Automatic structural matching of 3D image data

Svjatoslav Ponomarev[†ab], Vadim Lutsiv*[abc], Igor Malyshev[a]
[a]The Vavilov State Optical Institute, 5-2 Kadetskaya line, Saint Petersburg, 199053, Russia; [b]The State University of Information Technologies, Mechanics, and Optics, 49 Kronverksky avenue, Saint Petersburg, 197101, Russia; [c]The State University of Aerospace Instrumentation, 67 Bolshaya Morskaya street, Saint-Petersburg, 190000, Russia

ABSTRACT

A new image matching technique is described. It is implemented as an object-independent hierarchical structural juxtaposition algorithm based on an alphabet of simple object-independent contour structural elements. The structural matching applied implements an optimized method of walking through a truncated tree of all possible juxtapositions of two sets of structural elements. The algorithm was initially developed for dealing with 2D images such as the aerospace photographs, and it turned out to be sufficiently robust and reliable for matching successfully the pictures of natural landscapes taken in differing seasons from differing aspect angles by differing sensors (the visible optical, IR, and SAR pictures, as well as the depth maps and geographical vector-type maps). At present (in the reported version), the algorithm is enhanced based on additional use of information on third spatial coordinates of observed points of object surfaces. Thus, it is now capable of matching the images of 3D scenes in the tasks of automatic navigation of extremely low flying unmanned vehicles or autonomous terrestrial robots. The basic principles of 3D structural description and matching of images are described, and the examples of image matching are presented.

Keywords: Image matching, 3D structural description, 3D structural matching, automatic navigation

1. INTRODUCTION

The task of automatic analysis of images of natural environment is urgent for several decades, and it becomes ever more important at present. On the one hand, the amount of remote sensing and reconnaissance information transferred to the Earth by modern space and airborne video-sensors increases in avalanche-like mode (especially in the cases of hyperspectral image data), thus a preliminary classification of such data by onboard computer becomes urgent. On the other hand, the onboard image analysis is a base of automatic navigation of unmanned vehicles. In the cases of Earth surface observation from long distances (e.g. from the spacecrafts or high flying aircrafts), the image sensors "look" almost downwards, and the observed surface may be considered as approximately flat. The high-precision long-focus objectives are usually applied in such situations, thus the images of Earth surface projected onto the sensor matrices of photo cameras are subjected to geometric transforms corresponding approximately to the similarity or affine groups, and such transforms are homogeneously applied to the whole image area. Almost similar image transforms are encountered in some types of low-flying unmanned vehicles equipped with downwards looking electro-optical system[1,2]. However, quite different kinds of image distortions appear when a video-camera mounted on low-flying vehicle is directed forwards or in the cases of side-looking cameras. In such situation, it may be considered that the vehicle moves inside a 3D scene in which each separately observed object surface is inclined by different angle with respect to optical axis of camera lens. Thus, the image of each such surface projected on sensor matrix of camera is subjected to geometric transform of different model. In this case, no common model of geometric transform may be applied homogeneously to the whole image of scene[1,2].

It should be stressed additionally, that the images of natural environment are subjected to hard day-time and season changes which create additional difficulties for image matching algorithms. Besides, in the image based navigation systems, the pictures acquired by camera could be matched with the video-data formed by sensors of differing kind (IR, radar, or range images, raster-type or vector maps, and so on).

* vluciv@mail.ru
† slavs2006@bk.ru

Electro-Optical Remote Sensing, Photonic Technologies, and Applications IX, edited by Gary Kamerman, Ove Steinvall, Keith L. Lewis, John D. Gonglewski, Proc. of SPIE Vol. 9649, 96490M
 · doi: 10.1117/12.2194312

It has been understood already several decades ago that the traditional correlation techniques and methods of pattern recognition in feature spaces were not robust and powerful enough for dealing with such hard image distortions. Thus, the researchers turned their attention to the structural matching algorithms[3]. However, just the object-specific algorithms of structural analysis can be usually found in the most of expert systems at present day. Application of such specific algorithms restricts the sphere of efficient practical application of developed technical solutions, in particular, in the tasks of navigation on terrain and earth monitoring because they deal in general with infinite diversity of observed images[1].

A powerful object-independent image matching technique was developed by the automatic image analysis team at the Vavilov State Optical Institute in the beginning of this century[4,5,6]. It applies an alphabet of simple contour structural elements for hierarchical description and matching of 2D images of flat enough surfaces, and it is capable of reliable matching the pictures of natural landscapes taken in differing seasons from differing aspect angles by differing sensors (the visible optical, IR, and SAR pictures, as well as the depth maps and geographical raster-type and vector-type maps). The structural descriptions built are very robust with respect to distortions mentioned above because the contour structural elements applied are extremely simple (the pieces of straight lines, angles between straight lines, pieces of curves of second order). However, such extremely simple structural elements are poorly distinguishable from the very similar other ones in the large multitudes of structural elements detected in images of natural environment. For overcoming such ambiguity, a strong additional geometric limitation is applied: the images to be matched may have only 2D affine mutual geometric transformation applied homogeneously to the whole image area (though the strong enough local deviations from this common affine transform are allowed for dealing with images of not perfectly flat surfaces).

This structural matching algorithm is a nice tool for dealing with aerospace pictures acquired by downwards looking cameras. However, a quite different technique is needed for matching the images of environment acquired in the course of travelling inside 3D scenes, e.g. in the tasks of image based navigation of terrestrial robots or some types of low-flying unmanned vehicles. In such cases, the images of a lot of object surfaces having different orientation with respect to optical axis of camera are projected onto the plane of sensor matrix, thus no common 2D geometric transform may be homogeneously applied to the whole image area[1,2]. This problem is solved in the object-independent algorithms based on matching the descriptions of local image regions. A lot of differing versions of this technique can be found in scientific papers[7,8,9,10], however they exploit in general very similar ideas and seem to have no substantial advantages as compared with the world-famous algorithms SIFT[11] and SURF[12] (let us mention also the affine SIFT[13,14] as we wanted to deal with affine transforms). Thus, we will refer to the properties of SIFT considering the pros and contras of technique based on local regions.

As compared to the extremely simple contour structural elements (based on 1D lines), the structural elements built on the base of local regions are substantially more unique because their descriptions can incorporate all information contained in 2D local areas. Thus, the juxtaposition of descriptions of the structural elements of image under analysis with the ones of template images can be substantially less ambiguous, and no additional geometric restrictions may be applied in this case. Really, the SIFT algorithm described in the well known paper[11] does not involve any geometric limitations concerning the mutual spatial positions of matched key-points for getting nice matching results, and, from this standpoint, the “classical” SIFT is not a structural matching algorithm. However, the sizes of regions chosen as structural elements are rather small, and, in fact, each region is described based on information of pixel-level (based on the properties of texture contained in it). Unfortunately, the appearance of textures of rough natural surfaces varies substantially as a result of changes in direction of observation and in conditions of illumination. It is not crucial in the cases of key-point matching indoors or in the urban environment because the man-made articles have usually rather smooth surfaces. On the contrary, the SIFT algorithm often failed in our experiments to match the images of natural scenes, e.g. the pictures taken in bush from differing aspect angles or for differing directions of illumination. At least, it was true for the SIFT program that could be free downloaded from the Lowe’s site.

