From-edges-to-regions approach and multilayered image processing

Ales V. Michtchenko

Department of Computer Science, Moscow State University

Moscow, Russia

Abstract

Image Processing and Computer Vision has been traditionally considered on a different levels of abstraction. This paper presents a scheme, connecting all levels into a multilayered structure of modules. Each module is responsible for a certain subtask of processing. The information flow between modules corresponds to a different features and hypotheses.

The paper is concentrating on a feature extraction from edges and regions. It proposes methods for extraction of geometry of the scene (positions, orientations and surface properties of 3D objects) and resolving corresponding geometrical ambiguities.

Keywords: *Image Processing, Computer Vision, Edge Classification, Specular Reflection, Geometry, Ambiguity.*

1. INTRODUCTION

1.1 Discontinuities in image processing

Intensity and colour discontinuities were a subject of intense research during the last few years. They were proven to be the most informative parts of an image [1] and were used both as boarders of objects in a region-based image processing and as objects themselves in an edge-based image processing.

In this paper we are trying to develop a combined, from-edges-to-regions approach, using discontinuities as the seed regions, starting points for all subsequent image processing. We are implementing this approach as two modules (feature extractor and feature extrapolator), which are the core of multilayered image processing (see section 2).

1.2 Feature extraction

Big informativity of discontinuities and, hence, an advantage of starting an image processing from them, is due to close interconnection between them and a particular features of the scene.

Geometric properties of the scene, such as depth of objects, can be calculated from occlusions and shadows (see section 3).

Other geometric properties, such as orientations and curvatures of surfaces can be calculated from a position and a form of specular reflectance regions (see section 4).

Photometric properties, such as fraction of surface's specular and diffuse reflection, can be calculated from the degree of specks blurring (see section 4).

Textural properties, such as material, colour, albedo, etc, can be calculated using photometric invariants, such as reflectance ratios of adjacent regions (see [2]).

Therefore, each discontinuity can be associated with a particular feature (depth, material of object, norm to the surface, etc) and can be used as a starting point for extrapolating this feature across the continues regions of an image.

1.3 Paper organisation

The following section 2 describes a multilayered image processing, concentrating on the first step of from-edges-to regions approach: extraction of features, as a result of feature extractor work and its communication with the neighbouring image processing layers.

The sections 3 and 4 concentrate on a geometrical features extraction (3D from 2D) and on resolving the corresponding ambiguities.

2. FROM-EDGES-TO-REGIONS MULTILAYERED IMAGE PROCESSING

2.1 Model extraction and model matching (top-down and down-up information flows)

Fig. 1 shows the principal scheme of implementation of from-edges-to-regions image processing. Each module, communicating with its higher and lower neighbours, creates up-down and down-up information flows.

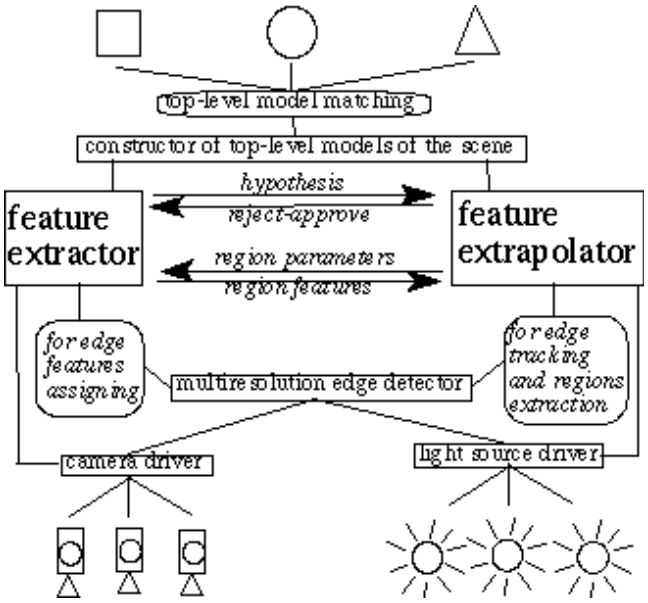

Figure 1: Multilayered image processing

The down-up information flow corresponds to a following sequence of image processing: detection of informative regions (discontinuities), extraction the corresponding features from them, extrapolating these features across the whole image, and, finally, constructing the top-level model of the scene.

