2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

Three-dimensional Gaussian Mixture Texture Model

Michal Haindl and Vojtěch Havlíček
Institute of Information Theory and Automation of the CAS
Prague, Czech Republic 182 08
Email: {haindl,havlicek}@utia.cas.cz

***Abstract*—Visual texture modeling based on multidimensional mathematical models is the prerequisite for both robust material recognition as well as for image restoration, compression or numerous physically correct virtual reality applications. A novel multispectral visual texture modeling method based on a descriptive, unusually complex, three-dimensional, spatial Gaussian mixture model is presented. Texture synthesis benefits from easy computation of arbitrary conditional distributions from the model. The model is inherently multispectral thus it does not suffer with the spectral quality compromises of the spectrally factorized alternative approaches. The model is especially well suited for multispectral textile textures and it can also describe the most advanced textural representation in the form of a bidirectional texture function (BTF).**

I. Introduction

Human observer's visual scene recognition is based on shapes and materials. Unfortunately, the surface material appearance vastly changes under variable observation conditions which significantly complicates and negatively affects its mathematical synthesis as well as machine analysis. Reliable computer-based interpretation of visual information which would approach human cognitive capabilities is very challenging and impossible without significant improvement of the corresponding sophisticated visual information models capable to handle huge variations of possible observation conditions. The appropriate paradigm for such a surface material reflectance function models is a multidimensional visual texture.

Texture synthesis approaches may be divided primarily into sampling and model-based methods. Sampling methods [1], [2], [3], [4], [5], [6] rely on sophisticated sampling from real texture measurements while the model-based techniques [7], [8], [9], [10], [11], [12], [13] describe texture data using multidimensional mathematical models and their synthesis is based on the estimated model parameters only.

Generative visual texture models are useful not only for modelling physically correct virtual objects material surfaces in virtual or augmented reality environments, image restoration or compression but also for contextual recognition applications such as segmentation, classification or image retrieval.

Physically correct surface material reflectance model (RM) is sixteen-dimensional function [14]

$$RM(\lambda_i, x_i, y_i, z_i, t_i, \theta_i, \psi_i, \lambda_v, x_v, y_v, z_v, t_v, \theta_v, \psi_v, \theta_{i,T}, \theta_{v,T}).$$

RM describes incident light with spectral value λ_i illuminating surface location x_i, y_i, z_i in time t_i under spherical reflectance angle θ_i, ψ_i and observed at time t_v from surface location x_v, y_v, z_v under spherical reflectance angle θ_v, ψ_v and spectrum λ_v. $\theta_{i,T}, \theta_{v,T}$ are the corresponding transmittance angles. The model height parameters z_i, z_v indicate that even radiance along light rays is not constant but depends on the height. Such a RM model is too complex and there neither exist any measurement of such data nor any mathematical representation allowing its synthesis. One of the early compromised attempts to capture real material appearance was done by Nicodemus et al. [15] and later elaborated by Dana et al. [16] in the form of *Bidirectional Texture Function* (BTF). Even if a BTF model assumes several strong simplifying assumptions [17], [18], [14] its measurement, compression and synthesis is on the leading edge of current mathematical modelling and technological capabilities. BTF is a seven-dimensional function [14] which considers not only measurement dependency on planar material position and spectral channel but also its dependence on illumination and viewing angles:

$$\mathrm{BTF}_{\theta_i,\phi_i,\theta_v,\phi_v}(\tilde{r}) \tag{1}$$

where θ, ϕ are elevation and azimuthal angles of illumination and view direction vector, the multiindex $\tilde{r} = [r_1, r_2, r_3]$ specifies planar horizontal and vertical position in a material sample image and r_3 is the spectral index. Reliable parameters estimation of such a seven-dimensional stochastic model is very difficult not only because it requires very demanding numerical optimization but because the learning textural data are always too limited to obtain robust and reliable estimates. The solution is to factorize the original seven-dimensional measurement space into a set of less dimensional textural factors. The realistic modeling strives not necessarily to recover the exact pixel-wise correspondence with some original target texture but rather a texture which is visually indiscernible from the original one.

In our previous paper [19] we have introduced three two-dimensional probabilistic mixture models, where a measured 3D multi-spectral texture had to be spectrally factorized and the corresponding multivariate mixture models were further learned from single orthogonal mono-spectral components and used to synthesise and enlarge these mono-spectral factor components. The presented model (BTF-3DGMM), on the contrary, is fully multispectral and thus it does not need to compromise spectral modeling quality in multicoloured textures. We applied this model for simpler task of high quality texture restoration [20] where the model can exploit information from corrupted textural data. The presented application to BTF (color / multispectral) texture synthesis is more demanding due to lack of any guiding local spatial information and the requirement to enlarge synthetic texture to any requested unobserved size.

