Information-theoretic approach to image description and interpretation

Alexey S. Potapov, Vadim R. Lutsiv

Center for Image Processing of the Vavilov State Optical Institute, Birzhevaya line 12, 199034, St. Petersburg, Russia
E-mail: vist@rbcmail.ru

ABSTRACT

We present an information-theoretic approach to the image interpretation problems. In the context of this approach such tasks as contour extracting, constructing the most informative image features and image matching are described as a single unified problem. Our approach is based primarily on the interpretation of the image (or image set) representation problem as a Minimum Description Length (MDL) task. The images matching turns out to be a generally adopted method of images alignment by maximization of their mutual information. However, instead of using the pixels intensities themselves a more condensed data representation form can be used to reduce the dimensionality of input data and to extract the invariant information: hierarchical image structural description. Though we developed and successfully applied the information-theoretic approach for the images matching, it can be extended to the other tasks, e.g. the changes detection.

Keywords: image interpretation, image matching, minimum description length.

1. INTRODUCTION

Computer analysis of images is an important branch of computer science. An image represents a series of measurements, performed on some set of physical objects called “the scene”. The images contain information about certain measured quantities (generally, the intensity of reflected light) and spatial arrangement of scene. The whole volume of information is presented in the raw images, but not in a useful form. Making the image contents explicit is one of the major problems in computer vision [4]. This is the problem of image description and interpretation.

Different levels of image description exist, ranging from the raw image data to a symbolic description using a domain dependent alphabet [4, 5]. Many particular approaches to image representation were developed, which made use of a priori restrictions typical for the application at hand. A number of generally applicable image analysis techniques were also proposed [4, 6, 7]. However, the theoretical approaches are usually restricted to the low-level image representations, and the higher level (structural and symbolic) descriptions of images are obtained by various heuristic methods. Thus, development of a single overall theory of image interpretation is far from completion. Such the theory is necessary for constructing a general purpose visual system, comparable in its abilities with that of humans.

Here the construction of image description is handled in the context of image matching problem. This problem consists in obtaining such spatial transformation, which maps one image to another in such the way, that the corresponding points of the images coincide [1]. There are a number of reasons for images being under alignment to be different: the changes in illumination, the season and day-time variations, the differences in sensor type, and so on. As a result, extraction of reliable information from the images is a supreme problem in the task of image matching. That’s why the quality and robustness of matching methods can be considered as the performance criterions of the image description approaches used in these methods.

There are many image matching techniques based on different levels of image representation [1, 5]. However, all these methods deal separately with the task of image alignment and the task of description of images being aligned. Here we interpret image matching as the Minimum Description Length (MDL) problem [2] for a pair of images. In the context of such interpretation the hierarchical matching techniques and the methods of alignment by maximization of the mutual information are united. In this paper an interdependence between different tasks of image analysis is established and the necessity of simultaneous solutions is shown. The image description possessing a minimum length is shown to be optimal for a number of particular cases. The software for image matching was constructed and applied using the proposed model with certain simplifying assumptions.

2. IMAGE MATCHING AS MDL-PROBLEM

Two images are given in the image matching problem: $u: G_1 \to \Re$ and $v: G_2 \to \Re$, where the images are defined in the regions $G_1, G_2 \subset \Re^d$, d is the number of dimensions of the images. It is necessary to find the spatial

transformation $T: G_1 \to G_2$, which transforms one image with respect to the other one so that the corresponding points in the two images coincide.

We consider the image matching problem as a task of construction of a description of minimum length for two images simultaneously. Such the approach permits to generalize the methods of image matching by maximization of mutual information.

The mutual information based matching algorithms have been accepted as one of the most accurate and robust methods [3], which have, however, rather narrow application range. In these methods an estimate of the transformation $\overline{T}$ is sought for, that matches the reference image u and test image v by maximizing their mutual information [6],

$$\overline{T} = \arg\max_T i(u(x), v(T(x))) \ . \quad (1)$$

The mutual information i of two registered images u and v is defined in terms of entropy H in the following way:

$$i(u(x), v(T(x))) = H(u(x)) + H(v(T(x))) - H(u(x), v(T(x))) . \quad (2)$$

In the discrete case the entropy of a random variable is defined as $H(x) = -\sum_x p(x) \log_2 p(x)$, and the joint entropy of two random variables x and y is $H(x, y) = -\sum_x \sum_y p(x, y) \log_2 p(x, y)$.

