INFORMATICA, 2015, Vol. 26, No. 2, 241–258

DOI: http://dx.doi.org/10.15388/Informatica.2015.47

A Normalized Parameter for Similarity/Dissimilarity Characterization of Sequences

Algis DŽIUGYS, Robertas NAVAKAS*, Nerijus STRIŪGAS
Laboratory of Combustion Processes, Lithuanian Energy Institute
Breslaujos 3, 44403 Kaunas, Lithuania
e-mail: robertas.navakas@lei.lt

Received: April 2014; accepted: December 2014

Abstract. We propose a normalized parameter for characterization of similarity/dissimilarity of two sequences providing a smoothly varying measure for varying symmetry score. Such a parameter can be used for analysis of experimental data and fitting to a theoretical model, mirror symmetry estimation with respect to a selected or presumed symmetry axis, in particular, in symmetry detection applications where the selected symmetry parameters must be evaluated multiple times. We compare the proposed parameter, as well as several of the well-known distance and similarity measures, on an ensemble of template functions morphing continuously from symmetric to antisymmetric shape. This comparison allows to evaluate different similarity and symmetry measures in a more controlled and systematic setting than a simple visual estimation in sample images.

Key words: distance measure, similarity measure, dissimilarity measure, symmetry, antisymmetry, sequence similarity, symmetry quantification.

1. Introduction

Distance and similarity measures between data and patterns are the key elements in a variety of applications in different fields, including classification and clustering (Finch, 2005; Stanujkic *et al.*, 2013), signal analysis, image analysis and retrieval (Androutsos *et al.*, 1998; Baušys and Kriukovas, 2012; Berman and Shapiro, 1997), speech recognition (Akila and Chandra, 2013; Lileikytė and Telksnys, 2013), time series analysis (Vlachos *et al.*, 2003; Ding *et al.*, 2008), probability density functions (Cha, 2007). Comprehensive surveys of various distance measures are included in Eiter and Mannila (1997), Monev (2004), Cha (2007, 2008), Seung-Seok *et al.* (2010). Certain applications, e.g., image retrieval from databases, require multiple calculations of distance measures. In these cases, the distance measures that are not computationally expensive are desirable (Berman and Shapiro, 1997). Selection of the proper distance measure is largely application dependent, e.g., how sensitive should it be to small deviations from perfect similarity, should it be

*Corresponding author.

invariant under certain transformations of the considered sequences, like rotations of one of the function pairs (Vlachos *et al.*, 2004), etc.

Distance and similarity measures are closely related to evaluation and detection of symmetries in patterns, objects and images. Symmetry is a distinct feature of many natural and artificial objects that can be exploited for multiple applications: visual perception and structure identification in both natural (Cohen and Zaidi, 2013; Treder, 2010; Machilsen *et al.*, 2009) and artificial vision (Wagemans, 1995; Gesu *et al.*, 2010; Liu *et al.*, 2010), medical imaging and diagnostics (Stegmann *et al.*, 2005; Mancas *et al.*, 2005; Ruppert *et al.*, 2011), traffic analysis (Zielke *et al.*, 1992), reconstruction of 3D shapes (Basri and Moses, 1999; Shimshoni *et al.*, 2000) and scenes (Köser *et al.*, 2011) from a single image.

Optical tomography is an attractive tool for remote analysis in various fields (Haisch, 2012). The inverse problem of reconstruction of 3D structures from the 2D images obtained by the imaging system might benefit from the symmetries of the object of interest; in particular, for the axially symmetric objects, a single projection is sufficient for reconstruction (Bracewell, 1956; Deans, 2000). The projection in this case exhibits the mirror symmetry, therefore, the degree of symmetry of 2D images thus obtained must be estimated before attempting reconstruction.

Formally, a symmetry means self-similarity under certain transformations, such as translations, rotations and reflections. Detection of symmetry involves comparison between the area of interest and its counterpart under the respective transformation, therefore, the appropriate similarity or distance score must be defined.

Here, we analyze the notion of symmetry as applicable to 2D images. A large class of symmetry analysis, such as analysis of 3D shapes (Kazhdan *et al.*, 2004; Martinet *et al.*, 2006), might be reduced to analysis of 2D images in case of application of imaging and optical sensing techniques. Many image processing and detection applications deal exclusively with 2D images.

Given the importance of symmetry detection in image processing and analysis applications, in most reported cases, the methods of symmetry detection and thus the underlying symmetry estimation criteria are tested using certain selected images (e.g., photographs), and the detected symmetrical features are then evaluated visually. For more systematic estimation, a specifically constructed function (equivalent to an abstract image) with predetermined symmetry properties might provide a formal template for comparison of symmetry scores. We propose a normalized similarity parameter, with values in the range $[-1, 1]$, for a pair of sequences providing a symmetry score that varies smoothly with continuously varying symmetry degree and compare it to a number of known similarity/distance measures using a test function with continuously varying symmetry degree.

2. Similarity of 1D Functions

2.1. *Distance and Similarity Measures*

Distance or similarity measures considered here are defined for two sequences or vectors $\mathbf{g} = \{g_i\}$, $\mathbf{h} = \{h_i\}$ containing n elements, $1 \leqslant i \leqslant n$. These definitions apply as well for

discrete or continuous functions $g_i = f(x_i)$, $g = f(x)$ respectively, where in the latter case both functions g and h are defined in the same range $x \in [x_1, x_n]$. Symmetry applications considered here (estimation of mirror symmetry) are concerned with comparing one part of a sequence, such as a row of pixels in an image, located on one side of the potential symmetry axis, to its counterpart on another side of the symmetry axis, mirrored over the symmetry axis. In the analysis below, we therefore treat the symmetry estimation analogously to comparison of two sequences. Since evaluation of symmetry might be reduced to comparison of similarity of the potentially symmetric counterparts of a 1D function, a number of the standard distance measures can be considered. Most of these measures have been applied in a variety of fields:

- Minimum ratio (Goshtasby, 2012):

$$m_r = \frac{1}{n}\sum_i r_i \tag{1}$$

  with $r_i = \min\left(\frac{h_i}{g_i}, \frac{g_i}{h_i}\right)$;
- Canberra distance:

$$d_{\text{canberra}} = \frac{1}{n}\sum_i \frac{|g_i - h_i|}{|g_i| + |h_i|}; \tag{2}$$

- Bray–Curtis distance:

$$d_{\text{bc}} = \frac{\sum_i |g_i - h_i|}{\sum_i ||g_i| + |h_i||}; \tag{3}$$

- Pearson correlation:

$$r = \frac{\sum_i (g_i - \langle \mathbf{g} \rangle)(h_i - \langle \mathbf{h} \rangle)}{\sqrt{\left[\sum_i (g_i - \langle \mathbf{g} \rangle)^2\right]\left[\sum_i (h_i - \langle \mathbf{h} \rangle)^2\right]}}, \tag{4}$$

