

# LEARNING NON-HOMOGENOUS TEXTURES AND THE UNLEARNING PROBLEM WITH APPLICATION TO DRUSEN DETECTION IN RETINAL IMAGES

Noah Lee<sup>†</sup>, Andrew F. Laine<sup>†</sup>, and Theodore R. Smith<sup>‡</sup>

Columbia University, New York, NY, USA

<sup>†</sup>Department of Biomedical Engineering

<sup>‡</sup>Department of Ophthalmology

## ABSTRACT

*In this work we present a novel approach for learning non-homogenous textures without facing the unlearning problem. Our learning method mimics the human behavior of selective learning in the sense of fast memory renewal. We perform probabilistic boosting and structural similarity clustering for fast selective learning in a large knowledge domain acquired over different time steps. Applied to non-homogenous texture discrimination, our learning method is the first approach that deals with the unlearning problem applied to the task of drusen segmentation in retinal imagery, which itself is a challenging problem due to high variability of non-homogenous texture appearance. We present preliminary results.*

**Index Terms**— Selective Learning, Unlearning Problem, Probabilistic Boosting, Texture.

## 1. INTRODUCTION

In recent years a vast amount of models have been proposed for image segmentation and classification including geometric, non-parametric, and statistical models. In depth reviews can be found in [3-12]. What one can distil from the literature is that given a set of data samples ad-hoc model assumptions are hypothesized for model derivation and parameters are tweaked to optimize an error criteria in a rather heuristic manner. In segmentation and classification problems we often evaluate with respect to ground truth data obtained by expert knowledge. No quantitative information and theoretic foundation of the algorithm performance for future datasets are reported. In contrary, supervised learning algorithms optimizing for a maximum-margin decision provide theoretical bounds on the generalization performance [1]. However, the problem of unlearning in applied areas of learning algorithms has not been addressed in depth. Given a learning algorithm the usual procedure is to train a model on a given dataset and generalization performance is evaluated in a single dataset

scope. If new knowledge is available batch learning algorithms have to be retrained. To ideally adapt to the workflow of a certain application it is desirable to inject new a posteriori knowledge to the existing learning algorithm over time without the need to relearn the intelligence over the complete batch. In large knowledge domains this might be an expensive task. To give a specific example in the field of computer aided medical diagnosis a doctor might inject new knowledge into a learning algorithm in sparse time frames where an online learning over time would be highly desirable. In online learning we keep the intelligence obtained so far without complete batch retraining while introducing new knowledge into the learning algorithm [13]. We do not want to unlearn what we have learned so far.

In this paper we propose a novel approach for learning non-homogenous texture discrimination with the following characteristics: i) theoretical bounds on generalization performance of the learning algorithm, ii) preservation of intelligence after new injection of a posteriori domain knowledge, iii) no need for complete batch retraining, iv) possible knowledge contribution through collaborative filtering, and v) computational efficient selective subspace learning. We apply our approach on non-homogenous texture discrimination for drusen detection in retinal images. The unlearning problem is solved through a structured similarity clustering on over-complete Lemarie wavelet frames and a probabilistic boosting approach.

The remainder of the paper is organized as follows: in section 2 we outline prior work in the domain of online learning algorithms closely related to the spirit of our work. Section 3 describes our approach using structured similarity clustering and probabilistic selective boosting. Preliminary experimental results are shown in section 4. Finally we conclude with future work and open issues.

## 2. PRIOR WORK

Closely related to the spirit of preventing the unlearning problem in a learning framework are online learning algorithms and incremental learning techniques. Various

approaches exist in realizing an online learning algorithm. Phan et al. propose an online learning algorithm using an asymmetric boosted classifier [14]. Their method builds up on the work of [2] and balances the skewness of the training data for equal weighting to deal with knowledge asymmetry. Additional constraints on the weight distribution achieve faster convergence rate. Graber et al. propose an online boosting algorithm for feature selection and apply their method to various vision problems [15]. Here a fixed set of weak classifiers perform computation of the weight distribution. Their work is inspired by Oza et al. who proposed online learning algorithms for boosting and bagging for generating decision hypotheses [16]. In his work he provides proofs for comparable convergence rates between offline and online boosting for a large number of iterations.

Our approach differs in the following aspects. Rather than following a complete online learning approach we aim for the ability to selectively learn a subspace of our domain knowledge to mimic human learning of fast memory renewal targeted to a specific application task. We investigate the feasibility on the problem of non-homogenous texture classification with application to drusen segmentation in retinal images. Our method is able to keep the intelligence while adapting to sparse injections of new domain knowledge over time. Furthermore, our method is ideally suited to be integrated into collaborative filtering framework for fast distributed domain knowledge contribution.