However, for justice's sake, it should be mentioned that, when the “classical” SIFT and SURF descriptions were supplemented with the “structural information” about the mutual spatial positions of structural elements inside local grouping areas, the image matching results became substantially more robust[15] which enabled application of these algorithms for analysis of aerospace photographs[16]. At the same time, we would like to notice that these improved versions of SIFT[15,16] still were applied for simplified cases of image matching. The aerospace photographs to be matched seemed to be taken of Southern Europe in which the season changes of terrain are not so strongly pronounced. Besides, we did not find in these papers any mentioning of matching the images acquired by differing kinds of sensors (e.g. the optical vs. radar image data).

The advantages of supplementing the SIFT descriptions with more pronounced structural information are also mentioned by other authors. In particular, in the experiments reported by Roman Malashin[17], the additional use of information about the mutual spatial positions of matched key-points enabled successful matching the aerial pictures acquired in differing seasons by differing kinds of sensors (e.g. the optical and radar pictures). Nevertheless, it should be understood that even such enhanced SIFT algorithms cannot match any pictures vs. contour sketches or geographic maps because the last ones usually have no substantially pronounced texture in the vicinities of chosen key-points.

Taking into account the described above pros and contras of image description based on local regions and description using simple contour structural elements, we chose the last ones. They still remain most robust and have strong analogue in living vision systems (as described by David Marr[18]). However, the 2D spatial models applied in these algorithms for image description and matching should now be changed to 3D models. Thus, the remaining part of this paper is organized in the following way. The basic principles of 2D image structural description and matching based on simple contour structural elements (as they were presented by Lutsiv at al.[4,5,6]) are explained in the 2nd chapter. The methods that we applied for passing from 2D to 3D models of structural description and matching and the results of matching of 3D image data are explained in the 3rd and 4th chapters respectively. The ways of further improvement of our 3D structural matching model are discussed in the 5th chapter. The results of research and development carried out are briefly summarized in the section of conclusion.

2. OBJECT-INDEPENDENT PROCEDURE FOR STRUCTURAL MATCHING OF 2D IMAGES

The previously developed technique of 2D image matching based on simple contour structural elements will be considered in two aspects. First, we will describe the procedure of building the robust contour structural descriptions. Then, we will consider the structural juxtaposition procedure in itself, and its description will be illustrated with the image processing results at different stages of matching.

2.1 Contour detection and structural description

The observed properties of object surfaces vary substantially as a result of day-time and season changes of natural environment, while the positions of object borders remain rather stable. That is why the special neural structures aimed at border detection were found in the living vision systems[18]. Thus, just the image contours marking the positions of object borders were chosen as a source of structural description. The Deriche filtering[19] was applied for contour detection. It corresponds to a filter of second order and is best in our opinion. An example of aerial photographs and contours detected in it are shown in the fig. 1.

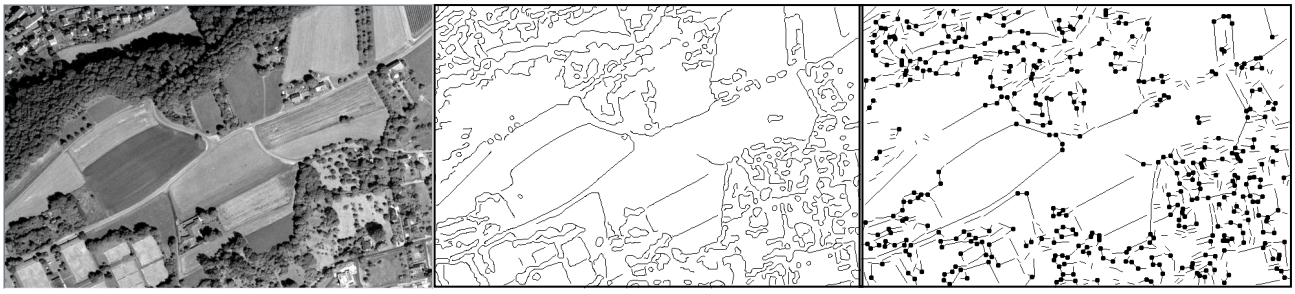

Figure 1. Image contour detection and structural description: initial aerospace photograph – at the left, Deriche contours – in the center, contour structural description – at the right (the angles between straight linear segments are marked with bold dots).

The shapes of arbitrary contours (such as shown in fig. 1) are still not sufficiently stable with respect to noise and natural changes of terrain. Besides, they are not invariant to affine and projective transforms appearing in electro-optical systems. This problem is solved by means of approximating the rough detected contours with the segments of straight lines. On the one hand, the straight linear shape is invariant to affine and projective transforms (only the parameters of line change). On the other hand, the longer is approximating rectilinear segment, – the stronger is its stability with respect to high-frequency noise causing the local variation of contour shape, thus, the possibly longest rectilinear segments should be applied for contour approximation.

The contour structural descriptions composed solely of straight lines are not suitable for reliable structural matching because all rectilinear segments are very simple geometric figures. Thus, they are very similar to each other, and

matching of such structural elements can be ambiguous the more so that the large multitudes of similar rectilinear contour segments are detected in images of natural environment as it can be seen in fig. 1. The problem becomes ever harder, because the parameters of such elements change substantially (and often – chaotically) as a result of global and local geometric transformations of pictures taken from different aspect angles. Thus, for making the structural matching less ambiguous, the alphabet of structural elements should be extended. Of course, the geometric structures composed of straight lines also are invariant to affine and projective transformations. In particular, the angles, T-junctions, Y-junctions, X-crossings remain themselves under these transforms. Thus, such structures could be included into contour structural description alphabet together with segments of straight lines. However, as it can be seen in fig. 1, there are almost no contours forming the T-junctions, Y-junctions and X-crossings among the detected contour structures. It results from the contour detection strategy applied. In particular, we chose rather strong contour smoothing for suppressing the noise and excessively large numbers of accidentally appearing small insignificant contour segments. As we learned in experiments, such smoothing made the contour descriptions more stable. Thus, only the angles were additionally included into the alphabet of structural elements. Besides, the cases may be encountered when the images to be described contain a lot of smoothly rounded borders. For such cases, the alphabet of contour structural elements was optionally supplemented with curves of second order that also are invariant to affine transforms and partially (in small pieces) invariant to projective transforms. An example of contour description with straight lines and angles is shown in the fig. 1. The rectilinear contour segments and curves are described with coordinates of their centers and their lengths. The line descriptions are supplemented with their spatial orientations, and the curves are additionally described with the directions and values of their curvatures. The angles are described with the coordinates and directions of their vertices as well as with the angle values. If the ends of neighboring structural elements connect, it is also fixed in structural description.

