

Flexible Global Constraint Extension for Dynamic Time Warping

Tomáš Kocyan^{1(✉)}, Kateřina Slaninová^{1,2}, and Jan Martinović^{1,2}

¹ IT4Innovations National Supercomputing Center,
VŠB – Technical University of Ostrava, 17. listopadu 15/2172,
708 33 Ostrava, Czech Republic

{tomas.kocyan,katerina.slaninova,jan.martinovic}@vsb.cz

² Faculty of Electrical Engineering and Computer Science,
VŠB – Technical University of Ostrava, 17. listopadu 15/2172,
708 33 Ostrava, Czech Republic

Abstract. Dynamic Time Warping algorithm (DTW) is an effective tool for comparing two sequences which are subject to some kind of distortion. Unlike the standard methods for comparison, it is able to deal with a different length of compared sequences or with reasonable amount of inaccuracy. For this reason, DTW has become very popular and it is widely used in many domains. One of its the biggest advantages is a possibility to specify definable amount of benevolence while evaluating similarity of two sequences. It enables to percept similarity through the eyes of domain expert, in contrast with a strict sequential comparison of opposite sequence elements. Unfortunately, such commonly used definition of benevolence cannot be applied on DTW modifications, which were created for solving specific tasks (e.g. searching the longest common subsequence). The main goal of this paper is to eliminate weaknesses of commonly used approach and to propose a new flexible mechanism for definition of benevolence applicable to modifications of original DTW.

Keywords: Dynamic time warping · Flexible global constraint · Longest common subsequence · Comparison of sequences

1 Introduction

Nowadays, searching and comparing time series databases generated by computers, which consist of accurate time cycles and which achieve a determined finite number of value levels, is a trivial problem. Main attention is focused rather on optimization of the searching speed. A non-trivial task occurs while comparing or searching signals with different length, which are not strictly defined and have various distortions in time and amplitude. As a typical example, we can mention the measurement of functionality of human body (ECG, EEG) or the elements (precipitation, flow rates in riverbeds), that does not contain any accurate timing for signal generation. Therefore, comparison of such sequences is significantly difficult, and almost impossible while using standard functions for

similarity (distance) computation [2], such as *Euclidean distance* [3], *cosine measure* [8], *Mean Estimate Error* [16], etc. Examples of such signals are presented in Fig. 1. A problem of standard functions for similarity (distance) computation consists in sequential comparison of the opposite elements in the both sequences (comparison of elements with the identical indices). Fortunately, such lack of commonly used approach can be easily eliminated by the *Dynamic Time Warping* algorithm, which is able to percept similarity through the eyes of a domain expert, in contrast with a strict sequential comparison. However, such commonly used definition of benevolence cannot be applied on DTW modifications, which were created for solving specific tasks (e.g. searching the longest common subsequence).

The main goal of this paper is to eliminate weaknesses of commonly used approach and to propose a new flexible mechanism for definition of benevolence applicable to modifications of the original DTW. It is organized as follows: First, the DTW algorithm for comparing two distorted sequences and its several modifications will be described in Sect. 2. In Sect. 3, commonly used approaches for definition of benevolence will be introduced. It will be followed by a proposal of a new *Flexible Global Constraint*. Finally, an effect of the algorithm's settings will be visualized and the proposed solution will be discussed.

2 Dynamic Time Warping

Dynamic Time Warping (DTW) is a technique for finding the optimal matching of two warped sequences using pre-defined rules [11]. Essentially, it is a nonlinear mapping of particular elements to match them in the most appropriate way. The output of such DTW mapping of sequences from Fig. 1 can be seen in Fig. 2. At first, this approach was used for comparison of two voice patterns during an automatic recognition of voice commands [13]. Since this time, it was widely used in many domains, e.g. for efficient satellite image analysis [12], in analysis of student behavioral patterns [17] or in protein fold recognition [9]. As it is correctly noted in [5], a common problem of many DTW applications lies in the fact, that the DTW is too computationally expensive. In order to speed up the algorithm run, several lower bounding methods [4] or parallelization techniques were created [14, 15]. Moreover, the DTW was modified many times for solving specific tasks (e.g. searching the longest common subsequence [7]) or for better algorithm behavior (e.g. *Derivative Dynamic Time Warping* [6]). Since the proposed approach is also an extension of this algorithm, the original DTW algorithm will be described in more detail for better understanding.

