Motion Regeneration using Motion Texture and Autoencoder

Suekyeong Nam
Department of Computer Science and Engineering
Kyunghee University, Republic of Korea
akasha989@khu.ac.kr

Seungkyu Lee
Department of Computer Science and Engineering
Kyunghee University, Republic of Korea
seungkyu@khu.ac.kr

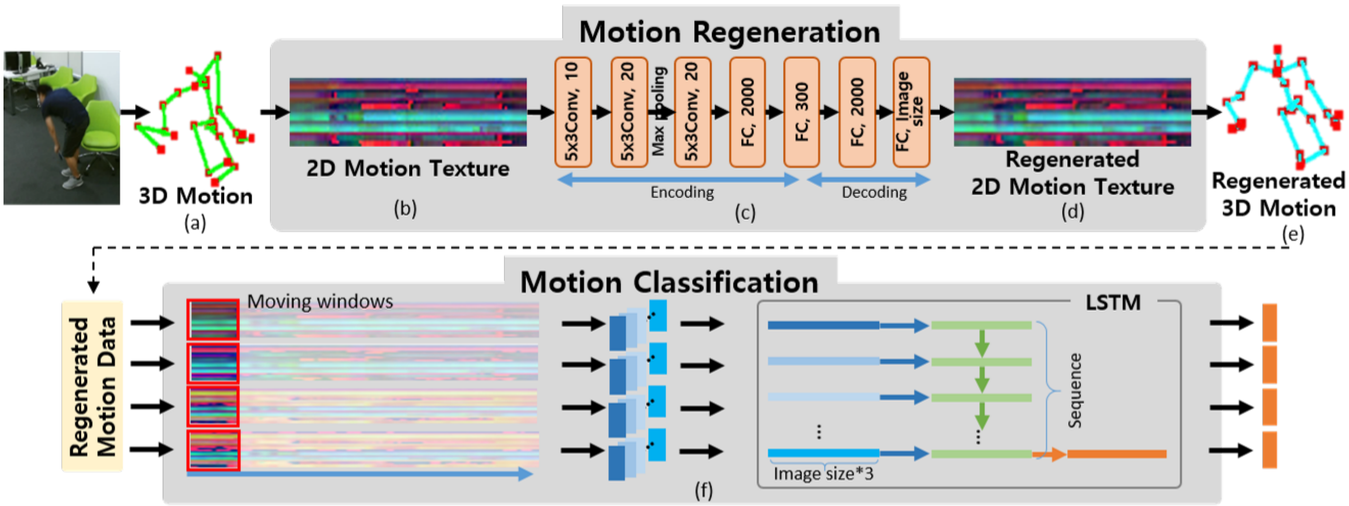


Figure 1: Proposed framework of regeneration and classification. Top row shows motion regeneration with motion texture. 3D motion described by skeletal dataset are converted to 2D motion textures shown in (b). Regeneration using autoencoder synthesizes new motion texture removing statistically meaningless features as shown in (d). Bottom row shows evaluating framework of proposed regenerated motion. We perform classification using Long-Short term memory(LSTM)(f).

ABSTRACT

Motion analysis and recognition frequently suffer from noisy motion capture data not only because of systematic noises of imaging devices but also because of motion dependent non-systematic errors such as self occlusions and motion dynamics extraction failure from visual data. In this work, we propose a motion regeneration method that extracts only statistically significant and distinct characteristics of human body motion and synthesizes a new motion data. To this end, we convert 3D human body motion to 2D motion texture that is easily applicable to well-trained deep convolutional network. An autoencoder is trained with our 2D motion textures to learn only essential characteristics of human body motion in encoded space discarding systematic noises and unexpected non-systematic errors that are nothing to do with the description of particular motion. For the verification of the effectiveness of our regenerated motion, we perform motion classification test on public body motion dataset using our Long-Short Term Memory(LSTM) based method.

CCS CONCEPTS

•**Computing methodologies** → Activity recognition and understanding; Dimensionality reduction and manifold learning; Motion capture;

KEYWORDS

Motion regeneration, Motion texture, Autoencoder

ACM Reference format:
Suekyeong Nam and Seungkyu Lee. 2018. Motion Regeneration using Motion Texture and Autoencoder. In *Proceedings of SA '18 Posters, Tokyo, Japan, December 04-07, 2018,* 2 pages.
DOI: 10.1145/3283289.3283334

1 INTRODUCTION

Human body motion has been represented, recognized, and synthesized based on body joint dynamics. Extracting joint movements from video is a challenging task due to the difficulty of extracting and tracking body joint locations from visual data. Recently, Kinect has been widely used to recognize and understand body motion that directly obtains each body joint movement. Its quality is limited because of errors in depth estimation, joint fitting on body volume, and self occlusion. However, its cheap and simple hardware compared to conventional expensive motion-capture devices makes it attractive and popular in motion capturing and synthesis industry such as animation and interactive games. Body joint movement obtained by Kinect contains two types of errors: systematic noises of the device and non-systematic errors depending on particular body motion or environment such as self occlusions. In order to remove the errors, previous methods employ supervised learning. However they are not applicable when ground truth motion data is not given. In this paper, we propose a motion regeneration method extracting statistically significant characteristics of each joint movement and synthesizing a new motion data. Body joint motion represented by 3D location is converted to 2D motion texture [Ke et al. 2017]. And then, convolutional autoencoder is trained with motion textures of all motion types which removes systematic noise [Holden et al. 2015] as well as statistically meaningless motion segment in the motion textures.