The MDL-principle can be formulated in the following way: choosing the model that gives the shortest description of data [2]. Let's show, that applying this principle to the tasks of image matching one can obtain the mutual information based methods as the particular cases. Applying MDL-principle to the matching problem one can obtain the best spatial transformation

$$\overline{T} = \arg\min_T \left[L(u, v \mid T) \right], \quad (3)$$

where $L(u, v \mid T)$ is the length of common description of two images registered together using the spatial transformation T.

Let's consider the following particular case. Suppose that the intensity values of individual pixels are the statistically independent samples of a certain random variable. In accordance with the Shannon's source coding theorem [2] the description length of N samples of a random variable x in an average sense is $L(x^N) = N \cdot H(x)$, thus $L(u, v \circ T) = N \cdot H(u(x), v(T(x)))$, where N is the number of pixels in the reference image. Recall the equation (2) to obtain:

$$H(u(x), v(T(x))) = H(u(x)) + H(v(T(x))) - i(u(x), v(T(x))) \Rightarrow$$
$$L(u, v \circ T) = N \cdot H(u(x), v(T(x))) = N \cdot H(u(x)) + N \cdot H(v(T(x))) - N \cdot i(u(x), v(T(x))) \Rightarrow$$
$$L(u, v \circ T) = L(u) + L(v \circ T) - I(u, v \circ T) ,$$

where $I(u, v \circ T) = N \cdot i(u(x), v(T(x)))$ is the amount of common information in the reference image and transformed target image. For symmetry one can write $L(u, v \mid T) = L(u) + L(v) + L(T) - I(u, v \mid T)$. The term $L(T)$ stands for the description length of the spatial transformation, which, however, is constant for the same model of spatial transformation. So maximizing the mutual information one achieves the minimization of description length.

Some assumptions about the image properties were made to get this result. The result shows the equivalence of the methods based on the mutual information and the ones based on the MDL-principle. That is, we applied a certain image model in an explicit form. As far as another image model can be used, the matching method based on the MDL-principle appears to be more general, than the approach based on maximization of the mutual information, and the former one includes the latter one as the particular case.

The approach of matching of images by maximization of their mutual information was extended in [3] to the usage of feature vectors (or tie points) instead of the initial intensity values as it was done in [6]. Thus, the image representations by arbitrary image features ranging from the intensity values to the symbolic descriptions can be used in the matching methods based on maximization of mutual information. One can reformulate these methods in terms of MDL-problem in the same way as it was done above. To do this one should only introduce a corresponding image model.

If no explicit algorithm is present which converts an image into its model, then an optimal model belonging to a given class Θ can be also sought for, using the MDL criterion: $\overline{\theta} = \arg\min_{\theta \in \Theta} \left[L(u \mid \theta) \right]$, where θ is the set of the model

parameters. As far as we deal with image matching, such a model is necessary which can be applied to a pair of registered images: $\overline{\theta} = \arg\min_{\theta\in\Theta}\left[L(u, v\circ T \mid \theta)\right]$. Then the equation (3) is specified as:

$$\overline{T} = \arg\min_{T}\left[\min_{\theta\in\Theta}\left[L(u, v\circ T \mid \theta)\right]\right]. \qquad (4)$$

The equation (4) can be reformulated if the description length of the registered images is expressed in the terms of the description lengths of the images taken separately and their mutual information. The tasks of image description and image matching are then united into the common problem:

$$\{\overline{\theta}_1, \overline{\theta}_2, \overline{T}\} = \arg\min_{\theta_1,\theta_2,T}\left[L(u \mid \theta_1) + L(v \mid \theta_2) + L(T) - I(u, v \mid \theta_1, \theta_2, T)\right]. \qquad (4')$$

As it is evident from the equation (4), the models are to be constructed for a pair of registered images using every spatial transformation, and the best transformation is to be chosen, on the base of which the model with the description possessing a minimum length can be built. That is, the tasks of image description and image matching are to be solved simultaneously. Therefore the unification of these tasks in the equation (4') is correct. In other words, the image model, which is optimal for a single image, could be nonoptimal in the case of joint description of a pair of images. The reason originates from a discrepancy of the descriptions, which will be discussed below (see, for example, the figure 2). Thus, the resulting models of the images are to correspond not only to the images, but also to each other. This can be treated as the fusion of data taken from different sources.

To implement this approach, one should choose a certain image model. Of course, the image model is to reflect the properties of the physical objects, which are expected to be present in the scene, and, probably, the properties of the image sensor used.

