Recognition of Partially Occluded and Deformed Binary Objects

Ondřej Horáček, Jan Kamenický, and Jan Flusser [1,1,2]

Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 4, 182 08 Prague 8, Czech Republic

Abstract

A method dealing with recognition of partially occluded and affine transformed binary objects is presented. The method is designed for objects with smooth curved boundary. It divides an object into affine-invariant parts and uses modified radial vector for the description of parts. Object recognition is performed via string matching in the space of radial vectors.

Key words: Image recognition, partial occlusion, affine transformation, inflection point, radial vector

1 Introduction

Recognition of objects under partial occlusions and deformations caused by imaging geometry is one of the most difficult problems in computer vision. It is required always when analyzing 2-D images of a 3-D scene. Although many methods trying to solve this task have been published, it still remains open. Clearly, there is no universal algorithm which would be "optimal" in all cases. Different methods should be designed for different classes of objects and for different groups of assumed deformations.

Email addresses: horacek@utia.cas.cz, kamenik@utia.cas.cz, flusser@utia.cas.cz (Ondřej Horáček, Jan Kamenický, and Jan Flusser).

[1] Ondřej Horáček and Jan Kamenický were supported by the Czech Ministry of Education under the project No. 1M6798555601 (Research Center DAR).

[2] Jan Flusser was supported by the Grant Agency of the Czech Republic under the project No. 102/04/0155.

This paper is devoted to objects with complicated curved boundary. Such a boundary cannot be approximated by a polygon without loss of accuracy, so we do not employ polygonal-based methods at all. Furthermore, we assume the objects are deformed by an unknown affine deformation. When photographing a planar object arbitrary oriented in 3-D space, the precise image deformation would be a perspective projection. It is well known it can be approximated by affine transform when the object-to-camera distance is large comparing to the size of the object. We use the approximation by affine transformation because it is easy to handle, mainly because it is linear and its jacobian is constant in the image.

We introduce a method developed for the recognition under the above mentioned conditions. First, the object is divided into parts which are defined by means of inflection points of the object boundary. Then the shape of each part is described by a special kind of radial vector. Finally, the parameters of the affine deformation are estimated and classification is performed by string matching in the space of radial vectors. The performance of the method is demonstrated by experiments.

2 Overview of current methods

Current methods can be classified into two major categories. The methods of the first group divide the object into affine-invariant parts. Each part is described by some kind of "standard" global invariants, and the whole object is then characterized by a string of vectors of invariants. Recognition under occlusion is performed by maximum substring matching. Since inflection points of the boundary are invariant to affine (and even projective) deformation of a shape, they become a popular tool for the definition of the affine-invariant parts. This approach was used by Ibrahim and Cohen (1998), who described the object by area ratios of two neighboring parts. As a modification which does not use inflection points, concave residua of convex hull could be used (Lamdan, 1988). For polygon-like shapes, however, inflection points cannot be used. Instead, one can construct pats defined by three or four neighboring vertices. Yang and Cohen (1999) used area ratios of the parts to construct affine invariants. Flusser (2002) further developed their approach by finding more powerful invariant description of the parts. A similar method was successfully tested for perspective projection by Rothwell (1992).

The methods of the second group are "intrinsically local", i.e. they do not divide the shape into subparts but rather describe the boundary in every point by means of its small neighborhood. In that way they transform the boundary to a so-called signature curve which is invariant to affine/projective transform. Recognition under occlusion is again performed by substring matching in the space of signatures. Typical representatives of this group are differential invariants. They were discovered hundred years ago by Wilczynski (1906) who proposed invariant signatures based on derivatives of up to 8-th order. Weiss (1992) introduced differential invariants to the computer vision community. He published a series of papers on various invariants of orders from four to six (Weiss, 1988; Bruckstein at al., 1997). Although differential invariants seemed to be promising from theoretical point of view, they have been experimentally proven to be extremely sensitive to inaccurate segmentation of the boundary, discretization errors and noise.

Following methods dealing with recognition of transformed object could be relevant to our conditions, too. Mokhtarian and Abbasi (2002) used inflection points themselves to characterize the boundary. They constructed so-called Curvature Scale Space and traced the position of inflection points on different levels of image pyramid. The trajectories of the inflection points then served as object descriptors. Lamdan (1988) used mutual position of four "interesting" points for the recognition. To verify the received match, normalized concave areas were described by the radial vector. There are also methods based on wavelet transform of the boundary. Tieng at al. (1995) introduced wavelet-based boundary representation, where affine invariance was achieved by enclosed area contour parametrization. A similar approach was used by Khalil at al. (2001). However, the use of the wavelet-based methods in case of partial occlusions is questionable.

