Computational Modeling of Human Social Intention

Mohamed Daoudi[1]
http://pagesperso.telecom-lille.fr/daoudi/

Maxime Devanne[1]
http://maxime-devanne.com/

Francois Quesque[2]
https://sites.google.com/site/francoisquesque/

Yann Coello[2]
http://perso.univ-lille3.fr/~ycoello/

[1] Télécom Lille
   CRIStAL Laboratory (UMR CNRS 9189)

[2] University of Lille 3
   SCALab (UMR CNRS 9193)

Abstract

In this paper, we propose to analyze the kinematic of human arm to predict social intention (personal vs. social intention). We propose to use infrared markers placed on the arms of two human subjects. The trajectories of different markers are defined as shape spaces and subsequently analyzed as a Riemannian manifold for the purposes of characterizing the intention (personal vs. social). The first results show that the best results are obtained with the wrist marker (69.2%).

1 Introduction

In the last decades, understanding social interaction has attracted many researchers from various fields of research including cognitive psychology and neurosciences, medical studies, and computer sciences. In this respect, computer vision researchers have developed powerful algorithms to differentiate and recognize human gestures. Such approach proved to be particularly useful for human-machine interaction systems. Besides, some cognitive sciences expert focused their studies on the communicative aspects of non-verbal human interactions. In particular, recent studies have focused on interaction gestures to demonstrate the effect of social intentions on kinematic variables, and the capacity of observers to use motor deviants for their own motor production. Based on the possibility to benefit from both computer vision and cognitive science expertise, we developed a study with the aim to combine efficient computer vision technique and kinematic analysis in a motor task implying different social intention. Our objective is to propose a statistical shape analysis framework of trajectories in $\mathbb{R}^3$ of human social intention. This framework should be able to classify human social intention.

2 Social Intention in Motor Actions

Understanding what a conspecific is doing (for instance recognising other's action) is crucial for the control of our everyday social interactions. Understanding the reasons that drive the observed behaviour (for instance identifying other's intention) is however much more complicated. Previous literature has shown than when we observe a confederate grasping a bottle of water, we can anticipate whether the bottle is grasped to drink from it or to throw it away, based on variation of arm movement kinematics (e.g. [5]). But, what happens if the spatial constraints of the task are similar (same object, same location, same movement), but only the social intention changes (for example, moving the bottle on a table for a subsequent movement performed by either the actor or someone else). Recent researches in cognitive psychology have suggested that even in that case movement kinematics is affected. In particular, these studies showed that when we perform an action with a social instead of a personal intention, we amplify the spatial and temporal parameters of the motor action [8]. Furthermore, an observer is able to use this kinematic deviants and anticipate social intention in motor actions performed by others, in order to act in a complementary way ([6]; see [7] for a review). A challenge for the future is to determine whether these kinematic deviants can be registered and processed by a classifier in order to allow virtual and artificial systems (avatars, robots ...) to distinguish between different social goals in interactive context, including either humans or other artificial systems.
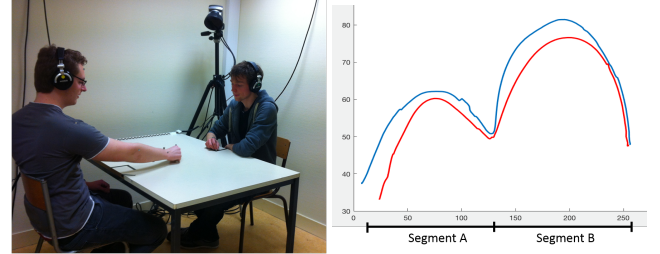

Figure 1: Illustration of the protocol. (Left) A screen-shot of the acquisition session. (Right) Example of two wrist trajectories. The red trajectory corresponds to personal intention and the blue trajectory corresponds to the social intention. Motions can be decomposed in the reaching phase (Segment A) and the putting phase (Segment B).

3 Data Acquisition

Two participants seated at a table, facing each other, and participating in a short interactive game which consisted in displacing a small object to different locations. Their sequential actions were time-locked to a series of broadcasted sounds. The first move of the game was always performed by the same member of the dyad (named here, the actor) and consisted in displacing the object from an initial location to a central location. After this preparatory action, a subsequent action was performed either by the actor (personal condition, half of the trials) or by the partner (social condition, half of the trials), depending on the pitch of the sound triggering the action. 40 trials were performed in a random order. Meanwhile, the actor's movements were recorded with 5 Oqus infrared cameras (Qualisys system). 9 infrared reflective markers were placed on the index (tip), the thumb (tip), the hand, the wrist (scaphoid and pisiform), the right-elbow, the shoulders (right and left), and the head of the actor. The calibration of the cameras provided the means to reach a standard deviation smaller than 0.2 mm, at a 200 Hz sampling rate.

