韓國 軍事運營分析 學會誌
第17卷, 第1號, 1991.6.30.

DYNAMIC TIME WARPING METHOD AND ITS APPLICATION

Youn, Sang-Youn*
Kim, Woo Youl**

Abstract

Dynamic Time Warping (in short DTW) is a kind of sequence comparison method. It is widely used in human speech recognition. The timing difference between two speech patterns to be compared is removed by warping the time axes of the speech pattern by minimising the time-normalised distance between them. In the process of finding the minimum time-normalised distance, the efficient method is dynamic programming problem. This paper describes the concept of dynamic time warping method, mathematical formulation and an application.

1. INTRODUCTION

Dynamic time warping is a widely used method in human speech recognition. In human speech, the speaking rate is different from person to person. The different speaking rate causes a variation in time axes when the spoken word is expressed as a waveform feature vector in accordance with time. The elimination of speaking rate variation, or speaking

* Combined Forces Command.
** Dong-Shin University.

time-normalisation, has been one of key problems in human speech recognition. The removal of variation can be achieved by introducing an appropriate transformation of time axis. This transformation is called time warping. In an early stage of human speech recognition linear transformation was used. But several reports showed that the linear transformation, which strictly changes the time scale by linear compression or expansion, is not only insufficient for dealing with the waveform speech feature vectors but it does not give a high rate of correct recognition.

To compensate for the local compression or expansion of the time scale nonlinear time warping is required. For this purpose, the algorithm known as "Dynamic Time Warping" method has been developed and improved the accuracy of speech recognition. In DTW, timing difference between two speech patterns to be compared is removed by warping the time axes of the speech pattern by minimising the time-normalised distance, the efficient method is a dynamic programming. So we call the algorithm Dynamic Time Warping.

2. FORMULATION OF DTW ALGORITHM

DTW works by mapping the axis of a test pattern onto the axis of a reference pattern by minimising the resulting dissimilarity or distane between two patterns. The problem is to find the best way to minimise the dissimilarity between two patterns by scale change of a test pattern relative to the reference pattern. The scale is changed by appropriate local compression, expansion and matching of a test pattern relative to the reference pattern. In the course of finding the best local compression and expansion DTW finds a one-to-one corresponding axis which achieves the resulting minimum distance between two patterns. The connection of the one-to-one corresponding axis constitutes a path on the grid point set. Thus, DTW can be formulated as a path finding problem over a finite set of grid points such as shown in figure 1, where the path parallel to m-axis, n-axis and diagonal path represent expansion, compression and matching, respectively.

Let us consider two patterns of R and T which represent reference and test pattern, respectively.

$$R = \{r_1, r_2, ---, r_m, ---, r_M\} \quad (1a)$$

$$T = \{t_1, t_2, ---, t_n, ---, t_N\} \quad (1b)$$

The goal of DTW is to find the optimal path,

$$P = c(1), c(2), ---, c(k), ---, c(K). \quad (2)$$

which is a sequence of grid points in the (m, n) plane which minimises the weighted total distance D(R, T) of the form

$$D(R, T) = \sum_{k=1}^{K} w(k) d(c(k)) \quad (3)$$

where $c(k) = (i(k), j((k)), d(c(k))$ is the local distance between two vectors $r_i(k)$ and $t_j(k)$ and $w(k)$ is the weight assigned to $c(k)$. In figure 1(a) the global path P is a sequence of local path segments which are in monotonic increasing on the grid points.

$$P = p(1), p(2), ---, p(7), p(8)$$

where

p(1) = segment from grid point (0, 0) to grid point (1, 1)

p(2) = segment from grid point (1, 1) to grid point (2, 2)

.

.

p(8) = segment from grid point (5, 6) to grid point (6, 7)

If we ignore the starting grid points and write

c(1) = (1, 1)

c(2) = (2, 2)

.

.

c(8) = (6, 7)

then equation (2) holds.

Now, assign a weight to each path segment. Then, the equation (3) for the path of figure 1(a) becomes

$$D(R, T) = w(1)d(r_1, t_1) + w(2)d(r_2, t_2) + , \cdots, + w(8)d(r_6, t_7)$$

If weights $w(i)$, $i = 1, 2, ---, 8$, and distance function $d(.)$ are defined we can obtain the total distance D(R, T). However, the problem is how we can find the optimal path P. There are so many possible routes to reach grid point (6, 7) from grid point (0, 0). Among them we can find the optimal path efficiently by dynamic programming formulation suitably define the possible path (local path type), legal region of the path (barrier constraint) and weights. Details are described in the subsequent sections.

2.1 WARPING FUNCTION AND GLOBAL PATH

The path P forms a sequence of points on the grid of points in the (m, n) plane. Each grid point on the path P is matched by mapping the reference pattern axis via a function f (.), that is

$$n = f(m) \tag{4}$$

The function (4) is called "warping function".

Let us consider the formulation (4) and also consider the path P in (2). The warping function f(.) makes a sequence of points in (m, n) plane, where reference pattern R and test pattern T are depicted in m-axis and n-axis, respectively. We seek to create a common axis k, expressing the axes m and n as a function of k.

$$\left.\begin{aligned} m &= i(k),\ k = 1, 2, ---, K \\ n &= j(k),\ k = 1, 2, ---, K \end{aligned}\right] \tag{5}$$

where K is the length of common axis, (K = 8 in figure 1)

The path P, which is a sequence of points of (i(k), j(k)) with length K of common axis, and which is created by the warping function f(.), is called "global path." We can obtain the relationship between warping function f(.) and the common axis k. Setting equal two functions of (4) and (5) yields

$$j(k) = n = f(m) = f(i(k)) \tag{6}$$

Thus, in DTW, to obtain the best global path P in (3) is equivalent to obtain the best function f(.) in (4). To obtain the optimal path or best warping function in (m, n) plane, we are to consider several factors affecting DTW in the subsequent sections.

2.2 CONSTRAINTS ON DTW

ASSUMPTION

Given a reference pattern R and a test pattern T, both patterns being measured in fixed intervals of time, distance, etc., then the comparison or matching of these two patterns is a problem of scale alignment in order to give the best similarity between two patterns.

