



HAL
open science

Statistical gesture models for 3D motion capture from a library of gestures with variants

Zhenbo Li, Patrick Horain, André-Marie Pez, Catherine Pelachaud

► **To cite this version:**

Zhenbo Li, Patrick Horain, André-Marie Pez, Catherine Pelachaud. Statistical gesture models for 3D motion capture from a library of gestures with variants. Gesture in Embodied Communication and Human-Computer Interaction: 8th International Gesture Workshop, GW 2009, Bielefeld, Germany, February 25-27, 2009, Springer, 2010, Lecture notes in artificial intelligence. hal-00473788

HAL Id: hal-00473788

<https://hal.science/hal-00473788>

Submitted on 16 Apr 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Statistical Gesture Models for 3D Motion Capture from a Library of Gestures with Variants

Zhenbo LI¹, Patrick HORAIN¹, André-Marie PEZ², Catherine PELACHAUD^{2,3}

¹ Institut Telecom, Telecom SudParis, 9 rue Charles Fourier, 91011 Evry Cedex, France

² Institut Telecom, Telecom ParisTech, 46 rue Barrault, 75634 Paris Cedex 13, France

³ CNRS, LTCI, 46 rue Barrault, 75634 Paris Cedex 13, France

zhenboli@gmail.com, Patrick.Horain@Telecom-SudParis.eu,

andremarie.pez@telecom-paristech.fr, catherine.pelachaud@telecom-paristech.fr

Abstract. A challenge for 3D motion capture by monocular vision is 3D-2D projection ambiguities that may bring incorrect poses during tracking. In this paper, we propose improving 3D motion capture by learning human gesture models from a library of gestures with variants. This library has been created with virtual human animations. Gestures are described as Gaussian Process Dynamic Models (GPDM) and are used as constraints for motion tracking. Given the raw input poses from the tracker, the gesture model helps to correct ambiguous poses. The benefit of the proposed method is demonstrated with results.

Keywords: Gaussian Process; 3D motion capture; gesture model; gesture library.

1 Introduction

Avatars are virtual self representations. They evolve in a 3D world and interact with other virtual entities on our behalf. Avatars are animated by their human counterpart. One difficulty is the control of their behaviors. Selecting behaviors from a menu or by using icons is tedious. Moreover the avatar animation is not lively as it moves only on command. Some attempts have been made to alleviate the user's role and endow avatars with some autonomy [1]. These approaches are promising but require intensive computation to parse what the user aims to say and derive an appropriate nonverbal behaviors that accompany this text. In our project we aim at letting the user fully control his avatar [2]. 3D user motion is captured through a plain webcam using computer vision algorithms, and rendered by his avatar.

Vision-based human body tracking allows inexpensive, non-obtrusive marker-less motion capture [3]. Research in this field has been motivated by numerous target applications: human-computer interfaces, animation, interaction with virtual environments, video surveillance, games, etc. Monocular vision based markerless 3D motion capture is a difficult problem because of the ambiguities resulting of the lack of depth information, partial occlusion of human body parts, high number of degrees

of freedom, variations in the proportions of the human body and various clothing [3,4,26]. It has an intrinsic limitation of poor precision and robustness.

In this paper, we address the issue of improving motion capture with statistical gesture models. These models require training on some relevant gesture databases. Most existing databases as CMU Graphics Lab Motion Capture Database (<http://mocap.cs.cmu.edu/>) and HumanEva (<http://vision.cs.brown.edu/humaneva/>) gather data on human actions such as running, jumping and the like, but very few include communicative gestures. The variety of communicative gestures is a challenge by itself. Even though some gestures with a defined shape can be linked to a precise meaning, most of the communicative gestures are creative [5] that is they are created on the spot. Thus, gesture shape can vary a lot depending on their meaning and discourse context. Building a library of such gestures is a tremendous enterprise.

