Paper report on

# Compression of Motion Capture Databases

*Sornum Kabilen Kadirvelen, Reg. No. G0701666C, NTU.*

**Reference Paper: ARIKAN, O. 2006. Compression of Motion Capture Databases.** *Computer Graphics, Annual Conference Series, ACM SIGGRAPH 2006.*


## ***Abstract***

*This report summarizes a lossy compression algorithm/method used for large databases of motion capture data. This method of compression of motion capture databases has been presented at the ACM SIGGRAPH annual conference series 2006.*


## 1. Background of Research

In past and current technologies, we have noticed that data acquisition technologies such as motion capture can produce large volume of data. If this data is to be applied in commercial applications we would like to accommodate as much as possible in a minimal amount of memory.

Established compression techniques involving audio and video make use of perceptual models. A good compression method involves study of features which can discarded and features which are fundamental to preserve. Similarly for motion compression essential features of human motion is important.

The main purpose of a compression method is to create a compressed representation with maximum compression ratio, and that is affectively as close to the original. Also, we would like compression and decompression to be as fast as possible and memory efficient.

## 2. The Proposed Algorithm

## 2.1 Compression

### 2.1.1 Overview

The motion database is assumed to be a single long motion which is represented as a vector valued function *M(*t), where *t* refers to the frame number. Each frame *M(t*) contains degrees of freedom which control the character (global position/orientation and a set of joint angles). The motion database is divided into clips of *k* subsequent frames.


Each clip is rotated and translated so that the character is located at the origin in a standard representation in the first frame of the clip. These transformations are stacked so that they can be popped back during decompression. The degrees of freedom are converted into a positional representation (lossless). Cubic Bezier curves are applied to this lossless representation. Locally linear models are applied to the Bezier control points to reduce the clips' amplitude.

Essential environmental contacts are encoded using a separate JPEG-like technique.

### 2.1.2 Motion Representation

To maintain proper environmental contacts, 3D trajectories of joints are used rather than joint angles. We convert rigid coordinates frames for each bone into a set of virtual markers placed at fixed positions. This is represented by the following matrix:

$$T \times R \times \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{vmatrix} = [\, a \,|\, b \,|\, c \,]$$

### 2.1.3 Smooth Approximation

A 3D cubic Bezier curve using least squares is fit over the *k* frames and each of the trajectories is represented with their control points. For each virtual marker we store only 12 numbers (4 control points x 3 coordinates) rather than 3 x *k.*

### 2.1.4 Clustering and Projection

Similar looking clips are grouped distinctively by clustering. The Nystrom approximation is used in the spectral clustering. This method is computationally efficient in our high dimensional space. Some motions are repeated in big motion databases. These common clips belonging to the same cluster have their projected coordinates quantized into 16 bits. The result are integers having low entropy and are compressed using entropy encoding using Huffman codes.

### 2.1.5 Environmental Contacts

Contacts with environment are important and a foot should not move when it is in contact (assuming that there is no slipping). The feet are compressed differently and we make sure the contacts are maintained during decompression. Discrete Cosine Transform(DCT) is used to transform the virtual markers into a frequency space. The resulting coefficients are quantized into a finite no. of bits. Due to the high importance of feet-ground interaction this algorithm compresses the entire trajectory of both feet.

## 2.2 Decompression

Each clip is decompressed individually. Entropy decoding is performed and the quantization is undone for the cluster coordinates and for virtual markers of the feet. The Bezier control points are obtained for the virtual markers of the feet. To return to the original representation rigid body transformation is applied between 2 adjacent bones (say *i* and *j*) by

$$T_{ij} = T_i^{-1} T_j$$

The rotation part of $T_{ij}$ is converted to joint angles. As the feet may not reach the desired position, IK (Jacobian based) is performed on the foot positions to restore originals.

Finally, we translate and rotate the clip to position it in the global coordinate frame.

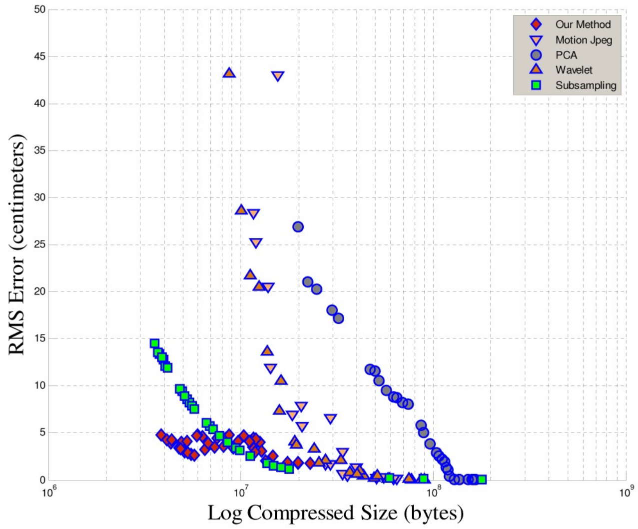
Discontinuities between 2 adjacent clips is solved by 'C1 continuous merge' method.

## 3. Results of Algorithm

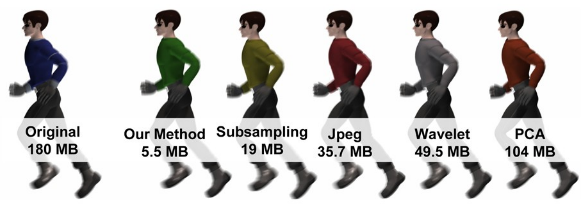
The results of the algorithm were demonstrated on 2 datasets. The 1st dataset consisted of different kinds of locomotion and amounted to 180 MB of uncompressed storage. The second dataset is the motion capture collection maintained at Carnegie Mellon University and amounted to 1085 MB of uncompressed storage. The RMS error was computed by:

$$RMS = \sqrt{\frac{1}{n}\sum_{t=1}^{n}|M(t) - M_c(t)|^2}$$

where $n$ is the no. of frames in the database and $M_c(t)$ is compressed version of motion. The figure below illustrates the RMS errors as compared to baseline methods like motion JPEG, PCA, wavelet and subsampling. The compression results and ratios is shown in the following table:




| | Us | Sub | JPEG | Wavelet | PCA | ZIP |
|---|---|---|---|---|---|---|
| CMU<br>1085MB | 35.4<br>30:1 | 92.1<br>12:1 | 237.2<br>5:1 | 184.7<br>6:1 | 520.4<br>2:1 | 788<br>1.4:1 |
| Sony<br>180MB | 5.5<br>32:1 | 17.9<br>10:1 | 35.71<br>5:1 | 49.48<br>4:1 | 104.23<br>1.7:1 | 165<br>1.1:1 |




## 4. Strengths of proposed method

- Method is fast for compression and decompression. It is 7 times faster than real time.
- Offline compression can be achieved, i.e without holding the entire database in memory.
- Method is not sequential, i.e any clip can be decoded without the need of decompressing others.
- The compression ratio achieved with minimal loss is quite high. Avg. compression ratio of 33:1 was obtained for the 2 datasets.

## 5. Limitations

- The major limitation of this method is that the contacts with external elements need to be known. Annotation is required to keep track of the contacts as compared to feet contact which is usually in contact with the environment.