Formally, the main goal of DTW method is a comparison of two time dependent sequences x and y , where $x = (x_1, x_2, \dots, x_n)$ and $y = (y_1, y_2, \dots, y_m)$, and finding an optimal mapping of their elements. To compare partial elements of sequences $x_i, y_j \in \mathbb{R}$, it is necessary to define a local cost measure $c : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$, where c is small if x and y is similar to each other, and otherwise it is large. Computation of the local cost measure for each pair of elements of sequences x and y results in a construction of the cost matrix $C \in \mathbb{R}^{n \times m}$ defined by $C(i, j) = c(x_i, y_j)$ (see Fig. 3(a)).

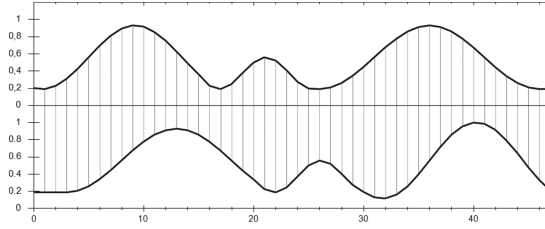


Fig. 1. Standard metrics comparison

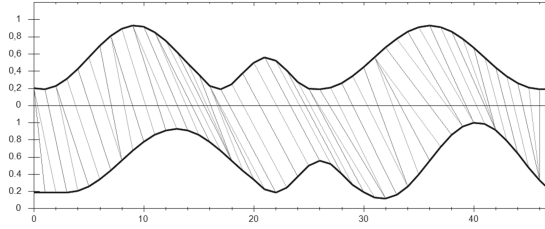


Fig. 2. DTW comparison

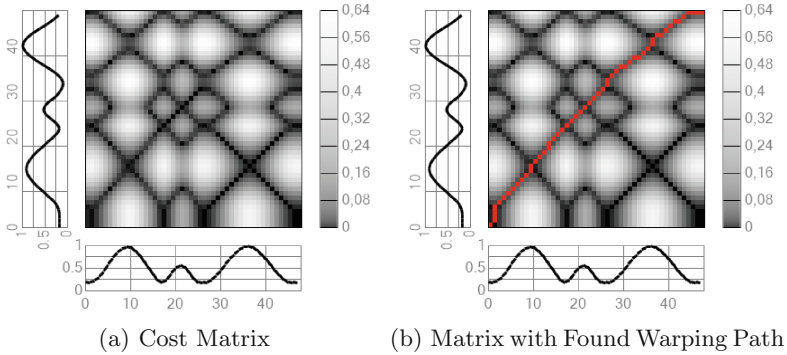


Fig. 3. DTW cost matrices

Then the goal is to find an alignment between x and y with a minimal overall cost. Such optimal alignment leads through the black valleys of the cost matrix C , trying to avoid the white areas with a high cost. Such alignment is demonstrated in Fig. 3(b). Basically, the alignment (called warping path) $p = (p_1, \dots, p_q)$ is a sequence of q pairs (warping path points) $p_k = (p_{kx}, p_{ky}) \in \{1, \dots, n\} \times \{1, \dots, m\}$. Each of such pairs (i, j) indicates an alignment between the i th element of the sequence x and j th element of the sequence y .

Retrieval of optimal path p^* by evaluating all possible warping paths between sequences x and y leads to an exponential computational complexity. Fortunately, there exists a better way with $O(n \cdot m)$ complexity based on dynamic programming. It involves the use of an accumulated cost matrix $D \in \mathbb{R}^{n \times m}$ described in [11].

Accumulated cost matrix computed for the cost matrix from Fig. 3(a) can be seen in Fig. 4(a). It is evident that the accumulation highlights only a single black valley. The optimal path $p^* = (p_1, \dots, p_q)$ is then computed in a reverse order starting with $p_q = (n, m)$ and finishing in $p_1 = (1, 1)$. An example of such found warping path can be seen in Fig. 4(b).

