Model-based Estimation of Texels and Placement Grids for Fast Realistic Texture Synthesis

Dongxiao Zhou
CITR, Department of Computer Science,
Tamaki Campus, The University of Auckland,
Auckland, New Zealand
Email: dzho002@ec.auckland.ac.nz

Georgy Gimel'farb
CITR, Department of Computer Science,
Tamaki Campus, The University of Auckland,
Auckland, New Zealand
Email: g.gimelfarb@auckland.ac.nz

***Abstract*— This paper describes the basic steps of a new technique, called bunch sampling, that enables the realistic synthesis of spatially homogeneous textures. A geometric shape of the bunch (acting as a texel) and spatial placement grid governing relative positions of the bunches are estimated from the training texture by using a generic Gibbs random field texture model with multiple pairwise pixel interactions. During the synthesis, the bunches are randomly sampled from the training texture and placed into the large-size goal image with due account of their spatial interdependence.**

I. Introduction

Due to diversity and complexity of natural textures, their analysis and synthesis is generally a complicated problem. Most of present approaches to texture synthesis are based on Markov random field (MRF) models and can be roughly divided into two groups, namely, model based probabilistic synthesis and non-parametric sampling.

The model based methods [1], [3], [4], [7]–[9], [17] identify first a particular MRF model of a given training texture specified usually by a joint Gibbs probability distribution (GPD) of image signals, and generate textures using Markov Chain Monte Carlo (MCMC). Since both the model identification (parameter estimation) and the image generation are computationally intensive processes, these approaches are less feasible for synthesising large-size textures. Much faster synthesis is obtained with non-parametric sampling [5], [6], [12]. Also assuming Markovianity of the texture, these techniques treat the training image as a source of random signal samples related to the underlying marginal GPDs. They synthesise new textures by direct sampling of image signals and permuting and replicating those samples into a goal image. Since the synthesis accounts for no explicit texture model, the synthetic texture may have false borders between the permuted samples and verbatim replication of training singularities. Also, there is no theoretically justified technique to select the proper size of the samples for various textures.

This paper details an alternative approach to fast texture synthesis, called bunch sampling in [10], [16], that aims to bridge gaps between the model based synthesis and non-parametric sampling by combining the strength of the both approaches. The structural approach to texture analysis [11] considers a homogeneous texture as a regular and repetitive spatial arrangement of its specific primitive elements (texels). From the statistical viewpoint, the same texture is a sample of a spatially homogenous MRF describing a particular set of local interactions between the image signals. Combining the both two viewpoints, bunch sampling interprets a texture with two types of signal interactions, namely, intra-texel and inter-texel ones. The intra-texel interactions govern signal co-occurrences inside a single texel, while the inter-texel interactions specify spatial interdependence between texels. During the texture synthesis, bunch sampling keeps both the intra- and inter-texel interactions revealed in the training prototype in order for the synthetic texture to inherit statistical features thus overall texture appearance. The intra-texel interactions are preserved by texel-based sampling, that is, a whole texel is retrieved at each sampling step. The inter-texel interactions are kept by placing texel into the synthetic texture in accord with the placement rule that specifies spatial relation of the adjacent texels of the training image.

II. Bunch Sampling: Estimation of a Bunch

Assuming that all its texels have the same geometric shape, a texture is characterised by the geometric shape of, and spatial relations between, the texels. Bunch sampling estimates the geometric shape from the characteristic structure of pairwise pixel interactions for a generic Gibbs random field (GGRF) model [7]–[9].

A. Generic Gibbs Random Field Texture Model

The GGRF texture model involves characteristic translation invariant second order clique families, each family containing pixel pairs with the same relative spatial displacement. Each clique family, $\mathbf{C}_{\xi,\eta} = \{(x,y),(x+\xi,y+\eta) : (x,y) \in \mathbf{R}; (x+\xi,y+\eta) \in \mathbf{R}\}$, has the same strength of interactions between every two grey levels in the pixels, which is given by a Gibbs potential. The GGRF model belongs to the exponential families of distributions, and has the grey level co-occurrence histograms (GLCH) for each clique family as its sufficient statistics. The exponent of the GPD for the GGRF specifies the total interaction energy:

$$\mathbf{E} = \sum_{(\xi,\eta)\in\mathbf{A}} \mathbf{V}_{\xi,\eta} \bullet \mathbf{F}_{\xi,\eta}(\mathbf{g}) \qquad (1)$$

