Mimicry Engine Using Kinect

Garima Goyal, Onkar Borgaonkar, Avinash More, Abhijeet Nirmal

Modern Education Society's College Of Engineering , Pune. Computer Department.

Summary

This paper presents a system for matching of 3 Dimensional actions using DTW. This system uses a depth sensing device to generate a skeleton of the user. For the training part, any Expert can train the system for a specific action sequence. The User can perform a similar action and the system will match the actions to calculate the percent match. The intention is to make the system independent of the user size and position by matching the joint angles instead of joint coordinates. At the moment the system is time variant. To further increase the tolerance of error in the user action, we use a range of actions performed by the expert and calculate the similarity of the intended action and user action based on a user specified strictness level. This system will calculate the accuracy of the user's actions and help improve the action and stance if used in tutoring software. This will be a generalized system which can be used as any action matching and action improving engine. The recognition system is able match the candidate pattern with that of the expert.

Keywords

DTW (Dynamic Time Warping), depth sensing, joint angles, time variant, similarity, range of actions.

1. Introduction

Computer vision is a field that includes methods for acquiring, processing, analysing, and understanding images and, in general, high-dimensional data from the real world in order to produce numerical or symbolic information, e.g., in the forms of decisions. Basically this field deals with the duplication of human vision by electronically perceiving and understanding an image. This image understanding can be done using the symbolic information from image or the models constructed using different techniques. Computer vision is that enterprise where the processes and representations are collected together and processed for vision perception. It has a wide range of applications such as autonomous vehicles, face recognition, smart camera, object detection etc. In machine learning, pattern recognition is the assignment of a label to a given input value. This is a separate domain which can be merged in with many domains viz. computer vision, pattern matching, digital signal processing, information security, image processing etc. various techniques and algorithms are available in order to apply pattern recognition in varied domains of technology. The algorithms used in pattern recognition try to provide an answer to the given set of inputs with all possible matches of inputs.

The system implemented in this paper is based on a low cost tutor or rather a mimicry engine which finds applications in general scenarios. The engine builds interactive environment with its users to analyze and critique the actions performed. This could prove to be an intelligent tutor provided we have an intelligent person or virtuoso to train the system. The algorithm used is DTW in order to perform matching of the two action sequences. We try to make a time variant system i.e. both the actions shall have to be performed in the same time frame for them to be matched, if not, then the similarity of the two action sequences shall be affected. After matching the two sequences the Accuracy Result will be displayed at the end of the matching module.

2. Related Work

An interesting question to ponder upon is what the next generation computer application would be like? To answer this, let's recall the first computers, with a command-line interface. Then came the GUI based computers. In which calculating, playing music, playing games, word processing and other operations were interactive using hardware devices. For the next generation computer interaction, we expect to use more interactive methods. For example can a computer identify me by looking at my face or even my gait? Can a computer know where I'm looking at and what I'm doing? Can a computer tell what is living and what is non-living? Can computers learn something by themselves? Can a computer summarize a video for me? While depth sensors are not conceptually new, sensors such as Kinect have been made accessible to all. In [1] this work represents an action recognition algorithm for matching dance sequences. [5] In order to align and evaluate dance performances, Kinect depth maps from the associated grand challenge data set are considered. [1] In this paper the point based and angle-based skeleton model required for the body and joint angel trajectory matching respectively.

3. Problem Description

The proposed system is developed in order to help the user to improve his actions. We focus on generating a set of the results for actions performed by the user in comparison with the actions performed by an expert. We present an

Manuscript received February 5, 2015
Manuscript revised February 20, 2015

approach for measuring the similarity between the two action sequences for any generalized applications. In this system we have concentrated not on one factor but it deals with many generalized actions. This is a more generalized system which will provide a helping and supporting hand in learning any form of actions involved in day to day life using skeletal information obtained from any depth sensor, and propose an easy to implement algorithm that can solve this problem in real time. The main problems involved are to stabilize and normalize the anthropometric and temporal variations. Though such differences can't be avoided when acquiring human motion data, we try to avoid and mitigate such differences while recognizing the sequence in following ways:

- By using an angle based skeleton model that will reduce and avoid anthropometrical dependencies, while it reduces the obtained candidate dimensionality without any loss of important information of sequence.

-we introduce a sample similarity measure based on our system algorithm in order to overcome the temporal differences and transformations which may vary in speed and distance from the camera. The similarity is bounded, even though the length of the user action sequence may vary; as the matching is performed only after the user had completed his action. Thus the user action sequence length is known before the matching process initiates.