  $\langle \cdot \rangle$ means averaging over the function range, i.e., either $\langle \mathbf{v} \rangle = \frac{1}{n}\sum_{i=1}^{n} v_i$ or $\langle v_i \rangle = \frac{1}{n}\sum_{i=1}^{n} v_i$, depending on whether the expression inside the brackets operates on the whole vector or element-wise;
- Pearson's absolute value dissimilarity:

$$d_{\text{p}} = \sqrt{\frac{n}{n-1}\left[d_{\text{ne}}(g,h)^2 - q(g,h)^2\right]}, \tag{5}$$

  where $d_{\text{ne}}(x, y) = \frac{1}{\sqrt{n}}\sum_i (g_i - h_i)^2$ is the normalized Euclidean distance, $q(g,h) = \frac{1}{n}\left(\sum_i g_i - \sum_i h_i\right)$;
- Extended Jaccard distance:

$$d_{\text{ej}} = \frac{\mathbf{g} \cdot \mathbf{h}}{\|\mathbf{g}\|^2 + \|\mathbf{h}\|^2 - \mathbf{g} \cdot \mathbf{h}}, \tag{6}$$

  here $\|\mathbf{v}\| = \sqrt{\sum_i v_i^2}$ stands for the vector norm;

- Cosine:

$$d_{\cos} = \frac{\mathbf{g} \cdot \mathbf{h}}{\|\mathbf{g}\| \cdot \|\mathbf{h}\|}. \tag{7}$$

For the purposes of the image analysis and symmetry estimation, a number of more specialized measures have been proposed.

Zielke *et al.* (1992) define the symmetry measure

$$S(x_l, w) = \frac{\int E_n(x, x_l, w)^2\,dx - \int O(x, x_l, w)^2\,dx}{\int E_n(x, x_l, w)^2\,dx + \int O(x, x_l, w)^2\,dx}, \tag{8}$$

with values $S = 1$ for ideal symmetry, $S = 0$ for asymmetry and $S = -1$ for ideal antisymmetry. This measure is based on decomposition of the considered function $G(x)$ into the even and odd components

$$\begin{aligned} E(x, x_l, w) &= \big[G(x - x_l) + G(x_l - x)\big]/2, \\ O(x, x_l, w) &= \big[G(x - x_l) - G(x_l - x)\big]/2 \end{aligned} \tag{9}$$

defined at the interval $|x - x_l| \leqslant w/2$ ($E = O = 0$ for $|x - x_l| > w/2$). The symmetry is evaluated for the selected point of origin x_l and the interval width w. Since the mean value of the odd function is zero, and the even function has, in general, a nonzero value, a normalized even function with the zero mean value is introduced:

$$E_n(x, x_l, w) = E(x, x_l, w) - \frac{1}{w}\int E(x, x_l, w)\,dx. \tag{10}$$

For the case of discrete sequences, this parameter can be rewritten as

$$S_Z = \frac{\sum_i E_{n,i}^2 - \sum_i O_i^2}{\sum_i E_{n,i}^2 + \sum_i O_i^2} \tag{11}$$

with normalized even component of the function

$$E_{n,i} = E_i - \frac{1}{n}\sum_i E_i = E_i - \langle E_i \rangle. \tag{12}$$

Zabrodsky *et al.* (1995) introduce a Symmetry Distance defined as the mean of the squared displacements of each point of an original figure required to make this figure symmetric:

$$\mathrm{SD} = \frac{1}{n}\sum_{i=0}^{n-1} |P_i - \hat{P}_i|^2. \tag{13}$$

Kiryati and Gofman (1998) introduce a symmetry measure based on decomposition of an arbitrary function $f(x) = f_s(x) + f_{as}(x)$, $x \in [-L, L]$, into the symmetric and asymmetric parts, respectively:

$$S\{f\} = \frac{|f_s|^2}{|f_s|^2 + |f_{as}|^2} \tag{14}$$

and the reflectional correlation coefficient

$$C\big\{f(x)\big\} = \frac{\int_{-L}^{L} f(x) f(-x) dx}{\int_{-L}^{L} f^2(x) dx} = 2S\big\{f(x)\big\} - 1. \tag{15}$$

Based on these parameters, a probabilistic genetic algorithm is implemented for symmetry detection by global optimization.

Androutsos *et al.* (1998) investigate a few vector distance measures for image retrieval from a database based on image color differences; among them, the Czekanowski coefficient

$$d_{\text{Czekanowski}} = 1 - \frac{2\sum_k \min(g_k, h_k)}{\sum_k (g_k + h_k)}, \tag{16}$$

where k denotes the vector component; the angular distance

$$\theta = 1 - \frac{2}{\pi} \cos^{-1}\left(\frac{\mathbf{g} \cdot \mathbf{h}}{|\mathbf{g}| \cdot |\mathbf{h}|}\right) \tag{17}$$

and a distance measure combining the angle between vectors and their magnitude difference:

$$d_N = 1 - \theta \cdot \left[1 - \frac{|\mathbf{g} - \mathbf{h}|}{\sqrt{3 \cdot 255^2}}\right]. \tag{18}$$

This measure is constructed for RGB color vectors with 3 components and 255 levels. By definition, these vectors can assume only positive values.

Kazhdan *et al.* (2004) introduce a reflective symmetry descriptor for an arbitrary 3D voxel model with respect to all planes through the model's center of mass, generalized even for the planes that are not the symmetry planes. This parameter can describe the general object shape and identify the symmetry planes or absence thereof.

Loy and Eklundh (2006) propose a symmetry detection method based on grouping the feature points in the image into the symmetric constellations. Symmetry of the feature pairs is characterized based on their relative orientation, scale and distance; the corresponding symmetry measure is then constructed from appropriately weighted partial parameters.

Milner *et al.* (2007) introduce a symmetry parameter specifically for characterization of bifurcating shapes having a tree-like structure. This approach is applied for characterization of leaves of trees growing under different environmental conditions.

Akila and Chandra (2013) introduce a slope finder distance between two vectors **g** and **h**:

$$d_{\mathrm{SF}} = \sum_i \frac{g_{i-1} - g_i}{h_{i-1} - h_i} \tag{19}$$

which is used in Dynamic Time Warping pattern-matching algorithm for speech recognition.

Medical diagnostics techniques, such as MRI or PET, yield 3D images. Symmetries must be estimated with respect to the mid-sagittal planes. The symmetry measures applied to 2D images can be generalized to 3D images; Ruppert *et al.* (2011) use the correlation-based symmetry measure:

$$S = \frac{\sum_i^w \sum_j^h \sum_k^d I^l_{ijk} I^r_{ijk}}{\sqrt{\left(\sum_i^w \sum_j^h \sum_k^d I^l_{ijk} I^l_{ijk}\right) \cdot \left(\sum_i^w \sum_j^h \sum_k^d I^r_{ijk} I^r_{ijk}\right)}}, \tag{20}$$

where w, h, d are width, height and depth of the 3D image, $I^{(l,r)}_{ijk}$ is the binary value $\{0, 1\}$ of a voxel at coordinate (i, j, k) of the half image to the left or right from the mid-sagittal plane, respectively.

