3D High-quality Textile Reconstruction with Synthesized Texture

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 108C (2017) 355–364

Procedia Computer Science

International Conference on Computational Science, ICCS 2017, 12-14 June 2017, Zurich, Switzerland

3D High-quality Textile Reconstruction with Synthesized Texture

Pengpeng Hu[1], Taku Komura[2], Duan Li[1], Ge Wu[1] and Yueqi Zhong[1, 3, *]

[1]*Donghua University, Shanghai 201620, China*

[2]*Edinburgh University, Edinburgh EH8 9AB, UK*

[3]*Key Lab of Textile Science and Technology, Ministry of Education, China*

Abstract:

3D textile model plays an important role in textile engineering. However, not much work focus on high-quality 3D textile reconstruction. The texture is also limited by photography methods in 3D scanning. This paper presents a novel framework of reconstructing a high-quality 3D textile model with a synthesized texture. Firstly, a pipeline of 3D textile processing is proposed to obtain a better 3D model based on KinectFusion. Then, convolutional neural networks (CNN) is used to synthesize a new texture. To our best knowledge, this is the first paper combining 3D textile reconstruction and texture synthesis. Experimental results show that our method can conveniently obtain high-quality 3D textile models and realistic textures.

Keywords: Textile texture,3D scanning,Convolutional neural networks

1 Introduction

3D scan technology has been widely used in textile industry to capture the 3D geometric shapes of clothes and human body. Texture, as one of the most important features for textiles, is still a challenging task in today's textile industry. One important parameter of textile is the shape in three dimensions, which can give human more intuitive information than that from two dimension images. With the rapid development of commercial depth cameras, e.g. Kinect, commercial systems for human body scanning based on depth cameras have been proposed [1,2]. However, not much previous work focus on high-quality 3D textile reconstruction, and the texture usually is captured by the RGB camera, which is limited by the photography method. To visualize different textures for 3D textile is very interesting and useful in textile engineering.

In today's textile industry, the texture is mainly designed by the artist which is expensive and time-consuming. An increasing number of people, especially young people, like having their own pattern on the T-shirt which is called cultural T-shirt. Hence, there is a permanent interest in the development of rapid and automatic texture generation method for the textile. However, human is extremely sensitive to the texture of textiles. It is not trivial that a realistic 3D textured textile model should be obtained.

*Corresponding author: zhyq@dhu.edu.cn

1877-0509 © 2017 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the scientific committee of the International Conference on Computational Science
10.1016/j.procs.2017.05.004

This paper presents a novel framework for high-quality 3D textile reconstruction with synthesized texture. Our framework is shown in Fig.1. An efficient pipeline is designed for 3D textile reconstruction based on KinectFusion. The convolution neural networks (CNN) is used to build a texture model to synthesis a new textile texture. To our best knowledge, it is the first paper to combine 3D textile reconstruction with textile texture synthesis.

2 Overview of the Framework

Our goal is to develop a framework of reconstructing 3D textile models with synthesized textures, as shown in Fig.1. We start by capturing the 3D textile model using KinectFusion with the help of a turntable (Section 3.1). Then a pipeline is designed for 3D textile mesh processing to increase the quality of textile shape (Section 3.2). We use the feature space provided by CNN to build the parametric texture model (Section 3.3 and Section 3.4), and then a new textile texture will be synthesized from the original texture (Section 3.5).The rest of this paper is structured as follows: Section 4 shows the results; Section 5 is the discussion and conclusion.

3 Methodology

3.1 3D Scan

KinectFusion is a popular single-Kinect 3D scanning system, which can be used for Kinect V1 and Kinect V2 [3]. Kinect V2 is chosen in our work as it has a higher depth resolution. We ensure the distance between Kinect v2 and the object to be no more than 1.0 m to obtain good depth resolution according to the work of Butkiewicz et al.[4]. As KinectFusion always collapses during scanning large objects, like the human body, and moving the Kinect at a constantly slow speed will obtain a better reconstructed 3D model. A turntable is used as the platform, which the object is located on, to increase the accuracy and robustness of KinectFusion. It costs 30 seconds to scan the whole object while the turntable rotates a round.