3. IMAGE MODEL

Solving the MDL-problem for an image depends on the class of image model, i.e. on the parameter space which θ_1 and θ_2 belong to. In the simplest case, an image is represented by initial intensity values as the samples of a random variable. Application of other image models can help to reduce the dimensionality of the data sets to be matched, to reduce the noise, and to extract the invariant information. Many approaches of representation of images by different features were proposed [1], among which the intensity edges, texture segments, various geometric primitives, and others are presented. However, few of theoretical bases exist for using certain types of features in certain image interpretation applications. Here we present a formalized image model in the context of the MDL-principle.

Recall that in the equation (2) an image is represented as a set of samples of a random variable with a single probability density function. However this is not the case for the real images. The image in the Figure 1 is divided into two regions with essentially different probability distributions. Moreover, if one estimates the entropy for these regions and calculates a minimum description length for them and for the initial image, he obtains that the summed description length of the regions is less then that of the whole image. As far as an image represents a set of physical objects, it is naturally to assume that these objects have different properties, so the image $u: G \to \Re$ can be described as a set of regions $\cup G_i = G$ corresponding to the samples of different random variables $u_i(x) = u(x)|_{G_i}$ with different probability density functions $P(u_i(x))$. If this assumption is true, it is possible to prove that for any other division of the image into the same number of regions $\cup g_i' = G$ the following inequality is true:

$$\sum_{i=1}^{K} L_H(u(x)|_{g_i'}) \geq \sum_{i=1}^{K} L_H(u_i(x)), \qquad (5)$$

where L_H is the description length obtained from entropy: $L_H(u|_R) = card(R)\cdot H(u(x)|_R)$, $card(R)$ is the cardinality (i.e. the number of points) of the set R. Here we also assume implicitly, that the objects constituting the scene are opaque, and their surfaces are not mirror-like reflecting. This is correct in the case of aerospace photographs.

Guided by the MDL-principle one can accept optimal regions to be:

$$\{\overline{G_i}\}_{i=1}^{K} = \arg\min_{\{G_i\},K}\left[\sum_{i=1}^{K}\left(L_H(u_i) + L(c_i)\right)\right]. \qquad (6)$$

Not only the content of regions is to be described but also their shapes, so the $L(c_i)$ terms are present, which stand for the descriptions of regions' boundaries. If one assumes, that the regions are simply connected, then regions are uniquely

described by their boundaries. In practice these terms prevent a trivial division, in which each pixel is represented by an individual region.

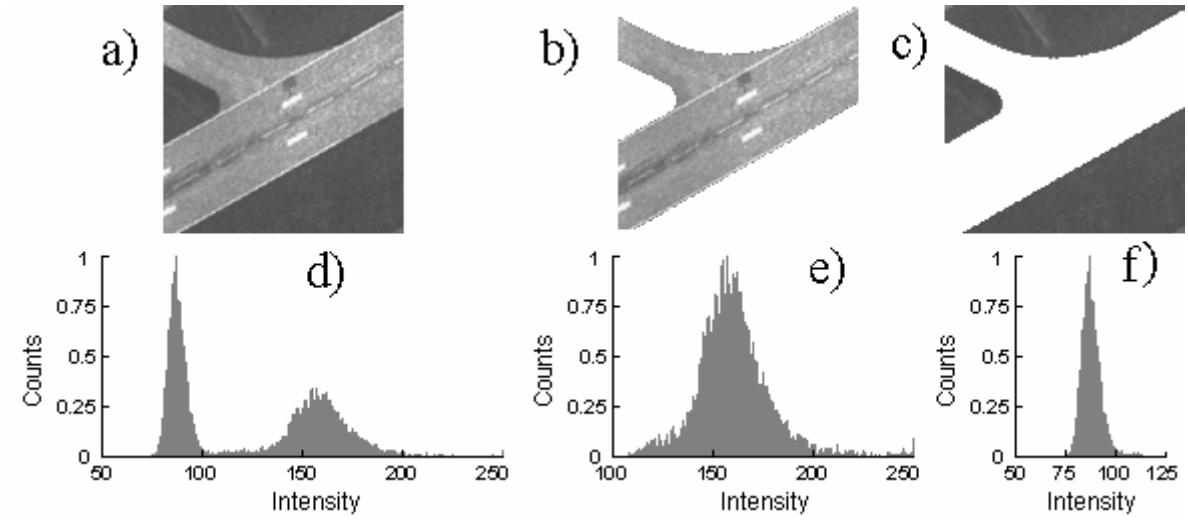

Fig. 1. a) is an initial image, b) and c) are two regions on the image, d) is the histogram of the whole image, e) and f) are the histograms of the regions. The values of entropy of the whole image and of the regions are 4.375, 4.297, and 3.032 respectively. The description lengths are 59845, 30910 and 19671.