The up-down information flow corresponds to a matching higher-level feature models with a lower level image features and finding the best matching model. This is applied to each two neighbouring levels:

Matching a detected edge with a set of edge types (occlusion edge, shadow, boarder between specular and diffuse reflection and boarder between textures and materials);

Matching a region, detected by boarder-tracking, with a set of region types (foreground object, specular reflectance region - speck, region of a particular texture, etc,).

Matching features, extrapolated from the different discontinuities to a feature-model (the depth field, extrapolated from the edges, have to correspond to the norm field, extrapolated from specular reflections, etc.);

Matching the top-level model of the scene with a set of known models (image understanding).

2.2 The second layer: feature extractor

The edge detection techniques have been widely discussed in a number of works (see, for example, [3]). In this work we will start to describe the work of from-edges-to-regions approach and the whole multilayered image processing scheme from the lowest introduced layer, the feature extractor.

The feature extractor receives the edge-related and region-related information from, correspondingly, lower and upper layers, classifies each edge and region, assigns a feature to it and passes this feature up for extrapolation.

In the subsections 2.2.1 and 2.2.2. we will outline edge classifying and feature assigning as a result of feature extractor communication with either edge-detector or feature-extrapolator layers.

In sections 3 and 4 we will discuss the work of this module on extraction the geometry of the scene from edges and specular reflection regions.

2.2.1 Classification of discontinuities

We will classify each edge in accordance with how it can be distinguished from the other edges:

Occlusion (the position and appearance of these discontinuities depend on position of a camera)

Shadow (the position and form of these discontinuities depend on position of a light source)

Specular reflection and other deviations from the Lambertian law (these discontinuities change their position and form with both camera and/or light-source position)

Textural and material discontinuities (they do not depend on neither camera, nor light source position and are detected totally by analysis of different colours intensities - by applying colour-filters to the camera)

The edge classifying can be resolved on a higher or on a lower layers: On a higher level, as a guess-reject/guess-approve communication between the feature-extrapolator and the feature-extractor. On a lower level, as sending commands to hardware-drivers to change positions of camera and/or light source.

2.2.2 Handling ambiguities while extracting the feature from the detected discontinuity.

For geometrical features, discussed in sections 3 and 4, the common ambiguities will be: Ambiguity of objects depth (if the texture is repetitive or if the edge is parallel to epipolar line); the norm-depth ambiguity (the position of speck on the image plane will remain the same, if simultaneously move the object away from the camera and decrease the angle between the norm and a view-direction).

These ambiguities, again, can be resolved on a higher or on a lower layers: On a higher level, as a communication with feature extrapolator (matching features, extrapolated from the different discontinuities to a feature-model with subsequent rejection or approval of the feature model - see 2.1). On a lower level, as communication with the

hardware-drivers (commands for changing the position of camera and/or light source - see sections 3 and 4).

3. EXTRACTION OF DEPTH FROM OCCLUSIONS; DEPTH AMBIGUITIES

The good example of features, extracted on a edge-level of processing is extraction of objects depth from their edges.

All the edge-related information is received by feature extractor from the edge detector. Detection of type (see 2.2.1) and resolving depth ambiguities (see below) are performed in a process of communication with the neighbouring layers.

3.1 Depth of foreground and foreground-background depth difference

There are two types of depth, that can be extracted from each occlusion edge: Depth of occluding (foreground) object and the depth difference between occluding (foreground) and occluded (background) objects. Traditionally most of research activity was concentrated on the depth of occluding object. The major reason for this is simplicity: the depth calculation requires the only one operation (division over the difference of edge position in two dual binocular images).

Human brain, probably, uses both techniques, since we are able to estimate both depths: Depth of foreground (estimating the difference of edge position in left and right eye images) and foreground-background depth difference (estimating the difference of edge position in respect to background texture).

3.2 Ambiguities of foreground depth estimation

It is impossible to find the position difference for edges, parallel to the epipolar line. Moreover, it can be shown, that a deviation of estimated depth and true depth grows then the angle between the epipolar line and the edge decreases.

In accordance with our approach to resolving ambiguities by feature extractor communication with lower or upper module, this can be handled in two possible ways:

Guess-reject/guess-approve communication with feature extrapolator (there was already a number of works on this subject; the most widely used extrapolation is edge-tracking, so that the feature-model is contours of edges).

Sending a position-change commands to a camera controlling driver (see 3.4).

