

Scaling and Time Warping in Time Series Querying

Ada Wai-chee Fu¹ Eamonn Keogh² Leo Yung Hang Lau¹ Chotirat Ann Ratanamahatana²

Department of Computer Science and Engineering¹
The Chinese University of Hong Kong
Shatin, Hong Kong
adafu,yhlau@cse.cuhk.edu.hk

Department of Computer Science and Engineering²
University of California
Riverside, USA
eamonn,ratana@cs.ucr.edu

Abstract

The last few years have seen an increasing understanding that Dynamic Time Warping (DTW), a technique that allows local flexibility in aligning time series, is superior to the ubiquitous Euclidean Distance for time series classification, clustering, and indexing. More recently, it has been shown that for some problems, Uniform Scaling (US), a technique that allows global scaling of time series, may just be as important for some problems. In this work, we note that for many real world problems, it is necessary to combine both DTW and US to achieve meaningful results. This is particularly true in domains where we must account for the natural variability of human action, including biometrics, query by humming, motion-capture/animation, and handwriting recognition. We introduce the first technique which can handle both DTW and US simultaneously, and demonstrate its utility and effectiveness on a wide range of problems in industry, medicine, and entertainment.

1 Introduction

We propose to query time series with both the accommodation of a scaling factor and the consideration of time warping effects. In this section we justify our proposal with some examples.

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the VLDB copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Very Large Data Base Endowment. To copy otherwise, or to republish, requires a fee and/or special permission from the Endowment.

**Proceedings of the 31st VLDB Conference,
Trondheim, Norway, 2005**

1.1 Justifying the Need for Uniform Scaling and DTW

Because time series are near ubiquitous, and are becoming increasingly prevalent as our ability to capture and store them improves, there is increasing interest in the database community in techniques for efficiently indexing large time series collections [7, 19]. It is found that in most domains, it is necessary to match sequences with tolerance of small local misalignments, and Dynamic Time Warping has been shown to be the right tool for this [5, 12, 21, 24]. For example, in speech comparison, small fluctuation of the tempo of the speakers should be allowed in order to identify similar contents. More recently, it has been shown that in many domains it is also necessary to match sequences with the allowance of a global scaling factor [14]. In this work, we argue that for most real world problems, it is necessary to be able to handle both types of distortions simultaneously. In fact, even a casual glance of existing literature confirms this. For example, in query-by-humming systems, it is well understood that we must allow for uniform scaling in addition to DTW. The current solution is to simply do DTW at many resolutions that span the possible range of tempos. For example, Meek and Birmingham [17] reports “*We account for the phenomenon of persons reproducing the same tune at different speeds ... allow(ing) for nine tempo mappings.*” However, repeating the query nine times clearly slows the system down. Furthermore, it is possible that the best match occurs somewhere in-between the nine discrete scalings. In [15], the authors also note that in addition to the local problems handled by DTW, “*(people can) perform faster or slower than usual.*” They again deal with this with multiple scaled queries, achieving reasonable performance only by the use of parallel processing.

Dynamic Time Warping is frequently used as the basis of gait recognition algorithms, but even in this highly structured domain, it is recognized that uniform scaling is also needed. For example, [11] notes “*different gait cycles tend to have unequal lengths.*” In fact, even if we discount human variability, it is well

understood that parallax effects from cameras (static or pan-and-tilt) can produce apparent changes in uniform scaling [10].

The need for uniform scaling has been noted in bioinformatics. Moeller-Levet et al. [18] noted that previous work that addressed *only* local scaling (with DTW) is inadequate, and they stressed that “(uniform) scaling factors in the expression level hide similar expressions and have to be eliminated or not considered when assessing the similarity between expression profiles” [11].

