (as usual, 'frequency' here is the conventional name of discrete argument of DFT-image of a function of one variable; it is opposite to the 'time' as the conventional name of argument of the original function.).

This observation justifies the name of the 3FM.

The key property that leads to the 3FM is the fact that any criterion of shift invariance in a 2D-image is equivalent to a criterion of periodicity of a function of one variable. More specifically, for the above lattice $\{F_{k,l}(\alpha)\}_{k=1,\ldots,K;l=1,\ldots,L}$ its local shift invariance along the *K*-side is equivalent to periodicity with period *L* of the function $\{f(t)\}_{t=1,2,\ldots,LK}$.

Theoretic foundation for the criterion of periodicity is presented by the following

Theorem. Given a real function *f(x)* of period *T*, which for any *x* has expansion in Fourier series, a natural number *K*, a small real number δ, the functional Ω_K defined as

$\Omega_K(f,\delta) = \frac{1}{KT}\int_0^{KT} f(x)e^{-i(1+\delta)x\frac{2\pi}{T}}dx$, then:

(a) $|\Omega_K(f,0)|$ is a constant independent of K;

(b) for any $\delta\neq0$, it is true that $|\Omega_K(f,\delta)|\to0$ as $K\to\infty$;

(c) $\Omega_K(f, 1/K)= \Omega_K(f, -1/K)=0$.

Let us understand why Theorem leads to the criterion of the 3FM presented by expression (2).

In practical terms, a combination of assertions (a) and (b) of Theorem means that, for sufficiently large *K*, $|\Omega_K(f,0)|$ is a local maximum of $|\Omega_K(f,\delta)|$ or, say, quite close to the maximum. On the other hand, on the basis of the same theorem, if the period of periodic function *f* is unknown, we can try to estimate it as $T(1+\delta^*)$, where $\delta^*= argmax\,|\boldsymbol{\Omega_K}(\boldsymbol{f}(\boldsymbol{x}),\delta)|$.

In the case under consideration, *f* is not an arbitrary function, but a 1D-row obtained from the 2D-table *F*. If *F* is shift-invariant along its *K*-side, then *f* is periodic and its period is known a priori as *L*. So in this case $\delta^* = 0$ is a local maximum of $|\Omega_K(f(x),\delta)|$. As it was mentioned just above, this maximum can be detected by variation of δ, but it is easy to see that, if *F* is shift-invariant along its *K*-side, the variation of δ at δ=0 affects the values of functional $|\Omega_K(f(x), \delta)|$ in the same way as the variation of orientation of the window that yields table *F*. This terminates semi-intuitive justification of the first half of criterion (2).

The usage of S_{-1} and S_1 in the second half of (2) is a straightforward application of Theorem (c).

In the rest of this section, we prove Theorem.

Proof. Affirmation (a) follows from *T*-periodicity of $f(x)e^{-ix\frac{2\pi}{T}}$. Let us prove (b).

Let $f(x) = \sum_m c_m \exp(imx2\pi/T)$ be Fourier expansion of f. Then $\Omega_K(f, \delta) = \frac{1}{KT}\sum_m c_m A_m$, where $A_m = \int_0^{KT} e^{imx\frac{2\pi}{T}} e^{-i(1+\delta)x\frac{2\pi}{T}}\,dx = \frac{T}{i(m-1-\delta)2\pi}(e^{i(m-1-\delta)2\pi K}-1) = \frac{T}{i(m-1-\delta)2\pi}(e^{i(-\delta)2\pi K}-1)$.

Thus

$$\Omega_K(f, \delta) = \frac{1}{KT}\sum_m c_m \frac{T}{i(m-1-\delta)2\pi}(e^{i(-\delta)2\pi K}-1) \quad = \frac{e^{i(-\delta)2\pi K}-1}{Ki2\pi}\sum_m \frac{c_m}{(m-1-\delta)} \quad . \qquad \textbf{(3)}$$

The series $\sum_m c_m$ gives a value of *f*(0) and hence converges due to assumptions of Theorem. Due to classic properties of convergence of series (Lang, 1983), the combination of this convergence with the one $\frac{1}{(m-1-\delta)} \xrightarrow[m\to\infty]{} 0$ leads to convergence of the series in the right-hand side of (3). Its sum is independent of *K*, meanwhile the module of the *K*-dependent expression $e^{i(-\delta)2\pi K} - 1$ is bounded. This obviously implies (b).

