THE UNIVERSITY OF AUCKLAND
LIBRARIES AND LEARNING SERVICES

ResearchSpace@Auckland

Version
This is the Accepted Manuscript version. This version is defined in the NISO recommended practice RP-8-2008 http://www.niso.org/publications/rp/

Suggested Reference

Versteegen, R., Gimel'farb, G., & Riddle, P. J. (2013). Learning Generic Third-order MGRF Texture Models. In *Proceedings 2013 28th International Conference of Image and Vision Computing New Zealand (IVCNZ 2013)* (pp. 5-12). Wellington: IEEE. doi:10.1109/IVCNZ.2013.6726984

Copyright

Items in ResearchSpace are protected by copyright, with all rights reserved, unless otherwise indicated. Previously published items are made available in accordance with the copyright policy of the publisher.

© 2013 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

http://www.ieee.org/publications_standards/publications/rights/rights_policies.html

https://researchspace.auckland.ac.nz/docs/uoa-docs/rights.htm

Learning Generic Third-order MGRF Texture Models

Ralph Versteegen, Georgy Gimel'farb and Patricia Riddle
Department of Computer Science, The University of Auckland, Auckland, New Zealand
Email: rver017@aucklanduni.ac.nz, {g.gimelfarb, p.riddle}@auckland.ac.nz

***Abstract*—Descriptive abilities of translation-invariant Markov-Gibbs random fields (MGRF), common in texture modelling, are expected to increase if higher-order interactions, i.e. conditional dependencies between larger numbers of pixels, are taken into account. But the complexity of modelling grows as well, so that most of the recent high-order MGRFs are built to a large extent by hand. At the same time it is very difficult (if possible) to manually choose an efficient structure and strengths of pixel interactions for modelling a particular texture. This paper explores a possible extension of a computationally feasible framework for learning generic translation-invariant second-order MGRFs onto generic third-order models. Several information-theoretical heuristic methods for automatic learning of the latter are compared experimentally on a large and diverse database of realistic textures in application to a practically important problem of semi-supervised texture recognition and retrieval. However it is shown that better generative abilities of learnt models do not necessarily imply their higher discriminative power, and also the increased difficulty of learning third order order models may lead to worse generative performance.**

I. Introduction

Translation-invariant Markov random fields are popular for texture modelling and other computer vision tasks, under which texture classes are regarded as stochastic processes with Gibbs probability distributions (GPD); hence termed Markov-Gibbs random fields (MGRF). Given a graphical structure of local conditional dependencies (interactions) between pixels the GPD is factorised into potentials (factors), which are functions of signals on complete subgraphs (cliques) of mutually interacting pixels. So the conditional distribution of a pixel depends only on the characteristic neighbourhood set of interacting pixels: the union of all cliques it is a member of. The true interaction structures are highly texture-dependent (e.g. periodic and semi-periodic textures have long-range interactions between repeated units), so the structure of an MGRF texture model should be learnt; it is difficult to manually choose the interactions.

While in earlier work MGRFs with second order (pairwise) potentials were used nearly exclusively, recent research increasingly uses MGRFs with higher-order interactions, which are now widely recognised to be necessary for more expressive models of natural images and textures [1–5], as opposed to merely using larger neighbourhoods. Higher-order interactions provide context not available at lower orders; thus it is necessary to be able to infer higher-order interactions directly rather than hoping to recover them after finding the most important pairwise interactions (for example, three uniformly distributed binary random variables related by the parity function $a = b+c \mod 2$ are pairwise independent). But higher orders increase the difficulty of modelling as well, so that structures for most of the recent popular high-order MGRF models for image modelling, including those using responses of banks of linear fixed-support filters (e.g. [4–6]), and those using "local patterns" (e.g. [7]), were selected manually.

Outside of computer vision, inference of (non-spatially invariant) MGRF interaction structures has recieved considerable attention, though there has been very little application to image modelling. These two approaches are combined below to study third-order MGRF texture models with learnt translation-invariant structures, as a first step to higher orders. We consider computationally feasible methods based on information-theoretic heuristics for learning of generic third-order models, in straightforward extension from earlier methods [1, 8, 9].

This paper quantitatively compares different structure estimation procedures in application to semi-supervised texture discrimination where one texture class, represented by a training sample, is separated from other classes. This task is used here in an attempt to evaluate empirically the sensitivity of learning procedures to relevant visual features (statistics) of textures, hence performance is not compared to state-of-the-art texture discrimination algorithms, which usually focus on a few distinguishing features. Likewise, it is shown that the learnt third order models are considerably worse at discrimination than simpler second order models, despite improved texture synthesis results. Our focus is on fundamental aspects of learning, rather than practical problems such as slight non-homogeneity across an image (in contrast, rotation, scale, etc.).

A. Related work

Almost all practical algorithms for structure learning in MGRFs have been restricted to pairwise models, as the number of candidate factors grows exponentially in the order. For higher orders it seems necessary to use heuristics to guide the search and allow only a small amount of computation per clique family considered, excluding the majority of theoretically stringent selection criteria.

A common structure learning approach is to search in a space of candidate structures to maximise a score (usually penalised likelihood). To reduce computational costs, a pseudolikelihood-based score was used in a rare application of MGRF structure learning for texture modelling [10]; however its practicality at higher orders is unclear. Another practical score-based algorithm [11], introduced in a rather different context, iteratively creates more complex features (GPD factors) by compounding an atomic set of features. This is similar to many other iterative feature/potential selection procedures with the innovation of gradually building higher-order factors, an idea which may be applicable to MGRF texture modelling.

