Unsupervised Dynamic Textures Segmentation

Michal Haindl and Stanislav Mikeš

Institute of Information Theory and Automation
of the ASCR, Prague, Czech Republic
{haindl,xaos}@utia.cz

Abstract. This paper presents an unsupervised dynamic colour texture segmentation method with unknown and variable number of texture classes. Single regions with dynamic textures can furthermore change their location as well as their shape. Individual dynamic multispectral texture mosaic frames are locally represented by Markovian features derived from four directional multispectral Markovian models recursively evaluated for each pixel site. Estimated frame-based Markovian parametric spaces are segmented using an unsupervised segmenter derived from the Gaussian mixture model data representation which exploits contextual information from previous video frames segmentation history. The segmentation algorithm for every frame starts with an over segmented initial estimation which is adaptively modified until the optimal number of homogeneous texture segments is reached. The presented method is objectively numerically evaluated on the dynamic textural test set from the Prague Segmentation Benchmark.

Keywords: dynamic texture segmentation, unsupervised segmentation.

1 Introduction

Many automated static or dynamic visual data analysis systems build on the segmentation as the fundamental process which affects the overall performance of any analysis. Visual scene regions, homogeneous with respect to some usually textural or colour measure, which result from a segmentation algorithm are analysed in subsequent interpretation steps. Dynamic texture-based (DT) image segmentation is an area of novel research activity in recent years and several algorithms were published in consequence of all this effort. Different published methods are difficult to compare because of incompatible assumptions (gray-scale, fixed or known number of regions, segmentation or retrieval, constant shape and/or location of texture regions, etc.), lack of a comprehensive analysis together with accessible experimental data. Gray scale dynamic texture segmentation or retrieval was addressed in few papers [1–5], while colour texture retrieval based on VLBP [6] or DT segmentation [7], based on the geodesic active contour algorithm and partial shape matching to obtain partial match costs between regions of subsequent frames, were addressed to even lesser extent. However all available published results indicate that the ill-defined dynamic texture

R. Wilson et al. (Eds.): CAIP 2013, Part I, LNCS 8047, pp. 433–440, 2013.

segmentation problem is far from being satisfactorily solved. Spatial interaction models and especially Markov random fields-based models are increasingly popular for texture representation [8, 9], etc. Several researchers dealt with the difficult problem of unsupervised segmentation using these models see for example [10–13] or [14] which is also generalized to dynamic textures and addressed in this paper.

The contribution of the paper is a novel unsupervised dynamic multispectral texture segmentation method with unknown and variable number of texture classes, and regions (with dynamic texture) which can in addition change their location as well as their shape. Thus the method relaxes most of the alternative approaches [1–5] limitations (gray-scale textures, fixed or known number of regions, fixed regions shape and locations) which prevent their practical applications.

The outline of this paper is as follows. Section 2 presents our Markovian multispectral texture representation. Section 3 outlines the unsupervised segmenter, followed by the experimental verification in the subsequent Section 4 and concluding Section 5.

2 Dynamic Texture Representation

Dynamic multispectral textures would require a four dimensional (4D) model or some of its lower dimensional approximation such as a set of spectrally factorized 3D models. However we assume to model each dynamic texture frame separately and thus a 3D static smooth textural model is sufficient for its adequate representation. We assume that single multispectral frame textures can be locally modeled using a 3D simultaneous causal auto-regressive random field model (AR3D). This model can be expressed as a stationary causal uncorrelated noise driven 3D auto-regressive process [15]:

$$Y_r = \gamma X_r + e_r \quad , \tag{1}$$

where $X_r = [Y_{r-s}^T : \forall s \in I_r^c]^T$ is a vector of the contextual neighbours Y_{r-s}, I_r^c is a causal neighbourhood index set of the model with the cardinality $\eta = card(I_r^c)$, $\gamma = [A_1, \ldots, A_\eta]$ is the $d \times d\eta$ parameter matrix containing parametric sub-matrices A_s for each contextual neighbour Y_{r-s}, d is the number of spectral bands, e_r is a white Gaussian noise vector with zero mean and a constant but unknown covariance, and $r, r-1, \ldots$ is a chosen direction of movement on the image index lattice I. The selection of an appropriate model support (I_r^c) is important to obtain good texture representation for realistic texture synthesis but less important for adequate texture segmentation which works only with site specific parameters. Both, the optimal neighbourhood as well as the Bayesian parameters estimation of the AR3D model can be found analytically under few additional and acceptable assumptions using the Bayesian approach (see details in [15]). The local model parameters can be advantageously evaluated using the recursive Bayesian parameter estimator for every DT frame as follows:

$$\hat{\gamma}_{r-1}^T = \hat{\gamma}_{r-2}^T + \frac{V_{x(r-2)}^{-1} X_{r-1} (Y_{r-1} - \hat{\gamma}_{r-2} X_{r-1})^T}{(1 + X_{r-1}^T V_{x(r-2)}^{-1} X_{r-1})} \quad , \tag{2}$$

where the data accumulation matrix is

$$V_{x(r-1)} = \sum_{k=1}^{r-1} X_k X_k^T + V_{x(0)} \quad . \tag{3}$$

Thus the parameter matrix estimate can be easily upgraded after moving to a new lattice location $(r - 1 \longrightarrow r)$. The model is very fast, hence the local texture for each pixel can be represented by four directional parametric vectors corresponding to four distinct models. Each vector contains local estimations of the AR3D model parameters. These models have identical contextual neighbourhood I_r^c but they differ in their major movement direction (top-down, bottom-up, rightward, leftward), i.e.,

$$\tilde{\gamma}_{r,o}^T = \{\hat{\gamma}_{r,o}^t, \hat{\gamma}_{r,o}^b, \hat{\gamma}_{r,o}^r, \hat{\gamma}_{r,o}^l\}^T \quad , \tag{4}$$

where $o = 1, \dots, n$ is the DT frame number.

3 Gaussian Mixture Segmenter

Multispectral texture segmentation is done by clustering in the AR3D parameter space Θ_o defined on the lattice I for every frame o where

$$\Theta_{r,o} = \bar{\gamma}_{r,o}^T \tag{5}$$

is the decorrelated parameter vector (4) computed for the lattice location r (the frame index is further left out to simplify notation). We assume that this parametric space can be represented using the Gaussian mixture model with diagonal covariance matrices due to the previous CAR parametric space decorrelation. The Gaussian mixture model for AR3D parametric representation is as follows:

$$p(\Theta_r) = \sum_{i=1}^{K} p_i \, p(\Theta_r \,|\, \nu_i, \Sigma_i) \quad , \tag{6}$$

$$p(\Theta_r \,|\, \nu_i, \Sigma_i) = \frac{|\Sigma_i|^{-\frac{1}{2}}}{(2\pi)^{\frac{d}{2}}} \, e^{-\frac{(\Theta_r - \nu_i)^T \Sigma_i^{-1} (\Theta_r - \nu_i)}{2}} \quad . \tag{7}$$

The mixture model equations (6),(7) are solved using a modified EM algorithm. The algorithm is initialised, for the first DT frame, using ν_i, Σ_i statistics estimated from the corresponding rectangular subimages obtained by regular division of the input texture mosaic. An alternative initialisation can be random choice of these statistics. For each possible couple of rectangles the Kullback Leibler divergence

$$D\left(p(\Theta_r \,|\, \nu_i, \Sigma_i) \,||\, p(\Theta_r \,|\, \nu_j, \Sigma_j)\right) = \int_{\Omega} p(\Theta_r \,|\, \nu_i, \Sigma_i) \log \left(\frac{p(\Theta_r \,|\, \nu_i, \Sigma_i)}{p(\Theta_r \,|\, \nu_j, \Sigma_j)} \right) d\Theta_r \tag{8}$$

is evaluated and the most similar rectangles, i.e.,

$$\{i,j\} = \arg\min_{k,l} D\left(p(\Theta_r \,|\, \nu_l, \Sigma_l) \,||\, p(\Theta_r \,|\, \nu_k, \Sigma_k)\right) \tag{9}$$

are merged together in each step. This initialization results in K_{ini} subimages and recomputed statistics ν_i, Σ_i . $K_{ini} > K$ where K is the optimal number of textured segments to be found by the algorithm. All the subsequent DT frames are initialized either from the corrected statistics $\hat{\nu}_{i,o-1}, \hat{\Sigma}_{i,o-1}$ for $i = 1,\ldots,K$ computed from the trimmed segmented regions in the previous frame $o-1$ or with random parameter values $\hat{\nu}_{i,o-1}, \hat{\Sigma}_{i,o-1}$ $\quad i = K+1,\ldots,K_{ini}$ for possibly newly (re)appearing regions. Two steps of the EM algorithm are repeating after initialisation. The components with smaller weights than a fixed threshold $(p_j < \frac{0.1}{K_{ini}})$ are eliminated. For every pair of components we estimate their Kullback Leibler divergence (8). From the most similar couple, the component with the weight smaller than the threshold is merged to its stronger partner and all statistics are actualised using the EM algorithm. The algorithm stops when either the likelihood function has negligible increase $(\mathcal{L}_t - \mathcal{L}_{t-1} < 0.05)$ or the maximum iteration number threshold is reached.