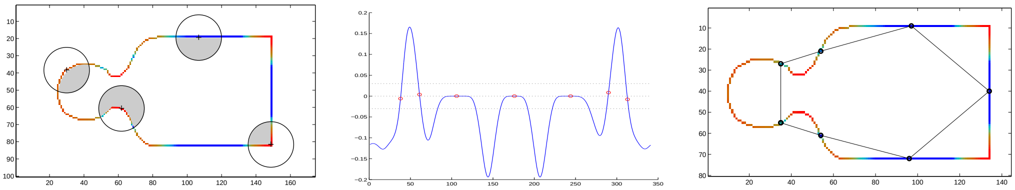
3 Definition of affine-invariant parts

Both inflection points and central points of straight lines are affine invariant, i.e. the properties "to be an inflection point" and "to be a central point of a straight line" are preserved under arbitrary nonsingular affine transform. Thus, we use these points (called "cut points" in the sequel) for the construction of affine-invariant parts. We connect each pair of neighboring cut points by a line. This line and the corresponding part of the object boundary form a convex region which may but need not to lie inside the original object (in Fig. 1c). The sequence of such parts carries efficient information about the object.

Detection of inflection points of discrete curves has been discussed in numerous papers. Let us recall that in the continuous domain an inflection point is defined by a constraint $\ddot{x}(t)\dot{y}(t) - \dot{x}(t)\ddot{y}(t) = 0$, where $x(t), y(t)$ represent a parametrization of the curve and the dots denote derivatives with respect to t. When this definition is directly converted to the discrete domain, it becomes very sensitive to sampling and noise. Thus, we propose a new robust method of curvature estimation.

For each boundary point, we construct a circle with its center in this point and having fixed radius (see Fig. 1a). We estimate the object boundary curvature as $\frac{object\ covered\ area}{whole\ circle\ area} - \frac{1}{2}$. The curvature is negative for convex parts of the object boundary (less than a half of the circle is covered by the object), positive for concave parts, and equals zero for inflection points and straight lines. To suppress small fluctuation of the curvature value, we apply a smoothing of the curvature series by convolution with a narrow gaussian kernel. We get a smoothed curvature graph, such as in Fig. 1b.

Now we construct the division of the original object into parts. Zero crossing points of the curvature and middle points middle points of approximately zero-value segments on the curvature graph serve as cut points, it means points separating the object parts. We connect neighboring cut points by straight line (see Fig. 1, which defines the object into parts. The parts, the area of which is less than a given threshold, are not considered.



a) Curvature estimation. The boundary color illustrates absolute value of the curvature.

b) The curvature graph with the cut points.

c) Division of the object into parts.

Fig. 1. Definition of affine-invariant parts.

4 Description of the parts

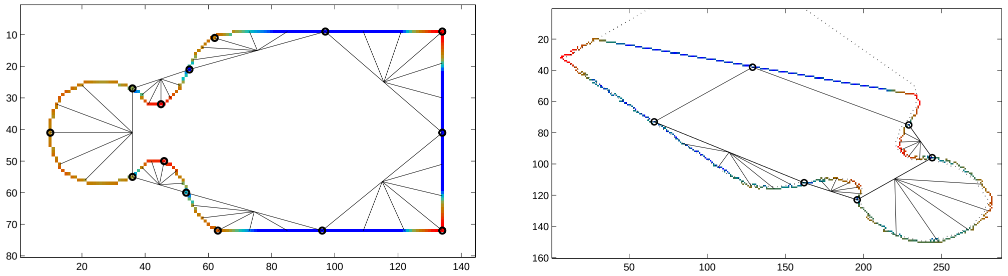
The object is represented by the parts defined in the previous section. By adding a description of the shape of the individual parts we get a description of the whole object which is robust to occlusion. Robustness to occlusion means that if some part of the object boundary is missing or changed, only few elements of the feature vector are changed. This is an important attribute. Note that traditional global methods, for instance description of the object by moment invariants or Fourier descriptors, do not have this property.

It would be possible to describe each part individually and eliminate the impact of the deformation by using proper affine invariants (moment invariants or Fourier descriptors for instance). In such case, however, we do not employ important information that all the parts were deformed by *the same* transformation. Including this consistency information in the object description can significantly increase the recognition performance. Thus, we propose the following description of the parts by a modified radial vector, with included position of control points. See complete demo object description in Fig. 2a.

