IEEE International Workshop on
Analysis and Modeling of Faces and Gestures, ICCV 2003

Inference of Human Postures by Classification of 3D Human Body Shape

Isaac COHEN, Hongxia LI
Institute for Robotics and Intelligent Systems
Integrated Media Systems Center
University of Southern California
icohen@usc.edu

Abstract

In this paper we describe an approach for inferring the body posture using a 3D visual-hull constructed from a set of silhouettes. We introduce an appearance-based, view-independent, 3D shape description for classifying and identifying human posture using a support vector machine. The proposed global shape description is invariant to rotation, scale and translation and varies continuously with 3D shape variations. This shape representation is used for training a support vector machine allowing the characterization of human body postures from the computed visual hull. The main advantage of the shape description is its ability to capture human shape variation allowing the identification of body postures across multiple people. The proposed method is illustrated on a set of video streams of body postures captured by four synchronous cameras.

1. Introduction

Multimodal interaction systems represent a considerable shift from classical windows, icons, menus and pointing (WIMP) interfaces. Gesture and speech represent the main component of such interface as they correspond to the foundation of natural human communication. While speech recognition systems are commercially available, gesture recognition is still in its infancy. This is partially due to the fact that speech modality is linear and very structured while gesture is a spatial modality that is still challenging to capture and interpret.

Human body motion tracking and analysis has received a significant amount of attention in the computer vision research community in the past decade. This has been motivated by the ambitious goal of achieving a vision-based perceptual user interface in which the state and the action of the user(s) are automatically inferred from a set of video cameras. The objective is to extend the current mouse-keyboard interaction techniques in order to allow the user to behave naturally in an immersed environment, as the system perceives and responds appropriately to user actions. Understanding human action in an environment is a challenging task as it involves different granularity in its analysis and description according to the targeted application. For example, describing a human activity in term of its trajectory constitutes a first level of representation, which may be satisfactory for surveillance applications but remains quite insufficient for understanding human gesture in an interactive environment. Indeed, in such situations richer descriptions are required in order to understand the human activity and recognize the performed gestures. In this paper we will focus on the capture and the description of human body 3D shape for the identification of human body posture. We believe that posture recognition is a first step towards gesture recognition. Indeed gestures can be decomposed into a set of "basic" postures that characterize temporal evolution of the performed gesture.

1.1. Previous Work

Various methods have been proposed for the estimation and analysis of full-body structure (see [8] and references therein). The objective is to develop real-time interactive systems with more sophisticated 2D and 3D tracking and representations [11][17]. Understanding the human motion from a monocular image sequence is challenging since only the 2D projection of these arbitrary motions is captured. Recently several researchers focused on the inference of 3D body model from a monocular camera using a human body model [6][15] or temporal templates [25]. The main drawback of these techniques is that roughly one third of the degrees of freedom of the human model are nearly unobservable due to motion ambiguities and self-occlusion. Multiple views are therefore required to disambiguate or identify the human motion.

Several approaches have been proposed for estimating human postures in the 3D case. These approaches rely on two to an array of cameras to capture the human shape and motion [7][12] or use 3D body scanners [19]. The body postures are then characterized through the use of shape descriptors or by characterizing body joints configurations. While the use of an articulated body model provides accurate measurements of body joints configurations, it requires intensive computing and state of the art techniques still lack robustness and accuracy for rapid hand motion.

1.2. Outline of the Proposed Approach

We present in this paper a shape posture identification technique based on the classification of the human body shape. The human body 3D shape inferred from a set of silhouettes and the corresponding visual hull is characterized by a shape description defined by a distribution. We introduce in this paper a 3D shape description allowing identifying various body postures. The shape description accounts for variability in people's body proportions and provides invariance to translation, rotation and scale. Continuity properties are also satisfied providing a robust shape descriptor that exhibits localized variation of the distribution for localized 3D shape variation. Moreover, the 3D shape descriptor we propose can selectively encode privileged axis of symmetry or desired rotation invariance. These properties are important for human posture identification, since the human body possesses a symmetry axis.

Identifying body postures from its 3D shape is challenging as the 3D description of the shape has to account for shape variability in characterizing a posture. Indeed, several people will perform similar posture differently and therefore identifying a posture from the 2D/3D shape descriptions will require a learning step. We present an appearance-based, learning formalism that is view point independent and uses a 3D shape descriptor of the visual-hull for classifying and identifying human posture. The proposed method does not require an articulated body model fitted onto the reconstructed 3D geometry of the human body. In fact, it complements the articulated body model since we can define a mapping between the observed shape and the learned descriptions for inferring the parameters of the articulated body model. In the following section we will present the shape descriptor considered and the learning algorithm based on Support Vector Machine (SVM).Our approach is based on the integration of 2D silhouettes captured by two or more cameras and the description of the human body shape using a 3D shape descriptor generated from the visual hull of the human body. An overview of the proposed approach is given in Figure 1.

