Dynamic Shapes Average*

Pierre Maurel

École Normale Supérieure
45 rue d'Ulm
75005 Paris, France
e-mail: Pierre.Maurel@ens.fr

Guillermo Sapiro

Electrical and Computer Engineering
University of Minnesota
Minneapolis, MN 55455
e-mail: guille@ece.umn.edu

Abstract

A framework for computing shape statistics in general, and average in particular, for dynamic shapes is introduced in this paper. Given a metric $d(\cdot,\cdot)$ on the set of static shapes, the empirical mean of N static shapes,$C_1,\ldots,C_N$, is defined by $\arg\min_C \frac{1}{N}\sum_{i=1}^{N} d(C,C_i)^2$. The purpose of this paper is to extend this shape average work to the case of N dynamic shapes and to give an efficient algorithm to compute it. The key concept is to combine the static shape statistics approach with a time-alignment step. To align the time scale while performing the shape average we use *dynamic time warping*, adapted to deal with dynamic shapes. The proposed technique is independent of the particular choice of the shape metric $d(\cdot,\cdot)$. We present the underlying concepts, a number of examples, and conclude with a variational formulation to address the dynamic shape average problem. We also demonstrate how to use these results for comparing different types of dynamics. Although only average is addressed in this paper, other shape statistics can be similarly obtained following the framework here proposed.

1 Introduction

Understanding shape and its basic empirical statistics is important both in recognition and analysis, with applications ranging from medicine to security to consumer photography. The basic metrics and statistics of static shapes have been the subject of numerous fundamental studies in recent years, see for example [1, 3, 4, 6, 8, 13, 15, 16] and references therein. In particular, given a metric on the set of static shapes (distance between two samples), the empirical mean shape of N static shapes, as well as other basic statistics, can be defined and computed. These are then used for diverse shape studies, from the recognition of particular objects to the detection of abnormalities in medical data. The purpose of this work is to extend this to dynamic shapes. This is fundamental for studies such as those involving gait, behavior, growth patterns, and all problems involving motion, deformations, and time-varying shapes.

Given N dynamic shapes $\Gamma_1(t),\ldots,\Gamma_N(t)$ (t stands for the time parameter, see Figure 1), we want to find $M(t)$, the empirical mean of these shapes. This basic computation will be used throughout this paper as an example of how to perform statistics on dynamic shapes. One idea could be to simply perform static average among $\Gamma_1(t_1),\ldots,\Gamma_N(t_1)$, for each time instance t_1, process that is clearly not efficient for every king of data. Indeed, the initial shape series might not be time-aligned (e.g., due to different growth rates in medical applications and different motion speeds in gait analysis). The dynamic shapes need to be properly aligned before any kind of shape statistics technique is applied.[1] This is exactly the role of the *dynamic time warping* (*DTW*), see Figure 1. This process is commonly used in speech recognition in order to time-align speech patterns to account for differences in speaking rates across speakers. It has also been used by a number of authors for gait analysis, but limited to the 1D path obtained by the tracking of particular joints. In this work we propose to combine DTW with results on static shape analysis to compute basic statistics on dynamic shapes. The framework here proposed is independent of the particular choice of static shape metric. This work deals with discrete time instances, while the extension to a continuous framework is discussed in the conclusions section.

2 Static Shape Averaging

In order to compute the mean of N static shapes, a distance on the set of the shapes is necessary (for examples of such metrics see [3, 7, 13, 14] and references therein). Once the metric is given, the empirical mean shape can be defined:

*PM performed this work while visiting the University of Minnesota. We thank Renaud Keriven and Olivier Faugeras for suggesting and encouraging this visit. We thank Facundo Mémoli, Alvaro Pardo, Renaud Keriven, Olivier Faugeras, and Paul Thompson for interesting discussions on shape statistics. This work was supported by a grant from the Office of Naval Research ONR-N00014-97-1-0509, the Presidential Early Career Award for Scientists and Engineers (PECASE), and a National Science Foundation CAREER Award.

[1] The topic of time alignment appears also in video (see for example [2] and references in there). The goals and techniques used there are completely different from the ones here presented. The use of our proposed framework for video alignment is the subject of future research.