To prove (c), using as above Fourier expansion of *f*, integral $\int_0^{KT} f(x)e^{-Nx\frac{2\pi}{KT}}\,dx$ can be re-written as $\sum_m c_m \int_0^{KT} e^{imx\frac{2\pi}{T}} e^{-iNx\frac{2\pi}{TK}}\,dx$. After change of variables, *y=x/K*, each integral of this series becomes equal to $I_m = \gamma\int_0^T e^{imKy2\pi/T} e^{-iNy2\pi/T}$, where γ is a common quotient for all *m*. Computation of the last integral is trivial and analysis of its results allows us to conclude that $I_m = 0 \Leftrightarrow mK \neq N$. But, if $N = K \pm 1$, then $mK \neq N$ for any integer *m*, Q.E.D.

3. Post-Processing Filter for 'Slope-wise' Image Fragments

Let us consider an input image represented by a 'slope-wise' brightness function, that is, by an arbitrary linear function of two arguments with non-zero gradient. For such kind of image, the 3FM signals with positive response at any point of the image frame because by construction all image locations are shift invariant along the direction normal to the image gradient.

Detection of such objects may be regarded as undesirable in some application contexts.

To keep under control acceptance/rejection of the 'slope-wise' image fragments, a post-processing filter is introduced by (Khachaturov and Moncayo-Muños, 2004) to analyze each positive response of the 3FM.

The idea of the filter is as follows. Let $\{F_{k,l}(\alpha^*)\}_{k=1,\dots,K;l=1,\dots,L}$ be an image fragment represented in the same way as above for a positive response of the 3FM. Then $\{F_{k,l}(\alpha^*)\}$ is subjected to a standard statistical test, so-called verification of 0-hypothesis (Johnson et al, 1997), applied to the hypothesis that the correct functional model for mathematical expectation of $F_{k,l}(\alpha^*)$ is given by the function $\varphi_{\lambda,\mu}(k,l)=\lambda+\mu l$, $k=1,\dots,K;l=1,\dots,L$, for some unknown real λ and μ.

More specifically, the filter algorithm provides the following steps according to standard verification of 0-hypothesis:

- Construct the least square approximation (linear regression) of $\{F_{k,l}(\alpha^*)\}_{k=1,\dots,K;l=1,\dots,L}$ by $\varphi_{\lambda,\mu}$. That is, given the least square functional $\Psi(\lambda,\mu)=\sum_{k,l}[F_{k,l}(\alpha^*)-\varphi_{\lambda,\mu}(k,l)]^2$, the components of pair $\{\lambda^*,\mu^*\}$ are constructed in standard way to satisfy $\{\lambda^*,\mu^*\}=\arg\min\ \Psi(\lambda,\mu)$;
- If $\Psi(\lambda^*,\mu^*)<\delta_{\text{post\_3FM}}$, where $\delta_{\text{post\_3FM}}>0$ is a threshold, then reject the current positive response of the 3FM.

Threshold $\delta_{\text{post\_3FM}}$ controls the share of 'slope-wise' fragments passing the filter. Its influence is studied empirically. Fig. 1 illustrates action of the filter for different values of $\delta_{\text{post\_3FM}}$.

[Fig. 1 refers to 'synthetic image' constructed by a "data structures-to-image" transform. This is an inverse transform with respect to the processing of the original image and it is constructed as follows. By a positive response $\{\boldsymbol{v}, \boldsymbol{w}\}$ stored in a data structure, a function of two arguments is constructed so that its support coincides with the window that cuts table $\{F_{k,l}(\alpha^*)\}_{k=1,\ldots,K;l=1,\ldots,L}$ from the original image. Inside the window, this function is constant along direction $\boldsymbol{w}$ and represents a part of normal distribution along the normal direction $\boldsymbol{w_n}$ with maximum at $\boldsymbol{v}$. Then, a sum of such functions is constructed for all positive responses registered at all scales. And finally, the synthetic image represents a normalized negative of this sum.]

