Texture Classification with Fisher Kernel Extracted from the Continuous Models of RBM

Tayyaba Azim and Mahesan Niranjan
School of Electronics and Computer Science, University of Southampton, Southampton, U.K.

Keywords: Fisher Kernel, Factored 3-way RBM, Gaussian Bernoulli RBM, Texture Classification, Brodatz, Emphysema.

Abstract: In this paper, we introduce a novel technique of deriving Fisher kernels from the Gaussian Bernoulli restricted Boltzmann machine (GBRBM) and factored 3-way restricted Boltzmann machine (FRBM) to yield better texture classification results. GBRBM and FRBM, both, are stochastic probabilistic models that have already shown their suitability for modelling real valued continuous data, however, they are not efficient models for classification based on their likelihood performances (Jaakkola and Haussler, 1999; Azim and Niranjan, 2013). We induce discrimination in these models with the help of Fisher kernel that is constructed from the gradients of the parameters of the generative model. From the empirical results shown on two different texture data sets, i.e. Emphysema and Brodatz, we demonstrate how a useful texture classifier could be built from a very compact generative model that represents the data in the Fisher score space discriminately. The proposed discriminative technique allows us to achieve competitive classification performance on texture data sets, without expanding the size of the generative model with large number of hidden units. Also, comparative analysis shows that factored 3-way RBM is a good representative model of textures, giving rise to a Fisher score space that is less sparse and efficient for classification.

1 INTRODUCTION

Texture analysis and classification is one of the most widely explored research problems in computer vision. Since textures form an important feature of the objects in an image, and the perception of textures has played an important role in the human visual system for recognition and interpretation, there has been a great interest in developing artificial recognition systems that deploy texture based features for classification (Hangarge et al., 2013; Zhang et al., 2005). There are a number of different approaches: statistical, geometrical and model based, that have become state of the art techniques for texture classification. The statistical techniques compute the local features at each point in an image and derive a set of statistics from the distribution of local features, for example, the co-occurrence features (Haralick et al., 1973) or the gray level differences (Weszka et al., 1976). The geometrical methods assume that the building blocks of textures are textons that govern the spatial organization of the textures. These primitives are usually extracted by edge detection filters such as Laplacian-of-Gaussian or difference-of-Gaussian (Marr and Vaina, 1982; Poggio et al., 1988; Tuceryan and Jain, 1998), by adaptive region extraction (Tomita and Tsuji, 1990) or mathematical morphology (Serra, 1983; Matheron, 1967). After the primitives have been identified, some statistics of the primitives such as intensity and area are computed for analysis. Model-based methods hypothesize the underlying texture process, constructing a parametric generative model, which can create the observed intensity distribution of the image. Some of the examples of this method are pixel based models and region based models (Cristani et al., 2002).

In this work, we introduce a novel approach for texture classification that models the textures through a probabilistic statistical model and then uses the gradients of the model parameters as features for classification. The energy based probabilistic models used for capturing the pixel intensity variations present in the continuous images are Gaussian Bernoulli restricted Boltzmann machine (GBRBM) and factored 3-way restricted Boltzmann machine. Both the models have been successfully used previously to capture the pixel intensity variations in real valued images like natural scenes and textures (Kivinen and Williams, 2012; Ranzato et al., 2010; Cho et al., 2011). These variants of RBM provide a solution to the poor mod-

Azim T. and Niranjan M..
Texture Classification with Fisher Kernel Extracted from the Continuous Models of RBM.
DOI: 10.5220/0004857506840690
In *Proceedings of the 9th International Conference on Computer Vision Theory and Applications* (VISAPP-2014), pages 684-690
ISBN: 978-989-758-004-8

elling strength of the classical RBM for real valued data. The GBRBM, although useful, is much slower to train (Krizhevsky, 2009) and is not a good model of the covariance structure of an image because it does not capture the fact that the intensity of a pixel is almost exactly the average of its neighbours. Also it lacks a type of structure that has proven very effective in vision applications. These challenges have been addressed in factored 3-way RBM that uses the states of its hidden units to represent the abnormalities in the local covariance structure of an image.