The inequality (5) holds if only the samples of $u_i(x)$ are uncorrelated inside the region G_i. This assumption can be invalid in some cases (particularly, in the analysis of 3-D scenes, where the brightness of a surface depends on its orientation, so the non-planar surfaces are subjected to systematic variations of intensity). In a general case an image in each region should fit a model, which describes the nonrandom variations of intensity.

Recall the equation (6) to consider the term, which stands for the boundary description length. A boundary of a simply connected region is a closed curve. Given a point on this curve, the coordinates of every next point can be obtained specifying the direction from the previous point (usually, there are 4, 6, or 8 possible directions). Thus, a curve can be described by the coordinates of initial point and by a set of numbers for all the other points. Assuming these numbers to be the samples of a random variable, their description with the shortest length can be calculated in terms of entropy. As it was done for the images, the curves can also be divided into the regions to minimize the description length. In every region a curve can be approximated by a model (the simplest models are the segments of straight lines or arc of circles). Supposing the deviations from the models to be a noise makes the corners separating the regions on the curves to be the most informative points in images. However, for the curve description the models of its segments are also to be used (especially, if different types of models can be present). Consequently, to construct an image description one should substitute (7) into the equation (6):

$$L(c_i) = \sum_{j=1}^{M_i} \left[L_H(c_i|_{S_{ij}} - m_{ij}^c) + L(m_{ij}^c) + L(c_{ij}^c) \right], \qquad (7)$$

where $c_i|_{S_{ij}}$ is j-th segment of i-th curve, m_{ij}^c is the model of this segment, and c_{ij}^c stands for the boundaries (i.e., the corner points) of this segment. In order to construct a description of an image having a minimum length, one should solve simultaneously the tasks of dividing the image into the regions and of describing the regions' boundaries. These are the tasks of image segmentation (or contour extraction) and structural elements construction.

Now one can apply this image model to describe the registered pair of images and find an optimal spatial transformation according to the equation (4). The description length of the "image" $(u(x), v(T(x)))$ is the cost function of the transformation T, which generalizes the mutual information criterion. It is supposed that the regions of images correspond to the appropriate surfaces of the objects of scene, thus the following strategy of search for the correct spatial transformations is applied. Different combinations of correspondences between the regions of two images are considered. Every such combination roughly defines a spatial transformation which can be improved by searching for the correspondences between the structural elements and contour points. In order to skip solving the complete problem for

every suppositional transformation one can construct a suboptimal description on the base of those of the original images. The descriptions of the initial images $\theta_u^* = \arg\min_{\theta\in\Theta}\left[L(u \mid \theta)\right]$ and $\theta_v^* = \arg\min_{\theta\in\Theta}\left[L(v \mid \theta)\right]$ are to be constructed only once, and then one can correct these descriptions to adjust them to each other for a given hypothesis of spatial transformation.

In order to carry out this procedure the shapes of presumably corresponding regions should be modified in such a way, that the regions would exactly overlap. Division of the contours (borders of the regions) into the segments can also be obtained from those of the initial images. As far as the description is hierarchical, and the cost function is a sum of the cost functions for the separate description levels, the task of adjustment of the entire model can be divided into the subtasks of correction of single-level descriptions united by the common cost function. The information about the way, in which the adjustment is to be carried out, passes from the higher level of the hierarchy to the lower levels. At the same time the changes in the lower level description automatically cause the changes in the higher levels (for example, the contour shifts result in the shifts of structural elements). This strategy has much in common with the adaptive resonance theory [8].

However, the description of images using the model described above is a complex problem that has not been solved yet in full volume, so the simplified models are being used in experiments.

4. EXPERIMENTAL RESULT

The acquired theoretic results were practically implemented in improvement the structural matching algorithm described in [7]. This algorithm uses the straight lines and corners constructed on the base of contours (see fig. 2). Such image description is a good approximation to the image model described above. However it gives no information what region a certain contour belongs to, so the structural elements were divided into groups on the base of their spatial proximity. To achieve the better results the dependence between the guess transformation and image description was introduced in accordance with the equation (4').