To avoid this problem, we have selected the context of a recruiting interview and we have gathered video data from 4 interviews, each about 30 minutes. We have generated gesture by mimicking real user motion and their variations using the Greta expressive conversational agent [6]. The resulting library of gestures has then been described with a statistical model suitable for constraining tracking and disambiguating monocular 3D people tracking.

The rest of the paper is organized as follows. Section 2 introduces the related works about gesture modeling in a low-dimensional space. In section 3, we present our real-time monocular motion capture system. Then, we present our work toward building a library of communicative gestures in section 4. We describe also how we encompass user's communicative distinctiveness in our library. In section 5, we propose using Gaussian Process Dynamic Models (GPDM) [7] to statistically describe gesture models and use them as guides for motion tracking. The interest of this approach is demonstrated with results.

2 Related Work

Although people can perform very large variation of complex motions, their movements can be represented in a low-dimensional space. Pullen *et al.* [8] observed human motions have certain cooperative relationships especially when people do some specific movements like walking, swimming, etc. This relationship can be used to reduce the parameter space dimensionality while performing human motion analysis. Safonova *et al.* [9] have demonstrated that many dynamic human motions can be adequately represented with only five to ten parameters. Elgammal *et al.* [10] and Grochow *et al.* [11] observed that human activities can be described in a latent space. So, human motion can be modeled in a low-dimensional latent space.

Building the latent motion space from existing motion data consists in defining a subspace with a lower dimension than the full motion capture data. Because human motion is non-linear, basic dimensionality reduction methods such as principal component analysis (PCA) are inadequate to describe non-linear human motion [4]. Other methods such as Locally Linear Embedding (LLE) [27] and Isomap[28] either do not provide invertible mapping from the low dimensional latent space to the

original pose space or do not provide probability distribution over data in latent space. They are not suitable to build low dimensional gesture models [7].

Locally Linear Coordination (LLC) [12], Gaussian Process Latent Variable Models (GPLVM) [13] and later appeared approaches like Gaussian Process Dynamic Models (GPDM) [7] and Laplacian Eigenmaps Latent Variable Model (LELVM) [14] can learn a non-linear mapping between the human motion parameter space and a latent space and they provide an inverse mapping. They allow describing human motion in a low-dimensional latent space.

Recently, latent gesture models have been used as prior constraints to help 3D human motion tracking. Urtasun [15, 16] used GPLVM and GPDM to learn prior models for tracking 3D human walking. She achieved good results even in case of serious occlusion. Raskin *et al.* [17] presented an approach to combine annealed particle filter tracker with GPDM that allows reducing the state vector and that enhances tracking stability. Moon and Pavlovic [19] investigated the effect of dynamics in dimensionality reduction problems on human motion tracking. Lu *et al.* [20] used LELVM as constraints in the probabilistic sigma point mixture tracker for robust operation with missing data, noisy and ambiguous image measurements. In these approaches, gesture models were combined with the image-based likelihood in the tracker to reduce the number of state sampling particles [21].

3 Real-Time 3D Motion Capture with a Webcam

Our baseline is a real-time webcam-based system for 3D motion capture that was previously developed in our team [18]. It works by registering an articulated 3D model of the human body on a video stream (Fig. 1).

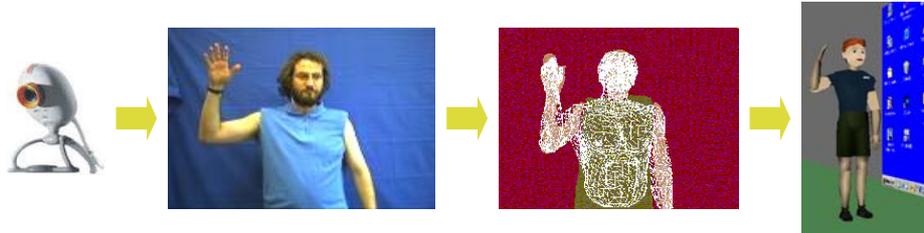


Fig. 1. Real time motion capture by monoscopic computer vision and virtual rendering.