The proposed recognition system is a part of assessing whether the action sequence performed by users approaches the expert action sequence, which we pre-recorded with the help of an expert during the training of the system. So we do not adopt algorithms that involve statistical recognition methods, thereby eliminating unnecessary set of data to be processed, so as to match two action sequences. The structure of the system is illustrated in the figure below:

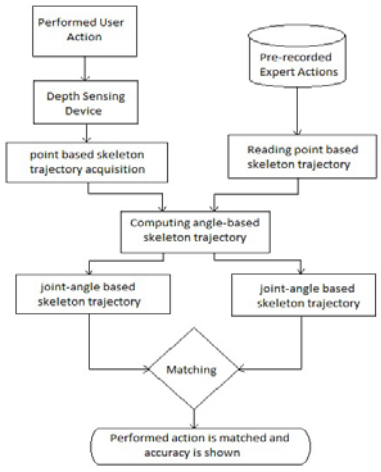

Fig. Block Diagram of working of Mimicry Engine

The system consists of four modules: - i) System Training Phase by using a depth sensing device for skeletal tracking. ii) User Skeleton tracking by acquiring the point based skeleton of the actions performed. iii) Feature extraction module where the body trajectory is converted from 3D point based skeleton to angle based skeleton which consist of the thirteen most important joint angles. iv)An angel based skeleton trajectory module comparison.

An important feature of our contribution is that, unlike the majority of sequence matching algorithms, ours does not need to segment a candidate pattern in the incoming motion data prior to recognition. In the first phase of system training the engine is trained with the help of an expert. These actions performed by expert are recorded thrice. All the three training sets of action sequences are stored in point based and joint-angle based skeleton format. In the second phase which is user skeleton tracking we are going to track point based skeleton of candidate and will be stored in the system in the form of angle-based skeleton trajectory. Here we are obtaining the user skeleton by using the depth sensing device. The output of that depth sensing device will be point-based skeleton trajectory. To eliminate the bottlenecks that are caused due to point based skeleton trajectory, we need to convert it into joint-angle based skeleton trajectory. The conversion of the point based skeleton trajectory to joint-angle based skeleton trajectory is accomplished in the feature extraction phase. In this phase we are obtaining 13 angles from user's skeleton, this overcomes the problem of anthropometric differences.

The problem that we are facing in getting a sequence similarity measure, while comparing a very large set of data, is to eliminate the delay factor involved while making the comparison. We have provided user with the number of options from very strict to lenient matching in which the tolerance level of the difference between frames (candidate and baseline) can be varied so as to give space for human error.

The sequence of the paper is that Section 4. will give the description about the point based and joint angle based body trajectory. Section 5 will provide the recognition algorithm strategy applied and the angle based trajectory comparison, and Finally Section 6 describes the results and conclusion.

4. Action Recognition and Feature Extraction

4.1 Point Based Skeleton

The point based skeleton is a set of 15 points obtained from depth sensing device, which are depicted by ellipses E= {1, 2, and 15}.

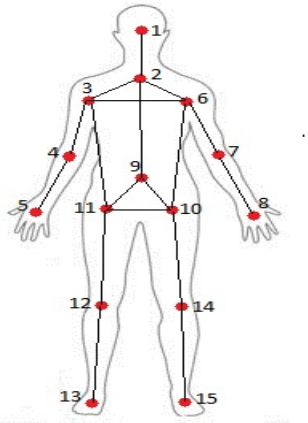

Fig: Front view of 3D Skeleton of Human body.

The set of points that are obtained are in a co-ordinate format. Each point from the set E can be represented as follows:

Suppose Xi is any point from set E then, Xi = {xi, yi, zi} where x, y, z are the three co-ordinates of the point X. The skeleton set S can be represented S={X1, X2....., X15}. Although this representation can describe every pose of human body, it is heavily dependent on the body size and proportion of the user. These anthropometric differences can be eliminated by using the angle-based data instead of using the 3D-point data. For this we define a function

$$f(k,l,m,n) = arccos\frac{<l-k,n-m>}{||l-k||||n-m||} \quad (1)$$

where k, l, m, n are the points from skeleton. Using the above function we create a feature vector set containing 13 angles.

