

Effective Object Localization in Images by Calculating Ratio and Distance Between Pixels

Rykhard Bohush¹ [0000-0002-6609-5810], Sergey Ablameyko² [0000-0001-9404-1206],
Yahor Adamovskiy¹ [0000-0003-1044-8741], Dmitry Glukhov¹ [0000-0003-4983-2919],

¹Polotsk State University, Blokhina st., 29, Novopolotsk, Republic of Belarus, 211440
{ r.bogush, adamovskiy.y, d.gluhov}@psu.by

²Belarusian State University, Nezavisimosti avenue 4, Minsk, Republic of Belarus
ablameyko@bsu.by

Abstract. In this paper, two novel similarity functions which consider the spatial and brightness relations between pixels for object localization in images are presented. We explore different advantages of our functions and compare them to others that use only spatial connection between pixels. It is shown, that one of them is robust to linear change in pixel brightness levels of the compared images. Comparison of computational cost and localization accuracy of shifted object for our similarity functions with others is given in the paper. The presented experimental results confirm the effectiveness of the proposed approach for object localization.

Keywords: similarity functions, object localization, distance

1 Introduction

Object localization is one of the critical tasks of computer vision. The task is used for the solution of applied problems, such as the automatic search for defects over images in industrial and medical diagnostics, the search for specified objects in inquiry and communication systems, the discovery and localization of bench marks in satellite Earth's surface images, the establishment of correspondence between the conjugate points of two and more images during the procedure of their binding, the tracking of targets in airborne radar systems, etc. [1, 2]. For this reason, a lot of works are devoted to the development of methods for localization of objects on static images and video sequences. Localization is notably more challenging than image classification because it involves generation of precise object locations [3]. Furthermore, the magnitude of the problem increases manifold in real-world situations where objects vary in position, scale and outlook, surrounded by cluttered scenes.

Unfortunately, there currently does not exist an optimal distance measure or similarity function whose application would provide the maximum efficiency for object localization by different features in images. Therefore, traditional similarity metrics are improved or new functions are provided considering expansion of applications of image processing, and different similarity functions are selected for various applica-

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). IntelITSIS-2020

tions [4, 5]. Image recognition applications require more shift and rotation robust measures. But registration and tracking applications require better localization and noise tolerance

In the general case for object detection and localization in an image, the similarity value is calculated for its features and the features of all the image fragments. The decision about the presence of an object is made by comparing the values with a threshold. If the comparison condition is met, the decision about the correspondence of a fragment to a standard is made. It is obvious that precise localization requires the similarity function value to exceed a threshold only for the correct location of an object. However, to provide the minimum probability of missing an object, it is necessary to decrease the threshold value, but this leads to an increase in the number of discovered image fragments, which do not correspond to the sought object, including the fragments near its correct location in an image. For this reason, the next step consists in determining the precise coordinates for the location of an object. Inaccuracy complicates the precise estimation of coordinates for the true location of an object due to an excess over the threshold function value for the standard and the set of neighbouring fragment values. Ambiguity is characterized by an excess over the similarity function threshold for a limited number of image fragments, which essentially overlap each other, at a certain distance between their centres. Therefore, the localization accuracy also depends on the similarity function.

We propose a new similarity functions for object detection and localization in image and video based on ratio and distance calculation between pixels. These functions can be used for arbitrary features of images and form a normalized similarity value. It is shown that it is possible to improve the exact localization of objects. One of the functions is robust to a linear change of analysed features.

2 Similarity functions based on ratios or spatial relations between the pixels

To compare two images $O = \{o_{ij}\}$ and $B = \{b_{ij}\}$, $N \times N$ size, several similarity functions are often used. There exist some known functions, which calculate the ratios of compared features and are applied to estimate the similarity in other applied areas, e.g., environmental science [6]. Relationship calculation between descriptors will better emphasize local differences compared to subtraction. Among such functions are Wave Hedges, Ruzicka, and Czekanowski.

