Content based 3D Human Document Retrieval using Latent Semantic Mapping

Yohan Jin
Data Science, Tapjoy Inc.
111 Shutter St.
San Francisco, CA 94014, USA
Yohan.jin@gmail.com

Balakrishnan Prabhakaran
University of Texas at Dallas
MS EC 31, PO Box 830688
Richardson, TX 75080, USA
bprabhakaran@utdallas.edu

Abstract

There has been an enormous increase of 3D human motion data in various fields, such as 3D gaming (such an EA sports) and medical fields (physical medicine and rehabilitations). We need an effective content-based 3D human motion retrieval scheme supporting human-level language queries. However, there is a big semantic gap between these two media since the 3D Human motion data and text are heterogeneous forms. In this paper, we propose a cross-media retrieval framework that reduces the semantic gap by semantic spatiotemporal dimensional reduction and reformulates 3D human motion data to HMDoc (Human Motion Document) representation, which is quite applicable for a traditional information retrieval technique such as Latent Semantic Indexing. After mapping complex 3D human motion matrix data into semantic space, we can achieve 88.72% precision, 86.98% recall accuracy with 14 different motion categories that consists of 370,294 frames. Our proposed approach (HMDoc) extracts the semantic characteristics of human motion capture data. This semantic feature compact representation outperformed other works such as weighted motion feature vector and LB_KEOGH's method, Geometric feature representation.

1. Introduction

3D motion capture data is 'multi-dimensional' time-series data. Each row of 3D human motion capture data matrix corresponds to a single frame that consists of information for 29 segments (corresponding to different parts of human body) and each element has the value of the degree of freedom. Human body consists of unit body part (UBP), which includes several segments. For example, the torso consists of 7 segments (with degree of freedom in parenthesis) namely root (6), lower back (3), upper back (3), thorax (3), lower neck (3), upper neck (3), and head (3) segments. The arms and legs consist of 7 and 4 pairs of segments respectively. In the 3D motion capture data matrix, column is corresponding to each body segment and each row is time-series. Because of motion-length variance and huge number of dimensions (62 dimensions), there is huge semantic gap between human description and real 3D human motion matrix. Normally, the level of details in the 3D human motion capture data is designed for rendering purpose. 3D human motion consists of a highly dense matrix, which is for rendering vivid human motions effectively.

However, to efficiently process for 3D human motion classification and retrieval purpose, we might not need to keep original amount of high dimensional matrix data. In our previous approach [8], we demonstrated that semantic dimensional reduction of 3D Human motion capture data can keep each motion class' characteristics. Although quantized symbolic sequences from 62 dimensional original motion data look too simple, we showed that it is quite enough and more useful representation for motion classification and retrieval purposes. Our main contribution is bridging the gap from high-dimensional time-series data to the semantic representation in 3d human motion domain. Especially, to recognize time-series human motion, time-variance is huge barrier to find similar motions accurately. In this paper, as discover "repeated" patterns as "rules", we show that human motion can be represented as time-invariant form- we can call it as 'Eigen-human-motion' vector form.

From an observation that each symbolic value (Torso, Arm, Legs and so on) has semantic meaning, we get an idea of dealing these human motion sequences as textual symbols. Through transformation from high-dimensional multimedia data (human motion) to symbolic representation, now we can apply some of very useful textual information retrieval techniques. Here, we will show how we can retrieve human motion data by using LSI (Latent Semantic Indexing) technique.

Some of cross-media retrieval approaches [10, 19, 23, 24, 25] tried to apply the traditional information retrieval technique, such as LSI (Latent Semantic Indexing) since LSI technique is already known for its usefulness in terms of retrieval speed and scalability. These approaches transformed image [19, 23], video [24, 25] and music clips [10] into text words. Then, these newly constructed multimedia documents can use the semantic indexing technique. Souvannanong et al. [23] tried to apply LSI technique to video data for categorization, with some difficulty in segmenting visual low-level features for quantization. Our intention is to explore information retrieval approaches for 3D human motions database.

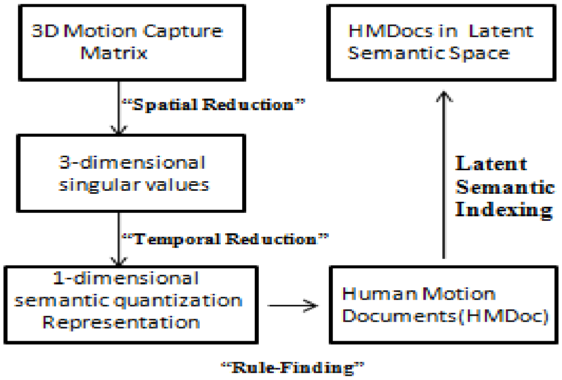

3D Motion Capture Matrix
"Spatial Reduction"
3-dimensional singular values
"Temporal Reduction"
1-dimensional semantic quantization Representation
Human Motion Documents(HMDoc)
"Rule-Finding"
Latent Semantic Indexing
HMDocs in Latent Semantic Space

Figure 1. Overall Framework for Cross-media Retrieval with 3D Human Motion Capture Data