3.3 Ambiguities in estimation of foreground-background depth difference

Since this depth estimation is based on a difference of occluding edge position on a background texture, it is ambiguous, if background texture patterns are not unique.

3.3.1 The texture repetition period along a epipolar vector

For simplicity, let us consider the case of rectangular repetition pattern with sides D_X and D_Y. Choosing a coordinate system with axes, parallel to these sides, the period of texture repetition $D(\vec{E})$ along arbitrary epipolar vector $\vec{E}$ will be equal to:

$$D(\vec{E}) = \frac{D_Y}{\sin(a\tan\frac{E_Y}{E_X})} \quad \text{if} \quad \frac{E_X}{E_Y} \le \frac{D_X}{D_Y}$$

$$D(\vec{E}) = \frac{D_X}{\cos(a\tan\frac{E_Y}{E_X})} \quad \text{if} \quad \frac{E_X}{E_Y} > \frac{D_X}{D_Y}$$

$D(\vec{E}) = D_Y$ if $E_X = 0$

If the repetition pattern has more complicated shape (if there are texture repetitions inside the rectangular), it may be necessary to triangulate it or to choose another type of coordinate system (affine, etc).

3.3.2 Occlusion difference sequence and depth difference sequence, corresponding to a given texture difference

As soon as we know a period of texture repetition, we can represent the set of all possible occlusion differences $\{D_O\}$ as an arithmetic sequence:

$D_O^i = D_{MIN} + i \cdot D(\vec{E})$; $i = \overline{0,\infty}$; where D_{MIN} is the distance between nearest texture patterns, corresponding to 2 occlusions.

In the simplest case, then the view directions of dual binocular images are parallel, the foreground-background depth difference D_{F-B} is connected to occlusion difference D_O as follows: $D_{F-B} = \frac{D_O \cdot D_F}{D_{CAM}}$, where D_{CAM} is the distance between two camera positions, D_F is the depth of foreground object.

If the dual view-directions are not parallel, it is possible to change two corresponding images, as they would be seen from a parallel view-directions (image rectification).

Therefore, the corresponding sequence of possible depth differences $\{ D^i_{F-B} \}$ will be:

$$D^i_{F-B} = D_F \cdot \frac{D_{MIN} + i \cdot D(\vec{E})}{D_{CAM}};\ i = \overline{0,\infty}$$

It is clear, that the less period $D(\vec{E})$ along a particular epipolar vector $\vec{E}$, the closer possible depth differences are to each other. In the limit case of monochromatic texture $D(\vec{E})$ = 0 along $\vec{E}$, the corresponding set of possible depth differences D_{F-B} degenerates to $R_{\geq 0}$ (set of non-negative real numbers).

The most advantageous cameras configuration, on the contrary, are the ones providing the maximum $D(\vec{E})$, namely, with camera position difference parallel to the diagonal of rectangular:

$\vec{E}$ || (D_X, D_Y) or $\vec{E}$ || $(-D_X, D_Y)$.

3.3.3 Multiple views and sparsing the depth difference sequence.

Consider now $N+1$ views on the same occlusion edge. As it was shown, each occlusion difference D_O can be associated with a depth difference D_{F-B}. Since this relation is linear, only N linear independent occlusion differences has to be chosen from the total number of $\frac{N \cdot (N+1)}{2}$ occlusion differences.

Thus, this occlusion edge will be represented by

N sequences of occlusion difference $D^{i,j}_O;\ i = \overline{0,\infty}$ and N camera position differences D^j_{CAM} with

corresponding N epipolar vectors $\vec{E}_j$; $j = \overline{1,N}$

The most probable occlusion differences will be, obviously, the ones, corresponding to the same depth difference.

The feature extractor will scan all the N sequences, finding the first term in each, corresponding to the same depth difference.

Suppose, that the feature extractor found these N terms to have the cardinal numbers i_1, i_2,…, i_N (first terms from each sequence), and all of them correspond to a same depth difference D^1_{F-B}:

$$D^1_{F-B} = D_F \cdot \frac{D^1_{MIN} + i_1 \cdot D(\vec{E}_1)}{D^1_{CAM}} =$$

$$D_F \cdot \frac{D^2_{MIN} + i_2 \cdot D(\vec{E}_2)}{D^2_{CAM}} = \ldots$$

That makes D^1_{F-B} the first possible depth difference (the first term in the depth difference sequence).