The aim of introducing the synthetic image was a qualitative estimation of informational impact that contributes all discrete structures constructed during image processing. The experiments show a strong resemblance between the original and the synthetic image. Hence those data structures can be efficiently used to extract any kind of semantically relevant information that a human is able to extract observing a synthetic image.

In addition, an unexpected effect can be observed by the experiments of Fig. 1: the boy's face in the last synthetic image looks like a usual shading portrait, then synthetic images in the counter clock-wise direction make visible some muscles under skin, and finally, the first image develops some skull bones. In other words, variation of threshold $\delta_{\mathrm{post\_3FM}}$ works as virtual focusing of depth to develop invisible details inside the boy's head.

Informally, imitating the name of well known computer vision technique – 'shape from shading', see a survey by (Zhang et al, 1999) – this example shows a kind of 'volume from multi-scale image-sketch'-technique.

4. Numerical Complexity

An algorithm that supports the presented approach must contain the following principal block operations:

- Given α, construction of an instance of table $\{F_{k,l}(\alpha)\}$ presented in Section 2;
- Computing three functions of α, $\{S_n(\alpha)\}$, $n = -1,0,1$ according to equation (1) of Section 2;
- Application of the rule of expression (2) of Section 2;
- Post-processing filter of Section 3.

They must be taken into consideration for an estimate of numerical complexity of the method.

Analyzing complexity of these items in the rest of this section, we focus our attention on the number of multiplications only.

Now let us consider the computation of $\{F_{k,l}(\alpha)\}$ and $\{S_n(\alpha)\}$.

The table $\{F_{k,l}(\alpha)\}$, in practice, can be computed in a slightly different way than the one described in Section 2: instead of mean values inside a lattice cell, the value of $F_{k,l}(\alpha)$ can be interpolated by four image pixels close to the center of (k,l)-cell.

This way coincides with the one described in Section 2 for the size of lattice cells to be equal to the image pixel size. For lattices of a larger size we also adopted this modified rule, however in such a case, it is applied not to the original image but to its convolution with a Gaussian averaging weight mask. After convolution, each of four pixels that participates in the interpolation conveys information of many pixels of the original image. The larger size of the cell, the wider range of the bell-wise mask of the unit summary weight. Complexity of the convolution is ignored in actual analysis. (Section 5.2 presents a justification of this decision.)

Then, in the same double cycle (with k and l as parameters of the cycle), the computation of an element $F_{k,l}(\alpha)$ can be combined with using this element for computation of $\{S_n(\alpha)\}$, $n = -1,0,1$.

Our code written in language C follows this scheme and carries out interpolation of $F_{k,l}(\alpha)$ by four closest pixels and then computes $\{S_n(\alpha)\}$, $n = -1,0,1$.

Given pair $\{k,l,\}$, the number of multiplications in the body of the inner cycle of our code is 19.

Thus, for this part of the whole algorithm, *the total number of multiplications for a single act of application of the 3FM is equal to* $19KLN_\alpha$, where N_α stands for the number of 'α's.

For instance, in experiments by (Khachaturov and Moncayo-Muños, 2004), N_α is optionally equal to 15 or 7 for, respectively, the general application of the 3FM and for an application in the process of tracing a curve after a positive response of the 3FM.

The rest of block-operations mentioned at the beginning of this section has a minor contribution into complexity.

