

A Generative Model of Textures Using Hierarchical Probabilistic Principal Component Analysis

(Regular Research Paper)

Aiga SUZUKI*[†], Hayaru SHOUNO[‡]

* University of Tsukuba

1-1-1 Tennodai, Tsukuba, Ibaraki, 305-0006, Japan

[†] National Institute of Advanced Industrial Science and Technology

1-1-1 Umezono, Tsukuba, Ibaraki, 305-0045, Japan

Email: ai-suzuki@aist.go.jp

[‡] University of Electro-Communication

1-5-1 Chofugaoka, Chofu, Tokyo, 182-8585, Japan

Email: shouno@uec.ac.jp

Abstract—Modeling of textures in natural images is an important task to make a microscopic model of natural images. Portilla and Simoncelli proposed a generative texture model, which is based on the mechanism of visual systems in brains, with a set of texture features and a feature matching. On the other hand, the texture features, used in Portillas' model, have redundancy between its components came from typical natural textures. In this paper, we propose a contracted texture model which provides a dimension reduction for the Portillas' feature. This model is based on a hierarchical principal components analysis using known group structure of the feature. In the experiment, we reveal effective dimensions to describe texture is fewer than the original description. Moreover, we also demonstrate how well the textures can be synthesized from the contracted texture representations.

Index Terms—texture modeling, texture synthesis, dimensionality reduction, probabilistic principal component analysis

I. INTRODUCTION

From the microscopic point of view, all of natural scenes, providing our visual stimulus, can be seen as a patchwork of various texture patterns. As a biological knowledges, our visual system recognize the objects by their shapes in primary visual cortex. In addition to this, recent studies have shown that their textures also be important property for recognizing and segmentation of the objects in our higher-level visual cortex [1]. Thus, availability of texture description is important for modeling of natural images in the field of computer vision. Especially, our research focus is to construct a generative model for various texture patterns.

Most of recent studies of texture modeling, achieving a good benchmark results, are based on the Markov Random Field (MRF) and Gibb's sampling [2]–[6]. These MRF based models assume that textures are characterized as just a 2-dimensional probabilistic distributions. Therefore, the most MRF based texture models would have lack of consideration for “a perceptually equivalence”.

The perceptually equivalence is our perceptual property considering that “the pattern A and B are came from same



Fig. 1. The macro photographs of same surfaces but different light sources from “Moss” class of CURET texture database [7]. (Left) front faced and low light amount. (Right) oblique faced and high light amount.

texture”. The perceptually equivalence is not referred explicitly from structural modeling, such as MRF models, but also naturally required as in a part of modeling of natural images. Figure 1 shows the example of the perceptually equivalence, which are the pictures of same materials but different illuminance. We can find that the structural/geometric differences, that are overexposures and shadow directions, came from the different light source, on the other hand, we can also decide that the both of them are same textures.

Portilla and Simoncelli proposed a texture feature, which considers such a perceptual aspects, and a generative model on the basis of feature matching reconstruction [8], [9]. Hereafter, we call their texture feature, “a Portilla-Simoncelli statistics”, PSS for short. The PSS is influenced by biological knowledges, the receptive field of the primary visual cortex and Gabor-like filters especially in the area V1. The PSS is based on wavelet-like multi-scale image decomposition method, which has translation-invariance and rotation-invariance, known as a steerable filter pyramid [10]. The PSS is consists of a marginal statistics over decomposed images, which act as constraints of the texture structure. Simoncelli et al. showed that the PSS could characterize and generate various textures [8], [9], [11], [12]. In recent, moreover, some studies report that the PSS could express the selective neuronal activity in area V4 of

macaque [13] and in area V2 of humans' cerebral cortex [14]. These reports were implying that the PSS could be appropriate representation of our texture recognition mechanisms.

Considering the modeling of textures in natural images, hereafter we call it as "natural textures", the PSS would be good representation, however, the PSS would have redundancy between its elements, because the natural textures typically have certain structure. Therefore, the PSS extracted by natural textures could be phrased with more simplified representation.

This paper proposes a dimension reduction method to grasp the latent factors in the PSS of the natural textures. Our method is based on Probabilistic Principal Component Analysis (PPCA) and focusing on known correlation of the PSS. We achieve a 88.8[%] dimension reduction from raw PSS to phrase natural texture dataset, which is hard to apply a plain PCA.