In this paper, we contribute by proposing a new framework using information retrieval technique for content-based 3D human motion retrieval. Through our proposed framework (Fig. 1), we claim and demonstrate that 3D motion capture data can be transformed into multimedia documents. Previously, Gutemberg et al. [6][7] tried to construct "human action" language model for human motion analysis. We show that 3D motion capture can be effectively quantized while maintaining its semantic characteristics through spatial-temporal semantic quantization process [8]. This process transforms 3D human motion data matrix into a 'human motion string'. Next, we segment the human motion string into "rules" by finding repeated string sequence as "rules". As Zhai et al. [28] explored structures in continuous video streams using clustering technique and fuzzy analysis, we try to detect "repeated" hierarchical pattern in continuous motion streams. Now, a single motion clip consists of several rules. So, we reformulate each 3D human motion clip to a HMDoc (**H**uman **M**otion **Doc**ument) by considering '*rules*' as '*terms*' in textual document and considering one 3D human motion clip as a document.

From HMDoc construction, we can apply LSI (Latent Semantic Indexing) technique to HMDoc for content-based 3D motion retrieval by mapping 3D human motion data into semantic space. For some motion classes overlapping in k-semantic space (k=2), we can see the accuracy improvement with increasing the dimensions of semantic space (k=6).

We can achieve 88.72% precision, 86.98% recall accuracy. It implies that our proposed cross-media framework can keep the semantic characteristics while we transform 3D complex motion data into low-dimensional one and map that into semantic space with newly constructed Human Motion Documents. It should be observed that the proposed approach, perhaps with some modifications, can be used for mapping normal (i.e., 2D) video of human motions for retrieval as well.

2. Related Works

Li et al. [14] extracted geometric structure as exposed by SVD of matrices of human motion data and index using interval-tree based index structure, similarity, Li et al. [15] classified human motions apply SVM on geometric extracted motion vectors and Guodong et el. [17] selected small set of leading eigenvectors as principle features and tried to represent motion frames as simplified "cluster transition signature", which is conceptually similar to 1-dimensional quantization representation in this paper. Other approaches utilized hierarchical trees for indexing 3D human motions. Gaurav et al. [4] used hierarchical structure of the human body segments to increase the searching speed and accuracy. Each level of index tree is associated with the weighted feature vectors of a body segment. Feng et al. [16] proposed content-based motion retrieval (CBMR) by building motion-index tree on hierarchical motion description, which serves as a classifier to determine a sub-library that contains promising similar motions to the query example. For dealing with temporal invariance between similar motions, [16] used "elastic match", a combination of DTW (Dynamic Time Warping) and dynamic programming. To overcome the limit of DTW technique ('local scaling') for time-series data comparison, Keogh et al. [11] proposed a uniform scaling , which can scale globally and showed that it can speed up indexing using bounding envelopes. Most recently, Muller et al. [20] contributed content-based human motion retrieval through "qualitative" geometric description for bridging the numerical and perception human motion similarity gap. However, a user has to select suitable features in order to obtain high-quality retrieval results. Totally different from other approaches for content-based 3D motion retrieval, this paper shows the way of mapping each complex and time series human motion data into semantic space through Latent Semantic Indexing on semantically quantized representation of 3D Human Motion.

3. Segmented Semantic Representation

Spatial Features: We consider one human motion as characterized by different combinations of three main body parts: torso, arms and legs. For extracting spatial relationships and reducing dimensions (from 62 to 3 dimensions) among the 3 different body components, we separate a motion data matrix ($M_{f\times m}$) into three sub matrices ($M^{\alpha} = M_{f\times k}$, $M^{\beta} = M_{f\times j}$, $M^{\gamma} = M_{f\times r}$,

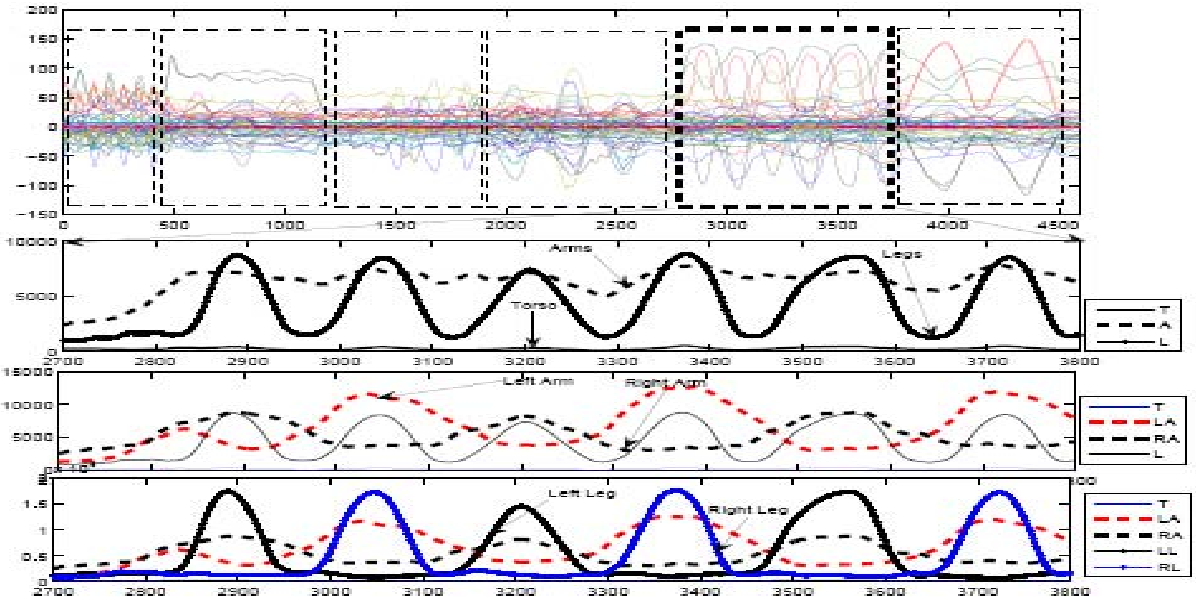


Figure 2. SVD Spatial Feature Dimension Reduction with 'Side-Twist' motion.