The spokes of the modified radial vector come from the middle of the cutting line and they divide the part into subparts of equal area. For each object part, they are constructed as follows.

(1) Let n be the required number of the spokes (i.e. the length of the radial vector).
(2) For each boundary point, do the following.
(3) Calculate the area of the triangle between the current boundary point, the neighboring boundary point, and the midpoint of the cutting line.
(4) If the cumulated area exceeds $k/(n-1)$ fraction of the total part area, we put the ending of the k-th spoke in the current boundary point.
(5) The radial vector consists of the n spokes lengths. For each part, we store its radial vector. It describes the part shape completely.
(6) We store also absolute position of the part. It is represented by position of its cut points and position of the mid-spoke end-point. This affine variant, complementary information will allow us to recover the transformation later.

The introduced modified radial vector divides the part invariantly under affine transformation. Note that a classical radial vector with constant-angle spokes distribution or constant-boundary length spokes distribution has not such a favorable property.



a) Object description by radial vectors. b) Sequence of similar parts.

Fig. 2. Description and matching of the demo object.

5 Matching

The image is classified by finding the longest and best matching section of the border (in Fig. 2b). This is realized by comparing sequences of parts, represented by their radial vectors, between the classified image and database objects.

(1) Initialize minimal required match length P to 1 and similarity threshold S_{thre} to its minimal required value S_{min} (S_{min} is a user defined parameter, we recommend $S_{min} = 0.8$).
(2) For each part of the database object take a sequence of this parts and $P - 1$ next parts.
(3) For each part of the image object take a sequence of this parts and $P - 1$ next parts.
(4) Calculate affine transformation T that transforms the image parts sequence to the database sequence. The least square fit is applied to their cut points and mid-spoke end-point.
(5) Transform the spokes of the image radial vectors by the transformation T and calculate their length, i.e. get their radial vectors.
(6) Compare each radial vector with the one from the database by means of the similarity measure S (described bellow).
(7) If $S > S_{thre}$, these two sequences match. Mark these sequences as the best ones, denote their length as P_{best} and its similarity as S_{best}. Now try to make the sequence even longer, set $P = P + 1$, $S_{thre} = S_{min}$ and continue by step 4).
(8) Otherwise reset the sequence length and similarity threshold to the last best values $P = P_{best}$ and $S = S_{best}$ and continue the by loop 2), resp. 3).

There are many choices how to measure similarity between who radial vectors $u = u_1, \ldots, u_n$ and $v = v_1, \ldots, v_n$. It may be misleading to use ℓ_2 norm. We introduce original similarity measure $S \in \langle 0, 1 \rangle$, that we have found to perform well in the practical experiments.

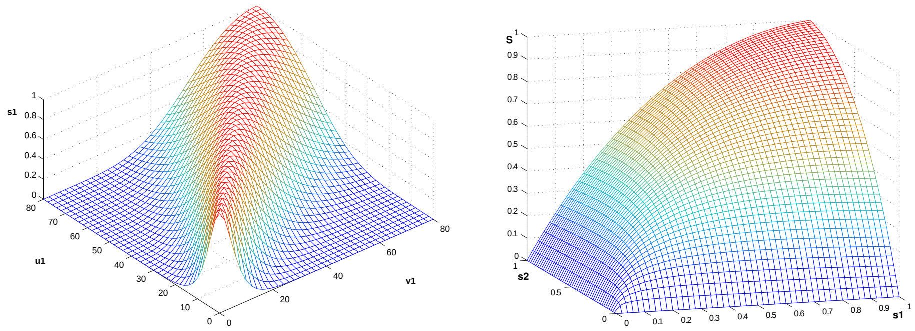
Before defining concrete S, we put some general constraints on it. We require $S = 1$ only if $u = v$, S decreases to zero for growing vector difference. The single similarity measure s_i of the i-th spoke lengths u_i, v_i is a Gaussian quantity of the $u_i - v_i$ difference (in Fig. 3a)

$$s_i = e^{-\frac{1}{\sigma_i^2}\left(\frac{u_i - v_i}{2}\right)^2}, \qquad \sigma_i = k_1 + k_2 \left|\frac{u_i + v_i}{2}\right|,$$

where k_1 and k_2 are user-defined parameters.