Our 3D human model has 3 global position parameters and 20 joint angles of the upper-body (bust, arms, forearms, hands, neck and head), so a body-pose is represented by a vector of 23 parameters. For each input image we search for the model pose that best matches the image. Image features (color regions, edges) are extracted and matched with model features (colored limbs, occluding edges) projected in the candidate pose. For each captured image, optimal registration is searched with respect to the pose parameters by iteratively maximizing the color region overlap and by minimizing the distance between the image edges and the projected occluding edges of the model (Fig. 2) [18].

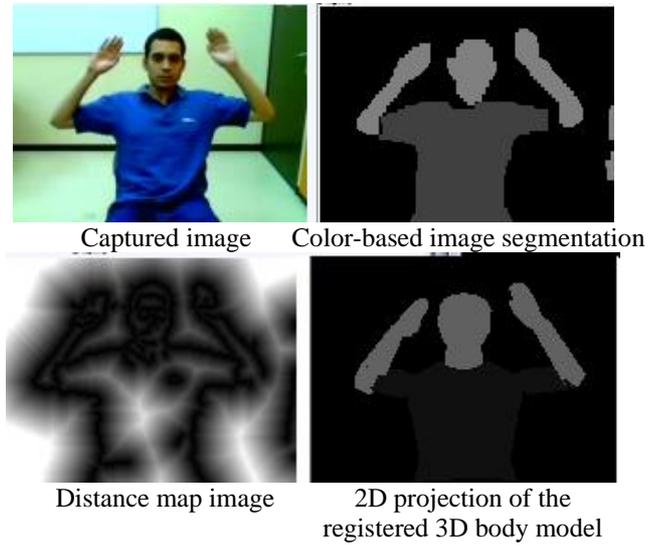


Fig. 2. The captured image, the color-segmented image, distance map image edges in the foreground mask and finally the projection of the 3D human body model in the optimal matching pose [18].

Joint angles are then output in real-time over the network as low bandwidth MPEG-4 body animation parameters (BAPs). The captured motion is rendered remotely by animating the user avatar in the virtual space (Fig. 3).



Fig. 3. Real-time motion capture using a single camera for each user and virtual rendering. Demonstration videos are available at <http://MyBlog3D.com>.

4 Creating a Library of Gestures

Several dictionaries of emblematic gestures have been gathered in our project. They are bound to a given culture. These dictionaries provide detailed information about the gesture shape and its associated meaning. Other attempts have looked to describe gesture shape in association with their physical meaning [22]. Raised hand with palm facing one's interlocutor carries the meaning to stop something or somebody to do something. It can be viewed as symbolizing a wall between both interlocutors.

In our work we are interested at all types of gestures as we aim to track any hand and arm movements done while communicating. We do not aim to recognize gestures or to interpret them. We are interested in detecting their shape and following their movement in view in reproducing them by the avatar. This reproduction does not require understanding the meaning of the gestures. With such an aim we decided to consider only one feature of the gestures: their shape.

Most of the existing databases of motion capture data gather data on action movements such as walking, jumping or running. Almost no databases are centered on communicative gestures. In our work, our aim is to track people gesturing while conversing. To be able to train our tracking algorithm, we need to gather data on nonverbal gestures. Rather than creating a database of motion capture data, we have decided to use virtual agent technology as it is cheaper. We have created a database with the animation of the synthetic character. It is difficult to select which gestures to consider. Communicative gestures are often 'creative'. It is not possible to create an extensive library of communicative gestures. At a first step of our work, we have decided to focus on one conversational domain: recruiting interviews. We have gathered data of real people going through job interview and we have searched for the gestures shapes that are the most frequent. The agent was made to replay those gestures. To ensure to gather gesturing variability of users, we have applied a set of expressivity parameters that modulate the animation of the virtual agent.