II. A PORTILLA-SIMONCELLI STATISTICS

This section gives an overview of the PSS used in our study as a texture feature.

A. A steerable filter pyramid

A steerable filter pyramid [10], [11] is multi-scale and directive image decomposition method which partially imitates the orientation selectivity of humans' visual systems. The steerable filter pyramid could be "steered" its decomposition traits by 2 parameters, a number of decomposition scales N and a number of decomposition orientations/directions K , thus, this name was given. These properties originated in complex orthogonal Wavelet transform and Gabor filters bank.

The steerable filter pyramid was defined in Fourier domain, in similar to wavelet transform. Figure 2 shows a Fourier domain block diagram of a steerable filter pyramid interpreted as a linear system. Let $I(x, y)$ be an input image in spatial domain, and $\tilde{I}(r, \theta)$ be a polar representation of an input image in Fourier domain, each transfer elements in Figure 2 would be given by:

$$L(r, \theta) = \begin{cases} 2 \cos\left(\frac{\pi}{2} \log_2\left(\frac{4r}{\pi}\right)\right), & \left(\frac{\pi}{4} < r < \frac{\pi}{2}\right) \\ 2, & \left(r \leq \frac{\pi}{4}\right) \\ 0, & \left(r \geq \frac{\pi}{2}\right) \end{cases} \quad (1)$$

$$B_k(r, \theta) = H(r)G_k(\theta), \quad k \in 0, \dots, K-1, \quad (2)$$

the band-pass transfer element $B_k(r, \theta)$ was decomposed into radial part $H(r)$ and angular part $G_k(\theta)$ defined as

$$H(r) = \begin{cases} \cos\left(\frac{\pi}{2} \log_2\left(\frac{2r}{\pi}\right)\right), & \left(\frac{\pi}{4} < r < \frac{\pi}{2}\right) \\ 1, & \left(r \leq \frac{\pi}{4}\right) \\ 0, & \left(r \geq \frac{\pi}{2}\right) \end{cases} \quad (3)$$

$$G_k(\theta) = \begin{cases} \alpha_K \left\{ \cos\left(\theta - \frac{\pi k}{K}\right) \right\}^{K-1}, & \left(\left|\theta - \frac{\pi k}{K}\right| < \frac{\pi}{2}\right) \\ 0, & \text{(otherwise)} \end{cases} \quad (4)$$

α_k denotes a coordinate of B_k which depends on number of orientation K , given by

$$\alpha_K = 2^{K-1} \frac{(K-1)!}{\sqrt{K(2(K-1))!}} \quad (5)$$

Low-pass transfer element L_0 had to be defined to reject a band exceeding the Nyquist limit, to make this decomposition a complete system.

$$L_0(r, \theta) = \frac{L(r/2, \theta)}{2} \quad (6)$$

Refer to Eq.6, the high-pass transfer element H_0 follows the low-pass residual as

$$H_0(r, \theta) = H\left(\frac{r}{2}, \theta\right) \quad (7)$$

As shown in Figure 2, the steerable filter pyramid would consist of recursion of subsystem $R^{(n)}(r, \theta)$. The multi-scaleness of the steerable filter pyramid was provided by the recursion of them, and the orientation selectivity was provided by a parallelism of them.

B. Construction of the PSS

The PSS is constructed as a list of 10-types of statistics $C_1 \sim C_{10}$, came from the steerable filter pyramid.

1) *Descriptive statistics of pixel values (C_1):* Mean, variance, skewness, kurtosis, minimum and maximum values of pixel value of the input image. (6 dimensions)

2) *Descriptive statistics of each scales (C_2):* Skewness and kurtosis of reconstructed images with each scales of steerable filter pyramid, including a low-pass residual \tilde{I}_{LR} . ($2(N+1)$ dimensions)

3) *Auto-correlation of each decompositions (C_3):* M -neighbor auto-correlation of reconstructed images with each decompositions (scales and orientations) of the steerable filter pyramid. ($N \cdot K \cdot M^2$ dimensions)

4) *Auto-correlation of each scales (C_4):* M -neighbor auto-correlation of reconstructed images with each scales of steerable filter pyramid, including a low-pass residual \tilde{I}_{LR} . ($M^2 \cdot (N+1)$ dimensions)

5) *Cross-correlations between decomposed images for each scales (C_5):* Cross-correlation of decomposed image with each scales between each orientations. ($N \cdot K^2$ dimensions)