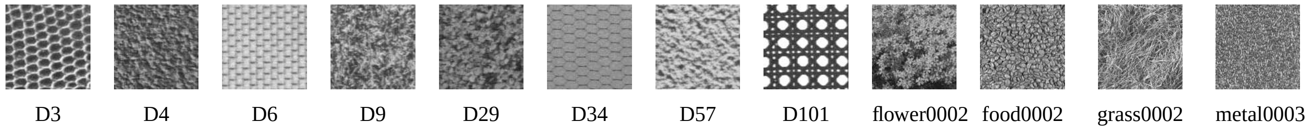


Fig. 1. Training textures (128×128) taken or cut from the digitised Brodatz album [2] and the MIT VisTex texture database [13].

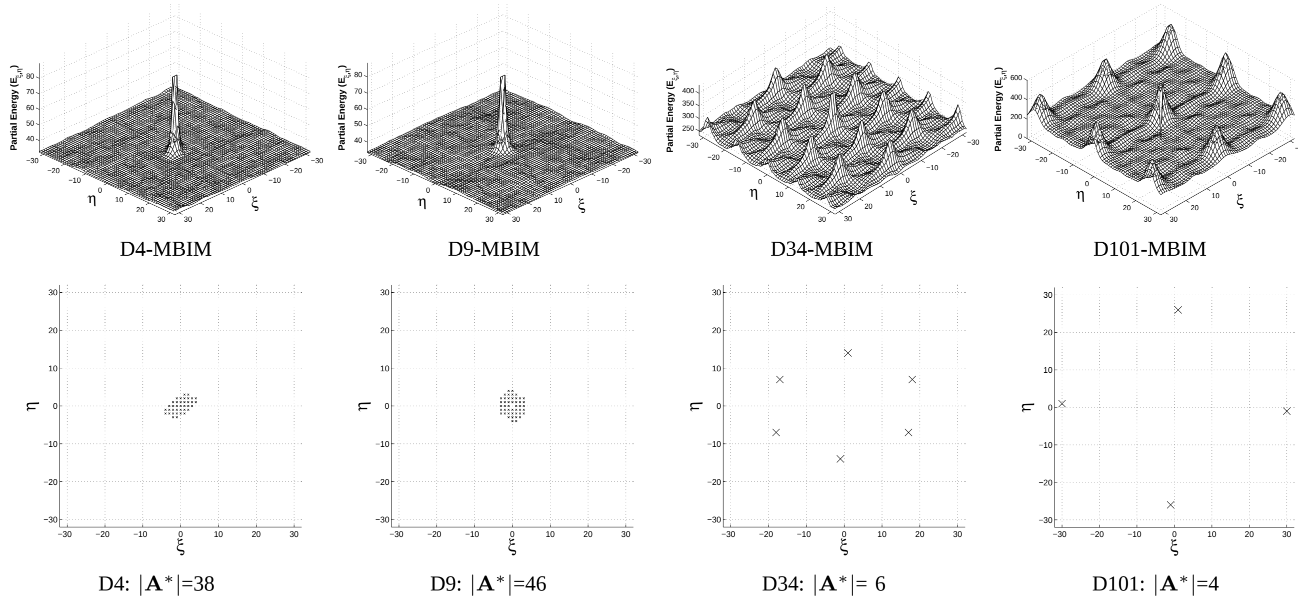


Fig. 2. MBIMs and the estimated bunches.

where $\mathbf{g}$ is a digital image, $\mathbf{A}$ denotes a set of the characteristic clique families indexed with the corresponding inter-pixel displacements, $\mathbf{V}_{\xi,\eta}$ and $\mathbf{F}_{\xi,\eta}(\mathbf{g})$ are, respectively, the potential vector and the normalised GLCH vector for the clique family $\mathbf{C}_{\xi,\eta}$ collected over the image $\mathbf{g}$, and $\bullet$ denotes the dot product. The partial energy of the clique family $\mathbf{C}_{\xi,\eta} \in \mathbf{A}$,

$$E_{\xi,\eta}(\mathbf{g}|\mathbf{V}_{\xi,\eta}) = \mathbf{V}_{\xi,\eta} \bullet \mathbf{F}_{\xi,\eta}(\mathbf{g}) \tag{2}$$

specifies its contribution to the overall interaction energy, and the higher the partial energy, the more characteristic the family in that texture. The rank of the partial energy is used to recover the most characteristic interaction structure of pairwise interactions [9]. As shown later, this interaction structure relates to the geometric shape of the texels.

