Spatio-Temporal Registration of Multiple Trajectories

Nicolas Padoy and Gregory D. Hager

The Johns Hopkins University, Baltimore, MD, USA
{padoy,hager}@jhu.edu

Abstract. A growing number of medical datasets now contain both a spatial and a temporal dimension. Trajectories, from tools or body features, are thus becoming increasingly important for their analysis. In this paper, we are interested in recovering the spatial and temporal differences between trajectories coming from different datasets. In particular, we address the case of surgical gestures, where trajectories contain both spatial transformations and speed differences in the execution. We first define the spatio-temporal registration problem between multiple trajectories. We then propose an optimization method to jointly recover both the rigid spatial motions and the non-linear time warpings. The optimization generates also a generic trajectory template, in which spatial and temporal differences have been factored out. This approach can be potentially used to register and compare gestures side-by-side for training sessions, to build gesture trajectory models for automation by a robot, or to register the trajectories of natural or artificial markers which follow similar motions. We demonstrate its usefulness with synthetic and real experiments. In particular, we register and analyze complex surgical gestures performed by tele-manipulation using the da Vinci robot.

1 Introduction

Time series data is gaining importance in medicine and poses new challenges. Whereas historically the spatial aspect of data was analyzed, for instance to build atlases, we now need to consider both the spatial and the temporal aspects together. Trajectories are natural spatio-temporal data descriptors. Our interest in this paper focuses specifically on tool trajectories that allow to model and analyze the performance of surgeons. Models of surgical performance are used to develop training systems [1] and evaluation methods [2]. With the rapid advancement of minimally invasive surgery, there is a growing need for such systems that can improve the learning curve of trainees. Surgical models can also be used to segment gestures [3] and to recognize the surgical workflow [4], in order to develop context aware assistance systems. More recently, expert gesture information has been used in robotics to make surgical robots learn how to perform complex movements like knot tying [5].

Gesture datasets generally consist of tool trajectory data recorded from different users, possibly operating in different environments. As a result, the data contains variations. For example, rigid spatial transformations arise from different patient and system locations. Different operator styles and environment constraints cause not only rigid and deformable spatial deformations, but also non-linear speed differences in the executions. A common way to recover spatial differences is to use markers. However, such a setup may be cumbersome in some scenarios or non-available in some datasets.

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 145–152, 2011.

Analyzing multiple trajectories representing the same gesture requires therefore the recovery of both the rigid spatial transformations and the non linear time warpings. This way, multiple instances can be both spatially and temporally superimposed, allowing the comparison of variations solely caused by style, experience and environment specificities. Note that this recovery also provides the temporal synchronization between videos observing the scenes. Such information is very valuable for surgeons, as it provides an intelligent way to mine surgical databases, e.g. to extract and compare side by side surgical performances during training seminars.

In this paper, we introduce two cost functions, one stating the spatio-temporal registration problem for a pair of trajectories, and the second stating the spatio-temporal registration problem for multiple trajectories. We then propose optimization approaches that combine dynamic time warping (DTW) [6] with rigid spatial registration [7]. In the case of multiple trajectories, the registration approach assumes that they are all transformations of a generic gesture template to be computed. It iteratively computes the rigid transformation and temporal warping between each trajectory and the template, which is updated. The recovery of the spatial transformations permits the computation of an accurate template that keeps the same shape as all trajectories. Also, rotation invariant features like curvature are then not necessary to drive the temporal synchronization.

Usually, a surgical gesture spatially overlaps over time within a small 3-dimensional volume. Therefore, methods like iterative closest point [8], which uses only spatial information, or dynamic time warping [6], which does not model spatial transformations, are not accurate for registration. We show that the spatio-temporal registration provides better results. Our approach share similarities with canonical time warping (CTW) [9], which uses DTW and canonical correlation analysis to jointly synchronize and extract common features from a pair of time-series having possibly different dimensions. Two main differences are: 1) we address the joint registration of multiple time-series and 2) due to the nature of our problem where time-series represent trajectories, we estimate a rigid transformation between two time-series instead of projection matrices for each of them. A synthetic experiment in section 3 will show how the choice of the appropriate modeling for the transformation affects the registration.