2.2.1 ENDPOINT CONSTRAINTS

Let M_I and M_F be the initial and final points of the reference pattern and N_I and N_F denote those of the test pattern. Then the warping function defines the alignment of two endpoints.

2.2.1.1 STRICT ENDPOINT CONSTRAINTS (STANDARD DTW METHOD)

If we assume that all endpoints (both initial and final points) are clearly known then the endpoints are matching and can be written in the form.

$$\left.\begin{aligned} f(M_I) &= f(1) = N_I = 1 \text{ (for the initial point)} \\ f(M_F) &= f(M) = N_F = N \text{ (for the final point)} \end{aligned}\right] \quad (7)$$

Or

$$\left.\begin{aligned} i(1) &= M_I = 1,\ j(1) = N_I = 1 \text{ (for the initial point)} \\ i(K) &= M_F = M,\ j(K) = N_F = N \text{ (for the final point)} \end{aligned}\right] \quad (7')$$

Under these endpoint constraints, DTW works from the low left initial grid point $c(1) = (1, 1)$ and reaches the top right final grid point $c(K) = (M, N)$.

2.2.1.2 RELAXED ENDPOINT CONSTRAINTS

If the endpoints are not clearly defined, then imposing strict endpoint constraints may cause errors in endpoint matching. Thus the alignment of test and reference patterns may not be perfect if the strict endpoint constraints are applied. In speech recognition, if certain words have similar sounds and particularly if the sounding differences between the words exist at the beginning or the end, then the strict endpoint constraints may decrease correct recognition rate. In order to reduce such inaccurate matching Rabiner et al.(3) considered endpoint relaxed methods.

UNCONSTRAINED ENDPOINTS

In this method, the endpoints of (7) are relaxed for a certain number of frames. So,the new endpoint constraints are as follow

$$1 \leq f(1) \leq 1 + \delta \quad (8a)$$

$$N - \delta \leq f(M) \leq N \quad (8b)$$

where δ represents the maximum anticipated range of mismatching in time frames between two patterns to be compared.

UNCONSTRAINED ENDPOINTS WITH LOCAL MINIMUM

This is a version in which both of the endpoints are relaxed. The beginning endpoint constraint is the same as (8a), but no constraint is directly introduced in the final frame matching. The final frame matching is dependent on the global path. The global path is allowed to follow the legal region defined by n* and δ for each reference frame m, where n* is the test frame number which achieves the minimum cumulative distance to the grid point (m - 1, n), n = 1, 2, - - -, N. Thus, for each reference frame m DTW works in the legal region defined

$$n^* - \delta \leq n \leq n^* + \delta \qquad (9)$$

where n* = (n; Dc = min [Dc (m - 1, n)]), δ is the specified range and Dc is the cumulative distance from the beginning grid point to the grid point (m - 1, n)

So, the final frame matching has flexibility which is dependent on the locally minimum global path.

TWO PATTERNS RELAXED ENDPOINT CONSTRAINTS

In real life, a pattern may subject to contraction and/or expansion. The stretching or contraction especially at the endpoints may be expected during the data acquisition. We want to compensate the stretched and/or contracted endpoints on DTW. If a pattern is expanded at the initial-end or final-end we may cut off the expanded endpoint by a few frames. On the contrary, if both or either endpoint is contracted we do not need to match the amount of contracted frames of the pattern. Thus, we may ignore a few frames of reference profile. We may reflect these basic considerations in our DTW formulation.

In standard DTW (S-DTW) two endpoints should be matched. That is, DTW begins from grid point (1, 1) and finishes on the grid point (M, N). But if we allow flexibility on endpoints by relaxation then the beginning and finishing grid points are not fixed. It is determined by the data, ie, it is dependent on the test pattern structure. There are several alternatives; either both endpoints are expanded or contracted, or one is expanded and one is contracted, or none of them are expanded or contracted. For example, if we assume a test pattern is expanded by 2 frames at the initial-end also that it is contracted by 2 frames at the final-end, then DTW should be designed to find the minimum distance between two patterns from grid point (1, 3) to grid point (M-2, N). However, in general, we do not know which patterns are expanded or contracted by how much. Thus, DTW may be modified to find the minimum distance between two profiles including all the above possible cases. i.e., we assume any pattern may be contracted or expanded at any endpoints. But for convenience of calculation we assume that the maximum amount of contraction and/or expansion is limited to a specified number of frames.

Our new endpoint relaxation constraints are written in the form

$$1 \leq f\ (1+\alpha) \leq 1 + \beta \tag{10a}$$

$$N-\delta \leq f\ (M-\gamma) \leq N \tag{10b}$$

were

$0 \leq \alpha,\ \beta,\ \gamma,\ \delta \leq z$, z is a fixed integer indicating the maximum amount of frames for relaxation. and

if $\alpha \geq 1$ then $\beta = 0$

if $\beta \geq 1$ then $\alpha = 0$

if $\gamma \geq 1$ then $\delta = 0$

if $\delta \geq 1$ then $\gamma = 0$

Thus, DTW starts from one of the bottom-left grid points

(1, 1) (2, 1) (3, 1) - - - (α, 1)
(1, 2) (1, 3) - - - (1, β)

and finishes one of the top-right grid points

(M, N) (M-1, N) (M-2, N) - - - (M-γ, N)
(M, N-1) (M-N-2) - - - (M, N-δ)

Figure 2 shows different type of global path according to the different type of endpoint constraints.

2.2.2 LOCAL CONSTRAINTS

The global path P is a succession of local paths. The local compression and/or expansion of the test pattern with relative to the reference pattern depends on how the local paths are constructed. The local constraints are used to avoid excessive local expansion or compression of the scale and to form a connected path from the initial grid point to the final point.

2.2.2.1 CONTINUITY CONSTRAINTS

The DTW path links the grid points (figure 2(a) global path). The grid points on the global path are connected under restriction by a set of continuity constraints of the form.

$$
\begin{aligned}
& f(m+C_1) - f(m) \leq C_2 \\
& C_1,\ C_2 \geq 0,\ \text{integer} \\
& m = 1, 2 ---, M - C_1
\end{aligned}
\qquad (11)
$$

The bounds C_1 and C_2 define the amount of local compression or expansion allowable in the warping. Too large a value of C_1 and too small C_2 may cause severe expansion. On the contrary too large value of C_2 and too small C_1 may cause severe compression. Figure 3 shows an example of how the grid points are connected and how the warping works using the constraint (11).