The Wave Hedges function obtains an unnormalized value by calculating the ratio between the minimum and maximum values of all pairs with further summation of the obtained results over the entire space of compared features and can be calculated for a pair of images as:

$$S^{WHs} = \sum_{i=1}^N \sum_{j=1}^N \left(1 - \frac{\min(o_{ij}, b_{ij})}{\max(o_{ij}, b_{ij})} \right) \quad (1)$$

The Ruzicka function differs from the previous one in that the ratio is calculated not between each pair of compared features, but only once after the summation of their minimum and maximum values:

$$S^{Rzk} = \frac{\sum_{i=1}^N \sum_{j=1}^N \min(o_{ij}, b_{ij})}{\sum_{i=1}^N \sum_{j=1}^N \max(o_{ij}, b_{ij})} \quad (2)$$

A specific trait of the Czekanowski function is the calculation of differences and sums for the corresponding pairs of features with further summation of obtained results and the estimation of the ratio between their final values. For compared images, this function can be written as:

$$S^{Cz} = \frac{\sum_{i=1}^N \sum_{j=1}^N |o_{ij} - b_{ij}|}{\sum_{i=1}^N \sum_{j=1}^N (o_{ij} + b_{ij})} \quad (3)$$

In [7] two similarity functions relationship based on calculation between the minimum and maximum values for all pairs of analyzed features are presented:

– normalized minimax additive similarity p-function ($MMADD^p$):

$$S^{MMADD^p} = \frac{1}{NN} \sum_{i=1}^N \sum_{j=1}^N \frac{\min(o_{ij}^p, b_{ij}^p)}{\max(o_{ij}^p, b_{ij}^p)}, p \in \mathbb{Z}, p \geq 1 \quad (4)$$

– normalized zero-mean minimax additive similarity p-function ($ZMMADD^p$):

$$S^{ZMMADD^p} = \begin{cases} \frac{1}{NN} \sum_{i=1}^N \sum_{j=1}^N \left(\frac{b_{ij} - \bar{b}}{o_{ij} - \bar{o}} \right)^2, & \text{if } |o_{ij} - \bar{o}| \geq |b_{ij} - \bar{b}| \\ \frac{1}{NN} \sum_{i=1}^N \sum_{j=1}^N \left(\frac{o_{ij} - \bar{o}}{b_{ij} - \bar{b}} \right)^2, & \text{if } |o_{ij} - \bar{o}| < |b_{ij} - \bar{b}| \end{cases} \quad (5)$$

where \bar{o} - image O mean value; \bar{b} - image B mean value.

Minimum or maximum attribute is necessary to determine when searching a relationship between them, therefore, the proposed functions are called normalized minimax similarity functions. High resistance to noise achieved through the use of the summation when obtaining a complex normalized value.

To decrease the sensitivity to the deformation and displacement of images, it is reasonable to take into account not only the brightness levels of corresponding pixels, but also the spatial distance between them. The authors [8] propose the approach, in which the function of the distance between all the pixels is calculated and used to determine the Euclidian distance for compared images:

$$S^{IMED} = 1 - \sqrt{\sum_{i=1}^{NN} \sum_{j=1}^{NN} g_{ij} (o_i - b_i)(o_j - b_j)} \quad (6)$$

where $G = \{g_{ij}\}$ is the symmetric matrix, which has a size of $NN \times NN$ and is found by calculating the distances d between pixels P_i and P_j , which monotonically decrease with increasing distance, and the elements of this matrix are determined with the Gaussian function:

$$g_{ij} = \frac{1}{2\pi\sigma^2} \left(\frac{-d(P_i, P_j)^2}{2\sigma^2} \right) \quad (7)$$

It is obvious that this approach appreciably increases computational efforts. With this distance measure, smaller deformation causes smaller changes in the distance.

In the paper [9], it is proposed to take into account the spatial distances between pixels to find the normalized correlation function:

$$S^{IMNCC} = \frac{\sum_{i=1}^{NN} \sum_{j=1}^{NN} g_{ij} o_i b_j}{\sqrt{\sum_{i=1}^{NN} \sum_{j=1}^{NN} g_{ij} o_i o_j \sum_{i=1}^{NN} \sum_{j=1}^{NN} g_{ij} b_i b_j}} \quad (8)$$

The normalized averaged correlation function with consideration for the distance between pixels is:

$$S^{IMZNCC} = \frac{\sum_{i=1}^{NN} \sum_{j=1}^{NN} g_{ij} (o_i - \bar{o})(b_i - \bar{b})}{\sqrt{\sum_{i=1}^{NN} \sum_{j=1}^{NN} g_{ij} (o_i - \bar{o})(o_j - \bar{o}) \sum_{i=1}^{NN} \sum_{j=1}^{NN} g_{ij} (b_i - \bar{b})(b_j - \bar{b})}} \quad (9)$$

IMED, IMNCC and IMZNCC functions are smoother in comparison with the basic ones.