The remainder of this paper is as follows: we present our approach in section 2 and show synthetic and real experiments in section 3. Conclusions are given in section 4.

2 Methods

Sections 2.1 and 2.2 present two standard approaches to register two trajectories either temporally or spatially. In section 2.3, we state the spatio-temporal registration problem for two time-series and explain how to combine the two previous approaches to address it. Section 2.4 extends the registration approach to multiple trajectories.

2.1 Temporal Registration

Let X and Y be two time-series of length n_x and n_y representing two 3-dimensional trajectory instances of the same gesture: $X \in \mathbb{R}^{3,n_x}$ and $X_t \in \mathbb{R}^3$ is a trajectory point for $1 \leq t \leq n_x$. In this section, we assume that one time-series can be derived from the

other by a time-warping $\phi = (\phi^x, \phi^y, m)$, modulo additional small spatial variations that are for instance due to style or environment specificities. The warping ϕ models a temporal alignment between the time-series X and Y, using a common timeline of length m: X_i is aligned with Y_j if and only if there exists $1 \leq t \leq m$ such that $\phi^x_t = i$ and $\phi^y = j$. The warping verifies the constraints: $0 \leq \phi^a_t - \phi^a_{t-1} \leq 1$ (*continuity*), $\phi^a_1 = 1$, $\phi^a_m = n_a$ (*boundaries*), for $a \in \{x, y\}$. DTW [6] is an algorithm based on dynamic programming that optimally computes a warping ϕ minimizing the cost:

$$C_{warping}(\phi) = \sum_{t=1}^{m} \|X_{\phi^x_t} - Y_{\phi^y_t}\|^2. \tag{1}$$

An example of alignment is shown in fig. 1(d). If the trajectories are spatially very different, this approach fails. An alternative is to express the total cost in terms of features invariant to certain spatial transformations. Instead, we propose to jointly integrate the spatial registration in the process, as will be needed for template generation.

2.2 Rigid Spatial Registration

Let $\{U_i\}_{1 \leq i \leq n_u}$ and $\{V_j\}_{1 \leq j \leq n_v}$ be two 3D point sets describing a similar shape. Assuming that p point correspondences $\{(U_{\psi^u(l)}, V_{\psi^v(l)}) \mid 1 \leq l \leq p\}$ are known, where $\psi : \{1, \ldots, p\} \longrightarrow \{1, \ldots, n_u\} \times \{1, \ldots, n_v\}$, [7] optimally recovers the rigid spatial transformation $[R, T] \in SO^3 \times \mathbb{R}^3$ between the two shapes using the singular value decomposition (SVD) to minimize the sum of squared errors:

$$C_{rigid}(R, T) = \sum_{l=1}^{p} \|RU_{\psi^u(l)} + T - V_{\psi^v(l)}\|^2. \tag{2}$$

Note that the scaling factor is omitted, as the gestures we analyze interact with an environment of constant size and we do not want an additional scaling transformation.

In the case where point correspondences are not known, the iterative closest point (ICP) algorithm has been proposed [8] to iteratively estimate correspondences and the spatial registration. This approach can fail however if the initial shapes are far from the solution and the point clouds do not describe neatly the shape, as it occurs for a trajectory that overlaps in a small spatial volume. Trajectories are not simple point clouds but have a temporal dimension. In the next section, we will use this additional information to improve point assignments and thereby the registration.