It can be noticed that the correlation-based symmetry measures are insensitive to scaling and shifting of a half- function with respect to its symmetric counterpart. Invariance of the distance/similarity measures with respect to certain transformations, such as scaling, rotations etc., as well as insensitivity to noise and distortions, might be a desirable feature for, e.g., image comparison, search and retrieval from databases etc. On the other hand, symmetry estimation must take into account the transformations of a considered sequence with respect to its potentially symmetric counterpart. We therefore consider here the measures that directly compare the analyzed sequences pointwise in their defined ranges.

We seek to introduce a similarity parameter for a pair of sequences; a "sequence" might refer to a function $v_i = f(x_i)$ defined over a certain range $x_{\min} \leqslant x_i \leqslant x_{\max}$ or a vector $\mathbf{v} = \{v_i\}$ having n elements, $1 \leqslant i \leqslant n$. Such parameter must be continuous with the varying degree of similarity of the considered function and allows a fast calculation for real-time applications, potentially, in embedded systems with limited computing resources, or in symmetry detection applications requiring many evaluations of the symmetry parameter. Estimation of symmetry in many cases involves estimation of similarity of two functions, located on both sides of the potential symmetry axis. Therefore, the terms "symmetry" and "similarity" are used somewhat interchangeably in the following study. For performance considerations, it is desirable to avoid calculation of "slow" functions, such as trigonometric functions or roots.

2.2. *Value and Function Similarity*

We start by defining the value difference $s_{p,q}$ for a pair of points p, q with respect to a certain selected line (referred to as a "midline"), Fig. 1. The "points" refer here to the

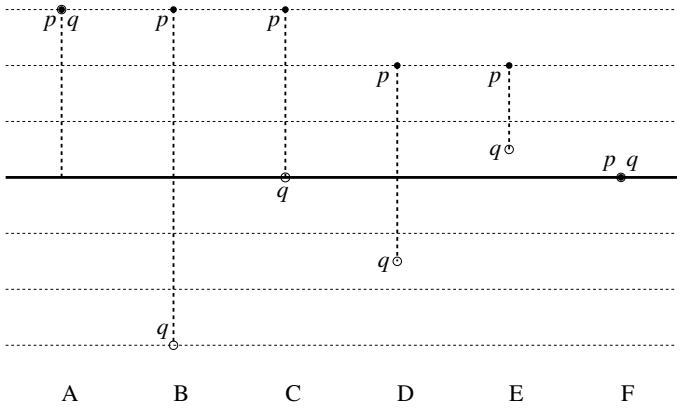


Fig. 1. Different cases of value similarity s_p for points p (solid) and q (hollow) with respect to the midline (solid line): A) p and q coincide, $s_{p,q} = 1$; B) p and q are equidistant from the midline at different sides, $s_{p,q} = -1$; C) one of the points (q) is located on the midline, $s_{p,q} = 0$; D) general case (p, q at different distances from midline), $-1 \leqslant s_{p,q} \leqslant 0$; E) general case, $0 \leqslant s_{p,q} \leqslant 1$, F) both p and q are located on the midline, $s_{p,q} = 1$.

argument–value pairs of the considered functions, such as $p_i = f(x_i)$. In the image analysis context, p, q would mean the values of the selected pixel pair, such as intensity or RGB value, rather than their geometric positions in the image. By definition, for special cases

$$s_{p,q} = \begin{cases} 1, & p \text{ and } q \text{ coincide,} \\ 0, & \text{either } p \text{ or } q \text{ is on the midline (but not both),} \\ -1, & p \text{ and } q \text{ are on the opposite sides of the midline.} \end{cases} \tag{21}$$

In the latter case, $s_{p,q} = -1$ if the points are equidistant from the midline to the opposite sides, i.e., one point is the mirror image of another one over the midline. Using these conventions, the similarity parameter of two points p, q is defined as

$$s_{p,q} = \begin{cases} 1, & p = q = 0, \\ \operatorname{sign}(p) \cdot \operatorname{sign}(q) \cdot \left[1 - \frac{||p|-|q||}{|p|+|q|}\right], & \text{otherwise,} \end{cases} \tag{22}$$

where $\operatorname{sign}(p)$ is the sign function:

$$\operatorname{sign}(p) = \begin{cases} -1, & p < 0, \\ 0, & p = 0, \\ 1, & p > 0. \end{cases} \tag{23}$$

Using the value difference parameter, two sequences $g_i = f_1(x_i)$, $h_i = f_2(x_i)$ can be compared pointwise. Treating the sequences as 1D functions, the midline is naturally parallel to the x axis. Having in mind the configuration shown in Fig. 1, in some cases it is convenient to normalize the sequences in such a way that their midline is located on the

x axis with $f(x) = 0$. This might be implemented by shifting the sequences

$$g' = g - \big(\langle g\rangle + \langle h\rangle\big)/2,$$
$$h' = h - \big(\langle g\rangle + \langle h\rangle\big)/2. \tag{24}$$

The difference parameter for the sequences is now defined as the average of the value differences:

$$S_{\mathbf{g}',\mathbf{h}'} = \frac{1}{n}\sum_i s_{g_i',h_i'}, \tag{25}$$

where $s_{g_i',h_i'}$ is given by Eq. (22). This expression can be applied to continuous functions as well by replacing the sum with the integral.

By definition (25), any function is "antisimilar" to its counterpart obtained by rotating the original function around the midline by the angle π, i.e., $s_{g_i,h_i} = -1$ for any respective value pair g_i, h_i in such a function pair. This property is illustrated in Fig. 2: the zero-order Bessel function $f_1(x) = j_0(x)$ in the range $0 \leqslant x \leqslant 10$ is taken as an original function (this particular function is selected for illustration purposes only, and the approach can be applied to any function in general), and two "antisimilar" functions are constructed: $f_{2,a}(x) = -f_1(-x)$, $f_{2,b}(x) = 2 \cdot \max[f_1(x)] - f_1(x)$. In the latter case, the maximum is calculated in the selected range; this produces reflection of the original function over the axis $y = \max[f_1(x)]$. The function pairs f_1, $f_{2,a}$ and f_1, $f_{2,b}$ are then shifted according to Eq. (24) and the resulting function pairs are shown in Figs. 2(a) and 2(b), respectively. These two configurations might be treated as different types of "antisimilarity" of function pairs.