3 New similarity function based on ratio and distance

A distinctive trait of the proposed two similarity functions is the calculations of the ratios between the minimum and maximum values for all the pair pixels and distance calculation between pixels. We use locally weighted normalized ratios for each pixel pair, where the weight is decreases with the increasing distance from the current pixel. If we change the elements of the matrix G , then the influence of distances on the result of similarity will also change.

Then the image normalized minimax additive similarity function (IMMMADD^P) is calculated as:

$$S^{IMMADD^p} = W^{-1} \sum_{i=1}^{NN} \sum_{j=1}^{NN} g_{ij} \frac{\min(o_i^p, b_i^p) \min(o_j^p, b_j^p)}{\max(o_i^p, b_i^p) \max(o_j^p, b_j^p)} \quad (10)$$

where g_{ij} is defined as in IMED; W is the normalizing coefficient determined as:

$$W = \sum_{i=1}^{NN} \sum_{j=1}^{NN} g_{ij} \quad (11)$$

The normalized minimax averaged additive similarity function (IMZMMADD^p) taking into account the distances between features is calculated as:

$$S^{IMZMMADD^p} = W^{-1} \begin{cases} \sum_{i=1}^{NN} \sum_{j=1}^{NN} g_{ij} \left(\frac{b_i - \bar{b}}{o_i - \bar{o}} \frac{b_j - \bar{b}}{o_j - \bar{o}} \right)^p, \text{ if } |o_i - \bar{o}| \geq |b_i - \bar{b}| \ \& \ |o_j - \bar{o}| \geq |b_j - \bar{b}| \\ \sum_{i=1}^{NN} \sum_{j=1}^{NN} g_{ij} \left(\frac{b_i - \bar{b}}{o_i - \bar{o}} \frac{o_j - \bar{o}}{b_j - \bar{b}} \right)^p, \text{ if } |o_i - \bar{o}| \geq |b_i - \bar{b}| \ \& \ |b_j - \bar{b}| > |o_j - \bar{o}| \\ \sum_{i=1}^{NN} \sum_{j=1}^{NN} g_{ij} \left(\frac{o_i - \bar{o}}{b_i - \bar{b}} \frac{o_j - \bar{o}}{b_j - \bar{b}} \right)^p, \text{ if } |b_i - \bar{b}| > |o_i - \bar{o}| \ \& \ |b_j - \bar{b}| > |o_j - \bar{o}| \\ \sum_{i=1}^{NN} \sum_{j=1}^{NN} g_{ij} \left(\frac{o_i - \bar{o}}{b_i - \bar{b}} \frac{b_j - \bar{b}}{o_j - \bar{o}} \right)^p, \text{ if } |b_i - \bar{b}| > |o_i - \bar{o}| \ \& \ |o_j - \bar{o}| \geq |b_j - \bar{b}| \end{cases} \quad (12)$$

It is obvious that the presented functions are universal, as they make it possible to find a normalized value from (0) to (1) for randomly selected characteristics of a pair of images for IMMADD^p and value from (-1) to (1) for IMZMMADD^p. Table 1 shows a comparison of computational costs for similarity functions for images with size $N \times N$ pixels.

Presented functions are universal, as IMMADD^p returns a normalized value from (0) to (1) and IMZMMADD^p returns a value from (-1) to (1) for randomly selected features of a pair of images. Table 1 shows a comparison of computational costs for similarity functions for images with size $N \times N$ pixels.

Two sets are considered as an example: image B 4×4 size and its fragment O 3×4 size. For this data there are two combinations: coincidence and relative displacement by one pixel. Thus, in the first case, the maximum value of the functions is expected. In the second case, the function value will show its resistance to displacement.