2.2 Structural matching of images

The task of structural matching the sets of structural elements in two images is solved unambiguously if each structural element of first image is put into correspondence to a single element in second image, or corresponds to no elements in it, and vice versa, if each element of second image corresponds to a single element or to no elements of first image. In the previously developed algorithm[5,6] of 2D structural matching, the task of juxtaposition of two sets of contour elements is solved by walking through a tree of possible hypotheses of structural correspondences as it is shown in fig. 2. A full way from the root to some leaf (some final sprig) of the tree corresponds in the fig. 2 to setting an unambiguous correspondence of the elements *A…Z* of 1st image to the elements *a…z* of 2nd image. If the 1st and 2nd images to be unambiguously matched have equal numbers of elements $N_1=N_2=N$, the number K of variants of mutual correspondences of structural elements in the images is the number of permutations ($K=P_N=N!$) that is enormously large for real large N.

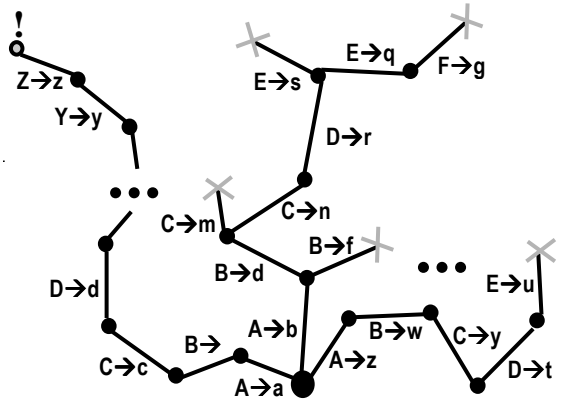

Figure 2. Juxtaposition of contour structural elements of two images by optimized travelling through a tree of juxtaposition hypotheses. The elements of first and second images are marked with capital letters and lowercase letters, respectively. The "wrong juxtaposition branches" are truncated ASAP and are not walked through completely.

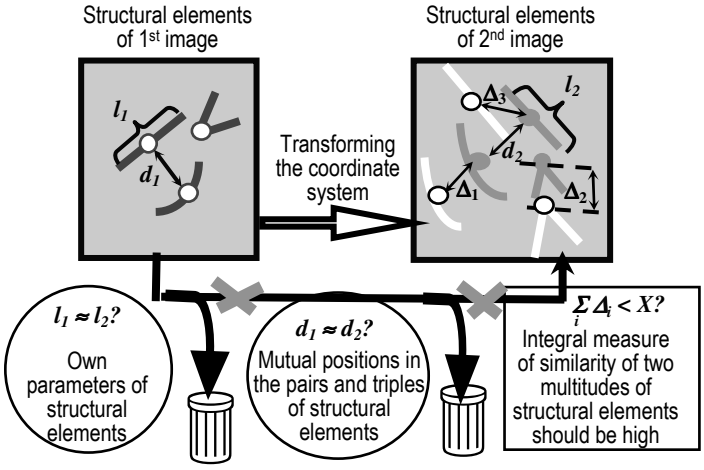

Figure 3. Cascade checking the suitability of each newly added sprig in the course of walking through the tree of structural juxtaposition hypotheses. The simplest checking (e.g. comparison of single own parameters of two structural elements) is performed earlier. If the current check was not passed, the further checking corresponding to more expensive computations is truncated.

The problem of avalanche-like growth of complexity of structural matching for the really encountered large numbers of structural elements is solved by optimized walking through the tree of matching hypotheses. Any way through the tree is not walked up to its respective final sprig and is cancelled instantly (see in the fig. 2) if an unacceptable matching of two structural elements is encountered in it, e.g. if a segment of straight line is matched with angle, or if the values of angles mismatch hardly, or if the spatial positions of newly matched elements do not conform to those of the elements already matched. Besides, the tree walking procedure is additionally accelerated using a cascade style of matching of each newly considered structural element as it is illustrated in fig. 3. According to this cascade strategy, the matching procedure for newly chosen structural element is accomplished in several stages, and it can be stopped at each of stages if the structural correspondence was classified as inadequate. And moreover, the computationally simple analysis of structural correspondence is accomplished at the very initial stages of cascade, while the computationally expensive operations are performed at final stages. Thus, in a lot of practical cases, a wrong structural correspondence can be rejected using computationally simple checking, and the computationally expensive operations will be skipped at all. For example, a simple checking whether the types of matched structural elements coincide (angle should be matched with angle, and line should be matched with line) is accomplished at the first stages of cascade as well as checking whether the values of matched angles or lengths of matched lines are quite similar (see in the fig. 3). At the further stages, the more complex operations are executed related to matched pairs of structural elements. E.g., the relative distances between them or their relative angular positions should roughly match, where the distance can be related with line length, while the direction can be related with orientation of structural element. At the next stage of cascade, a triple of elements is analyzed. A pair of newly matched structural elements and the elements matched at previously walked sprigs of tree compose a pair of triples in two images, thus a matrix of affine transform can be calculated that should have positive determinant (the negative determinants corresponding to specular reflection are prohibited). And at the final stages of cascade, the positions of structural elements newly chosen for matching in the pair of images should correspond roughly to the mutual affine transform of these images calculated based on positions of previously matched elements.

In spite of optimized tree walking and the cascade technology applied as described above, the computational complexity of structural matching still remains too heavy for really formed multitudes of structural elements. For overcoming this difficulty, the structural matching is implemented in hierarchical mode[5] as it is shown in fig. 4. The images to be matched are split into the sets of spatially compact regions smaller in size, and the whole multitude of structural elements in each image is divided into the groups corresponding to these regions. At the 1st hierarchical level, the structural elements belonging to each group of 1st image are matched independently to elements belonging to each group in 2nd image. At the 2nd hierarchical level the correctness of mutual spatial positions is checked for the pairs of groups independently matched in two images. Thus, if the multitude of N structural elements of image was divided into M groups of equal size, the maximum complexity of structural matching (the number of matching hypotheses to be checked) could be estimated as $L = M^2 P_{N/M} + P_M = M^2[(N/M)!] + M!$, where the 2nd summand corresponds to computational complexity of structural matching of groups at the 2nd hierarchical level. Let's have 1000 structural elements that will be divided into 10 groups of equal size. Then, the maximum complexity of not hierarchical matching would be $K=1000! \approx 4 \cdot 10^{2567}$, while, for the hierarchical case, the maximum complexity $L=10^2 \cdot 100! + 10! = 10^2 \cdot 9.33 \cdot 10^{157} + 3.6 \cdot 10^6 \approx 10^{160}$. Thus, for such example, the hierarchical strategy of matching would decrease the maximum computational complexity roughly $4 \cdot 10^{2407}$ times.

As it is shown in fig. 4, the hierarchical strategy of matching provides an additional useful possibility. According to the principle of adaptive resonance borrowed from living neural systems[20,21], the structural descriptions and the structural matching procedure of lower hierarchical levels are iteratively corrected based on the preliminary matching results reached at the higher levels. An example result of correction of contour structural descriptions based on the preliminary matching results of first hierarchical level is presented in the fig. 5. As it can be seen, such correction could enhance the structural matching results substantially. Similarly, the structural descriptions of higher hierarchical level (the shapes of grouping regions and the sets of contour structural elements grouped in each of them) are iteratively corrected based on preliminary matching results of higher hierarchical level.