Definition 1 *Let $d(\cdot,\cdot)$ be a distance on the set of shapes and $C_1,\ldots,C_N$, N static shapes. The empirical mean $M(C_1,\ldots,C_N)$ is given by*

$$M(C_1,\ldots,C_N) \triangleq \arg\min_{C} \sum_{i=1}^{N} d(C,C_i)^2$$

The goal of this paper is to present a natural and easy to compute extension of this definition for dynamic shapes. Before doing this, let us briefly recall the second fundamental component of our approach, dynamic time warping.

3 Dynamic Time Warping

Dynamic time warping (DTW) is principally used in speech recognition to time-align speech patterns in order to account for differences in speaking rates across speakers. A distance between two speech patterns can then be computed by this technique in order to be able to compare them (see [9]). DTW can be adapted to deal with other types of signals as done in this paper for shapes.

When the two signals (A,B) to be matched are defined as sampled time functions, $A = a_1,\ldots,a_L; B = b_1,\ldots,b_M$, the basic problem in DTW is to find two *time warping functions* f and g such that

$$\sum_{t=1}^{T} d(a_{f(t)}, b_{g(t)})^2$$

is minimized (here $d(\cdot,\cdot)$ stands for the function measuring the discrepancy between two samples).

Computing these warping functions can be viewed as the process of finding a minimum-cost path through the lattice of points $(a_i,b_j)_{(i,j)\in\{1,\ldots,L\}\times\{1,\ldots,M\}}$, starting from $(1,1)$ and ending at (L,M) (see Figure 2),[2] where the cost of a path is defined by:

$$D(f,g) \triangleq \sum_{t=1}^{T} d(a_{f(t)}, b_{g(t)})^2$$

and f and g are subject to the following constraints:

1. f and g must be monotonic:

$$f(k) \geq f(k-1) \text{ and } g(k) \geq g(k-1)$$

2. f and g must match the endpoints of A and B:

$$f(1) = g(1) = 1, \quad f(T) = L \text{ and } g(T) = M$$

3. f and g must not skip any points:

$$f(k) - f(k-1) \leq 1 \text{ and } g(k) - g(k-1) \leq 1$$

4. A limit in the maximum amount of warp is fixed by

$$|f(k) - g(k)| \leq Q, \quad Q \text{ being the given "window width"}$$

In the example in Figure 2, the *time warping functions* are:

$$\begin{aligned} f: 1\to 1,\ 2\to 2,\ 3\to 3,\ 4\to 4 \\ 5\to 5,\ 6\to 6,\ 7\to 6,\ 8\to 6 \\ g: 1\to 1,\ 2\to 2,\ 3\to 2,\ 4\to 2 \\ 5\to 3,\ 6\to 4,\ 7\to 5,\ 8\to 6 \end{aligned}$$

At first glance, it would seem as if $D(f,g)$ would have to be evaluated for a prohibitively large number of possible paths. Fortunately, *dynamic programming* brings this problem under control by noting that the best path from $(1,1)$ to any given point is independent of what happens beyond that point. Hence, if we call $D(i_k,j_k)$ the total cost of the best path from $(1,1)$ to (i_k,j_k), this is the cost of the point (i_k,j_k) itself plus the cost of the cheapest path to it:

$$D(i_k,j_k) = d(i_k,j_k)^2 + \min_{\text{legal}(i_{k-1},j_{k-1})} D(i_{k-1},j_{k-1})$$

By the subscript "legal (i_{k-1},j_{k-1})" we mean the minimum over all permissible predecessors of (i_k,j_k). By constraints 1 and 3 above, there are only three legal predecessors: (i_k-1,j_k), (i_k,j_k-1) and (i_k-1,j_k-1). Therefore we need to consider only three possibilities per lattice point (this if further constrained by point 4 above).

Dynamic programming for solving the DTW problem (finding f and g) then proceeds in incremental stages (see [9] for the complete algorithm), achieving an optimal time complexity of $\mathcal{O}(P\,Q)$ (P is the number of initial frames and Q the "window width" from constraint 4). It means that we need to compute $d(\cdot,\cdot)$, the distance between two static shapes, only $\mathcal{O}(P\,Q)$ times.