4.1 Gesture Variability

While communicating, people show large variability not only in their intentions but also in their way of expressing them. One can be characterized by a signature, a style one carries along in, basically, all circumstances [23].

We have developed a model of distinctive agent that encompasses variability in the modality preference used to communicate a given intention and on the behavior expressivity [24]. In particular when modulating the last set of parameters, the agent can display gestures more or less extent, more or less fast and powerful, etc. These variations occur at the level of execution of behaviors and not on the type of behaviors to be displayed.

4.2 Gesture Clustering

In our corpus we have gathered data from 8 interlocutors. The data was annotated using ANVIL [25]. Around 800 communicative gestures were found in the data. We gathered them into classes of gestures looking alike in their shape and movement.

Our aim is to create a library to train our gesture tracking algorithm [2]. To ensure robustness of our tracking algorithm over a large population of users, we have enhanced the library of gestures of the training phase with gesture variability using the model of distinctive agent [24]. The gesture of each class found in our corpus is reproduced by a virtual agent (Fig. 4). The library contains the reproduced gesture as well as the same gesture with different expressivities. Thus each gesture in a class and its variations are present.



Fig. 4. Rendering by the virtual agent of a gesture belonging to a gesture class.

5 Gesture Modeling and 3D Motion Capture

5.1 Gaussian Process Dynamical Models

Gaussian Process Dynamic Models (GPDMs) are a powerful approach for probabilistically modeling high dimensional time related data through dimension reduction that make it possible to learn probability gesture models from small training data sets [7]. We propose hereafter a brief introduction to GPDMs.

A GPDM consists of 1) a continuous mapping between the full-dimensional data space (joint angles) and a low-dimensional latent space and 2) a dynamical model in

the latent space. It is obtained by marginalizing out the parameters of the two mappings, and optimizing the latent coordinates of training data.

GPDMs aim at modeling the probability density of a sequence of vector-valued states $y_1, \dots, y_t, \dots, y_N$ with discrete-time index t and $y_t \in \mathbb{R}^D$ where D is the number of parameters that define a body pose in the full dimensional parameters space. These two mapping can be expressed as:

$$x_t = f(x_{t-1}; A) + n_{x,t} \quad (1)$$

$$y_t = g(x_t; B) + n_{y,t} \quad (2)$$

where $x_t \in \mathbb{R}^d$ denotes the d dimensional latent coordinates at time t . $n_{x,t}$ and $n_{y,t}$ are zero-mean white Gaussian noise processes. f and g are (nonlinear) mappings defined as linear combinations of basis functions ϕ_i and φ_i with weight vectors A and B :

$$f(x; A) = \sum_i a_i \phi_i(x) \quad (3)$$

$$g(x; B) = \sum_j b_j \varphi_j(x) \quad (4)$$

where $A = [a_1, a_2, \dots]$, $B = [b_1, b_2, \dots]$.

The specific forms of f and g will be marginalized out in GPDM. With an isotropic Gaussian prior on each b_j , we can marginalizing over B in closed form [29] to yield a multivariate Gaussian data likelihood:

$$p(Y/X, \bar{\beta}) = \frac{|W|^N}{\sqrt{(2\pi)^{ND} |K_Y|^D}} \exp\left(-\frac{1}{2} \text{tr}(K_Y^{-1} Y W^2 Y^T)\right) \quad (5)$$

where $Y = [y_1, \dots, y_N]^T$ is a matrix of training poses, $X = [x_1, \dots, x_N]^T$ is a matrix of latent positions, K_Y is a kernel matrix, and $\bar{\beta} = \{\beta_1, \beta_2, \dots, W\}$ comprises the kernel hyperparameters. $W \equiv \text{diag}(w_1, \dots, w_D)$ is a scaling matrix used to account for the different variances in the different data dimensions. The kernel matrix elements are defined by a kernel function $(K_Y)_{i,j} = K_Y(x_i, x_j)$. For the latent mapping, $X \rightarrow Y$, the Radial Basis Function (RBF) kernel is used.

$$k_Y(x, x') = \beta_1 \exp\left(-\frac{\beta_2}{2} \|x - x'\|^2\right) + \beta_3^{-1} \delta_{x,x'} \quad (6)$$

Hyperparameter β_1 represents the overall scale of the output function, while β_2 corresponds to the inverse width of the RBFs. The variance of the noise term $n_{y,t}$ is given by β_3^{-1} .