For instance, our implementation of the rule (2) works in two steps. Firstly, it revises three sparse tables $\{S_n(\alpha_i)\}, n= -1,0,1$, $i=1,\ldots, N_\alpha$ to finds such three successive values i_1, i_2 , i_3 that interval $[\alpha_{i_1}, \alpha_{i_3}]$ should contain, if any, α^* that satisfies condition (2). At this moment the precise value of α^* is unknown yet. It is determined at the next step as $\alpha^* = \mathit{argmax}\ S_0(\alpha)$ on the basis of a simple quadratic interpolation of $\{S_0(\alpha_{i_1}), S_0(\alpha_{i_2}), S_0(\alpha_{i_3})\}$.

The number of arithmetic operations needed for this block-operation is linear in N_α, but it does not depend on *KL*.

The number of arithmetic operations required for the post-processing filter is linear in *KL*, but it does not depend on N_α

Both these numbers are negligible compared to the number found in previous item and may be omitted from future consideration.

5. Comparison of the 3FM with Edge Detectors

5.1. *Meaning of Detected Features*

Unlike the strict mathematical meaning of shift invariance, the notion of edge in Computer Vision does not correspond yet to a single commonly accepted meaning.

Most edge detection methods deal with a 2D grey-scale function of brightness and treat edges as local maxima of gradient's magnitude of this function. The direction normal to the gradient at an edge location is declared as its direction.

Most techniques apply mathematical properties of maxima of the gradient function to develop an algorithm that would signal this property in digital images. In turn, 'digital image' means a discrete sample of the brightness function represented typically as a rectangular table of pixels.

Two important lines in development of edge detectors can be found in surveys by (Mlsna and Rodriguez, 2005), and by (Acton, 2005). The former (and more traditional) line refers to the gradient and Laplacian methods, whereas the latter refers to those that involve partial differential equations.

Likewise detectors of the former line, the presented detector of shift invariance belongs to the low level of a traditional image processing architecture that includes the low-level, intermediate-level, etc.. For the latter line, edges appear as a by-product of processing that extracts some larger objects, for instance, 'snakes' by (Kass et al, 1987). This processing is normally organized as an iterative scheme, and mathematically corresponds to optimization of a functional over input image.

As to semantic interpretation of output, notice that for the latter approach, perhaps, there is no 'intuitive edge' at a location marked as an 'edge' found as a part of snake. It occurs because smaller details can be restored during construction of a larger object under a global optimization criterion. So, this approach may lead to construction of phantom edges or change position of real edges.

Frequently, researchers use simulated images for studying properties of an edge detector. Typically, a function of one variable – 'step-wise', 'delta-wise', 'roof' function, etc. (Nalwa, 1993) – is used to construct a test-bed 2D image of edge. The image value along the direction of a test-bed edge is set to a constant. So, by construction, such test-bed images are shift-invariant. The 3FM yields a positive response to any of such test-bed images and in this meaning, it is able to detect a wider class of objects than edge detectors.

To compare furthermore notions of shift invariance versus 'edges', let us go into some formal details.

Let us regard image as function $F\colon R^2 \rightarrow R_+$.

Definition. Given domain $D \subset R^2$, the local shift invariance of image F inside D means that there exists a unit vector $w \in R^2$ such that the equation $F(u)=F(u+\varepsilon w)$ holds for all sufficiently small $\varepsilon \in R_+$ and all $u \in D$.

The feature that describes shift invariance for domain D is formed by pair $\{v, w\}$, where $v \in R^2$ is a vector of location (say, center) of D and $w \in R^2$ is a unit vector that represents the direction of shift invariance .

In general, a shift invariant domain D described by feature $\{v, w\}$ generates a two-parametric family of shift invariant domains $\{D_{s,u}\}_{s,u \in R}$, so that for all sufficiently small s and u, $D_{s,u}$ satisfies the same property of shift invariance as D, and is described by feature $\{v+s\boldsymbol{w_n} +u\boldsymbol{w}, \boldsymbol{w}\}$, where $\boldsymbol{w_n}$ is unit vector normal to w.