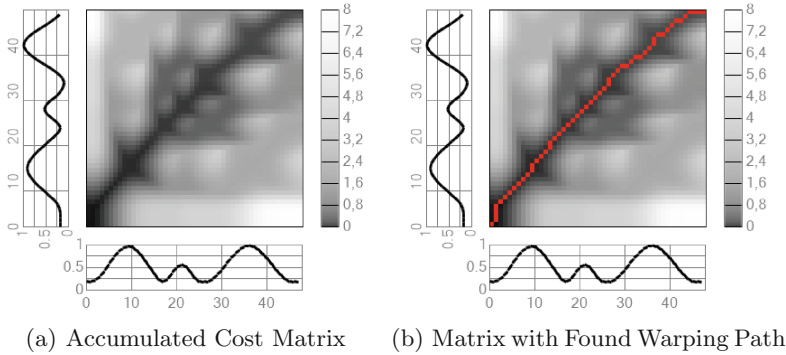


Fig. 4. DTW accumulated cost matrices

The final *DTW cost* can be understood as a quantified effort for the alignment of the two sequences (see Eq. 1).

$$DTW(x, y) = \sum_{k=1}^q C(x_{p_{kx}}, y_{p_{ky}}) = D(n, m) \tag{1}$$

2.1 Subsequence DTW

In some cases, it is not necessary to compare or align the whole sequences. A usual goal is to find an optimal alignment of a sample (a relatively short time series) within the signal database (a very long time series). This is very usual in situations, in which one manages with a signal database and wants to find the best occurrence(s) of a sample (query). Using the slight modification [11], the DTW has the ability to search such queries in a much longer sequence. The basic idea is not to penalize the omission in the alignment between x and y that appears at the beginning and at the end of the sequence y . Suppose we have two sequences $x = (x_1, x_2, \dots, x_n)$ of the length $n \in \mathbb{N}$ and $y = (y_1, y_2, \dots, y_m)$ of the much larger length $m \in \mathbb{N}$. The goal is to find a subsequence $y_{a:b} = (y_a, y_{a+1}, \dots, y_b)$ where $1 \leq a \leq b \leq m$ that minimizes the DTW cost to x over the all possible subsequences of y . An example of such searching the best subsequence alignment can be seen in Fig. 5. Both constructed matrices including the found warping path are then shown in Fig. 6.

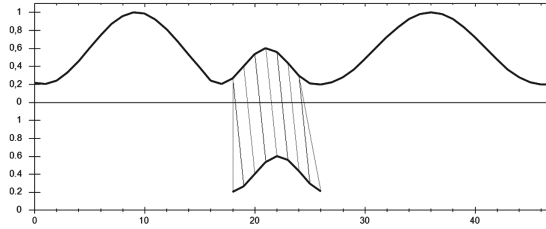


Fig. 5. Found DTW subsequence

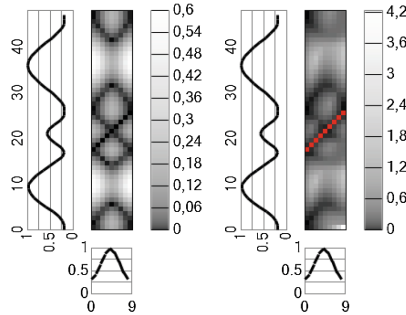


Fig. 6. Cost matrix and accumulated cost matrix for searching subsequence

Despite the fact that the DTW has its own modification for searching subsequences, it works perfectly only in case of searching an exact pattern in a signal database. However, in real situations, exact patterns are not available because they are surrounded by additional values, or even repeated several times in a sequence (see Fig. 7). Unfortunately, the basic DTW is not able to handle these situations and it fails or returns only a single occurrence of the pattern. To deal with this type of situations, several DTW modifications were created and described for example in [7] or [10] in detail.

