Data reduction using cluster sampling

Yeseung Park*, Mingyu Jang, Jungwoo Huh, Kyoungoh Lee, and Sanghoon Lee*
* Yonsei University, Seoul, Korea
E-mail: {pys940617, jmg1002, gjwjddn9, kasinamooth,slee}@yonsei.ac.kr

***Abstract*—It is natural to use larger and more diverse datasets to get better performance in pose study. Learning with a large scale is essential to improve the model performance to a level similar to human recognition, but there is a problem with gradually increasing learning time and data redundancy. This can also lead to a lack of data storage. Our study proposes a new way to solve these problems: Data shaping Using Cluster Sampling (DUCS). In this paper, we propose a sampling framework that clusters a pose dataset and extracts only a small number of random frames from each cluster. To ensure the consistency of pose data, the data is normalized, and a preprocessing process of aligning the entire joint based on the pelvic joint is performed, and an optimal parameter search in DBSCAN is proposed to improve the performance of clustering. This process can greatly reduce the redundancy due to the specific posture bias. To demonstrate the effectiveness of our method, we trained a 3D pose estimation model with sampled datasets of Human3.6M and shown competitive results despite the drastic compression rate of over 95%.**

I. Introduction

People's attempts to understand and utilize human posture through computers have continued to this day [16-18], and to achieve better results than ever in the human pose field, most studies use datasets [1-4] that contain a lot of human pose data[19]. It is common practice to gather more data to achieve better performance. However, the size of the actual dataset does not always guarantee better performance [5], [6]. Human3.6M [2] is one of the most widely used dataset for pose research today and is a 'Large scale Dataset' that contains about 3 million frames with pose annotations. However, Human3.6M may be an 'empty suit' in some respects. This aspect can be found in the composition of the database's pose. In the case of Human3.6M, a large proportion of each action sequence is the walking pose (e.g. Walking, Walkdog, Phoning, Eating, Purchases, etc.). Due to the characteristics of this database configuration, it is very likely that the model learned with this dataset will be overfitted to the walking pose and fail to estimate other distinct poses [7]. Even if the size of the database increases, if it is already in the existing database, it will be boring data for the model. Also, since the motion of people is captured with MoCap devices with high frequency, adjacent frames have very similar poses, which causes data redundancy [5], [6]. In this case, it can be counterproductive when learning with these duplicated data.

To solve this problem, most of the studies using pose data have applied various sampling methods to reduce the size of the dataset to improve learning efficiency [8-13]. In the previous study, a study was conducted to reduce the database by randomly extracting frames[13], but did not obtain satisfactory results. Also, the authors of [8], [11] proposed a downsampling method to reduce the amount of data without losing structural characteristics of the data.

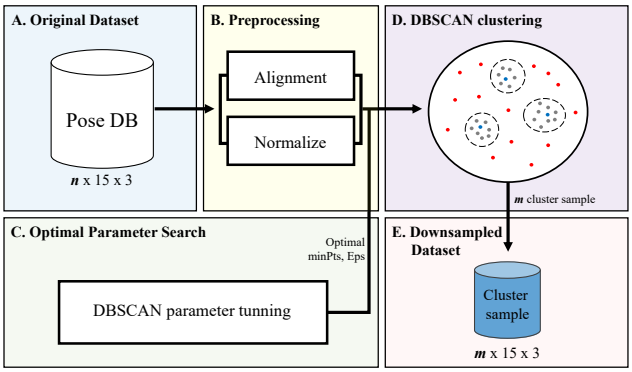

Fig. 1. The overall framework for this study. input data is the original dataset of $n*15*3$(A). Input data is aligned and normalized for data consistency(B). After that, DBSCAN is applied to the data set, and the optimal parameter search process is carried out to select the parameters that produce the most clusters. A small number of data was sampled in the cluster created by the process. The final output is the cluster sample dataset in $m*15*3$ dimension.

We note that this series of sampling processes can eliminate some degree of data redundancy, but it is difficult to eliminate the problem of highly biased data in a specific pose. In our paper, we propose DUCS (Data reduction Using Cluster Sampling), a method of clustering data using DBSCAN clustering and sampling data from clusters with similar characteristics. Specifically we suggest how to find the two most important parameters: Eps (the radius of the cluster) and MinPts (the minimum data objects requested inside the cluster) through the structural characteristics of DBSCAN, which has high pose estimation performance when the number of clusters is high. Despite this intuitive method, models trained with our sampling method show better performance than models trained through other sampling methods. Also, it has a strong advantage that performance does not deteriorate significantly compared to learning was performed using the original database. Finally we conducted a subjective test on how similar the data in the sampling dataset were to confirm that the dataset extracted with the parameters of our choice not only improved the performance of the model but also was similarly accepted human perception.