II. Spatial 3D Gaussian Mixture Model

A static homogeneous three-dimensional textural factor Y is assumed to be defined on a finite rectangular $N_1 \times N_2 \times d$ lattice I, $r = (r_1, r_2, r_3) \in I$ denotes a pixel multiindex with the row, columns and spectral indices, respectively. Let us suppose that Y represents a realization of a random vector with a probability distribution $P(Y)$. The statistical properties of interior pixels of the moving window on Y are translation invariant due to assumed textural homogeneity. They can be represented by a joint probability distribution and the properties of the texture can be fully characterized by statistical dependencies on a sub-field, i. e., by a marginal probability distribution of spectral levels on pixels within the scope of a window centered around the location r and specified by the index set:

$$I_r = \{r + s : |r_1 - s_1| \le \alpha \wedge |r_2 - s_2| \le \beta\} \subset I \ .$$

The index set I_r depends on a modeled visual data and can have any other than this rectangular shape. $Y_{\{r\}}$ denotes the corresponding matrix containing all Y_s in some fixed order arrangement such that $s \in I_r$, $Y_{\{r\}} = [Y_s \ \ \forall s \in I_r]$, $Y_{\{r\}} \subset Y$, $\eta =$ cardinality$\{I_r\}$ and $P(Y_{\{r\}})$ is the corresponding marginal distribution of $P(Y)$.

A. 3D Gaussian Mixture

If we assume the joint probability distribution $P(Y_{\{r\}})$, in the form of a normal mixture

$$\begin{aligned} P(Y_{\{r\}}) &= \sum_{m \in \mathcal{M}} p(m)\, P(Y_{\{r\}} \,|\, \mu_m, \Sigma_m) \qquad Y_{\{r\}} \subset Y \ , \\ &= \sum_{m \in \mathcal{M}} p(m) \prod_{s \in I_r} p_s(Y_s \,|\, \mu_{m,s}, \Sigma_{m,s}) \end{aligned} \qquad (2)$$

where $Y_{\{r\}} \in \Re^{d \times \eta}$ is $d \times \eta$ matrix, μ_m is $d \times \eta$ mean matrix, Σ_m is $d \times d \times \eta$ a covariance tensor, and $p(m)$ are probability weights and the mixture components are defined as products of multivariate Gaussian densities

$$P(Y_{\{r\}} \,|\, \mu_m, \Sigma_m) = \prod_{s \in I_{\{r\}}} p_s(Y_s \,|\, \mu_{ms}, \Sigma_{ms}) \ , \qquad (3)$$

$$\begin{aligned} p_s(Y_s \,|\, \mu_{ms}, \Sigma_{ms}) &= \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_{m,s}|^{\frac{1}{2}}} \\ &\exp\left\{ -\frac{1}{2} (Y_r - \mu_{m,s})^T \Sigma_{m,s}^{-1} (Y_r - \mu_{m,s}) \right\} \ , \end{aligned} \qquad (4)$$

i. e., the components are multivariate Gaussian densities with covariance matrices (8).

The underlying structural model of conditional independence is estimated from a data set $\mathcal{S}$ obtained by the step-wise shifting of the contextual window I_r within the original textural image, i. e., for each location r one realization of $Y_{\{r\}}$.

$$\mathcal{S} = \{Y_{\{r\}} \ \forall r \in I, \, I_r \subset I\} \qquad Y_{\{r\}} \in \Re^{d \times \eta} \ . \qquad (5)$$

1) Parameter Estimation: The unknown parameters of the approximating mixture can be estimated using the iterative EM algorithm [21]. In order to estimate the unknown distributions $p_s(\cdot \,|\, m)$ and the component weights $p(m)$ we maximize the likelihood function corresponding to the training set (5):

$$L = \frac{1}{|\mathcal{S}|} \sum_{Y_{\{r\}} \in \mathcal{S}} \log \left[\sum_{m \in \mathcal{M}} P(Y_{\{r\}} \,|\, \mu_m, \Sigma_m)\, p(m) \right] \ .$$

The likelihood is maximized using the iterative EM algorithm (with non-diagonal covariance matrices):