To the best of our knowledge, Fisher kernel has not been extracted from the GBRBM and factored 3-way RBM before. In this work, we'll first show how this extraction is possible, and then discuss the advantages one may get for texture classification, if the probabilistic models used for Fisher kernel is factored 3-way RBM rather than a binary-binary RBM (BBRBM) or GBRBM. We observed that the texture classification accuracies are better if the underlying probabilistic model represents the data in gradient space discriminatively and with less sparsity. This manuscript is organized as follows: Section 2 describes the Fisher kernel framework, Sections 3 and 4 explain the generative models used for deriving the Fisher score space, Section 5 discusses the experimental design and the results obtained on benchmark texture data sets. We then conclude with a discussion on the obtained results and future work.

2 THE FISHER KERNEL

The Fisher kernel provides a generic framework for deriving a kernel from a generative probability model, $p(\mathbf{x}|\theta)$ by computing Fisher scores which are the gradients of the log likelihood of the data with respect to the model parameters, θ. Thus, the Fisher kernel function is mathematically expressed as:

$$K(\mathbf{x}_i,\mathbf{x}_j) = [\nabla_\theta \log p(\mathbf{x}_i|\theta)]^T \mathbf{U}^{-1} [\nabla_\theta \log p(\mathbf{x}_j|\theta)],$$
$$\text{where } \nabla_\theta \log p(\mathbf{x}|\theta) = \phi_\mathbf{x} \text{ and } \mathbf{U} = \mathrm{E}_\mathbf{x}[\phi_\mathbf{x}\phi_\mathbf{x}{}^T].$$

The magnitude of the Fisher score, ϕ_x specifies the extent to which each parameter contributes in generating the data. The Fisher scores obtained for each independent parameter are arranged in a vector of fixed dimension. The Fisher information matrix, U is the covariance matrix of the score vectors ϕ_x and is approximated as an identity matrix in this work to avoid computational complexity (Taylor and Cristianini, 2004). Once a kernel function is derived from a generative probability model, it could be embedded into any discriminative classifier such as support vector machines (SVM), linear discriminant analysis (LDA), etc. We have used SVM as a classifier to classify the images in benchmark data sets. The two probabilistic generative models from which the Fisher kernel has been derived in this work are Gaussian-Bernoulli restricted Boltzmann machine and factored 3-way restricted Boltzmann machine, both discussed below.

3 GAUSSIAN BERNOULLI RESTRICTED BOLTZMANN MACHINE (GBRBM)

An RBM is a bipartite graph in which the visible units that represent the observations are connected to binary stochastic hidden units using undirected weight connections. The hidden units allow the network to discover interesting features that represent complex regularities in the observations fed to the visible layer during training. The connectivity of the units is restricted with no visible-visible or hidden-hidden connections, thus allowing us to update all the units in the same layer in parallel. Moreover, biases are connected as an external input to each of the unit in the network. In a GBRBM, the visible units have a Gaussian activation function that modifies the energy of the RBM in the following way:

$$E(\mathbf{v},\mathbf{h}) = \sum_{i=1}^{V} \frac{(v_i - b_i^v)^2}{2\sigma_i^2} - \sum_{j=1}^{H} b_j^h h_j - \sum_{i=1}^{V}\sum_{j=1}^{H} \frac{v_i}{\sigma_i} h_j w_{ij},$$

where b_i^v and b_j^h are the biases attached to the visible, v and hidden units, h and w_{ij} refers to the weight interaction between the visible unit i and hidden unit j. Since, there is no direct connection of the units of each layer with each other, it is easy to infer samples via the following conditional distributions:

$$p(\mathbf{v}|\mathbf{h}) = \prod_{i=1}^{V} \mathcal{N}\left(b_i^v + \sum_{j=1}^{K} h_j W_{ij}, \sigma_i^2\right),$$
$$p(\mathbf{h}|\mathbf{v}) = \prod_{j=1}^{H} \text{sigmoid}\left(b_j^h + \sum_{i=1}^{V} W_{ij}\frac{v_i}{\sigma_i}\right),$$

where $\mathcal{N}(.,\sigma^2)$ denotes the probability density function of the Gaussian distribution with mean, μ and variance, σ^2 and sigmoid(x)=$\frac{1}{1+\exp(-x)}$. The gradient to update the model parameters are:

$$\frac{\partial L}{\partial W} = \left\langle \frac{1}{\sigma}\mathbf{v}\mathbf{h} \right\rangle_{data} - \left\langle \frac{1}{\sigma}\mathbf{v}\mathbf{h} \right\rangle_{model},$$
$$\frac{\partial L}{\partial \mathbf{b}^v} = \left\langle \frac{1}{\sigma^2}(\mathbf{v}-\mathbf{b}^v) \right\rangle_{data} - \left\langle \frac{1}{\sigma^2}(\mathbf{v}-\mathbf{b}^v) \right\rangle_{model},$$
$$\frac{\partial L}{\partial \mathbf{b}^h} = \langle \mathbf{h} \rangle_{data} - \langle \mathbf{h} \rangle_{model}.$$

The angle brackets represent the expectation under the probability distribution specified by the subscript. The first expectation under the data is calculable whereas the second expectation over the model distribution is intractable and can be approximated by drawing samples through a Markov chain Monte Carlo algorithm running for a very short time, i.e. 1 step, as proposed in Contrastive Divergence 1 (CD-1) algorithm (Hinton, 2002). GBRBM in general is known as difficult to train and this difficulty arises from learning standard deviations σ_i of the visible neurons. Unlike other parameters, the standard deviations are constrained to be positive. However, with an inappropriate learning rate, it is possible for the obtained gradient update rule to result in a non-positive standard deviation which may result in an infinite energy of the model (in case of σ_i=0) or to an ill-defined conditional distribution of the visible neuron (in case of σ_i=0). Since, all gradients other than that of the hidden biases are scaled by the standard deviation, inappropriate learning of it affects the learning of other parameters too. In this context, (Krizhevsky, 2009) suggested using a separate learning rate for the standard deviations which should be 100 to 1000 times smaller than that of the other parameters. However, there has been a general consensus to update the weights and the biases only, and use fixed, possibly unit standard deviation that results in impressive performances (Hinton and Salakhutdinov, 2009; Krizhevsky, 2009; Mohamed et al., 2010). For this reason, we have taken $\sigma = 1$ in our work. The Fisher score is derived from the log likelihood of the model as:

$$\nabla_\theta \log p(\mathbf{x}_n|\theta) = \nabla_\theta L = \left[\frac{\partial L}{\partial W} : \frac{\partial L}{\partial \mathbf{b}^v} : \frac{\partial L}{\partial \mathbf{b}^h}\right],$$
$$\text{where } \theta = \{W, \mathbf{b}^v, \mathbf{b}^h\}.$$

4 FACTORED 3-WAY RESTRICTED BOLTZMANN MACHINE (FRBM)

Ranzato et al. (Ranzato et al., 2010) proposed that an RBM's visible and hidden units can be modified to incorporate three-way interactions so that the covariance of the visible units is captured. This modified RBM which allows the hidden units to modulate pairwise interactions between the visible units is called *three-way RBM* shown in Figure 1.

Capturing the interactions between the visible units has far too many parameters, therefore, to keep their count under control and make learning efficient

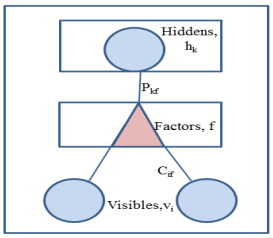


Figure 1: A graphical representation of the factored 3-way RBM in which the triangular symbol represents a factor that computes the projection of the input image whose pixels are denoted by v_i with a set of filters (columns of matrix C). The square outputs of the visible units are sent to the binary hidden units after projection with a second layer matrix (matrix P) that pools similar filters.

in practice, it is necessary to factorize these 3-way interactions. The energy function is redefined in terms of the three-way multiplicative interactions between the two visible binary units, v_i, v_j and one hidden binary unit, h_k as:

$$E(\mathbf{v},\mathbf{h}) = -\sum_{i,j,k} v_i v_j h_k W_{ijk}. \quad (1)$$

For real images, we expect the lateral interactions in the visible layer to have a lot of regular structure, therefore the three-way tensor can be approximated as a sum of factors:

$$W_{ijk} = \sum_f B_{if} C_{jf} P_{kf}, \quad (2)$$

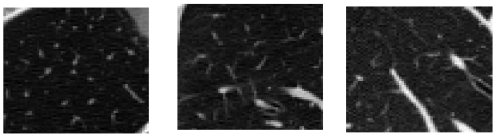
where i refers to the number of visible units, f refers to the number of factors and k refers to the number of hidden units. The matrix P is regarded as the *factor-hidden* or *pooling matrix* and the matrix C_{if} is known as *visible to factor matrix*. It is sensible to assume that matrix $B = C$ in Eq. 2 and the final approximation (Eq. 1) becomes :

$$E(\mathbf{v},\mathbf{h}) = -\sum_f \left(\sum_i v_i (C_{if})\right)^2 \left(\sum_k h_k P_{kf}\right).$$