The parametric vectors representing texture mosaic pixels are assigned to the clusters according to the highest component probabilities, i.e., Y_r is assigned to the cluster ω_j if

$$\pi_{r,j} = max_j \sum_{s \in I_r} w_s \, p(\Theta_{r-s} \,|\, \nu_j, \Sigma_j) \ , \tag{10}$$

where w_s are fixed distance-based weights, I_r is a rectangular neighbourhood and $\pi_{r,j} > \pi_{thre}$ (otherwise the pixel is unclassified). The area of single cluster blobs is evaluated in the post-processing thematic map filtration step. Regions with similar statistics are merged. Thematic map blobs with area smaller than a given threshold are attached to its neighbour with the highest similarity value. Finally, the resulting region classes are remapped to ensure their between frame consistency.

4 Experimental Results

The algorithm was tested on the natural colour dynamic textural mosaics from the Prague Texture Segmentation Data-Generator and Benchmark [16]. The benchmark (http://mosaic.utia.cas.cz) test mosaics with varying layouts and each cell texture membership are randomly generated and filled with dynamic colour textures from the Dyntex database [17]. The benchmark ranks segmentation algorithms according to a chosen criterion. The benchmark has implemented the majority of segmentation criteria used for both supervised or unsupervised algorithms evaluation. Twenty seven evaluation criteria (see their definition in [16]) are categorized into four groups: region-based (5+5), pixel-wise (12), consistency measures (2), and clustering comparison criteria (3) and permit detailed and objective study of any segmentation method properties. Tab.1 compares

Table 1. Dynamic A benchmark results for DTAR3D+EM (e+pp), DTAR3D+EM (pp), DTAR3D+EM; (Benchmark criteria [16]: **CS = correct segmentation; OS = over-segmentation; US = under-segmentation; ME = missed error; NE = noise error; O = omission error; C = commission error; CA = class accuracy; CO = recall - correct assignment; CC = precision - object accuracy; I. = type I error; II. = type II error; EA = mean class accuracy estimate; MS = mapping score; RM = root mean square proportion estimation error; CI = comparison index; GCE = Global Consistency Error; LCE = Local Consistency Error; dD = Van Dongen metric; dM = Mirkin metric; dVI = variation of information**). Arrows directions denote the required criterion motion, the criteria rank numbers are down-sized on the right with the average rank besides the method label. The bold numbers are the best criterion values, while italic numbers are the worst criterion values.

	Benchmark – Dynamic A		
	DTAR3D+EM e+pp (1.33)	DTAR3D+EM pp (1.86)	DTAR3D+EM (2.62)
↑ *CS*	**92.68** 1	60.75 2	*60.12* 3
↓ *OS*	*39.47* 3	20.37 2	**14.78** 1
↓ *US*	0.00 1	0.00 1	0.00 1
↓ *ME*	**0.00** 1	*35.77* 2	*35.77* 2
↓ *NE*	**0.00** 1	36.52 2	*36.76* 3
↓ *O*	**3.23** 1	10.07 2	*11.08* 3
↓ *C*	13.25 2	**12.45** 1	*14.56* 3
↑ *CA*	**87.03** 1	81.26 2	*80.10* 3
↑ *CO*	**92.68** 1	84.42 2	*83.69* 3
↑ *CC*	*94.01* 3	**95.85** 1	95.18 2
↓ *I.*	**7.32** 1	15.58 2	*16.31* 3
↓ *II.*	*1.32* 3	**0.76** 1	0.89 2
↑ *EA*	**92.80** 1	89.01 2	*88.30* 3
↑ *MS*	**89.07** 1	82.41 2	*81.35* 3
↓ *RM*	**2.56** 1	*5.54* 3	5.21 2
↑ *CI*	**93.07** 1	89.56 2	*88.86* 3
↓ *GCE*	**11.13** 1	12.39 2	*13.51* 3
↓ *LCE*	**7.02** 1	11.03 2	*12.21* 3
↓ *dD*	**7.27** 1	11.68 2	*12.37* 3
↓ *dM*	**4.95** 1	6.36 2	*6.80* 3
↓ *dVI*	**13.18** 1	13.93 2	*13.99* 3

the overall (average over all DT frames) benchmark performance of the proposed algorithm ($DTAR3D + EM(e + pp)$) with postprocessing (pp) and robust trimmed initialization (e) with its alternative versions. The results demonstrate very good performance on all criteria with the exception of over-segmentation tendency and slightly worse variation of information criterion. We could not compare our results with few published alternative DT segmenters [1, 2, 4] because neither their code, nor their experimental segmentation data are publicly available, however the static single-frame (AR3D+EM) version of the method was extensively evaluated and compared with several alternative methods (22

