SUPERVISED TEXTURE SEGMENTATION USING 2D LSTM NETWORKS

Wonmin Byeon[⋆] *Thomas M. Breuel*[†]

[⋆†] Department of Computer Science, University of Kaiserslautern, Germany
[⋆] German Research Center for Articial Intelligence (DFKI), Germany

ABSTRACT

Segmenting images into different regions based on textures is a difficult task, which is usually approached using a combination of texture classification and image segmentation algorithms. The inherent variability of textured regions makes this a difficult modeling task. This paper show that 2D LSTM networks can solve the texture segmentation problem, combining both texture classification and spatial modeling within a single and trainable model. It directly outputs per-pixel texture classes and does not require a separate feature extraction step. We first introduce a new blob-mosaics texture segmentation dataset and its evaluation criteria, then evaluate our approach on the dataset and compare its performance with existing methods.

Index Terms— texture; texture segmentation; supervised segmentation; 2D LSTM Recurrent Networks; blob-mosaics; texture dataset; segmentation quality measurement

1. INTRODUCTION

Image segmentation is a fundamental task for many applications such as object recognition, medical imaging, and scene analysis. An uniform region is defined by homogeneous material or between discontinuities in depth. The texture has major visual cues of the surface within and between the regions. In order to segment the disjoint uniform regions based on textures, the combination of texture classification and image segmentation algorithm are used. However, it is difficult to generalize the system in order to find a pattern without the knowledge of domain since it is affected by various external conditions, i.e., wide range of scale, illumination, rotation, as well as internal noise of the texture.

Texture analysis for segmentation has been focused on various texture descriptors to characterize a spatial repetition. The common methods are concerned with texture modeling by filters to enhance textural properties [1, 2, 3, 4]. Rather simpler and very popular approaches that were traditionally used for texture analysis are Tamura's feature [5], Haralick features [6] on Gray level co-occurrence matrix [7] and Local Binary Pattern(LBP) [8]. The discriminative potential of these statistical metrics are high under limited conditions. The extensions to multi-direction and multi-scale however need complex parameter tuning. Another popular statistical model to represent local characteristics is Gaussian Markov Random Field (GMRF) [9, 10, 11]. It is more robust under noise data than other approaches, but the locality fails in covering diversity of textures since it only captures the dependency of neighbors of each pixels. The common issues of above approaches are the needs of hand-engineering process and complex modeling to cover the variations of a texture. In addition, it is hard to generalize the sub-window size with unknown scale.

More recently, Kandaswamy [12] compared both the well-established earlier texture algorithms with more recent advancements under extreme conditions. In addition to the various textural feature-based methods, texel and semantic texton forests were used to improve the performance in segmentation [13, 14]. Furthermore, machine learning techniques for segmentation [15, 16, 17, 18] were proposed instead of using manually selected filters. The multichannel filters are generalized in an unified system which combine feature representation and classification task. Particularly, Convolutional Neural Networks (CNNs) utilize the convolution mask as feature extractors instead of complex filter bank. Although, the manual designed preprocessing step is not required, a specific texture filter and multi-level classifier is necessary to obtain the final segmented image.

In this paper, the standard single-layered 2D LSTM networks [19, 20] solve texture segmentation problem by per-pixel texture classification. It does not require any manually designed preprocessing or feature extraction step and outputs segmented image without postprocessing. Preliminary success of pixel-wise multi-dimensional LSTM recurrent neural network approach such as image labeling [19] and offline handwriting recognition [21] indicate the ability to perform pixel-level classification. However, both tasks are either under the limited conditioned images or for the specific task. Our network model is based on the idea of Graves's [19], but with a much simpler architecture on much complex and real-world data. First, a new texture segmentation dataset using blob mixture of natural texture images is proposed due to the lack of diversity and difficulty of standard texture segmentation database. Thus, it more closely deals with the true image

segmentation problem. The network precisely estimates the texture regions and automatically adapt the different scale, orientation, and shapes of texture regions in the image. We show a simple and direct way of applying LSTM networks to the problem of texture segmentation and compare the performance using area-based segmentation quality measurement on our blob-mosaics dataset.