In this example, C_1 and C_2 are defined as

$$
\begin{aligned}
& C_1 = 1 \\
& 0 \leq C_2 \leq 3
\end{aligned}
$$

The continuity constraint (11) can be written in another way by expressing it in terms of the common time axis k in the form

$$
\left.
\begin{aligned}
& 0 \leq i\ (k+1) - i\ (k) \leq C_1 \\
& 0 \leq j\ (k+1) - j\ (k) \leq C_2
\end{aligned}
\right] \qquad (12)
$$

From (12) the inequalities

$$
\left.
\begin{aligned}
& i(k+1) \geq i(k) \\
& j(k+1) \geq j(k)
\end{aligned}
\right] \qquad (13)
$$

hold and are called the monotonity conditions.

The paths in figure 3 can be written in the form of (12).

$$
\begin{aligned}
& i(k+1) - i(k) = 1 \\
& 1 \leq j(k+1) - j(k) \leq 3,\ \text{if } j(k) = (k-1) \\
& 0 \leq j(k+1) - j(k) \leq 3,\ \text{if } j(k) \neq (k-1) \\
& k = 1, 2, ---, K-1.
\end{aligned}
$$

Therefore, C=1 and C may take a value 0, 1, 2, or 3.

2.2.2.2 LOCAL PATH TYPE

Suitably chosen bound values in the continuity constraints guarantee a good matching. Different bound values may produce a different recognition result. So, they have to be determined to provide good results. The bound values in (11) or (12) allow movement along the grid points. They define the possible single grid movements from a grid point to next grid point. A local path type is the collection of all possible types of movement defined by the bound values in (11) or (12). DTW does not seek a global compression or expansion but a local compression or expansion allowable. The amounts of local compression and/or expansion are dependent on the local path type.

Figure 4 shows several local path types which have been proposed for use in DTW. The first four types were proposed by Sakoe and Chiba (5) and are called 'slope constraint' 0, 1/2, 1 and 2, respectively. Because if we define a and b as the number of consecutive movements in the direction m (or n) and diagonal direction, respectively, then s = b/a becomes the slope. For the type (a), the simplest type, the path movement may occur in three directions; parallel to n-axis, diagonal direction or parallel to m-axis. However, there is no restriction on the number of consecutive movements parallel to m-axis or n-axis. To prevent such unrestricted movements parallel to m or n-axis, another three types of path or more may be considered. Types (b), (c) and (d) are extensions of type (a). All these three types involve the intermediate points. The grid points (m-1, n) and (m, n-1) are intermediate points for type (c) and (d), and two more points (m-2, n) and (m, n-2) are intermediate points for type (b). The maximun expansion allowed by each of these four types is ∞, 3, 2 and 3/2, respectively. Or, inversely the minimum expansions are 0, 1/3, 1/2 and 2/3, respectively. The last type (type(h)) is proposed by Itakura (1). In this type, any two consecutive movements parallel to m-axis are prevented. However, no movement parallel to the n-axis is allowed.

The remaining three types (e), (f) and (g) are proposed by Myers et al. (2). Type (e) has the same initial and final grid points as type (c). But the paths reach directly to the grid point (m, n) without going through the intermediate grid point (m-1, n) or (m, n-1). Type (f) is an extended version of type (h). The difference is whether the movement from (m-1, n) to (m, n) is allowed or not when the best path to the grid point (m-1, n) is coming from the grid point (m-2, n). For type (h), whenever the best path reaching the grid point (m-1, n) comes from the grid point (m-2, n) it is not permitted to reach grid point (m, n) from the point (m - 1, n). Therefore, the path is completely deleted from all subsequent paths. However, for type (f), even if the best path reaching to the grid point (m - 1, n) comes from the grid point (m - 2, n); it is possible to reach to the grid point (m, n) from the grid point (m - 1, n). Type (g) is a more expanded version of type (f). Form type (e) to type (h), the maximum compressions are 2, 2, 3 and 2, respectively.

Sakoe and Chiba (5) reported that (c) showed the best discriminant results for the speech recognition. Three path movements, p1, p2 and p3, constitute local path type (c). Local path movement p1 represents a matching and an expansion, p3 represents a matching and a compression, p2 represents just a matching. Let us consider each local path movement separately. For local path movement p1, the (n - 1)th test element is matching to the (m - 2)th reference element, and the n th test element is matching to both the (m - 1)th and m th reference element, that is, the n th test element is expanded. For local path movement p2, the (n - 1)th test element is matching to the (m - 1) th reference element. For local path movement p3, the (n - 2)th test element is matching to the (m - 1) th reference element, and both the (n - 1) th and n th test element is matching to the m th reference element, that is, the (n - 1) th and n th test elements are compressed into a single element.

2.2.3 BARRIER CONSTRAINTS

The barrier constraints are used either to include certain parts of the (m, n) plane from the region where the optimal path may lie (Myers et al.(2)) or to avoid excessive compression or expansion between two patterns (Sakoe and Chiba (5)).

Let us assume that warping works from the grid point (1, 1) to the grid point (M, N). Then, the link of the local path grid point (1, 1) to grid point (M, N) forms a global path. Figure 5 shows an example of the legal region of DTW. Two straight lines a and a' represent, respectively, the boundaries formed by path with the maximum conpression and the maximum expansion from the initial grid point (1, 1). Whereas, two straight lines b and b' represent, respectively, the boundaries formed by paths reaching to the final grid point (M, N) with the maximum compression and the maximum expansion. Therefore, the parallelogram ABCD constitutes a legal region of the global path. From any grid point out of the region of the parallelogram, no combination of local paths can reach the final grid point (M, N) with the given path type.