The main contributions of this paper are as follows.
1. Using the DBSCAN algorithm, the size of the dataset was drastically reduced by eliminating the redundancy and bias of

the dataset.
2. Using DBSCAN's structural features, we found the optimal parameters that are well accepted by model performance and human perception.
3. Using a fast and intuitive method, it shows better performance than conventional random sampling methods.

II. Approach

This section introduces the approach used to proceed with DBSCAN Cluster Sampling.

A. Preprocessing

Due to the nature of 3D pose data, an identical pose can be expressed differently depending on the viewing angle and the height or shape of a person. This characteristic becomes a problem when learning because even in the same pose, it can be recognized as a different pose in the clustering process which leads to performance degradation of the model. Therefore, we proceed to normalize and align pose so that pose can be recognized under the same conditions before clustering. The formula expressing normalize is as follows.

$$norm = \frac{x_i - \min_x}{\max_x - \min_x} \tag{1}$$

x_i is input data, and $\min_x$ and $\max_x$ are min and max values of input data. We normalize the pose by applying the same ratio to all data. Also, we rotate the data based on the left and right pelvic joints of pose data to process all data to have the same angle. Through these methods, a problem in which a similar pose is recognized as a different pose due to an angle can be solved.

B. DBSCAN

Cluster analysis is an unsupervised learning method that classifies data according to their similarity to understand the characteristics of a large database. For most cluster analysis, the k-means clustering method that determines the number of clusters in advance is adopted frequently due to its simplicity. However, because the number of clusters cannot be determined in advance in the pose domain, the DBSCAN algorithm is used. DBSCAN has the advantage of creating an appropriate cluster on its own in the process of clustering without having to determine the size of the cluster in advance. However, the process of tuning parameters to fit dataset has continued in past studies because it is important to properly tune two parameters(minPts, Eps) to make DBSCAN perform well. We noted how to utilize the structural features of DBSCAN instead of learning models or optimizing parameters, and suggested ways to infer appropriate parameters through the variation of model accuracy according to two parameters.

C. Optimal Parameter Search in DBSCAN

We will give a brief explanation of how the number of clusters vary as Eps changes in the DBSCAN algorithm. If Eps is very small, all data are classified as noise because dots cannot be tied to other points. However, when Eps becomes larger, the noise points start to form a cluster, and the number of clusters gradually increases. At this time, if Eps becomes larger than a certain level, the number of noises continues to fall and approaches zero. But clusters that were not bound together join and form a bigger cluster and the total number of clusters also decreases. Also, minPts does not change the overall shape of the graph, and shifts the graph to the right so that the high point of the cluster occurs at a slightly higher Eps. This tendency is shown in Fig.2,3.

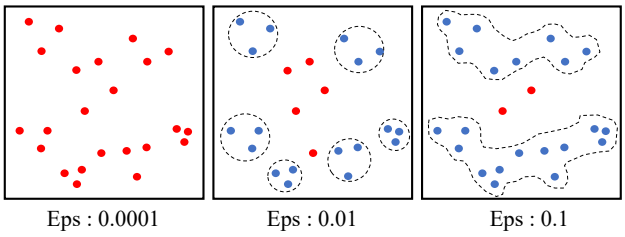

Fig. 2. The tendency of clustering according to Eps.

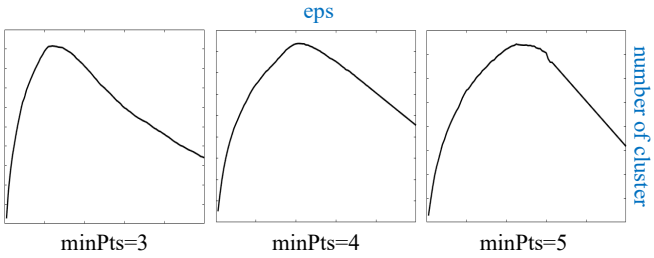

Fig. 3. The tendency of clustering according to minPts.

If Eps is small, the data in a cluster have considerable similarities, but problems arise during sampling because the total number of clusters is too small. In addition, if Eps is too large, the variance of the cluster increases, making it inconsistent with the goal of cluster sampling to eliminate highly redundant data. Thus, with the assumption that adjusting proper Eps would help learning, we tried to prove that the results of the pose estimation at that time were better than those of other parameters by using the DBSCAN parameter when the number of clusters was highest.

