In: Rolf P. Würtz, Markus Lappe (Eds.)
Dynamic Perception, Berlin: AKA Press, 2002, pp 115-120

The little difference: Fourier based synthesis of gender-specific biological motion

Nikolaus Troje

Fakultät für Psychologie, Ruhr-Universität-Bochum, 44780 Bochum
troje@uni-bochum.de, www.bml.psy.ruhr-uni-bochum.de

Abstract. A framework is outlined that can be employed to obtain gender and other characteristics of the agent from human motion patterns and subsequently use this information to synthesize motion with particular, well-defined biological and psychological attributes. The proposed model is based on the statistics of a data base of motion capture data. Based on linearization of the motion data, a motion space is defined which is spanned by the first few principal components obtained from the data base of input walkers. Using biological and psychological traits attributed to the input walkers, linear discriminant functions are computed which define vectors in the motion space that generalize the respective trait. These vectors are in turn used to generate walking patterns with the respective properties.

1 Introduction

Biological motion contains plenty of information about identity, personality traits and emotional state of the moving person. The human visual system is extremely sensitive to retrieve such information from motion patterns. We can recognize a familiar person by the way he or she walks and we can attribute gender and age as well as psychological attributes such as personality traits and emotions to an unfamiliar person with motion being the only source of information. We are also extremely sensitive in detecting deviations from natural behaviour. The high degree of perceived realism of modern computer graphics in animated movies and computer games is often disturbed by the fact that the animated movements are perceived to be unnatural. For modern atavars or in the case of virtual replacements of real actors (“virtual stunt men”) the observer is not supposed to even realize that the real actor is temporarily replaced by a digital character. To achieve the desired realism, there is considerable demand on methods to synthesize psychologically convincing biological motion.

I want to outline a framework that can be employed to obtain parameterizations of biological or psychological attributes from human motion. Subsequently, I will use this information to synthesize motion with the respective attributes. Gender classification is used as the main example, but I also present examples of how the framework can be applied to other attributes.

The data material to start with is raw motion capture data, i.e. the three-dimensional trajectories of discrete points on a persons body. The primary goal is to transform those data into a representation that would allow us to apply standard methods from linear statistics and pattern recognition. Such representations have been termed "morphable models" [1-3] in the computer vision community, expressing the fact that the linear transition from one item to a second item of the data set represents a well defined, smooth metamorphosis. Another term that has been used for the same class of models in the context of human face recognition is "correspondence-based representations" [4,5]. This term focuses on the fact that morphable models rely on establishing correspondence between features across the data set resulting in a separation of the overall information into range specific information on the one hand and domain specific information on the other hand [6].

The procedure developed in the present study contains elements of earlier work on parameterizations of animate motion patterns [1,7-10]. Unuma [7] showed that blending between different motions works much better in the frequency domain. At least for periodic motions, such as most locomotion patterns, Fourier decomposition can be used to achieve efficient, low-dimensional, linear decompositions. In fact, decomposing the time series of postures of a single walking person by means of principal component analysis reveals components, which are almost similar to Fourier components [10]. This demonstrates that Fourier decomposition of walking data is nearly optimal in terms of covering a maximum of variance with a minimum of components.

The focus of the current study is to obtain a system that is sensitive enough to extract biologically and psychologically relevant attributes. Based on the linearization of the motion data, a motion space is defined which is spanned by the first few principal components obtained from a set of input walkers. Within this space, linear discriminant functions are computed that generalize the respective trait. Those vectors are in turn used to generate walking patterns with the respective properties in a psychologically convincing manner.

2 Linearization of motion capture data

For the current study, twenty men and twenty women, most of them students and staff of the Psychology department of the Ruhr-University served as models to acquire motion data. A set of 38 retroreflective markers was attached to their body. Participants wore swimming suits and most of the markers were attached directly to the skin. Others, like the ones for the head, the ankles and the wrists were attached to elastic bands and the ones on the feet were taped onto the subjects' shoes.

Participants were then placed on a treadmill and were asked to walk. They could adjust the speed of the treadmill such that they felt most comfortable. To ensure that they did not feel too much under observation and that they did not "perform" in an unnatural manner, we let them walk for at least 5 minutes before we started to record 20 steps (i.e. 10 full gait cycles) from each of them.

Recording was done by means of a motion capture system (Vicon 512, Oxford Metrics). The system tracks the three-dimensional trajectories of the markers with spatial accuracy in the range of 1 mm and a temporal resolution of 120 Hz.

Based on the trajectories of the 38 original markers, we computed the location of "virtual" markers positioned at major joints of the body. The 15 virtual markers used for all the subsequent computations where located at the ankles, the knees, the hip joints, the wrists, the elbows the shoulder joints, at the center of the pelvis, on the clavicula and in the center of the head.

The walk of an individual subject can be regarded as a time series of postures. Each posture can be described in terms of the position of the 15 markers. Since three coordinates are needed for each position the representation of a single posture is a 45 dimensional vector $p=(m1_x, m1_y, m1_z, m2_x \ldots m15_z)^T$.