A set of relations can be obtained for expressing the boundaries of the legal global path, Let Smax be the maximum slope of the local path type and assume that the minimum slope is the reciprocal of the maximum slope. The straight lines a, a', b and b' are expressed as

$$\text{line a} : m = Smax\ (n-1) + 1 \tag{14a}$$

$$\text{line a'} : m = 1/Smax\ (n-1) + 1 \tag{14a'}$$

$$\text{line b} : m = Smax\ (n-N) + M \tag{14b}$$

$$\text{line b'} : m = 1/Smax\ (n-N) + M \tag{14b'}$$

By substituting m and n as i(k) and j(k), respectively, and combining both two equations (14a) and (14a'), (14b) and (14b') together a set of inequalities are obtained.

$$\frac{1}{Smax}\,(i(k)-1) + 1 \leq j(k) \leq Smax\ (i(k)-1) + 1 \tag{15a}$$

$$Smax\ (i(k)-N) + M \leq j(k) \leq \frac{1}{Smax}\,(i(k)-N) + M \tag{15b}$$

The inequality (15a) restricts the range of grid points which can be reached from the initial grid point (1, 1), whereas (15b) restricts the range of grid points from which the final grid point (M, N) can be reached.

Another barrier constraint to the global path, proposed by Sakoe and Chiba (5), takes the form

$$|\ i(k) - j(k)\ | \leq r \tag{16}$$

Where r is a nonnegative integer and is the maximum allowed frame difference between test and reference pattern. Inequality (16) is called by Sakoe and Chiba (5) th 'adjustment window condition.' This constraint (line c and c' in figure 5) is introduced to limit the maximum possible misalignment at any stage between two patterns, and cuts off the corners of the parallelogram. If we introduce the constraint (16), then DTW can be done within the hexagon AEFCGH of figure 5.

2.4 DISTANCE MEASURE

The dissimilarity between two patterns to be compared is defined by the distance. If one of the patterns is a reference pattern for a certain group this distance is used to recognise or to classify a test pattern into that group.

A general form of the distance function is

$$\bar{D}(c(K)) = \frac{\sum_{k=1}^{K} w(k)\, d(c(k))}{v(w)} \qquad (17)$$

where $c(k) = (i(k), j(k))$, $d(c(k))$ is the local distance between $r_i(k)$ and $t_j(k)$, $w(k)$ is the weighting function of the k the local path of the global path p and $v(w)$ is a normalisation factor which is a function of the weighting function $w(.)$. Therefor, $\bar{D}(c(K))$ is a function of a set of functions and reflects the normalised distance along the global path P of length K. The optimal path can be defined as the global path that minimises the distance $\bar{D}(c(K))$ of equation (17).

Let Dopt be the nomalised distance of the optimal path, then holds.

$$Dopt = \min_{\{c(k)\}} \bar{D}(c(K)) \qquad (18a)$$

Or

$$Dopt = \min_{P} \bar{D}(c(k)) \qquad (18b)$$

For the computation of (18) the functions: the local distance $d(.)$, weighting function $w(.)$ and normalisation factor $v(.)$ must be specified.

2.4.1 LOCAL DISTANCE

Several kinds of local distance measure have been used or can be used in DTW:

Chebyshev norm (city bolck distance)

Chebyshev norm with local difference

Euclidean norm

The Chebyshev norm, used by Sakoe and Chiba (5), finds the absolute value of difference between $r_i(k)$ and $t_i(k)$ and sums the entire differences along the global path. Instead of the absolute value of difference, the Euclidean norm finds the Euclidean distance between $r_i(k)$ and $t_j(k)$.

The Chebyshev norm with local difference is proposed one for this study. The Chebyshev norm does not give sufficiently good classification result in chromosome data. Because it does not measure the difference between two chromosome optical density profiles (patterns) efficiently. Even though test pattern looks similar to a reference pattern R, too low or too high optical density profile than the true group reference profile causes quite big distance between two optical density profiles. This may cause a misclassification. To recover from this situation, we *modified the local distance function as*

$$d(i, j) = | (r_i - r_{i-1}) - (t_j - t_{j-1}) | \qquad (19)$$

$$= | \Delta r_i - \Delta t_j |$$

This distance function measures the difference of local difference between two patterns. Now, let us call this local distance function as the Chebyshev norm with the local difference and DTW with the distance function as DTW with the local difference distance (in short LD-DTW).

2.4.2 WEIGHTING FUNCTION

The weighting function w depends on the movement of the local path : parallel to the n-axis, parallel to the m-axis or a diagonal movement. Several types of weighting functions have been used. Sakoe and Chiba (5) proposed four types of weighting functions :

Type 1 : $w1(k) = (i(k) - i(k-1) + (j(k) - j(k-1))$

Type 2 : w2 (k) = i(k) - i(k-1)

Type 3 : w3 (k) = max (i(k) - i(k-1), j(k) - j(k-1)).

Type 4 : w4 (k) = min (i(k) - i(k-1), j(k) - j(k-1))

The weighting function type (1) weights the arcs according to the sum of grid units moved to m and n directions. The weighting function type (2) weights the arcs according to the grid units moved along m-axis only. The weighting function type (3) weights the arcs according to the maximum grid movements along m or n axis, that is, the maximum expansion or compression. For weighting function type (4), the weights are determined according to the minimum grid increment to m-axis or n-axis. Therefore, all arcs parallel to any axes have zero weights.

The weighting function type (2) and (4) may cause 0 weight on some arcs. In such cases, the local distance does not contribute to the total distance. To prevent such nonphysical factor, a smoothing function on the weights may be applied. The smoothing function give each arc the average weights along the multiple arcs of the local path. Figure 6 shows an example of the original weights and smoothed weights on local path type (c).

2.4.3 NORMALISATION FACTOR

The normalisation factor is used to obtain the average local distance from the total cumulative distance. The resulting distances obtained from the comparison of a test pattern with several different group reference patterns with different length may be different because of the length. To eliminate the distance difference caused by the different length it is appropriate to find the average distance according to the lengths of test and reference patterns. It assures that the average local distance is independent upon the lengths of test and reference pattern and the length of any particular global path P. We define the normalisation factor v(w) as the sum of weights applied to each local path movement. Thus, it takes the form

$$v(w) = \sum_{k=1}^{K} w(k) \qquad (19)$$

For weighting function type (1) and (2)

$$v(w) = \sum_{k=1}^{K} (i(k) - i(k-1) + (j(k) - j(k-1))) \qquad (20)$$

$$= i(k) - i(0) + j(k) - j(0)$$

$$= M + N$$

$$v(w) = \sum_{k=1}^{K} (i(k) - i(k-1)) \qquad (21)$$

$$= i(k) - i(0) = M$$

For above two cases, the normalisation factors can be obtained easily and fixed. However, for type (3) and type (4), the normalisation factors dependent on the global path. Thus, last two types require extra calculation rather than first two types. So, $\overline{D}(.)$ can not be normalised enroute.

