Wavelet Moments for Recognizing Human Body Posture from 3D Scans

Naoufel Werghi Yijun Xiao
Department of Computing Sciences
University of Glasgow
Glasgow G12 8QQ
naoufelw@dcs.gla.ac.uk

Abstract

This paper addresses the problem of recognizing a human body (HB) posture from a cloud of 3D points acquired by a Human body scanner. It suggests the wavelet transform coefficients (WTC) as 3D shape descriptors of the HB posture. The WTC showed to have a high discrimination power between posture classes. Integrated within a Bayesian classification framework and compared with other standard moments, the WTC showed great capabilities in discriminating between close postures. The qualities of the WTC features were also reflected on its classification rate, ranked first when compared with other 3D features.

1. Introduction

Automatic identification of the HB posture from a 3D scan data, is of interest for many applications exploiting HB scanner technology. In Anthropological and medical applications, there is a need to segment the body scan data into areas corresponding to the functional parts of the body. In entertainment applications, HB scan are fitted to a generic model to produce realistic avatars of virtual actors that can be integrated in movies or video games. The approaches developed so far [1, 2, 3, 4] present a severe limitation because they assume a standard posture of the HB(Figure 1.b). Providing a tool for recognizing the posture will allow to tackle the segmentation and the model fitting problems in the general case of arbitrary postures.

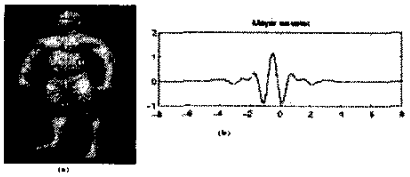

Figure 1. (a): The standard posture of a human body in a reference (x, y, z) attached to the scanner. A rotation of the whole HB is constrained to be around the z axis, affecting only the angle ϕ. (b):Meyer wavelet function.

2. 3D Shape descriptors

The posture data , is a set scattered 3D points representing the HB surface. Most HB scanners provide a complete data that covers a large area the body. Global features, particularly moments are suitable for that case. The moment features have been extensively used in image analysis and description [5, 6, 7, 8] However, less work has been done in the 3D case. One of the reasons, is that most of the 3D Imaging devices do not provide complete data in terms of surface covering, There have been, nevertheless,some attempts to define frameworks for 3D moments construction. Sadjadi *et al* [9] pioneered the development of 3D Geometric moment invariants. They built a family of three invariant moments with a degree up to the second-order. Using the notion of complex moments Lo *et al* [10] constructed a family of twelve invariant moments with orders up to the third degree. In these last works, moments were used mainly to estimate 3D transformations and their performances were not evaluated for classification tasks. Also, being not derived from a family of orthogonal functions, these moments were subject to correlation. Canterakis [11] extended Zernike moments to the 3D case, but their performances were not put into trial yet. In this work we propose to investigate the wavelet transform coefficients (WTC) . These features have been tested successfully as 2D shape descriptors [8]. We will show that the WTC can be utilized for 3D shape description. Derived from an orthogonal family of functions, these features are also less redundant and more informative than non-orthogonal family of moments [9, 10].

3 The WTC descriptors

The Wavelet Transform was introduced by Morlet [12] as a time-scale analysis tool for non-stationary signals. Then further developed by many authors [13, 14, 15]. The Wavelet Transform of a function $f(r)$ at the scale a ($a > 0$) and the shift b is: $C_{ab} = \int_{-\infty}^{\infty} f(r)\frac{1}{a}\psi(\frac{r-b}{a})dr$ where $\psi_{ab} = \frac{1}{a}\psi(\frac{r-b}{a})$ represent a family of functions derived from the mother wavelet $\psi(r)$. This function is characterized by

1051-4651/02 $17.00 © 2002 IEEE

a compact support both in the space and the frequency domain. The WTC embodies information about the spectrum of the frequency of $f(r)$ around the position b at a scale a. The coefficients C_{ab} constitute the set of projections of the function $f(r)$ on space spanned by the basis ψ_{ab}.