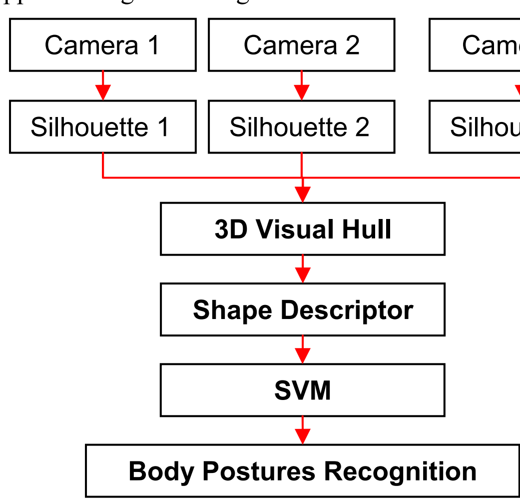


Figure 1: Overview of the proposed approach.

The paper is organized as follows; section 2 describes the 2D shape descriptor and its generalization to 3D. Section 3 presents the body postures classification and identification based on SVM approach. Section 4 exposes the experimental side of this work. It describes the definition of the model postures and illustrates the classification results. The paper is concluded by discussing the results, potential improvements and future work.

2. Human Body Shape Description

Shape descriptors have been well studied in various fields as they are used for determining the similarity between two shapes. The derived descriptors can be classified in terms of the shapes they characterize i.e. 2D contours, 3D surfaces, 3D volumes... For example, bending energy functions, spin images [10], shape context descriptors [1]. These descriptors were mainly used for shape matching and therefore focused on characterizing the local properties of the shape. Global models, assume a description of the objects into a set of features or parts segment. Common descriptions rely on parametric models [16], deformable regions [2][3], shock graphs [14] or wavelet decomposition [19] and spherical harmonics [22]. Shape similarity is then measured by comparing location of features and their spatial distributions. The performances of these approaches depend on the difficult task of segmenting the shape into its corresponding parts. These techniques perform well in the case of shapes of fixed configurations and are not suitable for modeling variability in the observed shapes such as the 3D visual hull of a gesturing person. Finally, a third description approach is based on modeling the geometric distribution of the shape properties such as histograms of angles [9] and distances between pair of arbitrarily sampled points [23], ... These descriptions are histogram based and do not perform well as the localization of the features is lost in the statistical representation used.

We present a statistical shape description model that preserves the localization of the geometric features considered. This global representation allows a robust description of shape that accommodates for variation of the shape. Indeed, as one would expect, small shape variation should induce a small change in the object description. Moreover, this variation is localized and does not interfere with the global representation of the object. These properties of the proposed shape description are crucial for efficiently representing the human shape and its variations.

The proposed shape descriptor describes a 3D polygon with regard to a reference shape. The reference shape considered characterizes the properties of the descriptor. Using a circle as reference for a 2D shape will guarantee rotation invariance and similarly a sphere for a 3D shape. This allows to selectively choosing the properties of the descriptor according to the desired application. The use of a cylinder as reference shape will guarantee rotation invariance only along its main axis and will enforce an axial symmetry. These properties are particularly interesting for human body shape descriptors. Indeed a human body shape has such properties.

Our shape description approach is based on the 3D visual hull reconstructed from the detected 2D silhouettes. Let's first describe the proposed description scheme in 2D and then generalize it to 3D visual hulls.

2.1. 2D Shape Description

Given a 2D silhouette of an object we compute a reference circle C_R defined by the centroid of the silhouette and its main axis. This circle is uniformly sampled into a set of control points P_i . We then consider a polar encoding of the projection of the silhouette onto the set of points $\{P_i\}$. For every point Q_j of the silhouette we compute the polar encoding of P_iQ_j, defined by $Q_j - P_i = r(\cos\theta, \sin\theta)$. For each control point P_i we infer a binned polar distribution where in each bin (r_k, θ_l). We then store the number of silhouette points $N_i(k,l)$ projected onto that bin. In Figure 2 we show the contribution of a single point P_i to the shape descriptor. The geometry of the bin is depicted in hashed green lines. The 2D shape description is obtained by adding and normalizing the set of descriptions derived from each point P_i:

$$N(k,l) = \frac{\sum_i N_i(k,l)}{\underset{k,l}{Max}(\sum_i N_i(k,l))}$$

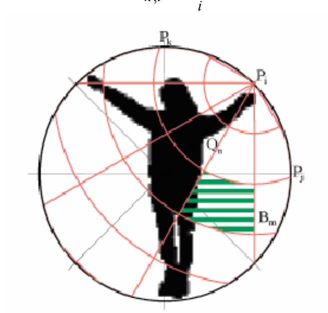


Figure 2: Illustration of the global shape descriptor of a 2D silhouette

The shape description derived here is independent of the scale of the silhouette as the description is normalized by the radius of the considered reference circle C_R . The translation invariance is obvious while the rotation invariance is also guaranteed. Indeed, rotating a silhouette in the image plane implies a cyclic permutation of the points P_i leaving the signature unchanged. In Figure 3 we show an example of such representation where the signature function, in polar coordinates was centered on the centroid and rendered in Cartesian coordinates. Note that this only a visualization of the representation map. The real shape signature is in the polar coordinate system.

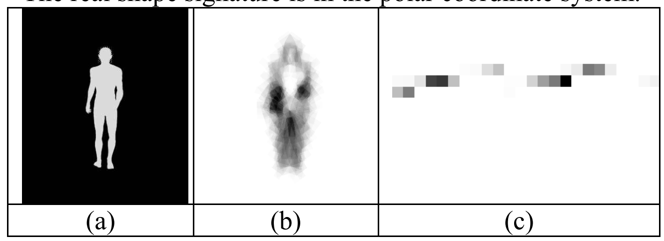


Figure 3: Example of a 2D global shape descriptor of a walking person. (a) the 2D silhouette. (b) 2D shape descriptor reprojected on the silhouette. (c) 2D shape descriptor in polar coordinates.

2.2. 3D Shape Description

The 2D shape description we described previously can be generalized to 3D without losing its properties (ie. scale, rotation and translation invariance). In the 3D case, instead of considering a single human body silhouette, a set of silhouettes are acquired synchronously. The 3D shape description is derived in two steps:

1. Construct the triangulated visual hull surface from the set of silhouettes.
2. Derive the 3D shape description of the human body based on the triangulated surface representation of the visual hull.

2.2.1. 3D human body visual-hull reconstruction

Integrating multiple silhouettes acquired simultaneously from different view points allows generating a 3D visual-hull of the human body. The visual-hull of an object is the closest approximation of the 3D object which can be obtained from the detected 2D silhouettes [13]. Assume that the person is viewed from a set of cameras. Each silhouette defines a cone characterized by all the rays from the camera origin to the points on the silhouette. The intersection of the cones generated by the multiple cameras defines the visual hull of the object.

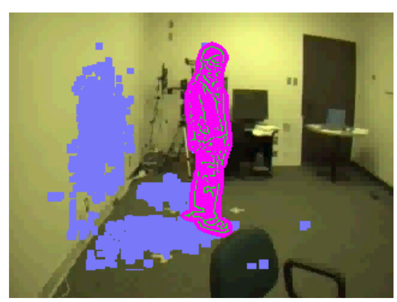


Figure 4: The body silhouette detection. The blue regions correspond to the detected and removed shadow, green line corresponds to the detected edges and the pink region represents the final human silhouette

Human silhouettes are extracted by calculating a Gaussian-based model of the static scene and then comparing each new observation to the background model distribution. Such approach detects pixels where a motion was observed, as well as the shadows and reflections. However, in indoor environments the shadows cast by a diffuse light do not have strong boundaries. Therefore, combining edge properties to color variations allows us to remove the pixels belonging to cast shadows and segment accurately the foreground pixels. In Figure 4 we show an example of such detection approach.

The integration of the detected silhouettes provides a 3D representation of the human body. A polyhedral representation of the detected silhouettes and their integration will provide a polyhedral approximation of the visual hull. The visual hull is computed using a polyhedral representation for the visual hull directly from the detected silhouettes [20]. If the number of silhouettes considered is large the visual-hull provides a good approximation of the 3D shape. This polygonal approximation of the shape can be computed in real time and also converted into a triangular description usable by a triangular processing framework. In Figure 5 we show an example of a 3D visual hull computed from three views.