where *m=k+j+r*) belonging to torso, arms and legs part respectively. From three sub matrices, SVD decomposes "singular" values [5].

$$M^i = U\Sigma V^T, M^i v_1 = \sigma^i v_1, i \in \{\alpha, \beta, \gamma\} \quad (1)$$

Now, three "singular" values ($\sigma^{\alpha}, \sigma^{\beta}, \sigma^{\gamma}$) which represent torso, arms and legs parts are the coefficient of each frame as the spatial feature, then we have a reduced matrix $M_{f\times 3}$ for a single human motion clip. If we split Arms columns into Left and Right Arms, then we can have more detailed representation about Arms motion. Likewise, we can separate leg columns into Right and Left parts. Then, we can increase spatial dimensions up to 5 dimensions in this case.

Temporal features: We map spatially extracted 3-dimensional singular values of each motion file into Gaussian Mixture Semantic Space [1] (see Fig. 2), this space is to find "latent" semantic quantization components (A(rms), T(orso), and L(egs)) which is corresponding to the given observation (i^{th} frame of one motion file). It computes $\Re_{ki}$ of the 3 body components, which is probability of "latent" component k's responsibility for observing the i^{th} frame O_i.

$$\Re_{ki} = P(k \mid O_i) = \frac{P(O_i \mid k)P(k)}{P(O_i)} = \frac{g(O_i;\mu_k)\Phi_k}{\sum_{k=1}^{K} P(O_i,k)} \quad (2)$$

Let $P(O_i \mid k)$ be the Gaussian function $g(O_i;\mu_k)$ of latent component k, and $P(k)$ be the mixing parameter Φ_k of latent component k. $P(O_i)$ is the "prior" probability that we can get from the marginalization of joint probability. Humans can express one action using more than one body component at the same time, so we need to extend Gaussian Mixture Semantic Space from three main body parts (Triangle) to a combination of the three main parts (Cube) (see Fig. 3). We add three combined "latent" components corresponding to each edge of Triangle, which is 'TL', 'AT' and 'AL' respectively. Thus, the overall number of "latent" component in Cube space is 8 including null (ϕ) and all (TAL) components. Each quantizing component has its semantic meaning: for example, if one frame window has a mixture value close to 'L', it means this frame window includes "legs intensive" actions. And if one frame re value is close to 'TAL', then it means that this frame has action using "legs" ,"arms" and "torso" actively. We can extend GMM with EM (Expectation Maximization) for finding local maximal values based on the initial GMM values of human motions. After iteratively running with GMM and EM model (Fig. 3 (b)), we get the locally maximized mixture value.

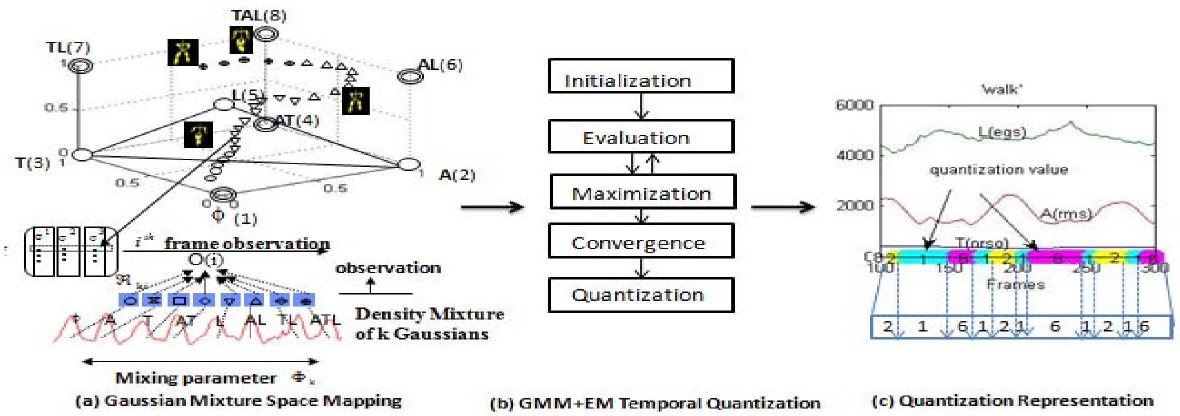


Figure 3. Semantic Gaussian Mixture Quantization Process.