Another class of algorithms operate by conducting independence tests; an example due to Margaritis and Bromberg [12] for determining pairwise MGRF structures achieves robustness

to noise in the test results by maintaining a population of candidate structures, and efficiency by choosing tests which maximise the expected information gain. A promising approach by Abbeel et al. [13] considers whether to add each possible factor, of arbitrary order, independently of all others, and deduces parameters analytically. However, for each candidate factor the neighbourhood must first be inferred, currently infeasible if it may be large. But this approach might be useful for refining or pruning structures initially selected by less discerning heuristics, such as those below.

A further pragmatic class of methods (see e.g. [14] for a number of examples), which are similar to those in this paper and are widespread in inferring neural wiring and genetic expression interaction networks, use heuristics based on mutual information to identify, without formal guarantees, probable interacting sets of variables. Again these are almost always second-order, though Margolin et al. [3] inferred third-order interactions under the simplifying assumption that two variables in each triple were conditionally independent.

Recently filter-based MGRFs [5] which also learn the filters (with predetermined fixed supports), have been successfully applied to texture modelling by Heess et al. [2]. The interaction structure is learnt implicitly (via filter coefficients), but because of learning difficulties and to keep computational feasibility the filters have been restricted to small sizes not capturing distant interactions, e.g. 7×7 in [2].

Selection of second-order clique families for MGRF texture models using model-based interaction maps [8] (see Section II-B) is a very simple and cheap approach appropriate for many classes of texture. Extending this to prevent selection of redundant clique families by estimating the secondary interactions induced by compounding of multiple families [1, 9] allows much finer characteristics of complex textures to be recovered, however degrades performance on other textures. The method in [1] is computationally cheap; this approach, with several modifications, is used here for second-order clique family selection and partially extended to the third order.

II. Selecting characteristic interactions

A. Notation

Let $\mathcal{R} \subset \mathbb{Z}^2$ and $\boldsymbol{\alpha} = \{\alpha_i : \alpha_i \in \mathbb{Z}^2; 1 \leq i \leq d\}$ be a finite arithmetic lattice with coordinates r of nodes (pixels) and a list of d coordinate offsets with $\alpha_1 = (0,0)$ fixed, respectively. An order d clique family $C_{\boldsymbol{\alpha}}$ in $\mathcal{R}$ is the set of all spatially repeated cliques, or configurations of mutually interacting pixels in $\mathcal{R}$, with the offset pattern given by $\boldsymbol{\alpha}$:

$$C_{\boldsymbol{\alpha}} := \{(r_1, \ldots, r_d) : \\ r_1, \ldots, r_d \in \mathcal{R}; r_i - r_1 = \alpha_i; i = 1, \ldots, d\}$$

Let $\mathbf{G} : \mathcal{R} \to \{0, \ldots, Q-1\}$ be an image on $\mathcal{R}$ with Q possible grey levels. Let $\mathbf{G}^{(\boldsymbol{\alpha})}$ denote the empirical distibution (normalised d-th dimensional histogram) of grey levels of $\mathbf{G}$ over the cliques of $C_{\boldsymbol{\alpha}}$.

A translation invariant generic MGRF model P of grey-scale images with multiple pixel interactions is specified by a Gibbs probability distribution over images on $\mathcal{R}$ given by

$$P(\mathbf{G}|\Lambda) = \frac{1}{Z(\Lambda)} \exp\left(-\sum_{\boldsymbol{\alpha} \in \mathbf{A}} \mathbf{V}_{\boldsymbol{\alpha}} \bullet \mathbf{G}^{(\boldsymbol{\alpha})} |C_{\boldsymbol{\alpha}}|\right) \tag{1}$$

where $\bullet$ denotes the dot product (momentarily treating the operands as vectors) and $\Lambda = (\mathbf{A}, \mathbf{V})$ denotes the model parameters: the interaction structure $\mathbf{A}$ of clique families and corresponding potentials $\mathbf{V} = \{\mathbf{V}_{\boldsymbol{\alpha}} \in \mathbb{R}^{Q^{|\boldsymbol{\alpha}|}} : \boldsymbol{\alpha} \in \mathbf{A}\}$. $Z(\Lambda)$ is a normalisation constant, and the negation of the expression inside the exponential is the energy $E_\Lambda(\mathbf{G})$ of $\mathbf{G}$. The order of the model is defined as the maximum order of any of its clique families. Given $\mathbf{A}$ and a training image $\mathbf{G}_{\text{obs}}$, the potentials of an MGRF P are to be estimated (e.g. using stochastic approximation) such that the expected values of $\mathbf{G}^{(\boldsymbol{\alpha})}$ under P are equal to $\mathbf{G}_{\text{obs}}^{(\boldsymbol{\alpha})}$ (the sufficient statistics). This is both the maximum entropy distribution given the $\mathbf{G}_{\text{obs}}^{(\boldsymbol{\alpha})}$ [15] and the Maximum Likelihood Estimate (MLE) of the $\mathbf{V}_{\boldsymbol{\alpha}}$ given $\mathbf{G}_{\text{obs}}$.