For every hypothesis T_i of spatial transformation obtained during the matching process the descriptions of both images are tuned to bring them into correspondence with each other:

$$\{\theta_{i1}(T_i), \theta_{i2}(T_i)\} = \arg\min_{\theta_1,\theta_2}\left[L(u \mid \theta_1) + L(v \mid \theta_2) + L(T_i) - I(u, v \mid \theta_1, \theta_2, T_i)\right]. \quad (8)$$

This adjustment was restricted to the reconstruction of the structural elements and their groups (the contour descriptions were not changed). Then the refined image models were used to obtain a more precise spatial transformation.

The refined algorithm was tested for a number of pairs of real images. The proposed solution of the united task of matching and description allowed not only to obtain the coinciding structural descriptions of two images (see fig. 3), but also to improve the precision and robustness of the matching algorithm (see fig. 4).

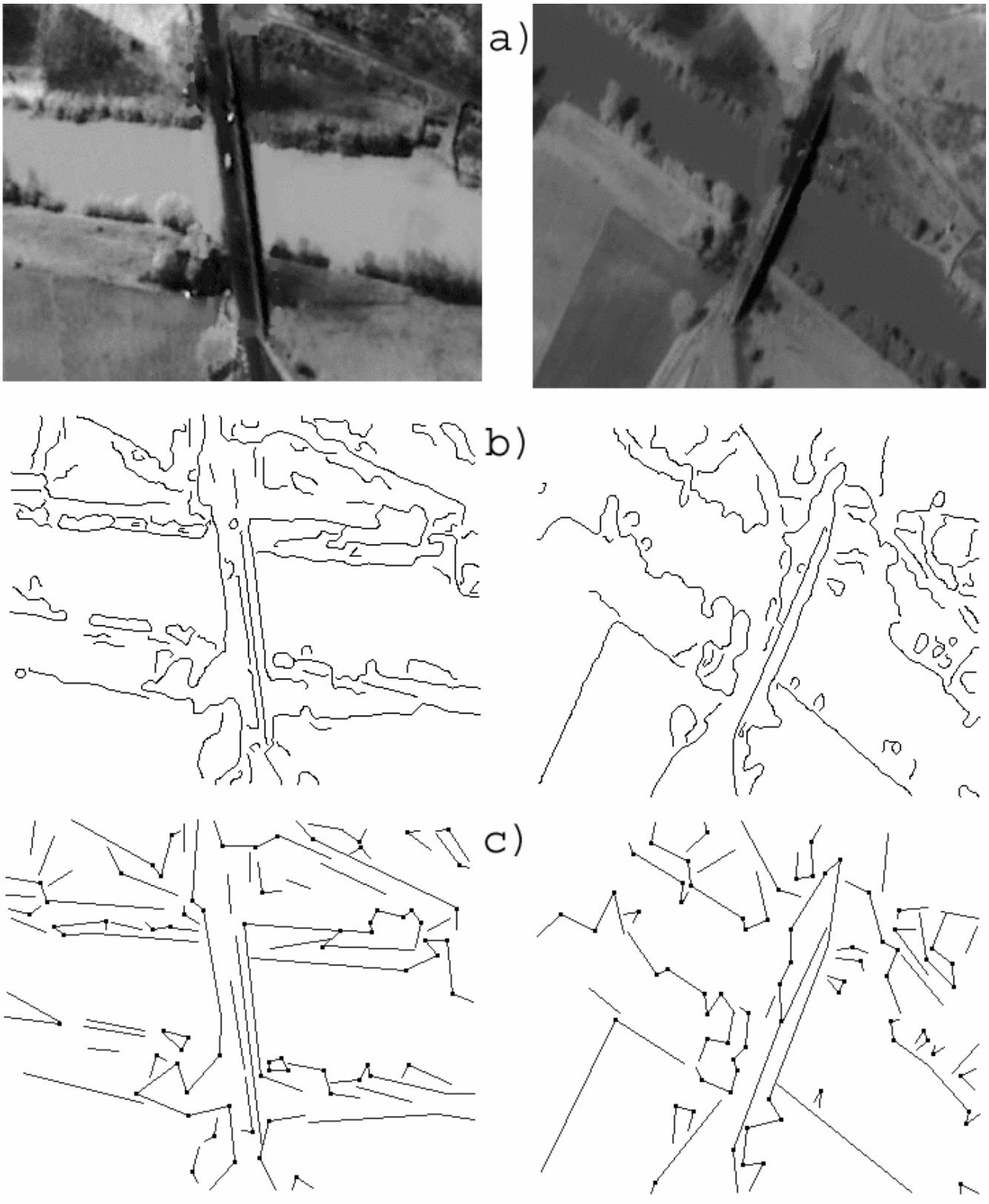

Fig. 2. Construction of the initial description of images: a) a pair of images, b) the initial contour descriptions, c) the initial structural descriptions.