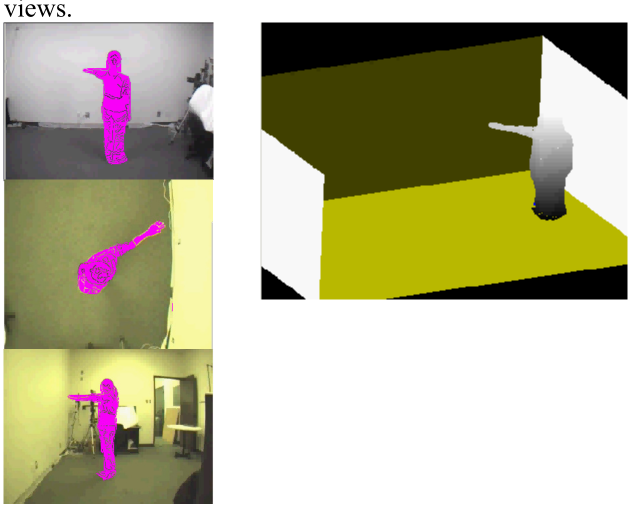


Figure 5: 3D visual-hull reconstructed from the 3 silhouettes shown on the left

2.2.2. 3D Shape Description

The generalization of the 2D shape descriptor to 3D is performed by defining a reference shape containing the visual hull surface and measuring the contribution of each triangle of the visual hull to the shape description inferred from a set of points lying on the surface of reference. The selection of the surface of reference allows choosing the properties of the descriptor according to the desired application. The use of a cylinder as reference shape will guarantee rotation invariance only along its main axis and will enforce an axial symmetry while a sphere will provide 3D rotation invariance and point symmetry. These properties are particularly interesting as they allow deriving an application dependent 3D shape description. For human body posture recognition, the use of a cylinder as reference shape enforces body axial symmetry while for a generic 3D object description, a sphere will be more appropriate.

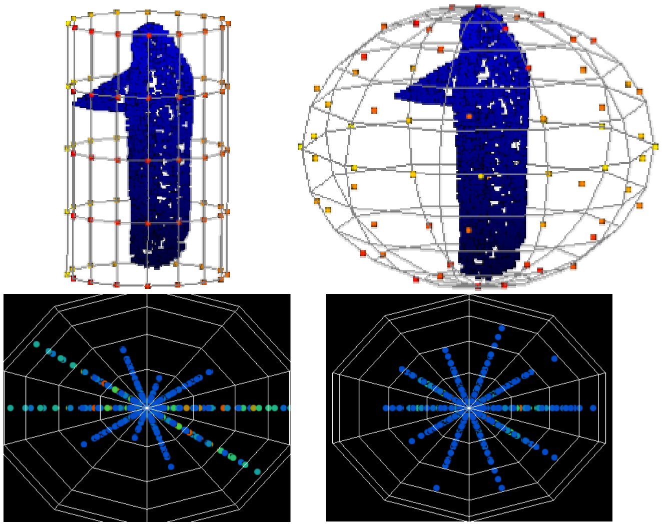
The computation of the 3D shape description is similar to the 2D case. The main difference between the 2D and 3D is the representation of the human body. In 2D, it is represented by all the points on the silhouette. But in 3D, the representation is based on the 3D triangular visual hull surface constructed from a set of silhouettes. In order to use the shape description method we proposed for the 2D case, we need to sample the visual hull surface into a set of points, which are dense enough to contain all the information of the visual hull. Because the size of the triangular visual hull surface is not uniform and depends on the polygonal approximation of the silhouettes, we need more sample points on the surface other than the triangle vertices. One could refine the triangular description by subdividing the 3D visual hull. This will increase dramatically the algorithmic complexity of the algorithm and create redundant mesh description. Instead, we have chosen to sample uniformly the triangles of the visual hull and encode the relationships between the sampled points and the reference surface considered.

Given a set of points $\{Q_j, j = 0..m\}$ corresponding to a sampling of the visual hull and a set of control points $\{P_i, i = 0..n\}$ sampled uniformly on the surface of the reference shape S_R. We compute, for each control point P_i, the distribution of 3D spherical projections of the set of points Q_j. The spherical encoding of P_iQ_j, defined by $Q_j - P_i = (r, \theta, \varphi)$. For each point P_i we then infer a binned spherical distribution where in each bin (r, θ, φ) we store the number of 3D points Q_j projected onto that bin. The 3D shape description is obtained by adding and normalizing the set of descriptions derived from each control point P_i.