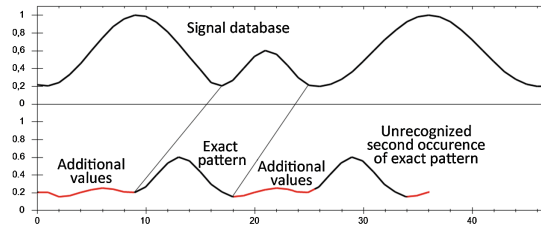


Fig. 7. Basic DTW subsequence inaccuracies

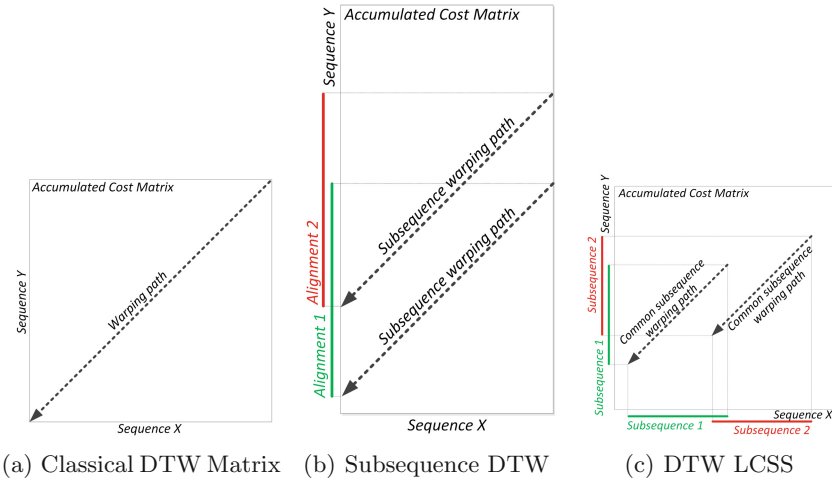


Fig. 8. Approach for searching the warping path

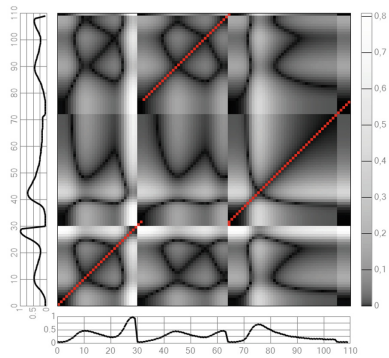


Fig. 9. Cost matrix with found warping paths

The biggest difference is in the approach for searching the warping path. In simple terms, the algorithm does not search the warping path from the upper right corner to the bottom left one (shown in the case of classical DTW in Fig. 8(a)) and also it does not connect the opposite sides of the matrix (shown in the case of subsequence DTW in Fig. 8(b)). The main idea is to find warping paths as long as possible from any element to another one, parallel to a diagonal, as it is outlined in Fig. 8(c). An example of such found common subsequences can be seen in Fig. 10. The corresponding warping paths are also visualized in the cost matrix in Fig. 9.

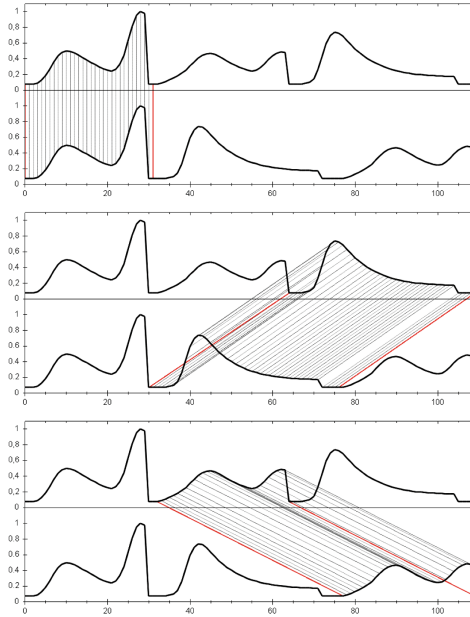


Fig. 10. Found common subsequences

3 Flexible Global Constraints

In the practical applications [1, 18–20], the construction of a warping path has to be controlled. The reason is possible uncontrolled high number of warpings, i.e. alignment of a single element to a high number of the elements in the opposite sequence [11]. In this manner, dissimilar sequences can get low *DTW Cost* and they can be evaluated as similar. This situation is demonstrated on sequences in Fig. 11, and on appropriate cost matrix in Fig. 12.