The potential $\mathbf{V}_{\xi,\eta}$ can be estimated using an MCMC process of stochastic approximation. However, since only the rank of the partial energy is of interest to recover the characteristic interaction structure, the analytic first approximation of the relative partial energy $E^{0}_{\xi,\eta}(\mathbf{g}^{0})$ turns out to be sufficient to rank the clique families in their energies [9]. The partial energy is proportional to the variance of the normalised GLCH vector collected for the clique family $\mathbf{C}_{\xi,\eta}$ over the training image $\mathbf{g}^{0}$:

$$E^{0}_{\xi,\eta}(\mathbf{g}^{0}) \propto (\mathbf{F}_{\xi,\eta}(\mathbf{g}^{0}) - \mathbf{F}_{\mathbf{IRF}}) \bullet \mathbf{F}_{\xi,\eta}(\mathbf{g}^{0}) \tag{3}$$

where $\mathbf{F}_{\mathbf{IRF}}$ is the normalised GLCH vector for the independent random field.

The partial energies of all the clique families in a large search window of displacements, $\mathbf{W} = \{(\xi,\eta) : |\xi| \leq \Delta, |\eta| \leq \Delta\}$ of the size $(2\Delta+1)\times(2\Delta+1)$, can be represented with a model based interaction map (MBIM). A coordinate point (ξ,η) in the MBIM corresponds to the clique family $C_{\xi,\eta}$, and its scalar value represents the partial energy $E_{\xi,\eta}(\mathbf{g})$ in Eq. (2). Figure. 2 shows the MBIMs for several textures with the search window $\mathbf{W}$ of the size 65×65.

B. Geometric Shape of a Bunch

As shown in Fig. 2, the clique families with the top-rank partial interaction energies form specific clusters in the MBIM. Stochastic textures such as D4 and D9 have only one central cluster indicating that the close range pixel interactions, which relate mainly to a uniform background, dominate in those textures. On the contrary, the peripheral clusters in the MBIMs of the regular mosaics like D34 and D101 in Fig. 2 reveal the repetitive patterns of these types of texture.

The essentially different nature of two texture types leads to different strategies of specifying the geometric shape of the bunches. For the stochastic textures, those clique families that form the central cluster of the MBIM define together the geometric shape of the bunches. For the highly structured regular mosaics, the spatial arrangement of the peripheral clusters is more important, and the bunch shape is better specified by the clique families corresponding to the peak

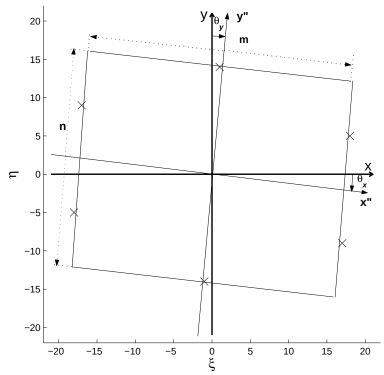


Fig. 3. Parameters for the placement grid for the texture D34.

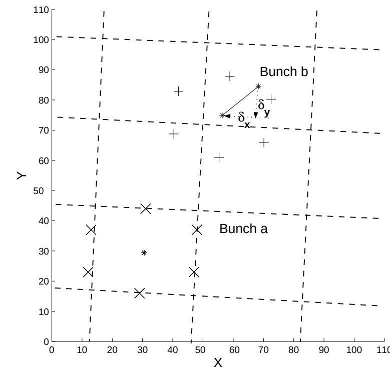


Fig. 4. Tessellation of the texture D34 and relative positions of the bunches: $(0,0)$ for the bunch 'a' and (δ_x,δ_y) for the bunch 'b'.

points of peripheral clusters that are the nearest neighbours of the centre of the MBIM.

The selected clique families, $\mathbf{A}^* \in \mathbf{A}$, for either the stochastic textures or the regular mosaics have the top-rank partial energies, so that one might assume that these families determine visual appearance of the texture. The size of bunches $|\mathbf{A}^*|$ for regular mosaics is usually very small, e.g., only 4 pixels in the bunch for the texture D101 in Fig. 2. But as shown later, even these simple structures are quite adequate to describe and synthesise the regular textures.