E:

$$q^{(t)}(m |\, Y_{\{r\}}) = \frac{\tilde{P}^{(t)}(Y_{\{r\}} \,|\, \mu_m, \Sigma_m)\, p^{(t)}(m)}{\sum_{j \in \mathcal{M}} P^{(t)}(Y_{\{r\}} \,|\, \mu_j, \Sigma_j)\, p^{(t)}(j)} \ ,$$

M:

$$p^{(t+1)}(m) = \frac{1}{|\mathcal{S}|} \sum_{Y_{\{r\}} \in \mathcal{S}} q^{(t)}(m \,|\, Y_{\{r\}}) \ , \qquad (6)$$

$$\mu_{m,s}^{(t+1)} = \frac{1}{\sum_{Y_{\{r\}} \in \mathcal{S}} q^{(t)}(m \,|\, Y_{\{r\}})} \sum_{Y_{\{r\}} \in \mathcal{S}} Y_s q^{(t)}(m \,|\, Y_{\{r\}}) \ . \qquad (7)$$

The $M\eta$ covariance matrices are:

$$\begin{aligned} \Sigma_{m,s}^{(t+1)} &= \frac{\sum_{Y_{\{r\}} \in \mathcal{S}, Y_s \in Y_{\{r\}}} q^{(t)}(m \,|\, Y_{\{r\}})}{\sum_{Y_r \in \mathcal{S}} q^{(t)}(m \,|\, Y_{\{r\}})} (Y_s - \mu_{m,s}^{(t+1)})(Y_s - \mu_{m,s}^{(t+1)})^T \\ &= \frac{\sum_{Y_{\{r\}} \in \mathcal{S}, Y_s \in Y_{\{r\}}} q^{(t)}(m \,|\, Y_{\{r\}})\, Y_s Y_s^T}{\sum_{Y_r \in \mathcal{S}} q^{(t)}(m \,|\, Y_{\{r\}})} \\ &- \frac{p^{(t+1)}(m)\, |\mathcal{S}|\, \mu_{m,s}^{(t+1)} \left(\mu_{m,s}^{(t+1)}\right)^T}{\sum_{Y_r \in \mathcal{S}} q^{(t)}(m \,|\, Y_{\{r\}})} \ . \end{aligned} \qquad (8)$$

The iteration process is stopped when the criterion increments are sufficiently small. The EM algorithm iteration scheme has the monotonic property: $L^{(t+1)} \ge L^{(t)}$, $t = 0, 1, 2, \ldots$ which implies the convergence of the sequence $\{L^{(t)}\}_0^\infty$ to a stationary point of the EM algorithm (local maximum or a saddle point of L).

B. Texture Synthesis

Textures without significant low frequencies such as Fig.3-bottom can be modeled using simple 3DGMM model. However, textures with substantial low frequencies (Figs.1,2,3-top,4) will benefit from a pyramidal model [14]. Such texture

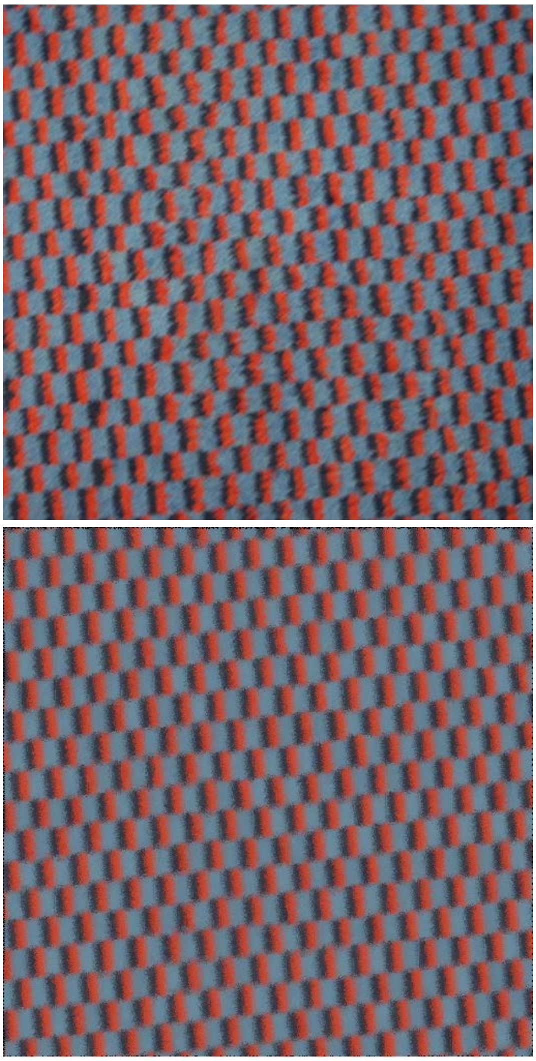

Fig. 1. A carpet textile (top) and its 3DGMM synthesis.