We have the following requirements for combining single component s_i to overall similarity measure S. We require $S = s_i$ if all s_i are equal, $S = 0$ if at least one $s_i = 0$, and S needs to be sensitive to all s_i. Moreover, it is reasonable to require S to be 0.75 if all but one s_i equal 1 and one s_i equals 0.5 (in Fig. 3b). One can construct many heuristic functions fulfilling the above constraints. After testing several possibilities we decided to use the following functions because of their simplicity and good performance

$$S = \frac{\sum_{i=1}^{n} w_i \cdot s_i}{\sum_{i=1}^{n} w_i}, \qquad w_i = \frac{n-2}{s_i} - (n-3).$$



a) Single spokes similarity measure s_i depends on difference of the spoke lengths.

b) Total similarity measure of radial vectors S is a combination of the spoke similarity measures.

Fig. 3. Similarity measure is introduced for radial vectors comparison.

6 Experimental results

The proposed method was tested on a set of 24 binary objects (Fig. 4) segmented from original color images. The objects were successively deformed by various affine transforms, their various regions were occluded and then the objects were matched against the original database. The sufficient number of matching parts is used as a criterion for match of the objects. This is in fact a well-known principle of string matching.

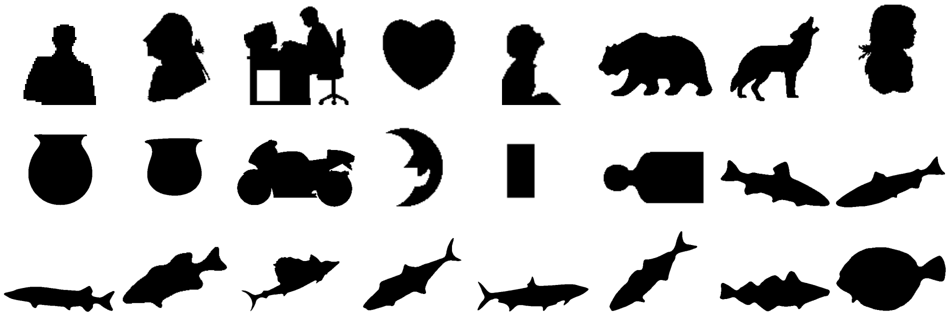
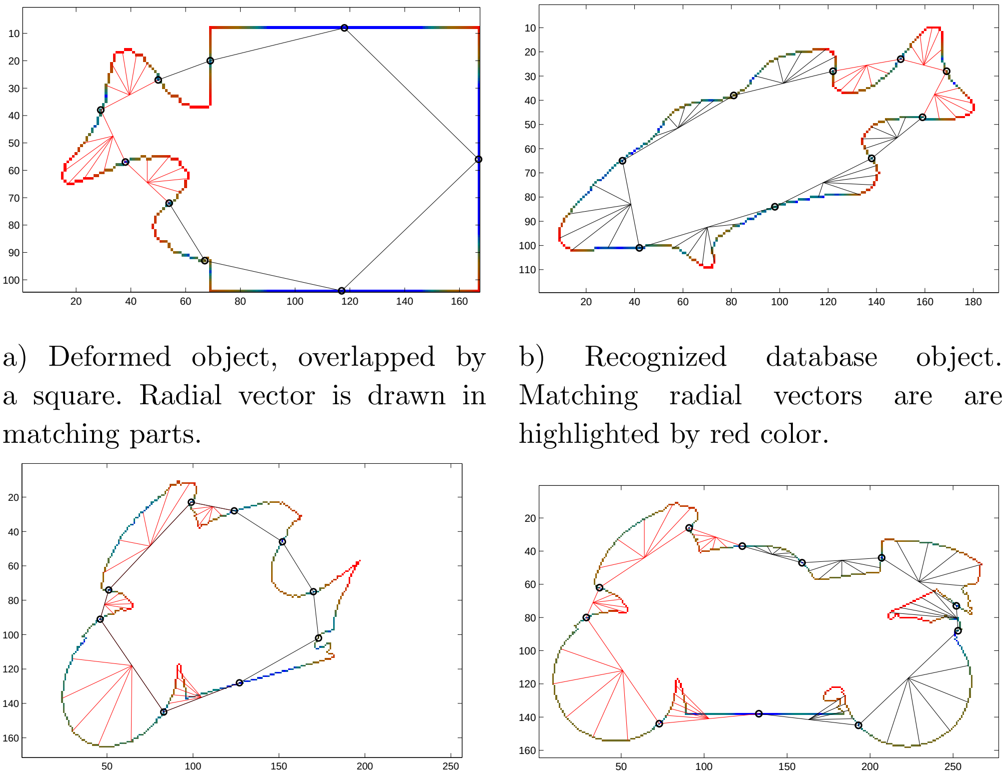


Fig. 4. Our object database. These 24 objects are represented by 204 object parts.