C. Placement Grid for Bunches

The bunch sampling specifies the placement rule in terms of relative positions of each signal bunch (or texel) with respect to others. In order to preserve the overall realistic visual appearance of the texture, every bunch must have the same relative position in both the synthetic and training textures.

Assuming that non-overlapping bunches are conditionally independent, the texture is first tesselated with a grid derived from the estimated geometric shape of the bunch. Each cell of the grid is a compact bounding parallelogram around the bunch that can be calculated by the method proposed in [14], [15]. The cell is specified with four parameters, $(\theta_x, \theta_y, m, n)$. The angles θ_x and θ_y give the guiding orientation of the cell sides with respect to the image coordinate axes, and the side sizes m and n are the maximum spans of the bunch along the guiding directions. Figure 3 shows how these four parameters relate to the bunch shape. With such a tessellation, the relative position of a bunch is defined as the shift of the bunch from the closest cell in the placement grid, as shown in Fig. 4.

III. Bunch Sampling: Texture Synthesis

Given the geometric shape and the placement grids for bunches, a new texture of an arbitrary size is synthesised by direct sampling of signal bunches from the training image with their subsequent replication and random placement. At each step, a bunch of signals, representing a particular texel, is randomly sampled from the training image using a mask of the estimated geometric shape. The bunch is then placed into the synthetic texture in a position so that it has the same relative shift with respect to the both placement grids tessellating the training and the synthetic textures. Usually more than one candidate position satisfies the constraint for every bunch, therefore a random position needs to be selected. Signal collisions may happen when a new bunch has to be placed into an area that has been partly or totally occupied by the previously placed bunches. As shown in [10], a simple heuristic rule of preserving the already placed signals resolves satisfactorily these collisions. A new texture is generated until the goal image canvas is fully covered by the image signals transferred from the training texture.

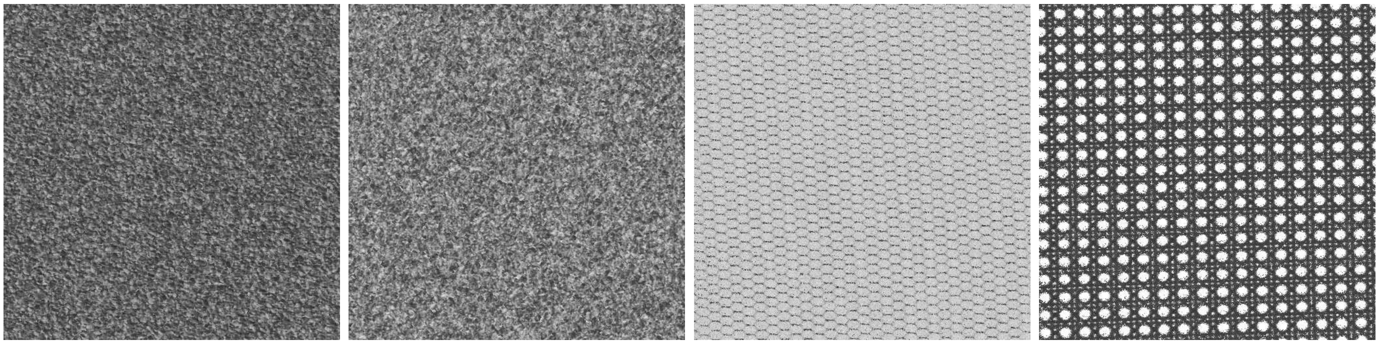
Such a synthesis is very fast because the computation time depends on the size of the synthetic image $|\mathbf{g}|$ and the size of bunch $|\mathbf{A}^*|$, which is $O\left(\frac{|\mathbf{g}|}{|\mathbf{A}^*|}\right)$.

IV. Results and Discussion

Figures 5 and 6 show the examples of textures synthesised from the training samples in Fig. 1, using bunch sampling. A few more training textures and the corresponding synthetic ones are presented in Figs 7 and 8, respectively. In most of these cases, the synthetic textures are visually quite similar to their prototypes.