That is, *a positive response of the 3FM with output {**v, w**} leads, in fact, to positive responses {**v*, w**} for any **v***$\in$**V**$_v$ where **V**$_v$ is a small neighborhood of v.*

The feature that represents an edge is described by a similar pair {v, w} with the same meaning of components as above. Let us remind again Canny's principles mentioned in Section 1, and more specifically the 'minimal response' principle, which means that any positive response {v, w} of a 'good' edge detector at pixel v should ban positive responses for pixels {$v+\varepsilon w_n$} closest in the transversal direction w_n to v.

A simple comparison of two previous paragraphs shows that 'minimal response' principle is meaningless for a detector of shift invariance.

Note however that application of Canny's edge detector to a 'delta-wise' (or 'roof-like') image fragment generates two positive responses: one on each side of the local image maximum. That means that the Canny's principles being combined with the definition of edges as 'maxima of gradient's magnitude' do not match in some details with our intuitive idea of edge.

5.2. *Comparison with Complexity of Canny's Edge Detector*

Let us compare complexities of exhaustive filtration of an image by a combination of the 3FM with the post-processing filter (this combination, for brevity, is denoted by *the 3FM**) and, on the other hand, by Canny's edge detector (denoted below by *CED*) (Canny, 1986).

These two detectors have many common properties. In particular, either of them

i. may perform preliminary convolution with Gaussian averaging mask and can be used in a multi-scale processing,
ii. evaluates direction of output features,
iii. allows tracing curves using a detected feature as seed; both approaches may perform tracing on the basis of the hysteresis principle: a higher level of threshold to catch a seed for tracing a curve, and a lower one to stop the tracing.

In the two respective detectors, the step (i) is applied for different purposes: in the 3FM*, it reduces the numerical complexity of construction of table $\{F_{k,l}(\alpha)\}$ and this step is omitted for small values of scaling parameter h, $h<2$; unlikely, in CED, it reduces noise.

For both approaches, a combination of step (i) and (ii) inside an exhaustive image filtration loop extracts a crude set of features. So steps (i) and (ii) should be regarded as a part of low-level processing.

Unlikely, the step (iii) should be regarded as a part of intermediate-level processing. After detection of a feature, both the 3FM* and CED may interrupt image filtering to trace the curve containing the feature, and later resume filtering. However, substantial difference in basic properties of outputs of the 3FM* and CED leads to quite different methods of tracing.

These observations explain why we focus comparison of complexities only on the step (ii) of both detectors. It was estimated already for the 3FM* in Section 4. Now let us evaluate the complexity of (ii) for CED.

Algorithmically, the core operations of CED are as follows. For each direction of a set of directions parameterized by index α (of the same meaning as for the 3FM), CED convolves an image fragment with all square mask of set $\{M_{\alpha_i}\}_{i=1}^{i=C_\alpha}$. Masks of the set are computed beforehand to be optimal in a certain meaning. CED chooses the mask with the highest response and then compares it with a threshold. So if the highest response is greater than the threshold, then an edge is detected and simultaneously its direction is estimated by the mask index.

Practically, it means that the number of multiplications is equal to $C_\alpha r^2$ for an instance of application of CED, where C_α is the number of directions similar to N_α of Section 4 and r corresponds to size $r\times r$ of each mask M_α.

Now let us compare complexities of the 3FM* and CED for exhaustive filtration of an X×Y-image.

Let us assume that K>L. For exhaustive filtration of an image by the 3FM*, there is no need to apply it at each pixel of the image: it suffices to apply it at each node of a lattice with cells of the size {L/3×L/3} inside the $X\times Y$-image frame. Due to Section 4, it gives $19KLN_\alpha XY/(L/3)^2 = 9\times 19\, XY\, (K/L)\, N_\alpha$ multiplications.

For filtering the same image by CED, $XY\, C_\alpha r^2$ multiplications are required. The ratio of these two numbers is $[9\times 19\, (K/L)\, N_\alpha] / [C_\alpha r^2]$.