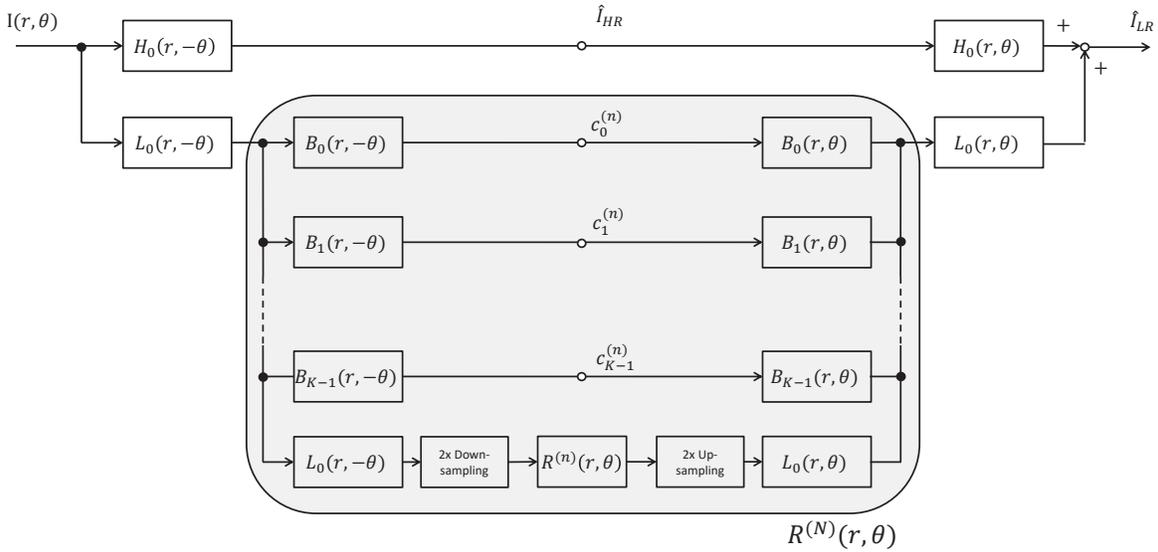


Fig. 2. Block diagram of steerable filter pyramid in Fourier domain. The white circle denotes a observation point.

6) *Cross-correlations between reconstructed images for each scales (C_6):* Cross-correlation of reconstructed images with each scales, including a low-pass residual \tilde{I}_{LR} , between each orientations.

($K^2(N + 1)$ dimensions)

7) *Cross-correrations between each reconstructed images (C_7):* Cross-correration of reconstructed images with each scales, including a low-pass residual \tilde{I}_{LR} , and each orientations.

($K^2 \cdot N(N + 1)$ dimensions)

8) *Cross-correlations between each decomposed images for each decompositions (C_8):* Cross-correlation of decomposed image for each decompositions (scales and orientations).

($N^2 \cdot K^2$ dimensions)

9) *Means of each reconstruted images (C_9):* Means of pixel values of reconstructed images with each decompositions, including a low-pass residual \tilde{I}_{LR} and high-pass residual \tilde{I}_{HR} .

($N \cdot K + 2$ dimensions)

10) *Variance of high-pass residual (C_{10}):* Variance of pixel values of reconstructed images with high-pass residual \tilde{I}_{HR} .

(1 dimension)

Dimensionality of the PSS will be determined by parameters of steerable filter pyramid N , K , and neighbor of pixel space M . In our experiments, we use parameters as $N = 4, K = 4, M = 7$ according to [9] consistently. The dimensionality of the PSS will be 1784 by these parameters.

Figure 3 shows a schematic diagram of texture reconstruction from the PSS. Texture reconstruction will be accomplished with iterative optimization from the Gaussian white noise. [9] shows this reconstruction would converge on desired PSS with in 50-iterations experimentally.

III. HIERARCHICAL PROBABILISTIC PRINCIPAL COMPONENT ANALYSIS

In this paper, we propose a probabilistic principal component analysis (PPCA) [15] based model for dimension

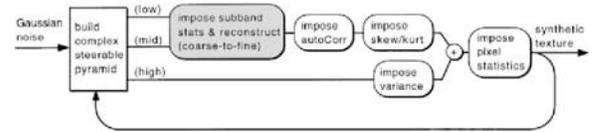


Fig. 3. Flowchart of texture reconstruction from the PSS, quoted from [9]. Texture reconstruction is based on coarse-to-fine optimization from the Gaussian white noise.

reduction, which focuses to known structure of the PSS. PPCA is a statistical method to estimate the latent variables which generate an input data essentially. It is similar to deterministic principal component analysis (PCA) and provides a same result usually, however, PPCA is a probabilistic model which assume that the input data would be generated under the Gaussian distribution and the Gaussian noise. Therefore, PPCA has a merit of an optimization under the high-dimensional input space.