2.3 Spatio-Temporal Time Warping (STW)

Let X and Y be two time-series representing two 3-dimensional trajectory instances of the same gesture, as in section 2.1. We make here the less restrictive assumption that one time-series can be derived from the other by a rigid spatial transformation $[R, T]$ and a time-warping $\phi = (\phi^x, \phi^y, m)$, modulo additional small variations. Our objective is to recover jointly the spatial transformation and the time-warping, in order to isolate and compare these variations. We propose to minimize the following registration cost

$$C_{stw}(\theta) = \sum_{t=1}^{m} \|RX_{\phi^x_t} + T - Y_{\phi^y_t}\|^2, \tag{3}$$

Algorithm 1. Spatio-temporal time warping (STW)

input : X,Y
output: $\theta = \{R, T, \phi\}$
initialization: $R = \mathbf{Id}$, $T = \mathbf{0}$;
repeat
  Compute temporal warping ϕ between $(RX + T)$ and Y using DTW;
  Compute $[R, T]$ using rigid registration and correspondence set $\{(X_{\phi_t^x}, Y_{\phi_t^y}) \mid 1 \leq t \leq m\}$ between trajectories X and Y;
until C_{stw} *converges* ;

with respect to the parameters $\theta = \{R, T, \phi\}$. We use a coordinate descent approach similar to [9], in which each step monotonically decreases the cost. We iteratively compute $[R, T]$ and ϕ using rigid registration and DTW. Even though convergence is only guaranteed to a local minimum, experiments show that the good correspondences obtained by temporal registration permit to recover large spatial transformations, even with trajectories containing variations. The approach is given in Algorithm 1.

2.4 Spatio-Temporal Registration of Multiple Trajectories

Let us now assume $X^1, \ldots, X^K$ to be K time-series of length $n_1, \ldots, n_K$, representing 3-dimensional trajectory instances of the same gesture. We make the assumption that all instances can be derived from a gesture template Y of length n_y by rigid spatial transformations $[R^k, T^k]$ and time-warpings $\phi^k = (\phi^{k,x}, \phi^{k,y}, m_k)$, modulo additional small variations due to style or environment specificities (see e.g. fig. 3(a)). Our objective is to recover all spatial transformations and time-warpings, in order to isolate and compare jointly these variations. We therefore introduce the following registration cost

$$C_{multi}(\omega) = \sum_{k=1}^{K} \sum_{t=1}^{m_k} \| R^k X^k_{\phi_t^{k,x}} + T^k - Y_{\phi_t^{k,y}} \|^2, \tag{4}$$

to be optimized on the set of parameters $\omega = \{R_k, T_k, \phi^k, Y \mid 1 \leq k \leq K\}$. During the optimization, we iteratively compute the warpings $\{\phi^k\}$, the template Y and the transformations $\{[R^k, T^k]\}$. If we suppose Y to be known, for instance to be one of the X^k, similarly to Algorithm 1 we can compute all $\{\phi^k\}$ and then all $\{[R^k, T^k]\}$ alternatively, to decrease monotonously $C_{multi}(\theta)$. The template needs however to be refined, as a particular sequence may contain anomalies or not be representative of the timeline. To compute the template Y, we use a temporal averaging technique presented in [10]. This approach synchronizes several time series to a common timeline whose length is the average length of all time series, using an initial template. Using this synchronization, the time-series are then averaged. This method is presented in Algorithm 2, where the monotonicity of the $\tilde{\phi}^{k,x}$ guaranties the invertibility of μ. Interpolation steps are needed in the computations of μ and Z, but omitted for better readability.

The final registration method is presented in Algorithm 3. Since the template computation also modifies the temporal warpings, monotonous decrease of the cost is not obvious. Convergence is however verified in our experiments.