4 Dynamic Shapes Averaging

With the basic concepts on the mean of static shapes and dynamic time warping, we are now ready to describe the framework for dynamic shape average.

4.1 Basic Idea

We first define a dynamic shape as a sequence of static shapes (represented by any possible characterization):

Definition 2 *Let S be a set of static shapes (using any existing representation). A dynamic shape Γ is an ordered sequence of static shapes $(C_1,\ldots,C_T) \in S^T$ ($T \in \mathbb{N}$ is the length of the dynamic shape).*

Although the above definition is given for discrete times, it can be extended to continuous space.

The idea now is to combine dynamic time warping and static shape averaging:

[2] Note that we use a_i and b_i both to denote the time positions and their corresponding values, the distinction clearly provided by the context.

Definition 3 *Given $d(\cdot,\cdot)$, a distance on the set of static shapes, and $\Gamma_1(t_1),\ldots,\Gamma_N(t_N)$, N dynamic shapes of respective length T_i (i.e. $t_i = 1,\ldots,T_i$), their empirical mean is defined as*

$$\text{for } 1 \le t \le T : \widehat{\Gamma}(t) \triangleq M(\Gamma_1(f_1(t)),\ldots,\Gamma_N(f_N(t)))$$

where $f_1,\ldots,f_N$ are N time-warping functions given by

$$(f_1,\ldots,f_N) = \arg\min_{f_1,\ldots,f_N} \sum_{t=0}^{T} \mu(\Gamma_1(f_1(t)),\ldots,\Gamma_N(f_N(t)))$$

with

$$\mu(C_1,\ldots,C_N) = \sum_{1\le i<j\le N} d(C_i,C_j),$$

and $M(C_1,\ldots,C_N)$ is the mean of static shapes (see Def.1).

In words, we start by finding (via DTW) optimal time-correspondences between static shapes and after that we compute the average of these static shapes per time instance. The warping is such that the metric is minimized.

This definition suggests to consider the "distance" between two dynamic shapes as follows (there is no triangle inequality here):

Definition 4 *Given two dynamic shapes Γ_1 and Γ_2, their "distance" is given by*

$$\delta(\Gamma_1,\Gamma_2) = \frac{1}{T}\sum_{t=0}^{T} d(\Gamma_1(f_1(t)),\Gamma_2(f_2(t))),$$

where $d(\cdot,\cdot)$ is the selected metric for static shapes and f_1 and f_2 are the optimal time warping functions.

This definition will be used later to compare human motions.

4.2 Basic Improvements

With the simple use of DTW, the mean shape's length T will be greater than or equal to the maximum of the individual time lengths $\{T_1,\ldots,T_N\}$. Therefore, the dynamic mean shape will always be longer than the initial shapes (in Figure 2, $T_1 = 6, T_2 = 6, T = 8$). In order to correct this, we add jumps in the final path.

When $N = 2$, define, for $i \in \{1,2\}$,

$$E_i \triangleq \{(\widehat{\Gamma}(t),\widehat{\Gamma}(t+1)) \mid f_i(t) = f_i(t+1)\}$$

where $\widehat{\Gamma}(\cdot)$ is the dynamic mean shape (from Definition 3). E_1 is the set of vertical segments and E_2 the set of horizontal segments in the graph representing the final path. In Figure 2, $E_1 = \{(\widehat{\Gamma}(6),\widehat{\Gamma}(7)),(\widehat{\Gamma}(7),\widehat{\Gamma}(8))\}$ and $E_2 = \{(\widehat{\Gamma}(2),\widehat{\Gamma}(3)),(\widehat{\Gamma}(3),\widehat{\Gamma}(4))\}$. These segments are responsible for the increase of the final length. Indeed, we have the simple relation (T is the length of the mean shape *without* jumps) :

$$T = T_1 + |E_1| = T_2 + |E_2|$$

We opt to replace every second pair in E_1 by its static average , then we do the same for the pairs in E_2 (see Figure 3). Each replacement decreases the length by one.