Let consider $f(r,\theta,\phi)$ a 3D binary representation for the cloud of 3D data points in the spherical coordinates. In its discrete form, $f(r,\theta,\phi)$ is a spherical voxel representation of the points. Consider a sphere of radius r, the points distribution at the sphere surface can be described by the spherical harmonics via the transformation: $F_{mn}(r) = \int_0^{2\pi}\int_0^{\pi} f(r,\theta,\phi)U_{m,n}(\theta,\phi)r^2 sin\theta d\theta d\phi, \quad 0 \leq m \leq n$ where $U_{m,n}$ are the spherical harmonics of order m and n defined on the unit sphere. they form an orthogonal family [16], expressed by $U_{m,n} = e^{jm\phi}V_n(\theta)$ where $V_n(\theta)$ is a polynomial function of order n in $cos\theta$ and $sin\theta$. $F_{mn}(r)$ define therefore a sort of moments that describe the distribution of points on the spherical surface of radius r. We considered the first four spherical harmonic functions namely, $U_{0,0} = 1$, $U_{0,1} = cos\theta$, $U_{1,1} = e^{j\phi}sin\theta$, $U_{1,2} = -3e^{j\phi}sin\theta cos\theta$.

Now what remains is to describe the variation of these moments in function of r to obtain a 3D description of the posture. This description should infers a multi-scale aspect since the differences in the posture distributions manifests at different scales. This can be seen, by examining for instance, the pairs of postures $(2,18)$ and $(6,8)$ in Figure 2. For the first pair, difference in data point distribution covers more than the half of the posture space, whereas it is limited to the space around the right arm for the second pair. Being a multi-scale operator, the wavelet transform satisfies this requirement. We define therefore the WTC of the moment functions $F_{mn}(r)$ as $C_{ab}^{mn} = \int_0^{\infty} F_{mn}(r)\psi_{a,b}(r)dr$ The mother wavelet function we utilized is Meyer's wavelet [15] (Figure 1.(a), it is highly regular and generates an orthogonal basis of function ψ_{ab}. We precise that the WTC descriptor used, is the module of C_{ab}^{mn}.

The invariance of the WTC with respect to translation and scale is obtained by preprocessing the data in the Cartesian space before passing to the spherical space. From the cloud of 3D data points a Cartesian voxel grid if formed. Then the origin of the voxel grid is shifted to the centre of mass of the data points. The scale invariance is obtained by affecting the 3D points' coordinates so that the data volume defined by the moment $m_{000} = \sum_x \sum_y \sum_z f(x,y,z)$ is equal to V_0, where V_0 is a predetermined value. The rotation of the whole HB within the scanner has only one degree of freedom that affects only ϕ. It can be shown easily that$\|C_{ab}^{mn}\|$ is invariant with respect to that rotation. The negative side is that pairs of symmetric postures have very close feature values. Such pairs have been associated to the same class and the ambiguity can be removed after the classification, by using simple geometric procedures.

A dyadic discretization is adopted for a and b, by choosing $a = (Scale)2^{-p}$, $p = 0,1,2,3$ and $b = qa/2$, $q = 0,1,...,2^{p+1}$, where $Scale$ is the radius of the sphere confining the data points. The scaling parameter a takes the values $Scale, Scale/2, Scale/4, Scale/8$. Values smaller than $Scale/8$ do not allow to extract significant information. The shifting parameter b is varied proportionally to the scale parameter within the range $[0, Scale]$. The number of (p,q) pairs is then equal to 34, which combined with 4 pairs (m,n) result in 136 WTC features C_{pq}^{mn}

4 Feature Selection and Classification

Since not all the features will contribute effectively in the classification. There is a need to select the most useful features that have a high discriminative power. The discriminative power is characterized by the interclass distance defined as metric for measuring the separation between two classes. A selection criterion based on that metric is therefore utilized in the search for the optimal set of features. This task was subject of intensive work in the literature [17]. There are mainly tow categories of techniques the first operates on feature vectors, the second treats each feature individually. We adopted a technique belonging to the second one, it is sub-optimal but quite efficient. The selection algorithm is as follows: Given a set of features $\{x_1, x_2, .., x_d\}$ and given a selection criterion J: 1) Compute the selection criterion value $J(k)$ for each feature x_k , 2) rank the features in descending order with respect to J, 3) select the top ranked features. The choice of the selection criterion is quite tight to the classification method in the sense that the interclass distance should be defined in the same framework of the classification scheme, so before setting the selection criterion let first examine the classification.