For illustration, two examples are shown in Fig. 5. On the left-hand side, one can see partially occluded and transformed objects. The corresponding database objects (which were successfully found in both cases) are shown on the right-hand side. The control (inflection) points are highlighted, their connecting lines define the division into parts. The spokes of the corresponding radial vectors are drawn inside the matched parts of the image.



a) Deformed object, overlapped by a square. Radial vector is drawn in matching parts.

b) Recognized database object. Matching radial vectors are are highlighted by red color.

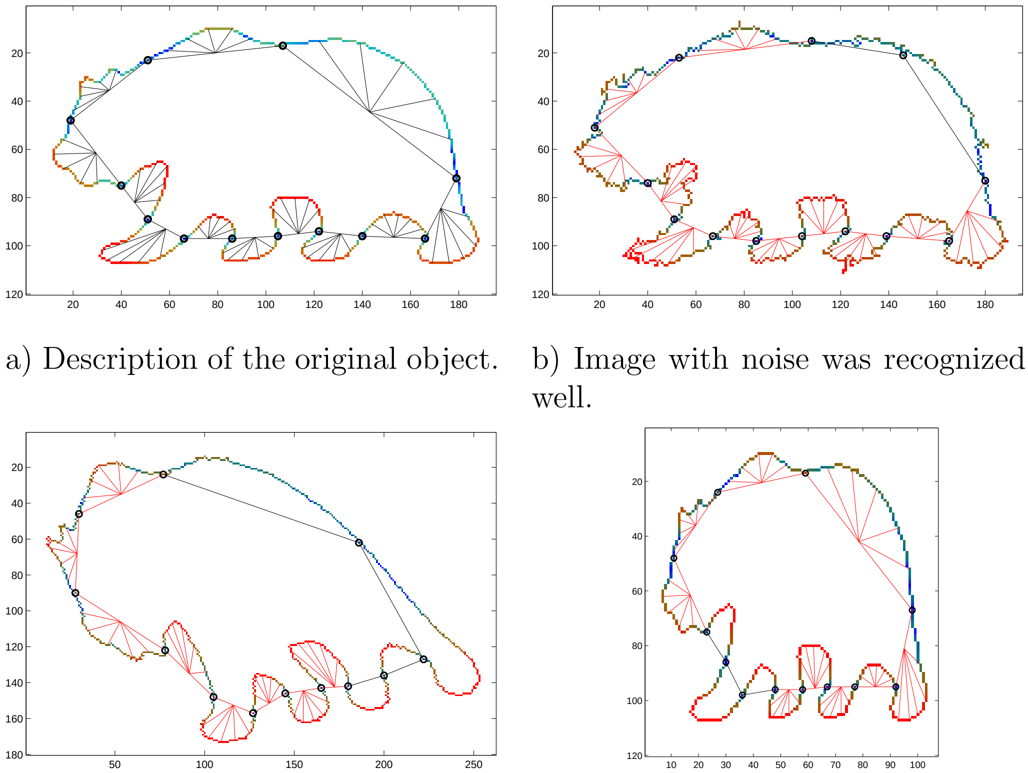
c) Partially occluded and deformed image.

d) Recognized database object. Note the radial vectors of this complicated shape.

Fig. 5. Recognition examples and description of recognized database objects.

The modified radial vector describes the boundary with a good precision, the tolerance to a shape perturbations is controlled by user-defined parameters/thresholds. This enables an optimization for various types of shapes. Surprisingly, the boundary does not need to be a smooth curve with well-defined inflection points. The method finds control points even on polygonal parts (in Fig. 5a). Furthermore, due to some tolerance threshold for the detection of inflection points, we can obtain even some non-convex parts. We are able to construct radial vector also for these non-convex parts (in Fig. 5d). Remind that our modified radial vector is created by dividing cumulative area while proceeding the part boundary.

The object description and the result of a recognition naturally depends on the conditions of the experiment: character of the shapes, amount of occlusion, degree of the transformation, and noise. Before summing up statistical experiments, let's focus on some situations in detail. As we can see in Fig. 6b), thanks to our robust curvature estimation and similarity measure, the proposed method's resistance to noise is quite good. It is possible to set the



a) Description of the original object.

b) Image with noise was recognized well.