4 Shape Analysis of Trajectories of Human Social Intention

Our hypothesis is that we can characterize the social intention using image data by studying the shapes of the trajectories of the markers. This requires proper mathematical representations of these trajectories and statistical models for studying their variability. In the last few years, many approaches have been developed to analyze shapes of 2D curves. We can cite approaches based on Fourier descriptors, moments or the median axis. More recent works in this area consider a formal definition of shape spaces as a Riemannian manifold on which they can use the classic tools for statistical analysis on tangent spaces. Motivated by the promising results obtained in 3D facial recognition [2] and human action recognition [1], we propose to use the shape analysis framework proposed by [9]. Each motion of markers is represented by a trajectory of markers. Formally, we start by considering a given trajectory as a continuous parameterized function $\beta(t) \in \mathbb{R}^3$, $t \in [0, 1]$. β is first represented by its *Square-Root Velocity Function* (SRVF), q, according to :

$$q(t) = \frac{\dot{\beta}(t)}{\sqrt{\|\dot{\beta}(t)\|}}, t \in [0, 1].$$

Then, with the $\mathbb{L}^2$-norm of the q functions scaled to 1 ($\|q\|=1$), the space of such representation: $\mathcal{C}=\{q:[0,1]\to\mathbb{R}^3,\|q\|=1\}$ becomes a Riemannian manifold and have a spherical structure in the Hilbert space $\mathbb{L}^2([0,1],\mathbb{R}^2)$. Given two curves β_1 and β_2 represented by their SRVFs q_1 and q_2 on the manifold, the geodesic path connecting q_1,q_2 is given analytically by the minor arc of the great circle connecting them on $\mathcal{C}$ (see [9] for further details). The distance between two elements q_1 and q_2 is defined as $d_{\mathcal{C}}(q_1,q_2)=cos^{-1}(\langle q_1,q_2\rangle)$. Such distance represents the similarity between the shape of two curves in $\mathbb{R}^2$. Basically, it quantifies the amount of deformation between two shapes. This distance called also elastic distance, is invariant to rotation, scaling and it takes into account the stretching and the bending of the curves [10].

5   Statistical Shape Analysis of Trajectories

The main goal of our study is to categorize the user intention among two classes c_k denote $\{personal, social\}$. For that we propose to learn representative distributions of trajectories for each class. An important advantage of our Riemannian approach is its ability to compute summary statistics of a set of trajectories. While a mean within a vector space is easy to compute, performing the task on a non-linear manifold such as $\mathcal{C}$ is not obvious.

For example, one can use the notion of Karcher mean [3] to define an average trajectories that can serve as a representative trajectories of a group of trajectories. Given a set of data points q_i on a the manifold $\mathcal{C}$, one way to define their geometric mean is via the minimization of a certain cost function defined by :

$$\mu=\underset{q_i}{\arg\min}\sum_{i=1}^{N}d_{\mathcal{C}}(\mu,q_i)^2 \qquad (1)$$

Since manifolds lack a vector space structure and other Euclidean structures such as norm and inner product, machine learning algorithms including principal component analysis (PCA) and Maximum Likelihood clustering algorithm cannot be applied in their original forms on the manifold $\mathcal{C}$. A common approach used to cope with its non-linearity consists in approximating the manifold valued data with its projection to a tangent space at a particular point on the manifold, for example, the mean of the data μ. Then, each sample shape q_i is mapped in the tangent space at the mean shape $T_\mu\mathcal{S}$ using the inverse exponential map [4] defined as:

$$v_i=exp_\mu^{-1}(q_i)=\frac{\theta}{sin\theta}(q_i-cos(\theta)\mu)\,, \qquad (2)$$

where $\theta=d_{\mathcal{C}}(\mu,q_i)$. The original shape q_i can be retrieved from the velocity vector v_i by using the exponential map operator [4] defined as:

$$q_i=exp_\mu(v_i)=cos(\|v_i\|)\mu+sin(\|v_i\|)\frac{v_i}{\|v_i\|}. \qquad (3)$$

Shapes are projected in the tangent space of the mean using the inverse exponential map (eq. **??**).

$$v_2^*=exp_{q_1}^{-1}(q_2^*)=\frac{\theta}{sin\theta}(q_2^*-cos(\theta)q_1) \qquad (4)$$

Such tangent space is a linear vector space which is more convenient to compute statistics. Hence, in order to learn the distribution of tangent vectors on the tangent space, we can first perform PCA to learn a principal subspace denoted $\mathcal{B}$. Then, the covariance matrix on this principal basis is computed as $\Sigma=\sum_{i=1}^{N}v_iv_i^T$, where v_i are the tangent vectors projected into the principal subspace $\mathcal{B}$.