2.5 CALCULATION OF DTW NORMALISATION DISTANCE

In order to do dynamic time warping, let us express the path finding problem as a standard dynamic programming recurrence relation. For any grid point (m, n) which lies within the barrier constraints, can be expressed as

$$\text{Minimize } D(m, n) = \min \sum_{k=1}^{K} [w(k)d(c(k))] \qquad (22)$$

Subject to

$c(k) = (i(k), j(k))$

$i(k) = m$

$j(k) = n$

$1 \le m \le M$ and $1 \le n \le N$

K : length of global path

where $d(c(k))$ is the local distance between two elements and $w(k)d(c(k))$ serves as the weighted local distance. Thus, $D(m, n)$ is the cumulative weighted distance and can be obtained by dynamic programming. For local path type (c) and type (1) weight we can write DP recursion as

$$D(M, N) = \min \begin{cases} D(m-2, n-1) + 2d(m-1, n) + d(m, n) \\ D(m-1, n-1) + 2d(m, n) \\ D(m-1, n-2) + 2d(m, n-1) + d(m, n) \end{cases} \tag{23}$$

If $m = M$ and $n = N$ then the normalised distance $\bar{D}(M, N)$ can be obtained by

$$\bar{D}(M, N) = \frac{D(M, N)}{v(w)} = \frac{\min \left[\sum_{k=1}^{K} w(k)d(c(k)) \right]}{v(w)} \tag{24}$$

If the normalisation factor $v(w)$ is independent of the global path, then $v(w)$ can be fixed as a constant r. Thus, (24) can be written as

$$\bar{D}(M, N) = \frac{1}{r} \min \sum_{k=1}^{K} \left[w(k)d(c(k)) \right] \tag{25}$$

The solution of (25) can be found by a procedure:

Step 1 : initialization

SET $D(1, 1) := w(1)d(1, 1)$

$= 2d(1, 1)$

Step 2 : recursion

computer $D(m, n)$ recursively for $1 \leq m \leq M$, $1 \leq n \leq N$ by the recursion (23)

Step 3 : termination

If $m = M$ and $n = N$ then find the normalised distance $\bar{D}(M, N)$

Step 4 : backtracking the optimal path

An optimal path is a global path which gives the best local compression, expansion and matching between two patterns or which gives the minimum distance between two patterns. We can obtain the optimal path by backtracking the local path movements along the global path which gives the minimum distance between a test pattern and a reference pattern. This is an optional step to find the correct normalisation factor for the endpoint relaxed DTW (section 2.2.1).

Step 4-1

Set I := M

J := N

Step 4-2 : backtracking the DP recursion (Eq. 23)

$$k := \begin{cases} 1 & \text{if } D(I, J) = D(I-2, J-1) + 2d(I-1, J) + d(I, J) \\ 2 & \text{if } D(I, J) = D(I-1, J-1) + 2d(I, J) \\ 3 & \text{if } D(I, J) = D(I-1, J-2) + 2d(I, J-1) + d(I, J) \end{cases}$$

i.e. k indicates the local path movement type which gives the minimum distance to the grid point (I, J) from three predecessors (I-2, J-1), (I-1, J-1) or (I-1, J-2).

IF k = 1

MPATH (I-1, J) = 1 {variable MPATH indicates the grid

MPATH (I-2, J-1) = 1 point which lies on the optimal path}

```
    I = I - 2
    J = J - 1
    GO TO STEP 4 - 3
IF k = 2
    MPATH (I-1, J-1) = 1
    I = I-1
    J = J-1
    GO TO STEP 4-3
IF k = 3
    MPATH (I, J-1) = 1
    MPATH (I-1, J-2) = 1
    I = I-1
    J = J-2
STEP 4-3
    IF I = J = 1
        obtained the optimal path
    ELSE GO TO STEP 4-2
```

By defining the variable MPATH we can easily distinguish the optimal path. At any grid point, if variable MPATH takes the value 1, then the grid point lies on the optimal path.

Ⅲ. APPLICATION

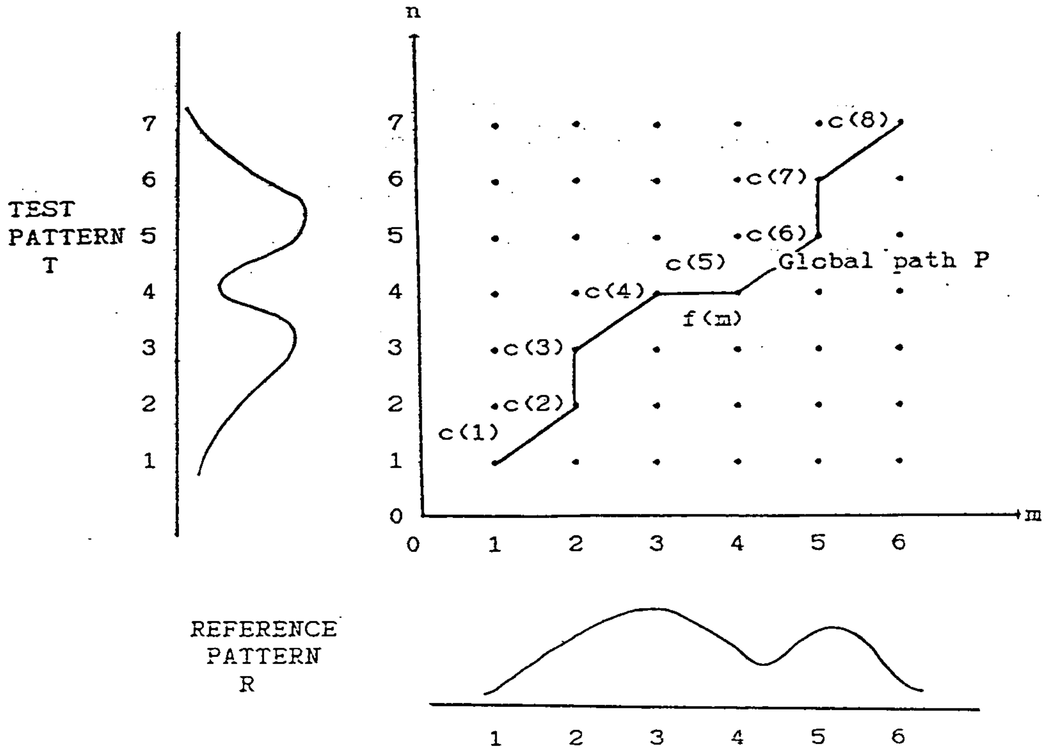
We applied DTW to human chromosome optical density profile comparison. An optical density profile is a kind of waveform feature vector whose shape is different according to the chromosome group. A normal human cell contains 46 chromosomes. The female has 23 pairs of chromosomes whereas the male has 22 pairs and an X and a Y chromosome. A test pattern was compared to several different group reference patterns. In each time the distance between a test pattern and a reference pattern was calculated by the method described above. Figure 7 shows an example of sequence (pattern or profile) comparison result where we can see that DTW well matches the peaks and valleys between two patterns. The length difference between two patterns is removed by the best local compression and expansion. Some parts of the test pattern are stretched while some other parts are compressed with relative to the reference pattern. DTW results a distance and sequence (pattern) matching results which indicate the corresponding (matching) elements between two patterns.