is down-sampled to several rough scale layers. The synthesis starts from the most down-sampled layer which is up-sampled to guide a finer layer synthesis and this process is repeated till the final fine resolution layer This frequency factorization based synthesis allow to use lower cardinality η 3DGMM models and consequently reduces the required learning data size for robust parameter estimation. The advantage of a mixture model is its simple synthesis based on the marginals:

$$p_{n\,|\,\rho}(Y_n\,|\,Y_{\{\rho\}}) = \sum_{m=1}^{M} W_m(Y_{\{\rho\}})\, p_n(Y_n\,|\,m) \ , \qquad (9)$$

where $W_m(Y_{\{\{\rho\}})$ are the a posterior component weights corresponding to the given sub-matrix $Y_{\{\rho\}} \subset Y_{\{r\}}$:

$$\begin{aligned} W_m(Y_{\{\rho\}}) &= \frac{p(m)P_\rho(Y_{\{\rho\}}\,|\,m)}{\sum_{j=1}^{M} p(j)P_\rho(Y_{\{\rho\}}\,|\,j)} \ , \qquad (10) \\ P_\rho(Y_{\{\rho\}}\,|\,m) &= \prod_{n\in\rho} p_n(Y_n\,|\,m) \ . \end{aligned}$$

There are several alternatives for the 3DGMM model synthesis [19]. The unknown multivariate vector-levels Y_n can be synthesized by random sampling from the conditional density (9) or the mixture RF can be approximated using the GMM mixture prediction.

The proposed method uses the 3DGMM model approximation by computing the conditional 3DGMM expectation:

$$\begin{aligned} E\{Y_n\} &= \int Y_n\, p_{n\,|\,\rho}(Y_n\,|\,Y_{\{\rho\}}) dY_n \\ &= \sum_{j=1}^{M} W_j(Y_{\{\rho\}})\, \mu_{jn} \ . \qquad (11) \end{aligned}$$

This is a fast non-iterative alternative for a 3DGMM model synthesis.

III. Experimental Results

Figs.1,2,3,4 illustrates the performance of our BTF-3DGMM model of selected textile BTF measurements. Unfortunately, the automatic texture quality evaluation is important but still unsolved problem and qualitative evaluation is for now possible only using impractical and expensive visual psychophysics. We have recently tested [23] on our texture fidelity benchmark (**http://tfa.utia.cas.cz**) several published state-of-the-art image quality measures and also one dedicated texture measure (STSIM) in several variants. These results clearly demonstrate that neither the standard image quality criteria (MSE, VSNR, VIF, SSIM, CW-SSIM) nor the STSIM texture criterion can be reliably used for texture quality validation (see for details [23]). It is easy to manifest failure counterexamples for each of these quality criteria. Thus our results can be checked only visually and all illustrations have also the corresponding original particular BTF measurement. All presented illustrations were chosen using the simple group viewers selection from possible alternative experimental parameters settings. Figs.1,4 were synthesised using four, Fig.3-top three, Fig.2 two, and Fig.3-bottom single layer models, respectively. The BTF-3DGMM model has also a weak restoration tendency (Fig.2-top right) thus the final results might require some additive noise (Fig.2-bottom). Fig.4 illustrates a sporadic modeling problem with complicated low frequency pattern for which there are no sufficiently large learning data set available.

The visual quality of the resulting textile synthetic textures generally surpasses the outputs of the previously published simpler MRF models [8], [22], [24], [18], [11]. This is also illustrated by comparing three BTF textile textures, blue knitted wool Fig. 2 and both coverings on Fig. 3, synthesized using the proposed model Figs. 2-bottom,3, and alternatively on Fig. 5 using the wide-sense Markovian models specified in