The entropy of a probability distribution $p : \Omega \to \mathbb{R}$ is defined (in the discrete case) as $H(p) := -\sum_{x \in \Omega} p(x) \log p(x)$. The mutual information I of two random variables measures the average amount of information gained about one variable by learning the value of the other. There are a number of different generalisations of mutual information. The simplest, the total correlation (also known as the multi-information or simply the mutual information) of a set of n random variables with marginal distributions p_i $(1 \leq i \leq n)$ and joint distribution p is defined as $I(p) = I(p_1, \ldots, p_n) := D_{\text{KL}}(p \parallel p_1 \ldots p_n) = \sum H(p_i) - H(p)$. Here, $D_{\text{KL}}(p \parallel q) := \sum_{x \in \Omega} p(x) \log \frac{p(x)}{q(x)}$ is the Kullback-Leibler divergence (KLD), which naturally measures the distance of an approximating probability distribution from the true distribution. However, the KLD is not an ideal measure of agreement if both p and q are empirical probability distributions: it is asymmetric, and $D_{\text{KL}}(p \parallel q)$ is only defined if $q(x) = 0 \implies p(x) = 0$, which is likely to be violated. The Jensen–Shannon divergence (JSD) defined as

$$D_{\text{JS}}(p \parallel q) := \frac{1}{2}\left(D_{\text{KL}}\left(p \parallel \frac{p+q}{2}\right) + D_{\text{KL}}\left(q \parallel \frac{p+q}{2}\right)\right)$$

is then more appropriate. The JSD measures the discriminability of two distributions, specifically, the average certainty with which a sample can be ascribed to one or the other.

B. Pairwise clique selection

Theoretically, the KLD between the true distribution and the maximum entropy distribution of an MGRF structure can be employed as a loss function for choosing $\mathbf{A}$. The priority is to select the most relevant (characteristic) families; redundant clique families may actually be helpful rather than harmful [1] but increase computation costs. Rather than penalising complexity, caps on the number of families were used. Our third order models are built by first selecting second order potentials to capture pairwise statistics of $\mathbf{G}_{\text{obs}}$, as considered here.

A model-based interaction map (MBIM) [8] assigns a score in $\mathbb{R}$ to each clique family in the width w search window of all order d offset patterns with bounded maximum differences δ_x, δ_y in any of the x, y offset coordinates; the set of unique candidate offset patterns used was

$$W_w^d := \{(\alpha_1 = (0,0), \alpha_2, \ldots, \alpha_d) : \\ \alpha_1, \ldots, \alpha_d \in \mathbb{Z}^2;\ \alpha_{i-1} < \alpha_i;\ i = 2, \ldots, d; \\ \delta_x(\{\alpha_i\}) \leq w;\ \delta_y(\{\alpha_i\}) \leq w\}$$

where the pairs are compared in the lexicographical ordering.

Originally the score function of the family $C_{\boldsymbol{\alpha}}$ used was the partial energy $\epsilon_{\boldsymbol{\alpha}}$: some estimate of the family's contribution to E_Λ if were it included in an MGRF, such as using the approximation $\mathbf{V}^*$ from [16] to the MLE of $\mathbf{V}$ for the MGRF with $\mathbf{A} = \{\boldsymbol{\alpha}\}$. The partial energies attempt to gauge the strength of interaction of a clique family $C_{\boldsymbol{\alpha}}$. These interactions might be quantified more directly with the mutual information (MI); $I(\mathbf{G}_{\text{obs}}^{(\boldsymbol{\alpha})})$. For some textures this better distinguishes weak interactions with low $\epsilon_{\boldsymbol{\alpha}}$ scores from background noise, while for many others there is no significant difference at all.

Given an MBIM, the simplest selection method is to take the families with highest scores. But as mentioned before, filtering out secondary interactions is necessary to capture weaker but important interactions, which may be done by sequential selection of cliques while attempting to account for compound effects of the already selected cliques by estimating them analytically [8] or by sampling [9]. Below, a variant of the faster analytic option with information-theoretic measures instead of partial energies is used. At step $n \geq 0$ of the algorithm with n already selected families F_n, for candidate family $C_{\boldsymbol{\alpha}}$ with $\boldsymbol{\alpha} = \{(0,0), a\}$ the expected marginal $f_n^{(\boldsymbol{\alpha})}$ of $C_{\boldsymbol{\alpha}}$ is approximated and the next clique family selected is $\arg\max_{\boldsymbol{\alpha} \in W_w^2} D_{\text{JS}}(\mathbf{G}_{\text{obs}}^{(\boldsymbol{\alpha})} \parallel f_n^{(\boldsymbol{\alpha})})$.[1] By finding sequences of offsets $(o_1, \ldots, o_k), o_i \in F_n$, which sum to a and treating these cliques as a Markov chain, the expected marginal due to this chain is easily computed from the pairwise marginals. If there is no such sequence, $f_n^{(\boldsymbol{\alpha})}$ is the product of its marginals. The influences of different chains ought to be combined somehow (e.g. taking the geometric mean), but currently only the path of length at most 4 minimising the divergence is found.