Therefore T', the length of the mean shape *with* the jumps, becomes:

$$T' = T - \frac{|E_1|}{2} - \frac{|E_2|}{2} = T_1 + \frac{|E_1| - |E_2|}{2}$$

$$T' = T_1 + \frac{T_2 - T_1}{2} = \frac{T_2 + T_1}{2}$$

The length of the final mean shape is then the average of the length of the two initial shapes.

In the general case (N dynamic shapes), it is also intuitive that we would like the length of the final mean shape to equal the average of the lengths of the N initial dynamic shapes. Therefore we now generalize the pairing process described above. Define, for any $A \subset \{1,\ldots,N\}$:

$$E_A \triangleq \{(\widehat{\Gamma}(t),\widehat{\Gamma}(t+1)) \mid f_i(t) = f_i(t+1) \Longleftrightarrow i \in A\}$$

and, for any $i \in \{1,\ldots,N\}$:

$$\mathcal{A}_i \triangleq \{A \subset \{1,\ldots,N\} \mid i \in A\}$$

Then the following relations, where T is the length of the mean shape *without* jumps, hold for any $i \in \{1,\ldots,N\}$:

$$T = T_i + \sum_{A\in\mathcal{A}_i} |E_A|$$

and

$$T - \frac{1}{N}\sum_{i=1}^{N} T_i = \frac{1}{N} \sum_{A\subset\{1,\ldots,N\}} |A|\cdot|E_A|$$

The right hand term of the previous equality can be eliminated by the following process: For every subset A of $\{1,\ldots,N\}$, we choose a number $\frac{|A|\cdot|E_A|}{N}$ of pairs belonging to E_A.[3] Then we replace each pair by their static average.The length of the final mean shape is then the average of the length of the N initial shapes. Figure 4 shows the mean shape for simple initial shapes and for $N = 3$.

5 Examples

For our experiments,we represented a static shape C by its distance function $\psi(x) = \min_{y\in C}\|x - y\|$ and we used the following simple metric on the set of static shapes:

$$d(C_1,C_2) = \sqrt{\int_\Omega (\psi_1(\omega) - \psi_2(\omega))^2\, d\omega}$$

where $\psi_i(x)$ is the distance function to the shape C_i. For this distance, $\widehat{C}$, the average shape, is the zero level set of

$$\widehat{\psi}(x) = \frac{1}{2}(\psi_1(x) + \psi_2(x))$$

[3] We choose these pairs uniformly spread in time.

As mentioned in the introduction, the framework here introduced is independent of the particular choice of the static metric $d(\cdot,\cdot)$, and we have selected this simple one for demonstration purposes only.

The segmentation of the input pictures is done by simple thresholding ([12, 10]), and the distance function ψ_i for each shape C_i is computed with the fast marching method (for details, see [5, 11, 17]). Figure 5 shows an example (one frame) of an initial dynamic shape.

In figures 6 and 8, we present a number of frames from two initial video clips (dynamic shapes), followed by sampled frames from the average dynamic shape computed without using DTW, and finally sampled frames from the average dynamic shape computed with our technique. Figures 7 and 9 show the corresponding DTW graphs.

In Figure 10, we present some frames from three initial video clips (three walking men), followed by sampled frames from the mean dynamic shape computed without using DTW, and finally with our technique.

Using Definition 4, we can compare different dynamics, such as running vs. walking men. As observed in the table below, this function is five times greater between one running men and one walking men than between two running or two walking men.

$\delta(\cdot,\cdot)/10^6$	walk 1	walk 2	run 1	run 2
walk 1	0	1.3	5.6	6.3
walk 2	1.3	0	5.2	6.7
run 1	5.6	5.2	0	1.1
run 2	6.3	6.7	1.1	0

6 Conclusion

A novel framework for performing shape statistics in dynamic shapes was described in this paper. The basic idea is to combine shape alignment with previously developed ideas from static shape studies. The shape alignment is based on dynamic time warping. The framework is independent of the metric between static shapes.

