

The Levenshtein distance as a measure of mirror symmetry and homogeneity for binary digital patterns

Godfried T. Toussaint

New York University Abu Dhabi, gt42@nyu.edu

Keywords: Mirror symmetry, Local symmetry, Global symmetry, Sub-symmetries, Levenshtein distance, Hamming distance, Pattern complexity, Krüger's symmetry and homogeneity measures, Perception, Image processing, Artificial intelligence

Abstract:

The complexity of a digital pattern, image, map, or sequence of symbols is a salient feature that finds numerous applications in a variety of domains of knowledge [1], [7], [10], [11]. Two features of patterns that form inherent components of pattern complexity, are mirror (reflection) symmetry and homogeneity [8], [9]. In the raster graphics representation mode, a pattern consists of a two-dimensional array (matrix) of elements (pixels, symbols). It is assumed here that the elements are binary-valued (black-white). With such a representation it is common to compute properties of 2-dimensional patterns, such as complexity, mirror-symmetry, and homogeneity, along the 1-dimensional rows, columns, and diagonals of the array [4]. In addition, within each row, mirror symmetries may be analysed either globally or locally [3]. A pattern that does not exhibit *global* mirror symmetry may still possess an abundant number of *local* mirror symmetries. Local symmetries permit graded measures of symmetry rather than all-or-nothing decisions. One powerful type of local symmetry is the *sub-symmetry*, a *contiguous* subset of elements of the pattern that is palindromic (has mirror symmetry). It has been shown empirically that the total number of sub-symmetries present in a pattern may serve as an excellent predictor of the perception of both *visual* pattern complexity [5], and *auditory* pattern complexity [6]. The present research project explores how two well-known measures of the distance between binary patterns and their inversions, correlate with sub-symmetries, as well as other measures of symmetry and homogeneity.

Given two equal-length binary symbol sequences (patterns) A and B , the Hamming distance between A and B is the total number of substitutions of individual symbols required to transform A into B . In other words, the Hamming distance measures the number of aligned corresponding elements in A and B that differ from each other. If two sequences A and B are mirror reflections of each other (palindromic) then the Hamming distance between A and B is zero. If A and B are not palindromic, and B is the inversion of A , then the number of substitutions required to convert A into B is a measure of the amount of mirror symmetry possessed by A . A relatively large number of substitutions indicates a relatively low value of symmetry. This measure of symmetry has been explored extensively as a measure of mirror symmetry (see for example [2] and [4]). In this research project on the other hand, preliminary experimental results are reported using the *Levenshtein* distance as a measure of symmetry, and compared with results obtained with other measures of symmetry and homogeneity, using several datasets composed of binary sequences.

Given two binary symbol sequences (patterns) A and B , the Levenshtein (also *edit*) distance between A and B is defined as the minimum number of *insertions*, *deletions*, and *substitutions* of individual symbols required to transform A into B . This measure generalizes the Hamming distance so as to include insertions and deletions of elements. Consider the two binary patterns A and B in Figure 1. Both exhibit global left-right mirror symmetry, as well as an abundant and equal number of sub-symmetries. As a result the two sequences are perceived to be quite similar. The edit distance between A and B is only 2, and thus captures well the similarity. On the other hand the Hamming distance is 9, and thus fails to capture the similarity. Therefore the edit distance appears to be a better model of the perceptual similarity of such patterns than the Hamming distance. The fact that similarity between a pattern and its mirror image constitutes a measure of mirror symmetry provides the motivation for the exploration of the Levenshtein distance in this context.

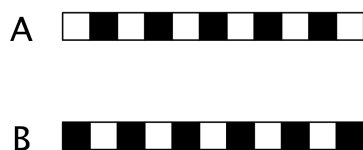


Figure 1. The two binary patterns A and B , which exhibit left-right mirror symmetry, are perceived to be similar.

One of the datasets tested in this emerging project was compiled by M. Krüger, and consists of 19 sequences, each of length 32 binary symbols [8]. Krüger analyzed these sequences with a family of measures of homogeneity and symmetry. Krüger's measure of homogeneity is defined as "the amount of invariance of a binary sequence under its

cyclic permutations.” His measure of symmetry is defined as “the amount of invariance of a binary sequence under the cyclic permutations of its inversion.” Both measures depend on a parameter k . Here we analyze the 19 sequences with the new proposed measure of symmetry based on the Levenshtein distance, and compare the results with those obtained with sub-symmetries and Krüger’s measures of symmetry and homogeneity for several values of k . The rankings of the 19 sequences according to the different measures are analyzed via Spearman rank correlations, and the results are shown in Table 1, along with their levels of statistical significance (p values). The labels S-2, S-4, and S-8 denote Krüger’s measures of symmetry for $k = 2, 4, 8$. The labels H-4 and H-8 denote Krüger’s measures of homogeneity for $k = 4$ and 8 . The value of k is an exponent in Krüger’s formula, and for $k = 2$, the symmetry and homogeneity measures turn out to be identical.