PPCA could be of use for dimension reduction in most cases, however, we could sometimes know the structure or correlation of the input data in advance. For example, construction of the PSS was built up gradually with 10-types of statistics. In other words, the PSS has a group-structure between its components. Considering such a known structure, it is possible to grasp more effective contracted representation of the input data.

We propose a novel architecture of PPCA, considering such a known structure of the input data, a “hierarchical probabilistic principal component analysis” (HPPCA). Figure 4 shows a schematic diagram of the HPPCA. The HPPCA applies a hierarchical dimension reduction which is based on a structure of the PSS. First, the HPPCA applies a dimension reduction for each classes $C_1 \sim C_{10}$ with distinct PPCA models. Second, reduced representations will be concatenated into an intermediate vector, finally, conclusive reduced representation

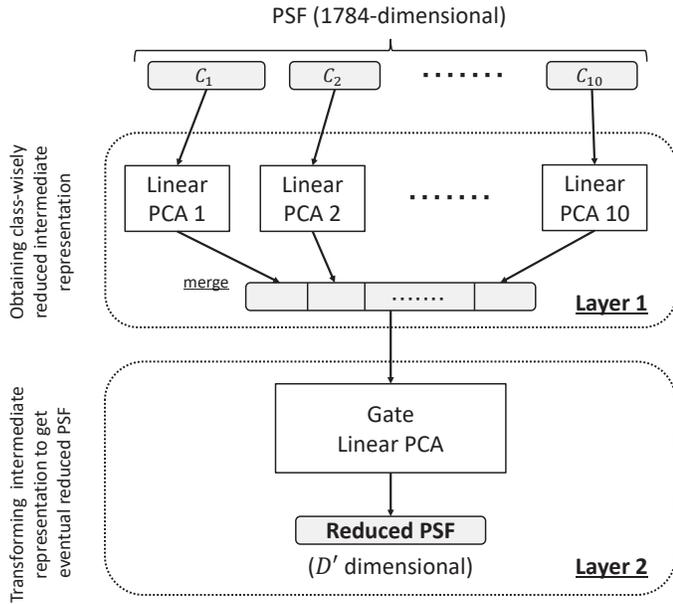


Fig. 4. Schematic diagram of the hierarchical probabilistic principal component analysis. Input will be divided into known groups and contracted with distinct linear-PPCA. After that, the final linear-PPCA makes a conclusive reduced representation.

will be given as an output of the final PPCA.

IV. MATERIALS

A. Texture Dataset

We evaluated our model on natural textures from Kylberg Texture Dataset v. 1.0 [16]. Figure 5 shows examples of texture images in Kylberg Texture Dataset. This dataset contains 28 classes of natural textures, which are the macro photographs of real-world surfaces. Each classes have 1920 patches of gray-scale images normalized with a mean value of 127 and a standard deviation of 40. The patches have a resolution of 576×576 pixels and resized into 128×128 pixel to adjust the model input. To evaluate the model performance, we used 1720 patches for training, and 200 patches for evaluation.

B. Performance index of texture similarity

To evaluate the performance of texture reconstruction, we chose a Texture Similarity Score (TSS) as an index of texture similarity [2]. For a source texture image \mathbf{x} and generated texture sample \mathbf{s} , the TSS will be given by:

$$\text{TSS}(\mathbf{s}, \mathbf{x}) := \max \left\{ \frac{\mathbf{x}_{(1)}^T \mathbf{s}}{\|\mathbf{x}_{(1)}\| \|\mathbf{s}\|}, \dots, \frac{\mathbf{x}_{(I)}^T \mathbf{s}}{\|\mathbf{x}_{(I)}\| \|\mathbf{s}\|} \right\} \quad (8)$$

$\mathbf{x}_{(i)}$ denotes patch i within the test region of the image and I is the number of possible unique patches in the test region. In other words, the TSS denotes the maximum cosine similarity between sample patches and possible source texture region, known as ‘‘the maximum normalized cross correlation’’.