A number of directions are suggested by the line of research here initiated. First of all, other more advanced static shape metrics need to be used, including those that incorporate landmarks, found to be fundamental for medical applications [15]. Once these advanced metrics are incorporated into our framework, we can proceed with more exhaustive experimentation, including 3D dynamic shapes. Of particular interest are the analysis and recognition of gait and the study of growth in medical applications.

In this paper we limited ourselves to the case of discrete time. In the continuous case, a variational formulation to address the dynamic shape average problem can be formulated as

$$\arg\min_{\Gamma, f_i} \int_0^T \sum_{1\le i\le N} [d(\Gamma(t), \Gamma_i(f_i(t))^2 \ + H(f_i(t)]dt$$

where f_i are the *time warping functions*, and H represents some constraints on them (such as continuity, monotonicity, acceleration, etc). To this we can add time domain landmarks (e.g., by splitting the domain). These topics are the subject of current efforts in our group.

References

[1] F.L. Bookstein. Size and shape spaces for landmark data in two dimensions. *Statistical Science*, 1:181–242, 1986.

[2] Y. Caspi and M. Irani. Parametric sequence-to-sequence alignment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[3] G. Charpiat, O. Faugeras, and R. Keriven. Approximations of shape metrics and application to shape warping and empirical shape statistics. Technical report, INRIA–RR-4820, 2003.

[4] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models: Their training and application. *Computer Vision and Image Understanding*, 61(1):38–95, 1995.

[5] J. Helmsen, E. G. Puckett, P. Collela, , and M. Dorr. Two new methods for simulating photolithography development in 3d. *Proc. SPIE Microlithography*, IX:253, 1996.

[6] I. L. IDryden and K. V. Mardia. *Statistical Shape Analysis*. John Wiley & Sons, New York, 1998.

[7] D. G. Kendall. Shape manifolds, procrustean metrics and complex projective spaces. *Bulletin of the London Mathematical Society*, 16:81–121, 1984.

[8] M. Miller and L. Younes. Groups actions, homeomorphisms and matching: A general framework. *International Journal of Computer Vision*, 41(1/2):61–84, 2001.

[9] T. Parsons. *Voice and Speech Processing*. McGraw-Hill, 1987.

[10] P. K. Sahoo, S. Soltani, A. K. C. Wong, and Y. C. Chen. A survey of thresholding techniques. *Computer Vision, Graphics, and Image Processing*, 41(2):233–260, February 1988.

[11] J. A. Sethian. A fast marching level-set method for monotonically advancing fronts. *Proc. Nat. Acad. Sci.*, 93:4:1591–1595, 1996.

[12] L. G. Shapiro and G. C. Stockman. *Computer Vision*. Prentice Hall, 2001.

[13] C. G. Small. *The Statistical Theory of Shape*. Springer, 1996.

[14] S. Soatto and A. J. Yezzi. Deformotion: Deforming motion, shape average and the joint registration and approximation of structures in images. *International Journal of Computer Vision*, 53(2):153–167, 2003.

[15] A. Toga. *Brain Warping*. Academic Press, 1998.

[16] A. Trouvé and L. Younes. Diffeomorphic matching problems in one dimension: Designing and minimizing matching functionals. In *ECCV'00*, pages 573–587, 2000.

[17] J. N. Tsitsiklis. Efficient algorithms for globally optimal trajectories. *IEEE Transactions on Automatic Control*, 40:1528–1538, 1995.