For this object O , the matrix G will be an array of 16×16 elements, as the square of the larger side of the object O . Normalizing coefficient $W = 67,285$.

$$O = \begin{bmatrix} 120 & 120 & 120 \\ 120 & 120 & 120 \\ 100 & 100 & 100 \\ 80 & 80 & 80 \end{bmatrix} \quad B = \begin{bmatrix} 120 & 120 & 120 & 140 \\ 120 & 120 & 120 & 140 \\ 100 & 100 & 100 & 120 \\ 80 & 80 & 80 & 100 \end{bmatrix}$$

$$G = \begin{bmatrix} 1 & 0.606 & 0.135 & 0.011 & 0.606 & 0.367 & 0.082 & 0.006 & 0.135 & 0.082 & 0.018 & 0.001 & 0.011 & 0.006 & 0.001 & 10^{-4} \\ 0.606 & 1 & 0.606 & 0.135 & 0.367 & 0.606 & 0.367 & 0.082 & 0.082 & 0.135 & 0.082 & 0.018 & 0.006 & 0.011 & 0.006 & 0.001 \\ 0.135 & 0.606 & 1 & 0.606 & 0.082 & 0.367 & 0.606 & 0.367 & 0.018 & 0.082 & 0.135 & 0.082 & 0.001 & 0.006 & 0.011 & 0.006 \\ 0.011 & 0.135 & 0.606 & 1 & 0.006 & 0.082 & 0.367 & 0.606 & 0.001 & 0.018 & 0.082 & 0.135 & 10^{-4} & 0.001 & 0.006 & 0.011 \\ 0.606 & 0.367 & 0.082 & 0.006 & 1 & 0.606 & 0.135 & 0.011 & 0.606 & 0.367 & 0.082 & 0.006 & 0.135 & 0.082 & 0.018 & 0.001 \\ 0.367 & 0.606 & 0.367 & 0.082 & 0.606 & 1 & 0.606 & 0.135 & 0.367 & 0.606 & 0.367 & 0.082 & 0.082 & 0.135 & 0.082 & 0.018 \\ 0.082 & 0.367 & 0.606 & 0.367 & 0.135 & 0.606 & 1 & 0.606 & 0.082 & 0.367 & 0.606 & 0.367 & 0.018 & 0.082 & 0.135 & 0.082 \\ 0.006 & 0.082 & 0.367 & 0.606 & 0.011 & 0.135 & 0.606 & 1 & 0.006 & 0.082 & 0.367 & 0.606 & 0.001 & 0.018 & 0.082 & 0.135 \\ 0.135 & 0.082 & 0.018 & 0.001 & 0.606 & 0.367 & 0.082 & 0.006 & 1 & 0.606 & 0.135 & 0.011 & 0.606 & 0.367 & 0.082 & 0.006 \\ 0.082 & 0.135 & 0.082 & 0.018 & 0.367 & 0.606 & 0.367 & 0.082 & 0.606 & 1 & 0.606 & 0.135 & 0.367 & 0.606 & 0.367 & 0.082 \\ 0.018 & 0.082 & 0.135 & 0.082 & 0.082 & 0.367 & 0.606 & 0.367 & 0.135 & 0.606 & 1 & 0.606 & 0.082 & 0.367 & 0.606 & 0.367 \\ 0.001 & 0.018 & 0.082 & 0.135 & 0.006 & 0.082 & 0.367 & 0.606 & 0.011 & 0.135 & 0.606 & 1 & 0.006 & 0.082 & 0.367 & 0.606 \\ 0.011 & 0.006 & 0.001 & 10^{-4} & 0.135 & 0.082 & 0.018 & 0.001 & 0.606 & 0.367 & 0.082 & 0.006 & 1 & 0.606 & 0.135 & 0.011 \\ 0.006 & 0.011 & 0.006 & 0.001 & 0.082 & 0.135 & 0.082 & 0.018 & 0.367 & 0.606 & 0.367 & 0.082 & 0.606 & 1 & 0.606 & 0.135 \\ 0.001 & 0.006 & 0.011 & 0.006 & 0.018 & 0.082 & 0.135 & 0.082 & 0.082 & 0.367 & 0.606 & 0.367 & 0.135 & 0.606 & 1 & 0.606 \\ 10^{-4} & 0.001 & 0.006 & 0.011 & 0.001 & 0.018 & 0.082 & 0.135 & 0.006 & 0.082 & 0.367 & 0.606 & 0.011 & 0.135 & 0.606 & 1 \end{bmatrix}$$

For this data there are coincidence and relative displacement by one pixel. For this data, it is possible to coincide O with first subimage of B on the left or second subimage after shifting to left one pixel. Thus, in the first case, the maximum value of the functions is expected. The function value for second subimage should differ as much as possible from maximal value and for this example ($p=1$): $IMED=0,972$; $IMNCC=0,999$; $IMZNCC=0,909$; $MMADD^p=0,9456$; $ZMMADD^p=0,465$; $IMMMADD^p=0,906$; $IMZMMADD^p=0,23$. We see that proposed similarity functions can allow us to obtain lowest similarity values when shift in image by one pixel object.