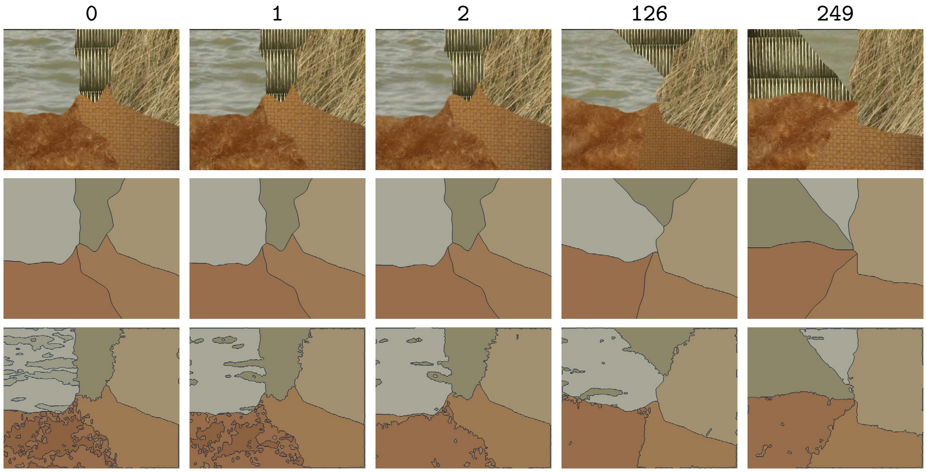

Fig. 1. Selected experimental dynamic texture mosaic frames (0, 1, 2, 126, 249), ground truth from the benchmark (middle row), and the corresponding segmentation results (DTAR3D+EM e+pp - bottom))

other leading unsupervised segmenters) in the segmentation benchmark. The static method proved its very good performance and outperformed most of these alternatives (see details in http://mosaic.utia.cas.cz). For example, the important correct region segmentation criterion (CS) is 25% better than for the HGS method [18], under-segmentation is low as well as missed and noise errors [19].

Fig.1 shows five selected (three from the beginning, one from the middle and one from the end of the sequence) 720×576 frames from the experimental benchmark mosaics created from five Dyntex dynamic colour textures (`47fa110` - *curtain*, `54aa110` - *curly hair*, `54ac210` - *straw*, `54pd110` - *escalator*, and `571b110` - *water*). While the first frame suffers with over-segmentation the contextual information propagated from previous frames significantly improves the segmentation consistency. Hard natural Dyntex textures were chosen for comparison rather than synthesised (for example using the generative AR3D model or some other Markov random field model) ones because they are expected to be more difficult for the underlying segmentation model. Resulting segmentation results are promising even if we could not compare them with alternative DT segmentation methods. The time for an unoptimized parameter estimation is 170 s and segmentation time is 10 s per frame. Our results can be further improved by an appropriate more elaborate postprocessing or frame model initialization.

5 Conclusions

We proposed novel method for fast unsupervised dynamic texture or video segmentation with unknown variable number of classes based on the underlying three dimensional Markovian local image representation and the Gaussian mixture parametric space models. Single homogeneous texture regions can not only dynamically change their location but simultaneously also their shape. Textural regions can also disappear temporarily or permanently and new regions can appear at any time. Although the algorithm uses the random field type data model it is very fast because it uses efficient recursive parameter estimation of the model and therefore is much faster than the usual Markov chain Monte Carlo estimation approach needed for Markovian models. Segmentation methods typically suffer with lot of application dependent parameters to be experimentally estimated. Our method requires only a contextual neighbourhood selection and two additional thresholds all of them having an intuitive meaning. The algorithm's performance is demonstrated on the extensive benchmark objective tests on natural dynamic texture mosaics. The static version of our method outperforms several alternative unsupervised segmentation algorithms and it is also faster than most of them. These dynamic texture unsupervised segmentation test results are encouraging and we proceed with more elaborate post-processing and some modification of the texture representation model.