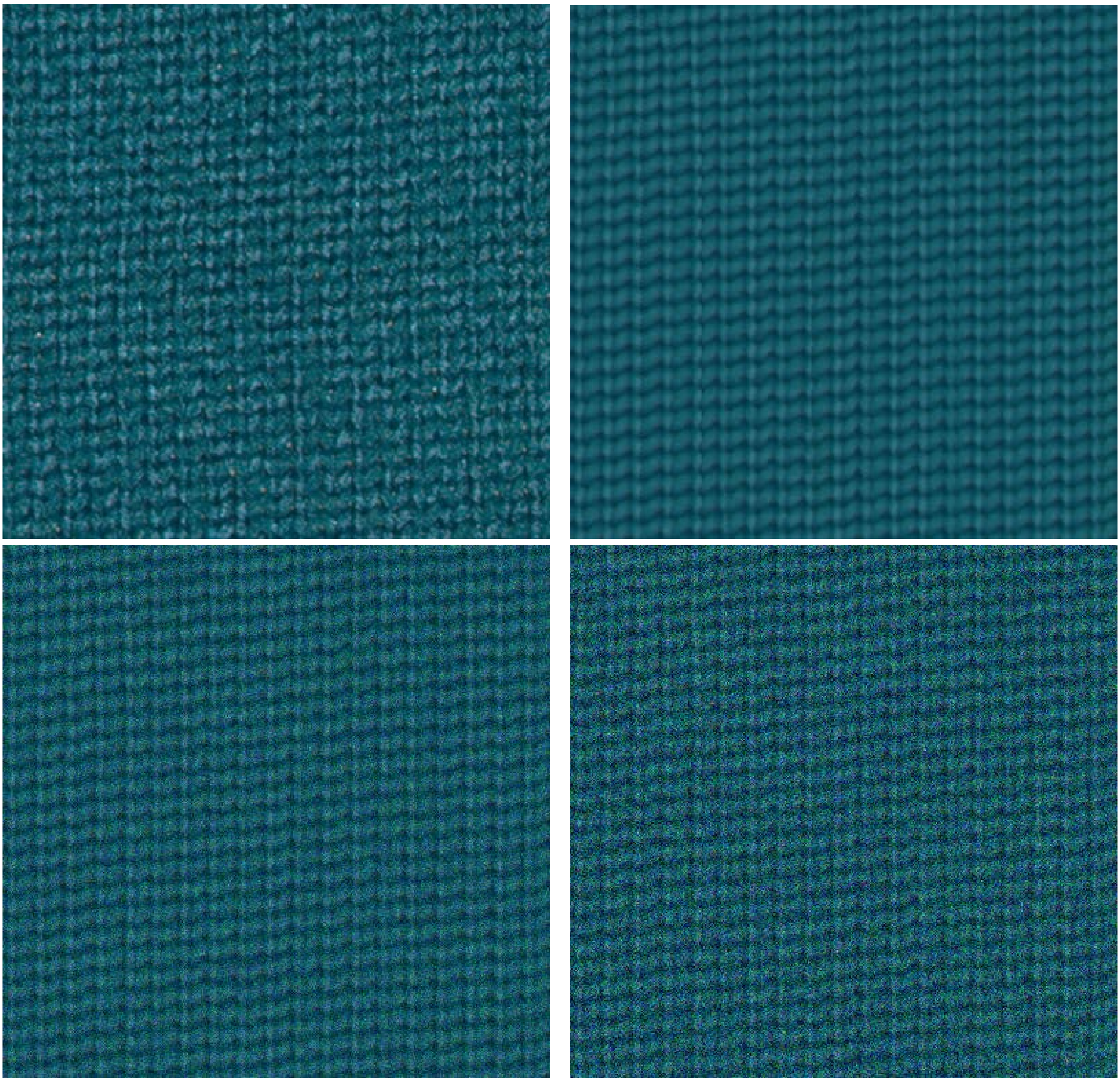

Fig. 2. An example of the BTF knitted wool measurement (top left), its BTF-3DGMM synthesis, and synthesis with additive Poisson (bottom left) and Gaussian noise.

[22]. Alternative 3D Gaussian Markov random field models [14] produce similarly unconvincing results.

IV. Conclusion

The proposed visual texture synthesis method is capable to simultaneously model and enlarge visual texture, to compress a measured texture or to restore its missing or noisy parts. The BTF-3DGMM model produces high quality results especially of regular or near-regular colorful BTF or multispectral textures, provided it has enough data to learn. Then it outperforms on these textures alternative methods based on Markovian random field models. The presented model has time consuming parameter estimation part and requires larger learning set than the alternative simpler 3DCAR method, which is not always available. The model synthesis is not iterative and thus relatively fast and the model can easily parallelized. Our future work will concentrate on optimal initialization of model's iterative equations.