Table 1. The Spearman rank correlation coefficients between the measures of symmetry based on the Levenshtein and Hamming distances, and the sub-symmetries and Krüger’s measures of symmetry and homogeneity.

| Spearman Rank Correlations Coefficients | | | | | | |
|---|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
| | Sub-Symmetries | Krüger’s S-2 | Krüger’s S-4 | Krüger’s S-8 | Krüger’s H-4 | Krüger’s H-8 |
| Levenshtein Distance | $r = -0.604$ $p < 0.003$ | $r = -0.481$ $p < 0.018$ | $r = -0.498$ $p < 0.015$ | $r = -0.592$ $p < 0.004$ | $r = -0.472$ $p < 0.021$ | $r = -0.465$ $p < 0.023$ |
| Hamming Distance | $r = -0.075$ $p < 0.381$ | $r = 0.009$ $p < 0.485$ | $r = 0.118$ $p < 0.316$ | $r = 0.087$ $p < 0.362$ | $r = 0.145$ $p < 0.275$ | $r = 0.190$ $p < 0.218$ |

From the results in Table 1 it is evident that the Hamming distance measure of symmetry (bottom row) does not correlate at all with any of the other measures tested. On the other hand, the Levenshtein distance measure of symmetry exhibits statistically significant moderately high correlation coefficients with all the measures tested. The maximum correlation is $r = -0.604$, obtained with the sub-symmetry measure. These results support the hypothesis that the Levenshtein distance measure of symmetry implicitly assigns greater weights to the sub-symmetries than does the Hamming distance. These results with Krüger’s data, which consist of symbol sequences that range widely between order and disorder, suggest that the often-used Hamming distance measure of symmetry tends to work well only with structured patterns. With unstructured patterns, distant elements may superimpose on each other by reflection, even if the elements in between them are not palindromic, thus inadvertently inflating the symmetry score when it may not be warranted. This tends not to happen with sub-symmetries. This property in effect makes the Hamming distance implicitly put more weight on global rather than local symmetries, and thus may explain the results in Table 1. Experiments with other datasets are being explored to explore this behavior, and determine more precisely the kind of data for which the Levenshtein distance measure of symmetry works better than the Hamming distance, and to ascertain how well it correlates with human judgments and with different types of measures of pattern complexity.

Quantitative measures of visual complexity are useful as guidelines for cartographers engaged in the production of maps that may serve a variety of different purposes [11]. In addition, topographic mapping authorities face an ongoing need to update maps. In this context, objective metrics of complexity such as the Levenshtein measure proposed here, may help to predict the amount of work involved in map revision planning [1].

1. D. Fairbairn, “Measuring map complexity,” *The Cartographic Journal*, 43(3), 2006.
2. A. Gartus & H. Leder, “Predicting perceived visual complexity of abstract patterns using computational measures: The influence of mirror symmetry on complexity perception,” *PLoS ONE*, 12(11), 2017.
3. G. T. Toussaint, “Objective stimulus features for predicting human judgments of visual pattern goodness: An empirical comparison,” in *Recent Advances in Computer Science, Proceedings of the 19th International Conference on Computers* (Part of CSCC'15). Zakynthos Island, Greece, July 16-20, 2015.
4. G. T. Toussaint, N. Onea, & Q. Vuong, “Measuring the complexity of two-dimensional patterns: Sub-symmetries versus Papentin complexity,” *Proceedings of the 14th IAPR Conference on Machine Vision Applications* (MVA 2015), Tokyo, Japan, May 18-22, 2015.
5. C. Alexander, & S. Carey, “Subsymmetries,” *Perception & Psychophysics*, 4, 1968.
6. G. T. Toussaint & J. F. Beltran, “Subsymmetries predict auditory and visual pattern complexity,” *Perception*, 42, November 2013.
7. F. Papentin, “Binary sequences. I. Complexity,” *Information Sciences*, 31, 1983.
8. M. Krüger, “Binary sequences. II. Homogeneity and symmetry,” *Information Sciences*, 31, 1983.
9. F. Papentin & M. Krüger, “Binary sequences. III. Complexity versus homogeneity and symmetry,” *Information Sciences*, 31, 1983.
10. F. Papentin, “On order and complexity. I. General considerations,” *Journal of Theoretical Biology*, 87, 1980.
11. A. K. Bregt & M. C. S. Wopereis, “Comparison of complexity measures of choropleth maps,” *The Cartographic Journal*, 27, 1990.