Algorithm 2. Template generation

input : $\{X^k\}_{1\leq k\leq K}$, initial template $\tilde{Y}$
output: new template Y and new warpings ϕ^k
for $1 \leq k \leq K$ **do**
  Compute warpings $\tilde{\phi}^k = (\tilde{\phi}^{k,x}, \tilde{\phi}^{k,y}, \tilde{m})$ between X^k and $\tilde{Y}$ using DTW;
Compute $\mu(t) = \sum_{k=1}^{K} \tilde{\phi}_t^{k,x}$, where $1 \leq t \leq \tilde{m}$;
Define $Y_t = \sum_{k=1}^{K} \tilde{Y}_{(\tilde{\phi}^{k,x}(\mu^{-1}(t)))}$, where $1 \leq t \leq \frac{1}{K}\sum_{k=1}^{K} n_k$;
for $1 \leq k \leq K$ **do**
  Compute warpings $\phi^k = (\phi^{k,x}, \phi^{k,y}, m)$ between X^k and Y using DTW;

Algorithm 3. Multiple trajectories registration

input : $\{X^k\}_{1\leq k\leq K}$
output: $\omega = \{R_k, T_k, \phi^k, Y \mid 1 \leq k \leq K\}$
Y is initialized as the X^k with median length;
for $1 \leq k \leq K$ **do**
  $R^k = \mathbf{Id}, T^k = \mathbf{0}$
repeat
  Update Y and $\{\phi^k\}$ using Algo. 2 with input $\{R^k X^k + T^k\}_{1\leq k\leq K}$ and Y;
  for $1 \leq k \leq K$ **do**
    compute R^k, T^k using rigid registration and correspondence set $\{(X^k_{\phi_t^{k,x}}, Y_{\phi_t^{k,y}}) \mid 1 \leq t \leq m_k\}$;
until C_{multi} *converges* ;

3 Experiments and Results

3.1 Synthetic Experiments

Pair of trajectories
Let X and Y be trajectories describing 80% of a circle (fig. 1(a)). Y is obtained by rotating X in the same plane by 60 degrees and translating it. A noise of 5% is added to the data points. The trajectories are performed at different speeds, as illustrated by the ground truth temporal warping in fig. 1(d). Y mostly describes the circle twice as fast, but in the end at the same speed ($n_x = 150$ and $n_y = 88$). Fig. 1(b) shows the results for ICP, which fails due to the noise and spatial overlap. DTW alone also fails (fig. 1(e,f)), as it assumes a time warping at the extremities and tends to associate temporally the points closest spatially. CTW[9] gives better results (fig. 1(g,h)), but also fails to recover the temporal alignment since it does not model the appropriate transformation. The affine deformation of the data is clearly visible in fig. 1(g). Finally, the STW approach recovers the correct spatio-temporal warping (fig. 1(i,j)).
Multiple trajectories
Figure 2 illustrates the registration of multiple trajectories using the approach of section 2.4. Three noisy partial circles with different time-warpings (lengths 88,150,250) are correctly registered, as shown in fig. 2(b). Note that the shape of the template is also a circle, which would not be the case if the rotations were not properly recovered.

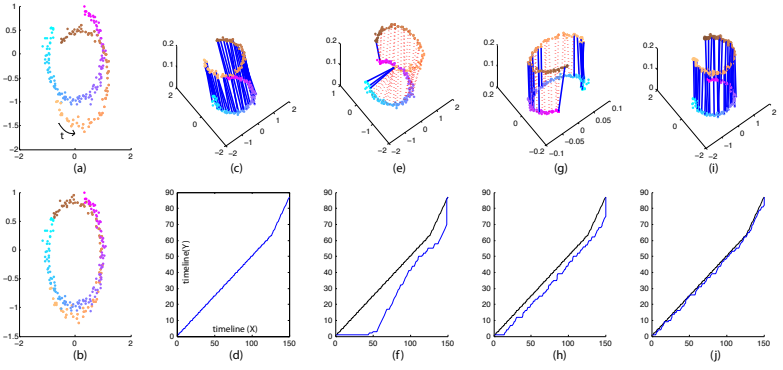

Fig. 1. Synthetic experiment. (a) two trajectories; (b) spatial registration using ICP; (c) Ground truth spatio-temporal registration, with trajectories in two different planes for better visualization. Lines indicate the spatial alignment; (d) corresponding time warping; (e) registration using DTW only; (g) registration using CTW; (i) registration using STW; (f),(h),(j) show the time warpings corresponding to (e),(g),(i). The dotted red lines in (e),(g),(i) indicate wrong alignment, with tolerance ± 5 frames.