The rest of the paper is organized as follows. Section 2 describes 2D LSTM network, new texture segmentation dataset, and evaluation criteria for texture segmentation. Section 3 discusses experimental results, and concluding remarks are given in Section 4.

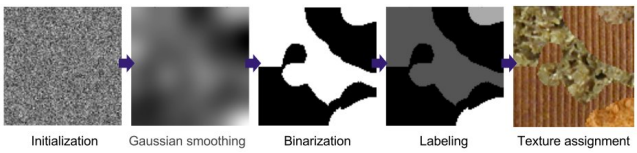
2. THE TEXTURE SEGMENTATION USING 2D LSTM NETWORKS

Network architecture: The network that we have applied for texture segmentation includes three layers: One input layer, one hidden layer, and one output layer. The networks receive a two-dimensional (2-D) array as input. The RGB value of pixels are fed directly to the hidden layer. The hidden layer consists of four recurrently connected units with LSTM subnets. In the subnet, memory cell is used as self-connection, and input, forget, and output gates control the storage of data. The recurrent connections access all directions of each pixel (left to right, right to left, top to bottom, and bottom to top; 2^n hidden unit for n-dimensional data), and accumulates the information. Hence, the global information of the pixel's all surroundings contributes to the final decision. Finally, the contextual information of all directions from the hidden layer is sent to a single output layer. The size of output layer corresponds to the number of classes and each output unit provides a probability of each texture class per pixel.

Network training: To find an optimal network model of the task, a different size of hidden unit is first trained on a constant size of input pixels and determined empirically. The size of input and output layer depend on the number of pixel and the texture class respectively. The network is trained using 4000 random 100×100 samples with the fixed learning rate and momentum. The optimal parameters for training in our task will be discussed in Section 3.

Blob-mosaics database for texture segmentation: There are a few public databases for texture segmentation [22, 23]. The existing texture segmentation database generated by synthetic compositions obtained from the collection of polygon mask has some limitations. Since the mask consists of constant voronoi polygons and is used for all images, the framework might learn this static shape of the region instead of the actual signature of textures. Moreover, the boundary of each texture region includes strong edges and/or corners which affect the performance of the segmentation.

To avoid these issues, we propose a new database using 2D Gaussian blobs. The image is composed of random 2D Gaussian blobs and each blob is filled with random material



(a) Blob-mosaics image generation

Fig. 1: Blob-mosaics texture segmentation database: The procedure of creating blob-mosaics image is as follow: 1) Gaussian filtering is applied on a random image (100×100 pixels with normal distributed value [0-1]). 2) Thresholding is performed for binarization (Median is selected as a threshold value). Randomly shaped regions are generated in this step. 3) Texture images from dataset KTH-TIPS2-a [24] is randomly assigned to the regions. Note that, the dataset KTH-TIPS2-a itself includes material textures with various conditions (different scales, illumination directions, and poses). Thus, the final blob-mosaics images include random shape, regions, as well as textures under various conditions.

textures from KTH-TIPS2-a [24]. To generate Blob-mosaics images, Gaussian filtering ($\sigma = 10.0$) is first applied on a random image (100×100 pixels with normal distributed value [0,1]). The random Gaussian image is then binarized by median value. In this step, randomly shaped blob regions are generated and ready for applying textures on the regions. Finally, randomly selected textures from KTH-TIPS2-a dataset are assigned to each region. The database, KTH-TIPS2-a, consists of texture images from eleven distinct materials under varying illumination, pose and scale. Thus, final blob-mosaics images include randomly shaped regions, as well as textures under various conditions, so it provides diverse challenges for texture segmentation. The 4000 images are generated for training and the trained network is tested on 630 images. The figure 1 (a) illustrates the flowchart of blob-mosaics image generation and examples of generated images are shown below.