III. Experiment

In this study, we perform DBSCAN clustering on Human3.6M, a 3D human pose dataset. After performing the optimal parameter tuning process, we make a Cluster Sampling(CS) dataset by sampling from the clusters formed by our algorithm. Then, we create a Random Sampling(RS) dataset that samples the same number of frames as the CS dataset randomly in raw Human3.6M. Both CS and RS datasets are used to train Simple-net, a 3D human pose estimation deep neural network. We compare and analyze the performance of the two datasets using Simple-net. We have proved the feasibility and effectiveness of our method through empirical experiments.

Data.	MPJPE
full Dataset(389938)	**45.5**
CS[3,0.014,2](10390)	**49.3**
CS[4,0.014,2](8474)	49.4
CS[5,0.014,2](10390)	49.6
RS(10390)	53.4
CS[3,0.014,1](5195)	**52.3**
RS(5195)	56.4

TABLE I
DETAILED SIMPLE-NET RESULTS ON HUMAN3.6M. THE FULL DATASET SHOWS THE RESULTS OF LEARNING USING ALL THE DATA, CS MEANS A CLUSTER SAMPLING METHOD, AND RS MEANS RANDOM SAMPLING. THE VALUES IN SQUARE BRACKETS ARE [MINPTS, EPS, COUNT OF SAMPLING], THE VALUE IN PARENTHESES IS THE NUMBER OF DATA.

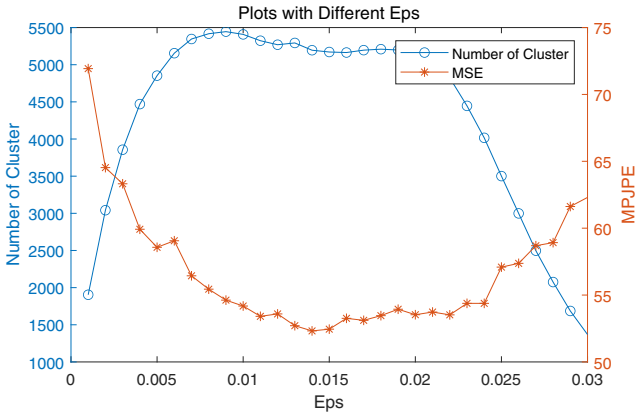

Fig. 4. The number of clusters and Pose estimation Loss according to the parameters of DBSCAN(Eps).

A. Experiment Setting

1) Human3.6M: The Human3.6M dataset is frequently used in 3D human pose estimation, which consists of 3.6 Million 3D poses(frames) of 11 subjects performing 15 different actions under 4 viewpoints. The subjects are recorded from 4 different views with RGB cameras, and the joint positions of the subjects are measured by a MoCap system. The calibration parameters are available for the RGB cameras and MoCap system. In our experiments, we use 5 subjects(S1, S5, S6, S7, S8) for training and validation and 2 subjects(S9, S11) for testing. When using the original Human3.6M dataset, we use about 1.56 million poses(frames) for training and about 0.55 million poses(frames) for testing.

2) Simple-net: To evaluate the performance of clustered data, we used Simple-net, a simple and easy to implement model of 3D pose estimation. Simple-net use normalization, dropout and Rectified Linear Units (RELUs), as well as residual connections. Simple-net basic block is consisted of a linear layer with 1024 hidden units, followed by batch normalization, dropout and a RELU activation. This is repeated twice, and the two blocks are connected with a residual connection.

B. Experiment Result

We compared MPJPE scores obtained through random sampling and cluster sampling. Eps was tested in increments of 0.001 from 0.001 to 0.03 and minPts was tested using three values 3,4,5. Table 1 used $minPts = 3$, $Eps = 0.014$ which obtained the best MPJPE value. As shown in Table 1, our method has recorded better performance than conventional random sampling, a method that has been widely used in all experiments, and even though we have reduced data by more than 95 percent compared to the full dataset, the performance have only dropped slightly. Selecting parameters and applying the DBSCAN algorithm to 390,000 takes only 5 minutes in a CPU:Intel I5-9400 environment. By applying a few parameters selectively, it is easy to find the highlands of DBSCAN, which can dramatically reduce the time required to check the feasibility of the pose estimation model. Qualitative pose estimation to this model results are shown in Fig. 4. Depending on the distribution characteristics of DBSCAN, the number of clusters tends to increase gradually and then decrease from a certain level. In our study, clustering through DBSCAN, and sampling in that cluster can reduce redundancy of the pose data. In the process, data with small redundancy is used for learning, so DBSCAN with appropriate parameters can improve the performance of the pose estimation. Fig.4 shows that the loss is optimized when the appropriate parameter is selected and the loss increases when it is not the appropriate parameter.