The hidden units of the model are conditionally independent given the states of the visible units and their binary states are sampled as:

$$p(h_k = 1|\mathbf{v}) = \sigma\left(\sum_f P_{kf}\left(\sum_i v_i C_{if}\right)^2 + b_k\right),$$

where σ is a logistic function and b_k is the bias of the k-th hidden unit. In contrast, given the hidden states, the visible units are not independent and therefore it is much more difficult to compute the reconstruction of the data from the hidden states. To resolve this, in practice, the hidden units are integrated out by



(a) Normal Tissue (NT) (b) Centrilobular Emphysema (CLE) (c) Paraseptal Emphysema (PSE)

Figure 2: Examples of different lung tissue patterns extracted through computed tomography are shown. NT represents the sample of a healthy tissue, CLE reveals a healthy smokers tissue and PSE shows the distorted tissue of a person suffering from chronic obstructive pulmonary disease (COPD).

calculating *free energy* function of the model, and the visible samples are inferred using the Hybrid Monte Carlo (HMC) sampling technique (Neal, 1996) that calculates gradient of the free energy w.r.t the visible vector as:

$$F(\mathbf{v}) = -\sum_k \log\left(1+\exp\left(\frac{1}{2}\sum_f P_{kf}\left(\sum_i C_{if}v_i\right)^2 + b_k\right)\right) - \sum_i b_i v_i,$$

$$\frac{\partial F(\mathbf{v})}{\partial \mathbf{v}} = -\sum_f C_{if}\sum_k P_{kf}\frac{\sum_i C_{if}v_i}{1+\exp(-0.5\sum_f P_{kf}(\sum_i C_{if}v_i)^2 - b_k)},$$

The gradients of the log likelihood function of free energy w.r.t each model parameter are given as (Ranzato et al., 2010):

$$L1 = -\frac{1}{2}\left(\sum_i C_{if}v_i\right)^2 \frac{1}{1+\exp(-0.5\sum_f P_{kf}(\sum_i C_{if}v_i)^2 - b_k)},$$

$$L2 = -v_i\sum_k P_{kf}\frac{\sum_i C_{if}v_i}{1+\exp(-0.5\sum_f P_{kf}(\sum_i C_{if}v_i)^2 - b_k)},$$

$$L3 = -\frac{1}{2}\frac{1}{1+\exp(-0.5\sum_f P_{kf}(\sum_i C_{if}v_i)^2 - b_k)},$$

$$L4 = -\frac{\sum_i^n v_i}{n}, \text{ where}$$

$$L1 = \frac{\partial F(\mathbf{v})}{\partial P},\ L2 = \frac{\partial F(\mathbf{v})}{\partial C},\ L3 = \frac{\partial F(\mathbf{v})}{\partial \mathbf{b}^h} \text{ and } L4 = \frac{\partial F(\mathbf{v})}{\partial \mathbf{b}^v}.$$

The F0isher score is derived from the log likelihood of this model as:

$$\nabla_\theta \log p(\mathbf{x}_n|\theta) = \left[\frac{\partial F(\mathbf{v})}{\partial P} : \frac{\partial F(\mathbf{v})}{\partial C} : \frac{\partial F(\mathbf{v})}{\partial \mathbf{b}^h} : \frac{\partial F(\mathbf{v})}{\partial \mathbf{b}^v}\right],$$

$$\text{where } \theta = \{P, C, \mathbf{b}^v, \mathbf{b}^h\}.$$

5 EXPERIMENTS AND RESULTS

We have carried out the classification experiments on two different kinds of texture data sets: a medical image database called Emphysema, and a famous texture data set called Brodatz. The experimental design and the results obtained on the two data sets are described as follows:

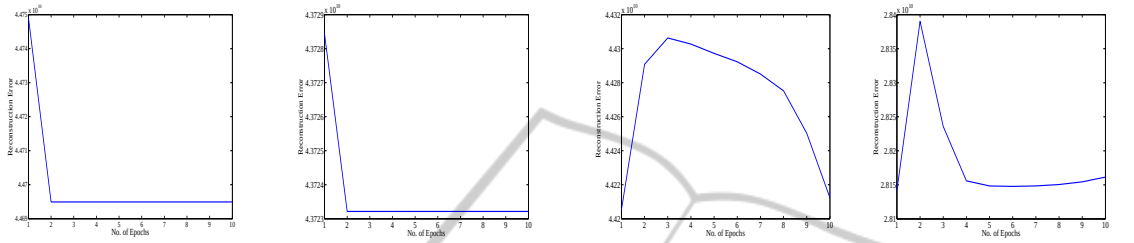
5.1 Emphysema Data Set

The Emphysema database (Sørensen et al., 2010) consists of 115 high-resolution computed tomography (CT) slices as well as 168 61×61 dimensional patches extracted from the subset of slices, and manually annotated for texture analysis techniques. Emphysema is a disease characterised by a loss of lung tissue and is one of the main reasons of chronic obstructive pulmonary disease (COPD). A proper classification of emphysematous - and healthy - lung tissue is useful for a more detailed analysis of the disease. The 61×61 pixel patches[1] are from three different classes: normal tissue (NT) with 59 observations, Centrilobular Emphysema (CLE) with 50 observations, and Paraseptal Emphysema (PSE) with 59 observations. The NT patches were annotated as never smokers, while the CLE and PSE region of interests were annotated as healthy smokers and smokers with COPD. These texture patterns serve as a good basis for assessing the modelling power of RBMs designed specifically for capturing pixel intensity variations present in the textures. As a preprocessing step, we crop 31×31 dimensional patch from the center of each 61×61 patch and threshold the pixel values in the dynamic range [-1000, 500]. The thresholding is based on the knowledge that the CT density values of lung parenchyma pixels are usually between the Hounsefield unit range [-1000HU, 500HU]. In order to classify these patches into 3 different classes, we have used Fisher kernel derived from three different probabilistic models: binary-binary RBM, Gaussian-Bernoulli RBM and factored 3-way RBM that model the data representations through different distributions. Once each of the generative model is trained, we calculate the gradients of the log likelihood function to form Fisher scores for the Fisher kernel. The Fisher kernel is then embedded into the SVM classifier that finally performs multi-class classification through one versus one training technique. The optimal value for hyperparameter C in SVM is decided via *grid search* method. In factored 3-way RBM, we maintained an average rejection rate of 6% with HMC sampling that used an adaptive step size to control the average acceptance rate of the drawn samples, thus yielding fast mixing rate. The summary of the classification results of Fisher kernel derived from different probabilistic models is shown in Table 1 (second col-

[1] http://image.diku.dk/emphysema_database/, http://www.ee.oulu.fi/research/imag/texture/image_data/Brodatz32.html

Table 1: Summary of classification results attained by different classifiers on the Emphysema and Brodatz texture data sets.

Classifier	Emphysema Performance (Acc)	Brodatz Performance (Acc)
k-Nearest Neighbour [Input=Image pixels, k=1]	$46.04 \pm 5.27\%$	$29.06 \pm 1.66\%$
Condensed Nearest Neighbour [Input=Image Pixels, 45% Data Retrieved | 25% Data Retrieved]	$46.06 \pm 5.19\%$	$28.11 \pm 2.01\%$
FK (Binary Binary RBM) [5 hid units]	$47.31 \pm 5.54\%$	$16.81 \pm 2.008\%$
FK (GaussianBinary RBM) [5 hidden units, $\sigma = 1$]	$47.85 \pm 4.83\%$	$16.96 \pm 2.40\%$
FK (Factored 3-Way RBM) [5 hid units, 32 factors]	$\mathbf{86.97 \pm 5.54\%}$	$\mathbf{65 \pm 4.6\%}$
k-Nearest Neighbour [Input=Local Binary Pattern features, k=1]	95.2%(Sørensen et al., 2010)	91.4%(**?**)