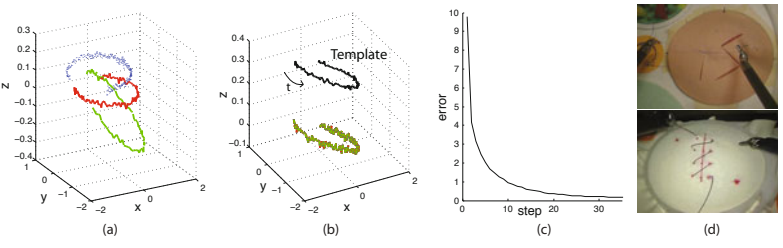

Fig. 2. Joint registration of 3 synthetic trajectories / setups for real experiments. (a) raw trajectories; (b) registration with generated template isolated on top for better visualization; (c) registration error; (d) setups for knot-tying (top) and 4-throw suture (bottom).

3.2 Experiments on Tele-Manipulation Data

We are interested in temporally synchronizing and spatially superimposing trajectories, to replay synchronously the videos or the registered gestures in a virtual environment.

Knot-tying analysis

We use six sequences of a knot tying task performed by the same user using a da Vinci robot (fig. 2(d)). Trajectories describe the left tool positions. In each recording, the task pod has a different position and orientation. In spite of the large spatial transformations and the multiple local variations (fig. 3(a)), e.g. in directions when pulling the thread, all trajectories can be jointly registered. Fig. 3(b) shows the qualitative result and the resulting template. Fig. 3(c) shows the decrease of the registration error C_{multi} during the optimization. The temporal synchronization between two trajectories is shown in

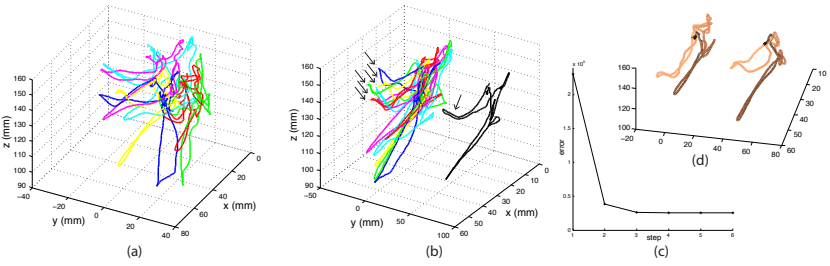

Fig. 3. Knot tying. (a) 6 raw trajectories; (b) registration with template isolated on the right for better visualization; arrows show a registered feature; (c) registration error; (d) temporal synchronization between two trajectories pictured by a color gradient.

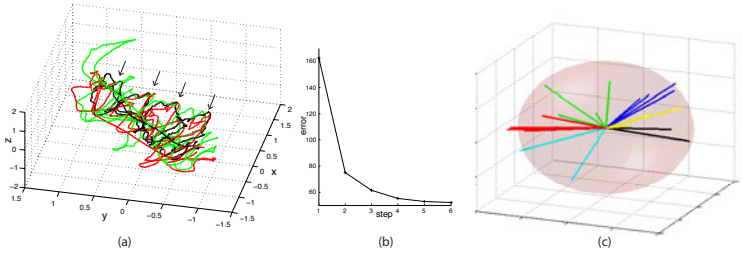

Fig. 4. Four-throw suture. (a) 3 of 19 registered trajectories with arrows showing the registered throws; (b) registration error; (c) rotation axis, colored per surgeon.

fig. 3(d) using a color gradient to represent time. By registering the sequences all together, the performances of surgeons can be synchronously compared (e.g. beginners vs experts). Side by side virtual replay can also teach a trainee the correct tool orientation.