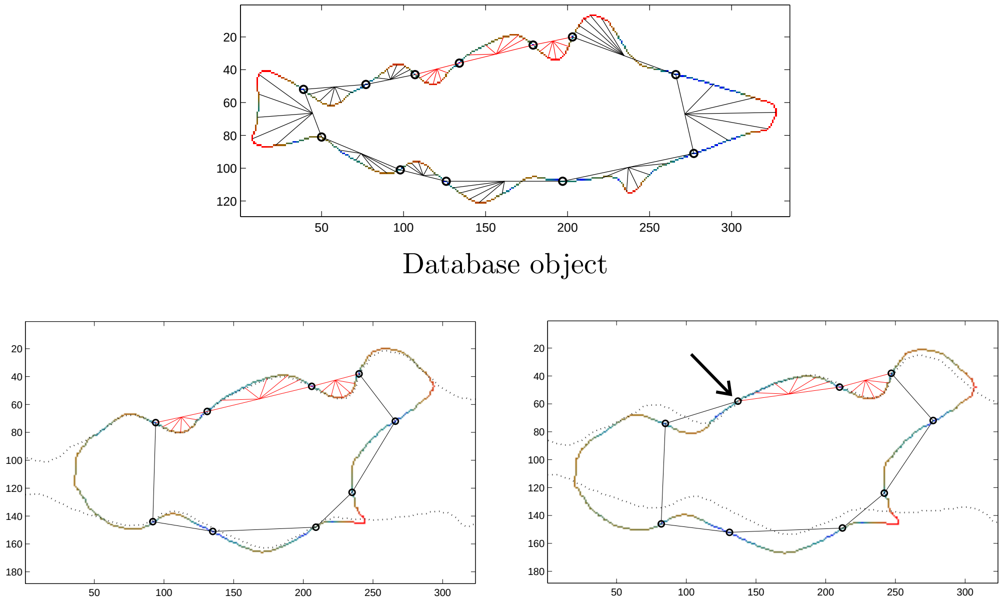
c) Transformed image was recognized, even if the back part did not match.

d) The match was shorten by image transformation.

Fig. 6. Influence of affine transformation and noise to recognition.

method parameters to be even more robust to noise than in the Fig. 6b), but the number of false positive matches would grow. Affine transformation was applied to the images in Fig. 6c) and 6d). The original object was recognized from the image, but some parts were not included in the match. We will explain this phenomenon on following example.

At the bottom of Fig. 7 are two overlapped objects, the second one was transformed by a slightly harder transformation. One can see that the position of the marked control point was evaluated differently due to changed curvature of the boundary. Therefore, one of the parts changed its shape and did not match with the pattern part. This leads to match reduction and worse position detection (the overlayed database object is drawn by dotted line). Control point instability is caused by unsuitable shapes (without clear inflection points), affine transformation (affects the curvature), or occlusion (inflection points originally ignored can become significant). Although our control point detection on a smooth boundary was improved comparing to traditional methods, it still remains a principal problem. In general, we can say that the recognition is as good as the stability of the critical points.



Database object

Control points are stable for this occlusion and transformation.

Instability of a control point leads to match reduction and worse position detection.

Fig. 7. The impact of instability of the control points.

Recognition under various conditions is summarized in Table 1. ”Image area” denotes the size of the visible part of the test object (in per cent), ”Constant scale of details” indicates whether or not the same thresholds were used for database and test objects when detecting inflection points, and ”Transformation” means the significance of the deformation measured by skewing. The transformation was chosen randomly to significant skewing and the occlusion was made automatically by straight white area. The image degradation is visualized on the square image in the table header. The table itself shows the maximum number of matching parts over all database objects. In all instances where the maximum number of matching parts was greater than two the test objects were recognized correctly. One or two matching parts does not ensure unique correct match, so the classification can be wrong. These not-recognized objects are represented inside the table by image. Their problem is caused by strong deformation or large amount of occlusion which leads to instability of control points.

We tested also the impact of a perspective transformation on the recognition rate. We took several photos of an object (trencher) with a camera in various positions. The object was segmented by single thresholding. Boundary of the segmented binary image was noisy and there was notable impact of the object thickness. When the camera was about 1 meter from the trencher, the perspective effect of the image transformation was not too strong and the object was recognized well (in Fig. 8). After we moved the camera to about 25cm

Table 1
Experiment of 8 object recognition under various conditions. Correct recognition is represented by the numbers of matching parts. Icons inside the table denote the cases where only an insufficient number of parts were found for unique recognition.