Acknowledgment

This research was supported by by the Czech Science Foundation project GAČR 14-10911S.

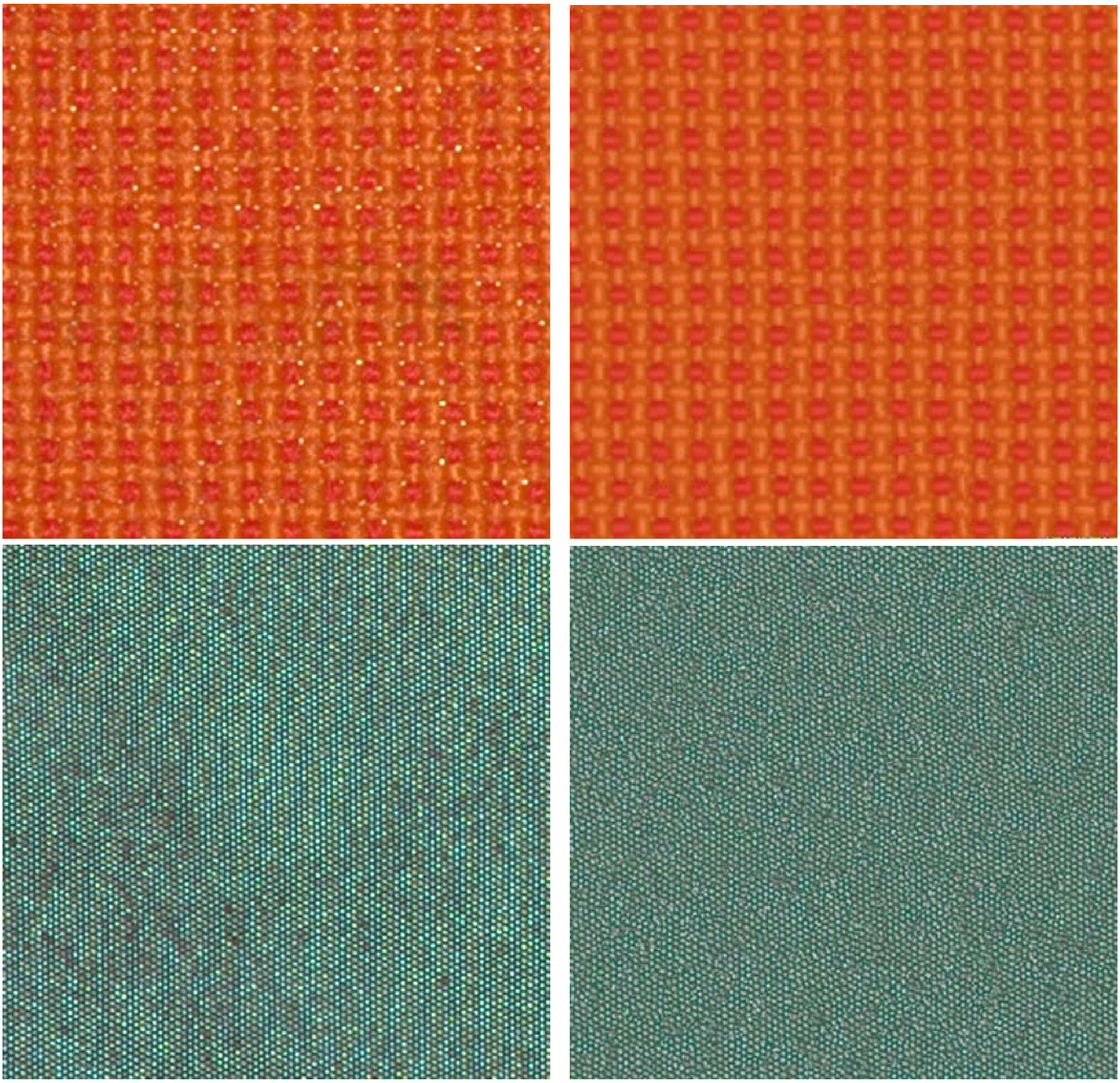

Fig. 3. An example of the BTF textile measurement (left) and its BTF-3DGMM synthesis.

REFERENCES

[1] J. De Bonet, "Multiresolution sampling procedure for analysis and synthesis of textured images," in *ACM SIGGRAPH 97*. ACM Press, 1997, pp. 361–368.