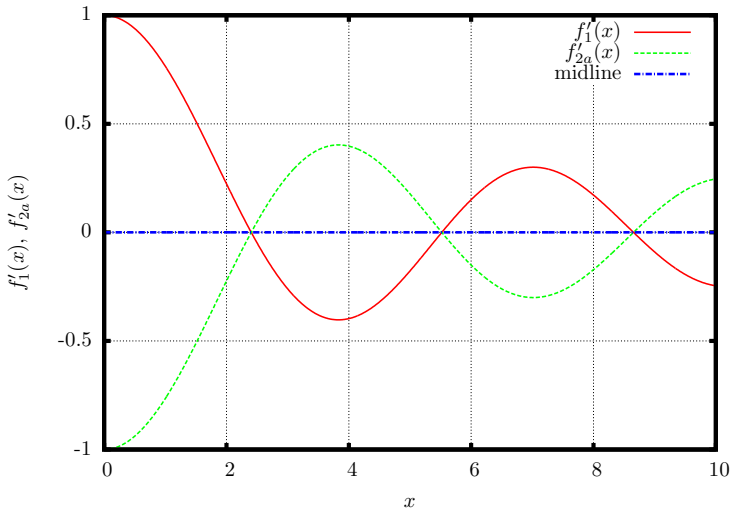
Since the pair of functions to be compared are centered around the midline by expression (24), the similarity parameter defined in this way might be referred to as centered value similarity. Other possibilities for normalization exist; here, we consider the following two normalizations, in addition to (24): value similarity with

$$g' = g,$$
$$h' = h, \tag{26}$$

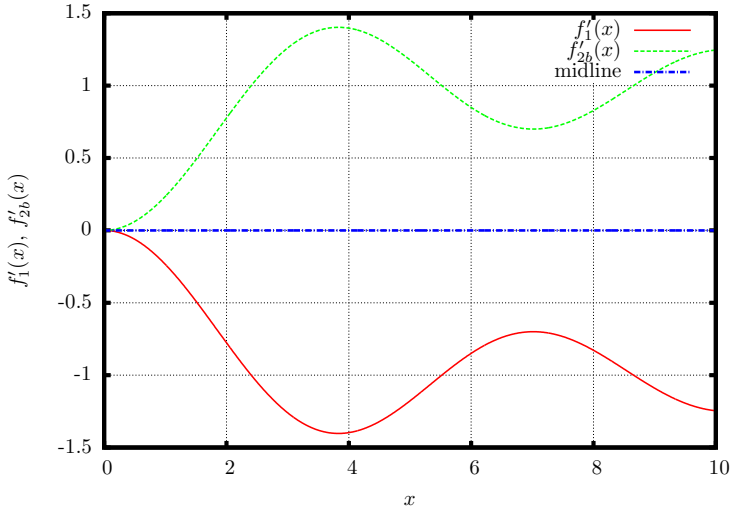
and individually centered value similarity

$$g' = g - \langle g\rangle,$$
$$h' = h - \langle h\rangle. \tag{27}$$

Function similarity is closely related to mirror symmetry, as the symmetry is estimated by comparing the two halves of the considered function located on the opposite sides of the symmetry line. The example of "antisimilar" functions in Fig. 2 can be recast into the case of antisymmetric functions (Fig. 3). By reflecting the functions f_{2a}, f_{2b}, two types



(a) "Type 1" antisimilarity



(b) "Type 2" antisimilarity

Fig. 2. Two types of "antisimilar" functions for which $S_{f_1,f_2} = -1$.

of antisymmetry arise from the respective "antisimilar" functions. The parameter (25) is therefore readily applicable for symmetry estimation.

The parameter based on definition (25) has analogy to the Canberra distance. Its values are in the range $[-1, 1]$, similar to the case of correlation coefficient. The latter, however, is not sensitive to scaling and shifting of the sequences being compared. There are distinct value ranges for similarity and antisimilarity: $S_{\mathbf{g},\mathbf{h}} > 0$ indicate a certain degree of similarity (symmetry), while the values $S_{\mathbf{g},\mathbf{h}} < 0$ indicate a degree of antisimilarity (antisymmetry). An attractive feature of the parameter defined in Eq. (25) is the possibility of quick calculation: it involves only the basic arithmetic operations, without the need for slower

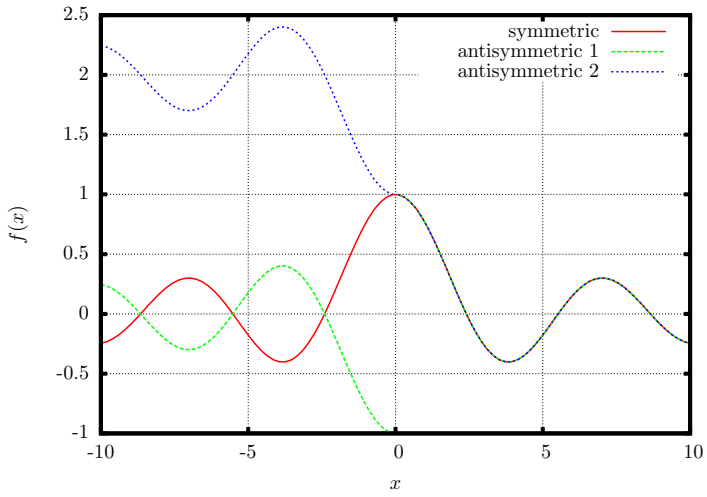


Fig. 3. Examples of antisymmetric functions for which $S_{f_1,f_2} = -1$ corresponding to those in Fig. 2. Two "types" of antisymmetry shown here correspond to the types of antisymmetry shown in Fig. 2.

calculation of functions such as square roots, therefore, it can be readily implemented on a less capable hardware (embedded systems) for real-time operation, especially for applications requiring multiple evaluations of the similarity parameter. This parameter can be extended and generalized by introducing, e.g., weighted averaging in (25), depending on the specific application.

3. A Test Case: Comparison of Similarity Parameters in a Continuous Symmetry Morphing

For image processing applications, outputs of the symmetry detection algorithms are usually estimated visually. For a more quantifiable estimation and comparison of the symmetry parameters, let us analyze a case of continuous morphing from "type 1" antisymmetric 1D function to "type 2" antisymmetric one, where the antisymmetry types are analogous to those shown in Fig. 3, including the symmetric case in between. For this purpose, we define an ensemble of test functions $f_{t,i}(x)$ in the range $x \in [-L_x, L_x]$. Half of the range, $x \in [0, L_x]$, is an expression that is the same for all the functions in the ensemble and serves as a template against which the expression defined on the other half range, $x \in [-L_x, 0]$, is compared. The expression in the range $x \in [-L_x, 0]$ varies in the ensemble. The whole ensemble of test functions $f_t(x)$ is arranged along the y axis, making the surface $f_t(x, y)$ each section of which at $y = y_i$ represents one of the 1D test functions $f_t(x, y_i)$ from the ensemble. They are sorted in such a way that the symmetry of the respective functions vary continuously along the y axis, morphing from "type 1" to "type 2" antisymmetry. For easy visualization, this ensemble of functions can be presented as a 2D surface (Fig. 4), however, symmetry of each section $f_t(x, y_i)$ is treated independently. For

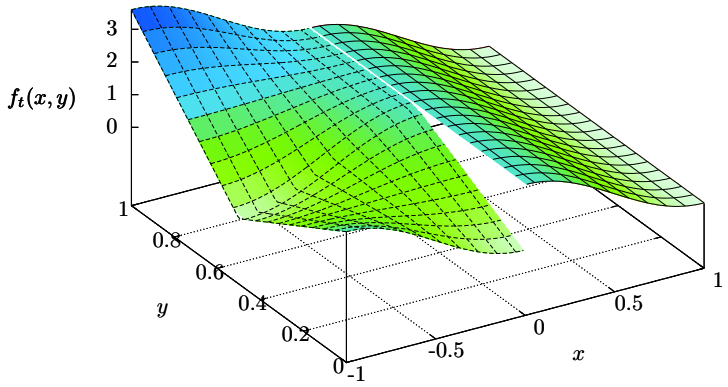


Fig. 4. The testing function with continuous similarity morphing.

each section, the corresponding symmetry parameter is calculated, yielding the variation $S(y)$ in the range $[Y_1, Y_3]$ with the varying asymmetry of the test function section.