For example, in experiments by (Khachaturov and Moncayo-Muños, 2004), the following values are used: $K=16$, $L=12$, $N_\alpha=15$. In literature (Mlsna and Rodriguez, 2005), for CED, the following values are mentioned: $C_\alpha=6$, $r=5$. For these values the ratio is $[9\times 19 \times (16/12) \times 15] / [6 \times 25] = 19 \times 2 \times 3 / 5 = 22.8$.

The *number_of_calls/sec* for both approaches is compared experimentally using function *clock*() of *C* time library. For the same data as above in analytical comparison, we have *1.7*e+6 for Canny's kernel vs *1.1*e+5 for the kernel of 3FM. So the experimental ratio is slightly lower than the theoretical one.

This ratio obviously can be reduced even more because no special optimization of parameters *K*, *L*, and N_α has been undertaken so far. Anyway, it may be asserted that filtration by CED (without construction of edge map, which corresponds to step (iii)) is about *10* times faster than by the 3FM*.

5.3. ***Precision, Noise, and Robustness***

CED is designed to satisfy Canny's 'good localization principle' — a pixel that represents a detected edge should be as close as possible to its prototype in the real image. Unlikely, the component *v* of an output {*v*, *w*} of the 3FM* is robust but imprecise due to the observations presented in the end of Section 5.1.

The directional component *w* of the 3FM* output is a unit vector computed from angle $\alpha^* \in [0, \pi/2)$. Due to the logic of construction of α^* described in Section 2, the precision of this real value is quite high. Unlikely, for CED, an edge direction is always rounded to a value of a few options.

A singularity of the 3FM is that it works on the basis of simultaneous fulfillment of three independent conditions. In contrast, most edge detectors (for example, CED) are based on revision of a single condition. Thus, the 3FM is much more noise protected than conventional edge detectors.

Let function N(S) denote estimated number of detected noise elements by a detector in the image of area S. The above singularity means that $N_{CED}(S) \sim [N_{3FM}(S)]^3$. In other words, the noise detected by CED is proportional to a cubic function of the noise detected by the 3FM*.

In the opinion of the author, these properties compensate completely the complexity drawback of the 3FM* presented in the previous item, because they simplify significantly logic of subsequent applications that use output of the 3FM*.

6. Experiments with simulated images

In addition to the experiment on complexity mentioned in Section 5.2, experiments with simulated images are undertaken as a quantitative study of the 3FM. The collected experimental data give clear evidence of specific advantages declared in the beginning of this paper: high precision, robustness, and low level of false detections.

Two kinds of images were used in our experiments: (*i*) a set of parallel dark and bright strips of the same width; (*ii*) images of sinusoid-wise brightness with respect to the direction normal to the one of shift-invariance. The results are quite similar for both options, so below only the tests with strip-wise images are presented.

The strip width is varied in experiments. The level of additive noise presents another variable. A sample of thousand images with different orientations is generated for each pair {*width, level of noise*}. In Fig. 2, the former of these variables is presented in an equivalent pictorial form, as a shot of content of the rectangular window that the 3FM operates with; the meaning of the window is described in Section 2. The latter is presented in the first text column of Fig.2.

In experiments, the one-byte dynamic range D of grayscale image is divided into three equal parts so that dark and bright strips have brightness equal, respectively, to $D/3$ and $2D/3$. Uniformly distributed noise is added independently to all pixels of an image. Values of noise belong to $[-LD/3, LD/3]$, where 'noise-level' L (the 1^{st} column of Fig.2) belongs to [0, 1]-range. For example, no noise is added as L=0 and the maximum noise value (for L=1) can even transform a "bright" pixel to a "dark" one, and vice versa.

The statistics corresponding to any instance of {*width, level of noise*} are presented in text columns 2-5 of Fig.2. The 2^{nd} column contains the share of negative responses (misses) of the 3FM. For any positive response, the angle error is computed as absolute value of difference between the estimated and the original direction of strips. Statistics of other columns of Fig.2 are based on this error: the 3^{rd} column gives the share of positive responses with too large errors (conventionally, greater than 15°), the 4^{th} gives the share of positive responses with not too large errors, and the 5^{th} column presents standard deviation of angle estimates. The standard deviation is computed only by positive responses with "not large" errors.