The classification problem is stated as follows: Given a set of posture classes $C_1,..C_N$ and given a query posture Q , find to which class the posture Q belongs ? The query posture is represented by an observation feature vector of dimension d, $X = [x_1, x_2, .., x_d]$. For each class C_i, consider the discriminative functions $d_i(X)$. The vector X is associated to the class C_i if $d_i(X) > d_j(X)$ for all$j \neq i$. The optimal discriminative function in the sense of Bayes, assuming $P(X|C_i)$ is a normal density $\mathcal{N}(\mu_i, \Sigma_i)$ and the the different classes have equal *priori* probability, can be brought to: $d_i(X) = -\frac{1}{2}(X-\mu_i)^T\Sigma_i^{-1}(X-\mu_i) - \frac{1}{2}ln|\Sigma_i|$. The statistics (μ_i, Σ_i) of class C_i are obtained from training process based using the standard EM technique [18]

Back to the selection criterion, the interclass distance between two classes C_i and C_j , having the conditional probability density functions $P(x_k, C_i) = \mathcal{N}(\mu_i^k, \sigma_i^k)$ and $P(x_k, C_j) = N(\mu_j^k, \sigma_j^k)$ with respect to the feature x_k, can be evaluated by the following probabilistic distance: $s_{ij}^k = \frac{1}{2}(\frac{\sigma_j^k}{\sigma_i^k} + \frac{\sigma_i^k}{\sigma_j^k} - 2) + \frac{1}{2}(\mu_i^k - \mu_j^k)^2(\frac{1}{(\sigma_i^k)^2} + \frac{1}{(\sigma_j^k)^2})$. This expression indicates that the larger the ratio of the means difference and

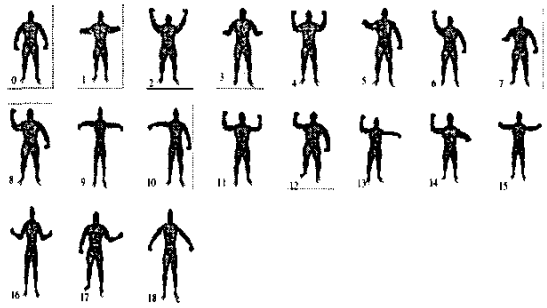

Figure 2. The posture models

the variances sum, the wider is the distance separating the two classes. The criterion that evaluates the discriminative power of the feature x_k is then the sum of the pairwise interclass distance: $J(k) = \sum_{i=1}^{d} \sum_{j=i+1}^{d} s_{ij}^{k}$. The larger the value of J, the better the feature x_k can discriminate between the classes.

This criterion is then used to rank the set of the 136 features $C_{p,q}^{m,n}$. the first three feature were $C_{1,1}^{0,1}$, $C_{1,2}^{1,1}$ and $C_{2,1}^{0,1}$. Because of the limited space we could not discuss in details the interpretation of the features rank. The main aspect is that the best WTC operate on the periphery of posture volume, area which is the most sensitive to posture changes caused by the arms' gesture in our models.

5. Experiments

The performance of the WTC features in terms of power discrimination and classification rate were assessed within a comparative study including the geometric moments developed by Lo *et al* [10] and the 3D Zernike moments [11]. The postures are generated from 3D Human body scan obtained from Cyberware cite in the Web [19]. The HB scan was then fitted to an articulated HB model. The orientations of the human body segments,set via the joint angles, define the parameters of the posture. By varying these parameters a variety of postures having a reasonable human shape can be obtained. Figure 2 shows the different posture models labeled from 0 to 18.

The statistic characteristics of each posture models are determined as follows, for each posture, 30 training data sets are generated, perturbing at each generation the posture parameters with a Gaussian noise and randomly rotating the full data in a direction that affects the ϕ coordinate. The mean and the variance of the model vectors are computed upon the 30 feature vectors associated to the training sets.

The discriminative power for each of the geometric moments, Zernike moments and the WTC is assessed by examining how well the best features of each category can discriminate between close postures. The best features for the geometric moments and Zernike moments were selected using the same scheme than for the WTC. From the posture models we selected pairs of postures which are close, namely $(0,7)$, $(2,11)$ and $(9,15)$ (Figure 2). Then for the 30 training samples of each posture in the pair, we plot the values of their corresponding best features. The distributions related to each posture could then be visually compared.