Table 1 shows that the consideration for the spatial distance between features leads to an appreciable increase in computational expenditures from $O(N^2)$ to $O(N^4)$ for all the functions, which use a similar approach.

Table 1. Computational costs in estimating the similarity between two images

Function type	Number of addition / subtraction operations	Number of multiplication / division operations	Number of comparison operations
Wave Hedges	$N^2+N(N-1)$	N^2	N^2
Ruzika	$2N(N-1)$	N^2	-
Czekanowski	$2N(N-1)$	1	N^2
IMED	$2N^2+N^2(N^2-1)+1$	$2N^4+1$	-
IMNCC	$3N^2(N^2-1)$	$6N^4+3$	-
IMZNCC	$3N^2(N^2-1)+6N^2$	$6N^4+7$	
$MMADD^p$ ($p=2$)	$N(N-1)$	$3N^2+2$	N^2
$ZMMADD^p$ ($p=1$)	$2N^2+3N(N-1)$	N^2+4	N^2
$IMMMADD^p$ ($p=1$)	$N^2(N^2-1)$	$2N^4+2N^2+3$	$2N^2$
$IMZMMADD^p$ ($p=1$)	$N^2(N^2-1)+2N(N-1)$	$2N^4+2N^2+6$	N^2

4 Experimental results

For the similarity functions analytical assessment when detecting and localizing objects in image the following parameters are used:

- function value calculated for object and subimage (A – main peak). Value should aim to 1;
- main peak variance (D_A). The value should aim to zero, meaning smaller value deviations A from expected value;
- amplitude coefficient (Q). This parameter will provide the possibility to determine the maximum amplitude outlier of the similarity function with respect to the mean-square value S_{ms} when comparing the object O with all subimages:

$$Q = \frac{S_{\max}}{S_{ms}} \quad (13)$$

where S_{\max} is the maximum value of the similarity function;

- function maximum value from all side peak (S_L) can be used to threshold T determine, which should be less than T .
- maximum side peak variance (D_{s_L}) should also be as less as possible;
- side peak number at levels higher than 0.95 (N_{s_L}). The parameter can be used to evaluate the possible false-positive detection results number if the threshold value is incorrectly selected.

All parameters were calculated for 20 different images without noise and distortion of 150×150 size. For each of them 20 reference objects of 15×15 size were used with obtained values averaging. Fig. 1 shows some examples of images and object.

Values $A=1$ and $D_A=1$ are obtained for all similarity functions. The resulting values of the remaining parameters are given in Table 2.

Table 2. Function characteristics

Function type	Q	S_L	D_{s_L}	N_{s_L}
Wave Hedges	1,51715	0,859	0,00055	277
Ruzika	1,52919	0,85987	0,00077	256
Czekanowski	1,31457	0,86406	0,00152	1644
IMED	1,19798	0,94581	0,00027	554
IMNCC	1,06136	0,9921	0,00005	10396
IMZNCC	4,20704	0,86102	0,00293	3
MMADD ^p (p=2)	2,16488	0,64973	0,00496	263
ZMMADD ^p (p=1)	11,54347	0,47403	0,01489	6
IMMMADD ^p (p=2)	2,22317	0,7404	0,00133	63
IMZMMADD ^p (p=1)	7,63841	0,3079	0,01148	337

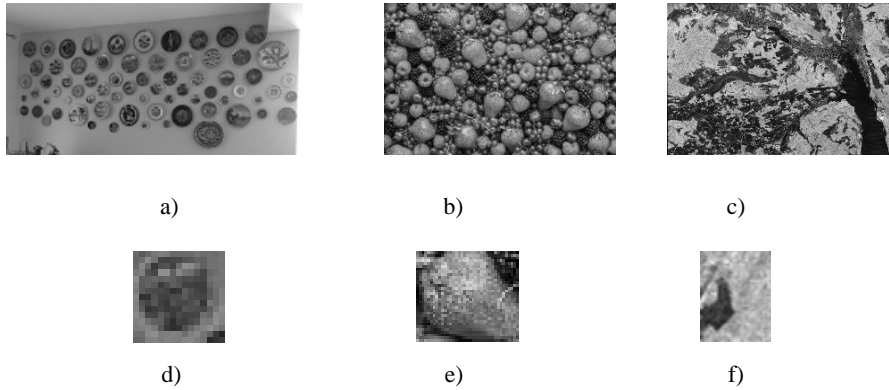


Fig. 1. Some test images: (a-c) input images; (d-f) object templates

The dependences of the similarity value on the shift of objects with respect to the central location for the studied functions are shown in Fig. 2. These values were found by averaging the results for 10 templates from the image shown in Fig. 1a.