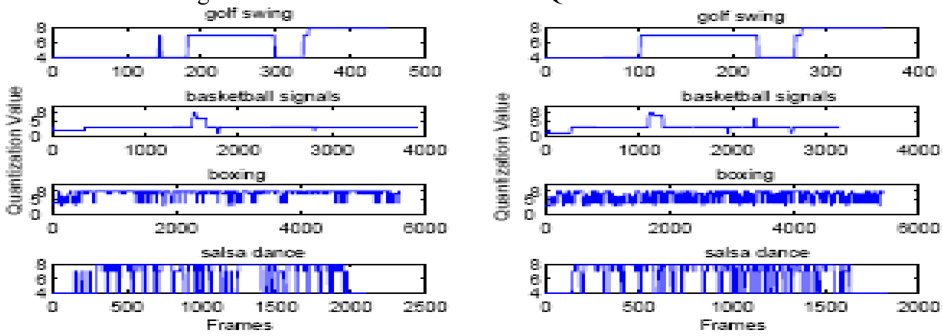


Figure 4. Quantization Value Representation of Similar Motions.

Then, we assign each maximized value to the closest quantization value ($\phi \leftarrow 1, A \leftarrow 2, T \leftarrow 3, AT \leftarrow 4$ and so on) in order to get the quantized representation of each frame window from spatially extracted feature vectors. It shows the effect of temporal segmentation as well as semantic quantization (see 'walk' example's quantization representation -'2-1-6-1-2-1-6-1-2-1-6 in Fig. 3 (c)). In Fig. 4, we observe that quantization value representation from semantic GMMEM has been segmented with temporally similar frames. Thus, finally extracted quantization value has spatial and temporal characteristics of a specific motion.

4. Finding Repeated Patterns in Quantized 3D Human Motion Data

From previous section, we show that 3D human motion capture data can be transformed into strings, where each literal has semantic meanings such as T(orso), A(rms), L(egs), TA, AL and so on. We can observe that quantization values of each motion class express semantic characteristics. For instance, semantically, 'golf swing' consists of pause, shot and relax, 'basketball signal' usually includes one specific signal after waving hands (see Fig. 4). Since human motions are repeated with some periodicity, we would like to detect those repeated semantic strings as rules.

We employed 'SEQUITUR' algorithm [11] that has been used to index large size of digital library by generating hierarchical phrases, which is a novel method for browsing. Through the Semantic Quantization process, we have translated from high-dimensional 3D motion capture data format into Human Motion Strings. Here, we call one-dimensional semantic representation of 3D human motion as Human Motion Strings. From the observation that 'SEQUITUR' algorithm [11] can segment real sequences with repetition units and deal with variant nature of similar sequence data, we get to know that 'SEQUITUR' algorithm is also applicable to Human Motion Strings. For similar motions, most of the hierarchical rules are commonly shared (see Fig. 5), but there is a variance problem to solve. For example, in the first root node rules, R3770 and R3793, R3793 subsumes R3770 as left child and has R3740 as right child. (R3793$\supset$R3770 $\supset$R3740). Such differences between similar motions are caused by length-variance; in this case, "basketball signal (b)" has more motion sequences than "basketball signal (a)" as amount of R3740 at the initial part in time order. About these variance, 'SEQUITUR' can find common parts as rule in hierarchical manner (in that, “basketball signal (a)" also has R3740 as found rule), thus similar motions can keep those maximal similarities as common rule even if there are variances among them.

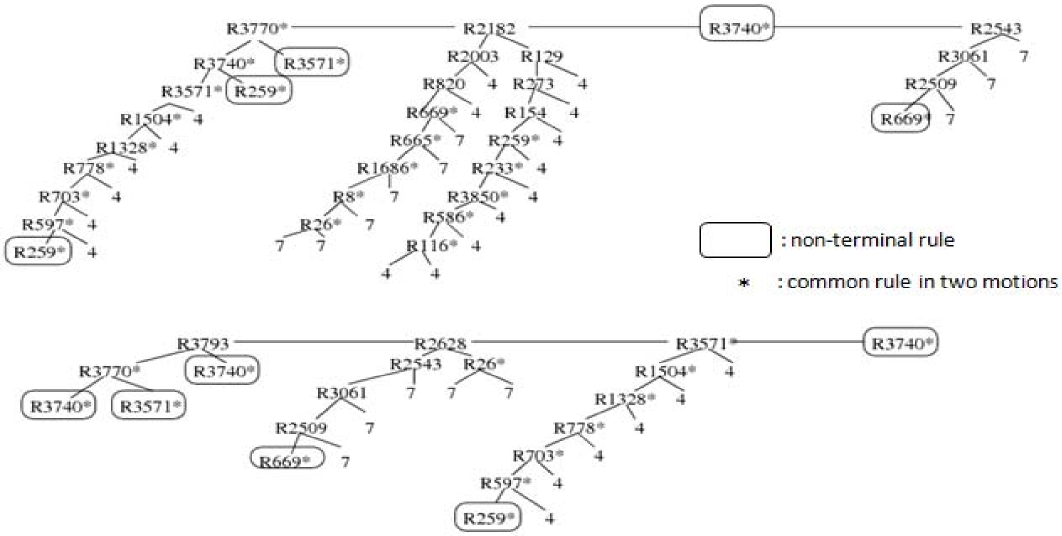


Figure 5. Found Hierarchical Rules between two similar "basketball signal" motions.