Generally, this can be easily fixed by definition of a global constraint region $R \subseteq D$. This region then determines the elements of the cost matrix, which can be used for searching the warping path. In the original paper about DTW [11], there are two global constraints for warping path mentioned - *Itakura parallelogram* (Fig. 13(a)) and *Sakoe-Chiba band* (Fig. 13(b)).

However, for purpose of searching subsequences and other DTW modifications, the Itakura parallelogram seems to be inappropriate, because it was designed to limit warpings at the start and end of the classical DTW warping path, where the first and last warping points are exactly known. Fortunately, the Saokoe-Chiba band looks more preferable. The warping path respecting this band for sequences from Fig. 11 is visible in Fig. 14.

However, one may ask what width of band to choose. The width essentially defines the maximal number of warpings in a found sequence. For this reason, it is almost impossible to define a universal number applicable both on shorter and

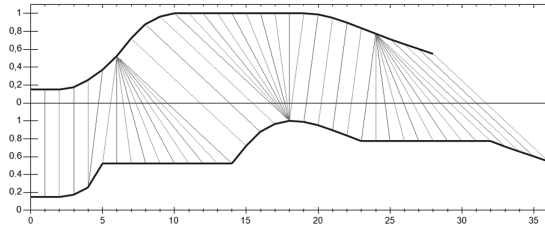


Fig. 11. Mapping of dissimilar sequences

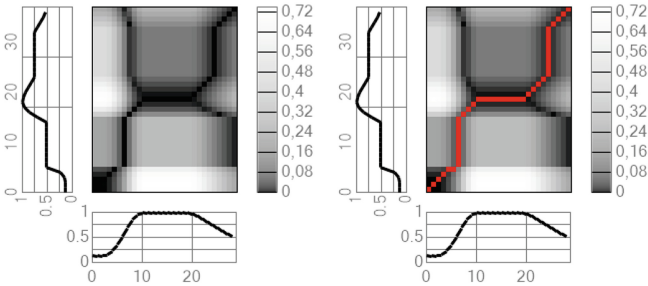


Fig. 12. Cost matrix of dissimilar sequences

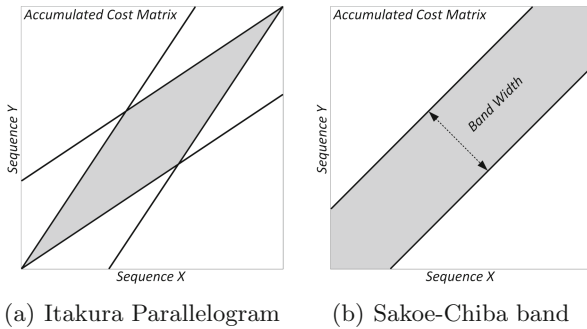


Fig. 13. DTW global constraint regions

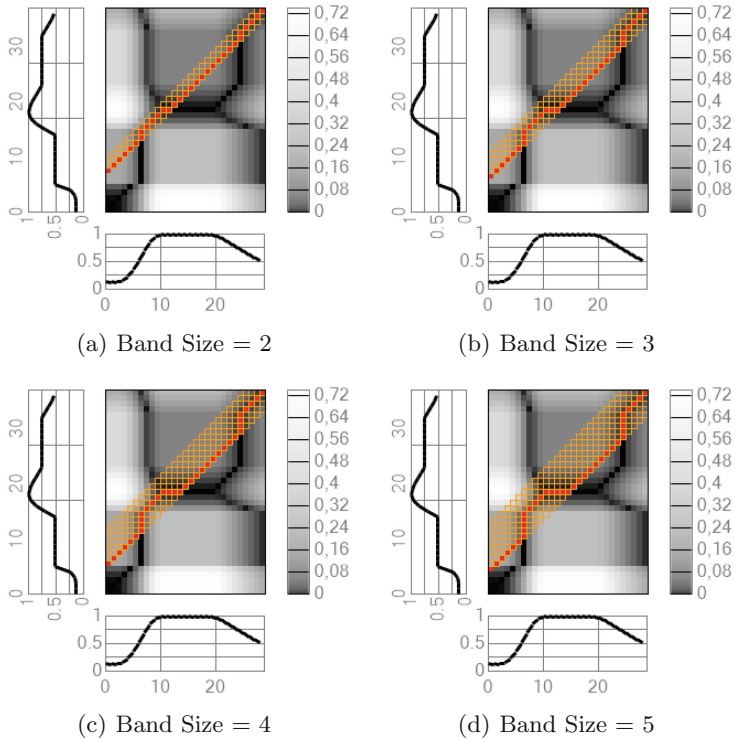


Fig. 14. Examples of applied Saoko-Chiba bands

longer sequences. It is evident that allowing five warpings on a path comparing sequences of the length ten or hundred has absolutely different meaning. In this example, the results look satisfactorily, but this belt was also designed for searching the warping path through the whole sequences. This inaccuracy is evident in the following example:

Lets have two sequences $x = (x_1, x_2, \dots, x_n)$ and $y = (y_1, y_2, \dots, y_{2n})$, where y is created by stretching x into the double length (i.e. $\forall i \in \{1, \dots, 2n\} : y_i = x_{i/2}$). The matrix will stretch in one dimension and the line of minima will slightly bend (see Fig. 15(a)). It causes some warpings, but it is still acceptable. Using the standard Saoko-Chiba band, the warping path cannot follow the minima trajectory and have to continue in straight direction, as shown in Fig. 15(b).