3.2 3D Mesh Processing

With the help of the turntable, 3D textiles can be easily obtained. Without loss of generality, we take reconstructing trousers model as an example to further explain our method. It is useless to scan a flat garment laid on the ground. Therefore, a human body dressing trousers is scanned and the trousers model is extracted next. The 3D model from KinectFusion is shown in Fig.2 (a). It can be seen that the 3D model is not perfect. It needs to be further processed, which is called 3D mesh processing. In our pipeline, as shown in Fig.1 (a), the noises and outliers should be removed firstly. Bilateral Mesh Denoising [5] is used in this step. As KinectFusion merges too many frames of depth images, redundant data cannot be avoided. Too large data is also not conveniently used in practice. Secondly, 3D model is subsampled. Poisson-disk sample [6] is used in this step. The occlusion is a generic issue in almost all 3D scanning work, especially when we scan garments and bags in our work. Thirdly, a hole-filling algorithm [7] is applied to get complete 3D textiles. Next, the surface of 3D textile is smoothed [8]. Universally, a pretty good 3D model can be obtained through these four steps, as shown in Fig.2 (b). However, due to features of textiles, these processes are not sufficient. Here, we just list part of textile specificities:

- **Diversity**: Containing various kinds of objects, e.g. garments, bags, cloths and so on. So textiles have many different topologies;

