Han, Zhi; **Xu, Zongben**; **Zhu, Song-Chun**

Video primal sketch: a unified middle-level representation for video. (English) Zbl 1343.94010

J. Math. Imaging Vis. 53, No. 2, 151-170 (2015).

Summary: This paper presents a middle-level video representation named video primal sketch (VPS), which integrates two regimes of models: (i) sparse coding model using static or moving primitives to explicitly represent moving corners, lines, feature points, etc., (ii) FRAME /MRF model reproducing feature statistics extracted from input video to implicitly represent textured motion, such as water and fire. The feature statistics include histograms of spatio-temporal filters and velocity distributions. This paper makes three contributions to the literature: (i) Learning a dictionary of video primitives using parametric generative models; (ii) Proposing the spatio-temporal FRAME and motion-appearance FRAME models for modeling and synthesizing textured motion; and (iii) Developing a parsimonious hybrid model for generic video representation. Given an input video, VPS selects the proper models automatically for different motion patterns and is compatible with high-level action representations. In the experiments, we synthesize a number of textured motion; reconstruct real videos using the VPS; report a series of human perception experiments to verify the quality of reconstructed videos; demonstrate how the VPS changes over the scale transition in videos; and present the close connection between VPS and high-level action models.

MSC:

94A08 Image processing (compression, reconstruction, etc.) in information and communication theory

Keywords:

middle-level vision; video representation; textured motion; dynamic texture synthesis; primal sketch

Software:

Steerable pyramid

Full Text: DOI

References:

[1] Adelson, E., Bergen, J.: Spatiotemporal energy models for the perception of motion. JOSA A \textbf{2}(2), 284-299 (1985)

[2] Bergen, J.R., Adelson, E.H.: In: Regan, D. (ed.) Theories of Visual Texture Perception. Spatial Vision. CRC Press, Boca Raton, FL (1991) · Zbl 1012.68698

[3] Besag, J.: Spatial interactions and the statistical analysis of lattice systems. J. R. Stat. Soc. Ser. B \textbf{36}, 192-236 (1974) · Zbl 0327.60067

[4] Black, M.J., Fleet, D.J.: Probabilistic detection and tracking of motion boundaries. IJCV \textbf{38}(3), 231-245 (2000) · Zbl 1012.68694

[5] Bouthemy, P., Hardouin, C., Piriou, G., Yao, J.: Mixed-state auto-models and motion texture modeling. J. Math. Imaging Vis. \textbf{25}(3) (2006)

[6] Campbell, N.W., Dalton, C., Gibson, D., Thomas, B. : Practical generation of video textures using the auto-regressive process. In: Proceedings of British Machine Vision Conference, pp 434-443 (2002)

[7] Chan, A.B., Vasconcelos, N.: Modeling, clustering, and segmenting video with mixtures of dynamic textures. PAMI \textbf{30}(5), 909-926 (2008)

[8] Chaudhry, R., Ravichandran, A., Hager, G., Vidal, R. : Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions. CVPR (2009)

[9] Chubb, C; Landy, MS; Landy, MS (ed.); etal., Orthogonal distribution analysis: a new approach to the study of texture perception, (1991), Cambridge, MA

[10] Comaniciu, D., Ramesh, V., Meer, P.: Kernel-based object tracking. PAMI \textbf{25}(5), 564-577 (2003)

[11] Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. CVPR (2005)

[12] Dalal, N., Triggs, B., Schmid, C. : Human detection using oriented histograms of flow and appearance. ECCV (2006)

[13] Derpanis, K.G., Wildes, R.P. : Dynamic texture recognition based on distributions of spacetime oriented structure. CVPR (2010)

[14] Doretto, G., Chiuso, A., Wu, Y.N., Soatto, S.: Dynamic textures. IJCV \textbf{51}(2), 91-109 (2003) · Zbl 1030.68646

[15] Elder, J., Zucker, S.: Local scale control for edge detection and blur estimation. PAMI \textbf{20}(7), 699-716 (1998)