C. 3rd order family selection

One would like to pick higher order families in the same way as above; however the quick growth of the search space prevents this from being practical. W_w^3 is a four-dimensional space, meaning that any local maxima are normally surrounded by tens to thousands of high scoring small pertubations; the large number of candidates should be reduced before considering more expensive selection criteria such as distribution divergences. The simple solution is to prevent families nearby (as points in Z^{2d}) to already selected families from being selected (Section III-C describes details). This roughly selects only local maxima or families near the largest maxima. Sequential, score-based selection similar to the 2nd-order method may then be used: at step n, with n 3rd order clique families G_n already selected, choose the family $\arg\max_{\boldsymbol{\alpha} \in W_w^3 \wedge \boldsymbol{\alpha} \notin X_n} S(\boldsymbol{\alpha})$ where S is a score function of 3rd order families and X_n is the set of all families within some fixed distance of a member of G_n (we used an L_2 distance of 3). Three score functions were tried: the simplest, $S_{\text{TC}} = I(\mathbf{G}_{\text{obs}}^{(\boldsymbol{\alpha})})$ using total correlation, as well as the L_1 distance, $S_1 = ||\mathbf{G}_{\text{obs}}^{(\boldsymbol{\alpha})} - g_n^{(\boldsymbol{\alpha})}||_1$, and the JSD, $S_{\text{JS}} = D_{\text{JS}}(\mathbf{G}_{\text{obs}}^{(\boldsymbol{\alpha})} \parallel g_n^{(\boldsymbol{\alpha})})$, between actual $\mathbf{G}_{\text{obs}}^{(\boldsymbol{\alpha})}$ and estimated $g_n^{(\boldsymbol{\alpha})}$ marginals. It may be practical to estimate the $g_n^{(\boldsymbol{\alpha})}$ through compositions of 2nd- and 3rd-order cliques, but the much simpler alternative used was to learn a 2nd-order MGRF model after selecting 2nd-order families using the sequential algorithm, sampling from that to approximate $g_0^{(\boldsymbol{\alpha})}$, and further assuming $g_n^{(\boldsymbol{\alpha})} \approx g_0^{(\boldsymbol{\alpha})}$.

[1] This means that instead of I as the score function as above an approximation $D_{\text{JS}}(p_{1,2} \parallel p_1 p_2)$ to $I(p_1, p_2) = D_{\text{KL}}(p_{1,2} \parallel p_1 p_2)$ is actually used. In practice these MBIMs are very nearly identical after non-linear scaling.

III. Experiments

A. Experimental setup

Experiments in this paper were conducted with a set of 136 grey-scale digitised photographs of natural and approximately spatially homogeneous textures sourced from several popular databases. MeasTex[2] is a framework of standardised testcases and procedures for evaluating texture discriminators, which includes a database ("NewTex") of natural textures, and also specifies a suitable subset of the MIT VisTex[3] database in its testcases. We used the 58 VisTex and 34 NewTex textures which were part of at least one MeasTex testset, and selected 44 textures from the Brodatz album [17].

Rather than create contrast invariant models using e.g. grey level difference or ordinal histograms, each image was simply preprocessed with the contrast-limited adaptive histogram equalization (CLAHE) [18] using the implementation available in `scikit-image`, with 16×16 tiles and a contrast clipping limit of 0.03, and then quantised to $Q = 8$ grey levels. Without CLAHE, misclassification rates were much lower due to the ease of distinguishing a training image from other images merely by their first-order histograms. After quantisation each image was split into 128×128 pieces and one of the centre blocks selected as the training piece. Modelling was not assisted by scaling the images to shorten interaction lengths.

Each clique family selection procedure investigated was executed given only the training piece of each texture as $\mathbf{G}_{\text{obs}}$. An MGRF $P(\mathbf{G}|\Lambda)$ was then obtained with the MLE potentials learnt by starting with $\mathbf{V}^*$ from [16], and its straightforward extension to third-order potentials, and fine tuning using stochastic approximation with Gibbs sampling from the MGRFs. The set of energies $e_i = \{E_\Lambda(\mathbf{G}_i) : \mathbf{G}_i \in \mathcal{T}\}$ of all the pieces $\mathcal{T}$ of the training texture is then used to define a simple classification rule: $\mathbf{G}$ is classed as belonging to the texture if $\min_i e_i \leq E_\Lambda(\mathbf{G}) \leq \max_i e_i$. The misclassification (false positive) rate of a model against all the pieces of every other texture image in the database was computed, and the whole procedure repeated to produce and evaluate one MGRF per texture four times. The mean misclassification rate across models and the range of the mean over the four runs are stated.

Although not totally realistic for practical texture classification this testing methodology is very challenging as it punishes overfitting (inability to cope with small differences across the texture image such as a slight twisting) by requiring a 100% recognition rate for the training texture, and requires generalisation with no negative examples provided.