A patch and sample of size 19×19 pixel was adopted to define the TSS in our experiments, according to previous work [2].

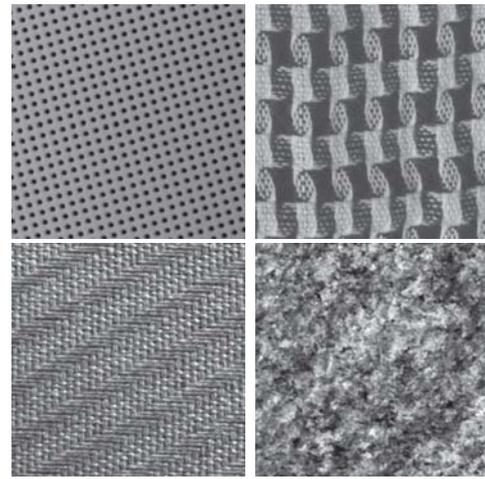


Fig. 5. Examples of texture images in Kylberg Texture Dataset. (top-left) the metal plate ceiling. (top-right) the woven scarf. (bottom-left) the woven fabric on chair. (bottom-right) Flat part of a granite.

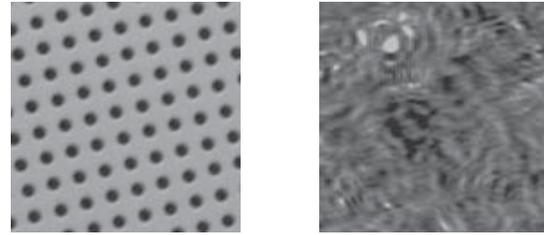


Fig. 6. Result of reconstruction of the texture image from ceiling1 with the plain linear-PPCA in failure. (left) source texture. (right) synthesized texture from 1000-dimensional representation.

V. EXPERIMENT

A. A preliminary experiment: dimension reduction with conventional PPCA

As a preliminary experiment, we tried to obtain a dimension-reduced representation with plain linear-PPCA directly. Figure 6 shows a reconstruction result by 1784 to 1000 dimension reduction. The reconstruction result does not reproduce the source texture structure by the little dimension reduction which expect to preserve the source PSS structure. This result implies that the plain PPCA could not grasp the latent variables of the PSS sufficiently.

B. Appropriate dimensionality of HPPCA

Dimension reduction with the HPPCA would be steered by 2 part, to get an intermediate representation and to obtain a conclusive representation, respectively. Because of the top-down dimension reduction, we have to determine a dimension reduction rate of intermediate PPCAs at first. We chose the cumulative contribution ratio as an index of dimension reduction of each PPCAs. We determined the dimensions of each PPCAs by variation of the TSS with cumulative contribution ratio.

Plot of the TSS vs cumulative contribution ratio of intermediate PPCAs is shown in Figure 7. It was shown that the TSS

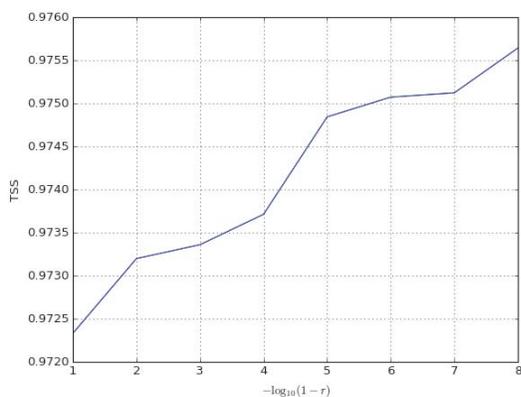


Fig. 7. Variation of the TSS by intermediate cumulative contribution ratio.

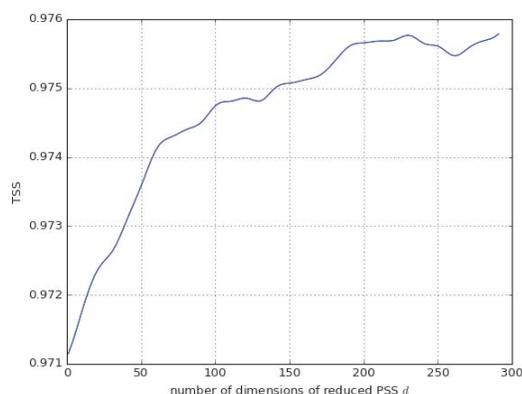


Fig. 8. Variation of the TSS by dimensionality of model outputs.

was monotonically increased as the cumulative contribution ratio, however, it exceeded machine epsilon of 32-bit float after $\log_{10}(1-r) = -8$, ($r = 0.99999999$). Thus, we chose $r = 0.99999999$ and 965-dimensional intermediate representation.

