Strong Markov Random Field Model

Rupert Paget, *Member, IEEE*

Abstract

The strong Markov random field (strong-MRF) model is a sub-model of the more general MRF-Gibbs model. The strong-MRF model defines a system whose field is Markovian with respect to a defined neighborhood and all sub-neighborhoods are also Markovian. A checkerboard pattern is a perfect example of a strong Markovian system. Although the strong Markovian system requires a more stringent assumption about the field, it does have some very nice mathematical properties. One mathematical property, is the ability to define the strong-MRF model with respect to its marginal distributions over the cliques. Also a direct equivalence to the Analysis-of-variance (ANOVA) log-linear construction can be proved. From this proof, the general ANOVA log-linear construction formula is acquired.

Index Terms

Markov processes, Contingency table analysis, Nonparametric statistics, Texture, Model development

I. Introduction

A MARKOVIAN system is most soundly modeled as a Gibbs distribution [1], [2]. If the Markovian system can not be modeled by an equivalent Gibbs distribution, then the MRF will not have a properly defined likelihood distribution [1]. However, in order to obtain the correct Gibbs distribution for a particular MRF, the neighborhood system needs to be known, and the "potential functions" for each "clique" are required. Apart from a simple binary auto-model [3], there is no exact solution to these parameters. In which case, a maximum likelihood estimation (MLE) is required [4], [5].

The process to estimate the maximum likelihood is to first choose a parameterized auto-model [4], and then perform updates on those parameters with respect to the MRF that the

Manuscript received June 24, 2001; revised October 11, 2002.

R. Paget is with the Computer Vision Group, Swiss Federal Institute of Technology (ETH), Zürich, Switzerland. E-mail: rpaget@vision.ee.ethz.ch .

parameterized model generates. The model is slowly refined until the marginal distributions of the generated field is comparable with the equivalent distributions of the training field. An alternative approach, as presented in this paper, is to model the MRF explicitly as a combination of the marginal distributions. The advantage of such an approach is that the texture can be quickly modeled from a training field as it bypasses the need for MLE. Another advantage is that the MRF model is not constrained by the functionality of arbitrary predefined potential functions.

In order to bypass the MLE and model the MRF as a combination of the marginal distributions, one major constraint is required of the MRF; the MRF needs to be of a subclass of MRFs known as "strong-MRFs" as defined by Moussouris [6]. In this subclass, the conditions on the MRF are a lot stronger. With a standard MRF, the state of a site is defined as being conditionally dependent on the states of its neighboring sites, given a particular neighborhood function. In the strong-MRF case we have the extra condition that if the state of a neighboring site is undefined, then the state of the site of concern is still only conditionally dependent on those neighboring sites that are defined. In the case of a standard MRF, this in general will not be true.

At first glance, the extra constraint imposed by the strong-MRF may not seem that restrictive. However, in a standard MRF, short range correlations modeled by the MRF can imply or induce long range correlations in the field. By the nature of the strong-MRF's definition the induced long range correlations are severely limited. This has not stopped the strong-MRFs being used to model texture. In fact most non-parametric MRF techniques use the strong-MRF model without drawing reference to it. If sites with undefined states are accommodated in the synthesis algorithm without modifying the neighborhood function, then a strong-MRF is implied. To get around the limiting behavior of the strong-MRFs, either a large neighborhood is used [7], [8], [9], or a multi-scale approach is used [10].

Simplifying complex mathematical problems by assuming an extra degree of independence is a common approach to solving intractable problems, even when there is no basis for the assumption. Likewise, we have assumed an extra degree of conditional independence so as to simplify the MRF model to a strong-MRF model. By demonstrating the equivalence between the strong-MRF model and the ANOVA log-linear construction [11], [12], we are able to use the estimation processes developed for the ANOVA log-linear construction to calculate the probability distribution for the strong-MRF model. In return, Moussouris's [6] strong-MRF formula gives the general formula for the ANOVA log-linear construction.

II. Standard Markov random field model

Denote a lattice as a set of sites $\{s \in S\}$. For each site $s \in S$ define a random variable $X_s = x_s$ where $x_s \in \Lambda \doteq \{0, \ldots, L-1\}$. A particular configuration of the lattice is given as $\{X_s = x_s, s \in S\}$ which will be abbreviated to $\mathbf{X} = \mathbf{x}$. The *configuration space* for the variable $\mathbf{x}$ is denoted by $\Omega = \prod_{s \in S} \Lambda$

Let Π be the Gibbs (joint) probability measure on Ω with $\Pi(\mathbf{X} = \mathbf{x}) > 0 \ \forall \ \mathbf{x} \in \Omega$. Besag [1] proved that a Gibbs (joint) distribution $\Pi(\mathbf{x})$ is uniquely determined by its Local Conditional Probability Density Function (LCPDF) $\Pi(X_s = x_s | X_r = x_r, r \neq s)$, which we will rewrite as $\Pi_s(x_s | \mathbf{x}_{(s)}), s \in S, \mathbf{x} \in \Omega$ where $\mathbf{x}_{(s)} = \{x_r, r \neq s\}$.