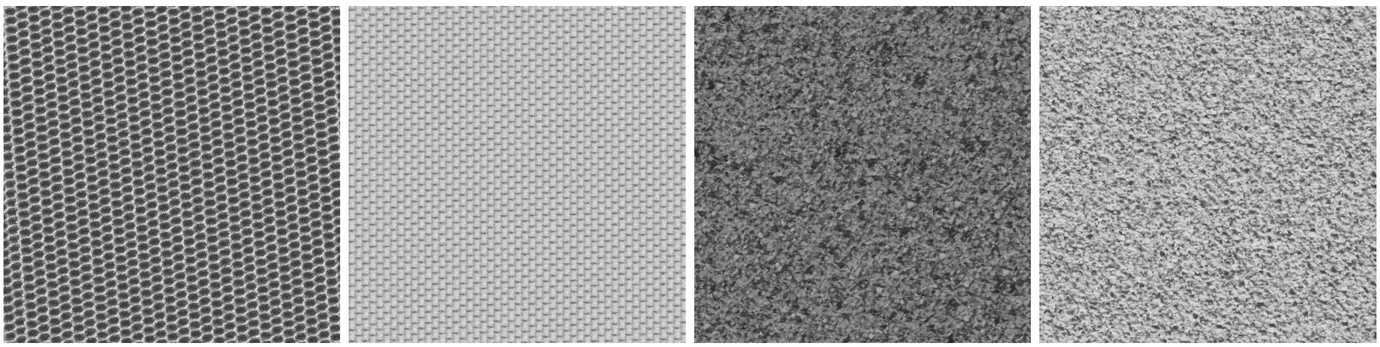
At the analysis stage, the bunch sampling exploits the GGRF texture model for deriving the most characteristic texel shape and spatial interdependence between the texels from the sufficient signal statistics. Based on these results, the very fast texture synthesis is achieved by copying, replicating and pasting the texels with due account of their spatial relationship.

In its basic idea, the bunch sampling is similar to an approach to the synthesis of the near-regular textures proposed in [18]. In this latter approach the periodicity of regular textures is recovered from the translation symmetries of the autocorrelation pattern of the texture. Since it is based on statistics of pairwise signal products over the clique families, the autocorrelation describes the interaction structure in a less definite way than the general statistics of pairwise signal co-occurrences in the GGRF model. Thus bunch sampling has less difficulties in deriving the texel shape, resulting in more detailed texels. This enables the derivation of an efficient tiling scheme for the synthesis of both regular and stochastic textures. However, the major limitation of bunch

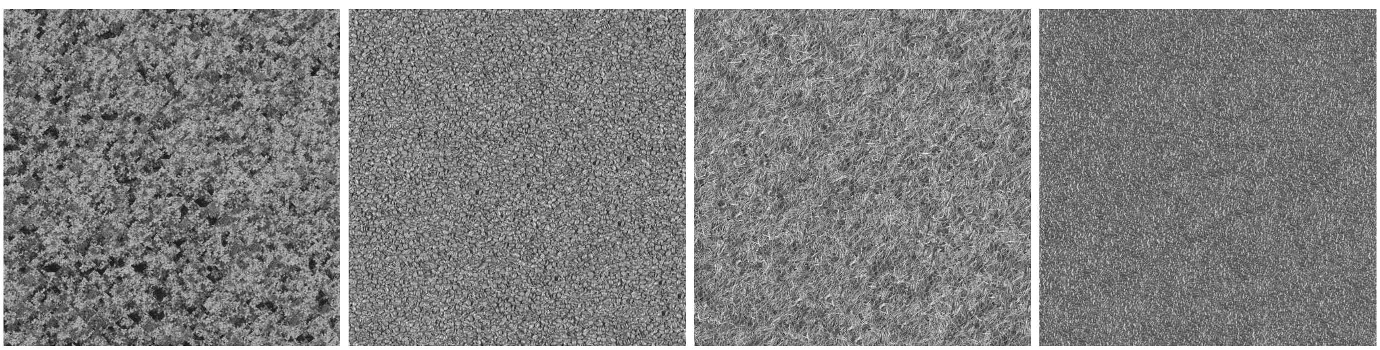


D4 D9 D34 D101

Fig. 5. Synthetic textures D4, D9, D34 and D101 (512×512).