Finally, the simultaneous need for both uniform scaling and DTW is well understood in the motion-capture and animation community. For example, Pullen and Bregler [20], explaining their motion-capture editing system, noted “we stretch or compress the real data fragments in time by linearly resampling them to make them the same length as the keyframed fragment . . . (then do DTW).” The computational difficulty of dealing with both uniform scaling and DTW at the same time has even led to practitioners abandoning temporal information altogether! Campbell and Bobick [3] used a phase space representation in which the velocity dimensions are removed, thus completely disregarding the time component of the data. This makes the learning and matching of motion patterns simpler and faster, but only at the cost of a massive increase in false positives.

Let us call “DTW with Uniform Scaling” **SWM**, which stands for *Scaled and Warped Matching*. In this paper, we study the combined effects of scaling and time warping in time series querying.

1.2 Motivating Examples

Below, we present two concrete examples that require SWM to produce meaningful and intuitive results.

Example 1 (Indexing video). *There is increasing interest in indexing sports data, both from sports fans who may wish to find particular types of shots or moves, and from coaches who are interested in analyzing their athletes performance over time. As a concrete example, we consider high jump. We can automatically collect the athlete’s center of mass information from video and convert the data into a time series (It is possible to correct for the cameras pan and tilt; see [6]). We found that when we issued queries to a database of high jumps, we got intuitive answers only when doing SWM. It is easy to see why if we look at two particular examples from the same athlete and consider all possible matching options, as shown in Figure 1. In this figure, we show four different ways to match two time series, the horizontal axis is the time axis. In each case, we have shifted one of the two series upward to show the way the points in the two series are matched. Each vertical line in the diagrams shows the*

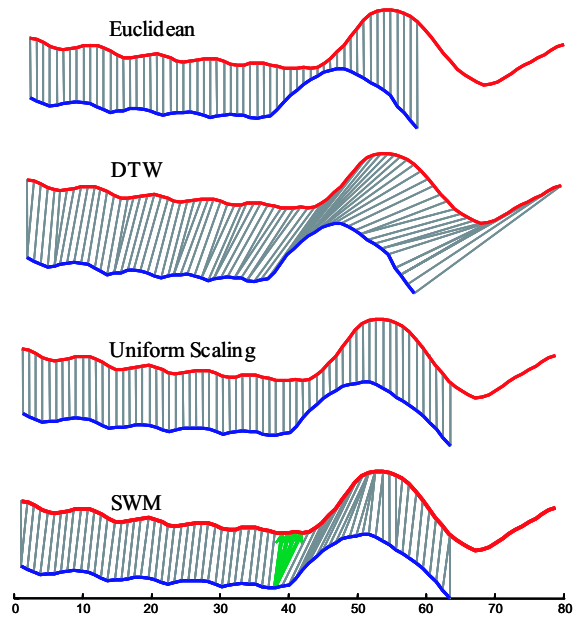


Figure 1: Two examples of an athlete’s trajectories aligned with various measures

matching of two points. Visually we can say that the two time series are similar, and hence the distance between them should be small. We want to see which of the four measurement can generate a result that gives a small distance as expected. From top to bottom:

- If we attempt simple Euclidean matching (after truncating the longer sequence), we get a large distance (which we can consider as error) because we are mapping part of the *flight* of one sequence to the *takeoff drive* in the other.
- If we simply use DTW to match the entire sequences, we get a large error because we are trying to explain part of the sequence in one attempt (the bounce from the mat) that simply does not exist in the other sequence. This problem can be corrected by constraints such as the Sakoe-Chiba Band, but without scaling, the matching will be poor.
- If we attempt just uniform scaling, we get the best match when we stretch the shorter sequence by 112%. However the local alignment, particularly of the *takeoff drive* and *up-flight*, is quite poor.
- Finally, when we match the two sequences with SWM, we get an intuitive alignment between the two sequences. The global stretching (once again at 112%) allows DTW to align the small local differences. In this case, the fact that DTW needed to map a single point in a time series onto 4 points in the other time series suggests an important local difference in one of these sequences. Inspection of the original videos by a professional coach

suggests that the athlete misjudged his approach and attempted a clumsy correction just before his takeoff drive.