2 PROPOSED METHOD

2.1 Motion Texture

2D motion texture is obtained by calculating relative joint positions from anchor joints holding absolute location invariance in the representation [Ke et al. 2017]. Four joints having minimal movements in all body motions are chosen as anchor joints. Four sets of relative position values are represented as respective 2D pattern (temporal and joint index axes). Since relative position is calculated in three (x,y,z) coordinates separately, we come to obtain three channels of motion textures.

2.2 Training Autoencoder

Now, a convolutional autoencoder is trained with motion textures of all motion types. The structure of autoencoder used in our framework is shown in figure 1 (c). Encoding part consists of convolution and max pooling layers reducing the dimension of motion textures. And then decoding part simply consists of fully connected layers restore original texture size from encoded space.

2.3 Motion Classification

LSTM for motion classification (figure 1 bottom row) uses sequential data for training. We take 22 sequential texture patches (24*12 size) out of regenerated motion texture with highlighted moving window in figure 1 (f). Each patch works as a very short-term motion frame. We concatenate joints in the texture patch resulting in 1-dimensional vector. Since we have four types of motion textures for one motion, we adopt multi-task learning networks [Ke et al. 2017].

3 EXPERIMENTAL RESULTS

Quantitative experimental evaluation is performed by motion classification test between original and regenerated motion textures using Long-Short Term Memory(LSTM). We modify multi-task learning networks [Ke et al. 2017] combined with LSTM. We use ntu rgb+d dataset [Shahroudy et al. 2016] which consists of 56,000 sequences, 60 classes with 80 views. Figure 2 shows original and regenerated motion texture samples. As indicated by the circles, noises and unexpected movement are removed. Usually in the motion texture, sudden color change along the temporal axis (horizontal axis) happens when there is unrealistic joint movement due to self occlusion. Figure 3 summarizes motion classification

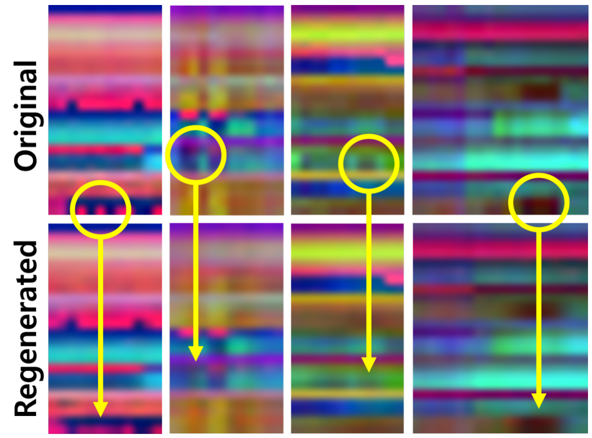


Figure 2: Original and regenerated motion texture samples using convolutional autoencoder: Yellow circle indicates removed or smoothed motion in original motion texture.

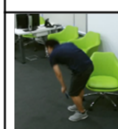
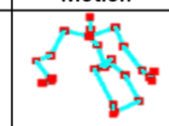
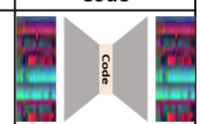
	Original Motion	Regenerated Motion	Regenerated Code
			
Accuracy	73.4%	75.8%	79.0%

Figure 3: Motion classification result using original and regenerated motion textures

results with original motion texture, regenerated motion texture, and encoded feature of the autoencoder. 3D motion sample in figure 3 shows that the pose errors in original data such as elbows and shoulders is corrected after regeneration. Regenerated motion texture obtains improved classification accuracy (73.4% $\rightarrow$ 75.8%). Although autoencoder is incapable of removing all such non-systematic errors, improvement of the classification accuracy shows that there has been improvement in the quality of motion data. We also conduct motion classification using encoded motion features showing further improved classification accuracy (75.8% $\rightarrow$ 79.0%). This is competitive accuracy compared to [Ke et al. 2017]. This means that autoencoder creates optimal set of low-dimensional features to classify body motions.

REFERENCES

Daniel Holden, Jun Saito, Taku Komura, and Thomas Joyce. 2015. Learning motion manifolds with convolutional autoencoders. In *SIGGRAPH Asia 2015 Technical Briefs*. ACM, 18.

Qiuhong Ke, Mohammed Bennamoun, Senjian An, Ferdous Sohel, and Farid Boussaid. 2017. A new representation of skeleton sequences for 3d action recognition. In *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*. IEEE, 4570–4579.

Amir Shahroudy, Jun Liu, Tian-Tsong Ng, and Gang Wang. 2016. NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.