D3: $|\mathbf{A}^*|$=6 D6: $|\mathbf{A}^*|$=4 D29: $|\mathbf{A}^*|$=82 D57: $|\mathbf{A}^*|$=72



flower0002: $|\mathbf{A}^*|$=122 food0002: $|\mathbf{A}^*|$=58 grass0002: $|\mathbf{A}^*|$ = 52 metal0003: $|\mathbf{A}^*|$=28

Fig. 6. Synthetic textures D3, D6, D29, D57, flower0002, food0002, grass0002 and metal0003 (512×300).

sampling is that it fails to handle the inhomogeneity in textures due to the use of bunches with rigid geometric shape, as exemplified by the synthetic textures “Brick wall” (D95) and “Brick” (brick0004) in Fig. 8. Our future work will pursue the goal of geometrically adapting the bunches (texels) with respect to each other as to “rectify” a weak homogeneity of natural periodic textures.

References

[1] J. E. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society*, Vol. B48, pp. 259–302, 1986.

[2] P. Brodatz. *Textures: A Photographic Album for Artists and Designers.* Dover, New York, 1966.

[3] R. Chellappa. Two dimensional discrete Gaussian Markov random field models for image processing. In *Progress in Pattern Recognition*, Vol. 2, pp. 79–112. Elsevier Science Publishers B. V., 1985.

[4] G. R. Cross and A. K. Jain. Markov random field texture models. *IEEE Trans. Pattern Anal. Machine Intell.*, Vol.5, pp. 25–39, 1983.

[5] A. A. Efros and W. T. Freeman. Image quilting for texture synthesis and transfer. In *SIGGRAPH 2001, Computer Graphics Proceedings*, pp. 341–346. ACM Press / ACM SIGGRAPH, 2001.

[6] A. A. Efros and T. K. Leung. Texture synthesis by non-parametric sampling. In *Proc. Int. Conf. on Computer Vision (ICCV’99), Kerkira, Corfu, Greece, 20–25 September 1999*, Vol. 2, pp.1033–1038, 1999.

[7] G. L. Gimel’farb. Non-Markov Gibbs texture model with multiple pairwise pixel interactions. In *Proc. Int. Conf. Pattern Recognition (ICPR 1996), Vienna, Austria, August 1996*, Vol. B, pp. 591–595, 1996.

[8] G. L. Gimel’farb. Texture modeling with multiple pairwise pixel interactions. *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 18, pp. 1110–1114, 1996.

[9] G. L. Gimel’farb. *Image Textures and Gibbs Random Fields*. Dordrecht, Kluwer Academic Publishers, 1999.

[10] G. L. Gimel’farb and D. Zhou. Fast synthesis of large-size textures using bunch sampling. In *Proc. Image and Vision Computing New Zealand (IVCNZ 2002), Auckland, New Zealand, 26–28 November 2002*, pp. 215–

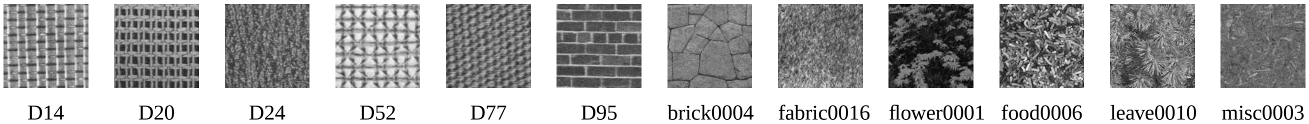


Fig. 7. More example of training textures (128×128)