Linearization of the data was achieved in two steps. In the first step, the series of postures obtained from a single walker j was decomposed into a second order Fourier expansion:

$$p_j(t) = p_{j,0} + p_{j,1}\sin(\omega_j t) + p_{j,2}\cos(\omega_j t) + p_{j,3}\sin(2\omega_j t) + p_{j,4}\cos(2\omega_j t) + err_j \quad (1)$$

The power carried by the residual term err is less than 3% of the power of the input data and we discard it from all further computations. A particular subject's walk is therefore approximated by specifying the average posture $p_{j,0}$, the four characteristic postures $p_{j,1}$, $p_{j,2}$, $p_{j,3}$, and $p_{j,4}$, and the fundamental frequency ω_j. Since each of the components is a 45 dimensional vector, the dimensionality of the model at this stage is 226=5∗45+1.

Although this number already reflects a considerable reduction in dimensionality as compared to the raw motion capture data the number of effective degrees of freedom within the database is probably much smaller. For classification purposes it is necessary to reduce the dimensionality of the representation such that the number of dimensions becomes much smaller than the number of items represented in the resulting space.

The advantage of the above representation is, that it provides the possiblity to successfully apply linear operations. Linear combinations of existing walking patterns result in new walking patterns which meaningfully represent the transitions between the constituting patterns [7,10]. We can therefore treat the 226 dimensional vector describing the walk w_j of walker j as a point in a linear space of the same dimension and apply linear methods.

This makes it also possible to use principal components analysis in order to further reduce dimensionality. Applying PCA to the set of walkers W results in a decomposition of each walker into an average walker v_0 and a weighted sum of Eigenwalkers v_k.

$$w_j = v_0 + \Sigma k_{i,j} v_i \quad (2)$$

or in Matrix notation:

$$W = V_0 + VK \quad (3)$$

V_0 denotes a matrix with the average walker v_0 in each column. The matrix V contains the Eigenwalkers as column vectors v_i. Matrix K contains the weights (or the scores) $k_{i,j}$ and is obtained by solving the linear equation system:

$$VK = W - V_0 \quad (4)$$

The variance of the first 15 components sums up to 80% of the overall variance. Truncating the expansion (Eq. 2) after the 15th term thus means loosing 20% of the overall variance. For all further computations we used a space spanned by just those first 15 Eigenwalkers.

3 Gender discriminant function

Given this relatively low-dimensional linear representation of human walking patterns, we can now construct a linear classifier c accounting for gender-specific differences in human walking. This is achieved by finding the best solution (according to a least-square criterium) of the overdetermined linear system

$$cK = r \qquad \textbf{(5)}$$

r is the row vector containing 80 values r_j accounting for the desired output of the classifier. r_j equals 1 if walker j is male and -1 if the walker is female. K is the matrix containing the coefficients of each walker in the 15-dimensional Eigenwalker space. The resulting row vector c contains the coefficients of the linear discriminant function best accounting for the gender of the walkers.
The invertibility of the representation can be used to visualize what is happening along this discriminant function by displaying walkers $w_{c,\alpha}$ corresponding to different points along this axis as point-light displays or stick figure animations:

$$w_{c,\alpha} = w_0 + \alpha V c^T \qquad \textbf{(6)}$$

As above, w_0 denotes the average walker. The matrix V contains the first few Eigenwalkers - one in each column. As α changes from negative to positive values the walker appears to change its gender. On our Web page (*http://www.bml.psy.ruhr-uni-bochum.de/Demos/WDP2002.html*), such animations can be viewed and interactively manipulated by changing the value of α.
We have therefore retrieved a vector c that generalizes the attribute “gender” in the obtained motion space. Adding or subtracting this vector from a given walker makes its appearance more male or more female, respectively. The same procedure can be used to extract vectors accounting for other attributes as well. For our database, we registered for every walker a number of easily available attributes such as sex, age and weight. In addition to being able to change the perceived gender of a walker, the above mentioned demonstration also visualizes a dimension obtained from using the weight of the walker to compute a respective discriminant function. Light and heavy walkers show clear differences which are easily extracted by our visual system.
Other attributes, however, are not directly available but have to be determined through psychophysical experiments. In such experiments, observers are presented with displays of the 80 walkers and have to rate them on a 6 point scale with respect to the respective attribute. Here, we report the results of rating two different emotional attributes: happiness vs. sadness, and nervousness vs. relaxedness.

4 Psychophysical determination of emotional attributes

The walking patterns were displayed on a computer monitor as point-light displays subtending 5 deg of visual angle. Each of the 15 markers that were used for the above computation was rendered as a white dot on a black background using orthographic projection from one of three different viewpoints (0 deg = frontal view; 30 deg; 90 deg). The display therefore shows the positions of the major joints of the body changing over time. This results in a vivid percept of a walking human body without providing any information about the person except the one carried by the motion itself [11]. Point-light displays have been widely used in experimental psychology in order to isolate biological motion from other cues about identity, psychological and emotional attributes of a person [12-17, to mention just a few of the classic papers].