Segmentation quality measurement: The pixel-based classification without any spatial information can especially result in imprecise or noisy segmented area. To measure and judge the robustness of these factors for the methods, segmentation accuracy was computed by area-based quality measurement proposed by Melendez [25]. Though the area-based accuracy is simply measured by the ratio between the predicted area and the area of corresponding ground-truth, there

is no direct way to map the predicted region onto the ground-truth. To find the best possible overlap between them, we first sort the predicted regions from large to small. It helps to avoid the double assignment of a region. The maximum overlapped region between the predicted and ground-truth image are matched and the overlapping ratio will be the area-quality of segmentation. Hence, the most probable similarities between the ground-truth and predicted image are found and the accuracy of the segmentation per image (Acc_I) is computed as follow:

$$Acc_I = \sum_{r=1}^{R} \frac{A_r^g \bigcap A_r^s}{A_r^g} \frac{A_r^g}{A_I} = \frac{1}{A_I} \sum_{r=1}^{R} A_r^g \bigcap A_r^s$$

where R is the number of region in the ground-truth, and A_I and A_r^g are the area of the whole image and label r in the ground-truth, respectively. A_r^s is the maximum portion of the correspoinding area for label r in the predicted image, and $A_r^g \bigcap A_r^s$ is the area of overlapping portion between A_r^g and A_r^s.

3. EXPERIMENTS

In order to validate the proposed model, the different approaches commonly used for segmentation are compared with our model; (1) patch-wise 6 Haralick features (gray) and Naive Bayesian classifier (Gray-Haralick+Naive Bayes) [26], (2) patch-wise 13 Haralick combined with color chroma features (color) and Naive Bayesian classifier (Color(HSV)-Haralick+Naive Bayes) [26, 27], and (3) the combination of Gaussian Mixture Model, ExpectationMaximization, and Hidden Markov Random Field (GMM-HMRF) [28]. All the experiments of LSTM networks have been run by using the RNNLIB library [29]

The first comparison method, 6 Haralick feature extraction (contrast, energy, homogeneity, correlation, dissimilarity, and angular second moment in four directions, 0°, 45°, 90°, and 135°) was performed on 9×9 patches and each pixel is classified by a Naive Bayesian classifier. The second one was 13 Haralick features (all except the 14th feature) considered with color (color-chroma) in HSV color space. For the last comparison approach, GMM-HMRF, the number of region (K = 3, 5) was initialized, and HMRF-EM was performed on RGB images with 20 EM iterations and 20 MAP iterations. For LSTM networks, a different size of hidden units (h = 10, 30, 50, 80) are tested. The input and output size are 3 (Red, green, and blue pixel) and 11 (the number of texture class). The learning rate of 1e-5 and a momentum of 0.9 are fixed for all of our experiments.

The best area-based segmentation quality (averaged over the 630 test samples) is compared in Table 1. The proposed technique, LSTM network, led to superior performance with the best K = 3 (for GMM-HMRF) and h = 30 (for LSTM networks). The performance with hidden size 30 and 50 are comparable (the difference was only about 0.4%). Segmentation results in Figure 2 show the effectiveness of our method. Particularly, as shown in Figure 3, most of other approaches have failed except LSTM networks under difficult blob-mosaics images.

Table 1: Accuracy comparison of texture segmentation on texture blob-mosaics images shown in Figure 1. 4000 random blob-mosaics images were used for training and the trained network was tested on 630 images. To compare the performance of segmentation, three different methods were selected —(1) patch-wise classification based method with gray texture features (Gray-Haralick+Naive Bayes), (2) patch-wise classification based method with gray and color texture features (Color(HSV)-Haralick+Naive Bayes), and (3) Gaussian mixture model+Expectationmaximization+Hidden Markov Random Field (GMM-HMRF). Haralick feature is one of the common texture features extracted from Grey level co-occurrence matrix (GLCM). The features were extracted on the patches (9×9), then sent it to the Naive Bayesian classifier. To consider the color information, color-chroma is considered with Haralick features. Another popular methods are ExpectationMaximization and Hidden Markov Random Field. The combination of Gaussian Mixture Model, ExpectationMaximization, and Hidden Markov Random Field (GMM-HMRF) is recently addressed in the literature [28]. (The best result on the table is with the initial region K=3). The accuracy is measured by region based quality measurement. The details of segmentation quality measurement was explained in Section 2. The proposed technique has obtained the highest average accuracy for texture segmentation (The best score is shown in bold).