We can see that 'SEQUITUR' algorithm with textual sequences and music score shows (see [11]) same effects when it is applied to Human Motion Strings; finding **repetition** units and dealing with **variance** by detecting the maximal common rules. Namely, even if there are variations due to persons, we can extract the maximal common expressions as the detected rules. Using the detected rules as 'terms' in the document, finally we can construct an HMDoc (**H**uman **M**otion **Doc**ument) from a single 3D motion capture clip. Thus, every HMDoc have the same length since each HMDoc consists of rule-frequency values throughout all the rules.

5. Using Latent Semantic Indexing in HMDoc

Originally, LSI technique [2] was introduced for textual document retrieval since "lexical matching" system shows limited performance and it did not extract semantic relations between query and words in document. Furthermore, when people make a text document, they use tremendously diverse words. In that, there are many ways to express a same concept (topic) of a document (synonymy) and most words have various meanings (polysemy). To overcome these weaknesses of "lexical matching" retrieval system, LSI uses conceptual indices which are statistically derived and it tries to disclose latent (hidden) structure in words usage that is obscured by word. We expect same effect with LSI technique for HMDocs in extracting conceptual indices and overcoming diverse rule variances, and apply LSI to HMDoc matrix.

Currently, we transformed 3D complex human motion matrix into rule-based human motion representation forms. After matching rules to terms and a human motion clip to a document, we can construct a matrix (M) of rules by HMDoc Database.

A matrix M consists of $r \times m$; where r is the number of found rules by 'SEQUITUR' and m is the number of HMDoc in the Database.

$$M = [h_{i,j}] \tag{3}$$

Here, $h_{i,j}$ denotes that the frequency of rule i occurring in the j^{th} HMDoc. Basically, to give less weight to those rules which frequently occur in many HMDocs, we apply local and global weighting to each cell of frequency values.

$$M' = [h_{i,j}] = loc(i,j) \times global(i) \tag{4}$$

$$loc(i,j) = rf_{ij} \tag{5}$$

$$global(i) = 1 - \sum_j \frac{p_{ij} \log(p_{ij})}{\log(|HMDocs|)} \tag{6}$$

$$p_{ij} = \frac{rf_{ij}}{gf_i} \tag{7}$$

Where, rf_{ij} is the rule frequency, which is the frequency of rule i in the j^{th} HMDoc and gf_i is the global frequency, which is the all the number of times that rule i appears in the whole HMDoc database. For global weighting, we used the entropy based global weighting, which takes the distribution of rules over HMDocs into account. Normally, HMDoc matrix is the sparse matrix since the number of rules found across all HMDocs is quite larger than the number of rules actually appearing at a particular HMDoc. Actually, term-document matrix is also sparse since every word does not occur in every document. A HMDoc matrix can be decomposed by LSI ($M' = U \sum V^T$) for conceptual indexing and avoiding lexical ('terms' ↔ 'rules') matching limit as Dumais et al. [2] mapped textual documents with LSI mapping. After decomposing into sub-matrices, we are interested only in the HMDoc vectors since we would like to map each HMDoc into the semantic space based on V^T (U, V^T matrix is corresponding to rule and HMDoc vectors respectively). This mapping is done as follow; if we choose the size of truncated matrix (k=2) in the k-semantic space, the first, second column vectors of V can be used as the x, y coordinates values of each HMDoc by multiplying rank-2 singular values (x-coordinate: $\sigma_1 v_1$ y-coordinate: $\sigma_2 v_2$). This multiplication between right singular vectors and the corresponding singular value is the approximating representation of the truncated original matrix $M_k^{'}$ [5]. For some motions, after mapping HMDocs into k-semantic space, other–unrelated motions may still be located closely. To overcome this limit, we try to expand the semantic space dimension size k up to 6. Previously, we only use the first two singular values and the right column vectors and get the x, y coordinates by multiplication ($\sigma_1 v_1$, $\sigma_2 v_2$). Here, we can add additional coordinate values with $\sigma_3 v_3$, $\sigma_4 v_4$ ... $\sigma_6 v_6$. Consequently, we have 6 dimensional coordinate representation of each motion clip. But, in this paper, we limit it to 2-dimensional result for visualization purpose.

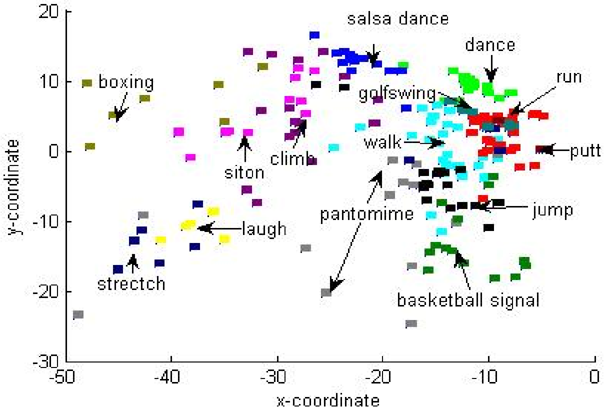


Figure 6. Mapping HMDocs into 2 dimensional Semantic Space (as 'Eigen-Human-Motion' Mapping).