More elegant solution is to allow a band to bend itself and provide a warping path with reasonable freedom. For this purpose, we designed a flexible band allowing configurable bending. The band is based on Saoko-Chiba band, but it changes its position and shape according the previously constructed warping path. The center of the original Saoko-Chiba band lies exactly on cost matrix's diagonal.

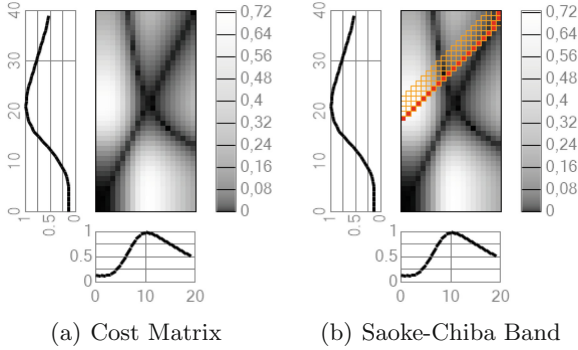


Fig. 15. Cost matrices for stretched sequence

Proposed Modifications to Saoke-Chiba Band. In our modification, the center of the band varies and passes through one of the previous points of the currently constructed warping path, called *control point*. Such control point is always located in the fixed distance from the currently processed point. This distance is called *control point distance* and it is defined as a number of *warping path points* preceding the currently processed point. The center of constructed band always moves to a newly established control point.

Formally, suppose we have a currently constructed warping path p defined as $p = (p_1, \dots, p_q)$ consisting of a sequence of q path points $p_k = (p_{kx}, p_{ky}) \in \{1, \dots, n\} \times \{1, \dots, m\}$, $p_1 = (n, m)$. Each such pair (p_{kx}, p_{ky}) indicates an alignment between the i th element of the sequence x and j th element of the sequence y . The path point (p_{kx}, p_{ky}) lies in the Saoke-Chiba band of a width w , if $|p_{kx} - p_{ky}| < w$. With the flexible band of the width w and with a control point distance d , the path point (p_{kx}, p_{ky}) lies in the band if $|(p_{kx} - p_{(k-d)x}) - (p_{ky} - p_{(k-d)y})| < w$. The distance d of such control point from the end of the warping path defines a rigidity of the band.

Figure 16 demonstrates how the increasing distance of the control point d causes higher toughness of the band, and how the ability to bend loses. The shorter distance makes the band more flexible, the higher distance causes inflexibility. It is especially evident from Fig. 16(d) (with $d = 4$), where the band became too much tough to follow the black valleys.