In Figure 6 we show the spherical shape distribution obtained by encoding a 3D visual hull with regard to a cylinder and a sphere. The shape distribution varies considerably as different shape properties are captured by the two reference surfaces.

2.3. Similarity Properties

The purpose of defining a shape description is the ability to characterize surfaces' local and global similarities, as well as comparing various 3D surfaces. The shape descriptor of a surface is defined by a distribution in the spherical coordinate space. Comparing shapes is therefore reduced to comparing the corresponding distributions.



(a) (b)

Figure 6: 3D visual hull and the shape distribution of the surface using a cylinder (a) and a sphere (b) as reference surface.

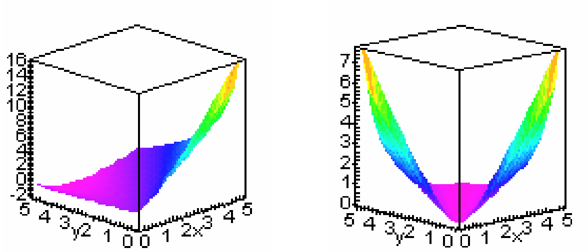
Measuring the similarity among distribution can be performed by various methods defining distance functions. We chose to characterize similarity between shape distributions by measuring the relative entropy of the distributions. Given two distributions f and g, the relative entropy, also called the Kullback-Leibler distance, is defined by:

$$d(f,g) = \int f \log \frac{f}{g}$$

Even though this distance does not satisfy the triangle inequality (it is not a true metric) it has many properties of distance functions and is equals to zero only if $f = g$. In order to guarantee symmetry of the similarity measure between two distributions we have derived the following distance:

$$D(f,g) = \frac{1}{2}\left(d(f,g) + d(g,f)\right) = \frac{1}{2}\int (f - g)\log \frac{f}{g}$$

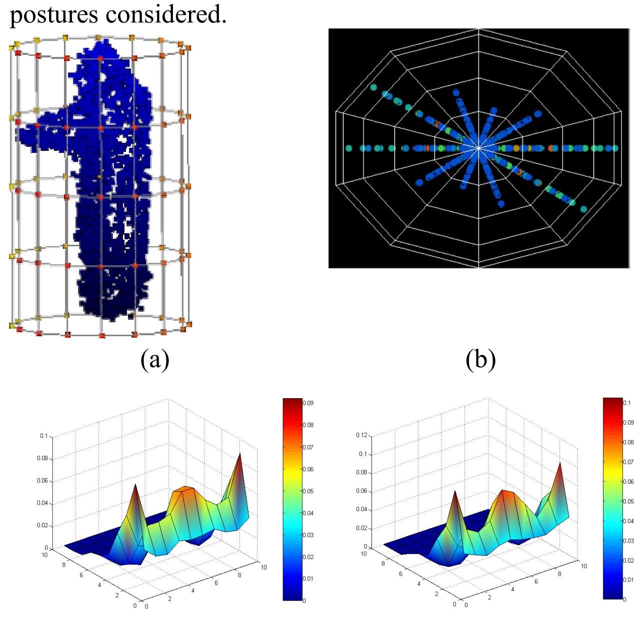
This new distance function illustrated in Figure 7 is used in the remaining of the paper for measuring the similarity between two surfaces.



(a) (b)

Figure 7: Graphs of the distance functions considered. (a) Kullback-Leibler distance and (b) its symmetric counterpart.

We have used the above distance function to show that the global descriptor is invariant to scale, translation, and rotation. Also small variations of the shape of the human body create small localized variations of the shape description (continuity). The robustness of the description with regard to noise in the computed visual hull is illustrated in Figure 8. We have added a Gaussian noise to the vertices of the triangles of the visual hull depicted in Figure 6.a and measured the similarity of the shapes. The noisy visual hull and the corresponding shape distribution are illustrated in Figure 8. A comparison of the similarity between the original visual hull and its noisy counterpart yielded a distance value less than 10^{-2} for the 12 body postures considered.