The properly chosen method of image description (contours corresponding to stable images of object borders and affine invariant geometric primitives approximating them) as well as the hierarchical matching procedure enhanced with the principle of adaptive resonance resulted in extremely high reliability and robustness of structural matching for the images of natural environment. The optical images can be matched with the SAR pictures (as it is demonstrated in fig. 6), with geographic and vector maps, or depth maps in spite of changes in observation distances and aspect angles and in spite of day time and season changes of terrain. The raster images can be matched even with the coarse contour sketches drawn by hand[22].

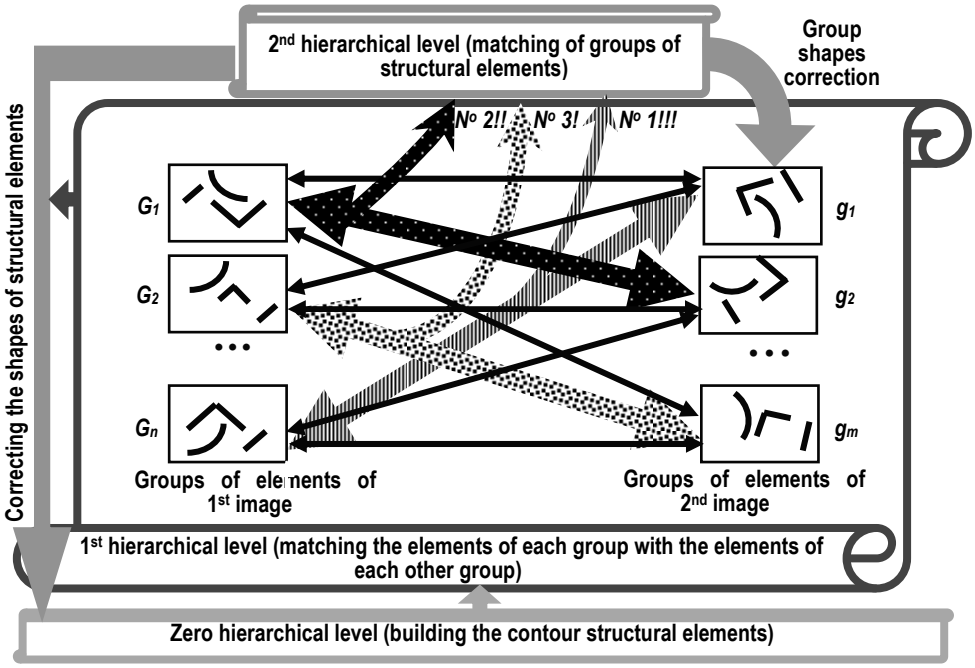

Figure 4. Hierarchical system for matching of 2D contour structural descriptions. The descending links adjust the lower level structural descriptions based on results of upper level structural matching (according to principle of adaptive resonance).

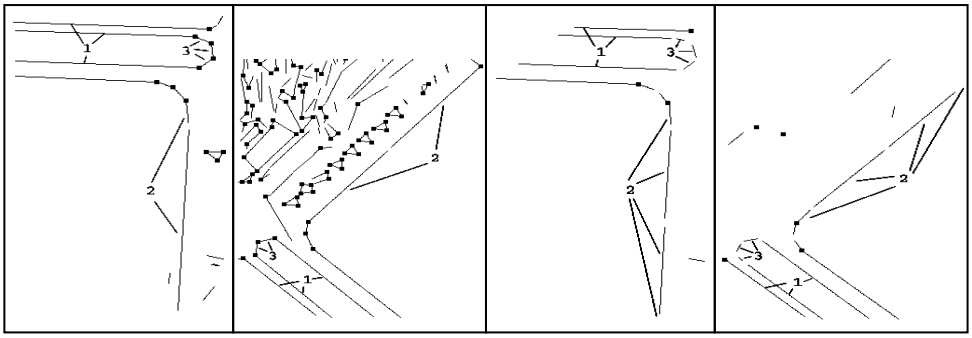

Figure 5. Correction of contour structural description in the course of structural matching using the principle of adaptive resonance: initial pair of contour descriptions to be matched – at the left, automatically corrected (adjusted) descriptions – at the right.

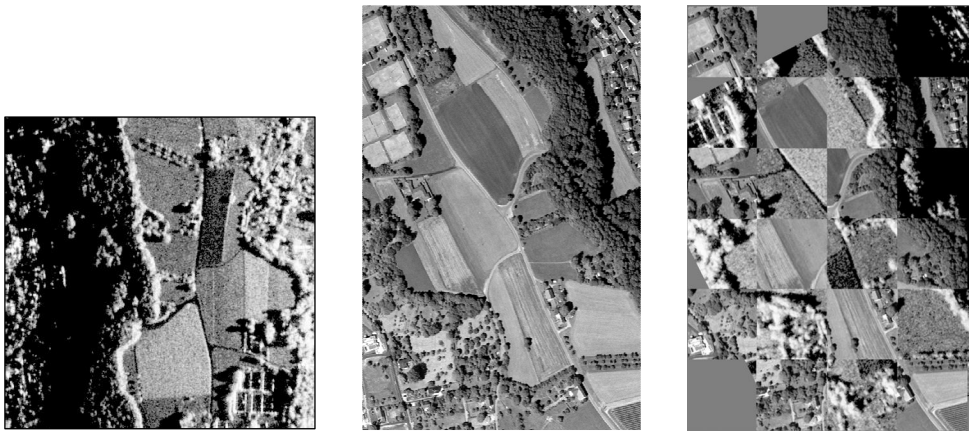

Figure 6. Example result of structural matching and registration based on 2D image analysis technique. SAR image – at the left, optical image of the same terrain – in the center, result of structural matching and registration – at the right. For convenience of visual perception, the result is presented in mosaic form: the optical and SAR image fragments alternate in adjacent mosaic cells.

3. PASSING FROM 2D TO 3D IMAGE DATA

Unfortunately, as it was described in the section above, the developed structural matching technique was aimed at dealing with images of relatively plain surfaces, e.g. with aerospace photographs or maps. In such cases, the restrictions corresponding to a single estimated 2D affine transformation model are applied in the structural matching procedure uniformly to the whole image area. It was correct for the pictures taken from satellites or from rather high flying aircrafts. However, if the images are acquired from a vehicle travelling inside a 3D scene, the image of each separate observed surface is subjected to 2D geometric transform described with its own separate model. One of the ways for solving this problem is splitting a picture into the images of separate surfaces (e.g. as it was proposed by Lutsiv at al.[23]) and applying individual models of geometric transformation to image of each surface. The additional difficulties arise in such case. On the one hand, a reliable technique should be available for correct detecting and separating the images of different surfaces. On the other hand, it is incomprehensible how to build 2D transformation model for image of surface of unknown 3D shape.

Fortunately, the modern image acquisition techniques are capable of estimating the 3rd spatial coordinates for image pixels. It could be accomplished by calculating the maps of stereo disparities, or using the technique of structured illumination (like in Kinect), or the distances may be directly measured by the radars, sonars, and scanning laser rangers. Involving the information on 3rd spatial coordinate enables transferring to 3D transformation models that are in many aspects free of the mentioned above drawbacks related to 2D image transforms. This paper is just devoted to enhancing the previously developed 2D contour structural matching technique by additional use of 3rd spatial coordinate for each image point.