For a pattern classification by using the DTW method the obtained distance (similarity) between two patterns are used. Normally a test pattern is classified into a reference pattern group which gives the minimum distance.

Ⅳ. DISCUSSION

In this paper we considered DTW method. In section 2, we considered DTW method and its mathematical formulation. Two improvements are described : new local distance measure and new endpoint relaxation method. In experiments our new local distance measure, the Chebyshev norm with local difference, finds more accurate similarity between two patterns. Especially when a test pattern is similar in shape to the reference one but the magnitudes are too low or too high relative to the reference one. New endpoint relaxed method also contributed to find more accurate similarity between two patterns. The expansion and/or contraction of an image may trivially appear in real life. Thus, our new endpoint relaxed DTW may usefully be applied to other application area.

(a)



(b)

common axis k	1	2	3	4	5	6	7	8
i(k) = m	1	2	2	3	4	5	5	6
j(k) = n	1	2	3	4	4	5	6	7
c(k) = (i(k), j(k))	(1, 1)	(2, 2)	(2, 3)	(3, 4)	(4, 4)	(5, 5)	(5, 6)	(6, 7)

m	1	2	3	4	5	6
f(m) = n	1	2, 3	4	4	5, 6	7

Figure 1.

(a) An example of grid point (centre)

A global path P = c(1), c(2), - - -, c(8), and warping function f(m).

(b) A numerical example of i(k), j(k) and c(k) as functions of common axis k.

And the warping function n = f(m)

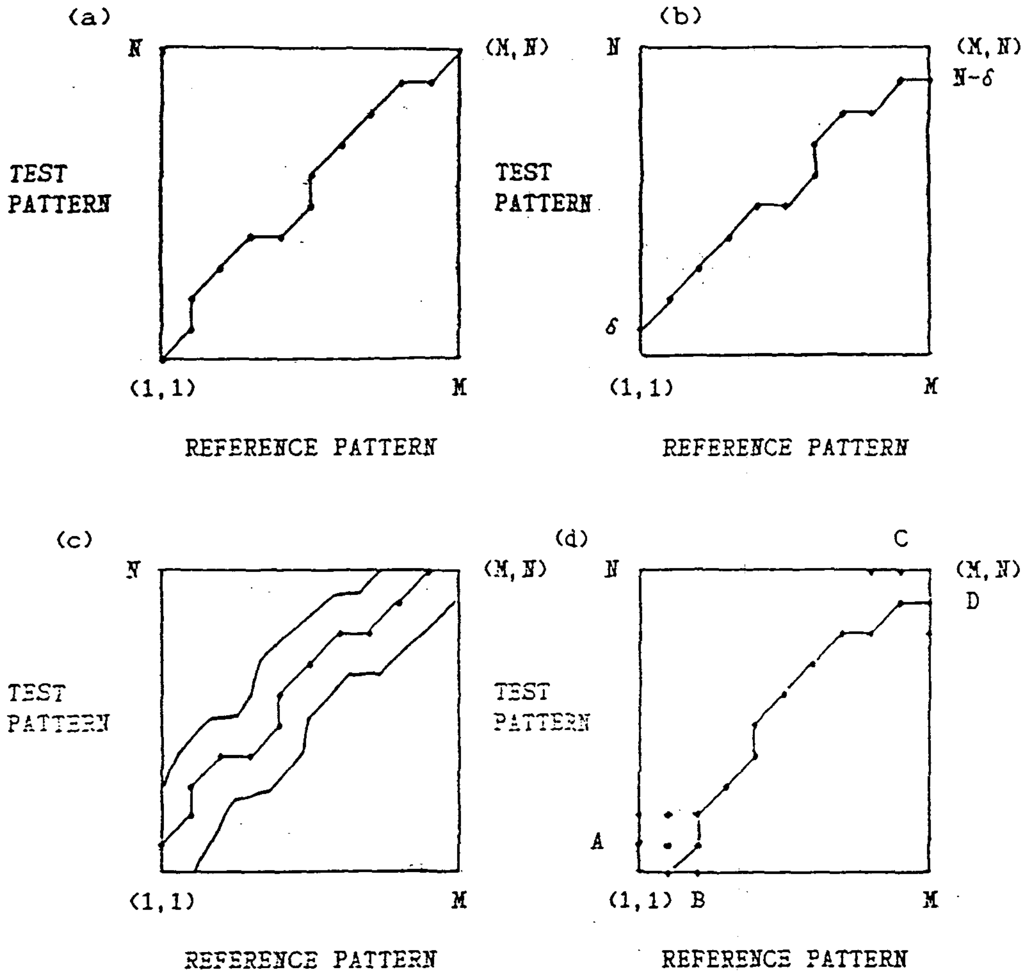


Figure 2. Illustrations of DTW method by different endpoint constraints

(a) Standard DTW (strict endpoint constraints)

(b) Unconstrained endpoint

(c) Unconstrained endpoint with local minimum

(d) New endpoint relaxation : final endpoint C, (M, N) of D can be reached from initial endpoint A, (1, 1) or B.