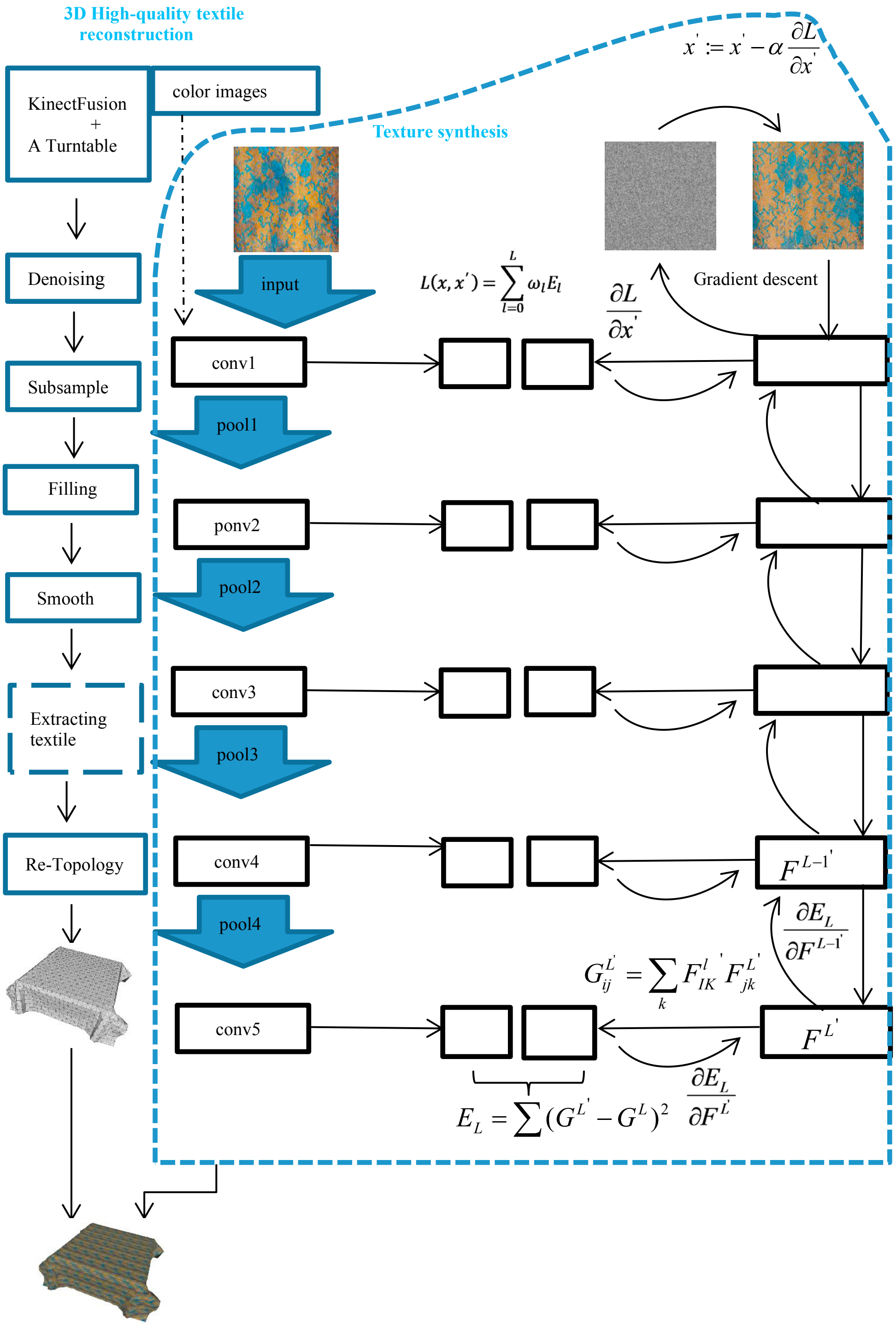


Figure 1. The framework of 3D high-quality reconstruction with synthesized texture. A better 3D textile model is obtained based KinectFusion using our method (left). The texture is generated using convolutional neural network (right). An initial noise image x is passed through the CNN and a loss function E_l is computed on each layer of

the texture model. L is a weighted sum of the contributions E_l from each layer. A new textile texture is found by producing the same Gram matrices $G_l^{'}$ as the original texture.

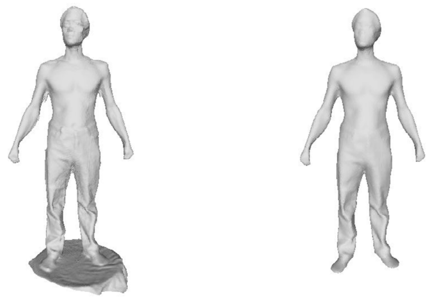
- **Non-rigidity**: Textiles are non-rigid, and once the geometry of the textile is changed by external force it is impossible to recover its original shape;
- **Utility:** Textile products are designed for meeting kinds of needs, like keeping warm, protecting and so on. So the 3D model of textile in usage is needed.

Majorities of textile products are scanned together with other accessories due to above features, e.g. the garment is dressed on a body for 3D scanning; a piece of cloth is laid on a platform to drape for 3D scanning. Therefore, differing from generic 3D mesh processing, the trousers extraction from the human body model is very important. As the data structure of 3D model is triangle soup, zigzag edges along the cutting line will happen. As shown in Fig.3 (a), the waist of trousers and leg openings have bad shape. Our solution to this issue is to do re-topology for the trousers model.

Meshing algorithms can be classified into local and global methods. The former are usually simple, robust and scalable. Global algorithms solve optimization problems whose size depends on the entire dataset. In our work, a re-topology method called Instant Field-aligned Meshing [9] is chosen. It compute a mesh that is globally aligned with a direction field using local orientation and position-field smooth operators. The mesh is then extracted from the fields and optionally post-processed. Comparing to other re-topology methods, Instant Field-aligned Meshing has following advantages:

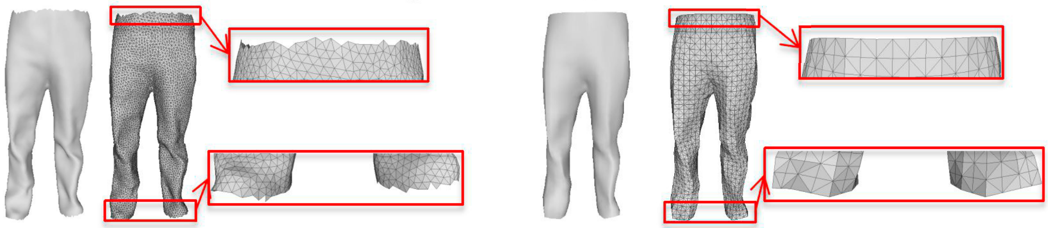
- This method is simple to implement and parallelize, and it can process a variety of input surface representations, such as point clouds, range scans and triangle meshes;
- This method can process extremely large meshes and point clouds with sizes exceeding several hundred million elements, as it avoids any global optimization;
- This method is interactive.

Due to these features, Instant Field-aligned Meshing algorithm is used to implement re-topology in our work. The final trousers model using our method is shown in Fig.3 (b). It can be seen that the waist of trousers and leg openings are perfect.



(a) 3D model from KinectFusion (b) 3D model after generic processing

Figure 2.3D scanning results



(a). Trousers using generic 3D mesh processing (b). Trousers using our method

Figure 3. 3D trousers model

3.3 Convolution Neural Network

Convolutional neural network has proven to be excellent in feature extracting. In our work, we apply the VGG-19 network, which is trained on objection recognition [10]. Here we give only a brief summary of its architecture.