Four-throw suture analysis

We use 19 sequences of a 4-throw suture performed by 6 different surgeons using the da Vinci tele-surgical robot (see fig. 2(d)). The trajectory data describes the cartesian positions of the master-manipulator corresponding to the right hand and directly controlling the right instrument. Fig. 4(a) shows a subset of the registered sequences for qualitative evaluation. Fig. 4(b) shows the quantitative registration error C_{multi}. This dataset illustrates another interesting application of the registration. For each surgeon, the same suturing pod is used, but different surgeons use different pods. Even though the data was meant to be recorded with the same setup, the videos show that between surgeons, the pods are rarely at the same location, but rotated. Fig. 4(c) shows the computed 3D axis of the axis/angle representations of the rotations R^k, colored per surgeon. We see that the registration is precise enough to cluster the 3D axis per surgeons (and setups), in spite of the fact that the long data sequences contain multiple atomic gestures and variations.

4 Discussion and Conclusion

Trajectories occur naturally in computer assisted surgery: they describe the movements of tools, markers and body features. We first propose an approach to jointly register spatially and temporally two trajectories differing by a rigid spatial transformation and a temporal warping, and containing possibly additional local variations. Such an assumption is valid for most complex gestures, since both the task to accomplish and the environment constrain the operator's freedom of movement. We then state the registration problem for multiple trajectories and propose an optimization approach that simultaneously generates a representative template. Synthetic experiments show the usefulness of combining spatial and temporal registration. Real experiments demonstrate its applicability for registering complex gestures, like knot tying and multi-throw suturing, for example for intelligent side-by-side comparison. Even though we register multiple trajectories of a single object (a tool), by stacking up the data, one can extend the approach to register multiple trajectories of multiple objects. We are interested in applying this idea to the registration of medical images acquired over time.

Acknowledgments. This work is funded by NSF grant CDI-0941362. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

1. Blum, T., Sielhorst, T., Navab, N.: Advanced Augmented Reality Feedback for Teaching 3D Tool Manipulation, ch. 25, pp. 223–236. Lupiensis Biomedical Publications (2007)
2. Lin, H.C., Shafran, I., Murphy, T.E., Okamura, A.M., Yuh, D.D., Hager, G.D.: Automatic detection and segmentation of robot-assisted surgical motions. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 802–810. Springer, Heidelberg (2005)
3. Varadarajan, B., Reiley, C., Lin, H., Khudanpur, S., Hager, G.: Data-derived models for segmentation with application to surgical assessment and training. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 426–434. Springer, Heidelberg (2009)
4. Padoy, N., Blum, T., Ahmadi, A., Feussner, H., Berger, M.-O., Navab, N.: Statistical modeling and recognition of surgical workflow. Medical Image Analysis (2010), doi:10.1016/j.media.2010.10.001
5. van den Berg, J., Miller, S., Duckworth, D., Hu, H., Wan, A., Fu, X.-Y., Goldberg, K., Abbeel, P.: Superhuman performance of surgical tasks by robots using iterative learning from human-guided demonstrations. In: ICRA 2010, pp. 2074–2081 (2010)
6. Sakoe, H., Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans. Acoust. Speech Signal Process. 26(1), 43–49 (1978)
7. Umeyama, S.: Least-squares estimation of transformation parameters between two point patterns. IEEE Trans. Pattern Anal. Mach. Intell. 13, 376–380 (1991)
8. Zhang, Z.: Iterative point matching for registration of free-form curves and surfaces. Int. J. Comput. Vision 13, 119–152 (1994)
9. Zhou, F., De la Torre, F.: Canonical time warping for alignment of human behavior. In: Advances in Neural Information Processing Systems Conference, NIPS (2009)
10. Wang, K., Gasser, T.: Alignment of curves by dynamic time warping. Annals of Statistics 25(3), 1251–1276 (1997)