[2] A. A. Efros and W. T. Freeman, "Image quilting for texture synthesis and transfer," in *ACM SIGGRAPH 2001*, E. Fiume, Ed. ACM Press, 2001, pp. 341–346. [Online]. Available: citeseer.nj.nec.com/efros01image.html

[3] A. A. Efros and T. K. Leung, "Texture synthesis by non-parametric sampling," in *Proc. Int. Conf. on Computer Vision (2)*, Corfu, Greece, 1999, pp. 1033–1038. [Online]. Available: citeseer.nj.nec.com/efros99texture.html

[4] D. Heeger and J. Bergen, "Pyramid based texture analysis/synthesis," in *ACM SIGGRAPH 95*. ACM Press, 1995, pp. 229–238.

[5] Y. Xu, B. Guo, and H. Shum, "Chaos mosaic: Fast and memory efficient texture synthesis," Redmont, Tech. Rep. MSR-TR-2000-32, 2000.

[6] S. Zelinka and M. Garland, "Towards real-time texture synthesis with the jump map," in *13th European Workshop on Rendering*, 2002, p. 99104.

[7] J. Bennett and A. Khotanzad, "Multispectral random field models for synthesis and analysis of color images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 327–332, March 1998.

[8] M. Haindl, "Texture synthesis," *CWI Quarterly*, vol. 4, no. 4, pp. 305–331, December 1991.

[9] M. Haindl and V. Havlíček, "Multiresolution colour texture synthesis," in *Proceedings of the 7th International Workshop on Robotics in Alpe-Adria-Danube Region*, K. Dobrovodský, Ed. Bratislava: ASCO Art, June 1998, pp. 297–302.

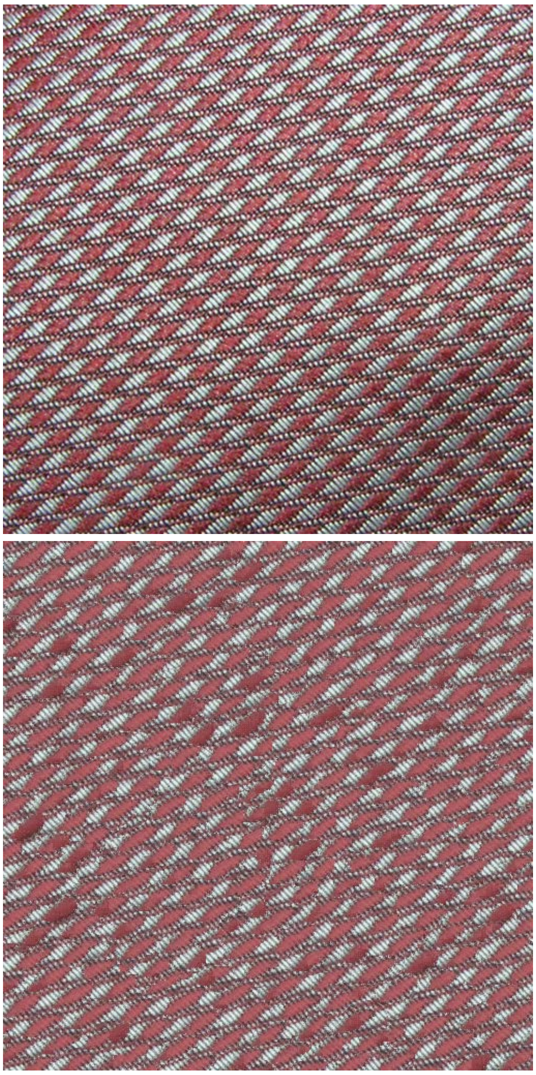

Fig. 4. A complicated large-pattern measured textile (top) and its 3DGMM synthesis.