(a) RBM-BB(Emphysema) (b) RBM-GB(Emphysema) (c) Factored 3-Way RBM (Emphysema) (d) Factored 3-Way RBM (Brodatz)

Figure 3: The reconstruction error shown after training different variants of RBM generative model on the Emphysema and Brodatz data set for 10 epochs. The error for each of these models drops after several epochs; for factored 3-way model on Emphysema, it first rises, stabilises and then drops.

umn) and the reconstruction error for each model is shown in Figure 3. Note that the best known performance on Emphysema data set has been achieved by (Sørensen et al., 2010), in which he used the leave one subject out methodology to test the classifier. Such a partitioning scheme did not reveal discriminative Fisher score space in our case, due to which we chose holdout estimation method to train models and draw Fisher scores. Consequently, the Fisher kernel derived from factored 3-way RBM does give competitive classification performance in the same league as shown by (Sørensen et al., 2010).

5.2 Brodatz Texture Data Set

The Brodatz textures (Valkealahti and Oja, 1998) data set consists of a subset of 32 different classes chosen randomly from the main Brodatz data set. These textures are histogram equalized and then 20 patches[1] of size 64×64 are drawn from random locations of each class database for further experimentation. Table 1 shows the classification performance of distance based approaches, i.e. k-NN and condensed NN on these preprocessed patches. The same patches are also fed to the generative probability models for representational learning. Once the models (binary-binary, gaussian-binary and factored 3-way) are trained, a Fisher kernel is extracted from them and then embedded into the SVM classifier. The SVM classifies these textures using one versus one training of the gradients learnt by different models. The hyperparameter C in SVM is once again decided via

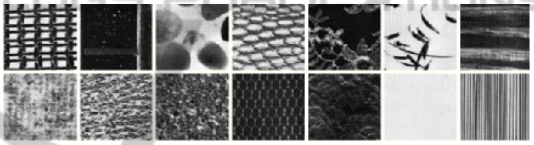


Figure 4: Samples of texture images from the Brodatz data set.

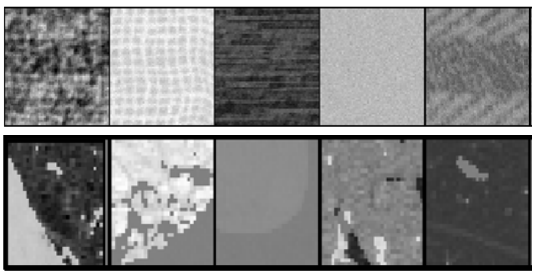


Figure 5: The visual factor filters, C_{if} learnt from the 64×64 size patches of Brodatz data set (1^{st} row) and 31×31 size patches of Emphysema data set (2^{nd} row).

grid search method. From the results obtained, we observe that the Fisher kernel derived from a factored 3-way RBM gives better classification performance in comparison to the other Fisher kernel based approaches and distance based classifiers on preprocessed images. The best performance on the data set is once again shown by local binary pattern features classified through k-NN and shown in Table 1.

6 DISCUSSION AND FUTURE WORK

In this paper, we present a novel approach of deriving a suitable classifier for texture classification that uses the gradients of the generative model to differentiate between different categories of textures. From the experiments conducted above, we observed that the performance of the Fisher kernel approach relies on the discriminative quality of the Fisher score space attained via maximum likelihood training of the generative models. On a comparative scale, the factored 3-way RBM proves better than the GBRBM and BBRBM since it was able to provide less sparse Fisher vectors that makes them suitable for discrimination in the dot product space. The dot product space is not suitable for learning distance metric similarities over sparse data, therefore Fisher vectors with zero or very small gradients donot provide a space discriminant enough for texture classification, as revealed for FK-BBRBM and FK-GBRBM in Table 1. It is also important to note that despite the availability of less sparse Fisher vectors, the Fisher kernel classification performance still does not beat the best known classification performance on the Brodatz data set. This follows us to the conclusion that a generative model which is trained well via maximum likelihood learning does not necessarily give rise to a representation that is well suited for classification tasks. In practice, the Fisher vectors for objects that have high probability under the model, will comprise of very small gradients that are less likely to form a discriminative basis for kernel functions. We would like to explore this in more detail by overcoming the gradient scaling problem through kernel normalization techniques in the future. Such a kernel should satisfy the rationale of achieving a discriminant Fisher score space by assigning similar gradients to two similar objects, and maintaining inter-class separability too. The impact of generative model's scale on the Fisher score space is also worth studying and will be pursued in future.