Example 2 (Query by Humming). *The need for both local and global alignment when working with music has been extensively demonstrated [4, 16, 17, 24]. For completeness, we will briefly review it here. Finding similar sequences of music has applications in copyright infringement detection, analyzing the evolution of music styles [4], automatic annotation, etc. (It is interesting to note that these studies are not confined to human endeavors; similar techniques have been used in animal “music”, especially in humpback whales and songbirds [16]). However, the vast majority of research in this area is used to support query by humming.*

The basic idea of query by humming is to allow users to search large music collections by providing an example of the desired content, by humming (or singing, or tapping) a snippet. Clearly, humans cannot be expected to reproduce an exact fragment of a song, so the system must be invariant to certain distortions. Some of these are trivial to deal with. For example, the query can be made invariant to key by normalizing both the query and the database to a standard key. However, two types of errors are more difficult to deal with; users may perform the query at the wrong tempo, and users may insert or delete notes. The former corresponds with uniform scaling, the latter with DTW. The music retrieval community has traditionally dealt with these two problems in two ways. The first is to do DTW multiple times, at different scalings [17]. However, this clearly produces scalability problems. The other common approach is to only do DTW with relatively short song snippets as queries believing that short sequences are less sensitive to uniform scaling problems than long sequences. While this is undoubtedly true, short snippets also have less discriminating power.

In Figure 2, we demonstrate the problems with the universally familiar piece of music, *Happy Birthday to You*. For clarity of illustration, the music was produced by the fourth author on a keyboard and converted into a pitch contour, however, similar remarks apply to other music representations. From top to bottom:

- Because the query sequence was performed at a much faster tempo, direct application of DTW fails to produce an intuitive alignment.
- Rescaling the shorter performance by a scaling factor of 1.54 seems to improve the alignment, but note for example that the higher pitched note produced on the third “*birth...*” of the candidate is forced to align with the lower note of the third “*happy...*” in the query.

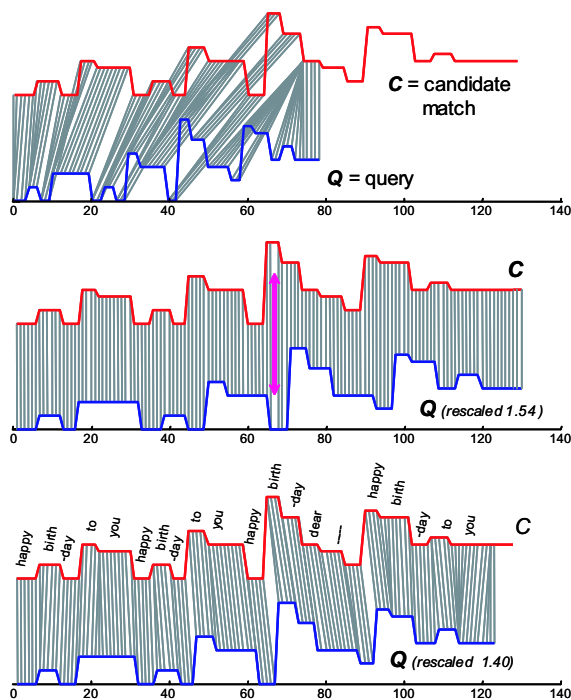


Figure 2: Two performances of *Happy Birthday to You* aligned with different metrics. Both performances were performed in the same key, but are shifted in the Y-axis for visual clarity.

- Only the application of *both* uniform scaling and DTW produces the correct alignment.

Having developed the intuition for DTW and US, and having demonstrated the need to handle both types of distortions simultaneously, we will next define the problem of similarity measurement under SWM more formally.