B. Second order models

It was found empirically that across different textures, background noise in the MI-based 2nd order MBIM is up to about 0.015 bits (for a 128×128 image). So the termination rule used for sequential selection was either once a clique

[2] http://www.texturesynthesis.com/meastex/meastex.html

[3] http://vismod.media.mit.edu/vismod/imagery/VisionTexture/

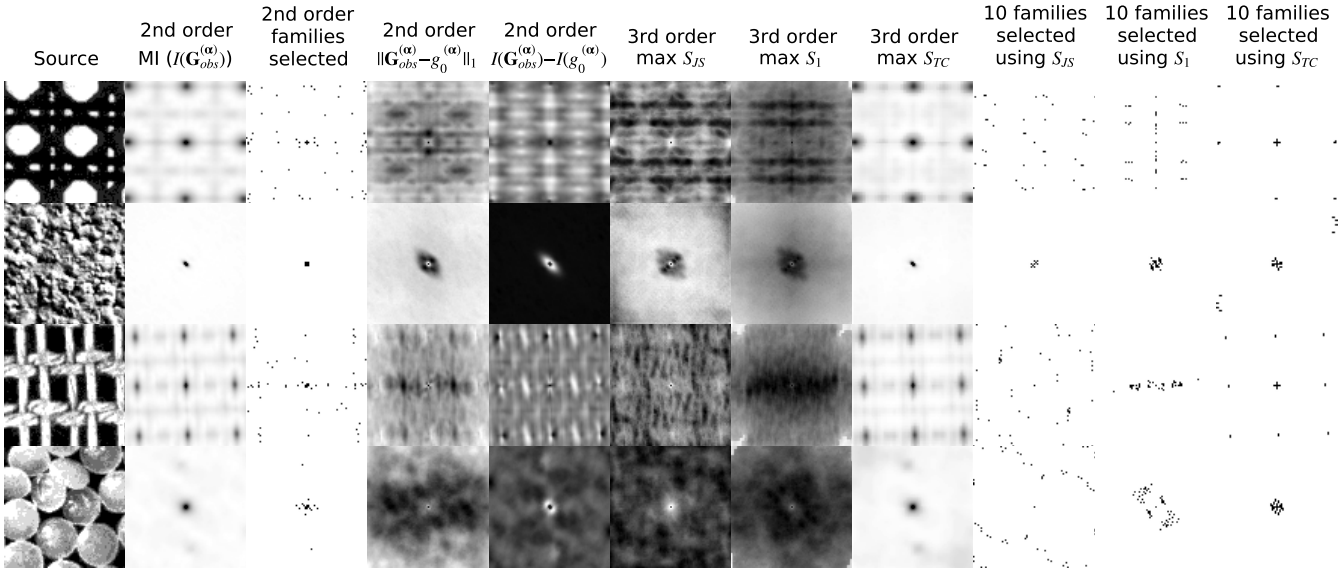

Fig. 1. Various second- and third-order MBIMs for a number of textures. All images are shown are 81×81 (w = 40); larger values are darker, using a linear scale. First column shows part of the training image used, after CLAHE preprocessing. Second column is the mutual information MBIM. The third column shows the 2nd order families selected sequentially (each appears as two dots). 2nd order MGRF models with these structures are then learnt and sampled from to estimate expected 2nd and 3rd order marginal distributions $g_0^{(\boldsymbol{\alpha})}$. Two other second order MBIMs follow, visualising errors in $g_0^{(\boldsymbol{\alpha})}$. Third order MBIMs and offset sets are visualised by flattening: each clique is drawn three times with each choice of point to place at the origin; so each appears as six points. For the MBIMs the maximum value plotted at each point is shown. The last three columns show ten 3rd order families selected using the three score functions to the left.

family with an MI below 0.015 bits is to be selected, or when the cap (set at 40) is reached. The sequential selection algorithm was compared (with a window size of $w = 40$) against several simpler baseline algorithms: firstly, a fixed number (30) of the families with the highest MI; secondly a simple threshold on the MI of $\mu + k * \sigma$ as in [8] with a cap of 40 clique families; $k = 4$ was appropriate. This selected on average 32 and at least 6 offsets. Discrimination results are presented in Table I. Although the sequential selection does not do better on average than the others, it selected on average only ten families, so had a third the computational cost.

C. Feasible selection of higher-order families

For third order families a width 40 window was again used, giving $|W^3_{40}| = 4,032,760$ candidates. The search for high scoring clique families can be sped up by first iterating over the search space with a step size of 2 pixels in each dimension (or similarly by scaling the training image and sampled images to 50%), reducing the space to $1/16$th the size, and then searching the vicinities of the highest–score solutions found.

However, this approach does not scale well to higher orders except by using increasingly larger step sizes, as the number of families grows as $O(w^{2(d-1)}/d!)$ where w is the window size (divided by the step size) and d is the order. Already at 4th-order and a step size of 2 about an hour of computation would be required, and any step size larger than 3 is likely to miss important families. Therefore, a far more scalable approach was investigated, of guiding the search by estimating the higher-order score functions with functions of scores of sub-cliques. For the clique family defined by $\boldsymbol{\alpha} = ((0,0), \alpha_2, \alpha_3)$ with sub-families $Y := \{((0,0), \alpha_2), ((0,0), \alpha_3), ((0,0), \alpha_3 - \alpha_2)\}$ two estimates/indicators for $S_X(\boldsymbol{\alpha})$ were used: the sum $S^{\Sigma}_X = \sum_{\boldsymbol{\beta} \in Y} S_X(\boldsymbol{\beta})$ and the maximum $S^{\max}_X = \max_{\boldsymbol{\beta} \in Y} S_X(\boldsymbol{\beta})$. Ideally the clique families with the highest scores would be (large) subsets of those with the highest indicators. Figure 2 shows that this is the case for S^{Σ}_{TC} and S^{Σ}_1, which provide excellent heuristics, but that S^{Σ}_{JS} is not as good a guide for S_{JS}. A better estimate of S_{JS} will therefore likely be needed. Other plots, not shown here, also show that the maximums are strongly correlated with the 3rd-order score for all three score functions, so a (weaker) threshold on this could also be used as an additional criterion for earlier pruning.