### 3. METHODOLOGY

To outline our approach we begin with the issue of building a texture vocabulary for feature generation. Prior to the presentation of our selective learning algorithm we describe each key component of the algorithm in more detail.

#### Texture Vocabulary Construction

To construct a texture vocabulary we focus on regional texture descriptors that are robust and discriminative for non-homogenous texture appearance. In particular we examine the granularity of a texture by employing morphological scale space analysis [21] together with other texture measures [19, 20]. Consider the multi-scale top-hat transform

$$\begin{aligned} \Psi_{\theta}(f) &= f - \gamma_{\theta}(f), \text{ with} \\ \gamma_{B^{\theta}}^{\theta}(X) &= \bigcup_x \{B_x^{\theta} | B_x^{\theta} \subseteq X\}. \end{aligned} \quad (1)$$

Dependent on the local structure of the texture,  $B_x^{\theta}$  can be adapted to arbitrary shape formations. The transformation  $\gamma_{\theta}(f)$  denotes the morphological opening at different scale resolutions of  $\theta$ . In our case we use circular flat shaped

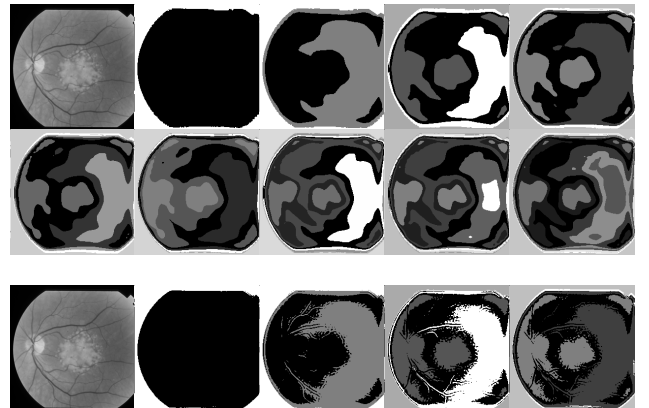
structuring elements for  $\theta_i$ , with  $i = 1 \dots k$ . We also incorporate other non-homogenous texture descriptors such as edge histogram features [19] and various other common texture measures [20]. We build a richer feature space by including additional interaction terms between the features. Feature selection for optimal generalization performance is then learned from our knowledge domain using boosting as described in [2]. Optimal features are selected minimizing the training error through a weighted combination of weak hypotheses with an upper bound of

$$\begin{aligned} \frac{1}{m} |\{i : H(x_i) \neq y_i\}| &\leq \prod_t Z_t, \text{ with} \\ \prod_t Z_t &= \frac{1}{m} \sum_i \exp(-y_i f(x_i)). \end{aligned} \quad (2)$$

This approach has the advantage that given our existing knowledge domain we compute a meaningful feature representation that are not based on heuristics with weak assumptions. Rather the obtained feature set comprises the best selection for our given learning task optimized for minimization of training and generalization error bounds through maximum-margin learning.

#### Structural Similarity Clustering

For meaningful texture information we make use of an adaptive multi-channel wavelet representation [18]. We start with a hierarchical approach to perform structural similarity clustering. By structural information we mean the texture information at multiple resolutions to cope with high variability in non-homogenous texture appearance. Texture regions are portioned into homogenous compositions through supervised multi-scale over-complete Lemarie wavelet frame learning. The clustering starts with a low frequency texture subspace to assign similarity assignments among classes and then gradually refine membership assignments by incorporating higher frequency subspaces during clustering.



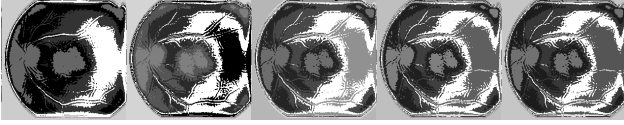


Fig. 1: Example of different wavelet subspaces sampled from over-complete Lemarie wavelet frames for texture based structural similarity clustering. Row I and II are obtained by using only the low frequency channels. Row III and IV show examples where only level 3 and 4 was discarded.

Once a representation is found for structural texture cues we perform unsupervised learning to separate the feature space into non-overlapping clusters to learn a partition function. Any standard unsupervised learning algorithm such as expectation maximization can be used for this purpose. Once we obtain a partition function of our knowledge domain we can assign the cluster membership for a new test sample by optimizing for the closest similarity match with highest probability. In this way we selectively learn a sub domain allowing more efficient learning without the need for complete batch relearning. A non-linear decision hypothesis is learned in this subspace using boosting.