2 Problem Definition

Assume that we are given a database D which contains M time series (note that this formulation does not preclude the subsequence matching case, since it may be trivially transformed into this formulation). Further, assume that we are given a query Q , and a scaling factor $l, l \geq 1$, which represents the maximum allowable stretching of the time series. The maximum allowable shrinking is implicitly set to $1/l$.¹ Hence the query can be shortened by a factor of up to $1/l$ or lengthen by a factor of up to l . Note that while $1/l$ is bounded below by zero, and l is bounded from above by infinity, such loose bounds would allow pathological solutions to certain problems, and in any case are surely impossible to support efficiently. We therefore

¹Such formulation assumes the maximum allowable stretching and shrinking is symmetric. If this is not the case for a specific application, it is trivial to add as an extra parameter: the maximum allowable shrinking s .

restrict our attention to scaling factors in the range $0.5 \leq 1/l \leq 1 \leq l \leq 2$. Note that this range encompasses the necessary flexibility documented in virtually every domain we are aware of. For instance, in [17], the authors reported excellent results with a query-by-humming system that allows “a (maximum) tempo scaling of 1.25.” [17] notes that in their experience, amateur singers can speed up their rendition of a song by as much as 200% or slow down to as little as 50%.

Recently it has been shown that for nearly all types of time series data, using an appropriate global constraints always improves the classification or clustering accuracy and the precision and recall of indexing [21]. Therefore a global constraint is typically enforced to limit the warping path to a roughly *diagonal portion* of the warping matrix.

Given N variable-length data sequences and a query sequence Q , we would like to find all data sequences that are “similar” to Q . Suppose the query sequence is $Q = Q_1, Q_2, \dots, Q_m$, where Q_i is a numerical value. We are interested in tackling the following problem.

Problem: Assume the data sequences can be longer than the query sequence Q . Find the best match to Q in database, for any rescaling in a given range, under the Dynamic Time Warping distance with a global constraint. By best match we mean the data sequence with the smallest distance from Q .

This problem has never been considered in the literature before. This problem is realistic in applications such as query by humming.

3 Preliminaries

In this section, we review separately time series querying with time warping distance and also querying with the scaling effect. For each case, we can apply a lower bounding technique for pruning the search space.

3.1 Time Warping Distance

Intuitively, *dynamic time warping* is a distance measure that allows time series to be *locally* stretched or shrunk before applying the base distance measure. Definition 1 formally defines time warping distance.

Definition 1 (Time Warping Distance (DTW)). Given two sequences $C = C_1, C_2, \dots, C_n$ and $Q = Q_1, Q_2, \dots, Q_m$, the time warping distance DTW is defined recursively as follows:

$$\begin{aligned} \text{DTW}(\phi, \phi) &= 0 \\ \text{DTW}(C, \phi) &= \text{DTW}(\phi, Q) = \infty \\ \text{DTW}(C, Q) &= D_{\text{base}}(\text{First}(C), \text{First}(Q)) + \\ &\min \begin{cases} \text{DTW}(C, \text{Rest}(Q)) \\ \text{DTW}(\text{Rest}(C), Q) \\ \text{DTW}(\text{Rest}(C), \text{Rest}(Q)) \end{cases} \end{aligned}$$

where ϕ is the empty sequence, $\text{First}(C) = C_1$, $\text{Rest}(C) = C_2, C_3, \dots, C_n$, and D_{base} denotes the distance between two entries.

Several metrics were used as the D_{base} distance in previous literature, such as Manhattan Distance [23] and squared Euclidean Distance [12, 22]. We will use squared Euclidean Distance as the D_{base} measure. That is,

$$D_{\text{base}}(C_i, Q_j) = (C_i - Q_j)^2$$

Note we deliberately omit the final square root function in our distance definitions. Such optimization speeds up computations without altering the relative ranking given by these distances, which is more important than the actual value in most applications. The same optimization has been used before in [14]. However, if such optimization is not desired, we can also consistently insert the final square root function without altering the essence of this work.