Figure 3.a shows that the first three WTC features, namely, $C_{1,1}^{0,1}$, $C_{1,2}^{1,1}$ and $C_{2,1}^{0,1}$ are clearly well separated for the three pairs. The separation of the two best Zernike moments $Z_{5,3}^{1}$ and $Z_{7,1}^{1}$ plotted in a same figure for each pair (Figure 3.b) is less clear although their distributions can be distinguished. A more mixed close distributions are noticed for the two other best Zernike moments. The distributions of the best geometric moment I_{22}^{2} in Figure 3.c, look very close, particularly for the two last pairs. The distributions of the two other best geometric moments have similar behavior. These results illustrate that the WTC features are more capable to distinguish between close postures than Zernike moments whereas the geometric moments are far less competitive. It is worth to mention that the three best WTC features were not selected specifically to discriminate between these particular close postures, since the selection process involved all the postures, yet these features still do separate them quite reasonably.

In the second part of the experiments, the performances of the WTC features, Zernike moments and geometrical moments are assessed by evaluating the rate of successful classifications. A set of query postures is matched with the posture models. Query postures were obtained with the same way than the posture models, that is a 30 randomly perturbed and rotated version for each artificially generated posture. The first experiment involved the three categories of features. The aim is to have a rough comparison between them. This test was carried out with the best four features of each category. The results are illustrated in Figure 3.(d). where the WTC has the best rating followed by Zernike moments then the geometric moments.

The other assessed aspect concerns how the classification rate evolves in function of the number of features. This gives an idea about the optimality of the selected set of features. In this experiment, only the WTC and the Zernike moments were assessed, as we decided not to carry with geometric moments based on the results of the previous experiment. This experiment used a query set of 30×19 samples. The experiment consist of many trials, in each one, the number of features involved in the classification is increased by one, starting by 5 features and ending by 35. The classification rate associated to the WTC and Zernike moments are mapped in Figure 3.(e). The Figure shows that the WTC have the best classification rate for all the numbers of features, with a maximum rate of 98% reached with 23 features. For Zernike moments the maximum rate of 94% is obtained with 28 features. Also we notice that with 11 WTC

can guarantee a classification rate of 95% whereas a lower rate of 93% needs 16 Zernike moments. Although there is an overall improvement of the classification rate as the number of features get increased, this improvement is not monotonous as some fluctuations appears from the 10th feature. The cause comes back to the feature selection process, where the used technique guarantee only a sub-optimal set of features. Also, after a certain number of features (25), the classification rate looks stagnating. Indicating that adding other features do not improve the classification.

6 Conclusion

The WTC features demonstrated very reasonable discriminative power compared to Zernike moments and geometric moments. This was reflected on classification rate of each category of features, where The WTC were top ranked whatever number of features is used. For some set of WTC features, the classification rate reached 98% whereas a larger number of Zernike moments is needed for the maximum rate of 94%. The database of the training samples can be enriched by adding a variety of human body shapes coming from different scan sources. Naturally the number of model postures we considered is far from being exhaustive. Many different postures can be added. This raises the question of what is the maximum number of different postures that could be successfully recognized. We believe that this is linked to what extent the recognition process could distinguish between two close postures and how to quantify the closeness of two postures. These issues will be examined in a future work.