On the top of Fig. 3, the left and central images are respective examples of "no noise" and "maximal noise" images. These two images represent instances of input for experiments corresponding, respectively, to the first and the last rows of the third (from top to bottom) strip width in Fig.2.

In the author's opinion, the results for the first three strip widths can be regarded as excellent. Directional estimates with an error about 1° seem quite acceptable for practical applications.

Standard deviations for experiments that correspond to the first width are somewhat higher than for the second and the third. This is natural because *non-uniformity* of the images with the first width is the lowest for these experiments. [Non-uniformity of images is introduced in (Khachaturov, 1999) as quadratic form with the matrix $Q_g = \int_S \; P(x)P^T(x)dx,$ where $g(x)$ is image represented as function defined inside the image frame, $P(x) = grad\ g(x)$, and S is a fragment in the image frame. For matrices A and B of the same dimension, the generally accepted notation '$A>B$' is used to indicate that inequality $\mathbf{z}^T A\mathbf{z} > \mathbf{z}^T B\mathbf{z}$ holds for any vector $\mathbf{z}$ and quadratic forms based on respective matrices. Then, it is proved that if two different patterns painted on the same real object yield two different images, $g_1(x)$ and $g_2(x)$, so that $Q_{g_1} > Q_{g_2}$ then potential precision of estimates (found by image processing) of any physical parameter related to the object is higher for $g_1(x)$ than for $g_2(x)$.]

For strip widths 4 and 5, the first two values of the 4^{th} column are from 61 to 73%. This is worse than in above experiments but, as the author discovered, not too bad yet and still acceptable for practice. Indeed, while preparing this article, the software package used for Fig.1 was revised and some bugs in that version were found. Then, it has been found that with those bugs and without noise, the 3FM has a similar efficiency of 70%. Nevertheless, this does not present any obstacle for obtaining the results related to Fig.1.

The gradual deterioration of the 3FM for higher frequencies of the strips can be explained by the border effects: as the rectangle window is rotated, a new portion of image enters inside the window and a portion of the window goes out. This violates some implicit assumptions in the justification of the 3FM presented in Section 2. The higher the spatial frequency, the stronger this violation.

The same experiments are undertaken over an image of a constant brightness. In such a case the estimate of direction is meaningless. The relevant statistic is presented by the number of phantoms (a phantom is an instance of a false positive detector response). The results are presented at the bottom of Fig.3. The right image on the top of Fig. 3 presents an instance of a constant brightness image corrupted by noise of the highest level. It is not surprising that about 17% of phantoms were found for such images. The number of phantoms decreases to 1% at the noise level 0.4 and is 0% for a lower noise level.

Note that while tracing long 1D objects, the phantoms detected with a low probability, say <10% , can be easily filtered out: rare single phantoms hardly would generate a long false curve because several successive phantoms have close directions with a negligible probability. So, even phantoms with frequency 17%, as in the last paragraph, do not present a serious obstacle for tracing long objects.

7. Conclusion

In contrast to conventional edge detectors that work with a higher precision of localization, the localization of an output of the presented Detector of Shift Invariant Locations (DSIL) is imprecise, but it yields very low noise and is robust. On the contrary, the directional component of DSIL is of a much higher precision than the one of edge detectors. DSIL is slower than edge detectors, but leads to a simple and fast curve tracing.

DSIL erases any information about the local structure of a detected element: a positive response may occur as for a single as well as for a double line; the transversal behavior at a detected line may be arbitrary, for example, step-function, delta-wise function, etc. This property is very useful for tracing extended 1D objects "at once", without study of irrelevant details of a local object structure.

Specific properties of DSIL open a wide range of possibilities for development of novel intermediate-level processing algorithms. Instead of image-to-image processing, which is typical for edge detectors, DSIL leads to simple conversion of an input image into some discrete data structures. The inverse transformation, from the discrete data structures to a synthetic image, was studied as well. The experiments show that the original images have a strong resemblance with the synthetic images; hence a significant amount of semantically relevant information is hidden in the data structures.