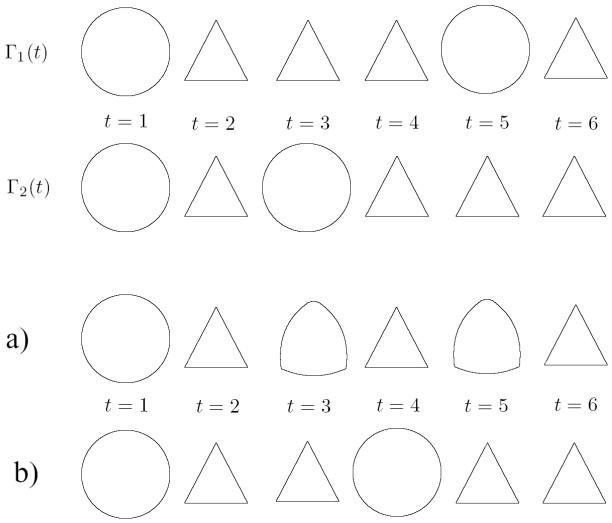

Figure 1: A simple example showing the importance of time alignment when performing shape statistics. The first two rows show two dynamic shapes $\Gamma_1(t)$ and $\Gamma_2(t)$. The following two rows show their mean: a) Computed without DTW-alignment, b) Computed with enhanced DTW-alignment. We clearly observe the need for the DTW step.

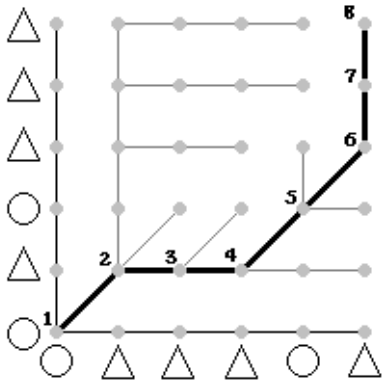

Figure 2: Dynamic time warping example.

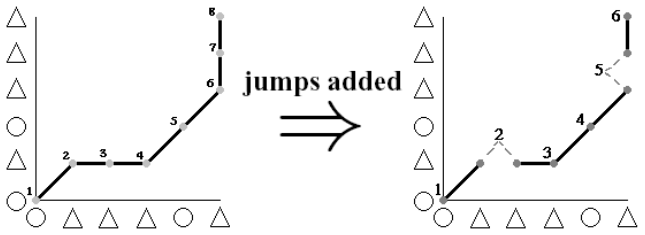

Figure 3: Example of the introduction of jumps in DTW for $N = 2$.

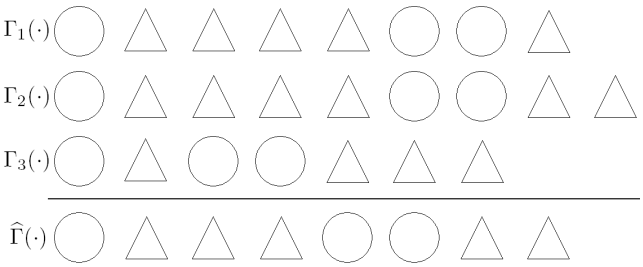

Figure 4: Example of mean shape, with jumps, for simple initial shapes and $N = 3$.

Figure 5: From left to right: Initial image, segmented shape, and distance function