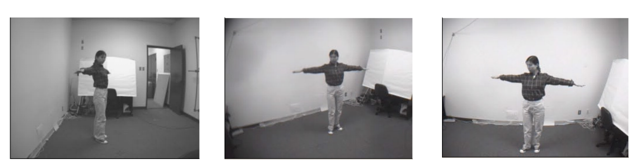


(a) (b)

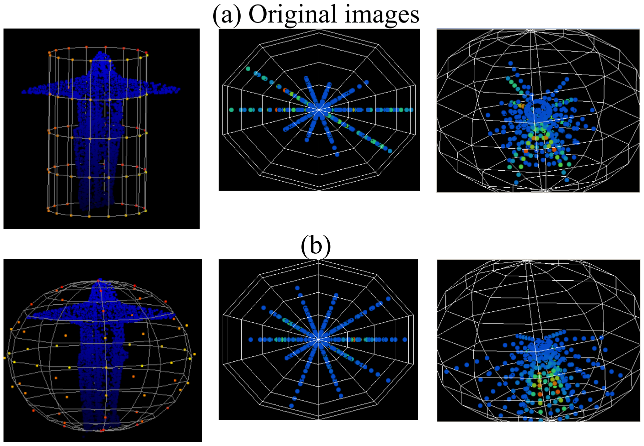
(c) (d)

Figure 8: Stability of the shape descriptor with regard to noisy information. (a) noisy visual hull ($\sigma = 2$) and its shape descriptor (b). (c) and (d) display the polar spherical distributions for a selected radius (r=0.1) of the noisy visual hull depicted in (a) and the original one illustrated in Figure 6.a.

In Figture.9, we show the 3D descriptor for another body posture. It depicts the variability of the shape signature according to the selected reference shape and its ability to capture body shape properties. In the remaining of the paper, we have considered the cylinder as a reference shape for deriving the shape descriptors of the computed visual hull. This permits to enforce the axial symmetry of the body shape.



(a) Original images



(c)

Figure 9: Example of shape descriptors derived using a cylinder and a sphere as reference shape. In (b) and (c) we display the visual hull and the spherical shape descriptor viewed from above and the side.

3. Human Body Posture Inference

Deriving the human posture from its silhouettes in 2D or from the reconstructed shape in 3D is a challenging task as it requires taking into account posture variability across people. A method commonly used relies on the articulated body model in order to infer the human posture. The recovery of an articulated body model still requires the interpretation of the 45plus-degree of freedom in order to infer the human posture. This interpretation has to take into account posture variability and errors in the estimation of the articulated model in order to perform an efficient analysis of the 45plus-D parameter space. In the following sections we show that the global 3D shape descriptor introduced in this paper can be used for human body posture inference. We use a Support Vector Machine formalism [18] to train and classify the set of heterogeneous information provided by the 3D shape-based descriptor. The main advantage of using a SVM is its ability to compress the information contained in the training set, since only support vectors are required for the classification. This allows us to reach near real-time performances.

3.1. SVM-based Classification

The main issues in using a machine learning approach are the selection of the features used as training data set and the choice of the data set for training the model. While an articulated body model provides a natural set of features (joints and limbs) to consider for training purposes, it is time consuming and it is difficult to acquire the 45plus degrees of freedom of the selected model. Conversely, 3D shape based descriptors are very easy to collect but a correct representation of the shape has to be selected in order to be significant for learning.

The problem we are addressing here is the definition of a decision function that from a set of observations $x \in X = \{x_i, i = 0 \ldots N\}$ and the corresponding labels $y \in Y = \{y_i, i = 0 \ldots M\}$ will make accurate classification of unseen values of x. A very successful approach for solving this supervised learning problem is the support vector machine (SVM) [18]. In this work we are interested in a classification of the observed human postures; therefore the set of available labels is limited to $Y = \{-1,1\}$ representing respectively non-posture and posture descriptions. The decision function is defined by the SVM is:

$$f(x) = \mathrm{sgn}\left(\sum_{i=0}^{l-1} \alpha_i^0 y_i K(x_i, x) + b\right)$$

where the coefficients α_i^0 are obtained by maximizing the functional:

$$W(\alpha) = \sum_{i=0}^{l-1} \alpha_i - \frac{1}{2} \sum_{i,j=0}^{l-1} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

under the constraints:

$\sum_{i=0}^{l-1} \alpha_i y_i = 0$ and $\alpha_i \geq 0$.

The coefficients α_i^0 define a maximal margin hyper-plan in the high dimensional feature space where the data are mapped through the non linear function ϕ such that $\phi(x_i) \bullet \phi(x_j) = K(x_i, x_j)$. Various kernels K are commonly used (linear, exponential, polynomial...) we will use a linear kernel K using therefore a linear mapping between the feature space (posture we defined) and the representation space (shape description).

3.2. Training and Classification

3.2.1. Selection of Body Postures

We have defined a set of 12 postures that need to be identified by the system. The selected postures are shown in Figure 10. The selected postures focus on hand gestures and were chosen in order to build a representative set of postures likely to be observed as people interact with a system using hand gestures.