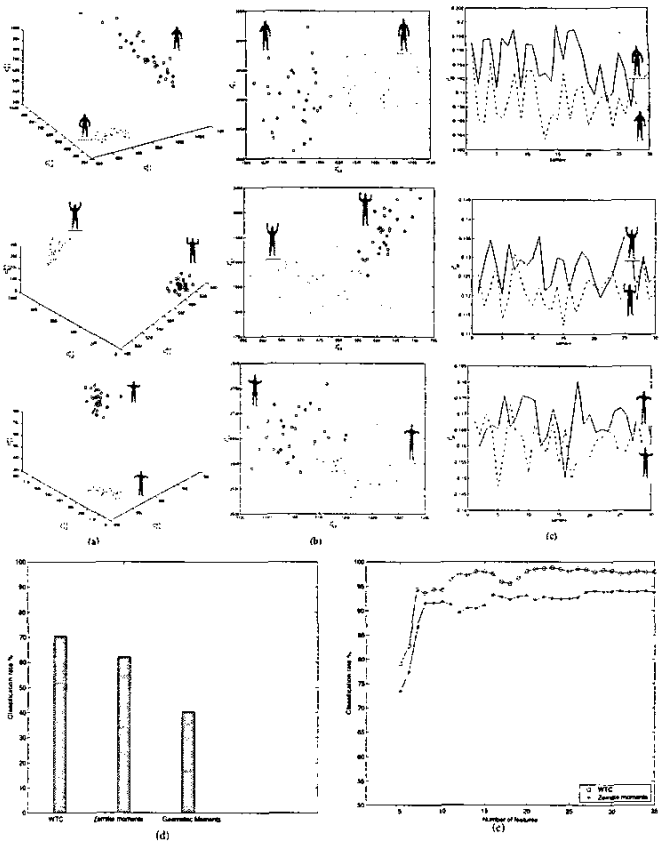

Figure 3. (a) variation of the three best WTC features $C_{1,1}^{0,1}$, $C_{1,2}^{1,1}$ and $C_{2,1}^{0,1}$. (b) variation of the two best Zernike moments $Z_{5,3}$ and $Z_{7,1}$. (c) variation of the geometric moment I_{22}^2. (d) classification rate associated to the 4 geometric moments, 4 best WTC and 4 best Zernike moments. (e) Classification rate of the WTC and the Zernike moments mapped in function of the number of features.

References

[1] J. Starck *et al*, "Human Shape Estimation in a Multi-Camera Studio". Proc. BMVC, pp.573-583, Manchester 2001.

[2] P. Jones, P.Li, K. Brook-Wavel, G. West, "Format of human body modelling from 3D body scanning", Int. Journal of Clothing Sciences, Vol.7, No.1, pp.7-16, 1995.

[3] R. Pargas *et al* , "Automatic measurement extraction for apparel for three-dimensional body scan", Journal of Optics and Laser engineering, Vol.28, PT2. , pp.157-172. 1996.

[4] L. Dekker *et al*, "Models for understanding the 3D human body form". Proc. IEEE workshop on Model-Based 3D Image analysis, pp.65-74, Bombay, India, 1988.

[5] M. Hu, "Visual pattern recognition by moment invariants", IRE Trans. On Information Theory, IT-8:pp. 179-187, 1962.

[6] M. Teague, "Image Analysis via the general theory of moments", Journal of Optical Society of America. pp. 920-930. 70, 1980.

[7] C.H. Teh *et al*, "On Image Analysis by the methods of moments", IEEE Trans. PAMI Vol.10 pp. 496-513, 1988.

[8] D. Shen *et al* , "Discriminative wavelet shape descriptors for recognition of 2-D patterns", Pattern Recognition, 32(1999), pp. 151-165

[9] F.A Sadjadi *et al*, "Three-Dimensional Moment Invariants", IEEE Trans. Pami, Vol.2 No. 2 March 1980.

[10] C. Lo *et al*, "3-D Moment Forms: Their Construction and Application to Object Identification and Positioning", IEEE Trans. Pami Vol.11, No.10, October 1980.

[11] N. Canterakis, "Fast 3D Zernike Moments and Invariants", Tech. Report, 1997, Inst. of Informatic, Univ. Friburg,Germany.

[12] J. Morlet*et al* "Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shape" SIAM, J.Math.Anla, Vol.15, No.4, pp.723-736, July 1984.

[13] S. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation", IEEE Trans. PAMI, Vol.11, No.7, pp.674-693, July 1989.

[14] I. Daubechies, "The Wavelet Transform, Time-Frequency Localization and Signal Analysis", IEEE Trans. Info.Theory, Vol.36, No.5, pp.961-1005, September 1990.

[15] Y. Meyer, "Wavelets: Algorithms & Applications", SIAM Ed. 1993

[16] N.M Ferrers, "Spherical Harmonics", MacMillan 1877.

[17] K. Fukunaga, "Introduction to Statistical Pattern Recognition", 2nd Ed., Academic Press, New York, 1990.

[18] R. Redner *et al*, "Mixture densities, maximum likelihood and the EM algorithm", SIAM Review, Vol 26, No.2, 1984.

[19] http://www.cyberware.com/