After finding the first D^1_{F-B}, each occlusion difference sequence is truncated (so that j_{th} occlusion sequence $\{ D^{i,j}_O \}$, $j = \overline{1,N}$ will begin from the number i_j), and operation of finding N occlusion differences, corresponding to a same depth difference D^2_{F-B} is repeated again.

Note, that if a period of texture repetition $D(\vec{E}_j)$ along one of the epipolar vectors $\vec{E}_j$ is zero, the corresponding occlusion sequence $\{ D^{i,j}_O \}$ is neglected in accordance with 3.3.2.

On the contrary, if a period of texture repetition $D(\vec{E}_j)$ is infinite (the texture is unique along $\vec{E}_j$) there will be the only one term in the corresponding occlusion sequence $\{ D^{i,j}_O \}$, and hence, the only one possible depth difference: $\{ D^i_{F-B} \} = D_{F-B}$.

3.4 The overview of feature extractor work and its communication with hardware drivers while extraction of depth from occlusions.

Summarising the all discussed in this section, the most advantageous camera positions are:

$\vec{E}$ || $\overrightarrow{(D_X, D_Y)}$ and $\vec{E}$ || $\overrightarrow{(-D_X, D_Y)}$

(for foreground-background depth estimation);

$\vec{E} \perp$ *occlusion edge* (for foreground depth estimation)

In accordance with this, we can define a weight $\omega_{type}(\vec{E})$, proportional to an angle between $\vec{E}$ and diagonals (D_X, D_Y), $(-D_X, D_Y)$ (for foreground-background type of depth) or to an angle between $\vec{E}$ and a

norm $\vec{n}$ to a occlusion edge (for foreground type of depth): $\omega_{F-B}(\vec{E})$ =

$$= \alpha \,|\, \vec{E} \wedge \overrightarrow{(D_X, D_Y)} \,| + (1-\alpha) \,|\, \vec{E} \wedge \overrightarrow{(-D_X, D_Y)} \,|,$$

$$\omega_F(\vec{E}) = \beta \,|\, \vec{E} \wedge \vec{n} \,|.$$

The work of the feature extractor module and its communication with the hardware-driver, controlling the camera, will be as follows:

From a given pair of views (or, more precisely, from a two sets of edges, detected by the lower module), it assigns the depth of foreground object and foreground-background depth difference to each edge.

If, for a certain subset of edges, a mistake in foreground depth estimation is above a certain threshold, or, if, for another subset of edges, a difference between the neighbouring terms in a depth difference sequence is above a certain threshold, it sends to the camera-driver the command to change the view position (in fact, to add a new view). The new camera position is chosen to minimise the weighted sum over these two subsets of ambiguous edges:

$$\vec{E} = \arg\{\min(\sum_{\substack{edges\_with\_undefinite \\ foreground\_depth}} \omega_F(\vec{E}) + \sum_{\substack{edges\_with\_undefinite \\ foreground-background \\ depth}} \omega_{F-B}(\vec{E}))$$

From a new, trinocular, set of views, it improves the binocular estimation of depth of foreground object and foreground-background depth difference.

If this, corrected estimation is still unacceptable, it adds the 4th best possible view and repeats calculation.

4. EXTRACTION OF SURFACES ORIENTATIONS AND CURVATURES FROM THE SPECULAR REFLECTION REGIONS; NORM-DEPTH AMBIGUITY

The good example of features, extracted on a region-level of processing is extraction of surfaces norm and curvatures from a position and form of specular reflectance region.

4.1 Communication between the feature extractor and feature extrapolator for assigning a region-level features.

During the extraction of edge-level features, all the edge parameters were received by feature extractor directly from the edge detector. Region parameters, necessary for region-level features extraction require the following pre-processing actions:

On a edge-level of processing, a corresponding edge has been classified (in our case - as an edge between specular and diffuse reflectance regions by a change-position communication between feature extractor and camera/light-source drivers).

Feature extrapolator tracked this edge and defined a boarders of the specular reflectance region. Finally, it passes the parameters of this region (coordinates of its centre and dimensions) back to the feature extractor.

Feature extractor will calculate the norm and curvatures and pass these features back to the feature extrapolator to be extrapolated across the diffuse reflectance regions.