The dynamic mapping on the latent coordinates X is:

$$p(X / \bar{\alpha}) = p(x_1) \frac{1}{\sqrt{(2\pi)^{(N-1)d} |K_X|^d}} \exp\left(-\frac{1}{2} \text{tr}(K_X^{-1} X_{out} X_{out}^T)\right) \quad (7)$$

Where $X_{out} = [x_2, \dots, x_N]^T$, K_X is the $(N-1) \times (N-1)$ kernel matrix constructed from $\{x_1, \dots, x_{N-1}\}$, and x_1 is assumed to have an isotropic Gaussian prior. Where $\bar{\alpha}$ is a vector of kernel hyperparameters. The dynamics can be modeled using the following ‘‘Linear + RBF’’ kernel:

$$k_X(x, x') = \alpha_1 \exp\left(-\frac{\alpha_2}{2} \|x - x'\|^2\right) + \alpha_3 x^T x' + \alpha_4^{-1} \delta_{x,x'}. \quad (8)$$

Hyperparameters α_1 , α_2 represent the output scale and the inverse width of the RBF terms, and α_3 represents the output scale of the linear term. Together, they control the relative weighting between the terms, while α_4^{-1} represents the variance of the noise term $n_{x,t}$.

Learning the GPDM from measurements $Y = [y_1, \dots, y_N]^T$ entails minimizing the negative log-posterior:

$$\begin{aligned} \ell &= -\ln p(X, \bar{\alpha}, \bar{\beta} / Y) \\ &= \frac{d}{2} \ln |K_X| + \frac{1}{2} \text{tr}(K_X^{-1} X_{out} X_{out}^T) + \sum_j \ln a_j \\ &\quad - N \ln |W| + \frac{D}{2} \ln |K_Y| + \frac{1}{2} \text{tr}(K_Y^{-1} Y W^2 Y^T) + \sum_j \ln \beta_j \end{aligned} \quad (9)$$

5.2 Gesture Modeling

Gesture models are learnt offline from the gesture library using GPDM. Here we use a 3D latent space as this appears to be the lowest dimension for robustly learning complex motions with stylistic variability [16]. We select gestures from each subject for training in order to get the mean trajectories and variances of gestures in the latent space. The input matrix for GPDM learning is pose parameters, which are 20 joint angles of the upper-body (bust: 3DOFs, left shoulder: 3DOFs, right shoulder: 3DOFs, left forearm: 1DOFs, right forearm: 1DOFs, left hand: 3DOFs, right hand: 3DOFs, neck and head: 3DOFs). Fig. 5 shows two conversational gestures from our library and their description as trajectories in a 3D latent space.

5.3 Gesture Modeling for 3D Motion Capture

In our model-based approach, tracking relies on evaluating how well some synthesized appearance of the human body model matches the input image, *i.e.* how