An effect of predefined toughness can be also easily quantified by the received *DTW Cost* defined in Eq. 1. With an original Saoke-Chiba Band (see Fig. 13), received *DTW Cost* = 3.6433. On the other hand, with using the proposed flexible constraint (distances of the control point d) and appropriately adjusted benevolence, the sequences can be evaluated as almost equal (*DTW Cost* = 0,0182). Table 1 illustrates how the received *DTW Cost* reflects the adjusted amount of benevolence (various distances of the control point d). In order to set the *control point distance* up correctly, it is necessary to have some domain knowledge. At this point, the domain expert has to define the benevolence for the evaluation.

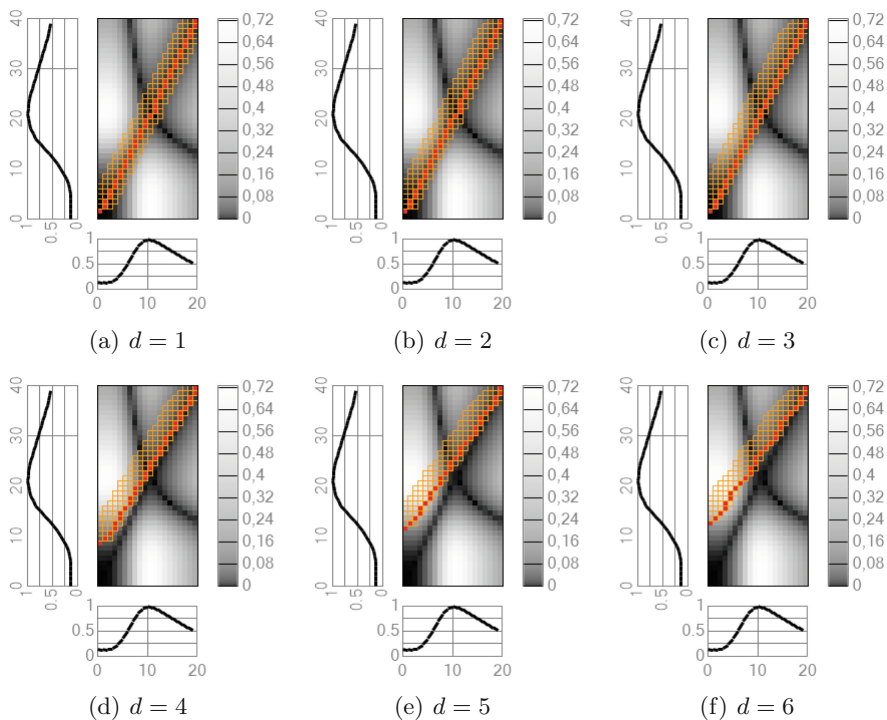


Fig. 16. Various distances of control point

Table 1. DTW costs for various control point distances

Control point distance	<i>DTWCost</i>
1	0.0182
2	0.0182
3	0.0182
4	1.0329
5	2.3110
6	2.9819
7	3.1504
8	3.9598
9	3.6433
10	3.6433
Without flexible band	3.9634

4 Conclusion

The *Dynamic Time Warping* algorithm has become widely used technique for comparing two sequences and evaluating their mutual similarity. Its many modifications, created for solving specific tasks, subsequently requested additional adjustments of partial steps of this algorithm. As a typical example, the DTW approach for searching the longest common subsequence can be mentioned. In this type of modification, none of commonly used constraints for construction of the warping path can be used. Therefore, the main goal of this paper was to provide a solution for such situations and to propose a new flexible mechanism for definition of the constraint applicable to the modifications of the original DTW. The proposed solution consists in a new flexible constraint, which is based on the original Sako-Chiba band. The constraint enables the control over the process of warping path construction and it generally offers more flexibility and predictable behaviour. Moreover, definition of its conduct (i.e. rigidity of the band) can be defined by a single number, which is not dependent on the length of the processed sequences. The use of the proposed solution is not limited only for searching the common subsequences, but it can be utilized in all DTW modifications, whose constructed warping paths are not defined by exactly beginnings and ends.

Acknowledgment. This work was supported by The Ministry of Education, Youth and Sports from the National Programme of Sustainability (NPU II) project ‘IT4Innovations excellence in science - LQ1602’.