4.2 The equations for extracting geometrical and photometrical properties of a surface element

Let us consider first a surface with a zero curvature (plane) and choose the following coordinate system (see fig.2): origin is at the centre of the speck, $\overrightarrow{OZ} = \vec{n}$ (norm to a plane), $\overrightarrow{OX}$ is an intersection of a plane, containing the speck with a reflection plane (the plane, containing incidence and reflected rays), $\overrightarrow{OY} = [\overrightarrow{OX} \times \overrightarrow{OZ}]$. Let the coordinates of the light source and camera be $\vec{S} = (S_X, 0, S_Z)$ and $\overrightarrow{D_O} = (D_X, 0, D_Z)$ correspondingly.

In this coordinate system, the most bright point is the origin (x=0, y=0) and the brightness fades with the growth of $|x| + |y|$.

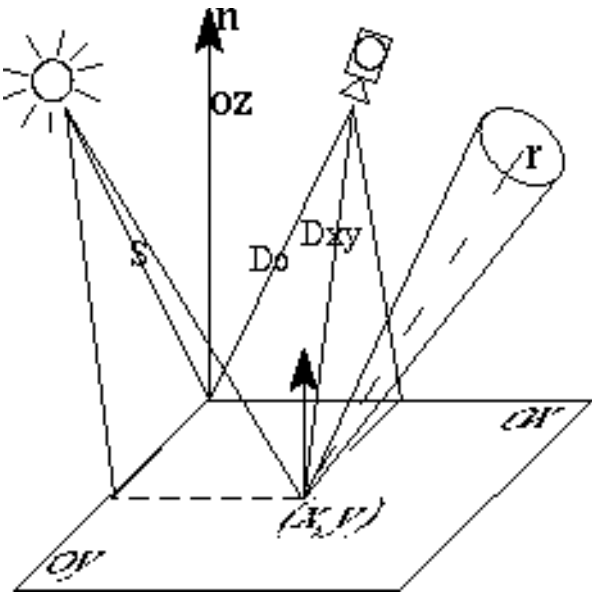

Figure 2: Geometry of specular reflection

The ray ($x - S_X$, y, $-S_Z$), hitting the plane at the point (x, y) is reflected to a solid angle (cone), centred around specular reflection direction $\vec{r}$=($x - S_X$, y, S_Z). If a camera position $\overrightarrow{D_O}$ fall inside this specular reflection cone, the point (x, y) is seen inside the speck. The bigger the angle between reflection to camera $\overrightarrow{D_{X,Y}}$ = ($D_X - x$, $-y$, D_Z) and the specular reflection axis $\vec{r}$, the less bright this point is seen.

In case of plane ($\vec{n}(x, y)$ = $\vec{n}(0,0)$ = $\overrightarrow{OZ}$), the specular reflectance radiance δI will be a linear function of

$$\cos(\overrightarrow{D_{X,Y}} \wedge \vec{r}) = \frac{\overrightarrow{D_{X,Y}} \cdot \vec{r}}{|\overrightarrow{D_{X,Y}}| \cdot |\vec{r}|} = C_O$$ (for details see below).

In a general case, $\vec{n}(x, y) \neq \vec{n}(0,0) = \overrightarrow{OZ}$.

Assuming the small size of the speck, let us decompose the norm difference in a Taylor sequence: $\vec{n}(x, y) - \vec{n}(0,0) = \overrightarrow{\Delta n}(x, y) = \frac{\partial \vec{n}}{\partial x} x + \frac{\partial \vec{n}}{\partial y} y$, so that

$\overrightarrow{\Delta r} = 2 \cdot |\vec{r}| (\frac{\partial \vec{n}}{\partial x} x + \frac{\partial \vec{n}}{\partial y} y)$ and

$$C_O \mapsto C = \frac{\overrightarrow{D_{X,Y}} \cdot (\vec{r} + \overrightarrow{\Delta r})}{|\overrightarrow{D_{X,Y}}| \cdot |\vec{r} + \Delta r|}.$$

In the case of a small speck and a smooth surface $|\frac{\partial \vec{n}}{\partial x}| + |\frac{\partial \vec{n}}{\partial y}| << |\vec{r}|$ the last equation can be written as

$$C = C_O + \Delta C = C_O + \frac{\overrightarrow{\Delta r}}{|\vec{r}|} \cdot \frac{\overrightarrow{D_{X,Y}}}{|D_{X,Y}|}$$