6. Implementation & Performance Evaluation

For experiments, we used the publicly used motion capture data files (CMU Motion Capture Database)[3] and chose 209 human motion clips (370294 frames) as HMDocs in our experiment. Among 209 HMDocs, there are 14 different semantic motion categories, such as 'dance', 'laugh', 'salsa dance', 'pantomime', 'siton & standsup', 'jump', 'golf swing', 'run', 'boxing', 'basketball signal', 'golf putting', 'walk', 'stretching', 'climb'. Semantic classification of these subtle actions is a quite challengeable problem. Some actions are semantically similar (e.g, 'salsa dance' and 'dance', 'golf swing' and 'golf putt') and the motions are length-variant and there are several different actors in the same motion class (e.g., 'pantomime' actions are done by 4 different actors). After translating quantization motion representation into rule-based motion document (which is HMDocs), we can map each HMDoc into semantic subspace using Latent Semantic Indexing technique.

From these results, semantic subspace mapping can recognize highly-semantic 3D human motion capture data quite well in conditions where data is very high-dimensional (62 dimensions) and length-variant data (overall precision and recall values for all action categories are 88.72% and 86.98% -see Table 1).

We compared our performance with Li et al. [15]'s work, which extracts vector feature values from each motion clip (we call it wMSV (weighted Motion Singular Vectors) measure). Although our approach make more compact dimensional representation, our approach shows much better accuracies than wMSV (see Table1). Keogh et al. [11] showed the way of indexing using linearly scaling the time-series data; especially they demonstrated its usefulness in terms of retrieval time efficiency with human motion database. Here, we would like compare with this work in terms of motion classification accuracy. For computing the most similar motion clip in the motion database, we scaled each candidate motion to the upper and lower bound representation by following to Keogh et al.'s method (LB_KEOGH) [11] and computed the distance between the query motion and upper and lower bounded values of each candidate motion data. Same as we did with previous comparison, for classification performance analysis, we used k-NN method (k=1).

This approach considers the time variant of each human motion data. For reducing this effect, Keogh et al. tried to scale two comparing time-series data. Different from our semantic dimensional reduction technique, this approach doesn't use the semantic characteristic of 3D human motion capture data and LB_KEOGH doesn't reduce the dimensions. For computing aspect, our approach can compute the similar motion faster than LB_KEOGH since we reduced the high-dimensional (62) human motion data into 1-dimensional quantization values as the preprocessing step. The classification accuracy of our proposed approach shows much better results than other two works (see Table 1). Muller et al.'s approach showed quite good results since it also extracts each human motion's geometric relational features, which is close to semantic characteristics of each human motion class. For instance, when a human walking, normally hand and leg geometric location has been switched periodically. Whereas Muller et al still keeps 31 dimensional values of a frame as feature vectors, our HMDoc approach can reduce even to smaller representation (as small as x, y coordinates) with showing almost similar precision and recall values. This process makes retrieval speed faster since we just have to compute each motion's coordinate values. Although we simplify human motion expressions, we can achieve more accurate results in precision and recall values (see Fig 7, 8).

7. Conclusion

We proposed a new cross-media retrieval framework for 3D human motion data in this paper. For the dimensionality reduction issue, we can reduce complex (62) dimensions to 1-dimensional quantization representation through semantic quantization process. To transform 1-dimensional human motion strings into latent

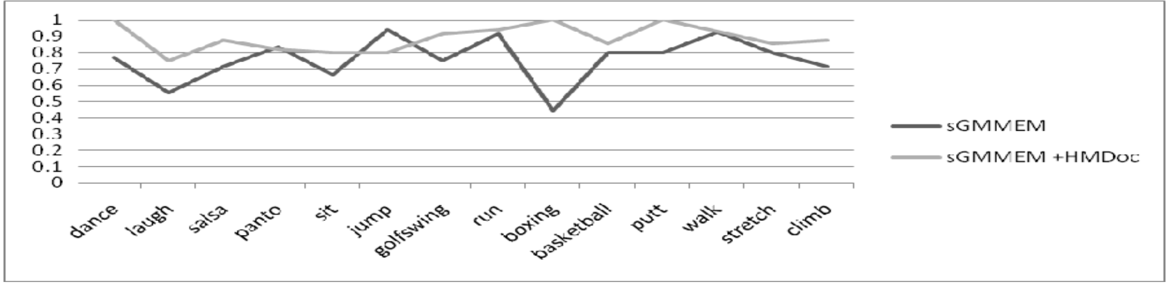


Fig. 7 Precision Improvement by applying HMDoc retrieval technique on semantic GMMEM representation.