The domain of the test function $f_t(x, y)$ is stitched from 4 areas in the x, y plane by combining the ranges $x \in [0, L_x]$, $x \in [-L_x, 0]$, $y \in [Y_1, Y_2]$, $y \in [Y_2, Y_3]$: $[-L_x, 0] \times [Y_1, Y_2] \cup [0, L_x] \times [Y_1, Y_2] \cup [-L_x, 0] \times [Y_2, Y_3] \cup [0, L_x] \times [Y_2, Y_3]$.

Let us define the function on the right-hand side of the area

$$g(x, y) = a \cdot \cos\left(\frac{\pi x}{L_x}\right), \quad x \in [0, L_x],\ y \in [Y_1, Y_3], \tag{28}$$

which is constant along the y axis, but still denoted here as $g(x, y)$ for consistency of notations. The left hand side transfers linearly from the symmetric to antisymmetric counterpart of the function g:

$$h_1(x, y) = \frac{y - \frac{1}{2}(Y_1 + Y_2)}{(Y_2 - Y_1)} \cdot g(x, y), \quad x \in [-L_x, 0],\ y \in [Y_1, Y_2], \tag{29}$$

$$h'(x, y) = -\frac{y - \frac{1}{2}(Y_3 + Y_2)}{(Y_3 - Y_2)} \cdot g(x, y), \quad x \in [-L_x, 0],\ y \in [Y_2, Y_3],$$

$$h_2(x, y) = h'(x, y) + g(0, y) - h'(0, y), \quad x \in [-L_x, 0],\ y \in [Y_2, Y_3]. \tag{30}$$

For brevity, the function comprising (28)–(30) will be denoted as

$$f_t(x, y) = \begin{cases} g(x, y) + b, & x \in [0, L_x], \quad y \in [Y_1, Y_3], \\ h_1(x, y) + b, & x \in [-L_x, 0], \ y \in [Y_1, Y_2], \\ h_2(x, y) + b, & x \in [-L_x, 0], \ y \in [Y_2, Y_3]. \end{cases} \tag{31}$$

The parameter values were selected as follows: the amplitude $a = 1$, the ranges $L_x = 1$, $Y_1 = 0$, $Y_2 = 0.5$, $Y_3 = 1.0$. An additional parameter b was added to explore the influence of a possible shift along the z axis; for results presented below, $b = 0.6 \cdot a$. The surface plot of the testing function f_t is shown in Fig. 4 and its sections at several values of y in Fig. 5.

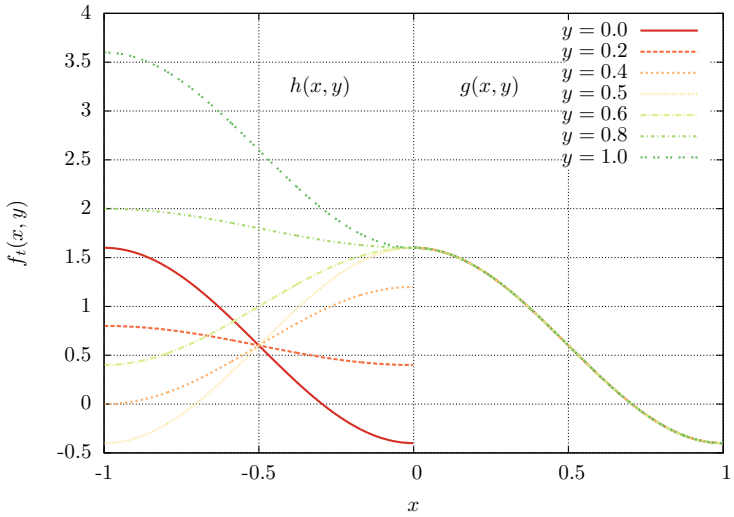


Fig. 5. Profiles of the test function at different y sections.

The symmetry measures are calculated with respect to the $x = 0$ axis for the cross-sections of the test function $y = y_i$, $y \in [y_1, y_3]$. A number of the known symmetry parameters were calculated for the shifted test functions in the ensemble (31). The expressions of the respective parameters have been slightly modified as applicable, so that their values are in the range $[-1, 1]$ to allow for comparison. For this reason, the respective expressions are presented in Table 1; this table lists the standard distance measures in the left-hand column and the modified expressions in the right-hand column. In addition to the parameters listed in Table 1, we also calculate the parameter (11) used by Zielke *et al.* (1992). Its values are in the range $[-1, 1]$, which allows for direct comparison with other parameters.

The resulting variation of the similarity parameters along the y axis are shown in Figs. 6–8. As seen, the correlation does not differentiate between the different degrees of (dis)similarity shown in Fig. 5 and assumes only the limit values -1 or 1, as expected from its analytic expression. Both versions of the minimum ratio parameter (original and modified) are nonmonotonous in the selected range and therefore do not allow for unambiguous estimation of the similarity/symmetry degree. Moreover, the parameters having the values in the range $[0, 1]$ (those based on Pearson absolute dissimilarity, Canberra, Bray–Curtis, extended Jaccard, cosine) do not allow to differentiate between the cases of partial similarity/symmetry and partial antisimilarity/antisymmetry. For such differentiation, the parameters having the values in the range $[-1, 1]$ ($[-1, 0]$ for antisimilarity and $[0, 1]$ for similarity) provide a more sensible estimation. Fig. 8 presents the variation of this type of parameters: Pearson correlation, centered and individually centered value similarities, and the parameter (11) by Zielke *et al.* (1992). Note than in the range $y \in [0, 0.5]$, the centered and individually centered value similarities coincide. Uncentered value similarity parameter is shown for comparison to centered and individually centered value similarity parameters. The behavior of the last three parameters is similar in the analyzed setting. From the application point of view, the performance of their evaluation might be taken into account.

Table 1
Expressions of tested symmetry parameters used for characterization of symmetries in the test function (31). The left-hand column lists the original expression upon which the calculated distance was based, and the right-hand column presents the modified expression used to calculate the symmetries of the test function.