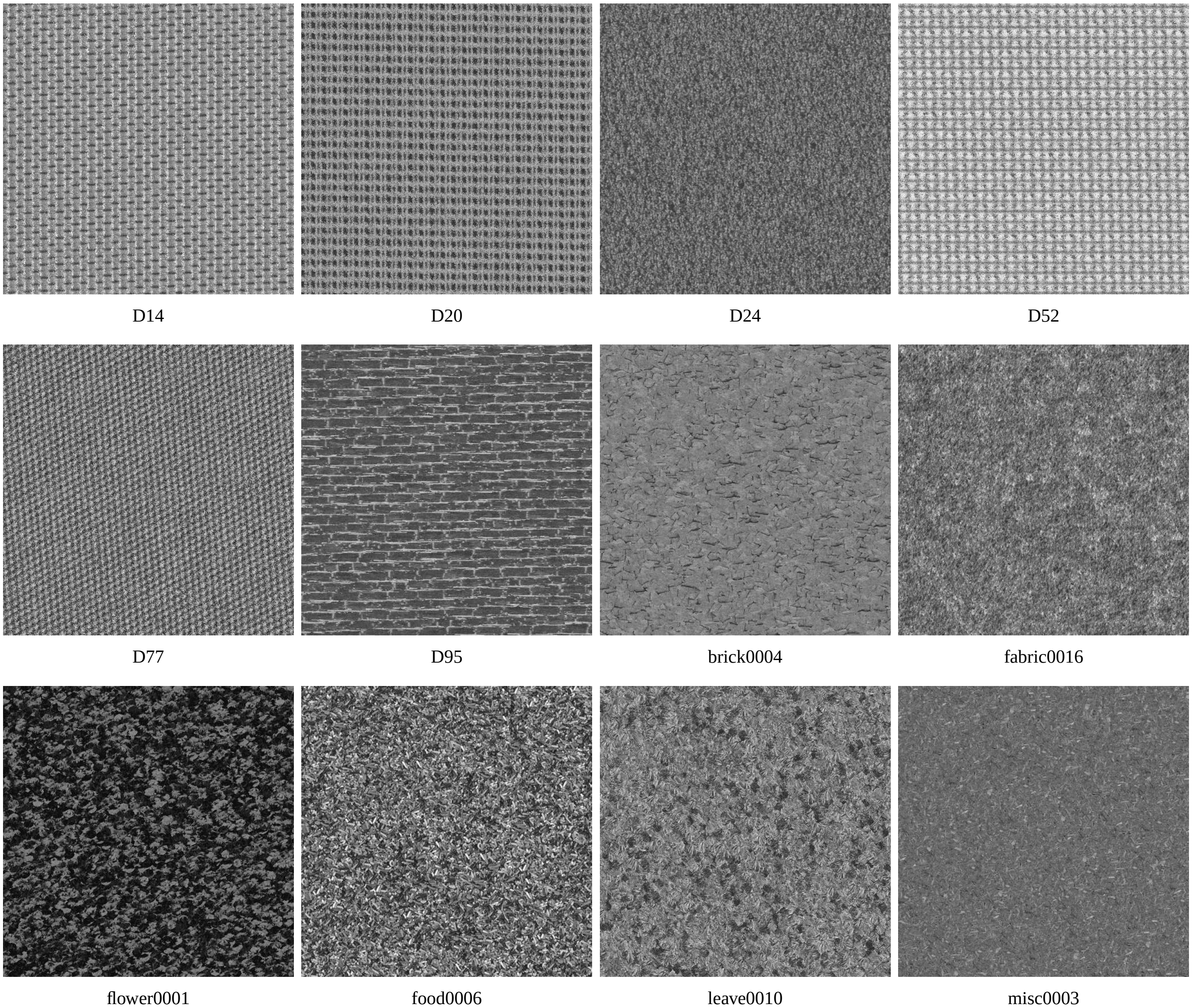


Fig. 8. Synthetic textures (512×512) from the training samples in Fig. 7

220, 2002.

[11] R. Haralick. Statistical and structural approaches to texture. *Proc. IEEE*, Vol. 67, no.5, pp. 786–804, 1979.

[12] L. Liang, C. Liu, and H. Y. Shum. *Real-time texture synthesis by patch-based sampling.* Technical Report MSR-TR-2001-40, Micorsoft Research, 2001.

[13] R. Picard, C. Graszyk, S. Mann, *et al*. *VisTex Database*. MIT Media Lab, Cambridge, Massachusettes, 1995.

[14] C. Schwarz, J. Teich, E. Welzl and B. Evans On finding a minimal enclosing parallelgram. *Technical Report TR-94-036, ICSI, Berkeley, CA*, 1994.

[15] K. Voss and H. Suesse. Invariant fitting of planar objects with primitives. *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 19, pp. 80–84, 1997.

[16] D. Zhou and G. L. Gimel'farb. Bunch sampling for fast texutre synthesis.. In *Proc. Int. Conf. Computer Analysis of Images and Patterns, Groningen, The Netherlands, 25–27 August 2003* [in print].

[17] S. Zhu, Y. Wu, and D. Mumford. Filters, random fields and maximum entropy (FRAME): To a unified theory for texture modeling. *International Journal of Computer Vision*, Vol. 27, no. 2, pp. 107–126, 1998.

[18] Y. Liu and Y. Tsin. The promise and perils of near-regular texture. *Texture 2002, the 2nd International Workshop on Texture Analysis and Synthesis, associated with the European Conference on Computer Vision 2002 (ECCV'02)*, 2002.