There are 16 convolutional and 5 pooling layers in the VGG-19 network. We just use parts of the full layers. The network's architecture is based on two fundamental computations:

(1) The convolution is linearly rectified with filters of size $3\times3\times k$ where k is the number of output per layer. The strider and padding are both set to one so that the output image has the same spatial dimensions as the input image, which satisficed the equation 1. Equation 1 describes the relationship among the hyper-parameters, where W denotes the dimension of input image, F is the dimension of filter, P represents the padding value, S means the stride, and n denotes the dimension of output image.

(2) Pooling layer is maximum pooling in non-overlapping 2×2 regions, which down-samples the feature maps by a factor of two. This is used for reducing the information to make the training more efficient.

$$n=\frac{W-F+2P}{S}+1 \qquad (1)$$

The overall architecture of VGG-19 is shown in Fig.1 (b). Convolutional layers and max-pooling layers are connected in an alternating manner. As we use only the convolution layers, the input images can be arbitrarily large without considering the input image dimension change. The first convolution layer has the same size as the image and for the following layers the ratio between the feature map sizes remains fixed. We usually believe that each convolutional layer defines a non-liner filter, and that is why this kind of neural network is called convolutional neural network.

The trained convolutional network is publicly available and supported by the Caffe-framework [11]. For texture generation, L Gatys et al. reports that using average pooling will improve the gradient flow and slightly cleaner results can be obtained [12]. Hence, the images shown blow were generated with average pooling. Finally, for practical reasons, the weights in the network were rescaled such that the mean activation of each filter over images and positions is equal to one.

3.4 Textile Texture Model

Similar to the texture model proposed by Portilla and Aimoncelli [13], we define the texture model. The main difference to their work is that instead of using linear filter bank and a set of carefully chosen summary statistics, we use feature space provided by convolutional neural network and only one spatial summary statistic: the correlations between feature responses in each layer of the network. We first extract features of different sizes homogeneously from original image. Then we compute a spatial summary statistic on the feature responses to get a stationary representation of the original image, shown in Fig.1 (b). Finally we synthesis a new image with the same stationary representation by performing gradient descent on a random image which has been initialized with white noise, as shown in Fig.1 (b).

The original texture is denoted as x in our model, we first pass x through the convolutional neural network and computer the activations for each layer l in the network. A layer with N_l distinct filters has N_l feature maps each of size M_l when vectorized. We reorganize these features in a matrix $F^l \in R^{N_l\times M_l}$, where F^l_{jk} is the activation of the j^{th} filter at position k in layer l. Textures are per definition stationary, so a texture model needs to be agnostic to spatial information. A summary statistic that discards the spatial information in the feature maps is given by the correlations between the responses of different features. These feature correlations are given by the Gram matrix $G^l \in R^{N_l \times N_l}$, where G^l_{ij} is the inner product between feature map i and j in layer l:

$$G^l_{ij} = \sum_k F^l_{ik} F^l_{jk} \quad (2)$$

A set of Gram matrices $\{G^1, G^2, ..., G^L\}$ from the L layers in the network responding a given texture provides a stationary representation of the texture, which fully specifies a texture in our model.

3.5 Textile Texture Generation

As we described above, we use gradient descent from an initializing white noise image to find a synthesized image that matches the Gram-matrix representation of the original image. This optimization is implemented by minimizing the mean-squared distance between the entries of the Gram matrix of the original image and the Gram matrix of the image being generated, as shown in Fig.1 (b).

Let x and $x^{'}$ be the original image and the generated image, then G^l and $G^{l^{'}}$ denote their respective Gram-matrix representations in layer l (Eq.2). The contribution of layer l to the total loss is then

$$E_l = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} (G^l_{ij} - G^{l^{'}}_{ij})^2 \quad (3)$$

and the total loss is

$$L(x, x^{'}) = \sum_{l=0}^{L} \omega_l E_l \quad (4)$$

Where ω_l are weights of the contribution of each layer to the total loss. The derivative of E_l with respect to the activations in layer l can be computed analytically:

$$\frac{\partial E_l}{\partial F^{l^{'}}_{ij}} = f(x) = \begin{cases} \frac{1}{N_l^2 M_l^2} \left(\left(F^{l^{'}} \right)^T \left(G^l - G^{l^{'}} \right) \right)_{ji} & if\ F^{l^{'}} > 0 \\ 0 & if\ F^{l^{'}} < 0 \end{cases} \quad (5)$$