### Probabilistic Selective Boosting

The first provable polynomial-time boosting algorithm was initially proposed by Schapire [1]. The underlying idea of boosting is to combine simple weak learners to form a learning ensemble of hypotheses with higher accuracy. If we have a set of weak learners  $h_1, h_2, \dots, h_T$  boosting builds a linear weighted ensemble hypothesis

$$f(x) = \sum_{t=1}^T \alpha_t h_t(x). \quad (3)$$

For details on how to compute  $f(x)$  we refer to [1]. Equation 3 aims to find an optimal weighted sum of  $h_t(x)$  by iteratively minimizing an exponential loss function

$$\sum_i \exp(-y_i f(x_i)) = \sum_i \exp\left(-y_i \sum_t \alpha_t h_t(x_i)\right). \quad (4)$$

The motivation for using a boosting approach is the strength of a confidence estimate in our prediction through bounds on generalization performance. Schapire showed that given the margin on example  $(x, y)$

$$\text{margin}_f(x, y) = \frac{y f(x)}{\sum_t |\alpha_t|} = \frac{y \sum_t \alpha_t h_t(x)}{\sum_t |\alpha_t|} \quad (5)$$

the generalization error is at most

$$\Pr[\text{margin}_f(x, y) \leq \theta] + O\left(\sqrt{\frac{d}{m\theta^2}}\right) \quad (6)$$

for any  $\theta > 0$  with high probability. In our case we compute a distribution over ensemble hypotheses  $H_i(x)$  to obtain a more robust learning estimate of the maximum-margin decision function. A summary of our learning approach is described below.

---

### Algorithm: Selective Learning to Prevent Unlearning

---

Given a knowledge domain  $X = (x_1, y_1)_t, \dots, (x_n, y_n)_t$  at time  $t$ , where  $y_i$  denotes expert knowledge and  $x_i$  image data

- Perform structural similarity clustering  $SSC$
  - Let the clustered knowledge domain  $X_C$  be  $(x_k, y_k)_t, \dots, (x_n, y_n)_t$  with  $k \in \{C_k | k = 1, \dots, n\}$
  - Find  $\{x_k, y_k\}_t \in C_k$ , s.t.  $\text{argmax}_k f(C_k, x_{n+1})$  with
$$f(C_k, x_{n+1}) = \text{similarity}(C_k, x_{n+1})$$
  - Learn a set of weak classifiers  $\{h_t\}$  using boosting  $\forall \{x_k, y_k\}_t$  and normalize hypothesis  $h(x) = \sum \alpha_t h_t$
  - Compute distribution over  $\hat{H}(x) = \frac{1}{n} \sum_{i=1}^n H_i(x)$
  - Apply  $\hat{H}(x)$  to  $x_{n+1}$  to get  $y_{n+1}$  and insert  $(x_{n+1}, y_{n+1})$  to  $X_C$
- 

## 4. EXPERIMENTS AND RESULTS

We evaluate our learning approach on a database of 179 images comprising normal cases and diseased cases with high phenotype variability in disease expression. We select 4 test samples for different texture types of drusen to evaluate generalization performance and mimic the application workflow it will be used in. Below is an example of an obtained cluster  $k$  showing examples of similar non-homogenous texture appearance.

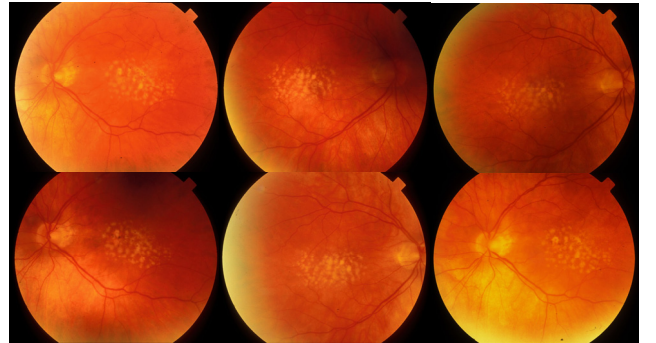


Fig. 2: Example set of similar texture types obtained from structured similarity clustering on over-complete Lemarie wavelet frames.

The table below shows detection results using our learning framework at four different time frames for two different texture complexities.

Table. 1: Example set of similar texture types obtained from structured similarity clustering on over-complete Lemarie wavelet frames.

	T=1	T=2	T=3	T=4
Texture I	0.867	0.813	0.922	0.896
Texture II	0.717	0.732	0.819	0.852

## 5. CONCLUSION

We have presented a learning approach for non-homogenous texture discrimination without facing the unlearning problem with application to drusen segmentation. Preliminary results seem promising. To our knowledge no such system was reported for non-homogenous texture discrimination. Future work is devoted to an in depth evaluation of our learning approach to other object targets as well as research towards a generalization of our learning algorithm to other application domains. Furthermore clinical validation studies will follow.