Figure 10: The set of 12 postures we defined in our system.

3.2.2. Posture training and classification

The SVM is trained using 3D shape descriptors defined in the previous section. The shape descriptor is inferred from the 3D visual hull obtained by integrating multi-view silhouettes acquired by 4 synchronous cameras. The shape descriptor is presented by a vector:

S= {(index of the bin, density of points in the bin)}.

The visual-hull corresponding to the detected silhouettes and their shape descriptors are computed at a frame rate of one frame per second. The system was trained on the chosen 12 postures by considering approximately 2000 samples per posture.

4. Experimental Results

We have used the proposed 3D shape description technique for identifying user's postures while performing specific gestures. The experimental settings contains four synchronized cameras allowing to extract in real time the human body silhouettes and infer the 3D visual hull at one frame per second. In Figure 11 we show the output of the system.

In our experiments, the number of control points used to generate the shape descriptor is 5×16. Less control points cannot accommodate all rotation variation, while more control points do not significantly improve the performances but require more computing. The selection number of bins depends on the complexity of the predefined postures. In our case we have used the following: $10(r)\times10(\theta)\times10(\psi)$.

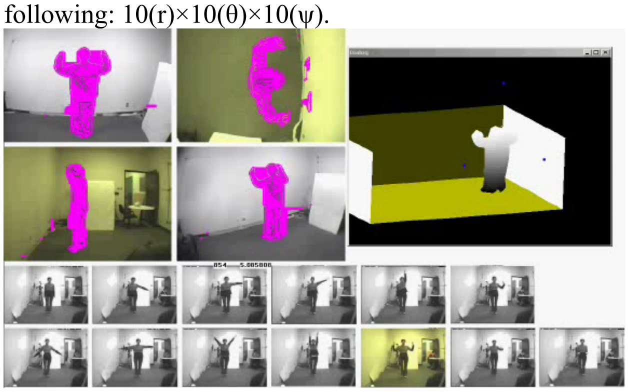


Figure 11: Illustration of the system's output. The thumbnail of the recognized posture is highlighted.

The detected silhouettes and the corresponding visual hull are displayed, while at the bottom, the system highlights the image thumbnail corresponding to the recognized posture. The performed postures are recognized by the system even though the visual hull is the detected silhouettes are corrupted by reflections on the wall. Moreover, the position and orientation of the considered persons are different from the one used for training the classifier. These results show the capability of the system to handle people that were not used for training and handle variations in body proportions and the person's pose while performing a specific gesture.

4.1. Identification Rate

We have evaluated the performances of the system for several people. None of the persons tested were considered in the training phase. In fact the SVM was trained using postures from one single person. Therefore, we expect the recognition rates to be improved as we broaden the set of people considered for training the SVM-based classifier. In table 1 we show the recognition rates obtained on 20 video sequences (containing each about 2000 frames) of 4 different persons. For each posture the rates displayed correspond to averages of the obtained recognition rates.

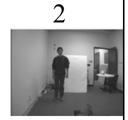
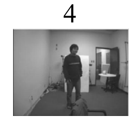
Person / Posture	**1** 	2 	3 	4 
0	**99.8**	98	98.3	98.1
1	**95.4**	96.5	96.0	94.4
2	**96.2**	99	86	83.0
3	**93.9**	98.6	95.4	92.1
4	**85.5**	90.1	87.3	85.7
5	**89.6**	86	88.6	97.3
6	**96.6**	92.8	89.2	89.0
7	**88.6**	94.2	86.0	84.5
8	**98**	97.3	95.2	94.2
9	**99**	91.7	96.4	91.7
10	**87.5**	82	90.8	81.3
11	**96.3**	94.2	80.0	92.5.

Table 1. Identification rates of the 12 postures for 4 different persons. Only person 1 was used for training the SVM.

5. Conclusion

Identifying user postures as a first step towards gesture recognition is a challenging task. The challenge here was to define a 3D shape description that allows a robust characterization across users without requiring a specific training for each person. The presented experimental results illustrated the ability of the system to recognize a variety of human body posture. We have started investigating the characterization of basic gestures or gestemes as a transition states models of some canonical body postures. Our first experimental characterizations of the temporal transitions are very encouraging and indicate a strong temporal structure in gesture inference.

Acknowledgement

The research has been funded in part by the Integrated Media Systems Center, a National Science Foundation Engineering Research Center, and Cooperative Agreement No. EEC-9529152.