A single rating session consisted of 80 trials with each walker shown once for 7 s in a randomized order. All walkers within one session were shown from the same viewpoint. In order to indicate their rating observers had to hit one of 6 buttons displayed on the top of the screen above the point-light display by using the computer mouse. An intertrial interval of 3 s, during which a blank screen was shown, separated the trials. Six observers participated in the experiments. For three observers the most left and right buttons were labeled “happy” and “sad”, respectively. The other three observers were presented with the labels “nervous” and “relaxed”. Each observer carried out three sessions, one for each viewpoint, with short breaks between the sessions. The order of the three sessions was counterbalanced across observers.

The average of the ratings (across the three observers in each group and across the three different viewpoints) was used to form a vector r which, in turn, was used to compute the respective discriminant function c according to Equation 5. The animation at *http://www.bml.psy.ruhr-uni-bochum.de/Demos/WDP2002.html* visualizes the results. Animations both along the happy-sad axis as well as along the nervous-relaxed axis give a clear percept of a change in the respective emotions of the walker.

5 Discussion

Visualizing the respective discriminant functions shows that we have really captured the particular attribute and that the resulting walker vividly changes its characteristics in accordance with the intended characteristic. In all four cases examined so far, changes are a complex composite of structural and dynamic properties of the walker. For instance the exaggerated male walker has wider shoulders than hips whereas in the female walker this ratio reverses. Male walkers display considerable lateral body sway whereas this is not the case for female walkers. Hip motion in male walkers is 180 phase shifted with respect to the hip motion in female walkers. The position of the elbows is very different in male and female walkers. Men tend to hold their elbows away from the body whereas women hold them close to the body. In general, the exaggerated man seems to attempt to occupy much more space than the exaggerated woman -- a display not unique to the human species.

The differences in walking between light and heavy walkers are much harder to describe. Heavy persons have a somewhat smaller gait frequency and vertical movement components seem to be more pronounced in light-weighted walkers as compared to heavy walkers. However, there remains a discrepancy between the clear percept of a change in weight and the ability to identify the sophisticated composite features that communicate this information. The power of the proposed method for generating characteristic motion, however, is that it is not necessary to specify the features that carry the impression of changing biological or emotional attributes explicitely. Instead, we can extract them in terms of the statistical features of a data base that contains variations along the dimensions of interest.

References

1. Giese, M.A., Poggio, T.: Morphable models for the analysis and synthesis of complex motion patterns. International Journal of Computer Vision 38 (2000) 59-73
2. Jones, M.J., Poggio, T.: Multidimensional morphable models - a framework for representing and matching object classes. International Journal of Computer Vision 29 (1999) 107-131
3. Shelton, C.R.: Morphable surface models. International Journal of Computer Vision 38 (2000) 75-91
4. Troje, N.F., Vetter, T.: Representations of human faces. In: Taddei-Ferretti, C., Musio, C.(eds.): Downward processing in the perception representation mechanism. World Scientific, Singapore (1998) 189-205
5. Vetter, T., Troje, N.F.: Separation of texture and shape in images of faces for image coding and synthesis. Journal of the Optical Society of America A 14 (1997) 2152-2161
6. Ramsay, J.O., Silverman, B.W.: Functional data analysis. Springer, New York (1997).
7. Unuma, M., Anjyo, K., Takeuchi, R.: Fourier principles for emotion-based human figure animation. Computer Graphics Proceedings of SIGGRAPH 95 (1995) 91-96
8. Bruderlin, A., Williams, L.: Motion signal processing. Computer Graphics Proceedings of SIGGRAPH 95 (1995) 97-104
9. Witkin, A., Popovic, Z.: Motion warping. Computer Graphics Proceedings of SIGGRAPH 95 (1995) 105-108
10. Troje, N.F. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. Journal of Vision (in press)
11. Johansson, G.: Visual perception of biological motion and a model for its analysis. Perception & Psychophysics 14 (1973) 201-211
12. Dittrich, W.H.: Action categories and the perception of biological motion. Perception 22 (1993) 15-22
13. Mather, G., Murdoch, L.: Gender discrimination in biological motion displays based on dynamic cues. Proceedings of the Royal Society of London Series B 258 (1994) 273-279
14. Kozlowski, L.T., Cutting, J.E.: Recognizing the sex of a walker from a dynamic point-light display. Perception & Psychophysics 21 (1977) 575-580
15. Barclay, C.D., Cutting, J.E., Kozlowski, L.T.: Temporal and spatial factors in gait perception that influence gender recognition. Perception & Psychophysics 23 (1978) 145-152
16. Dittrich, W.H., Troscianko, T., Lea, S., Morgan, D.: Perception of emotion from dynamic point-light displays represented in dance. Perception 25 (1996) 727-738
17. Cutting, J.E., Kozlowski, L.T.: Recognizing friends by their walk: Gait perception without familiarity cues. Bulletin of the Psychonomic Society 9 (1977) 353-356