By given intermediate layer, we chose a conclusive dimensionality of model output. Plot of the TSS vs conclusive dimensionality of the HPPCA is shown in Figure 8. The TSS seemed to saturate around d equal 150 to 200. We adopt $d = 200$ as conclusive reduced representation of PSS in our experiments.

C. Qualitative evaluation

This section shows that the contracted representation by our proposed method can reconstruct the textures with perceptually equivalence. The reconstructed textures by contracted PSS using the HPPCA discussed in previous section shown in Figure 9. The fine textures, that are ceiling1, cushion1, and blanket1, were well-reproduced by reduced PSS (9 first row). On the other hand, however, the coarse textures, which have large patterns, sometimes have trouble with a reproducing the continuous (9 second and third row). This could be due to the lack of reproduce the low-frequency component of reduced

PSS with the HPPCA. Nevertheless, the most of reconstruction results might be said to reproduce source texture structure and we could determine these textures were almost much the same.

VI. DISCUSSION AND CONCLUSION

The dimension reduction method of the Portilla-Simoncelli statistics, that is a perceptual texture feature, using the hierarchical probabilistic principal component analysis was introduced in this paper. We achieved a 88.8 [%] dimension reduction from the raw PSS preserving the source texture structures in reconstruction.

The HPPCA model could be read as the Gaussian-Gaussian restricted Boltzmann machine adopted a sparsity in its connections [17]. Such connectionism would be make the PPCA easier to grasp the latent structure of the input space. In future work, we intend to analyze mathematically the mechanism of the HPPCA model and build more generalized texture model via the large-scaled natural image datasets.

REFERENCES

- [1] E. H. Adelson, "On seeing stuff: the perception of materials by humans and machines," in *Photonics West 2001-electronic imaging*. International Society for Optics and Photonics, 2001, pp. 1–12.
- [2] N. Heess, C. K. Williams, and G. E. Hinton, "Learning generative texture models with extended fields-of-experts," in *BMVC*, 2009, pp. 1–11.
- [3] M. Ranzato, V. Mnih, and G. Hinton, "Generating more realistic images using gated MRF's," *Nips*, pp. 1–9, 2010.
- [4] H. Luo, P. L. Carrier, A. Courville, and Y. Bengio, "Texture Modeling with Convolutional Spike-and-Slab RBMs and Deep Extensions," *Proceedings of the 16th International Conference on Artificial Intelligence and Statistics (AISTATS)*, vol. 31, pp. 415–423, 2012. [Online]. Available: <http://arxiv.org/abs/1211.5687>
- [5] J. J. Kivinen and C. K. I. Williams, "Multiple Texture Boltzmann Machines," *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics (AISTATS 2012)*, vol. 22, pp. 638–646, 2012. [Online]. Available: <http://homepages.inf.ed.ac.uk/s0960152/papers/MTBM-AISTATS12.pdf>
- [6] M. Ranzato, V. Mnih, J. M. Susskind, and G. E. Hinton, "Modeling natural images using gated MRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 9, pp. 2206–2222, 2013.
- [7] K. J. Dana, B. Van Ginneken, S. K. Nayar, and J. J. Koenderink, "Reflectance and texture of real-world surfaces," *ACM Transactions on Graphics (TOG)*, vol. 18, no. 1, pp. 1–34, 1999.
- [8] J. Portilla, R. Navarro, O. Nestares, and A. Taberero, "Texture synthesis-by-analysis method based on a multiscale early-vision model," *Optical Engineering*, vol. 35, no. 8, pp. 2403–2417, 1996.
- [9] J. Portilla and E. P. Simoncelli, "Parametric texture model based on joint statistics of complex wavelet coefficients," *International Journal of Computer Vision*, vol. 40, no. 1, pp. 49–71, 2000.
- [10] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *Image Processing, 1995. Proceedings., International Conference on*, vol. 3. IEEE, 1995, pp. 444–447.
- [11] J. Portilla and E. P. Simoncelli, "Texture modeling and synthesis using joint statistics of complex wavelet coefficients," in *IEEE workshop on statistical and computational theories of vision*, 1999.
- [12] E. P. Simoncelli and J. Portilla, "Texture characterization via joint statistics of wavelet coefficient magnitudes," in *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on*, vol. 1. IEEE, 1998, pp. 62–66.
- [13] G. Okazawa, S. Tajima, and H. Komatsu, "Image statistics underlying natural texture selectivity of neurons in macaque V4," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 4, pp. E351–60, 2015.
- [14] C. Hiramatsu, N. Goda, and H. Komatsu, "Transformation from image-based to perceptual representation of materials along the human ventral visual pathway," *NeuroImage*, vol. 57, no. 2, pp. 482–494, 2011.