However recall that a collection of variables may be pairwise independent yet have non-zero total correlation, although for the small collection of textures used to create Figure 2 this effect did not appear. Thus S^{Σ}_{TC} can not be truely as reliable as indicated by the figure, though it may in practice for naturally occurring textures still be an effective indicator.

TABLE I. DISCRIMINATION WITH SECOND-ORDER MODELS.

Total average misclassification rate (range of averages across four runs): 30 families with highest I (*a*), 4σ thresholding of I (*b*), and sequential selection (*c*)		
(*a*) 13.4 (12.7-14.5)%	(*b*) 13.1 (12.8-13.4)%	(*c*) 13.0 (12.4-13.4)%

TABLE II. DISCRIMINATION WITH THIRD-ORDER MODELS USING k_3 THIRD-ORDER CLIQUE FAMILIES.

Total average misclassification rate (range of averages across four runs): Selection of the third-order cliques using: S_{TC} (*a*), S_1 (*b*), and S_{JS} (*c*)			
k_3	(*a*)	(*b*)	(*c*)
10	14.0 (14.0-14.1)%	16.4 (15.8-17.4)%	17.5 (17.3-17.6)%
20	14.8 (14.8-14.9)%	17.9 (17.6-18.5)%	19.5 (19.0-20.1)%
With 3rd order potentials split			
10	14.6 (14.4-15.0)%	13.7 (13.0-14.0)%	15.9 (15.4-16.3)%
20	16.2 (15.9-16.5)%	13.2 (12.8-13.4)%	18.5 (17.9-19.3)%

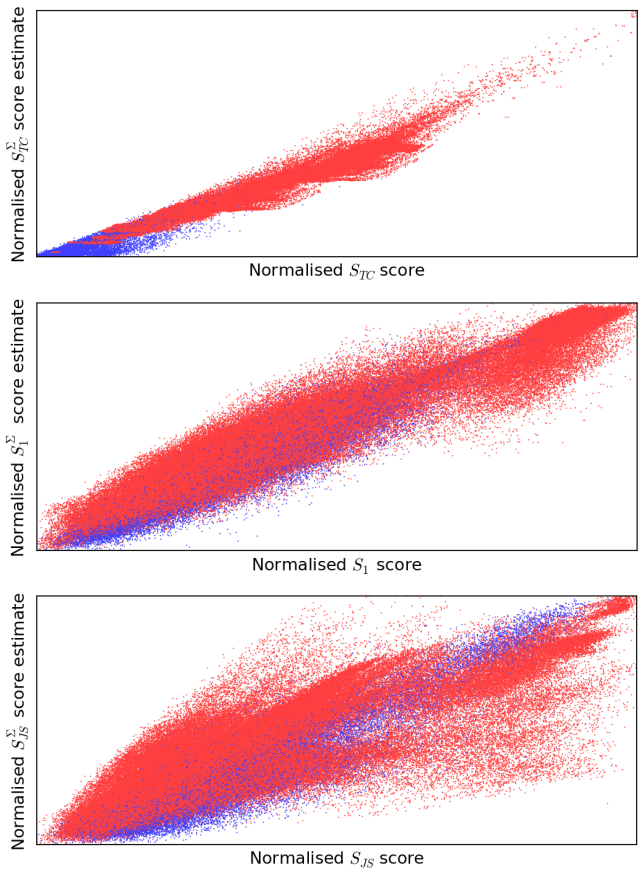

Fig. 2. Plots of estimates S^{Σ}_{TC}, S^{Σ}_{1}, S^{Σ}_{JS} against S_{TC}, S_1, S_{JS} for clique families chosen from MBIMs of ten different textures either randomly (blue dots) or using the step size 2 approach to find a superset of the highest scores (red dots). The true and estimated scores for each texture have been normalised to the same scale.

D. Third order models

Figure 1 visualises 2nd- and 3rd-order MBIMs computed using different scoring functions, and the families selected as a result. It can be seen that all scores pick out quite different offsets (see Figure 4 for more examples); actually S_{JS} prefers offsets on the edge of the MBIM, S_1 ones close to the origin, and S_{TC} chiefly only those where all sub-cliques have high MI. These tendencies suggest all the scores have undesired biases.

If third order clique families with a low I are allowed to be selected dramatic over-fitting can occur; Figure 3 shows an example of an energy map (showing the contribution to $E_{\Lambda}(\mathbf{G})$ due to each pixel) of a model with such families, with a complete failure to generalise due to learning random fluctuations in $\mathbf{G}^{(\boldsymbol{\alpha})}$. Based on empirical experiments we used a stopping condition like that for second order clique selection, requiring that $I(\mathbf{G}^{(\boldsymbol{\alpha})}) \geq 0.25$ bits. Before this cutoff is reached usually thousands of families qualify, so we compared performance with caps of 10 or 20 3rd-order families.