method	avg. acc.(%)
Gray-Haralick+Naive Bayes [26]	43.87
Color(HSV)-Haralick+Naive Bayes [26, 27]	49.34
GMM-HMRF [28]	71.20
LSTM networks	**90.88**

4. CONCLUSION

In this paper, we presented an entirely learning based texture segmentation using 2D LSTM networks. The network has a strong discrimination power with a single, per-pixel trainable model that does not require a separate feature extraction step. The network itself takes care of the pixel and its surrounding neighborhood and does not require separate spatial modeling. Thus, it can automatically learn texture patterns for each pixel. The proposed technique has been compared with the common and popular segmentation algorithms and the results in terms of segmentation quality, robustness, and simpleness have been favorable. An important indication of this paper is that, 2D LSTM networks analyze texture information as well as its location efficiently with single, directly and automatically learnable model that are suitable for many different tasks. For example, it can be easily extended to the task of natural image segmentation if we know labels to supervise the network training. In the future, it is also highly interesting to adapt to a wide range of situations such as outdoor and/or indoor scene classification, segmentation, and object localization.

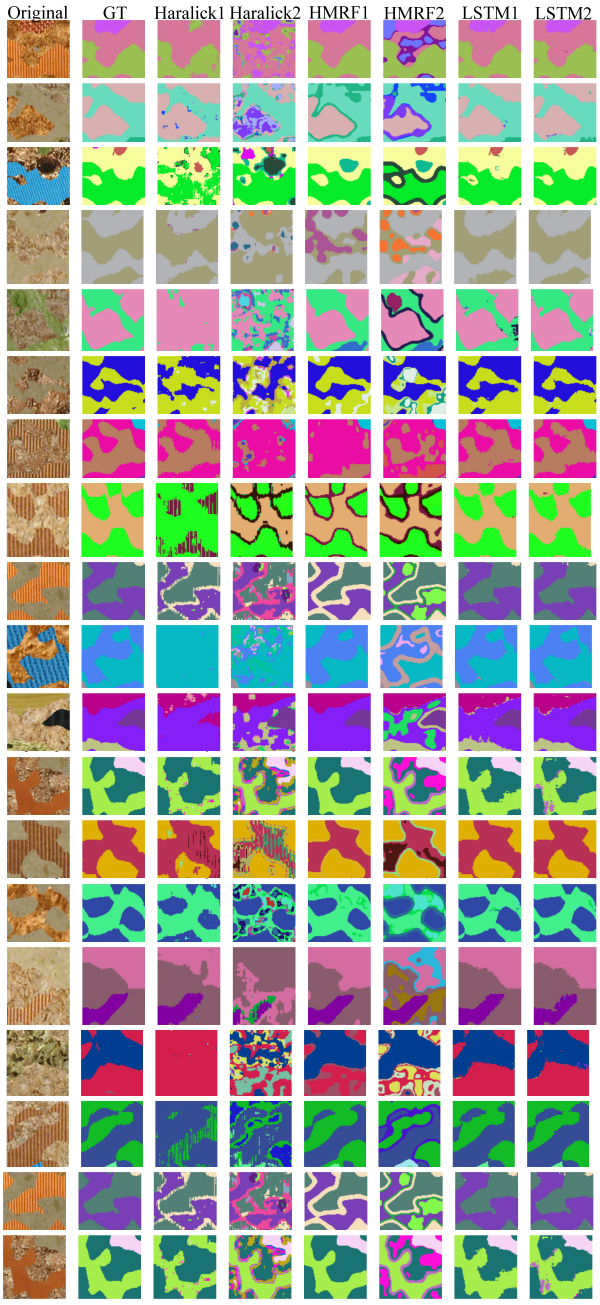


Fig. 2: Segmentation results by blob-mosaics images. From left to right column are original image (Original), ground truth (GT), Haralick gray features with Naive Bayesian classifier (Haralick1), Haralick color features with Naive Bayesian classifier (Haralick2), and HMM-HMRF (The initial region K = 3 (HMRF1) and 5 (HMRF2)), and LSTM networks (learning rate (lr) = 1e-5, hidden size (h) = 30 (LSTM1) and 50 (LSTM2)). The segmentation results show the superior performance of the proposed method.