It is well known that dynamic time warping distance can be computed by filling a *warping matrix* using a *dynamic programming algorithm* directly derived from the definition of time warping distance. A *warping path* can be identified by tracing the elements in the warping matrix that were used to compute the time warping distance. Formally, a warping path W for two sequences Q and C is a sequence of elements w_1, w_2, \dots, w_p so that $w_k = (i_k, j_k)$ is an entry in the warping matrix, where $i_k \geq i_{k-1}$ and $j_k \geq j_{k-1}$, $\max(|Q|, |C|) \leq |W| \leq |Q| + |C| - 1$.²

3.2 Constraints and Lower Bounding

In the previous section we have explained with examples the importance of having global constraints on time warping in order to give meaningful results. Keogh [12] suggested a lower bounding measure based on such *global constraints* on time warping. Two commonly used global constraints exist. The *Sakoe-Chiba Band* [22] limits the warping path to a band enclosed by two straight lines that are parallel to the diagonal of the warping matrix. The *Itakura Parallelogram* [9] limits the warping path to be within a parallelogram whose major diagonal is the diagonal of the warping matrix.

[12] viewed a global constraint as a constraint on the warping path entry $w_k = (i, j)_k$ and gave a general form of global constraints in terms of inequalities on the indices to the elements in the warping matrix,

$$j - r \leq i \leq j + r$$

where r is a constant for the Sakoe-Chiba Band and r is a function of i for the Itakura Parallelogram.

² $|X|$ denotes the length of a sequence X .

Incorporating the global constraint into the definition of dynamic time warping distance, Definition 1 can be modified as follows.

Definition 2 (Constrained DTW (cDTW)). Given two sequences $C = C_1, C_2, \dots, C_n$ and $Q = Q_1, Q_2, \dots, Q_m$, and the time warping constraint r , the constrained time warping distance cDTW is defined recursively as follows:

$$\text{Dist}_r(C_i, Q_j) = \begin{cases} D_{\text{base}}(C_i, Q_j) & \text{if } |i - j| \leq r \\ \infty & \text{otherwise} \end{cases}$$

$$\text{cDTW}(\phi, \phi, r) = 0$$

$$\text{cDTW}(C, \phi, r) = \text{cDTW}(\phi, Q, r) = \infty$$

$$\text{cDTW}(C, Q, r) = \text{Dist}_r(\text{First}(C), \text{First}(Q)) + \min \begin{cases} \text{cDTW}(C, \text{Rest}(Q), r) \\ \text{cDTW}(\text{Rest}(C), Q, r) \\ \text{cDTW}(\text{Rest}(C), \text{Rest}(Q), r) \end{cases}$$

where ϕ is the empty sequence, $\text{First}(C) = C_1$, $\text{Rest}(C) = C_2, C_3, \dots, C_n$, and D_{base} denotes the distance between two entries.

The upper bounding sequence UW and the lower bounding sequence LW of a sequence C are defined using the time warping constraint r as follows.

Definition 3 (Enveloping Sequences by Keogh [12]). Let $UW = UW_1, UW_2, \dots, UW_m$ and $LW = LW_1, LW_2, \dots, LW_m$,

$$UW_i = \max(C_{i-r}, \dots, C_{i+r}) \text{ and}$$

$$LW_i = \min(C_{i-r}, \dots, C_{i+r})$$

Considering the boundary cases, the above can be rewritten as

$$UW_i = \max(C_{\max(1, i-r)}, \dots, C_{\min(i+r, n)}) \text{ and}$$

$$LW_i = \min(C_{\max(1, i-r)}, \dots, C_{\min(i+r, n)})$$

These two sequences form an envelope which encloses the sequence C , as shown in Figure 3.

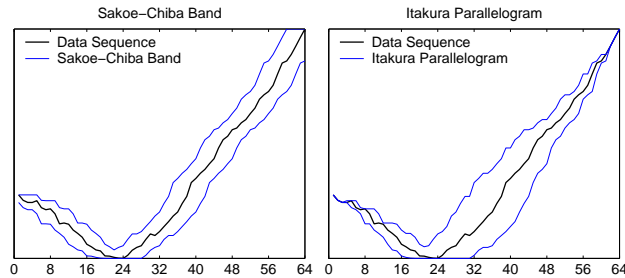


Figure 3: Enveloping sequences derived from two different constraints

The lower bounding measure by Keogh [12] bounds the time warping distance between two sequences Q and C by the Euclidean distance between Q and the envelope of C . Equation (1) below formally defines the lower bounding distance.

$$LB_W(Q, C) = \sum_{i=1}^m \begin{cases} (Q_i - UW_i)^2 & \text{if } Q_i > UW_i \\ (Q_i - LW_i)^2 & \text{if } Q_i < LW_i \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

3.3 Uniform Scaling

Consider a query sequence $Q = Q_1, \dots, Q_m$ and a candidate sequence $C = C_1, \dots, C_n$.