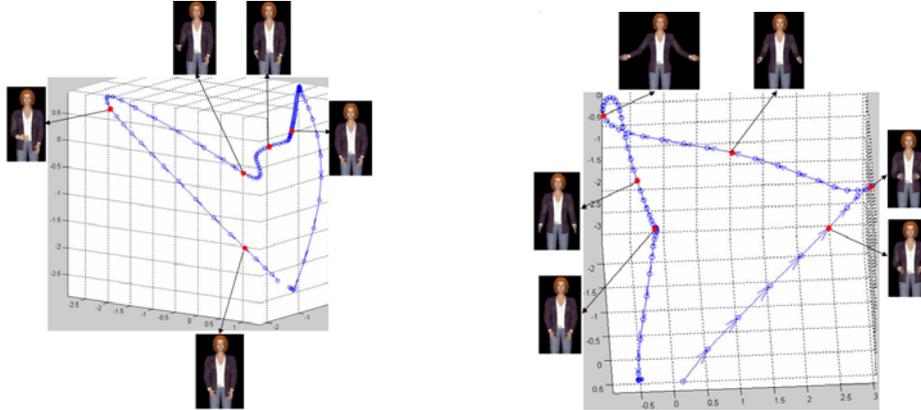


Fig. 5. Conversational gestures described in a 3D latent space. Each circle on the trajectories represents a body pose.

well some model instantiation explains the input image. However, because we only use a single camera, depth ambiguities can occur. Furthermore, because real-time processing implies limited computation power, only a limited subset of the human body degrees of freedom can be processed, so the hand pose cannot be captured, and image size may even be limited. We use the gesture model as a constraint to raise ambiguities and augment the captured motion with details (Fig. 6).

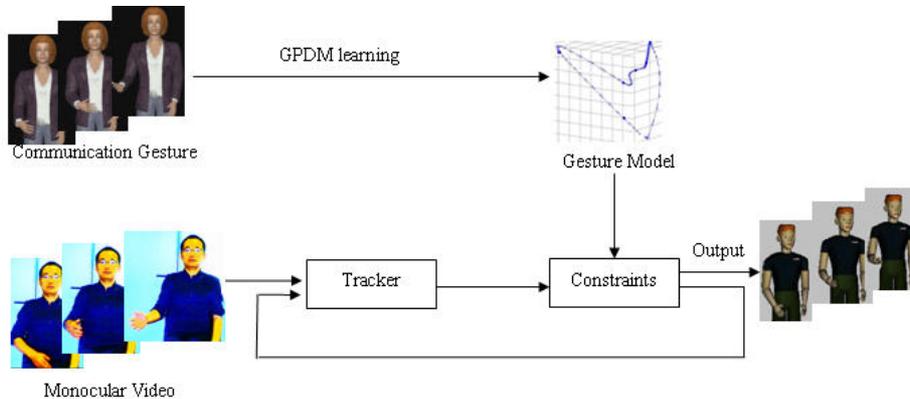


Fig. 6. Gesture model working as constraints in tracking

For each iteration, the motion tracker outputs a candidate pose to be constrained with the gesture model. That pose is projected from the full motion parameter space to the 3D latent model space and then replaced with the closest point on the latent motion trajectory. Since GPDM mapping is continuous, poses that are close in the full motion parameter space remain close in the latent space. The output constrained pose at each time step is the pose reconstructed into the full motion space. The point used

for pose reconstruction is the closest point from the projected point on the model trajectory in the latent space. The resulting pose is then used as the initial pose for tracking at the next image.

5.4 Experiments

We used those gesture models as constraints while tracking communicative gestures and actions gestures from videos. As we only using one webcam, the tracker can not always distinguish poses correctly due to lacking depth information (for example, hands orientation are sometimes incorrectness). In the gesture models based tracking, each candidate pose is projected into the latent gesture model space, so enforcing the captured motion regularity.

We tested our approach on 4 subjects of communication gestures (each test video is about 90 frames). Because the output poses lay on the gesture model internal trajectory, most of the monocular vision ambiguities can be solved. Impossible poses will not happen in the gesture model space, so heuristic biomechanical constraints, which otherwise must be used for pruning the pose space, can be replaced with the gesture model that constrain the output poses to be on the learnt motion trajectory (fig.7). Another benefit of this approach is that the poses reconstructed in the latent space include motion details that cannot be captured from the input video sequence (such as hand shapes) (fig.8).