[16] Fan, Z., Yang, M., Wu, Y., Hua, G., Yu, T. : Effient optimal kernel placement for reliable visual tracking. CVPR (2006)

[17] Gong, H.F., Zhu, S.C.: Intrackability: characterizing video statistics and pursuing video representations. IJCV \textbf{97}(33), 255-275 (2012) · Zbl 1235.68265

[18] Guo, C., Zhu, S.C., Wu, Y.N.: Primal sketch: integrating texture and structure. CVIU \textbf{106}(1), 5-19 (2007)

[19] Han, Z., Xu, Z., Zhu, S.C.: Video primal sketch: a generic middle-level representation of video. ICCV (2011) · Zbl 1343.94010

[20] Heeger, D.: Model for the extraction of image flow. JOSA A \textbf{4}(8), 1455-1471 (1987)

[21] Heeger, D.J., Bergen, J.R.: Pyramid-based texture analysis/synthesis. SIGGRAPH (1995) · Zbl 1012.68698

[22] Kim, T., Shakhnarovich, G., Urtasun, R.: Sparse coding for learning interpretable spatio-temporal primitives. NIPS (2010)

[23] Lindeberg, T., Fagerstrm, D.: Scale-space with casual time direction. ECCV (1996)

[24] Maccormick, J., Blake, A.: A probabilistic exclusion principle for tracking multiple objects. IJCV \textbf{39}(1), 57-71 (2000) · Zbl 1060.68629

[25] Mallat, S., Zhang, Z.: Matching pursuits with time-frequency dictionaries. IEEE TSP \textbf{41}(12), 3397-3415 (1993) · Zbl 0842.94004

[26] Marr, D.: Vision. W H Freeman and Company, San Francisco, CA (1982)

[27] Olshausen, B.A.: Learning sparse, overcomplete representations of time-varying natural images. ICIP (2003)

[28] Olshausen, B.A., Field, D.J.: Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381 (1996)

[29] Portilla, J; Simoncelli, E, A parametric texture model based on joint statistics of complex wavelet coefficients, IJCV, 40, 49-71, (2000) · Zbl 1012.68698

[30] Ravichandran, A., Chaudhry, R., Vidal, R.: View-invariant dynamic texture recognition using a bag of dynamical systems. CVPR (2009)

[31] Schuldt, C., Laptev, I., Caputo, B.: Recognizing human actions: a local svm approach. ICPR (2004)

[32] Serby, D., Koller-Meier, S., Gool, L.V.: Probabilistic object tracking using multiple features. ICPR (2004)

[33] Shi, K., Zhu, S.C.: Mapping natural image patches by explicit and implicit manifolds. CVPR (2007)

[34] Silverman, M.S., Grosof, D.H., Valois, R.L.D., Elfar, S.D.: Spatial-frequency organization in primate striate cortex. Proc. Natl. Acad. Sci. \textbf{86}, 711-715 (1989)

[35] Szummer, M., Picard, R.W.: Temporal texture modeling. ICIP (1996)

[36] Wang, Y.Z., Zhu, S.C.: Analysis and synthesis of textured motion: particles and waves. PAMI \textbf{26}(10), 1348-1363 (2004)

[37] Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error measurement to structural similarity. IEEE TIP \textbf{13}(4) (2004)

[38] Wildes, R., Bergen, J.: Qualitative spatiotemporal analysis using an oriented energy representation. ECCV (2000)

[39] Wu, Y.N., Zhu, S.C., Liu, X.W.: Equivalence of julesz ensemble and frame models. IJCV \textbf{38}(3), 247-265 (2000) · Zbl 1012.68692

[40] Yao, B., Zhu, S.C.: Learning deformable action templates from cluttered videos. ICCV (2009)

[41] Yuan, F., Prinet, V., Yuan, J.: Middle-level representation for human activities recognition: the role of spatio-temporal relationships. ECCVW (2010)

[42] Zhu, S.C., Wu, Y.N., Mumford, D.B.: Filters, random field and maximum entropy (FRAME): towards a unified theory for texture modeling. IJCV \textbf{27}(2), 107-126 (1998)