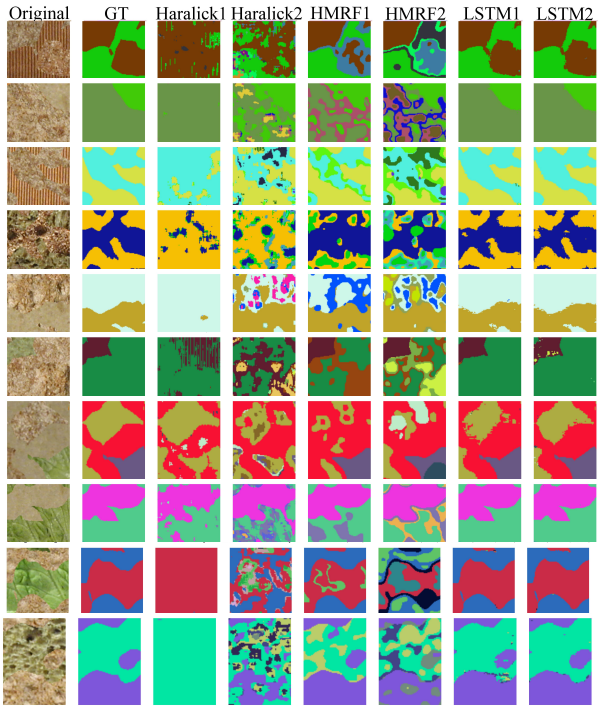


Fig. 3: Segmentation results under difficult blob-mosaics images. From left to right column are original image (Original), ground truth (GT), Haralick gray features with Naive Bayesian classifier (Haralick1), Haralick color features with Naive Bayesian classifier (Haralick2), and HMM-HMRF (The initial region K = 3 (HMRF1) and 5 (HMRF2)), and LSTM networks (learning rate (lr) = 1e-5, hidden size (h) = 30 (LSTM1) and 50 (LSTM2)). The most of other approaches result in the failed segmentation except LSTM networks.

5. REFERENCES

[1] David A Clausi and Huang Deng, "Design-based texture feature fusion using gabor filters and co-occurrence probabilities," *Image Processing, IEEE Transactions on*, vol. 14, no. 7, pp. 925–936, 2005.

[2] Olaf Pichler, Andreas Teuner, and Bedrich J Hosticka, "An unsupervised texture segmentation algorithm with feature space reduction and knowledge feedback," *Image Processing, IEEE Transactions on*, vol. 7, no. 1, pp. 53–61, 1998.

[3] Michael Unser, "Texture classification and segmentation using wavelet frames," *Image Processing, IEEE Transactions on*, vol. 4, no. 11, pp. 1549–1560, 1995.

[4] Hsi-Chia Hsin, "Texture segmentation using modulated wavelet transform," *Image Processing, IEEE Transactions on*, vol. 9, no. 7, pp. 1299–1302, 2000.

[5] Hideyuki Tamura, Shunji Mori, and Takashi Yamawaki, "Textural features corresponding to visual perception," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 8, no. 6, pp. 460–473, 1978.

[6] Robert M Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE*, vol. 67, no. 5, pp. 786–804, 1979.

[7] Robert M Haralick, Karthikeyan Shanmugam, and Its' Hak Dinstein, "Textural features for image classification," *Systems, Man and Cybernetics, IEEE Transactions on*, , no. 6, pp. 610–621, 1973.

[8] James C Bezdek, "Cluster validity with fuzzy sets," 1973.