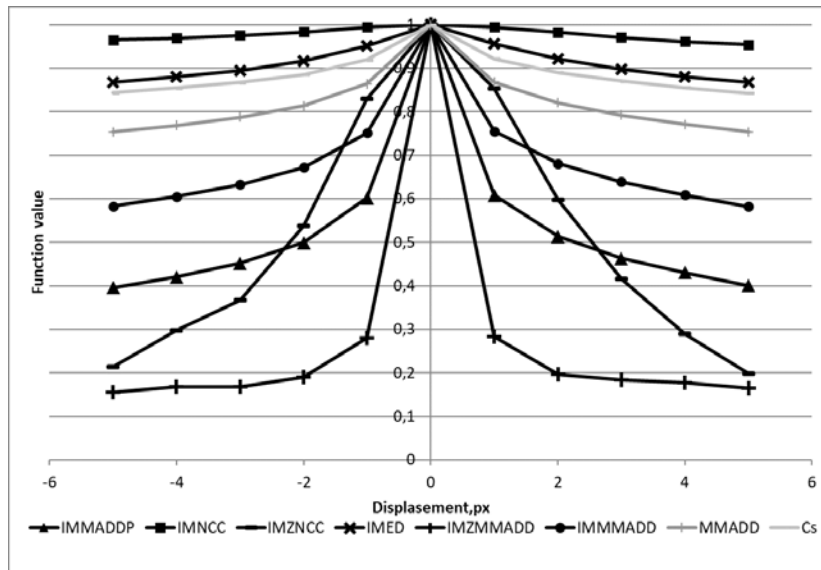


Fig. 2. Dependence of the similarity functions values on the shift of objects

As a result, the analysis of the Table 2 and Fig. 2 shows that the proposed $IMMMADD^P$ and $IMZMMADD^P$ functions have the best characteristics for accurately determining object coordinates in images. For them, amplitude coefficient is the largest among all functions, and function maximum value from all side peak is the minimum. Other parameters are satisfactory and less important. This means that function level is more likely to not exceed the specified threshold and localization will be more accurate.

Some 2D similarity functions for image (Fig. 1a) and object (Fig. 1d) are presented in Fig. 3.