References

1. Cheng, H., Dai, Z., Liu, Z., Zhao, Y.: An image-to-class dynamic time warping approach for both 3D static and trajectory hand gesture recognition. *Pattern Recogn.* **55**, 137–147 (2016)
2. Ding, H., Trajcevski, G., Scheuermann, P., Wang, X., Keogh, E.: Querying and mining of time series data: experimental comparison of representations and distance measures. *Proc. VLDB Endow.* **1**(2), 1542–1552 (2008)
3. Elmore, K.L., Richman, M.B.: Euclidean distance as a similarity metric for principal component analysis. *Mon. Weather Rev.* **129**(3), 540–549 (2001)
4. Keogh, E.: Exact indexing of dynamic time warping. In: *Proceedings of the 28th International Conference on Very Large Data Bases, VLDB 2002*, pp. 406–417. VLDB Endowment (2002). <http://dl.acm.org/citation.cfm?id=1287369.1287405>
5. Keogh, E.J., Pazzani, M.J.: Scaling up dynamic time warping for datamining applications. In: *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 285–289. ACM (2000)
6. Keogh, E.J., Pazzani, M.J.: Derivative dynamic time warping. In: *First SIAM International Conference on Data Mining SDM 2001* (2001)
7. Kocyan, T., Martinovič, J., Slaninová, K., Szturcová, D.: Searching the longest common subsequences in distorted data. In: *27th European Modeling and Simulation Symposium, EMSS 2015*, pp. 84–92 (2015)
8. Lee, D.L., Chuang, H., Seamons, K.: Document ranking and the vector-space model. *IEEE Softw.* **14**(2), 67–75 (1997)

9. Lyons, J., Biswas, N., Sharma, A., Dehzangi, A., Paliwal, K.K.: Protein fold recognition by alignment of amino acid residues using kernelized dynamic time warping. *J. Theor. Biol.* **354**, 137–145 (2014)
10. Movchan, A., Zymbler, M.L.: Time series subsequence similarity search under dynamic time warping distance on the intel many-core accelerators. In: *SISAP (2015)*
11. Müller, M.: *Information Retrieval for Music and Motion*. Springer-Verlag New York Inc., Secaucus (2007)
12. Petitjean, F., Weber, J.: Efficient satellite image time series analysis under time warping. *IEEE Geosci. Remote Sens. Lett.* **11**(6), 1143–1147 (2014)
13. Rabiner, L., Juang, B.H.: *Fundam. Speech Recogn.* Prentice-Hall Inc, Upper Saddle River (1993)
14. Rakthanmanon, T., Campana, B., Mueen, A., Batista, G., Westover, B., Zhu, Q., Zakaria, J., Keogh, E.: Addressing big data time series: mining trillions of time series subsequences under dynamic time warping. *ACM Trans. Knowl. Discov. Data* **7**(3), 101–1031 (2013)
15. Sart, D., Mueen, A., Najjar, W., Keogh, E., Niennattrakul, V.: Accelerating dynamic time warping subsequence search with GPUs and FPGAs. In: *2010 IEEE International Conference on Data Mining*, pp. 1001–1006, December 2010
16. Singh, J., Knapp, H.V., Arnold, J., Demissie, M.: Hydrological modeling of the iroquois river watershed using HSPF and SWAT. *J. Am. Water Resour. Assoc.* **41**(2), 343–360 (2005)
17. Slaninová, K., Kocyan, T., Martinovič, J., Dráždilová, P., Snášel, V.: Dynamic time warping in analysis of student behavioral patterns. In: *Proceedings of the Dateso 2012 Annual International Workshop on DATAbases, TExts, Specifications and Objects*. *CEUR Workshop Proceedings*, pp. 49–59 (2012)
18. Toyoda, M., Sakurai, Y.: Discovery of cross-similarity in data streams. In: *Proceedings - International Conference on Data Engineering*, pp. 101–104 (2010)
19. Xu, Q., Zheng, R.: Automated detection of burned-out luminaries using indoor positioning. In: *International Conference on Indoor Positioning and Indoor Navigation, IPIN 2015 (2015)*
20. Zhao, J., Liu, K., Wang, W., Liu, Y.: Adaptive fuzzy clustering based anomaly data detection in energy system of steel industry. *Inf. Sci.* **259**, 335–345 (2014)