We assume that m is not greater than n ($m \leq n$); hence, the query is typically shorter than the candidate sequence. We assume that the data can scale up or down by a factor of at most l , $l \geq 1$. The entry Q_m may be matched to C_{lm} when the data is shrunk by a factor of l . To simplify our discussion we shall assume that $lm \leq n$.

In order to scale time series $C = C_1, \dots, C_q$ to produce a new time series $C' = C'_1, \dots, C'_m$ of length m , we use the formula

$$C'_j = C_{\lceil j \cdot q/m \rceil} \text{ where } 1 \leq j \leq m$$

This is similar to the formula used in [14]. We target to find a scaled prefix in C to compare with Q . With a scaling factor of l , q can range from $\lceil m/l \rceil$ to lm .

Definition 4 (Uniform Scaling (US)). Given two sequences $Q = Q_1, \dots, Q_m$ and $C = C_1, \dots, C_n$ and a scaling factor bound l , $l \geq 1$. Let $C(q)$ be the prefix of C of length q , where $\lceil m/l \rceil \leq q \leq lm$ and $C(m, q)$ be a rescaled version of $C(q)$ of length m ,

$$C(m, q)_i = C(q)_{\lceil i \cdot q/m \rceil} \text{ where } 1 \leq i \leq m$$

$$\text{US}(C, Q, l) = \min_{q=\lceil m/l \rceil}^{\min(lm, n)} D(C(m, q), Q)$$

where $D(X, Y)$ denotes the Euclidean distance between two sequences X and Y .

Note that the ceiling function in the definition of $C(p, q)$ may be replaced by the floor function. The whole definition of $C(p, q)$ may also be replaced by some interpolation on the values of $C(q)_{\lfloor i \cdot q/p \rfloor}$ and $C(q)_{\lceil i \cdot q/p \rceil}$.

3.3.1 Lower bounding uniform scaling

We create two sequences $UC = UC_1, \dots, UC_m$ and $LC = LC_1, \dots, LC_m$, such that

$$UC_i = \max(C_{\lceil i/l \rceil}, \dots, C_{\lceil il \rceil})$$

$$LC_i = \min(C_{\lceil i/l \rceil}, \dots, C_{\lceil il \rceil})$$

These sequences bound the points of the time series C that can be matched with Q .

The lower bounding function, which lower bounds the distance between Q and C for any scaling ρ , $1 \leq \rho \leq l$, can now be defined as:

$$LB_S(Q, C) = \sum_{i=1}^m \begin{cases} (Q_i - UC_i)^2 & \text{if } Q_i > UC_i \\ (Q_i - LC_i)^2 & \text{if } Q_i < LC_i \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Lemma 1. *For any two sequences Q and C of length m and n respectively, for any scaling constraint on the warping path $w_k = (i, j)_k$ of the form $j/l \leq i \leq lj$, the value of $LB_S(Q, C)$ lower bounds the distance between C and Q under a scaling of C between $1/l$ and l , where $l \geq 1$.*

Proof. We can assume a matching path $w_k = (i, j)_k$ which defines a mapping between the indices i and j , so that each such mapping constitutes a term $(Q_i - C_j)^2$ to the required distance. We will show that each term t_{lb} in the square root of our lower bounding distance $LB_S(Q, C)$ can be matched with a term t resulted from the one-to-one mapping, with $t_{lb} \leq t$.