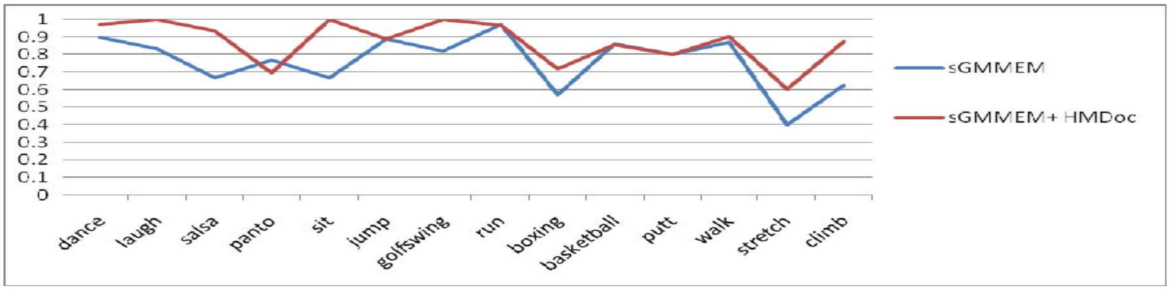
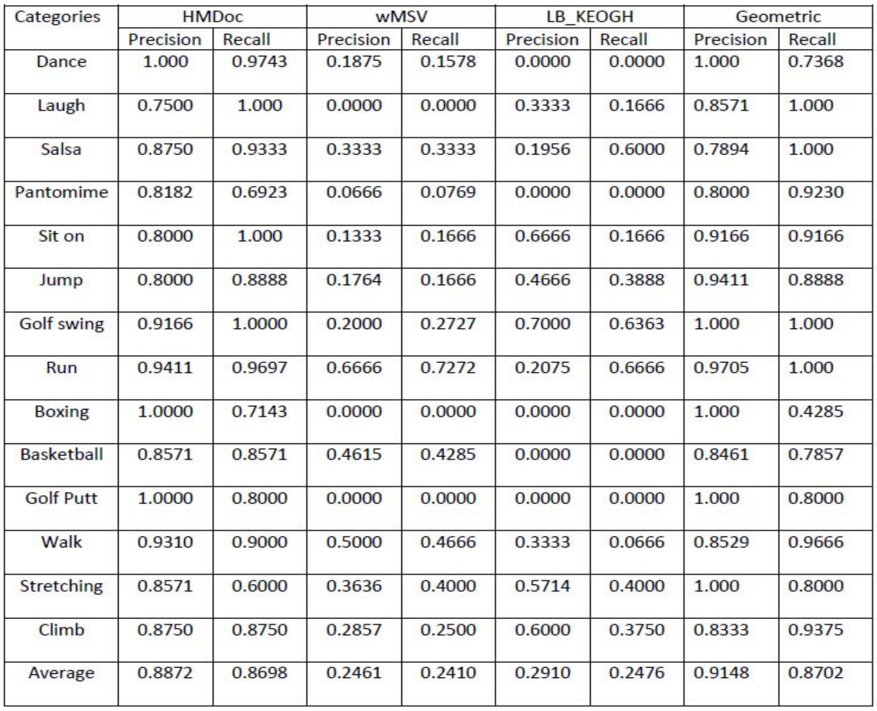


Fig. 8 Recall Improvement by applying HMDoc retrieval technique on semantic GMMEM representation.



Categories	HMDoc		wMSV		LB_KEOGH		Geometric	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
Dance	1.000	0.9743	0.1875	0.1578	0.0000	0.0000	1.000	0.7368
Laugh	0.7500	1.000	0.0000	0.0000	0.3333	0.1666	0.8571	1.000
Salsa	0.8750	0.9333	0.3333	0.3333	0.1956	0.6000	0.7894	1.000
Pantomime	0.8182	0.6923	0.0666	0.0769	0.0000	0.0000	0.8000	0.9230
Sit on	0.8000	1.000	0.1333	0.1666	0.6666	0.1666	0.9166	0.9166
Jump	0.8000	0.8888	0.1764	0.1666	0.4666	0.3888	0.9411	0.8888
Golf swing	0.9166	1.0000	0.2000	0.2727	0.7000	0.6363	1.000	1.000
Run	0.9411	0.9697	0.6666	0.7272	0.2075	0.6666	0.9705	1.000
Boxing	1.0000	0.7143	0.0000	0.0000	0.0000	0.0000	1.000	0.4285
Basketball	0.8571	0.8571	0.4615	0.4285	0.0000	0.0000	0.8461	0.7857
Golf Putt	1.0000	0.8000	0.0000	0.0000	0.0000	0.0000	1.000	0.8000
Walk	0.9310	0.9000	0.5000	0.4666	0.3333	0.0666	0.8529	0.9666
Stretching	0.8571	0.6000	0.3636	0.4000	0.5714	0.4000	1.000	0.8000
Climb	0.8750	0.8750	0.2857	0.2500	0.6000	0.3750	0.8333	0.9375
Average	0.8872	0.8698	0.2461	0.2410	0.2910	0.2476	0.9148	0.8702

Table 1. Performance Comparison with proposed approach (HMDoc) and other works.

semantic indexing applicable format (HMDocs), 'SEQUITUR' has been used since it is quite useful for finding hierarchically repeated rules from 3D human motion data. We observed that newly constructed HMDoc matrix sparse one like textual document data. From LSI, we can overcome lexical matching limit and index conceptually since human motion expressions have subtle variance. This also happens when we try to retrieve with textual document since human writes similar meaning with different words. From these procedures, we demonstrated that 3D human motion data gets benefit of information retrieval techniques fully, such as hierarchically repetition detection ('SEQUITUR') and conceptual indexing ('LSI'). When we increase the k-semantic space dimensions, we can achieve 88.72%, 86.98% precision and recall accuracies.

More than this, we don't need to train our dataset, but reduce the dimensions and map into semantic space. It promises the scalability effect for mapping newly taken motion clips into the pre-existing mapping space incrementally. Different from other multimedia format (images, video and music), with 3D motion data, human body parts are already mapped into columns of motion matrix. Thus, we can get much better effect in bridging the gap between human motion matrix values to higher level semantics.