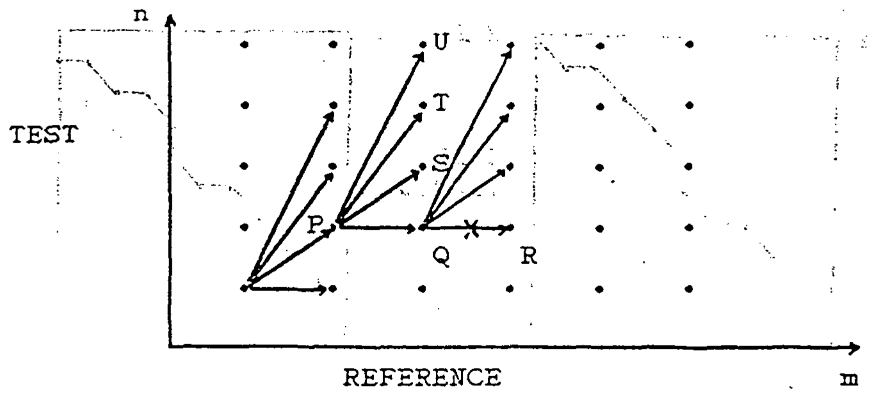


Figure 3. An example showing the continuity constraints (11).

Form point P, 4 points Q, S, T and U can be reached. From point Q, the movment to the grid point R is not allowed.

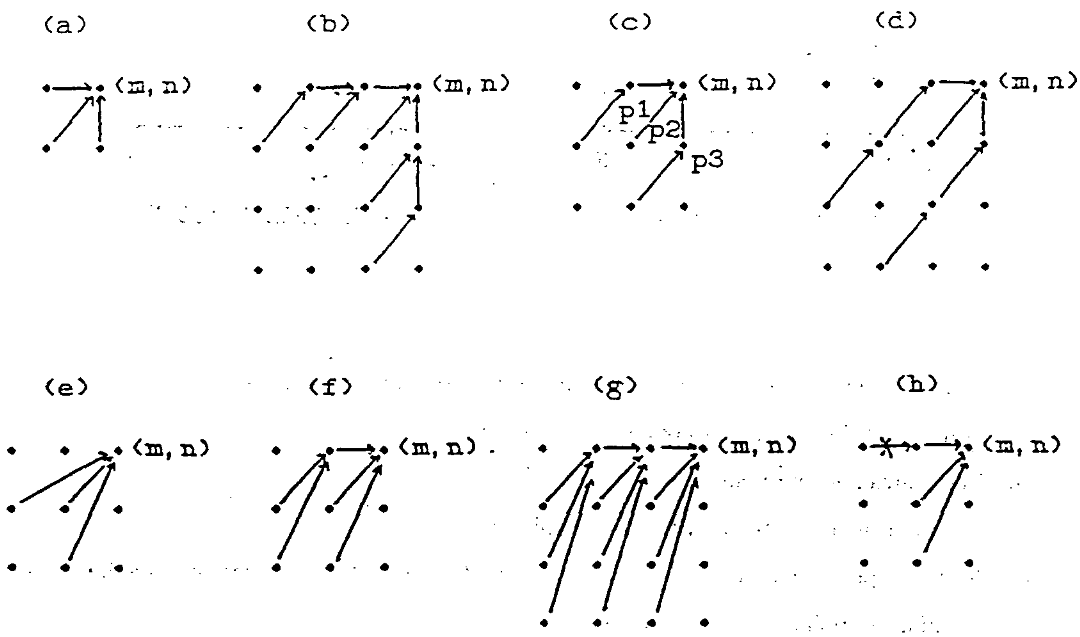


Figure 4. Local Path Types

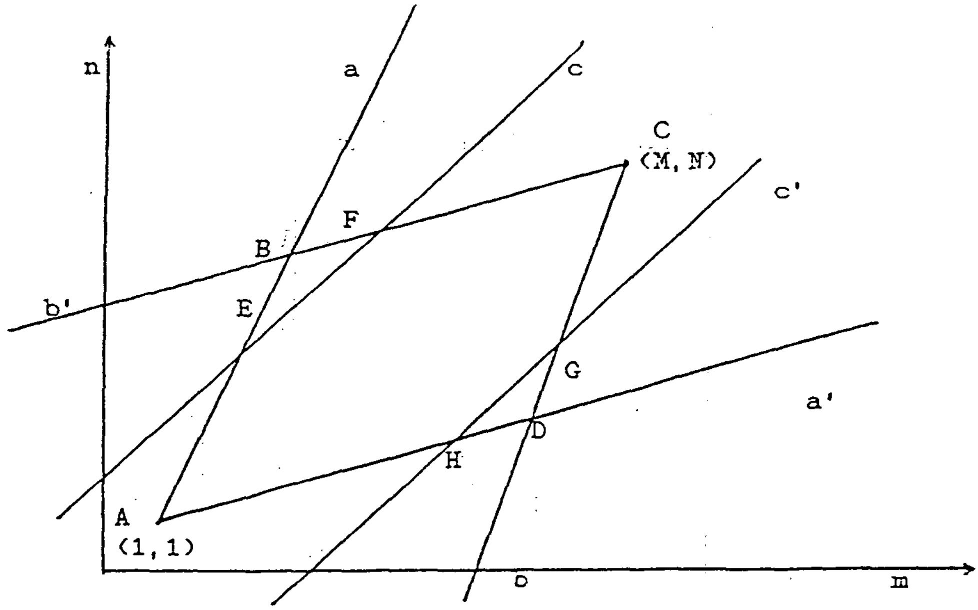


Figure 5. An example of the barrier constraints

weighting function type	original weight smoothed weight
$W1(k)=(i(k)-i(k-1)+j(k)-j(k-1))$	
$W2(k)= i(k) - i(k-1)$	
$W3(k)=\max\begin{cases} i(k)-i(k-1) \\ j(k)-k(k-1)\end{cases}$	
$W4(k)=\min\begin{cases} i(k)-i(k-1) \\ j(k)-j(k-1)\end{cases}$	