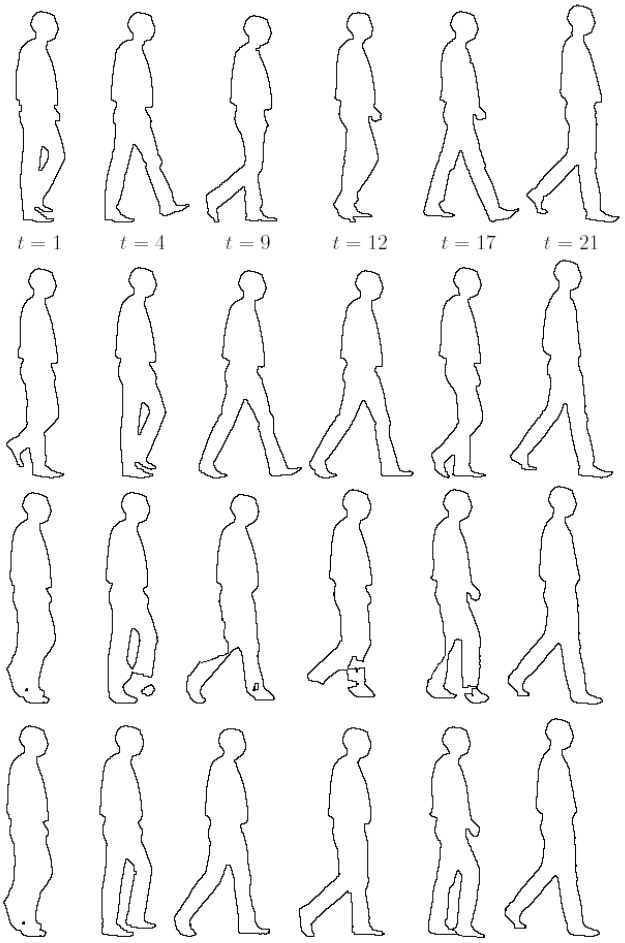

Figure 6: Example of two walking men. The two dynamic shapes are given first, followed by the mean without DTW (third row), and finally the mean with DTW (last row). Note how the lack of time alignment creates topological errors, not present in the average when DTW is used.

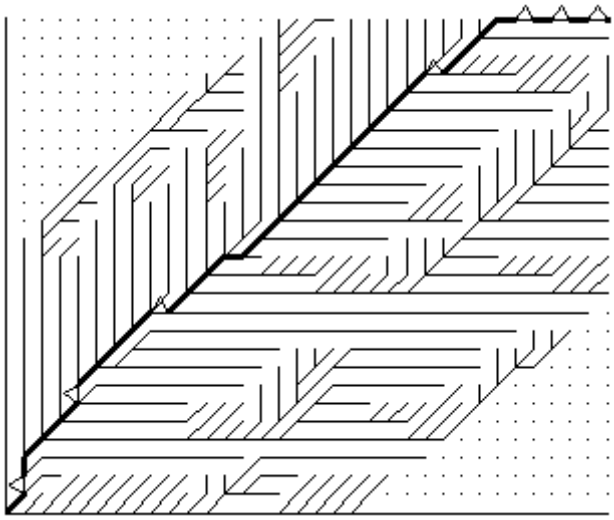

Figure 7: Graph corresponding to the DTW for the running sequences.

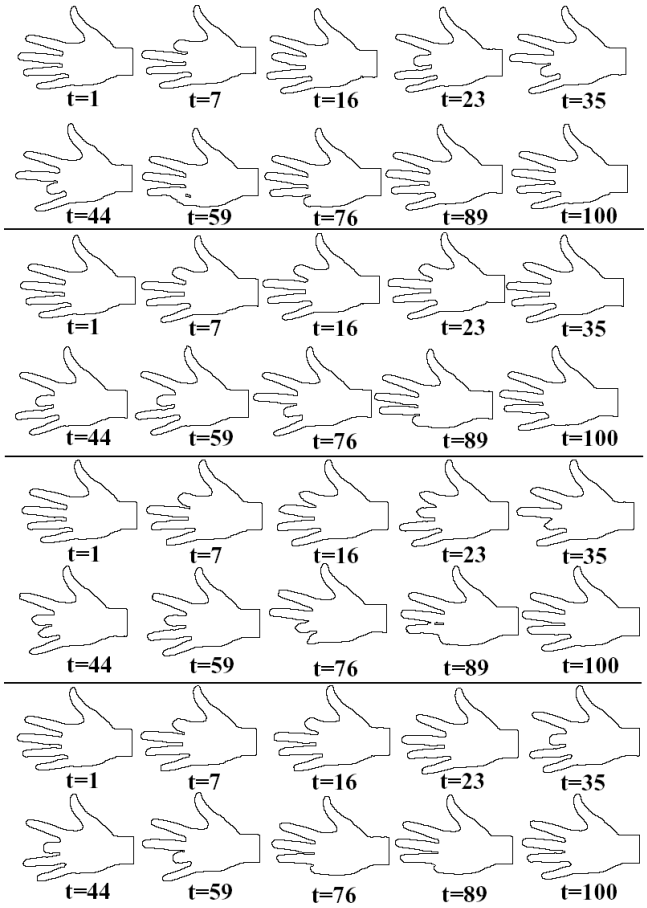

Figure 8: Same as Figure 6 for two hands in motion.