Based on the constraints on the scaling factor, we have the constraint $j/l \leq i \leq lj$ between i and j in $w_k = (i, j)_k$. From this, we have $i/l \leq j \leq il$ and by definition

$$\begin{aligned} UC_i &= \max(C_{\lceil i/l \rceil}, \dots, C_{\lceil il \rceil}) \\ LC_i &= \min(C_{\lceil i/l \rceil}, \dots, C_{\lceil il \rceil}) \end{aligned}$$

thus $UC_i = \max(C_{\lceil i/l \rceil}, \dots, C_j, \dots, C_{\lceil il \rceil}) \geq C_j$, or

$$Q_i - UC_i \leq Q_i - C_j$$

If $Q_i > UC_i$ then $Q_i - UC_i > 0$, hence

$$(Q_i - UC_i)^2 \leq (Q_i - C_j)^2$$

Similarly we can show that if $Q_i < LC_i$ then

$$(Q_i - LC_i)^2 \leq (Q_i - C_j)^2 \quad \square$$

4 Scaling and Time Warping

Having reviewed time warping, uniform scaling, and lower bounding, this section introduces *scaling and time warping* (SWM).

Definition 5 (Scaling and Time Warping (SWM)). *Given two sequences $Q = Q_1, \dots, Q_m$ and $C = C_1, \dots, C_n$, a bound on the scaling factor $l, l \geq 1$ and the Sakoe-Chiba Band time warping constraint r which applies to sequence length m . Let $C(q)$ be the*

prefix of C of length q , where $\lceil m/l \rceil \leq q \leq \min(lm, n)$ and $C(m, q)$ be a rescaled version of $C(q)$ of length m ,

$$\begin{aligned} C(m, q)_i &= C(q)_{\lceil i \cdot q/m \rceil} \text{ where } 1 \leq i \leq m \\ \text{SWM}(C, Q, l, r) &= \min_{q=\lceil m/l \rceil}^{\min(lm, n)} \text{cDTW}(C(m, q), Q, r) \end{aligned}$$

To simplify our discussion we shall assume that $lm \leq n$. We are interested in being able to scale the sequence and also to find nearest neighbor or evaluate range query by means of time warping distance. As noted in [14], a naïve search for the uniform scaling problem alone requires $O(|D| \cdot (a - b))$ time, where $[b, a]$ is the range of lengths resulting from scaling. Time warping computation alone requires $O(n^2)$ time for time series length of n . Hence we need to find a more efficient technique for the SWM problem.

In previous sections, we reviewed the lower bounding technique for each sub-problem. Here, we propose to combine these lower bounds to form overall lower bounds for the querying problem. Figure 4 illustrates this graphically.³

We apply time warping on top of scaling, i.e. we scale the sequence first, and then measure the time warping distance of the scaled sequence with the query. Typically, time warping with Sakoe-Chiba Band constrains the warping path by a fraction of the data length, which is translated into a constant r . Hence, if the fraction is 10%, then $r = 0.1|C|$. If the length of C is changed according to the scaling fraction ρ , that is, C is changed to ρC , then the Sakoe-Chiba Band time warping constraint is $r = 0.1|\rho C|$. Hence, we have $r = r'\rho$, where r' is the Sakoe-Chiba Band time warping constraint on the unscaled sequence, and ρ is the scaling factor.

The lower envelope L_i and upper envelope U_i on C can be deduced as follows: Recall that the upper and lower bounds for uniform scaling between $1/l$ and l is given by the following:

$$\begin{aligned} UC_i &= \max(C_{\lceil i/l \rceil}, \dots, C_{\lceil il \rceil}) \\ LC_i &= \min(C_{\lceil i/l \rceil}, \dots, C_{\lceil il \rceil}) \end{aligned}$$

and the upper and lower bounds for a Sakoe-Chiba Band time warping constraint factor of r for a point C_i is given by:

$$\begin{aligned} UW_i &= \max(C_{\max(1, i-r)}, \dots, C_{\min(i+r, n)}) \\ LW_i &= \min(C_{\max(1, i-r)}, \dots, C_{\min(i+r, n)}) \end{aligned}$$

Therefore, when we apply time warping on top of

³In this example, the scaling factor is $l = 1.5$, the time warping constraint is $r' = 10\%$ of the length of C .