6. References

[1] S. Belongie, J. Malik, and J. Puzicha, "Matching shapes.", In IEEE Proc. of the International Conference on Computer Vision, Vancouver, Canada, July 2001.

[2] I. Cohen, N. Ayache, and P. Sulger, "Tracking points on deformable objects using curvature information", Proc. of the Second European Conference on Computer Vision, Italy, 1992.

[3] I. Cohen and I. Herlin, "Curves matching using geodesic paths", IEEE Proc. of Computer Vision and Pattern Recognition, Santa Barbara, June 1998.

[4] I. Cohen and M.W. Lee, "3D body reconstruction for immersive interaction", Springer Verlag, editor, 2nd International Workshop on Articulated Motion and Deformable Objects, 2002.

[5] M.W. Lee. I. Cohen and S.K. Jung, "Particle filter with analytical inference for human body tracking. In IEEE Workshop on Motion and Video Computing, Orlando, Florida, 2002.

[6] D. DiFranco, T. Cham, and J. Rehg, "Reconstruction of 3d figure motion from 2d correspondences", In IEEE Proc. of Computer Vision and Pattern Recognition, December 2001.

[7] S. Iwasawa et al, "Human body postures from trinocular camera images", In International Conference on Automatic Face and Gesture Recognition, pages 326–331, 2000.

[8] A. Hilton and P. Fua., "Modeling people toward vision-based understanding of a person's shape, appearance, and movement", Computer Vision and Image Understanding, 81(3):227–230, 2001.

[9] K. Ikeuchi, T. Shakunaga, M. Wheeler, and T. Yamazaki, "Invariant histograms and deformable template matching for sar target recognition", In IEEE Proc. of Computer Vision and Pattern Recognition, 1996.

[10] A. E. Johnson and M. Hebert, "Using spin-images for efficient multiple model recognition in cluttered 3-D scenes", IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(5):433–449, 1999.

[11] I.A. Kakadiaris and D.Metaxas, "Three-dimensional human body model acquisition from multiple views", International Journal of Computer Vision, 30(3):227–230, 1998.

[12] T. Kanade, H. Saito, and S. Vedula, "The 3D room: Digitizing time-varying 3D events by synchronized multiple video streams", Technical report, CMU-RI, 1998.

[13] A. Laurentini, "The visual hull concept for silhouette-based image understanding", IEEE Trans. on Pattern Analysis and Machine Intelligence, 16(2):150–162, 94.

[14] K. Siddiqi, A. Shokoufandeh, S. J. Dickinson, and S. W. Zucker, "Shock graphs and shape matching", Computer Vision, pages 222–229, 1998.

[15] C. Sminchisescu and B. Triggs, "Covariance scaled sampling for monocular 3d body tracking", In IEEE Proc. of Computer Vision and Pattern Recognition, December 2001.

[16] F. Solina and R. Bajcsy, "Recovery of parametric models from range images: The case for superquadrics with global deformations", IEEE Transactions on Pattern Analysis and Machine Intelligence, 1990.

[17] M. Turk and G. Robertson, "Perceptual user interfaces", Communications of the ACM, March 2000.

[18] V.N. Vapnik, "Statistical Learning Theory", Wiley, New York, 1998.

[19] N. Werghi and Y. Xiao, "Recognition of human body posture from a cloud of 3d points using wavelet transform coefficients", In Proc. Of the fifth IEEE International Conference on Automatic face and gesture recognition, 2002.

[20] W. Matusik, C. Buehler, L. McMillan. "Polyhedral Visual Hulls for Real-Time Rendering", Proceedings of Eurographics Workshop on Rendering 2001.

[21] R.Bakis, "Coarticulation modeling with continuous-state HMMs", In Proc. IEEE Workshop Automatic Speech Recognition, pages 20-21, Arden House, New York, 1991.

[22] T. Frankhouser, P. Min, M. Kazhdan and J. Chen. "A Search Engine for 3D Models" ACM Transactions on Graphics, 22, 1, Jan. 2003, pages 83-105.

[23] R. Osada, T. Frankhouser, B. Chazelle and D. Dobkin. "Shape Distributions". ACM Transactions on Graphics, 21, 4, Oct. 2002, pages 807-832.

[24] R. Rosales and S. Sclaroff , "Inferring Body Pose without Tracking Body Parts", In proc. IEEE Computer Vision and Pattern Recognition, June 2000.

[25] J. W. Davis and A. F. Bobick, "The Representation and Recognition of Action Using Temporal Templates", IEEE Conference on Computer Vision and Pattern Recognition, 1997.