The standard error back propagation [14] can be used to compute the gradients of E_l with respect to the pixels $x^{'}$. In our practice we use L-BFGS [15], which can work well for the high-dimensional optimization problem. The whole procedure relies mainly on the standard forward-backward pass used to train the convolutional network. With the help of GPUs and performance-optimized toolboxes for training convolutional neural network [16], textile texture generation can be done in reasonable time.

4 Results

To evaluate the proposed method, we first scan some textile objects to obtain their 3D models. The original textures come from Texture website [17]. These images consists of 9 classes of textile texture, e.g. camouflage and leather textures. Also, our 3D scanning system can get RGB information. We capture some real sample cloths to enrich the textile textures. The final 3D textured trousers can be seen in Fig. 4.

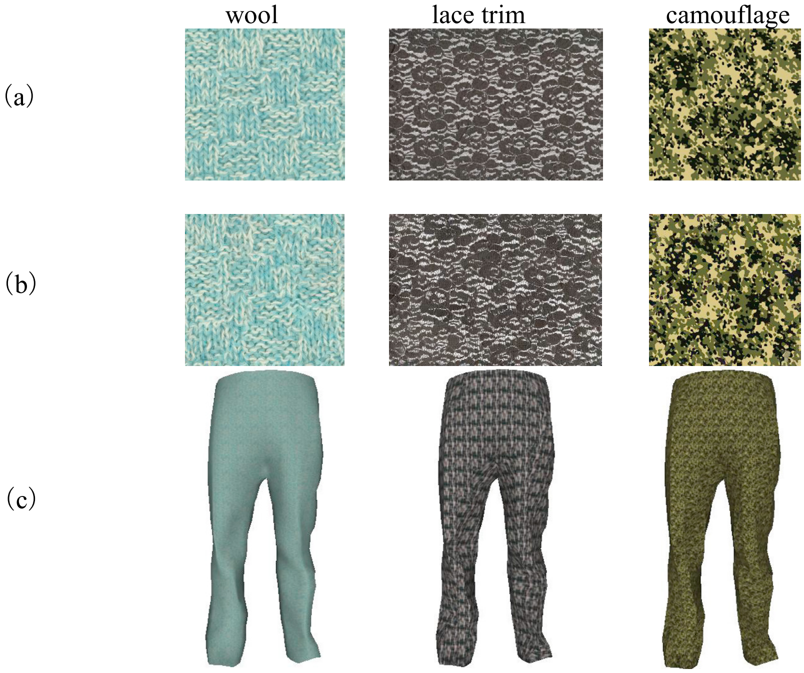
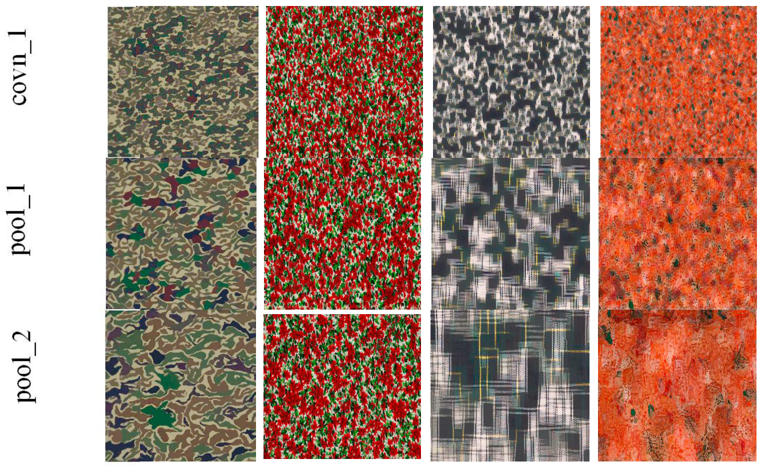


Figure 4. Images of the first row (a) are the original textile textures; Images of the second row (b) are the synthesized textile textures; (c) shows the 3D textured trousers model.