Image area	100%	*90%*	*50%*	*50%*	100%	100%	*50%*
Constant scale	yes	yes	*no*	yes	yes	yes	yes
Transformation	none	none	none	none	*medium*	*strong*	*medium*
Input image							
	12	7	4	3	10	7	3
	11	8		7	4	6	
	11	8	3	4	6	4	3
	11	9	3	4	3		3
	7	4			7	7	
	13	8			3		
	10	7		6	8		4
	9	6	4	4	4	8	4

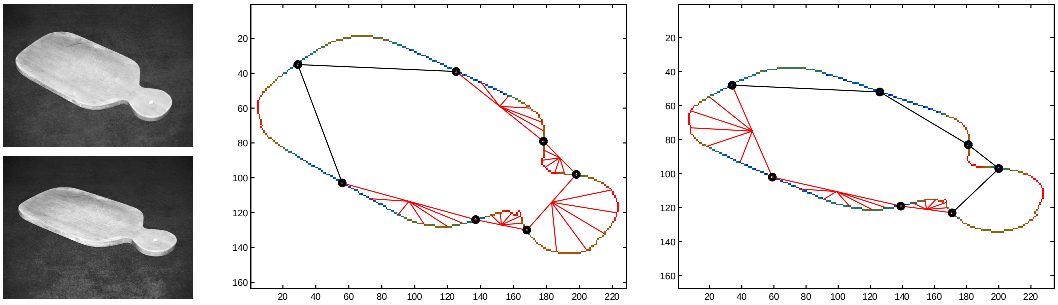


Fig. 8. Recognition under mild perspective projection. Both objects on the left were identified correctly.

from the object, the transformation became obviously nonlinear (see Fig. 9). This is in contradiction with our original assumption about the linearity of the deformation, and that is why the recognition may failed in some cases.

6.0.1 Comparison with area ratio method

The presented method is compared to Ibrahim and Cohen (1998) paper, which is based on area ratio of shape parts. Although their algorithm was originally

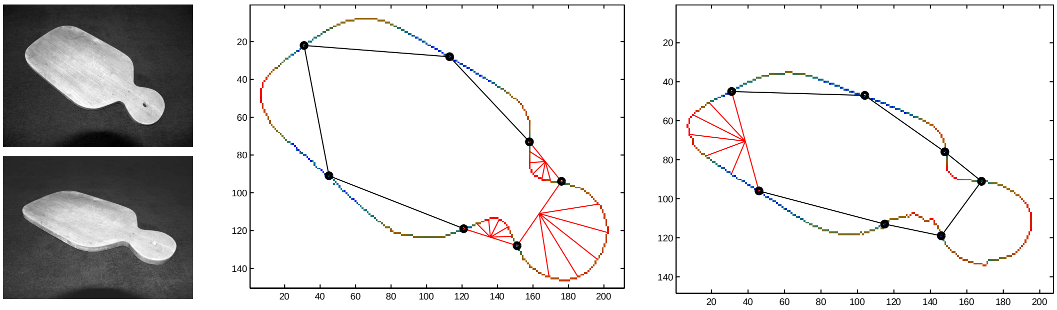


Fig. 9. Recognition under heavy perspective projection. The object top left was identified correctly, while the object bottom left was misclassified because of significant nonlinear deformation.

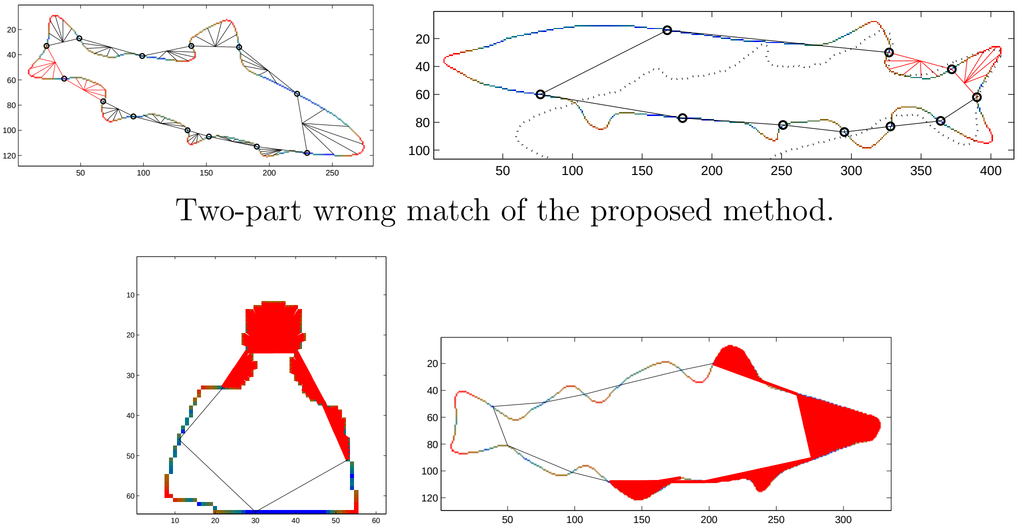
Table 2
The number of incorrect matches of area ratio and proposed method. Thresholds of Ali's method was set to classify our correct matches as good ones.