## 6. ACKNOWLEDGEMENTS

This work was founded by NEI (R01 EY015520-01), the NYC Community Trust (RTS), and unrestricted funds from Research to Prevent Blindness.

## 7. REFERENCES

[1] R. E. Schapire, "The Boosting Approach to Machine Learning – An Overview," MSRI Workshop on Nonlinear Estimation and Classification, 2002.

[2] P. Viola and M. Jones, "Robust Real-time Object Detection", 2<sup>nd</sup> International Workshop on Statistical and Computational Theories of Vision – Modeling, Learning, Computing, and Sampling, 2001.

[3] D. Cremers, M. Rousson, and R. Deriche, "A Review of Statistical Approaches to Level Set Segmentation: Integrating Color, Texture, Motion and Shape", IJCV Vol. 72, Issue 2, pp. 195-215, 2007.

[4] J.S. Suri, J.S., Kecheng Liu, S. Singh, S.N. Laxminarayan, Xiaolan Zeng and L. Reden, "Shape recovery algorithms using level sets in 2-D/3-D medical imagery: a state-of-the-art review", IEEE Transactions on Information Technology in Biomedicine, vol. 6, no. 1, pp. 8-28, 2002.

[5] N.R. Pal and S.K. Pal, "A review on image segmentation techniques," Pattern Recognition, vol. 26, no. 9, pp. 1277-1294, 1993.

[6] C. Xu, D. L. Pham, and J. L. Prince, "Medical Image Segmentation Using Deformable Models," Handbook of Medical Imaging – Vol. 2: Medical Image Processing and Analysis, pp. 129-174, 2000.

[7] M. Andreetto, L. Zelnik-Manor, and P. Perona, "Non-Parametric Probabilistic Image Segmentation", ICCV 2007

[8] T. F. Chan and J. Shen, "Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods", Society for Industrial and Applied Mathematics (SIAM) Philadelphia, 2005.

[9] Paragios N: Variational, geometric, and level set methods in computer vision: 3rd international workshop, VLSSM 2005, Beijing, China: proceedings. In Lecture notes in computer science, 3752, 2005.

[10] A. Baraldi and P. Blonda, "A survey of fuzzy clustering algorithms for pattern recognition", IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, Vol. 29, No. 6, pp. 778-801, 1999.

[11] J. Montagnat, H. Delingette and N. Ayache, "A review of deformable surfaces: topology, geometry and deformation", Image and Vision Computing, Vol. 19, No. 14, pp. 1023-1040, 2001.

[12] C. Cedras and M. Shah, "Motion based recognition: A survey", Image and Vision Computing, Vol. 13, No. 2, pp. 129-155, 1995.

[13] A. Blum, On-line Algorithms in Machine Learning, Lecture Notes In Computer Science; Vol. 1442, p. 306 – 325, 1998.

[14] M.-T. Phan and T.-J. Cham, "Online Learning Asymmetric Boosted Classifiers for Object Detection", Computer Vision and Pattern Recognition CVPR, pp. 1-8, 2007

[15] H. Grabner and H. Bischof, "On-line Boosting and Vision ", IEEE Computer Society Conference on Computer Vision and Pattern Recognition – Vol. 1, pp. 260-267, 2006.

[16] N. Oza and S. Russell, "Online Bagging and Boosting", In Proc. Artificial Intelligence and Statistics, pp. 105-112, 2001.

[17] B. Wu and R. Nevatia, "Improving Part based Object Detection by Unsupervised, Online Boosting", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2007.

[18] A. F. Laine and J. Fan, "An Adaptive Approach for Texture Segmentation by Multi-Channel Wavelet Frames", Technical Report TR-93-025, Center for Computer Vision and Visualization, Florida, 1993.

[19] C. S. Won, "Feature Extraction and Evaluation Using Edge Histogram Descriptor in MPEG-7", Advances in Multimedia Information Processing – PCM, Vol. 3333 pp. 583-590, 2004.

[20] P. Howarth, S. Rueger, "Evaluation of Texture Features for Content-Based Image Retrieval, Content-Based Image and Video Retrieval (II), Vol. 3115, pp. 326-334, 2004.

[21] P. T. Jackway, M. Deriche, "Scale-Space Properties of the Multiscale Morphological Dilation-Erosion", IEEE Transactions of Pattern Analysis and Machine Intelligence, Vol. 18, 1996.