Having the contour structural descriptions built as it was described above in section 2.1, we take additionally into account the information about 3rd spatial coordinates at the ends of 2D structural elements. Thus, as it is shown at the left in fig. 7, the contour structural elements are now described as 3D objects. E.g., instead of calculating the length d of line segment and its direction α based on 2D coordinates x_1, y_1, x_2, y_2 of its ends, we take into account additionally the depths z_1 and z_2 and form 3D description as length d' and angles of azimuth and elevation α' and β':

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad \rightarrow \quad d' = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2},$$

$$\alpha = \tan^{-1}\left[(y_1 - y_2) / (x_1 - x_2) \right] \quad \rightarrow \quad \alpha' = \tan^{-1}\left[(y_1 - y_2) / (x_1 - x_2) \right], \ \beta' = \tan^{-1}\left[(z_1 - z_2) / d \right].$$

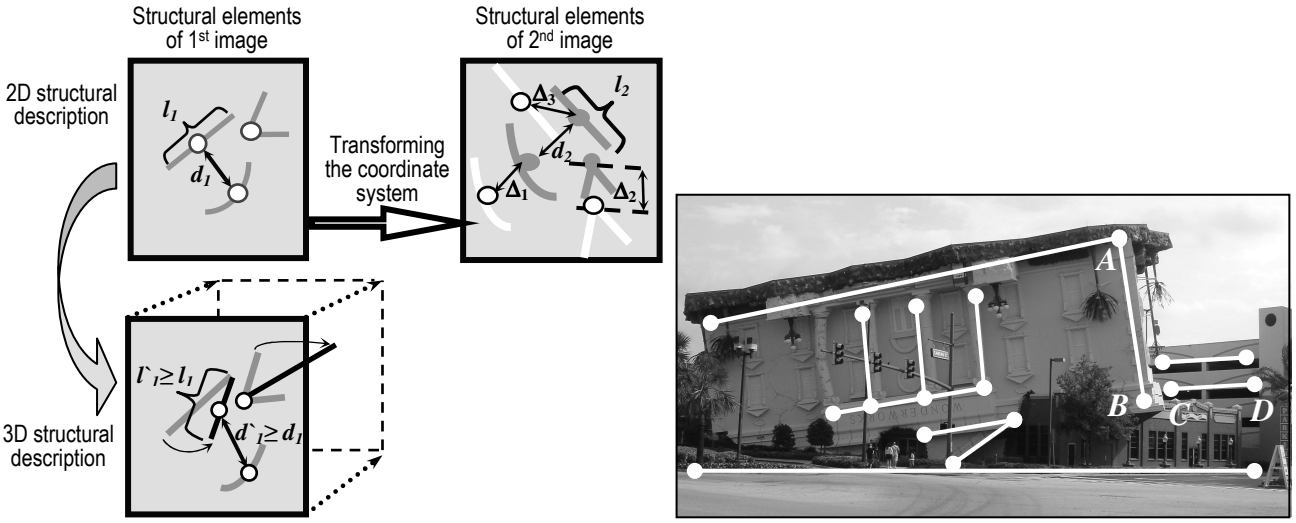

Figure 7. Passing from 2D to 3D structural descriptions (compare with fig. 3). Transfer from 2D to 3D at the left is shown only for the first image of the pair of matched ones. The ends of 3D contour structural elements (marked with black color) are now situated not in a single image plane, but at different depths. The rectilinear contour fragments *AB* and *CD* (at the right) lie in the planes located at differing depths, thus, they are skew lines and don't form real contour angles.

Of course, the direction of line segment may be also presented in the form of its three Euler angles. Similarly, the value δ of angle (angular contour structural element) is calculated via normalized scalar product of two 3D vectors $\vec{X}_1$ and $\vec{X}_2$ radiating from its vertex (3D rectilinear segments composing its sides):

$$\delta = \cos^{-1}\left(\frac{\vec{X}_1 \cdot \vec{X}_2}{\left\|\vec{X}_1\right\| \cdot \left\|\vec{X}_2\right\|} \right).$$

The direction of angle vertex is calculated as average direction of these vectors. Of course, such calculation would be correct if an angle really exists in 3D space, i.e. if the rectilinear segments composing it are not skew lines (as it is illustrated at the right in fig. 7), however, this problem will be considered later.

As it was mentioned above in section 2.2, calculating the parameters of mutual affine transform of two images is an important stage of structural matching. Passing from 2D to 3D case, we changed from 2D to 3D transformation model:

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & a_5 \\ a_3 & a_4 & a_6 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad \rightarrow \quad \begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & a_3 & a_{10} \\ a_4 & a_5 & a_6 & a_{11} \\ a_7 & a_8 & a_9 & a_{12} \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}.$$

In this case, nine transformation parameters may be estimated from five pairs of points put into correspondence in two 3D images (instead of three pairs of points in 2D images).

We carried out a series of experiments on structural matching the images of 3D scenes using the 3D matching technique described above in this and previous sections. The artificially simulated images and dense depth maps (like those shown in fig. 8) of scenes observed from different aspect angles were used in simulation. The changes of aspect angles around three Cartesian axes varied from 0° to 15° and were known from simulation procedure applied. The contour structural descriptions were built for the simulated pictures taken from different aspect angles, and the matrices of 3D affine transforms (resulting from changes of observation direction) were calculated as the results of structural matching of these descriptions. The rotation angles of scene observation direction were extracted from these matrices and compared with the rotation angles known from procedure of simulation. The errors calculated as differences of really known rotation angles and the ones calculated from structural matching did not exceed 1.5° which seems rather satisfactory. The 2D structural matching extended to 3D really worked! However, such precision was kept only for the rather narrow range of scene rotation angles as it was mentioned above. For larger rotation angles, the matching error substantially increased. Thus, we judged that the main source of increasing error was unsatisfactory treating the 2D projections of 3D scenes. In particular, a lot of contour angles detected in images of 3D scene did not correspond to any really existing angles of object borders. These were virtual angles between skew lines, and they should not be matched as really existing contour structural elements. Such matching did not cause errors for images of plain surfaces, but became crucial for 3D scenes.