[9] Philippe Andrey and Philippe Tarroux, "Unsupervised segmentation of markov random field modeled textured images using selectionist relaxation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 3, pp. 252–262, 1998.

[10] Santhana Krishnamachari and Rama Chellappa, "Multiresolution gauss-markov random field models for texture segmentation," *Image Processing, IEEE Transactions on*, vol. 6, no. 2, pp. 251–267, 1997.

[11] Giovanni Poggi, Giuseppe Scarpa, and Josiane B Zerubia, "Supervised segmentation of remote sensing images based on a tree-structured mrf model," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 43, no. 8, pp. 1901–1911, 2005.

[12] Umasankar Kandaswamy, Stephanie A Schuckers, and Donald Adjeroh, "Comparison of texture analysis schemes under nonideal conditions," *Image Processing, IEEE Transactions on*, vol. 20, no. 8, pp. 2260–2275, 2011.

[13] Sinisa Todorovic and Narendra Ahuja, "Texel-based texture segmentation," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 841–848.

[14] Jamie Shotton, Matthew Johnson, and Roberto Cipolla, "Semantic texton forests for image categorization and segmentation," in *Computer vision and pattern recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[15] Anil K. Jain and Kalle Karu, "Learning texture discrimination masks," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 18, no. 2, pp. 195–205, 1996.

[16] Kwang In Kim, Keechul Jung, Se Hyun Park, and Hang Joon Kim, "Support vector machines for texture classification," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 11, pp. 1542–1550, 2002.

[17] Fok Hing Chi Tivive and Abdesselam Bouzerdoum, "Texture classification using convolutional neural networks," in *TENCON 2006. 2006 IEEE Region 10 Conference*. IEEE, 2006, pp. 1–4.

[18] Jaime Melendez, Xavier Girones, and Domenec Puig, "Supervised texture segmentation through a multi-level pixel-based classifier based on specifically designed filters," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*. IEEE, 2011, pp. 2869–2872.

[19] Alex Graves, Santiago Fernández, and Jürgen Schmidhuber, "Multi-dimensional recurrent neural networks," in *Artificial Neural Networks–ICANN 2007*, pp. 549–558. Springer, 2007.

[20] Alex Graves, *Supervised sequence labelling with recurrent neural networks*, Ph.D. thesis, Technische Universitt Mnchen, 2008.

[21] Alex Graves and Juergen Schmidhuber, "Offline handwriting recognition with multidimensional recurrent neural networks," in *Advances in Neural Information Processing Systems*, 2008, pp. 545–552.

[22] Michal Haindl and Stanislav Mikes, "Texture segmentation benchmark," in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. IEEE, 2008, pp. 1–4.

[23] "The usc texture mosaic images," `http://sipi.usc.edu/database/database.php?volume=textures`.

[24] "KTH-TIPS and KTH-TIPS2 texture image database," `http://www.nada.kth.se/cvap/databases/kth-tips/download.html`.

[25] Jaime Melendez, Miguel Angel Garcia, Domenec Puig, and Maria Petrou, "Unsupervised texture-based image segmentation through pattern discovery," *Computer Vision and Image Understanding*, vol. 115, no. 8, pp. 1121–1133, 2011.

[26] Omar S Al-Kadi, "Supervised texture segmentation: a comparative study," in *Applied Electrical Engineering and Computing Technologies (AEECT), 2011 IEEE Jordan Conference on*. IEEE, 2011, pp. 1–5.

[27] Francesco Bianconi, Richard Harvey, Paul Southam, and Antonio Fernández, "Theoretical and experimental comparison of different approaches for color texture classification," *Journal of Electronic Imaging*, vol. 20, no. 4, pp. 043006–043006, 2011.

[28] Quan Wang, "GMM-based hidden markov random field for color image and 3d volume segmentation," *CoRR*, vol. abs/1212.4527, 2012.

[29] Alex Graves, "RNNLIB: A recurrent neural network library for sequence learning problems," `http://sourceforge.net/projects/rnnl/`.