Just one algorithm of the extraction of semantically relevant information from the data structures was studied experimentally and applied successfully to a well known computer vision problem: an instance of the correspondence problem, (Khachaturov and Moncayo-Muños, 2004).

It is a challenge to study furthermore the algorithms of construction and use of such data structures.

The author believes that a structural decomposition of image constructed by DSIL is useful for a wide range of computer vision applications such as: spatial analysis and reconstruction of 3D scenes with multi-layer overlapping, automatic construction of singular points for meshes used in computer representation of articulated objects, etc.

For instance, part-based recognition (Ullman, 1997) is still a challenge for development of various techniques for shape decomposition – the approach by (Pan et al, 2009) is a single recent example that can be mentioned here. The data structures constructed by DSIL give a quite new tool for this kind of problems.

Acknowledgements

The 3FM was discovered by the author in a process of experimental study of functional Ω_K. Anna Sotnichenko and Oleg Tsyganov helped to develop a user interface for that study. Then the author proposed Theorem as a hypothesis that would explain the experiments. Victor Zalgaller proposed the idea of proof of Theorem in a private discussion with the author.

The experimental study of the presented approach was done by means of a software package available on-line*. Its graphical user interface was developed by Rafael Moncayo-Muños.

This work was partially supported by CONACYT, Mexico, under the grant 400200-5-34812-A.

References

Acton, S.T., 2005, Diffusion partial differential equations for edge detection, In: *Handbook of Image and Video Processing*, 2nd ed., Al Bovik, ed. San Diego, CA: Academic Press, Sect. 4.14, pp. 555-571 (book chapter).

Canny, J., 1986. A Computational Approach To Edge Detection, IEEE Trans. Pattern Analysis and Machine Intelligence, 8:679-714.

Johnson, R.A. , 1997. *Miller and Freund's Probablitiy and Statistics For Engineers*, 5th Ed., Prentice-Hall.

* http://newton.uam.mx/xgeorge/RESEARCH_SOFTWARE_FOR_EXPERIMENTS/SHIFT_INVARIANCE_DETECTOR/

Kass, M., Witkin, A., Terzopoulos, D., 1987. Snakes: Active contour models, *International Journal of Computer Vision*, 1(4), , 321–331. Marr Prize Special Issue.

Khachaturov, G. , 1995. An Approach to Detection of Line Elements. In: *Proc. of the Second Asian Conference on Computer Vision (ACCV'95),* Vol. III, pp.559-563, Singapore.

Khachaturov, G., 1999. Non-uniformity of a pattern and "the best" single view 3D pose estimator, *Computación y Sistemas* (an international journal in Computer Science, published by CIC IPN, Mexico, ISSN 1405-5546), Vol. II, no.2 and 3, pp. 114-125.

Khachaturov, G.and Moncayo-Muños, R., 2004. Synthetic Image of Multiresolution Sketch Leads to New Features, In: *Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT 2004),* Thessaloniki, Greece. pp. 872-879, IEEE Computer Society, ISBN 0-7695-2223-8.

Lang, S. , 1983. *Undegraduate analysis*. Springer-Verlag.

Mlsna, Ph. A., and Rodriguez, J. J., 2005, Gradient and Laplacian Edge Detection. In: *Handbook of Image and Video Processing*, 2nd ed., Al Bovik, ed. San Diego, CA: Academic Press, Sect. 4.13, pp. 535-553 (book chapter).

Nalwa, V.S., 1993, *A guided tour to computer vision*, Addison-Wesley.

Pan, X., Chen, Q. H., and, Liu, Z., 2009. 3D shape recursive decomposition by Poisson equation, Pattern Recognition Lett. 30, pp. 11–17

Ullman, S., 1997. High Level Vision. MIT Press.

Zhang, R., Tsai, P.-S., Cryer, J., and Shah, M., 1999, Shape from Shading, A Survey, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 21, No. 8, pp. 690-706.