

Human Posture Recognition System Based on Depth Imaging

Amrutha S Nair

Assistant Professor

Department of Computer Science and Engineering
Mount Zion College of Engineering, Kadammanitta, India

Abstract— The proposed system is able to efficiently model human postures by exploiting the method includes 3 main steps: skeleton estimation, feature computation, and posture classification. Firstly, a 3D skeleton is estimated from depth images. Joint-based features are then computed for modeling the observed human pose the depth information captured by an RGB-D camera. Skeleton model is used to operate efficiently, regardless illumination conditions in an indoor environment, as it is based depth imaging using the infrared sensor of an RGB-D camera.

Keywords: RGB-D Camera, Depth Imaging, Posture Recognition, SVM

I. INTRODUCTION

Human skeleton configuration is then analyzed in the 3D space to compute joint-based features. Our feature set characterizes the spatial configuration of the body through the 3D joint pairwise distances and the geometrical angles defined by the body segments. Posture recognition is then performed through a supervised classification method. To evaluate the proposed system we created a new challenging dataset with a significant variability regarding the participants and the acquisition conditions. The experimental results demonstrated the high precision of our method in recognizing human postures, while being invariant to several perturbation factors, such as scale and orientation change.

II. PROPOSED SYSTEM

Our system uses a single RGB-D camera (Microsoft Kinect v2) to capture depth maps and build a body skeleton. We then use the 3D joint positions to compute posture features including

- 1) The 3D pairwise joint distances and
- 2) The geometrical angles defined by adjacent body segments.

A supervised classification method is finally applied to recognize static human postures. In summary, the system consists of three main modules (see figure 1): skeleton estimation, feature computation, and postures classification.

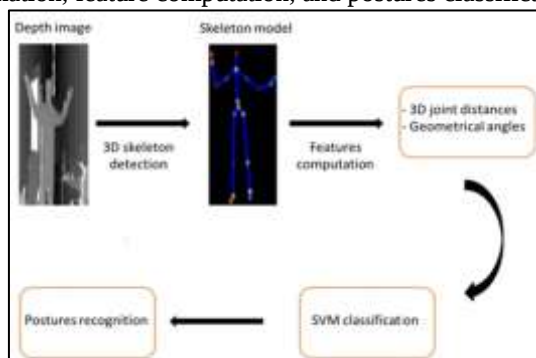


Fig. 1:

A. 3D Skeleton Estimation

To build the 3D model of human posture, we use the skeleton detection method to localize different body parts. This method is quite accurate in modeling the human body as an articulated structure of connected segments. This can be achieved in real-time by the only use of depth images. The method of Shotton et al. uses a random forest classifier to segment the different human body parts through a pixel level classification from single depth images. A mean shift mode detection is then applied to find local centroids of the body part probability mass and generate the 3D locations of body joints. The obtained 3D skeleton formed by 25 joints represents the human posture as illustrated in the figure 2

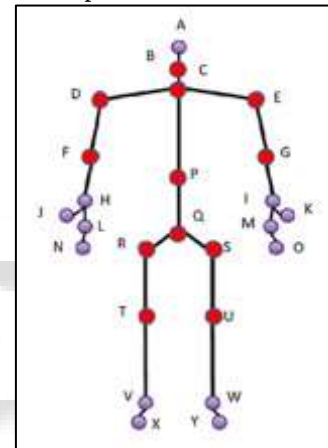


Fig. 2: The 3D joints defining the skeleton. Joints corresponding to the considered geometrical angles are represented in red.

B. Computing Features

Constructing a robust feature set is a crucial step in posture recognition systems. In our work, we used 2 types of features to represent human postures: the 3D pairwise joint distances and the geometrical angles defined by adjacent segments.

Considering the set of $N = 25$ joints (shown in figure 2), the 3D joint coordinates are extracted as:

$$X = \{J_i = (x_i, y_i, z_i), i = 1, 2, \dots, N\} \quad (1)$$

where, J_i is the joint number i obtained from the body skeleton and the triplet (x_i, y_i, z_i) represents the 3D coordinates of J_i . Once joint positions are extracted, we compute 2 types of features representing human body posture: 1) the 3D pairwise distances between joints, and 2) geometrical angles of adjacent segments.

1) 3D Joint Distances:

Given the body skeleton obtained from depth images, we calculate the relative 3D distances between pairs of joints. Such features were successfully used to model human body for action recognition [8]. We calculate the feature vector as:

$$D = \{\text{dist}(J_i, J_j) | i, j = 1, 2, \dots, N; i \neq j\} \quad (2)$$

Where $\text{dist}(J_i, J_j)$ is the Euclidean distance between two joints J_i and J_j . To ensure scale invariance and to remove the impact of body height variation, the 3D distances are normalized with respect to the person's height. The normalization process is performed by considering the distance BP (the 3D distance between the spine shoulder joint and the spine middle joint). The choice of this segment is based on experimental observations concluding that the distance BP is sufficiently stable with respect to several body deformations, in addition to its proportionality to the person's height.

C. Geometrical Angles

In addition to 3D joint distances, our joint-based features include geometrical angles corresponding to relevant joints. Based on our dataset, we selected experimentally a subset of joints according to their importance in characterizing the studied human postures. In the figure 2, the selected joints are represented in red. We considered $m = 12$ angles as follows: (CB, CP), (CP, CD), (CE, CP), (DC, DF), (EC, EG), (DR, DF), (ES, EG), (PC, PQ), (QT, QU), (RT, RQ), (SQ, SU) and (QR, QS).