	Basis parameter	Expression, $S_{\mathbf{g},\mathbf{h}} = \ldots$
1	Pearson correlation (Eq. (4))	$\frac{\sum_i (g_i - \langle \mathbf{g} \rangle)(h_i - \langle \mathbf{h} \rangle)}{\sqrt{\left[\sum_i (g_i - \langle \mathbf{g} \rangle)^2\right]\left[\sum_i (h_i - \langle \mathbf{h} \rangle)^2\right]}}$
2	Modified Pearson abs. dissim. (Eq. (5))	$1 - \frac{\sqrt{\frac{n}{n-1}\left[\langle (\mathbf{g}-\mathbf{h})^2 \rangle - \langle \mathbf{g}-\mathbf{h} \rangle^2\right]}}{\sqrt{\langle \mathbf{g}^2 \rangle + \langle \mathbf{h}^2 \rangle}}$
3	Extended Jaccard (Eq. (6))	$\frac{\mathbf{g}\cdot\mathbf{h}}{\|\mathbf{g}\|^2 + \|\mathbf{h}\|^2 - \mathbf{g}\cdot\mathbf{h}}$
4	Cosine (Eq. (7))	$\frac{\mathbf{g}\cdot\mathbf{h}}{\|\mathbf{g}\| \cdot \|\mathbf{h}\|}$
5	Based on Canberra (Eq. (2))	$1 - \frac{1}{n}\sum_i \frac{\lvert g_i - h_i \rvert}{\lvert g_i \rvert + \lvert h_i \rvert}$
6	Based on Bray–Curtis (Eq. (3))	$1 - \frac{\sum_i \lvert g_i - h_i \rvert}{\sum_i (\lvert g_i \rvert + \lvert h_i \rvert)}$
7	Minimum ratio (Eq. (1))	$\frac{1}{n}\sum_i \min\left(\lvert \frac{g_i}{h_i} \rvert, \lvert \frac{h_i}{g_i} \rvert\right)$
8	Minimum ratio, modified (Eq. (1))	$\frac{1}{n}\sum_i \left[\operatorname{sign}(g_i) \cdot \operatorname{sign}(h_i) \cdot \min\left(\lvert \frac{g_i}{h_i} \rvert, \lvert \frac{h_i}{g_i} \rvert\right)\right]$
9	Zielke *et al.* (1992) (Eq. (11))	$\frac{\sum_i E_{n,i}^2 - \sum_i O_i^2}{\sum_i E_{n,i}^2 + \sum_i O_i^2}$, $E_{n,i} = E_i - \frac{1}{n}\sum_i E_i = E_i - \langle E_i \rangle$, $E_i = (g_i + h_i)/2$, $O_i = (g_i - h_i)/2$
10	Value similarity (Eqs. (25), (26))	$\frac{1}{n}\sum_i s_{g',h'}$, $s_{g',h'}$ is given by Eq. (22) $g' = g$, $h' = h$
11	Value similarity centered (Eqs. (25), (24))	$\frac{1}{n}\sum_i s_{g',h'}$, $s_{g',h'}$ is given by Eq. (22) $g' = g - (\langle g \rangle + \langle h \rangle)/2$, $h' = h - (\langle g \rangle + \langle h \rangle)/2$,
12	Value similarity ind. centered (Eqs. (25), (27))	$\frac{1}{n}\sum_i s_{g',h'}$, $s_{g',h'}$ is given by Eq. (22) $g' = g - \langle g \rangle$, $h' = h - \langle h \rangle$

4. Computational Performance of Similarity/Distance Measures

Practical applications usually involve multiple evaluations of the appropriate similarity/distance measures. For numerically-intensive applications, like image analysis, the performance of evaluating the respective functions becomes important. We therefore compared the speeds of computation of some of the measures presented here. The results are presented here for illustration purposes only; they are largely dependent on particular implementations and platforms, besides, as the vectors were assigned random numbers, the calculation times varied slightly for different runs.

The functions calculating the distance measures were implemented in C and Fortran 95 using the GNU compilers `gcc` and `gfortran`, ver. 4.7.2, respectively. The implemen-

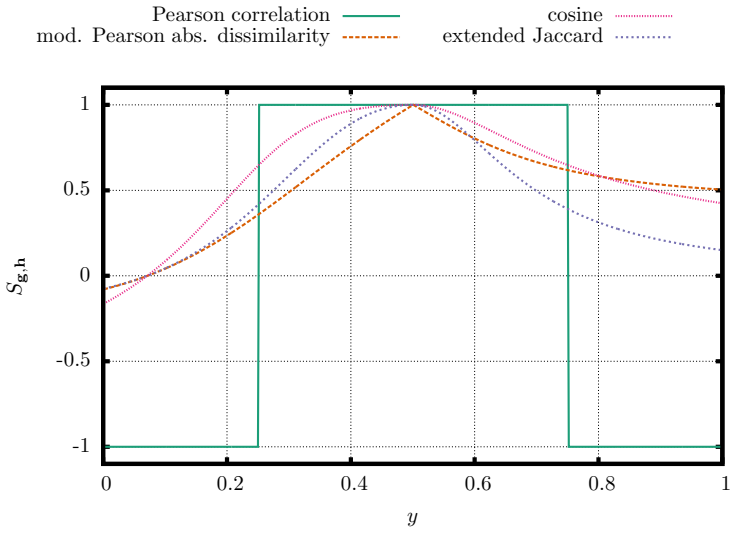


Fig. 6. Variations of function similarity measures $S_{\mathbf{g},\mathbf{h}}$: Pearson correlation, modified Pearson absolute dissimilarity, extended Jaccard distance, cosine distance.

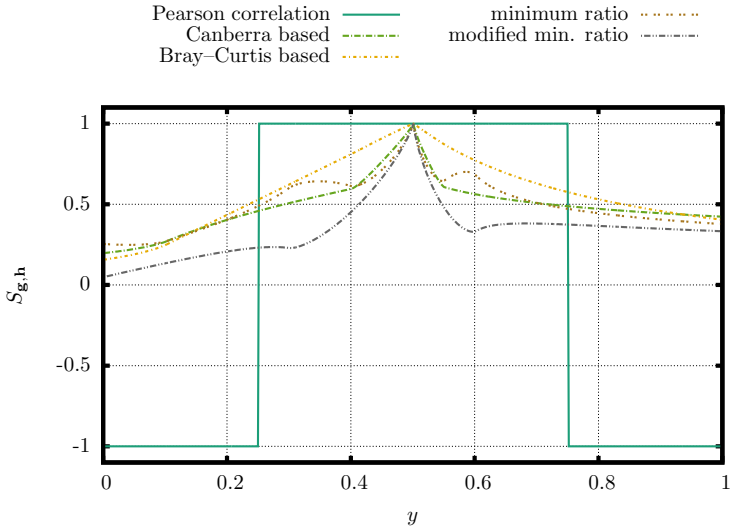


Fig. 7. Variations of function similarity measures $S_{\mathbf{g},\mathbf{h}}$: Pearson correlation, Canberra-based, Bray–Curtis-based, minimum ratio and modified minimum ratio.

tations followed closely the definitions of the respective parameters as listed in Table 1, without attempts to optimize, except for reasonable measures, e.g., using intrinsic vector operations, reusing the computed values, etc. The test program was compiled without the compiler optimization switches (e.g., "-O2" or similar; optimizations would inline the functions and prevent measuring the evaluation time). The parameters were evaluated 1000 times each for the vectors of length $N = 10^6$, filled with random numbers generated by a random number generator. Usually, the typical vectors encountered in image