Finally, the multivariate normal distribution of trajectory c_k, $p(v|c_k;\Sigma)$ is learned using the covariance matrix Σ computed from the set of v_i where $|\Sigma|$ is the determinant of the covariance matrix Σ.

$$p(v|c_k;\Sigma)=\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}e^{-\frac{1}{2}v^T\Sigma^{-1}v} \qquad (5)$$

In addition, the previous learned distributions can be employed so as to generate random trajectory shapes representing random trajectories.

6   Experimental results

To evaluate the effectiveness of our method we collect data from two different subjects. Each subject performs the two different movements 20 times resulting in 40 samples per subject. We use samples from subject 1 to learn distributions and samples and used them for subject 2 as test. The process is repeated with subject 2 for training and use for subject 1 as test. The average accuracy is reported in Table1. We propose to analyze each marker trajectory separately to identify which part of the arm is more important to distinct the two intentions. In addition, the whole movement can be decomposed in two phases as illustrated in Figure 1. The first phase (Segment A) corresponds to the reaching movement where the participant take the object. The second phase (Segment B) corresponds to the putting movement where the participant put the object on the table. During our experiments, we propose to analyze these segments separately as well as the full movement (Segment A+B).

Table 1: Categorization accuracy in % for each of the markers.

Markers	Hand	Wrist	Elbow	Shoulder
Seg A	55.3	60.5	60.5	64.1
Seg B	66.7	**69.2**	61.6	59.0
Seg A+B	61.6	60.5	64.1	50.0

From the table, we can first observe that we obtain our better result when analyzing the segment B of the trajectory corresponding to the wrist marker (69.2%). We can also see that analyzing the segment B is more discriminant than segment A for most of the markers. However, this observation is not valid for the shoulder marker. This shows that in order to guarantee a better categorization to the intentions, a different marker should be analyzed depending on the motion segments. This encourage us to consider this aspect as future work. In addition, we also would like to extend the analysis by analyzing several markers simultaneously instead of separately. This would allows to combine strength of each marker.

[1] M. Devanne, H. Wannous, S. Berretti, P. Pala, M. Daoudi, and A. Del Bimbo. 3D human action recognition by shape analysis of motion trajectories on riemannian manifold. *IEEE Trans. on Cybernetics*, 45(7):1340–1352, 2014.

[2] H. Drira, B. Ben Amor, A. Srivastava, M. Daoudi, and R. Slama. 3D face recognition under expressions, occlusions, and pose variations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(9):2270–2283, 2013.

[3] H. Karcher. Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30:509–541, 1977.

[4] S. Kurtek, A. Srivastava, E. Klassen, and Z. Ding. Statistical modeling of curves using shapes and related features. *Journal of the American Statistical Association*, 107(499):1152–1165, 2012.

[5] R.G. Marteniuk, C.L. MacKenzie, M. Jeannerod, S. Athenes, and C. Dugas. Constraints on human arm movement trajectories. *Canadian Journal of Psychology*, 41:365–378, 1987.

[6] F. Quesque and Y. Coello. Perceiving what you intend to do from what you do: Evidence for embodiment in social interactions. *Socioaffective Neuroscience and Psychology*, 5:365–378, 2015.

[7] F. Quesque, Y. Delevoye-Turrell, and Y. Coello. Facilitation effect of observed motor deviants in a cooperative motor task: Evidence for direct perception of social intention in action. *Quarterly Journal of Experimental Psychology*.

[8] F. Quesque, D. Lewkowicz, Y. Delevoye-Turrell, and Y. Coello. Effects of social intention on movement kinematics in cooperative actions. *Frontiers in Neurorobotics*, 7(14):1–10, 2013.

[9] A. Srivastava, E. Klassen, S. H. Joshi, and I. H. Jermyn. Shape analysis of elastic curves in euclidean spaces. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(7):1415–1428, 2011.

[10] A. Srivastava, E. Klassen, S.H. Joshi, and I.H. Jermyn. Shape analysis of elastic curves in euclidean spaces. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(7):1415–1428, 2011.