Length of possible match	Area ratio method	The proposed method
4-part string	12 wrong matches	0 wrong matches
3-part string	59 wrong matches	0 wrong matches
2-part string	249 wrong matches	10 wrong matches

not developed for recognition of partially occluded objects, it is suitable for these conditions too. Their parts are bounded by inflection points as well.

Recognition power and discriminability of the methods were tested by mutual matching of our 24 database objects (in Fig. 4). Note, the objects are represented by 204 parts and a match can be detected on each of their combination. Implementations of both algorithms use the same detection of inflection points, therefore we can set a threshold of the Ali's method to classify our correct matches as good ones. In Table 2, counts of incorrectly matched neighboring parts are compared.

It is clear that the area ratio carries much less information than our modified radial vector. Proposed method needs only 3 parts for unique correct match, while the area ratio method requires 5 parts. These numbers are relevant to our database, different number of matching parts could be required for unique object match on some other database. Both methods can be affected by control points instability. In Fig. 10 you can see one of the 10 worst two-part wrong matches of presented method and a sample of wrong four-part match of area ratio method.



Two-part wrong match of the proposed method.

Four-part wrong match of the area ratio method

Fig. 10. Example of too short matches for correct recognition, for both compared methods. Proposed method needs 3 parts for unique correct match, the area ratio method requires 5 parts.

7 Conclusion

We presented a method for recognition of partially occluded binary objects deformed by affine transformation. The method uses local affine-invariant description of the object boundary by means of inflection points and radial vectors. When working with digital boundary, the major limitation of the method is stability of inflection points. As the experiments demonstrated, if the curve has "prominent" inflection points, they are usually very stable under affine transformation and the method works perfectly. On the other hand, in the case of obscure boundary the inflection points may be detected at different positions depending on the particular transformation and/or occlusion and the recognition may fail.

Our experiment proved a good discrimination power of the method. On the given test set, we discovered that if the maximum number of matched boundary parts between the unknown object and the database element is greater than two, it always indicates a correct match. Thus, this threshold can be recommended for prospective real experiments too.

References

Bruckstein, A.M., Rivlin, E., Weiss, I., 1997. Scale-space Semi Local Invariants. Image and Vision Computing 15, 335–344

Guggenheimer, H.W., 1963. Differential Geometry. New York. McGraw-Hill

Flusser, J., 2002. Affine Invariants of Convex Polygons. IEEE Transactions on Image Processing 11

Ibrahim Ali W.S., Cohen, F.S., 1998. Registering Coronal Histological 2-D Sections of a Rat Brain with Coronal Sections of a 3-D Brain Atlas Using Geometric Curve Invariants and B-spline Representanion. IEEE Transaction on Medial Imaging 17

Khalil, M.I., Bayeoumi, M.M., 2001. A Dyadic Wavelet Affine Invariant Function for 2D Shape Recognition. IEEE Transaction on PAMI 23, 1152–1163

Lamdan, Y., Schwartz, J.T., Wolfson, H.J., 1988. Object Recognition by Affine Invariant Matching. Computer Vision and Pattern Recognition, 335–344

Mokhtarian, F., Abbasi, S., 2002. Shape Similarity Under Affine Transform. Pattern Recognition 35, 31–41

Rothwell, C.A., Zisserman, A., Forsyth, D.A., Mundy, J.L., 1992. Fast Recognition Using Algebraic Invariants. Geometric Invariants in Computer Vision. Cambridge, MIT Press, 398–407

Tieng, Q.M., Boles W.W., 1995. An Application of Wavelet-based Affine-invariant Representation. Patern Recognition Letters 16, 1287–1296

Weiss, I., 1988. Projective Invariants of Shapes. Proc. Image Understanding Workshop, 1125–1134

Weiss, I., 1992. Noise Resistant Invariants of Curves. Geometric Invariance in Computer Vision. Cambridge, MIT Press, 135–156

Wilczynski, E.J., 1906. Projective Differential Geometry of Curves and Ruled Surfaces. B. G. Teubner, Leipzig

Yang, Z., Cohen, F.S., 1999. Image Registration and Object Recognition Using Affine Invariants and Convex Hulls. IEEE Transitional on Image Processing 8