C. Subjective Test

A subjective test was conducted to verify that the dataset extracted with parameters found through the optimal parameter search(OPS) was properly clustered from the human perspective. Poses were extracted from a various set of eps between 0.001 and 0.03, and mean(μ) and variance(σ) of the cluster were calculated. After that, the poses were extracted as 5 steps of [μ-2σ, μ-σ, μ, μ+σ, μ+2σ] to evaluate how similar they were to the subject on a 5-point scale. Highly similar clusters will receive high scores because σ is not large, but clusters with unsimilar poses will receive low scores.

The above experiment was conducted on a total of 30 people, and each of the 30 questions was graded on a five-point Likert scale. The items of the scale are [1:Very unsimilar, 2:unsimilar, 3: Neither similar nor unsimilar, 4: similar, and 5: Very similar]. The results of Fig.6 determined that the subjects thought similar data were gathered up to a certain level of Eps. (the green area of Fig.6). On the other hand, when evaluating clusters with a certain level of Eps or higher, most of the subjects felt that the interior of the cluster was not similar, which means the clustering did not function properly.

IV. LIMITATION

In this study, we were limited in two parts, the first of which failed to apply the model to various datasets. Our main dataset is the Human3.6M dataset, which is a great dataset with very large and diverse information, but it is difficult to predict pose data in a variety of real world situations using only this dataset.

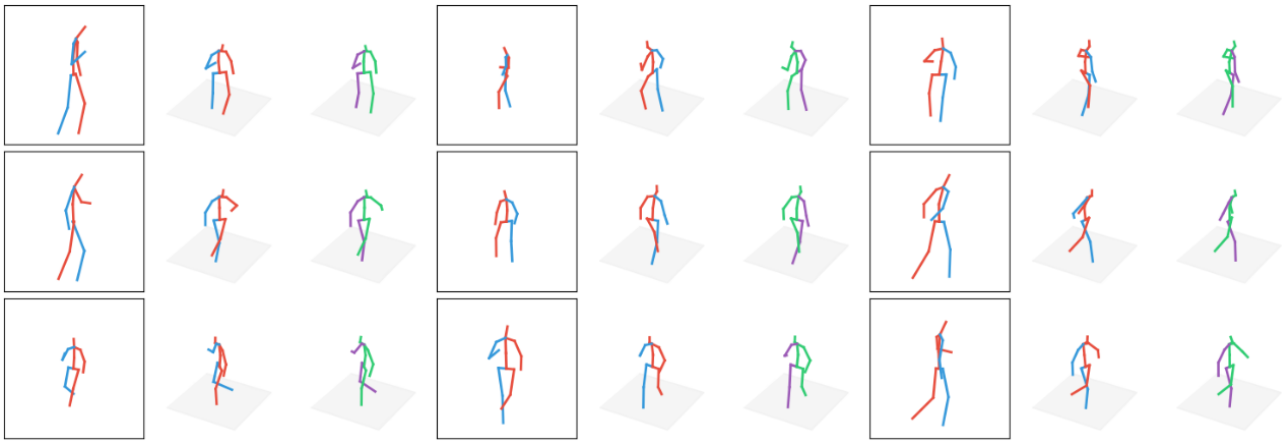

Fig. 5. Example output on the test set of 10390 Cluster Sample(CS) of Human3.6M.Left: 2d observation. Middle: 3d ground truth. Right(green): 3d predictions. It is a visualization of how well the learning data follows the GT, using the ground truth data information of the 3D pose as the correct answer.

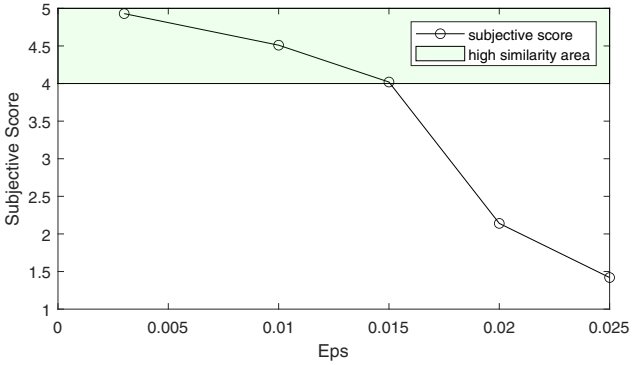

Fig. 6. Subjective score according to Eps. Fig.6 shows the results of subjective tests on the DBSCAN sampling dataset conducted with 5 Eps, and the subjects feel that the data in the cluster was similar up to Eps: 0.015 level, but over 0.015, they don't feel similarity.