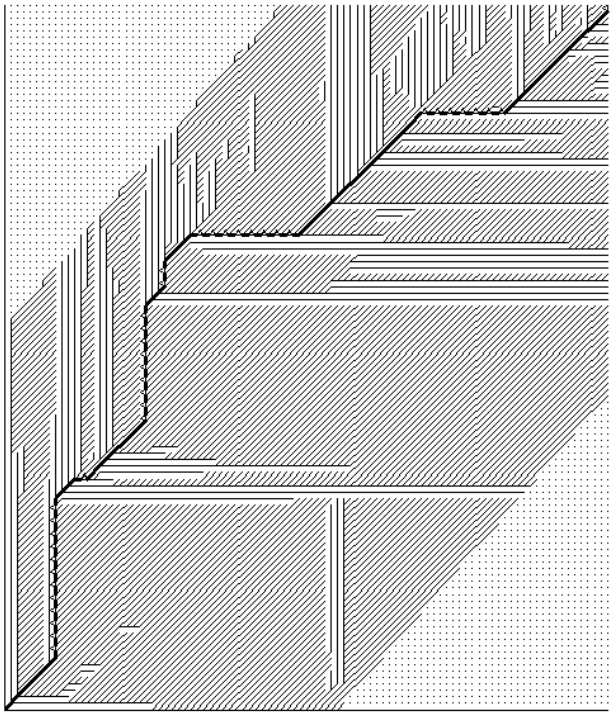

Figure 9: Graph corresponding to the DTW for the hands sequences.

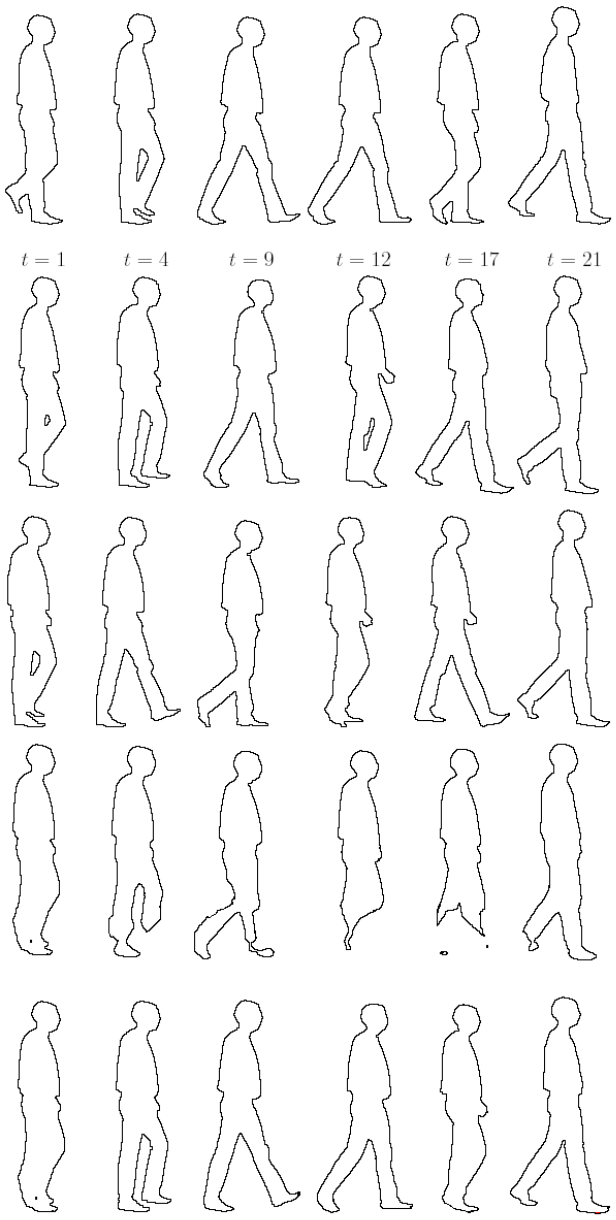

Figure 10: Example of three walking men. The three dynamic shapes are given first, followed by the mean without DTW (fourth row), and finally the mean with DTW, our proposed technique (last row). Once again, note the significant improvement when the time-warping is added to the shape statistics process.