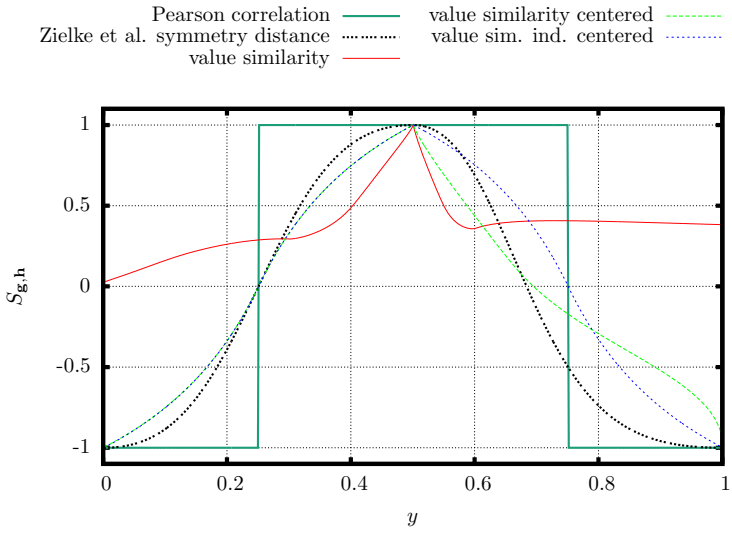


Fig. 8. Variations of function similarity measures $S_{\mathbf{g},\mathbf{h}}$: Pearson correlation, symmetry distance by Zielke et al., value similarity, value similarity centered and value similarity individually centered. In the range $y \in [0, 0.5]$, the value similarity and the individually centered value similarity coincide.

Table 2
The CPU time spent for 1000 evaluations of the respective parameters t_{CPU} for vector pairs of length $N = 10^6$ and times relative to evaluation of Pearson correlation parameter $\frac{t_{\mathrm{CPU}}}{t_{\mathrm{CPU}}^{\mathrm{Pearson}}}$, for implementations in Fortran 95 and C.

Parameter	Fortran 95		C	
	t_{CPU} (s)	$\frac{t_{\mathrm{CPU}}}{t_{\mathrm{CPU}}^{\mathrm{Pearson}}}$	t_{CPU} (s)	$\frac{t_{\mathrm{CPU}}}{t_{\mathrm{CPU}}^{\mathrm{Pearson}}}$
Pearson correlation	19.33	1.00	18.74	1.00
Modified Pearson abs. dissim.	14.15	0.73	22.31	1.19
Extended Jaccard	11.96	0.62	9.45	0.50
Cosine similarity	10.34	0.53	12.87	0.69
Canberra distance	9.83	0.51	13.12	0.70
Bray–Curtis distance	10.05	0.52	11.50	0.61
Minimum ratio	18.62	0.96	23.38	1.25
Minimum ratio modified	25.00	1.29	47.66	2.54
Zielke et al.	24.00	1.24	21.83	1.16
Value similarity	13.20	0.68	18.78	1.00
Value similarity centered	22.38	1.16	27.70	1.48
Value similarity ind. centered	22.63	1.17	27.77	1.48

processing applications are much shorter than that, however, long vectors in this performance measuring case enable better distinction between different function calling times.

The function calling times were measured by the GNU profiler `gprof`. The test program was run on an AMD Phenom II X4 945 CPU at 3 GHz under the Debian 7.0 OS (kernel ver. 3.2.0). The resulting times are presented in Table 2, as well as the time relative to evaluation time of the Pearson correlation. The factors influencing the computation

times are the performance of evaluation of arithmetic operations, i.e., division is usually noticeably slower than other arithmetic operations, use of "slow" functions in the parameter definitions. In this case, such functions are square roots and calculating absolute values, which is implemented as a sign transfer function in both Fortran 95 and C (`sign` and `copysign`, respectively).

5. Conclusion

We have introduced a simple parameter for characterization of similarity/dissimilarity of two sequences or vectors (that can be generalized for functions in a continuous case) defined on a selected numeric range, which can be readily extended to characterization of symmetry/asymmetry of two functions. This parameter has a clear and intuitively understandable geometric interpretation with respect to the above defined different types of antisymmetry.

The proposed parameter, as well as a number of other well-known distance parameters, was tested for the case of continuous variation of dissimilarity of the specifically constructed test function. This parameter varies monotonously and smoothly with the antisymmetry transferring from one type to another. In this respect, its characteristics are most similar to the symmetry distance parameter of Zielke et al. of all the tested parameters.

The proposed parameter was shown to have certain attractive properties: normalization in the range $[-1, 1]$, continuous variation with the continuously varying dissimilarity of the sequence and reasonably fast evaluation.

Acknowledgements. This research is funded by the European Social Fund under the project "Microsensors, microactuators and controllers for mechatronic systems (Go-Smart)" (Agreement No. VP1-3.1-ŠMM-08-K-01-015).

References

Akila, A., Chandra, E. (2013). Slope finder – a distance measure for DTW based isolated word speech recognition. *International Journal of Engineering and Computer Science*, 2(12), 3411–3417.

Androutsos, D., Plataniotis, K.N., Venetsanopoulos, A.N. (1998). Distance measures for color image retrieval. In: *International Conference on Image Processing.*, Vol. 2, Chicago, IL, October 1998, pp. 770–774.

Basri, R., Moses, Y (1999). When is it possible to identify 3D objects from single images using class constraints. *International Journal of Computer Vision*, 33(2), 95–116.

Baušys, R., Kriukovas, A. (2012). Pixel-wise tamper detection under generic blur/sharpen attacks. *Informatica*, 23(4), 507–520.

Berman, A.P., Shapiro, L.G. (1997). Efficient image retrieval with multiple distance measures. In: *Storage and Retrieval for Image and Video Database*, Vol. 3022, pp. 12–21.

Bracewell, R.N. (1956). Strip integration in radio astronomy. *Australian Journal of Physics*, 9(2), 198–217.

Cha, S.-H. (2007). Comprehensive survey on distance/similarity measures between probability density functions. *International Journal of Mathematical Models and Methods in Applied Sciences*, 1(4), 300–307.

Cha, S.-H. (2008). Taxonomy of nominal type histogram distance measures. In: *Proceedings of the American Conference on Applied Mathematics*, World Scientific and Engineering Academy and Society (WSEAS), pp. 325–330.

Cohen, E.H, Zaidi, Q. (2013). Symmetry in context: Salience of mirror symmetry in natural patterns. *Journal of Vision*, 13(6), 1–9.

Deans, S.R. (2000). Radon and Abel transforms. In: Poularikas, A.D. (Ed.), *The Transforms and Applications Handbook*, 2 ed. CRC Press, Boca Raton.

Ding, H, Trajcevski, G., Scheuermann, P., Wang, X., Keogh, E. (2008). Querying and mining of time series data: Experimental comparison of representations and distance measures. *Proceedings of the VLDB Endowment*, 1(2), pp. 1542–1552.

Eiter, T., Mannila, H. (1997). Distance measures for point sets and their computation. *Acta Informatica*, 34(2), 109–133.