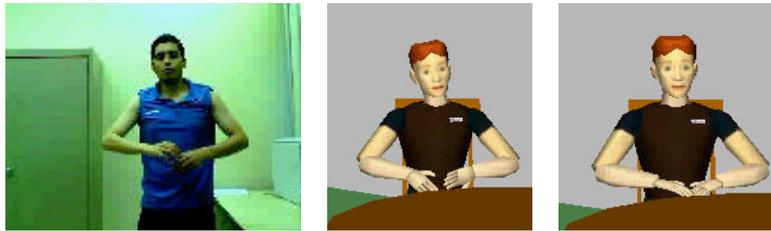


Fig.7. Motion tracking with biomechanical constraints vs. gesture model. **Left:** An input image where the left hand orientation can hardly be distinguished. **Middle:** Motion capture result with heuristic biomechanical constraints: the unlikely pose of the left hand is biomechanically possible, so it is accepted. **Right:** Tracking with gesture model instead of biomechanical constraints: the gesture model avoids the awkward pose.



Fig.8. Augmenting motion capture with gesture models. **Left:** Input image. **Middle:** Tracking without gesture model: the hand shape is not captured to meet real-time computation constraints. **Right:** The gesture model augments the motion capture with fingers movements.

6 Discussion and Future Work

We have created a library of communication gestures with variants and used it to learn gesture models with a GPDM statistical approach. These models allow both non-linear mapping for reconstruction and dimension reduction between the motion parameters space and a low-dimensional latent space, while being simple to learn from small training data. The main contribution of this work is used the low-dimensional gesture models as prior constraints for monocular motion capture while tracking communicative gestures. These gesture models help track ambiguous poses and render some motion details.

Our approach works on specific gestures. Future work is still required to be done to support broader classes of gestures.

Acknowledgement

The computer vision motion capture we used was widely developed by David Antonio Gómez Jáuregui.

References

1. Vilhjálmsón, H.: Avatar Augmented Online Conversation, Ph.D. thesis, Media Arts and Sciences, Massachusetts Institute of Technology, Media Laboratory, Cambridge, MA, 2003.
2. Horain, P., Marques Soares, J., Rai, P. K., Bideau, A.: Virtually enhancing the perception of user actions. In: 15th International Conference on Artificial Reality and Telexistence (ICAT 2005), Christchurch, New Zealand, pp. 245–246 (2005). DOI:10.1145/1152399.1152446.
3. Moeslund, T., Hilton, A., Kruger, V.: A survey of advances in vision-based human motion capture and analysis. *Computer vision and image understanding* 104 (2-3), pp. 90–126 (2006). DOI:10.1016/j.cviu.2006.08.002.
4. Poppe, R.W.: Vision-based human motion analysis: An overview. *Computer Vision and Image Understanding* 108 (1-2), pp. 4-18 (2007).
5. Poggi, I., Mind, Hands, Face and Body. A Goal and Belief View of Multimodal Communication, Weidler Verlag, Körper, Zeichen, Kultur, n° 19, 2007.
6. Bevacqua, E., Mancini, M., Niewiadomski, R., Pelachaud, C.: An expressive ECA showing complex emotions. In: AISB'07 Annual convention, workshop “Language, Speech and Gesture for Expressive Characters”, Newcastle UK, pp. 208–216 (2007).
7. Wang, J. M., Fleet D. J., Hertzmann A.: Gaussian Process Dynamical Models for Human Motion. *IEEE Transactions on PAMI*, 30 (2), pp. 283–298 (2008). DOI:10.1109/TPAMI.2007.1167.
8. Pullen, K., Bregler, C.: Motion capture assisted animation: Texturing and synthesis. In: SIGGRAPH 2002, pp. 501–508 (2002).
9. Safonova, A., Hodgins, J. K., Pollard, N. S.: Synthesizing physically realistic human motion in low-dimensional, behavior-specific spaces. *ACM Transactions on Graphics* 23 (3), pp. 524–521 (2004).