Table II presents results, which should be compared to the 13.0% misclassification rate before the 3rd order families are selected as a baseline. We also compare performance with models learnt with the same structures except with each 3rd-order clique family replaced with ('split' into) the three corresponding 2nd-order sub-clique families. Not only do all

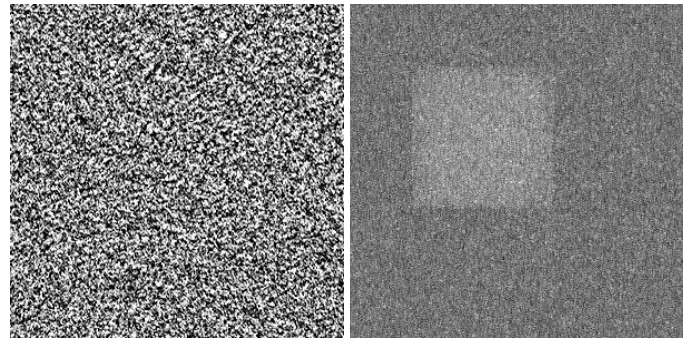

Fig. 3. Part of VisTex image Fabric.0018 and its energy map for a model with 20 poorly chosen 3rd order clique families. The visible square is the 128×128 piece of the image from which the model was learnt.

the models have higher than baseline misclassification rates despite typically much better generative abilities (see the next section), but the 'split models' perform better in the case of S_1 and S_{JS}. For S_{TC} the large overlap with existing 2nd-order families means that little is added over the baseline by the split cliques, so it seems likely that the split S_{TC} models show worse performance because of the removal of third order statistics.

E. Texture synthesis

Figure 4 shows samples of texture synthesis using models with up to 20 3rd-order clique families, as well as split models for a direct comparison of the effect of including third order statistics. Models using third order families selected with S_1 or S_{JS} nearly always outperform those selected with S_{TC}, which often do worse than the baseline, in reversal of the discrimination results. Splitting the 3rd order potentials may worsen (e.g. D66, D33) or have little effect (e.g. D101, D103) on the synthesis results, or actually show a significant improvement (e.g. D20). Selecting and then splitting third order potentials seems to find pairwise interactions which are useful yet would not be selected directly (e.g. D103, D20, D66). The apparent biases in the offsets preferred by the score functions seem to result in one outperforming another depending on the behaviour suiting each texture, rather than one being generally superior.

IV. Discussion and Conclusion

This paper explored a computationally feasible approach to learning generic third-order MGRFs for texture modelling. However, contrary to the common expectation, the third-order models constructed did not evenly outperform their second-order counterparts. It seems that dilution of the most significant statistical differences between textures with additional potentials harms discrimination in our case where it increases the range of e_i (see Section III-A). Also, for most of the evaluated third-order models the ability to discriminate between textures decreased on average even when compared to the 'split' models differing only in ability to capture third order statistics, indicating that the larger number of parameters per potential made parameter learning or generalisation more difficult, e.g. caused overfitting due to noise. However, incorporating third-order statistics is unnecessary for other textures. That the adequate order of an MGRF model depends on classes of textures involved and related computational problems to be solved

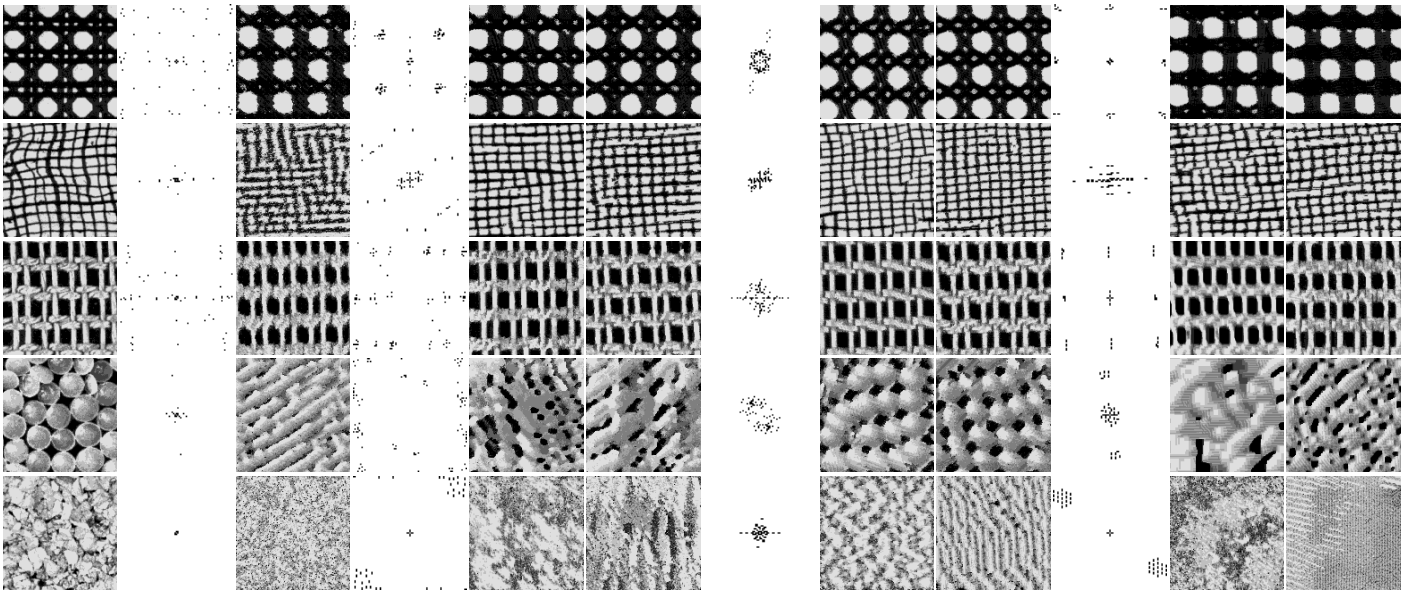

Fig. 4. Texture synthesis results. From left: training texture samples (of D101, D103, D20, D66, D33); selected 2nd order clique families and results; and for each of S_{JS}, S_1, S_{TC}: selected 3rd order cliques families (left), result (centre), and result with 2nd order models after breaking the triples into pairs (right).

is trivial common knowledge. Therefore, feasible learning of challenging higher-order MGRF models is of theoretical and practical interest irrespective to their successes or failures in applications to individual problems.