$$V = \begin{pmatrix} \theta ler \\ \theta ley \\ \theta lsr \\ \theta lsp \\ \theta rer \\ \theta rey \\ \theta rsr \\ \theta rsp \\ \theta lhp \\ \theta lkr \\ \theta rhp \\ \theta rkr \\ \theta tor \end{pmatrix} = \begin{pmatrix} f(x7,x6,x7,x8) \\ f(x6,x2,x7,x8) \\ f(x6,x10,x6,x7) \\ f(x10,x11,x6,x7) \\ f(x4,x3,x4,x13) \\ f(x3,x2,x4,x13) \\ f(x3,x11,x3,x4) \\ f(x11,x10,x3,x4) \\ f(x14,x10,x2,x9) \\ f(x14,x10,x14,x15) \\ f(x12,x11,x2,x9) \\ f(x12,x11,x12,x13) \\ f(x11,x10,x3,x6) \end{pmatrix}$$

which is the reduction of the original skeleton model S. The vector set has been created without any loss of important information. Here we have given the meaning of each angle with the logical names left elbow roll and similarly for the right part. The first letter can be l (left), r (right), t (torso). The second letter can be e (elbow), s (shoulder), h (hip), k (knee). The last letter can be r (roll), y (yaw), p (pitch). The last letter to stands for torso and r stands for roll.

4.2 Joint Angle Based Skeleton

Now one can define the joint angle based skeleton trajectory (J) set for any form of action using the features angles mentioned in the above equation. The trajectory J= {v1, v2, v13}. The set of points in joint angle based skeleton are free from anthropometric differences. Hence they provide a set of data which can be useful in comparing any set of actions performed by different people having varied physique. This is the main function which will be used many times in order to achieve the targeted aim in our project. One more point that should be noted is that the angle based skeleton does not provide us with the whole relative information so we consider and describe the whole global body displacement of the skeleton to be located at the centre of mass in the 3d space i.e. at the torso.

5. System Algorithm

This section describes the actual implementation and technique used to reach the targeted aim of our project.

The first part deals with the user tracking and recognizing the action performed by different users. This is done by providing namespaces to different users. Also the set of actions performed are saved in separate namespaces in the form of files. These files contain data related to joint angle based skeleton which is generated using the vector set in a continuous stream of frame data containing position of skeleton joints. System is trained thrice for the same action. Now we are having baseline pattern with us which can be used for comparison.

Once the actions of the expert are recorded then the actions of user are recorded from, which are displayed on a canvas next to a canvas showing the actions that are performed by the expert. This is done so as to help user to imitate them at ease. After the actions are done now begins the matching technique where the similarity is displayed based on mainly the tolerance adjusted with those of the strictness levels. Here the similarity measure has been scaled from lenient to very strict.

A separate list for selection of appropriate action set of the expert with which the actions of the users are to be matched is being provided.

5.1 Action Pattern Recognition

As illustrated in the block diagram, the goal of the system is to recognize a baseline skeleton trajectory B = {b1, b2, bM} in a candidate skeleton trajectory C = {c1, c2, cN}, which is captured in real-time. The recognition implies that both the corresponding joint-angles of the action sequences must match. To achieve this, we make the

following assumptions about the baseline and candidate trajectories:
– There is no priori knowledge about which parts of the candidate trajectory C contain important information. Also there is not one concentrated section in the action from where one can inference that the actions are matching.

5.2 Similarity Measure

A sample similarity measure has been defined in this paper, before actually applying the DTW algorithm. This sample similarity measure is based upon the feature (positions/angles) vectors that define the skeleton. However to make space for removing and normalizing the frame differences between the two feature vector sets without much affecting the similarity measure, we calculate an inner product of the two feature vector sets.
Next task is to compare baseline sequence and candidate sequence. For this we created sample similarity matrix.

$$S = 1 - v.w/|v||w| \quad (2)$$

where, S= sample similarity matrix. And from the sample similarity matrix we have calculated optimal warping path that will lead to the nearest correct solution. In the above equation we have used a 'inner product' of the two sequences i.e. the candidate and the baseline sequence. This measure has a property that it is very useful for determining the tolerance or the threshold value. This measure is used along with DTW to form a sequence similarity measure.

5.3 Angle Based Trajectory Comparison Strategy

Now a sample similarity matrix is achieved, so one has to use this entity in order to compare the two action sequences. This measure can be used to compare whether the candidate (user) pattern C is similar with the baseline (expert) pattern B. So a sample similarity threshold value T is introduced to decide the match capability value between the two considered sequences. For this the range from the baseline trajectory (the action sequence performed by the expert) is obtained which is used for comparing the action sequence of the user (candidate) with the baseline trajectory.