References

[1] Bishop C. M., "Pattern Recognition and Machine Learning", Springer, 2006.

[2] Berry M., Dumais, S. T. and O'Brien, G. W., "Using linear algebra for intelligent information retrieval", SIAM Review 37(4):573--595, 1995.

[3] CMU Motion Capture Library, http://mocap.cs.cmu.edu/,

[4] Gaurav P., Li C. and Prabhakaran B., "Hierarchical Indexing Structure for 3D Human Motion", Int'l Proceedings of Multimedia Modeling Conference (MMM) 2007,January 9-12, Singapore.

[5] Golub G. H. and Loan C. F., "Matrix Computations", Johns Hopkins University Press, Baltimore, Maryland, 1996.

[6] Gutemberg G. F. and Aloimonos Y., "A Sensory-Motor Language for Human Activity Understanding", In Proc.of the 6th IEEE-RAS International Conference on Humanoid Robots (HUMANOIDS'06), Genoa, Italy, pages 69-75.

[7] Gutemberg G. F. and Aloimonos Y, "A Language for Human Action", IEEE Computer Magazine, 40(5), 2007.

[8] Jin Y. and Prabharakan B., "Semantic Quantization of 3D Human Motion Capture Data through Spatial-Temporal Feature Extraction", Int'l Proceedings of ACM Multimedia Modeling Conference (MMM) 2008.

[9] K. Forbes and E. Fiume, An efficient search algorithm for motion data using weighted PCA, SCA '05: Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation, pp. 67--76

[10] Knees P., Pohle T., Schedl M. and Widmer G., "A Music Search Engine Built upon Audio-based and Web-based Similarity Measures", SIGIR07' Amsterdam, The Netherlands.

[11] Keogh E., T. Palpanas, V. B. Zordan, Gunopulos D., and Cardle M., "Indexing large human-motion databases", Proc. 30th VLDB Conference, pages 780--791, Toronto, Canada, 2004.

[12] Manning N., Witten I., "Identifying Hierarchical Structure in Sequences: a Linear-time Algortihm", Artificial Intelligence Research, Vol 7, 66-82, 1997.

[13] Li C., Gaurav P., Zheng S.Q. and Prabhakaran B., "Indexing of Variable Length Multi-attribute Motion data", In Proc. of the Second ACM International Workshop on Multimedia, Washington D.C., USA, pp. 75-84, November 2004.

[14] Li C., Kulkarni P.R. and Prabhakaran B., "Motion Stream Segmentation and Recognition by Classification", International Journal of Multimedia Tools and Applications (MTAP) by Springer-Verlag, Vol.35(1), October 2007.

[15] Li C., Zheng S. Q. and Prabhakaran B., "Segmentation and Recognition of Motion Streams by Similarity Search", The ACM Transactions on Multimedia Computing, Communications and Applications (ACM TOMCCAP), Vol. 3(3), August 2007.

[16] Liu F., Zhuang Y., Wu F., and Pan Y., "3D motion retrieval with motion index tree", Computer Vision and Image Understanding, 92:265--284, June 2003

[17] Liu G., Zhang J., Wang W., and McMillan L., "A system for analyzing and indexing human-motion databases", In Proc. 2005 ACM SIGMOD International conference on Management of data.

[18] Lucas Kovar and Michael Gleicher, Automated extraction and parameterization of motions in large data sets, SIGGRAPH 2004, pages 559--568.

[19] Monay F., Gatica-Perez D, "On Image Auto-annotation with latent space models", ACM Multimedia 2003, 275-278.

[20] Muller M., Roder T., and Clausen M., "Efficient content based retrieval of motion capture data", ACM Transactions on Graphics (TOG), 24:677.685, 2005.

[21] Meinard Muller and Roder T., Motion templates for automatic classification and retrieval of motion capture data, SCA '06: Proceedings of the 2006 ACM SIGGRAPH/Eurographics symposium on Computer animation, pp. 137—146

[22] NCBI, "National Center for Biotechnology Information", http://www.ncbi.nlm.nih.gov/

[23] Pecenovic Z., "Image retrieval using latent semantic indexing", Final year graduate thesis, AudioVisual Communications Lab, Ecole Polytechnique Federale de Lausanne, Switzerland" , June, 1997.

[24] Souvannanong F., Merialdo B. and Huet B., "Latent Semantic Analysis For An Effective Region-Based Video Shot Retrieval System", MIR'04, October 15-16, New York, USA. 2004.

[25] Souvannanong F., Merialdo B. and Huet B., "Latent Semantic Analysis For Semantic Content Detection of Video Shots", ICME 2004.

[26] Wang T.S., Shum H.Y., Xu Y.Q. and Zheng N.N., "Unsupervised Analysis of Human Gestures", In Proceedings of second IEEE Pacific Rim Conference on Multimedia, 174-181, 2001.

[27] Yasuhiko Sakamoto and Shigeru Kuriyama and Toyohisa Kaneko, Motion map: image-based retrieval and segmentation of motion data, SCA '04: Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation, pages 259-266

[28] Zhai Y. and Shah M., "Determining structure in continuously recorded videos", ACM Multimedia 2005, 495-498.