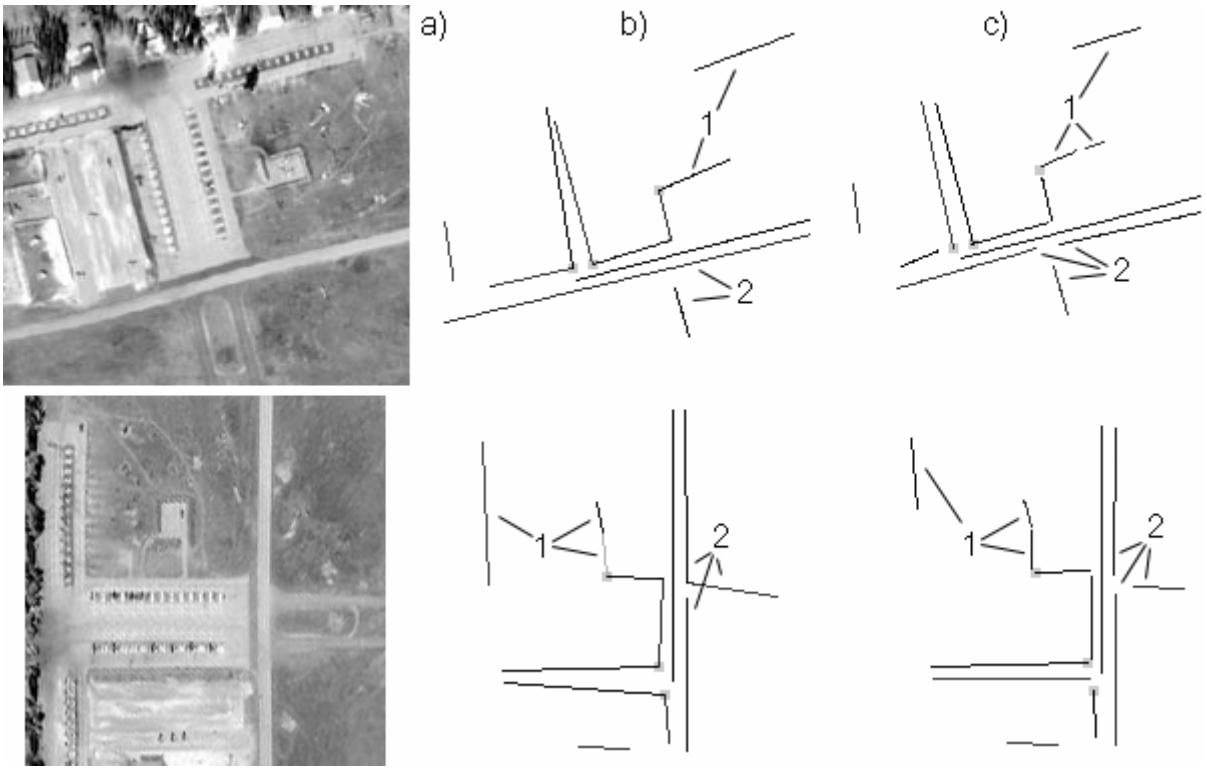

Fig. 3. a) A pair of fragments of initial images, b) some of structural elements in the initial descriptions, c) corrected structural elements corresponding to the true spatial transformation.