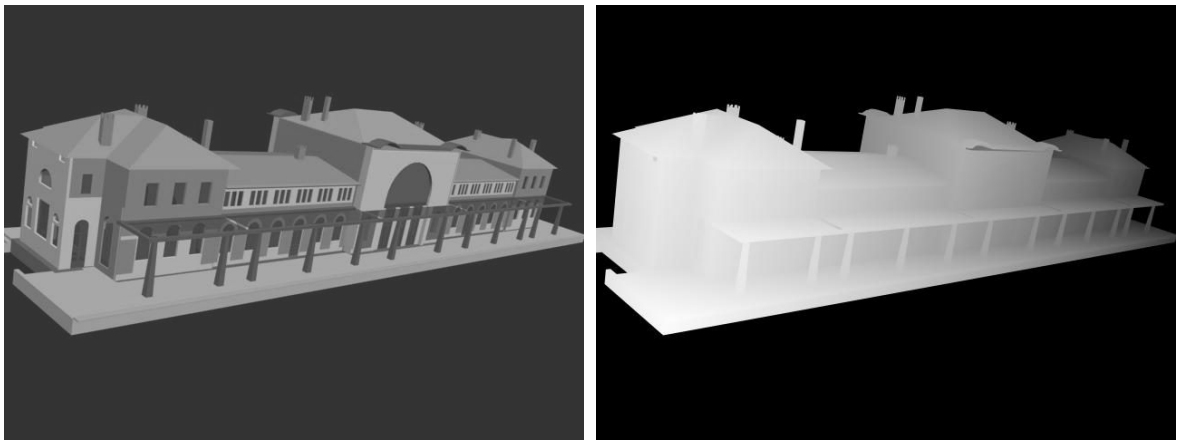

Figure 8. Example of simulated 3D scene applied in structural matching experiments: 2D image of scene – at the left and dense depth map – at the right.

The problem of virtual angles could be solved by dividing the whole multitude of contour structural elements into several groups corresponding to different surfaces of scene objects. We accomplished such segmentation keeping in mind that the condition (1) should hold for coplanar straight lines[24]:

$$\begin{vmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \\ x_4 & y_4 & z_4 & 1 \end{vmatrix} = \left(\vec{X}_3 - \vec{X}_1\right) \cdot \left[\left(\vec{X}_2 - \vec{X}_1\right) \times \left(\vec{X}_4 - \vec{X}_3\right)\right] = 0, \quad (1)$$

where $\vec{X}_1 = \begin{pmatrix} x_1 & y_1 & z_1 \end{pmatrix}^T$, $\vec{X}_2 = \begin{pmatrix} x_2 & y_2 & z_2 \end{pmatrix}^T$ are the vectors of 3D Cartesian coordinates of the endpoints of first line, and $\vec{X}_3 = \begin{pmatrix} x_3 & y_3 & z_3 \end{pmatrix}^T$, $\vec{X}_4 = \begin{pmatrix} x_4 & y_4 & z_4 \end{pmatrix}^T$ are the vectors of 3D Cartesian coordinates of the endpoints of second line, and *T* denotes the operation of transposition. We take as a seed any rectilinear segment in image contour structural description and test each other rectilinear segment detected in this image whether it holds the condition (1) with the seed segment. All the segments holding the condition (1) are marked as belonging to first plane surface and removed from further consideration. Then, step by step, the new seed segments are taken among the remaining contour elements, the elements coplanar with them are detected, marked with new sequential number of plane surface, and removed from further consideration. A set of enumerated (marked) multitudes of contour segments belonging to respectively enumerated (marked) different plane surfaces is created in this way. Now, if the angular structural elements are built of the lines belonging to each separately taken multitude, these will be for sure not virtual, but real angles.

4. EXPERIMENTS WITH REAL IMAGE DATA

The theoretical schemes described above were experimentally tested by structural matching of artificially simulated pictures and corresponding to them depth maps. Of course, it was interesting to experiment also with images of real scenes. However, in this case, the main problem was getting the sufficiently dense and precise depth maps corresponding to images of real objects. Two ways of building the depth maps were really available. On the one hand, we had a Kinect sensor at our disposal, thus we could acquire both the 2D images and related to them dense depth maps of sufficiently high quality without substantial additional efforts. However, the Kinect's depth sensor proceeds reliably only at the distances of several meters, therefore it usually could be efficiently applied indoors, and even in such case, the rooms could not be too large. As to the images and depth maps of real out of doors environment, the experiments could be carried out only in unrealistically restricted space. On the other hand, the stereo methods of depth measurement could be applied. They proceed at any distances, thus they are more suitable for out of doors experiments. However, the task of building the high quality dense depth maps is rather computationally heavy, and a crucial dilemma is usually encountered: either a depth map of better spatial resolution or less noisy map could be built. Besides, a thorough optical calibration of stereo-system should be accomplished, otherwise only the relative depths could be measured. This could influence the precision of measuring the values of angles, lengths of line segments, and parameters of mutual spatial position of such structural elements. Thus, it would be crucial for correct structural matching.

Keeping in mind the pros and contras of available techniques for depth measuring in real 3D scenes, we chose Kinect as a more suitable tool for experiments. It was more suitable at least at initial stages of practical investigation at which the errors and bottle necks of developed technique should be detected. The examples of image data really acquired indoors by Kinect from substantially differing aspect angles are shown in fig. 9 and fig.10. Just such images were applied in the experiments discussed below. The results of structural matching and registration accomplished for the image data mentioned above are presented in the fig. 11, fig. 12, fig. 13, and fig 14 and are illustrated by superposition of contours of registered images: black lines for first of them and grey lines for second one, respectively. For estimating the results reached using different matching modes, the noticeable important details of foreground and background of scene (the flowerpot and painting on a rear wall, respectively) are marked out with bold ellipses.

A registration result calculated by 2D matching engine is shown for comparison in the fig. 11 and gives the readers a possibility to estimate advantages of 3D structural matching. It can be seen in the fig. 11 that only a largest multitude of structural elements (corresponding to foreground flowerpots and plants) influenced the matching and registration results, thus the background contours corresponding to different affine (or projective) transform mismatch hardly.

As it was mentioned above, for rejecting the virtual angles composed of skew lines, the structural elements corresponding to different surfaces should be separated from each other and only the lines lying on same surface should be applied for composing the angles. For comparison, the 3D matching and registration results reached without separating the contour segments belonging to different surfaces are shown in the fig. 12. They are much better than the results of 2D structural matching discussed above. The results of registration at the background are substantially corrected, and there are some improvements in registration of long horizontal contours corresponding to nearest part of scene.

Figure 9. Example images of real indoor scene applied in structural matching experiments.