Future work will need to overcome first the problems which have been encountered with third-order models. More efficient non-parametric and/or parametric estimates of distributions (e.g. [19]) are likely to be helpful to represent higher-order Gibbs potentials instead of the vector-based representation used here. In this way their robustness to noise can be improved. Further options for scoring functions and changes to the iterative procedure used should be investigated, too.

Once these problems are adequately solved, our goal is to extend the feasible model identification framework, presented in this paper, onto fourth- and higher-order MGRF models. At present, scoring based on L_1 histogram distances seems to have the best results and ease of estimation. Combining the heuristic factor selection approaches evaluated in this paper as a pruning step before the use of otherwise impractical structure selection frameworks, such as in [13], is a particularly promising direction to be explored.

References

[1] G. Gimel'farb, "Estimation of an interaction structure in Gibbs image modelling," in *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, vol. 3, 2000, pp. 506–509 vol.3.

[2] N. Heess, C. K. I. Williams, and G. E. Hinton, "Learning generative texture models with extended Fields-of-Experts," in *British Machine Vision Conference, BMVC 2009, London, UK, September 7-10, 2009. Proceedings*, 2009.

[3] A. A. Margolin, K. Wang, A. Califano, and I. Nemenman, "Multivariate dependence and genetic networks inference," *Systems Biology, IET*, vol. 4, no. 6, pp. 428–440, 2010.

[4] S. C. Zhu, Y. N. Wu, and D. Mumford, "Minimax entropy principle and its application to texture modeling," *Neural computation*, vol. 9, no. 8, pp. 1627–1660, 1997.

[5] S. Roth and M. J. Black, "Fields of experts," *International Journal of Computer Vision*, vol. 82, no. 2, pp. 205–229, 2009.

[6] Y. W. Teh, M. Welling, S. Osindero, and G. E. Hinton, "Energy-based models for sparse overcomplete representations," *J. Mach. Learn. Res.*, vol. 4, pp. 1235–1260, Dec. 2003.

[7] T. Ojala, K. Valkealahti, E. Oja, and M. Pietikäinen, "Texture discrimination with multidimensional distributions of signed gray-level differences," *Pattern Recognition*, vol. 34, no. 3, pp. 727 – 739, 2001.

[8] G. Gimel'farb, *Image textures and Gibbs random fields*, ser. Computational Imaging And Vision. Dordrecht: Kluwer Academic Publishers, 1999, vol. 16.

[9] A. Zalesny, "Analysis and synthesis of textures with pairwise signal interactions," Katholieke Universiteit Leuven, Tech. Rep. Project 50542 (EC ACTS 074 - Vanguard, 1.3.1998-1.3.1999), 1999.

[10] C. Ji and L. Seymour, "A consistent model selection procedure for Markov random fields based on penalized pseudolikelihood," *The Annals of Applied Probability*, vol. 6, no. 2, pp. 423–443, 1996.

[11] S. Della Pietra, V. Della Pietra, and J. Lafferty, "Inducing features of random fields," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 4, pp. 380–393, 1997.

[12] D. Margaritis and F. Bromberg, "Efficient Markov network discovery using particle filters," *Computational Intelligence*, vol. 25, no. 4, pp. 367–394, 2009.

[13] P. Abbeel, D. Koller, and A. Y. Ng, "Learning factor graphs in polynomial time and sample complexity," *J. Mach. Learn. Res.*, vol. 7, pp. 1743–1788, Dec. 2006.

[14] P. Meyer, K. Kontos, F. Lafitte, and G. Bontempi, "Information-theoretic inference of large transcriptional regulatory networks," *EURASIP Journal on Bioinformatics and Systems Biology*, vol. 2007, no. 1, p. 79879, 2007.

[15] E. T. Jaynes, "Information theory and statistical mechanics," *Phys. Rev.*, vol. 106, pp. 620–630, May 1957.

[16] G. Gimelfarb and D. Zhou, "Accurate identification of a Markov-Gibbs model for texture synthesis by bunch sampling," in *Computer Analysis of Images and Patterns*, ser. Lecture Notes in Computer Science, W. G. Kropatsch, M. Kampel, and A. Hanbury, Eds. Springer Berlin Heidelberg, 2007, vol. 4673, pp. 979–986.

[17] P. Brodatz, *Textures: a Photographic Album for Artists and Designers*. New York, NY, USA: Dover Publications, 1966.

[18] K. Zuiderveld, "Contrast limited adaptive histograph equalization," in *Graphic Gems IV*. Academic Press Professional, 1994, pp. 474 – 485.

[19] D. W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*. New York, NY, USA: J. Wiley & Sons, 1992.