References

1. Doretto, G., Cremers, D., Favaro, P., Soatto, S.: Dynamic texture segmentation. In: Proceedings of the 9th IEEE International Conference on Computer Vision, vol. 2, pp. 1236–1242 (2003)
2. Péteri, R., Chetverikov, D.: Dynamic texture recognition using normal flow and texture regularity. In: Marques, J.S., Pérez de la Blanca, N., Pina, P. (eds.) IbPRIA 2005, Part II. LNCS, vol. 3523, pp. 223–230. Springer, Heidelberg (2005)
3. Chan, A.B., Vasconcelos, N.: Classifying video with kernel dynamic textures. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2007), pp. 1–6. IEEE Computer Society (2007)
4. Chan, A.B., Vasconcelos, N.: Layered dynamic textures. IEEE Transactions on Pattern Analalysis and Machine Intelligence 31(10), 1862–1879 (2009)
5. Chen, J., Zhao, G., Salo, M., Rahtu, E., Pietikinen, M.: Automatic dynamic texture segmentation using local descriptors and optical flow. IEEE Transactions on Image Processing (2012)
6. Zhao, G., Pietikäinen, M.: Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(6), 915–928 (2007)
7. Donoser, M., Urschler, M., Riemenschneider, H., Bischof, H.: Highly consistent sequential segmentation. In: Heyden, A., Kahl, F. (eds.) SCIA 2011. LNCS, vol. 6688, pp. 48–58. Springer, Heidelberg (2011)
8. Kashyap, R.L.: Image models. In: Young, T.Y., Fu, K.S. (eds.) Handbook of Pattern Recognition and Image Processing. Academic Press, New York (1986)
9. Haindl, M.: Texture synthesis. CWI Quarterly 4(4), 305–331 (1991)
10. Panjwani, D.K., Healey, G.: Markov random field models for unsupervised segmentation of textured color images. IEEE Transactions on Pattern Analysis and Machine Intelligence 17(10), 939–954 (1995)
11. Manjunath, B.S., Chellapa, R.: Unsupervised texture segmentation using Markov random field models. IEEE Transactions on Pattern Analysis and Machine Intelligence 13, 478–482 (1991)

12. Haindl, M.: Texture segmentation using recursive markov random field parameter estimation. In: Bjarne, K.E., Peter, J. (eds.) Proceedings of the 11th Scandinavian Conference on Image Analysis, Lyngby, Denmark, pp. 771–776. Pattern Recognition Society of Denmark (June 1999)
13. Haindl, M., Mikeš, S., Pudil, P.: Unsupervised hierarchical weighted multi-segmenter. In: Benediktsson, J.A., Kittler, J., Roli, F. (eds.) MCS 2009. LNCS, vol. 5519, pp. 272–282. Springer, Heidelberg (2009)
14. Haindl, M., Mikeš, S.: Unsupervised texture segmentation using multispectral modelling approach. In: Tang, Y.Y., Wang, S.P., Yeung, D.S., Yan, H., Lorette, G. (eds.) Proceedings of the 18th International Conference on Pattern Recognition, ICPR 2006, vol. II, pp. 203–206. IEEE Computer Society, Los Alamitos (2006)
15. Haindl, M., Šimberová, S.: A multispectral image line reconstruction method. In: Theory & Applications of Image Analysis, pp. 306–315. World Scientific Publishing Co., Singapore (1992)
16. Haindl, M., Mikeš, S.: Texture segmentation benchmark. In: Lovell, B., Laurendeau, D., Duin, R. (eds.) Proceedings of the 19th International Conference on Pattern Recognition, ICPR 2008, pp. 1–4. IEEE Computer Society (2008)
17. Péteri, R., Fazekas, S., Huiskes, M.J.: DynTex: A Comprehensive Database of Dynamic Textures. Pattern Recognition Letters 31(12), 1627–1632 (2010), http://projects.cwi.nl/dyntex/
18. Hoang, M.A., Geusebroek, J.M., Smeulders, A.W.M.: Color texture measurement and segmentation. Signal Processing 85(2), 265–275 (2005)
19. Hoover, A., Jean-Baptiste, G., Jiang, X., Flynn, P.J., Bunke, H., Goldgof, D.B., Bowyer, K., Eggert, D.W., Fitzgibbon, A., Fisher, R.B.: An experimental comparison of range image segmentation algorithms. IEEE Transaction on Pattern Analysis and Machine Intelligence 18(7), 673–689 (1996)
20. Kittler, J.V., Marik, R., Mirmehdi, M., Petrou, M., Song, J.: Detection of defects in colour texture surfaces. In: IAPR Workshop on Machine Vision Application, Tokyo, Japan, pp. 558–567 (1994)
21. Mitra, P., Murthy, C.A., Pal, S.K.: Unsupervised feature selection using feature similarity. IEEE Transactions on Pattern Analysis and Machine Intelligence 24, 301–312 (2002)
22. Mirmehdi, M., Marik, R., Petrou, M., Kittler, J.: Iterative morphology for fault detection in stochastic textures. Electronic Letters 32(5), 443–444 (1996)