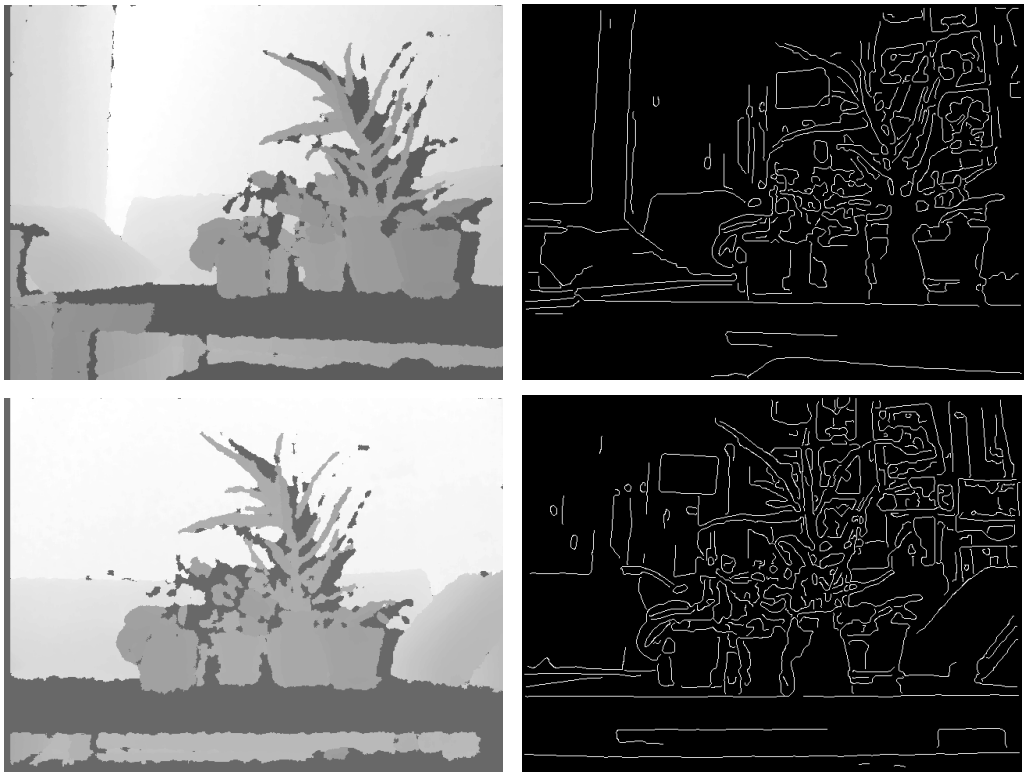

Figure 10. Dense depth maps and contours calculated for image data acquired from different aspect angles for indoor scene shown in fig. 9.

Now, let's pass to the case of separating the structural elements belonging to different surfaces. The registration results reached for this 3D matching mode are shown in the fig. 13 and fig. 14. The results shown in fig. 14 correspond to mutual affine transform calculated based on matched structural elements lying in three surfaces. The registration errors

are minimized both at the noticeable foreground and background details marked with bold ellipses, and the result looks almost excellent. We suppose that some spatial mismatch at the left part of background mostly results from the 3D affine transformation paradigm applied in this investigation. As a matter of fact, the photographed surfaces are rather near to the camera lens, thus a 3D projective transformation model should be rather applied.

For comparison, a matching and registration result calculated for structural elements belonging to a single surface is shown in the fig. 13. As it could be expected, it looks rather worse because the parameters of 3D affine transform cannot be correctly calculated from coplanar matched points. At least, a single point lying out of single surface should be taken into account. Thus, we suppose that the results shown in fig. 14 (calculated for structural elements belonging to three surfaces) could be further enhanced with the use of contour elements lying on remaining surfaces.

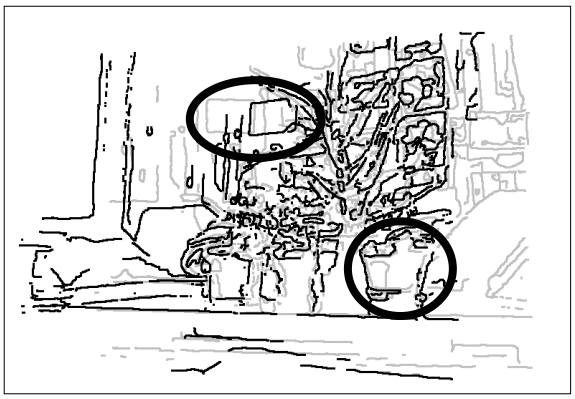

Figure 11. Result of structural matching and registration for the images of 3D scene shown in fig. 9. The 2D matching engine was applied.

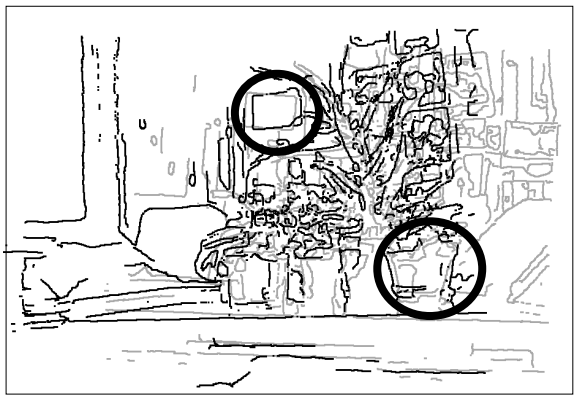

Figure 12. Result of structural matching and registration for the images of 3D scene shown in fig. 9. The 3D matching engine was applied without dividing the images into regions corresponding to separate surfaces.

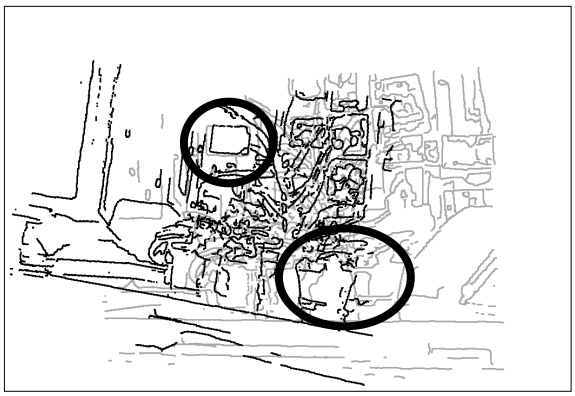

Figure 13. Result of structural matching and registration for the images of 3D scene shown in fig. 9. The 3D matching engine was applied to an image part corresponding to single surface (the rear wall of room).

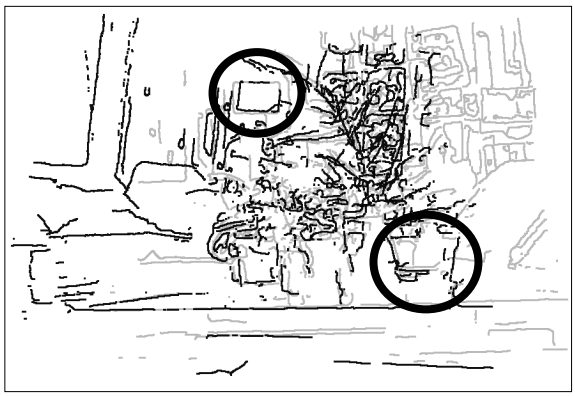

Figure 14. Result of structural matching and registration for the images of 3D scene shown in fig. 9. The 3D matching engine was applied to the image parts corresponding to three surfaces.

5. PROSPECTS OF FURTHER EVOLVING OF DEVELOPED 3D MATCHING TECHNIQUE

We discussed above our very first steps on the way of passing from 2D to 3D structural matching. The results already reached were encouraging, and they strengthened our hope that the extremely robust matching reported above for 2D image data (see the example presented in fig. 6) could be also reached for 3D scenes using our modified contour structural matching engine. For confirming this hypothesis, we should now carry out a series of experiments with the image data acquired for natural out of doors environment for the cases of varying aspect angles of observation and differing conditions of illumination. We suppose that the image data acquired by Kinect for rather short distances would be sufficient at least at the first steps of this investigation. However, then, we should pass to other techniques capable of

measuring longer distances, and it is rather difficult to get sufficiently representative sets of such image data (the illumination intensity images coupled with dense and sufficiently precise depth maps). A stereo equipment with realistically wide stereo-bases could be applied at least for sufficiently precise measuring of rather modest distances. However, for longer ranges, the time-of-flight cameras or gated laser illumination should be applied, and it would be a luck if we would have a chance to get a suitable set of such image data for real natural environment.