We further show the textile textures generated by our method using different number of layers that is used to constrain the gradient descent, as shown in Fig.5. It means that the images in the first row were generated only from the texture representation of the first layer (‘conv1_1’) of the VGG-19 network. The images in the second row were generated by jointly representations of the layers of ‘conv1_1’, ‘conv1_2’ and ‘pool1’. From Fig.5, we can see that the textile texture generated by constraining all layers to layer ‘pool4’ are nearly indistinguishable from the original texture (Fig.5, the last row).More results are shown in Fig.6.



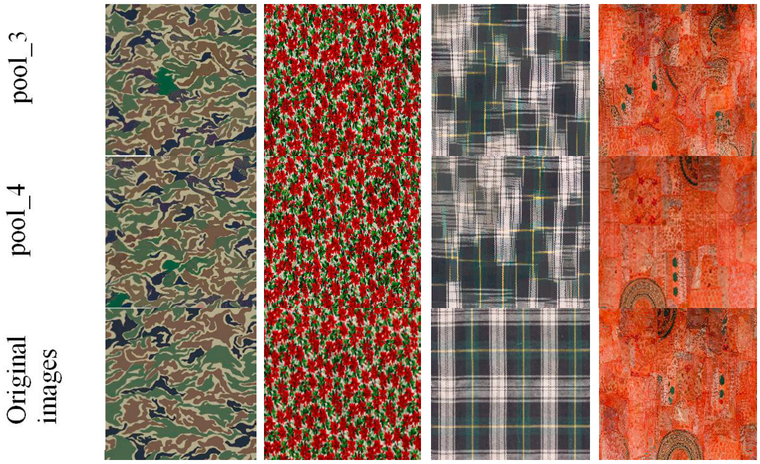
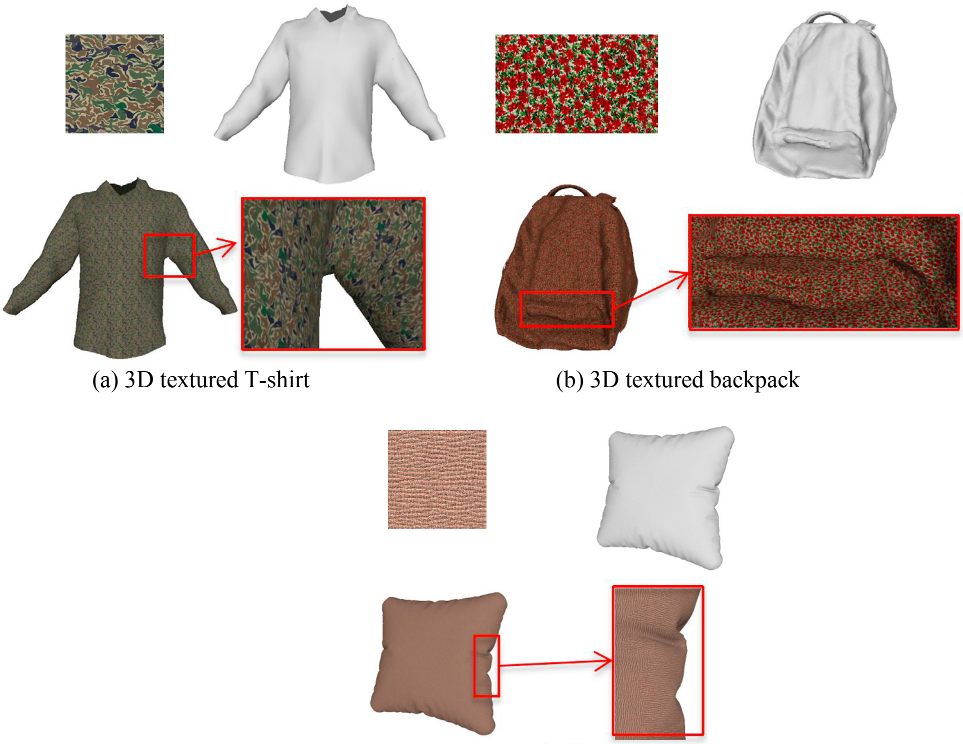


Figure 5. Each row corresponds to a different processing stage in the network



(a) 3D textured T-shirt (b) 3D textured backpack

(c) 3D textured pillow

Figure 6. More 3D textile models with synthesized textures using our method

5 Discussion and Conclusion

We introduced a novel framework of reconstructing high-quality 3D textile models with synthesized textures. Our main contributions are as follows:

- A novel pipeline is designed to obtain 3D high-quality textile models based on KinectFusion;
- Convolutional neural networks (CNN) is used for textile texture synthesis;

- To our best knowledge, this is the first paper combining 3D textile scanning with textile texture synthesis. It overcomes the restriction of photography texture in the area of 3D scanning to obtain realistic 3D synthesized textured textile models.