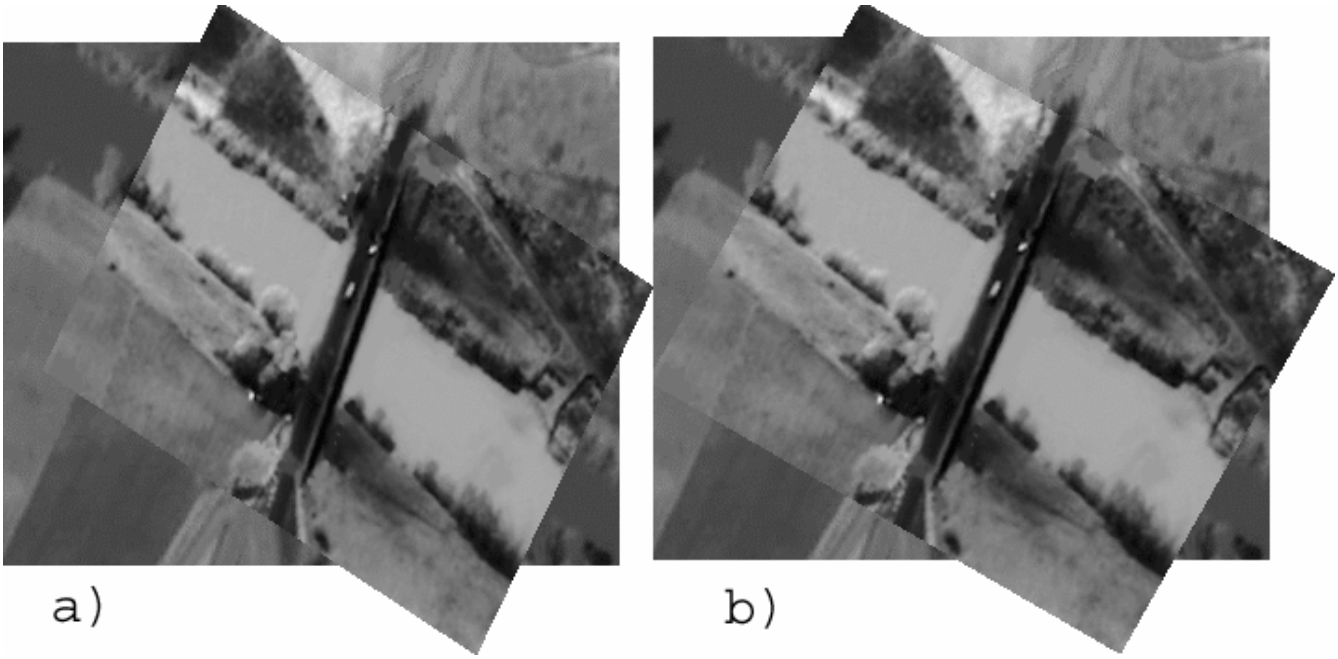

Fig. 4. Result of matching a pair of images presented in fig. 2: a) without model adjustment, b) with model adjustment.

5. CONCLUSIONS

In this article we reformulated the method of images matching by maximization of their mutual information in terms of the MDL-problem which allowed to unite the tasks of image description and image matching tasks. The formalized image model was proposed, which can be used in the context of the MDL approach. However our approach gives no algorithms for constructing the model, but it only explains existing empirically derived approaches. The results of this theoretic study were implemented in the image matching algorithm, which uses an empirical image model similar to the theoretically developed one. The improvement of the matching results practically reached shows the correctness of the proposed approach.

REFERENCES

1. L.G. Brown, “A survey of Image Registration Techniques”, ACM Computing surveys, vol. 24, pp. 325-376, 1992.
2. M. Hansen and B. Yu, "Model Selection and the Principle of Minimum Description Length," Technical Memorandum, Bell Labs, Murray Hill, N.J, 1998.
3. B. Ma and A. Hero, “Image Registration with Minimum Spanning Tree Algorithm”, ICIP’00, Vancouver, Sep. 2000.
4. P. Nacken, “Image Analysis Methods Based on Hierarchies of Graphs and Multi-Scale Mathematical Morphology”, PhD-thesis, University of Amsterdam, 1994.
5. A. Pinz, M. Prantl, and H. Ganster, “A Robust Affine Matching Algorithm Using an Exponentially Decreasing Distance Function”, Journal of Universal Computer Science, vol. 1, no. 8, 1995.
6. W.M. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, “Multi-modal volume registration by maximization of mutual information”, Medical Image Analysis, 1(1):35-51, 1996.
7. V.R. Lutsiv, I.A. Malyshev, V. Pepelka, A.S. Potapov, “Target independent algorithms for description and structural matching of aerospace photographs”, Proc. SPIE, Vol. 4741, pp. 351-362, 2002.
8. G.A. Carpenter and S. Grossberg, “ART-2: self-organization of stable category recognition codes for analog input patterns”, Applied Optics, Vol. 26, No. 23, pp. 4919-4930, 1987.