Finch, H. (2005). Comparison of distance measures in cluster analysis with dichotomous data. *Journal of Data Science*, 3(1), 85–100.

Gesu, V.D., Tabacchi, M.E., Zavidovique, B. (2010). Symmetry as an intrinsically dynamic feature. *Symmetry*, 2, 554–581.

Goshtasby, A.A. (2012). Similarity and dissimilarity measures. In: *Image Registration, Advances in Computer Vision and Pattern Recognition*. Springer, London, pp. 7–66.

Haisch, C. (2012). Optical tomography. *Annual Review of Analytical Chemistry*, 5, 57–77.

Kazhdan, M., Chazelle, B., Dobkin, D., Funkhouser, T., Rusinkiewicz, S. (2004). A reflective symmetry descriptor for 3D models. *Algorithmica*, 38, 201–255.

Kiryati, N., Gofman, Y. (1998). Detecting symmetry in grey level images: the global optimization approach. *International Journal of Computer Vision*, 29(1), 29–45.

Köser, K., Zach, C., Pollefeys, M. (2011). Dense 3D reconstruction of symmetric scenes from a single image. In: *Pattern Recognition. Lecture Notes in Computer Science*, Vol. 6835. Springer, bERLIN, pp. 266–275.

Lileikytė, R., Telksnys, L. (2013). Metrics based quality estimation of speech recognition features. *Informatica*, 24(3), 435–446.

Liu, Y., Hel-Or, H., Kaplan, C.S., Van Gool, L. (2010). Computational symmetry in computer vision and computer graphics. *Foundations and Trends in Computer Graphics and Vision*, 5(1–2), 1–195.

Loy, G., Eklundh, J.-O. (2006). Detecting symmetry and symmetric constellations of features. In: *Computer Vision, ECCV 2006*. Springer, Berlin, pp. 508–521.

Machilsen, B., Pauwels, M., Wagemans, J. (2009). The role of vertical mirror symmetry in visual shape detection. *Journal of Vision*, 9(12), 1–11.

Mancas, M., Gosselin, B., Macq, B. (2005). Fast and automatic tumoral area localisation using symmetry. In: *Proceedings of the IEEE ICASSP Conference*, Vol 1, pp. 725–728.

Martinet, A., Soler, C., Holzschuch, N., Sillion, F.-X. (2006). Accurate detection of symmetries in 3D shapes. *ACM Transactions on Graphics*, 25(2), 439–464.

Milner, D., Raz, S., Hel-Or, H., Keren, D., Nevo, E. (2007). A new measure of symmetry and its application to classification of bifurcating structures. *Pattern Recognition*, 40, 2237–2250.

Monev, V. (2004). Introduction to similarity searching in chemistry. *MATCH Communications in Mathematical and in Computer Chemistry*, 51, 7–38.

Ruppert, G.C.S., Teverovskiy, L., Yu, C.-P., Falcao, A.X., Liu, Y. (2011). A new symmetry-based method for mid-sagittal plane extraction in neuroimages. In: *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, March 2011, pp. 285–288.

Seung-Seok, C., Sung-Hyuk, C., Tappert, C.C. (2010). A survey of binary similarity and distance measures. *Journal of Systemics, Cybernetics & Informatics*, 8(1), 43–48.

Shimshoni, I., Moses, Y., Lindenbaum, M. (2000). Shape reconstruction of 3D bilaterally symmetric surfaces. *International Journal of Computer Vision*, 39(2), 97–110.

Stanujkic, D., Magdalinovic, N., Jovanovic, R. (2013). A multi-attribute decision making model based on distance from decision maker's preferences. *Informatica*, 24(1), 103–118.

Stegmann, M.B., Skoglund, K., Ryberg, C. (2005). Mid-sagittal plane and mid-sagittal surface optimization in brain MRI using a local symmetry measure. In: *Medical Imaging*. International Society for Optics and Photonics, pp. 568–579.

Treder, M. S. (2010). Behind the looking-glass: A review on human symmetry perception. *Symmetry*, 2(3), 1510–1543.

Vlachos, M., Hadjieleftheriou, M., Gunopulos, D., Keogh, E. (2003). Indexing multi-dimensional time-series with support for multiple distance measures. In: *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '03, New York, NY, USA, pp. 216–225.

Vlachos, M., Gunopulos, D., Das, G. (2004). Rotation invariant distance measures for trajectories. In: *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, New York, NY, USA, pp. 707–712.

Wagemans, J. (1995). Detection of visual symmetries. *Spatial Vision*, 9(1), 9–32.

Zabrodsky, H., Peleg, S., Avnir, D. (1995). Symmetry as a continuous feature. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(12), 1154–1166.

Zielke, T., Brauckmann, M., von Seelen, W. (1992). Intensity and edge-based symmetry detection applied to car-following. In: G. Sandini (Ed.), *Computer Vision, ECCV'92*, vol. 588. Springer, Berlin, pp. 865–873.

A. Džiugys holds a PhD degree in technical science and is a chief research associate in the Laboratory of Combustion Processes at Lithuanian Energy Institute. The main research interests include numerical simulation of dynamics of granular matter, computational fluid dynamics, combustion processes and numerical methods. He is the author (or a co-author) of more than 70 scientific papers.

R. Navakas obtained a PhD degree in Physics in 2001 from Vytautas Magnus University and the Institute of Physics. He is a senior research associate in the Laboratory of Combustion Processes of Lithuanian Energy Institute. His research interests include Discrete Element Modelling of granular matter, hydrodynamics, flame spectroscopy, applications of graph theory, simulations and high performance computing.

N. Striūgas is the head of the Laboratory of Combustion Processes of Lithuanian Energy Institute. He obtained a PhD degree in Power and Thermal Engineering from Lithuanian Energy Institute. His research interests include combustion and gasification processes of all types of feedstock for heat and electricity production; developing and testing various equipment for application in industrial combustion.

Normalizuotas parametras sekų panašumo/skirtumo charakterizavimui

Algis DŽIUGYS, Robertas NAVAKAS, Nerijus STRIŪGAS

Siūlomas normalizuotas parametras, charakterizuojantis dviejų sekų panašumą/skirtumą, kuris tolygiai kinta, kintant sekų simetrijos savybėms. Šis parametras gali būti naudojamas eksperimentinių duomenų analizei ir aproksimavimui, naudojant teorinį modelį, veidrodinės simetrijos įvertinimui pasirinktos arba numanomos simetrijos ašies atžvilgiu, tame tarpe simetrijos nustatymo uždaviniuose, kai simetrijos parametrą reikia skaičiuoti daug kartų. Siūlomas parametras, taip pat keli gerai žinomi atstumo ir panašumo parametrai, palyginami, naudojant šabloninių funkcijų ansamblį, kurios tolygiai kinta nuo simetrinės iki antisimetrinės formos. Toks palyginimas leidžia įvertinti skirtingas panašumo ir simetrijos metrikas labiau apibrėžtomis ir valdomomis sąlygomis, negu vertinant bandomojo atvaizdo simetriškumą vizualiai.