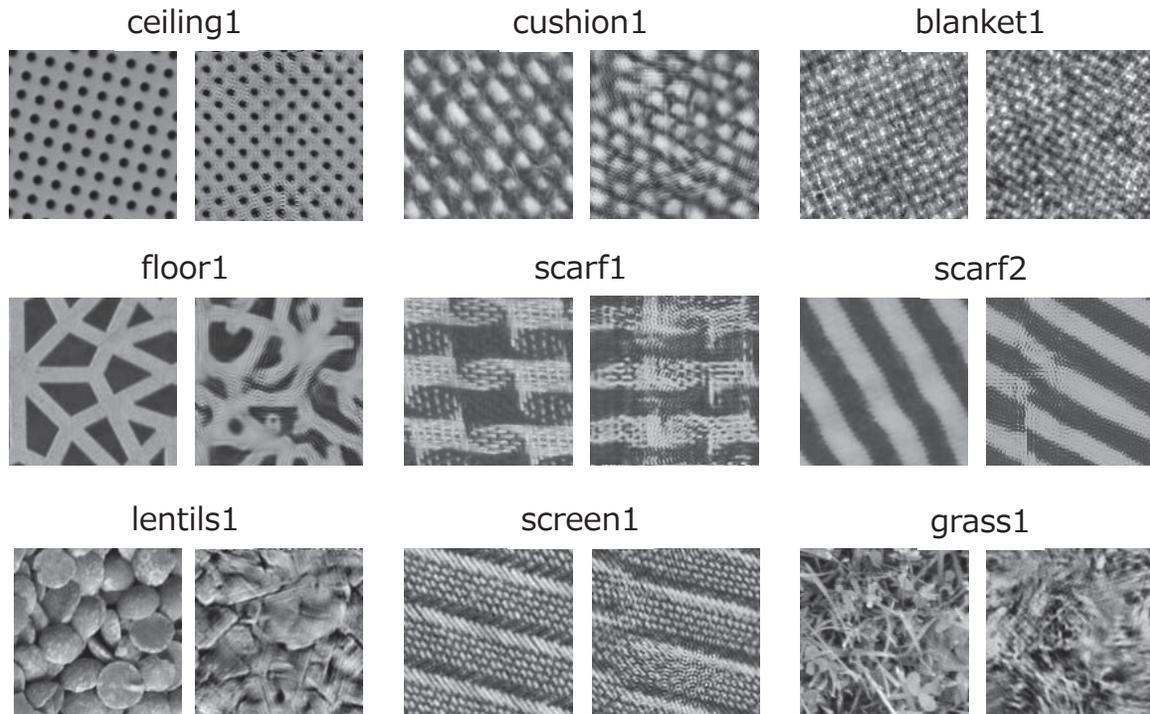


Fig. 9. Reconstruction results from 200-dimensional contracted TSS. The left side of each classes denotes the source texture, right side denotes the reconstructed texture.

- [15] M. E. Tipping and C. Bishop, "Probabilistic Principal Component Analysis," *Journal of the Royal Statistical Society, Series B*, vol. 21/3, pp. 611–622, jan 1999.
- [16] G. Kylberg, "The kylberg texture dataset v. 1.0," Centre for Image Analysis, Swedish University of Agricultural Sciences and Uppsala University, Uppsala, Sweden, External report (Blue series) 35, September 2011. [Online]. Available: <http://www.cb.uu.se/gustaf/texture/>
- [17] R. Karakida, M. Okada, and S.-i. Amari, "Dynamical analysis of contrastive divergence learning: Restricted boltzmann machines with gaussian visible units," *Neural Networks*, vol. 79, pp. 78–87, 2016.