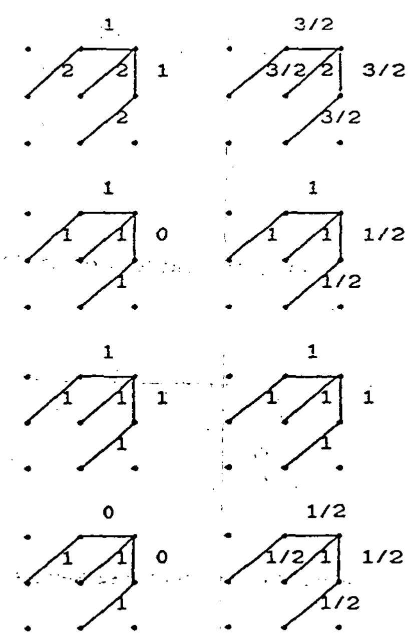


Figure 6. An example of weighting functions and smoothed weights

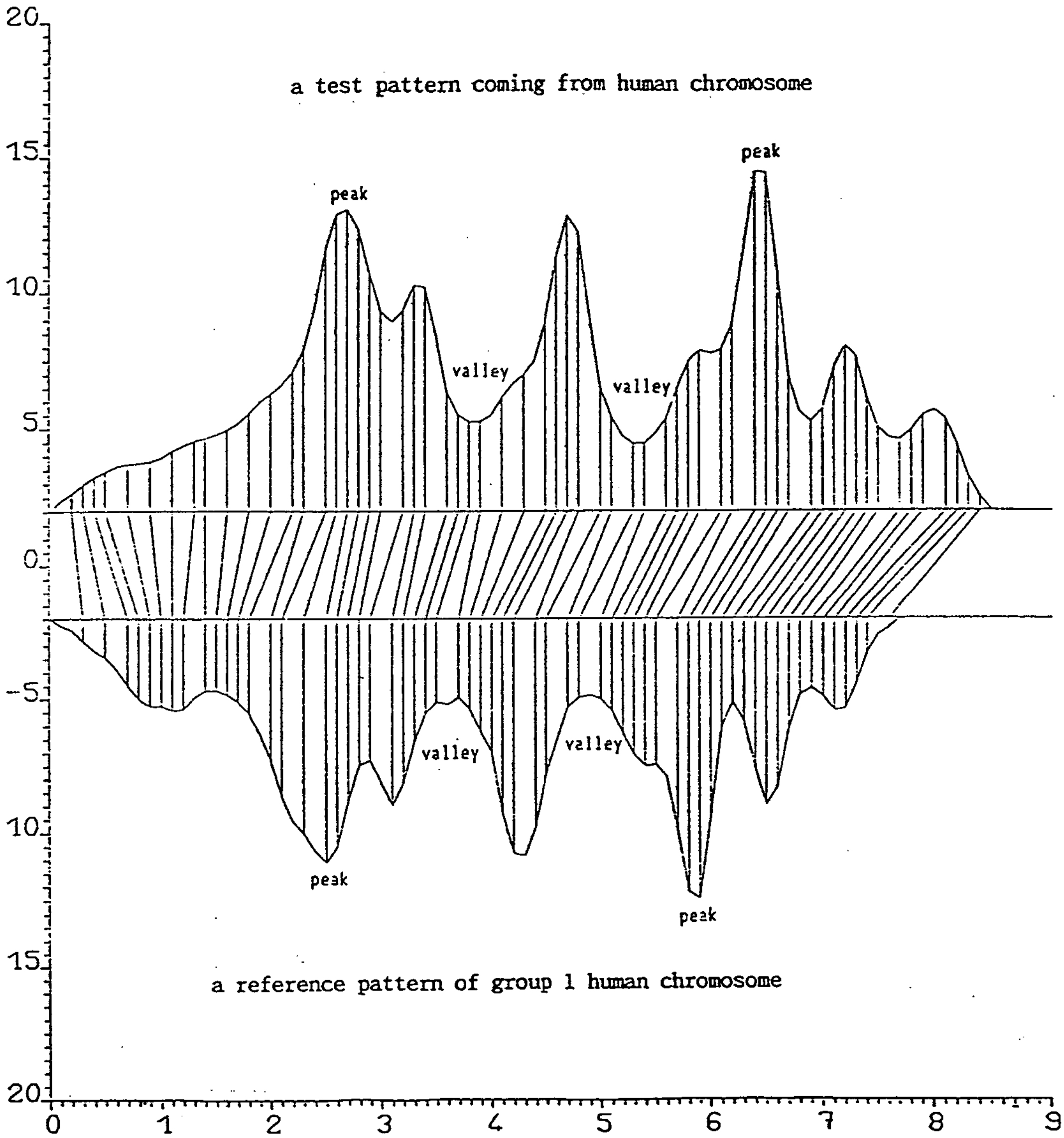


Figure 7. An example of showing sequence comparison (pattern matching).

REFERENCES

1. Itakura F.; "Minimum prediction residual principle applied to speech recognition", *IEEE Trans. on Acoustics, Speech, and Signal Processing ASSP*-23, 67-72, 1975.
2. Myers C., L.R. Rabiner and A.E. Rosenberg; "Performance tradeoffs in dynamictime warping algorithms for isolated word recognition", *IEEE Trans. on Acoustics, Speech, and Signal processing ASSP*-28, 623-635, (1980).
3. Rabiner L.R., A.E. Rosenberg and S.T. Levinson; "Considerations in dynamictime warping algorithm for discrete word recognition", *IEEE Trans. on Acoustics, Speech and Signal Processing ASSP*-26, 575-582, 1978.
4. Sakoe H. and S. Chiba; "*A dynamic programming approach to continuous speech recognition*", 7th Int. Congress on Acoustics, Budapest, 1971, 65-68.
5. Sakoe H. and S. Chiba; "Dynamic Programming Algorithm optimization for spoken word recognition", *IEEE Trans. on Acoustics, Speech, and Signal processing, ASSP*-26, 43-49, 1978.
6. Sankoff D. and J.B. Kruskal; Time Warps, String Edits, and Macromolecules: *The theory and practice of Sequence Comparisons*, Addison-wesley, 1983.
7. Youn S.Y.; "Dynamic Programming Based Automatic Image Analysis Systems", Ph.D thesis, Department of Mathematics, University of Manchester Institute of Science and Technology (UMIST), England, 1990.