We should have applied it to a wider variety of datasets to show that our model is robust to a variety of environments, but it was difficult to apply in this area. Another part is that we ignored the noise of DBSCAN when sampling. At the beginning of DBSCAN, noise is literally noise, but from the moment the cluster peaks, noise becomes 'unique' data that is not tied to any pose. We wanted to put noise in this model, but when Eps were low, the number of noise was too high to reduce data, and we didn't know which parameters made the optimal number of noise, so we wanted to find out in the next study.

V. Conclusions

We introduced a method of removing redundancy and bias in the dataset by sampling the highly redundant and biased Pose dataset by DBSCAN. Also, we studied the method of estimating optimal DBSCAN parameter. In particular, by analyzing the distribution characteristics of DBSCAN, we found that the higher the number of clusters generated by the DBSCAN algorithm, the more likely it is that various pose data are extracted. To prove this, we have numerically shown that our method is an effective method by comparing the 3D pose estimation performance when the optimal parameter is selected and when the other parameters are selected. However, this paper has two limitations. The first is that it has not been applied to various models and datasets, and the second is that it has failed to analyze noise, an important element of DBSCAN. These problems will be solved in the next study.

Acknowledgment

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No.NRF-2020R1A2C3011697).

References

[1] ANDRILUKA, Mykhaylo, et al. 2d human pose estimation: New benchmark and state of the art analysis. In: Proceedings of the IEEE Conference on computer Vision and Pattern Recognition. 2014. p. 3686-3693.
[2] IONESCU, Catalin, et al. Human3. 6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. IEEE transactions on pattern analysis and machine intelligence, 2013, 36.7: 1325-1339.
[3] MEHTA, Dushyant, et al. Monocular 3d human pose estimation in the wild using improved cnn supervision. In: 2017 international conference on 3D vision (3DV). IEEE, 2017. p. 506-516.
[4] LASSNER, Christoph, et al. Unite the people: Closing the loop between 3d and 2d human representations. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 6050-6059.
[5] REN, Kun; THOMSON, Alexander; ABADI, Daniel J. An evaluation of the advantages and disadvantages of deterministic database systems. Proceedings of the VLDB Endowment, 2014, 7.10: 821-832.
[6] NAJAFABADI, Maryam M., et al. Deep learning applications and challenges in big data analytics. Journal of Big Data, 2015, 2.1: 1.
[7] HAWKINS, Douglas M. The problem of overfitting. Journal of chemical information and computer sciences, 2004, 44.1: 1-12.
[8] ZHOU, Xiaowei, et al. Sparseness meets deepness: 3d human pose estimation from monocular video. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 4966-4975.
[9] CHEN, Ching-Hang; RAMANAN, Deva. 3d human pose estimation= 2d pose estimation+ matching. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017. p. 7035-7043.

[10] TOME, Denis; RUSSELL, Chris; AGAPITO, Lourdes. Lifting from the deep: Convolutional 3d pose estimation from a single image. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017. p. 2500-2509.
[11] LIN, Mude, et al. Recurrent 3d pose sequence machines. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 810-819.
[12] MEHTA, Dushyant, et al. Vnect: Real-time 3d human pose estimation with a single rgb camera. ACM Transactions on Graphics (TOG), 2017, 36.4: 1-14.
[13] MORENO-NOGUER, Francesc. 3d human pose estimation from a single image via distance matrix regression. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017. p. 2823-2832.
[14] ESTER, Martin, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In: Kdd. 1996. p. 226-231.
[15] RUMELHART, David E.; HINTON, Geoffrey E.; WILLIAMS, Ronald J. Learning internal representations by error propagation. California Univ San Diego La Jolla Inst for Cognitive Science, 1985.
[16] LEE, Kyoungoh; LEE, Inwoong; LEE, Sanghoon. Propagating lstm: 3d pose estimation based on joint interdependency. In: Proceedings of the European Conference on Computer Vision (ECCV). 2018. p. 119-135.
[17] Lee, Inwoong, et al. "Ensemble deep learning for skeleton-based action recognition using temporal sliding lstm networks." Proceedings of the IEEE international conference on computer vision. 2017.
[18] Kwon, Beom, Junghwan Kim, and Sanghoon Lee. "An enhanced multi-view human action recognition system for virtual training simulator." 2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA). IEEE, 2016.
[19] Kwon, Beom, and Sanghoon Lee. "Human Skeleton Data Augmentation for Person Identification over Deep Neural Network." Applied Sciences 10.14 (2020): 4849.