Geometric angles defined by adjacent segments are directly estimated from joint positions. We define the geometrical angle feature vector as:

$$A = \{\theta_k(u, v) | k = 1, 2, \dots, K\}$$

The angle θ_k between the two adjacent segments u and v is calculated using the 3D coordinates as:

$$\theta_k(u, v) = \arccos(u \cdot v / \|u\| \cdot \|v\|)$$

The two feature sets D and A are then concatenated in a single feature vector to form the final representation of human posture $F = [D, A]$.

D. Posture Classification

Once the body pose is modeled, posture recognition is performed in a supervised classification approach using SVM classifier. With its different kind of kernels, SVM has the ability to generate non-linearly as well as high-dimensional classification issue. The SVM classifier also supports multiple class problems by computing the hyperplane between each class and the rest.

III. EXPERIMENTS

Datasets: We created a new dataset that we named KSRD (Kinect Posture Recognition Dataset) using the Microsoft Kinect sensor v2 1. To the best of our knowledge, this is the first dataset created with the second version of the Kinect sensor (using Time of flight technology for depth sensing). Note that a recent study demonstrated that the second version of the Kinect sensor surpasses Kinect v1 significantly in terms of accuracy and precision [18]. In our experimental setup, the camera is mounted on a tripod placed in a corner of our laboratory.

We consider 5 classes of postures: standing, bending, sitting, walking, and crouching. Each posture was performed by 10 participants of different ages, gender, and morphological characteristics. In order to increase the variability of our dataset and evaluate our system in handling scale and orientation change, each posture is captured at 4 different orientations and 4 different distances with respect to the camera. The distance varies from 1 meter

to 4 meters while the orientation angle is changing between 0 and 360 degrees. Our dataset includes a total number of 800 observations that we used for training and testing the proposed system. The dataset is randomly divided into training and testing set using the K-Fold Cross-Validation method (with $K = 10$).

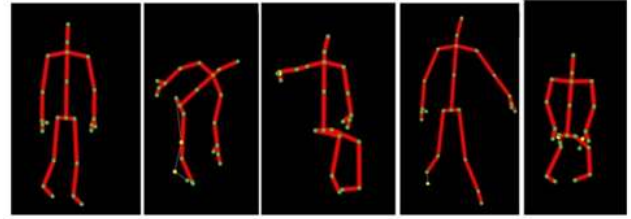


Fig. 3: shows examples of human body skeletons corresponding to the 5 classes

IV. RESULTS AND DISCUSSION

Designed the experiments in order to analyze the performance of our method according to several aspects. First experiment set aimed to identify the optimal choices regarding the two main system components:

- 1) The feature set and
- 2) The classification method.

We thus investigated several choices implying 5 classifiers (linear discriminant, quadratic discriminant, linear SVM, quadratic SVM, and cubic SVM) and 3 feature sets (geometrical angles, pairwise joint distances, and the combination of the 2 types of features). Remark that the best recognition rates of linear discriminant classifiers were obtained by using the geometrical angles as features. However, SVM classifiers clearly outperformed linear discriminant methods regardless the feature set. This investigation shows that the linear SVM applied to the entire feature set achieves the best recognition rate (88.3%). Therefore base our recognition system on

- 1) Linear SVM classification, and
- 2) A combination of the 2 types of features.

One of the major difficulties of pose recognition is the scale change. Method is designed to handle this problem through the normalization of 3D distance features. On the other hand, our dataset includes participants of different heights, whose poses are captured at 4 different distances ranging from 1 to 4 meters from the camera. In order to evaluate the scale invariance of the proposed system, we carried out another sequence of tests where we trained 4 linear SVM classifiers. Each classifier is trained using poses captured at 3 distances (by excluding 1 subset corresponding to 1 distance in each training procedure). The tests are then performed on the excluded subsets. Table IV shows that the proposed method is scale invariant, as it is able to achieve high recognition rates regardless the distance to the camera. In these experiments, the system performance was stable for the 4 tested distances, with a recognition rate greater than (81%). Note that the highest recognition rate was achieved at about 2 meters. V.

V. CONCLUSION

Presented a novel method for human posture recognition using an RGB-D camera. Our method includes 3 main steps:

skeleton estimation, feature computation, and posture classification. Firstly, a 3D skeleton is estimated from depth images. Joint-based features are then computed for modeling the observed human pose. Posture classification is finally carried out using a linear SVM. Our method was validated on a challenging dataset released in our laboratory. The proposed system demonstrated high recognition rates and a significant invariance to important perturbation factors, including scale, orientation, and illumination changes. Our future work will focus on developing appropriate mechanisms addressing other difficulties, such as partial occlusion and fast body movements. This also includes extending the dataset by considering these challenging conditions.

REFERENCES

- [1] B. Boulay, F. Bremond, and M. Thonnat, Human posture recognition in video sequence. IEEE International Workshop on VS-PETS, Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2003.
- [2] R. Gouiaa and J. Meunier, Human posture recognition by combining silhouette and infrared cast shadows. In Image Processing Theory, Tools and Applications (IPTA), 2015 International Conference on, proceedings, 2015