REFERENCES

Azim, T. and Niranjan, M. (2013). Inducing Discrimination in Biologically Inspired Models of Visual Scene Recognition. In *MLSP*.

Chen, J., Kellokumpu, V., and Pietikinen, G. Z. . M. (2013). RLBP: Robust Local Binary Pattern. In *BMVC 2013, Bristol, UK*.

Cho, K., Alexander, A., and R.Tapani (2011). Improved Learning of Gaussian-Bernoulli Restricted Boltzmann Machines. In *ANN - Volume Part I, ICANN, pages* 1017, Berlin, Heidelberg. Springer-Verlag.

Cristani, M., Bicego, M., and Murino, V. (2002). Integrated Region and Pixel Based Approach to Background Modelling. In *Workshop on MVC*, pages 3–8.

Hangarge, M., Santosh, K., Doddamani, S., and Pardeshi, R. (2013). Statistical Texture Features Based Handwritten and Printed Text Classification in South Indian Documents. *CoRR*, 1(32).

Haralick, R., Shanmugam, K., and Dinstein, I. (1973). Textural Features for Image Classification. *SMC*, 3(6):610–621.

Hinton, G. (2002). Training Products of Experts by Minimizing Contrastive Divergence. *Neural Computation*, 14:1771–1800.

Hinton, G. and Salakhutdinov, R. (2009). Semantic Hashing. *IJAR*, 50(7):969–978.

Jaakkola, T. and Haussler, D. (1999). Exploiting Generative Models in Discriminative Classifiers. *NIPS*, pages 487–493.

Kivinen, J. and Williams, C. (2012). Multiple Texture Boltzmann Machines. In *AISTATS*, volume 22, pages 638–646.

Krizhevsky, A. (2009). Learning Multiple Layers of Features from Tiny Images. Master's thesis.

Marr, D. and Vaina, L. (1982). Representation and Recognition of the Movements of Shapes. *Proceedings of the Royal Society of London.* Series B. Biological Sciences, 214(1197):501524.

Matheron, G. (1967). Representation and Recognition of the Movements of Shapes. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 214(1197):501–524.

Mohamed, A., Dahl, G., and Hinton, G. (2010). Deep Belief Networks for Phone Recognition. In *NIPS.*

Neal, R. (1996). *Bayesian Learning for Neural Networks*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Poggio, T., Voorhees, H., and Yuille, A. (1988). A Regularized Solution to Edge Detection. *Journal of Complexity,* 4(2):106 123

Ranzato, M., Krizhevsky, A., and Hinton, G. (2010). Factored 3-Way Restricted Boltzmann Machines for Modeling Natural Images. In *AISTATS*.

Serra, J. (1983). *Image Analysis and Mathematical Morphology*. Academic Press, Inc., Orlando, FL, USA.

Sørensen, L., Shaker, S., and de Bruijne, M. (2010). Quantitative Analysis of Pulmonary Emphysema Using Local Binary Patterns. *Medical Imaging*, 29(2):559–569.

Taylor, J. and Cristianini, N. (2004). *Kernel Methods for Pattern Analysis*. Cambridge University Press, New York, USA.

Tomita, F. and Tsuji, S. (1990). *Computer Analysis of Visual Textures*. Kluwer Academic Publishers, Norwell, MA, USA.

Tuceryan, M. and Jain, A. (1998). Handbook of Pattern Recognition & Computer Vision. *Chapter Texture Analysis*, pages 235276. River Edge, NJ, USA

Valkealahti, K. and Oja, E. (1998). Reduced Multidimensional Co-occurrence Histograms in Texture Classification. *PAMI*, 20(1):90–94.

Weszka, J., Dyer, C., and Rosenfeld, A. (1976). A Comparative Study of Texture Measures for Terrain Classification. *SMC*, 6(4):269–285.

Zhang, Y., X. Jian and J. Han (2005). Texture Feature-based Image Classification Using Wavelet Package Transform. In AIC, volume 3644 of LNCS, pages 165 173. Springer Berlin Heidelberg.