Assuming, that total radiance I of the surface consist from the diffusive (Lambertian) part I_O =const and the specular reflection part $\delta I = \delta I(\overrightarrow{D_{X,Y}} \wedge \vec{r})$, and assuming the Gaussian shape of the specular reflection peak, we arrive to the following connection between the coordinates (x, y) and the observed brightness $I(x, y)$ of the corresponding point:

$$I(x, y) = I_O + \delta I_{MAX} \cdot \exp\{-\frac{(\overrightarrow{D_{X,Y}} \wedge \vec{r})^2}{2 \cdot \sigma^2}\}.$$

Decomposing an exponent into a Taylor sequence and using an expression for C,

$$I(x, y) = I_O + \delta I_{MAX} \cdot \frac{1}{\sigma^2}(\sigma^2 - 1 + C)$$

This equation connects geometrical (C) and photometrical (I_O, δI_{MAX}, σ) properties of the surface (see also [4]).

4.3 Variables in radiance equation and a norm-depth ambiguity

If both camera $\overrightarrow{D_O}$ and light source $\vec{S}$ positions are known, then, in order to solve the radiance equation for 3 photometric (I_O, δI_{MAX}, σ) and 4 geometric ($\frac{\partial \vec{n}}{\partial x}$, $\frac{\partial \vec{n}}{\partial y}$) variables, it is necessary to write it for 7 points (x, y).

It is necessary to mention, that radiance equation can be resolved only if at least two quantities from ($\vec{n}(0,0)$, $\overrightarrow{D_O}$, $\vec{S}$) are known.

It is easy to show, that all the geometry, shown on fig. 2 will remain the same, if moving the origin of coordinate system away from the camera $\overrightarrow{D_O} \mapsto \overrightarrow{D_O} + \overrightarrow{\delta D}$ and rotating the coordinate system around $\overrightarrow{OY}$ axis so that $\overrightarrow{OZ} \,||\, \overrightarrow{D_O} + \vec{S} \mapsto \overrightarrow{OZ}' \,||\, \overrightarrow{OZ} + 2 \cdot \overrightarrow{\delta D}$.

This ambiguity (any vector $\overrightarrow{\delta D} \,||\, \overrightarrow{D_O}$, added to distance to a camera, and simultaneously twice to a norm of the surface, does not change the position of a speck) is a typical 3D$\mapsto$2D ambiguity and can not be resolved locally.

Similarly, if only $\vec{n}(0,0) = \overrightarrow{OZ}$ is known, it possible to move the origin along $\overrightarrow{OZ}$ without violation of reflection laws. And again, if only $\overrightarrow{D_O}$ is known, an angle of rotation around $\overrightarrow{OY}$ will be undefined.

Therefore, if quantities ($\vec{n}(0,0)$, $\overrightarrow{D_O}$, $\vec{S}$) are not known, the feature extractor resolves the radiance equation with

accuracy limited by a corresponding ambiguity and passes these results to the feature extrapolator. It then extrapolates all photometric and geometric variables, containing this ambiguity across continues regions of an image, and resolves this ambiguity then another discontinuity is met, for example, then it meets an edge with assigned depth.

5. CONCLUSIONS AND FUTURE WORK

In this paper we divided the Image Processing problem into a several subtasks, corresponding to a different modules in a multilayered processing scheme.

The most of investigation was directed to extraction of geometrical features of the scene (feature extractor module) and to a communication between modules.

The future research will outline a work of other relevant modules with subsequent implementation of the entire Image Processing system.

6. REFERENCES

[1] D.Geiger, A.Rudra, L.Maloney. *Features as Suffisient Statistics.* (Adv. In Neural Information Processing sys.10, 1997)

[2] Shree K. Nayar, Rund M. Bolle. *Computing Reflectance Ratios from an Image.* Department of C.S., Columbia University, New York, NY 10027

[3] H.F Hueckel. *An operator, which locates edges in digitized pictures.* J. Assoc.Comp.Mach. Vol 18, pp 113-125, 1971

[4] Ales Michtchenko *Uniform multiresolution light fields.* GraphiCon 1999, Moscow Aug.-Sept. 1999

Author:

Ales V. Michtchenko, PhD student of Moscow State University, Department of Applied Mathematics and Computer Science.

Address: Russia, Moscow 117292, Profsouznaya, 8-2-349
E-mail: ales.michtchenko@usa.net