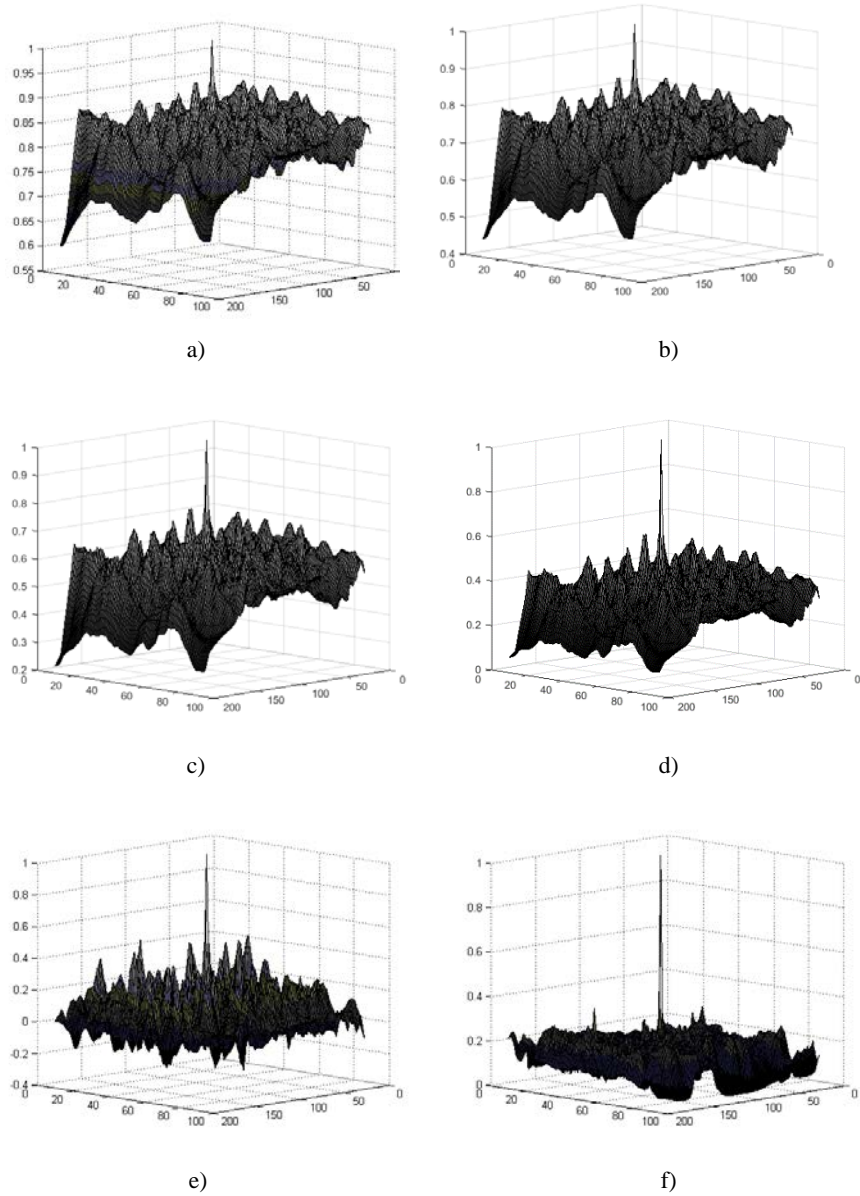


Fig. 3. Object and all subimages similarity for Fig. 1a and 1d based on: (a) Ceanowski function; (b) Wave Hedges function; (c) MMADD^P; (d) IMMADD^P; (e) ZMMADD^P; (f) IMZMMADD^P

Visual analysis of Fig. 3e and 3f shows that $IMZMMADD^P$ main peak is more contrasting than $ZMMADD^P$ main peak, however, according to Table 2, and $ZMMADD^P$ has a larger Q value. This contradiction is explained by the behavior of the functions outside of the object location: $ZMMADD^P$ function has wave-like shape, which reduces its square value, and $IMZMMADD^P$ save uniformity in almost the all range of definition.

5 Conclusion

Two new similarity functions for object detection and localization in image and video based on ratio and distance calculation between pixels is presented. Functions have good characteristics for accurately determining object coordinates in images. One limitation of is that consideration for the spatial distance between features leads to an appreciable increase in computational costs. Our future directions for this area research are construction efficient algorithm to reduce computations and estimate of stability of functions to noise.

References

1. Tschepnakis G., Xirouhakis Y., Delopoulos A. Main Mobile Object Detection and Localization in Video Sequences Advances in Visual Inf. Syst. Lect. Not. in Comp. Sc. (1929) pp. 84-95 (2000)
2. Long Y., Gong Y., Xiao Z., Liu Accurate Q.: Object Localization in Remote Sensing Images Based on Convolutional Neural Networks IEEE Trans. on Geosc. and Remote Sens. May 55 (5) pp. 2486-2498 (2017)
3. Choudhuri S., Das N., Sarkhel R. and Nasipuri M.: Object Localization on Natural Scenes: A Survey Int. J. Patt. Recogn. Artif. Intell. 32 (2) (2017)
4. Lv G.: A novel similarity measure for matching local image descriptors IEEE Access 6 pp 55315–25 (2018)
5. Kim J., Hyun Ch., Han H., Kim H.: Evaluation of Matching Costs for High-Quality Sea-Ice Surface Reconstruction from Aerial Images Remote Sensing 11 (9) pp 1055-72 (2019)
6. M. M. Deza and E. Deza, "Encyclopedia of Distances", Berlin, 2009.
7. Bohush R., Ablameyko S., Adamovskiy Y. Robust object detection in images corrupted by impulse noise Workshop on Computer Modeling and Intelligent Systems, Zaporizhzhia, Ukraine, April 27- May 1 (2020)
8. Wang L., Zhang Y., Feng J. On the Euclidean Distance of Images IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (8) pp 1334–39 (2005)
9. Nakhmani A., Tannenbaum A.: New Distance Measure Based on Generalized Image Normalized Cross-Correlation for Robust Video Tracking and Image Recognition Pattern Recogn. Lett. 34 (3) pp 315–21 (2013)