5.4 DTW Approach

The warping path that we have defined is under a range of values between the smallest and the largest vector values. Now these values in the range are compared with the candidate trajectory.
Here three conditions are possible:
When the action sequence of candidate trajectory lies within the range of baseline trajectory, i.e. this is near to ideal action sequence that the candidate has performed. Hence the accuracy computed by the system will be much higher but will depend on the chosen strictness level.
When the action sequence of candidate trajectory is above range values then in this case this sequence will be matched with the higher range values in the baseline trajectory. Hence the result which is computed depends upon the difference between the action sequence of candidate and higher range values of the baseline trajectory.
When the action sequence of candidate trajectory is below range values then in this case this sequence will be matched with the lower range values in the baseline trajectory.
The decision of the range values is computed as follows:
Since our system is trained thrice by the expert so we need to consider all the three baseline trajectories while deciding the range. While considering the feature vector values of these three trajectories we compare these values one by one among each other so as to separate out, which of these is the highest and the lowest values in different namespaces.

6. Results and Conclusions

In experiments, the algorithm is implemented on Windows platform using C# in Microsoft's Visual Studio 2012 environment.
In this paper OpenNI library is used for getting the required data from Microsoft Kinect Range Camera. This library helps in acquiring the point based skeleton from the Kinect. The input given to the system is pre-recorded baseline pattern which is captured by the Kinect. Various kinds of gestures are considered ranging from simple such as wave, bend, and step forward or backward to complex gestures which include full body motion. With the help of our system approximately 85% of similarity has been achieved by testing it manually. Yet we didn't find any automated tool to test engine. Acknowledgement
We are very thankful to TouchMagix Pvt. Ltd. for providing us with the necessary environment and guidance. The sole idea behind this project is formulated by TouchMagix. We are also very thankful to our professors, HOD and the guides from the company. Special thanks to P. R. Menaria, Mandar Kulkarni, Spurti Shinde and the College authorities for accepting our idea.

References

[1] Real-Time Dance Pattern Recognition Invariant to Anthropometric and Temporal Differences Meshia C´edric Oveneke, Valentin Enescu, and Hichem Sahli Vrije Universiteit Brussel Department of Electronics and Informatics (ETRO) Pleinlaan 2, 1050 Brussels, Belgium coveneke@vub.ac.be,escu,hichem.sahli}@etro.vub.ac.be
[2] Dance Pattern Recognition using Dynamic TimeWarping Henning Pohl, Aristotelis Hadjakos Telecooperation

Technische Universit¨at Darmstadt Darmstadt, Germany fhpohl@rbg, telis@tkg.informatik.tu-darmstadt.de
[3] View Invariant Human Action Recognition Using Histograms of 3D Joints Lu Xia, Chia-Chih Chen, and J. K. Aggarwal Computer & Vision Research Center / Department of ECE The University of Texas at Austin {xialu|ccchen}@utexas.edu, aggarwaljk@mail.utexas.edu
[4] Evaluating a Dancer's Performance using Kinect-based Skeleton Tracking. Dimitrios Alexiadis, Petros Daras Informatics and Telematics Institute, Thessaloniki, Greece {dalexiad, daras}@iti.grPhilip Kelly,Noel E. O'ConnorCLARITY: Centre for SensorWeb Technologies, Dublin CityUniversity, Ireland{philip.kelly,noel.oconnor}@dcu.ieTamy BoubekeurInstitut Telecom / Telecom ParisTech, Paris, France tamy.boubekeur@telecomparistech.frMaher Ben MoussaMIRALab, University of Geneva, Switzerland benmoussa@miralab.ch
[5] IEEE TRANSACTIONS ON CYBERNETICS, VOL. 43, NO. 4, AUGUST 2013 Inverse Dynamics or Action Recognition Al Mansur, Yasushi Makihara, and Yasushi Yagi.

Garima Goyal B. E. Computer Engineer degree from University Of Pune, in 2014 from Modern Education Society's College Of Engineering, Pune, India.

Onkar Borgaonkar B. E. Computer Engineer degree from University Of Pune, in 2014 from Modern Education Society's College Of Engineering, Pune, India.

Avinash More B. E. Computer Engineer degree from University Of Pune, in 2014 from Modern Education Society's College Of Engineering, Pune, India.

Abhijeet Nirmal B. E. Computer Engineer degree from University Of Pune, in 2014 from Modern Education Society's College Of Engineering, Pune, India.

(Project Guide) Prof Spurti Shinde Completed ME Information technology from PICT ,Pune under Pune university ,India. Her Field of research is Image Processing.She has 5 years of Experience in this field.