A 3D textile model with better geometry can be obtained using our method. And the proposed texture synthesis method can generate new texture for 3D textile model, which is overcoming the limitation of texture mapping from photography methods in 3D scanning. The final textured 3D textile model from our method is high-quality and realistic, as shown in Fig. 6.

There are still some limitations on our method. The 3D textile model cannot extracted automatically due to its complicated geometry, so the quality of 3D textile models highly depend on manual precision. As to the textile texture synthesis, an original texture should be given in our work, the synthesized texture has a similar style to the original textile texture.

In the future, we hope to generate the arbitrary textile texture without any input images. Also, it is interesting to synthesize the 3D textile geometry rather than using 3D scanning. A similar concept has been proven in the field of human body modeling called Body Talk [18]. Only a brief word description, e.g. “medium thin” and “tall”, can generate a good human body shape.

Acknowledgment

This work is supported by Natural Science Foundation of China (Grant No.61572124), and China Scholarship Council (File No.201506630055).

Reference

1 Izadi S, Kim D, Hilliges O, et al. KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera[C]//Proceedings of the 24th annual ACM symposium on User interface software and technology. ACM, 2011: 559-568.
2 Smisek J, Jancosek M, Pajdla T. 3D with Kinect[M]//Consumer Depth Cameras for Computer Vision. Springer London, 2013: 3-25.
3 Amon C, Fuhrmann F, Graf F. Evaluation of the spatial resolution accuracy of the face tracking system for kinect for windows v1 and v2[C]//Proceedings of the 6th Congress of the Alps Adria Acoustics Association. 2014.
4 Butkiewicz T. Low-cost coastal mapping using Kinect v2 time-of-flight cameras[C]//2014 Oceans-St. John's. IEEE, 2014: 1-9.
5 Fleishman S, Drori I, Cohen-Or D. Bilateral mesh denoising[C]//ACM transactions on graphics (TOG). ACM, 2003, 22(3): 950-953.
6 Dunbar D, Humphreys G. A spatial data structure for fast Poisson-disk sample generation[J]. ACM Transactions on Graphics (TOG), 2006, 25(3): 503-508.
7 Zhao W, Gao S, Lin H. A robust hole-filling algorithm for triangular mesh[J]. The Visual Computer, 2007, 23(12): 987-997.
8 Jones T R, Durand F, Desbrun M. Non-iterative, feature-preserving mesh smoothing[C]//ACM Transactions on Graphics (TOG). ACM, 2003, 22(3): 943-949.
9 Jakob W, Tarini M, Panozzo D, et al. Instant field-aligned meshes[J]. ACM Transactions on Graphics, 2015, 34(6): 15.
10 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

11 Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional architecture for fast feature embedding[C]//Proceedings of the 22nd ACM international conference on Multimedia. ACM, 2014: 675-678.
12 Gatys L, Ecker A S, Bethge M. Texture synthesis using convolutional neural networks[C]//Advances in Neural Information Processing Systems. 2015: 262-270.
13 Portilla J, Simoncelli E P. A parametric texture model based on joint statistics of complex wavelet coefficients[J]. International journal of computer vision, 2000, 40(1): 49-70.
14 LeCun Y, Bottou L, Orr G. Efficient BackProp in Neural Networks: Tricks of the Trade (Orr, G. and Müller, K., eds.)[J]. Lecture Notes in Computer Science, 1524.
15 Zhu C, Byrd R H, Lu P, et al. LBFGS-B: Fortran subroutines for large-scale bound constrained optimization[J]. Report NAM-11, EECS Department, Northwestern University, 1994.
16 Dong C, Loy C C, He K, et al. Learning a deep convolutional network for image super-resolution[C]//European Conference on Computer Vision. Springer International Publishing,
17 www.textures.com
18 Streuber S, Quiros-Ramirez M A, Hill M Q, et al. Body talk: crowdshaping realistic 3D avatars with words[J]. ACM Transactions on Graphics (TOG), 2016, 35(4): 54.