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 received 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 discriminating 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 induced by compounding of multiple families [1, 2]. 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 $I(p) := I(p_{1}, \ldots, p_{n}) := \sum_{x \in \Omega} p(x) \log \frac{p(x)}{\prod_{i=1}^{n} p_{i}(x)}$. Here $D_{KL}(p || q) := \sum_{x \in \Omega} p(x) \log \frac{p(x)}{q(x)}$ 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 $D_{KL}(p || q)$ is only defined if $q(x) = 0 \implies p(x) = 0$, which is likely to be violated. The Jensen–Shannon divergence (JSD) defined as

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

is then more appropriate. The JSD measures the discriminability of two distributions, specifically, the average certainty with which a sample can be assigned 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 $A$. 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 $G_{\text{obs}}$, 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 $\delta_x, \delta_y$ in any of the $x, y$ offset coordinates; the set of unique candidate offset patterns used was

$$W^{d}_{w} := \{(\alpha_{1} = (0, 0), \alpha_{2}, \ldots, \alpha_{d}) : \alpha_{1}, \ldots, \alpha_{d} \in \mathbb{Z}^{2} ; \alpha_{i-1} < \alpha_{i} ; i = 2, \ldots, d ; \delta_x(\{\alpha_{i}\}) \leq w ; \delta_y(\{\alpha_{i}\}) \leq w \}$$

where the pairs are compared in the lexicographical ordering.
Originally the score function of the family $C_\alpha$ used was the partial energy $e_\alpha$; some estimate of the family’s contribution to $E_\Lambda$ if we included it in an MGRF, such as using the approximation $V^*$ from [16] to the MLE of $V$ for the MGRF with $\Lambda = \{\alpha\}$. The partial energies attempt to gauge the strength of interaction of a clique family $C_\alpha$. These interactions might be quantified more directly with the mutual information (MI); $I(G_{\text{obs}})$. For some textures this better distinguishes weak interactions with low $e_\alpha$ 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 $C_\alpha$ with $\alpha = \{(0,0), \alpha\}$ the expected marginal $f_n(\alpha)$ of $C_\alpha$ is approximated and the next clique family selected is $\text{arg max}_\alpha \in \mathbb{W}_2 \ D_{JS}(G_{\text{obs}} \parallel f_n(\alpha))$. By finding sequences of offsets $(o_1, \ldots, o_k)$, $o_i \in F_n$, which sum to $\alpha$ 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, $f_n(\alpha)$ 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. $W_3^n$ is a four-dimensional space, meaning that any local maxima are normally surrounded by tens to thousands of high scoring small perturbations; 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 $Z^{2d}$) from 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 $\text{arg max}_{\alpha \subseteq X_n} S(\alpha)$ 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, $S_{TC} = I(G_{\text{obs}}) \lambda$ using total correlation, as well as the $L_1$ distance, $S_1 = \|G_{\text{obs}}(\alpha) - g_n(\alpha)\|_1$, and the JSD, $S_{JS} = D_{JS}(G_{\text{obs}}(\alpha) \parallel g_n(\alpha))$, between actual $G_{\text{obs}}$ and estimated $g_n(\alpha)$ marginals. It may be practical to estimate the $g_n(\alpha)$ 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 $g_n(\alpha)$, and further assuming $g_n(\alpha) \approx g_0(\alpha)$.

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. MeaTex[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 MeaTex 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 $16 \times 16$ 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 $128 \times 128$ 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 $G_{\text{obs}}$. An MGRF $P(G|\Lambda)$ was then obtained with the MLE potentials learnt by starting with $V^*$ 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 $\epsilon_i = \{E_\Lambda(G_i) : G_i \in \mathcal{T}\}$ of all the pieces $\mathcal{T}$ of the training texture is then used to define a simple classification rule: $G$ is classed as belonging to the texture if $\min_i \epsilon_i \leq E_\Lambda(G) \leq \max_i \epsilon_i$. 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 $128 \times 128$ image). So the termination rule used for sequential selection was either once a clique

---

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

2http://wwwtexturesynthesis.com/meaTex/meaTex.html

3http://vismod.media.mit.edu/vismod/imragery/VisionTexture/
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 \( |W_{40}| = 4,032,760 \) 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 1/16th 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 \( O(w^{2d-1}/d!) \) 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 \( \alpha = ((0, 0), \alpha_2, \alpha_3) \) with sub-families \( Y = \{((0, 0), \alpha_2), ((0, 0), \alpha_3), ((0, 0), \alpha_3 - \alpha_2)\} \) two estimates/indicators for \( S_X(\alpha) \) were used: the sum \( S_X^\Sigma = \sum_{\beta \in Y} S_X(\beta) \) and the maximum \( S_X^{\max} = \max_{\beta \in Y} S_X(\beta) \). 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 \( S_X^{\Sigma} \) and \( S_X^{\max} \), which provide excellent heuristics, but that \( S_X^{\Sigma} \) is not as good a guide for \( S_X^{\max} \). A better estimate of \( S_X^{\Sigma} \) 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 \( S_X^{\Sigma} \) can not be truly as reliable as indicated by the figure, though it may in practice for naturally occurring textures still be an effective indicator.
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 $E_A(G)$ due to each pixel) of a model with such families, with a complete failure to generalise due to learning random fluctuations in $G^{(\alpha)}$. Based on empirical experiments we used a stopping condition like that for second order clique selection, requiring that $I(G^{(\alpha)}) \geq 0.25$ 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 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.
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 $L_2$ 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