10. Elgammal, A. M., Lee, C.-S.: Inferring 3D body pose from silhouettes using activity manifold learning. In: Conference on Computer Vision and Pattern Recognition (CVPR'04), vol. 2, pp. 681–688 (2004).
11. Grochow, K., Martin, S. L., Hertzmann, A., Popovic, Z.: Style-based inverse kinematics. *ACM Transactions on Graphics* 23 (3), pp. 522–531 (2004).
12. Teh, Y. W., Roweis, S. T.: Automatic alignment of local representations. In: *Neural Information Processing Systems 15 (NIPS'02)*, pp. 841–848 (2003).
13. Lawrence, N. D.: Gaussian process latent variable models for visualisation of high dimensional data. In: Thrun, S., Saul, L., Schölkopf, B. (eds): *Advances in Neural Information Processing Systems*, MIT Press, Cambridge, MA, pp. 329–336 (2004).
14. Carreira-Perpiñán, M. Á., Lu, Z.: The Laplacian Eigenmaps Latent Variable Model. In: 11th International Conference on Artificial Intelligence and Statistics (AISTATS), Puerto Rico (2007).
15. Urtasun, R., Fleet, D. J., Hertzmann, A., Fua, P.: Priors for people tracking from small training sets. In: *International Conference On Computer Vision (ICCV'05)*, vol. 1, Beijing, China, pp. 403–410 (2005).
16. Urtasun, R., Fleet, D. J., Fua, P.: 3D people tracking with Gaussian process dynamical models. In: *Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 1, New York, NY, pp. 238–245 (2006).
17. Raskin, L., Rivlin, E., Rudzsky, M.: Dimensionality Reduction for Articulated Body Tracking. In: *3DTV07*, pp. 1–4 (2007).
18. Gómez Jáuregui, D. A., Horain, P.: Region-based vs. edge-based registration for 3D motion capture by real time monoscopic vision. In: Gagalowicz, A., Philips, W. (eds.) *MIRAGE 2009*. LNCS, vol. 5496, pp. 344–355 (2009). Springer, Heidelberg. DOI:10.1007/978-3-642-01811-4_31.
19. Kooksang Moon, Vladimir I. Pavlovic': Impact of dynamics on subspace embedding and tracking of sequences, in: *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 1, New York, NY, June, pp. 198 – 205 (2006).
20. Lu, Z., Carreira-Perpiñán, M. Á., Sminchisescu, C.: People Tracking with the Laplacian Eigenmaps Latent Variable Model. *Advances in Neural Information Processing Systems (NIPS 21)* (2007).
21. Isard, M., Blake, A.: Condensation-conditional density propagation for visual tracking. *Int. J. Computer Vision* 29(1), pp.5–28 (1998).
22. Calbris, G.: *The semiotics of French gestures*. University Press, Bloomington: Indiana (1990).
23. Gallaher, P. E.: Individual differences in nonverbal behavior: Dimensions of style. *Journal of Personality and Social Psychology*, 63(1), pp.133–145 (1992).
24. Mancini, M., Pelachaud, C.: Distinctiveness in multimodal behaviors. In: 7th International Joint Conference on Autonomous Agents and Multi-Agent Systems, AAMAS'08, Estoril Portugal (May 2008).
25. Kipp, M.: Anvil - A Generic Annotation Tool for Multimodal Dialogue. In: 7th European Conference on Speech Communication and Technology (Eurospeech), pp. 1367–1370, Aalborg (2001).
26. Davis, J., Agrawala, M., Chuang, E., Popovic, Z., Salesin, D.: A Sketching Interface for Articulated Figure Animation. *Eurographics/SIGGRAPH Symposium on Computer Animation (SCA)*, (2003).
27. Sam R., Lawrence S. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500) , pp.2323--2326.(2000).
28. Tenenbaum. J. B., V. de Silva and J. C. Langford. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science* 290 (5500), pp. 2319-2323. (2000).
29. Neal R. M.: *Bayesian Learning for Neural Networks*. Lecture Notes in Statistics No. 118. Springer-Verlag, (1996).