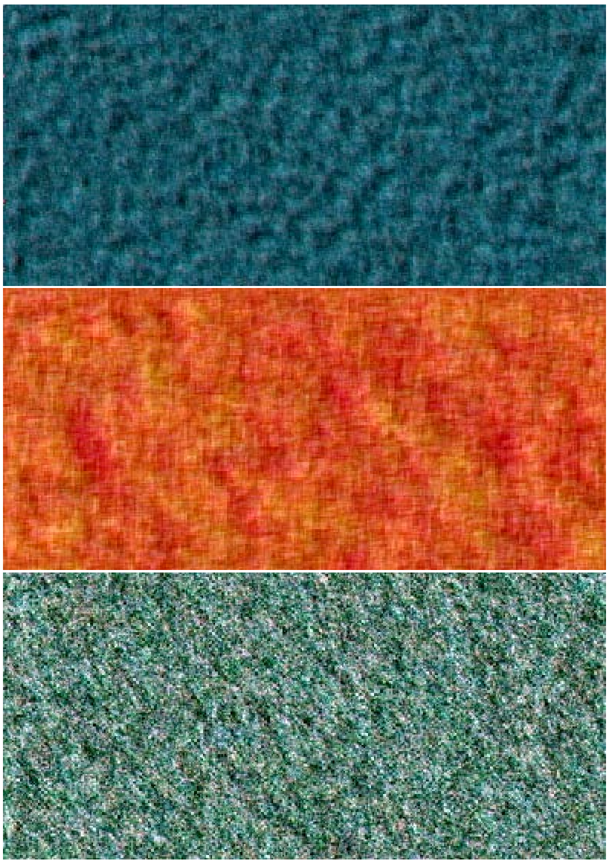

Fig. 5. Knitted wool and covering texture patterns from Figs.2,3 synthesized using wide-sense Markovian models [22].

[10] G. Gimelfarb, *Image Textures and Gibbs Random Fields*. Kluwer Academic Publishers, 1999.

[11] M. Haindl and V. Havlíček, "A multiresolution causal colour texture model," *Lecture Notes in Computer Science*, no. 1876, pp. 114–122, August 2000.

[12] R. Paget and I. D. Longstaff, "Texture synthesis via a noncausal nonparametric multiscale markov random field," *IEEE Trans. on Image Processing*, vol. 7, no. 8, pp. 925–932, 1998.

[13] S. Zhu, X. Liu, and Y. Wu, "Exploring texture ensembles by efficient markov chain monte carlo - toward a "trichromacy" theory of texture," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 6, pp. 554–569, June 2000.

[14] M. Haindl and J. Filip, *Visual Texture*, ser. Advances in Computer Vision and Pattern Recognition. London: Springer-Verlag London, January 2013.

[15] F. Nicodemus, J. Richmond, J. Hsia, I. Ginsburg, and T. Limperis, *Geometrical considerations and nomenclature for reflectance*, ser. NBS Monograph. Washington, D. C.: National Bureau of Standards, U.S. Department of Commerce, October 1977, no. 160.

[16] K. J. Dana, S. K. Nayar, B. van Ginneken, and J. J. Koenderink, "Reflectance and texture of real-world surfaces," in *CVPR*. IEEE Computer Society, 1997, pp. 151–157.

[17] J. Filip and M. Haindl, "Bidirectional texture function modeling: A state of the art survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 1921–1940, 2009.

[18] M. Haindl and J. Filip, "Extreme compression and modeling of bidirectional texture function," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 10, pp. 1859–1865, 2007. [Online]. Available: http://doi.ieeecomputersociety.org/10.1109/TPAMI.2007.1139

[19] M. Haindl, V. Havlíček, and J. Grim, "Probabilistic mixture-based image modelling," *Kybernetika*, vol. 46, no. 3, pp. 482–500, 2011. [Online]. Available: http://www.kybernetika.cz/content/2011/3/482/paper.pdf

[20] M. Haindl and V. Havlíček, "Color texture restoration." 445 Hoes Lane, Piscataway, NJ 08854, USA: IEEE, July 2015, pp. 13–18. [Online]. Available: http://dx.doi.org/10.1109/ICCIS.2015.7274540

[21] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society, B*, vol. 39, no. 1, pp. 1–38, 1977.

[22] M. Haindl and V. Havlíček, "A multiscale colour texture model," in *Proceedings of the 16th International Conference on Pattern Recognition*, R. Kasturi, D. Laurendeau, and C. Suen, Eds. Los Alamitos: IEEE Computer Society, August 2002, pp. 255–258. [Online]. Available: http://dx.doi.org/10.1109/ICPR.2002.1044676

[23] M. Haindl and M. Kudělka, "Texture fidelity benchmark," in *Computational Intelligence for Multimedia Understanding (IWCIM), 2014 International Workshop on*. Los Alamitos: IEEE Computer Society CPS, November 2014, pp. 1 – 5. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7008812&isnumber=7008791

[24] M. Haindl, V. Remeš, and V. Havlíček, "Potts compound markovian texture model," in *Proceedings of the 21st International Conference on Pattern Recognition, ICPR 2012*. Los Alamitos: IEEE Computer Society CPS, November 2012, pp. 29 – 32.