At the same time, as it was mentioned above, we should continue enhancing our 3D structural matching engine:

- The whole multitude of structural elements (not only those belonging to some chosen surfaces) should be involved in calculating the parameters of geometric transform. It could be accomplished by abandoning the idea of grouping the contour segments belonging to separate surfaces. Instead, the condition (1) should be checked for any angle to be included into structural description.
- No curves were included into structural descriptions. It should be thought how to deal with such 3D structural elements the more so that some curves can delusively look like rectilinear segments in images acquired from some aspect angles.
- A restricted transformation model (the 3D affine transform) was applied in the research carried out. It could be satisfactory only if the distances to scene objects were much longer than object sizes. We suspect, that this limitation has already been a source of some matching and registration errors noticed in the experiments described above. Thus, our further efforts should be also aimed at transferring from 3D affine to 3D projective transformation models in our modified contour structural matching engine.

6. CONCLUSION

A nice contour structural matching engine was developed two decades ago that was capable of matching the strongly differing images of approximately plane surfaces. That engine was a powerful tool for analysis of aerospace photographs. Unfortunately, it was not powerful enough for matching the images of 3D scenes acquired from differing aspect angles which is very urgent in the tasks of automatic navigation of unmanned vehicles. Thus, we undertook a task of enhancing that structural matching engine for making it capable of dealing with images of 3D scenes. In contrast to analysis of usual images from which the 2D object coordinates could be only extracted, we should now operate with 3D image data. Thus, a 2D image transformation model previously applied is now replaced with a 3D affine transformation model involving also additionally the third spatial coordinates of observed object surfaces. The additional technique should be applied in this case for acquiring the depth information for each image point. The Kinect device was available for getting the needed 3D image data. Unfortunately, it is capable of measuring only rather short distances, thus, at the reported stage of investigation the experiments were carried out with the images and dense depth maps acquired indoors. The 3D matching results reached for these image data are encouraging. They are noticeably better for images of 3D indoor scenes than the results demonstrated by previously developed 2D contour structural matching technique. The 3D matching results presently reached confirmed correctness of mathematical models that we applied for passing from 2D to 3D structural analysis, we also learned from them the bottle necks of approaches applied, thus, we could outline the ways of further research and development. In particular, one of the most important further steps could be transferring from 3D affine to 3D projective transformation models, and, of course, we will pass to analysis of out of doors image data.

ACKNOWLEDGMENTS

This work was supported by the Ministry of Education and Science of Russian Federation, and partially supported by the Government of Russian Federation, Grant 074-U01.

REFERENCE LINKS

[1] Lutsiv, V., Malyshev, I., “Image structural analysis in the tasks of automatic navigation of unmanned vehicles and inspection of Earth surface,” Proc. SPIE 8897, 0F1-0F15 (2013).
[2] Lapina, N. N., Lutsiv, V. R., Malyshev, I. A., Potapov, A. S., “Features of image comparison in problems of determining the location of a mobile robot,” Journal of Optical Technology 77(11), 677-683 (2010).
[3] Lutsiv, V. R., Malashin, R. O., “Object-independent structural image analysis: History and modern approaches,” Journal of Optical Technology 81(11), 642-650 (2015).

[4] Lutsiv, V. R., Malyshev, I. A., Pepelka, V. A., Potapov, A. S., "The target independent algorithms for description and structural matching of aerospace photographs," Proc. SPIE 4741, 351-362 (2002).
[5] Lutsiv, V. R., Malyshev, I. A., Potapov, A. S., "Hierarchical structural matching algorithms for registration of aerospace images," Proc. SPIE 5238, 164-175 (2003).
[6] Lutsiv, V. R., "Object-independent approach to the structural analysis of images," Journal of Optical Technology 75(11), 708-714 (2008).
[7] Kadir, T., Zisserman, A. and Brady, M., "An affine invariant salient region detector," Proceedings of the 8-th European Conference on Computer Vision, 404-416 (2004).
[8] Tuytelaars, T. and Van Gool, L., "Matching Widely Separated Views Based on Affine Invariant Regions," International Journal of Computer Vision 59(1), 61-85 (2004).
[9] Matas, J., Chum, O., Urban, M. and Pajdla T., "Robust wide baseline stereo from maximally stable extremal regions," BMVC, 384-393 (2002).
[10] Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T. and Van Gool, L., "A comparison of affine region detectors," International Journal of Computer Vision 65(1/2), 43-72 (2005).
[11] Lowe, D. G., "Distinctive Image Features from Scale-Invariant Keypoints," Int. J. of Computer Vision 60(2), 91-110 (2004).
[12] Bay, H., Tuytelaars, T., Van Gool, L., "SURF: Speeded Up Robust Features," Proc. 9th European Conf. on Computer Vision, 404-417 (2006).
[13] Morel, J. M. and Yu, G., "ASIFT: A New Framework for Fully Affine Invariant Image Comparison," SIAM Journal on Imaging Sciences 2(2), 438-469 (2009).
[14] Yu, G. and Morel, J. M., "A Fully Affine Invariant Image Comparison Method," Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 1597-1600 (2009).
[15] Ascani, A., Frontoni, E., Mancini, A., Zingaretti, P., "Feature group matching for appearance-based localization," IEEE/RSJ International Conference on Intelligent Robots and Systems, 3933-3938 (2008).
[16] Casetti, A., Frontoni, E., Mancini, A., et al., "A visual global positioning system for unmanned aerial vehicles used in photogrammetric applications," Journal on Intelligent Robot Systems 61, 157-168 (2011).
[17] Malashin, R. O., "Correlating images of three-dimensional scenes by clusterizing the correlated local attributes, using the Hough transform," Journal of Optical Technology 81(6), 327-333 (2014).
[18] Marr, D., [Vision: A Computational Investigation into the Human Representation and Processing of Visual Information], W.H. Freeman and Co., New York, (1982).
[19] Deriche, R., "Optimal edge detection using recursive filtering," Proc. 1st Int. Conf. Computer Vision, 501-505 (1987)
[20] Fukushima, K., "Neural network model for selective attention in visual pattern recognition and associative recall," Applied Optics 26(23), 4985-4992 (1987).
[21] Carpenter, G. A., Grossberg, S., "ART-2: self-organization of stable category recognition codes for analog input patterns," Applied Optics 26(23), 4919-4930 (1987).
[22] Lutsiv, V. R., Andreev, V. S., Gubkin, A. F., Iljashenko, A. S., Kadykov, A. B., Lapina, N. N., Malyshev, I. A., Novikova, T. A., Potapov, A .S., "Algorithms for automatically processing and analyzing aerospace pictures," Journal of Optical Technology 74(5), 307-322 (2007).
[23] Lutsiv, V., Potapov, A., Novikova, T., Lapina, N., "Hierarchical 3D structural matching in the aerospace photographs and indoor scenes," Proc. SPIE 5807, 455-466 (2005).
[24] Weisstein, E. W., "Coplanar," From MathWorld -- A Wolfram Web Resource. http://mathworld.wolfram.com/Coplanar.html (accessed 29 July 2015).