The property of an MRF is that the LCPDF is defined with respect to a neighborhood $\mathcal{N}_s \subset S$ of the site s.

$$\Pi_s(x_s | \mathbf{x}_{(s)}) = P(x_s | x_r, r \in \mathcal{N}_s) \qquad \forall \mathbf{x} \in \Omega, s \in S \tag{1}$$

The set of neighborhoods is the *neighborhood system* denoted as $\mathcal{N} = \{\mathcal{N}_s \subset S, s \in S\}$. Given a neighborhood system $\mathcal{N}$, a *clique* is a set $C \subseteq S$ if every pair of distinct sites in C are neighbors, or C is a singleton with just one site. Let $\mathcal{C}$ denote the set of cliques defined on S with respect to $\mathcal{N}$, and let $\mathcal{C}_s$ denote the *local clique set* for a neighborhood $\mathcal{N}_s$ such that $\mathcal{C}_s = \{C \in \mathcal{C}, s \in C\}$.

The *Hammersley-Clifford theorem*, which is also referred to as the *MRF-Gibbs equivalence theorem*, and proved in [1], [13], [14], [6], gives form to the LCPDF so as to define a valid joint distribution $\Pi(\mathbf{x})$ by expressing the MRF in terms of $\mathcal{N}$-potentials $V_C(\mathbf{x})$ defined on the cliques $C \in \mathcal{C}_s$ [2], [13],

$$P(x_s | x_r, r \in \mathcal{N}_s) = \frac{1}{Z_s} \exp\left\{ \sum_{C \in \mathcal{C}_s} V_C(\mathbf{x}) \right\}, \tag{2}$$

where Z_s is the *local normalizing constant* $Z_s = \sum_{\lambda_s \in \Lambda} P(\lambda_s | x_r, r \in \mathcal{N}_s)$. The MRF-Gibbs equivalence theorem implicitly requires the neighborhoods system to be symmetrical and self similar for stationary homogeneous MRFs.

Given a neighborhood system $\mathcal{N}$ and its corresponding set of cliques $\mathcal{C}$, an $\mathcal{N}$-potential V is defined such that,

$$V_C(\mathbf{x}) = 0 \qquad \text{if} \qquad C \notin \mathcal{C}. \tag{3}$$

A representation for the $\mathcal{N}$-potential V is given by Grimmett [14] and Moussouris [6], but a thorough proof is given by Geman [13]. Any $\Pi > 0$ is a Gibbs distribution with respect to

$\mathcal{N}$-potentials,

$$V_C(\mathbf{x}) = \sum_{C' \subseteq C} (-1)^{|C|-|C'|} \log \Pi(\mathbf{x}^{C'}), \qquad \forall \mathbf{x} \in \Omega, s \in S. \tag{4}$$

Moreover, for any element $s \in C$,

$$V_C(\mathbf{x}) = \sum_{C' \subseteq C} (-1)^{|C|-|C'|} \log \Pi_s(x_s^{C'} | \mathbf{x}_{(s)}^{C'}), \qquad \forall \mathbf{x} \in \Omega, s \in S, s \in C \tag{5}$$

where $C, C' \in \mathcal{C}$, and for $A \subset S$,

$$\mathbf{x}^A = \{x_s^A, s \in S\}, \qquad x_s^A = \begin{cases} x_s, & s \in A \\ 0, & s \notin A. \end{cases} \tag{6}$$

The $\mathcal{N}$-potential V representation of Eq. 4 is obtained via the Möbius inversion theorem [15], for which an elegant proof is given by Moussouris [6].

Theorem 1 (Möbius inversion theorem): For arbitrary real functions F and G defined on the subsets A, B and C of some finite set.

$$F(A) = \sum_{B \subseteq A} G(B) \qquad \text{iff} \qquad G(B) = \sum_{C \subseteq B} (-1)^{|B|-|C|} F(C) \qquad \forall A, B \tag{7}$$

or, equivalently,

$$F(A) = \sum_{B \subseteq A} \sum_{C \subseteq B} (-1)^{|B|-|C|} F(C) \tag{8}$$

where $|A| =$ number of sites in set A.

A factorization of the joint probability distribution $\Pi(\mathbf{x})$ can be obtained from Möbius inversion theorem[15] (see Moussouris [6]), giving,

$$\log \Pi(\mathbf{x}) = \sum_{C \in \mathcal{C}} \sum_{C' \subseteq C} (-1)^{|C|-|C'|} \log \Pi(\mathbf{x}^{C'}), \qquad \text{where} \qquad C, C' \in \mathcal{C} \tag{9}$$

From Eq. (9) Moussouris [6] gives the following decomposition,

$$\Pi(\mathbf{x}) = \prod_{C \in \mathcal{C}} \Pi(\mathbf{x}^C)^{n_{SC}} \qquad \text{where} \qquad n_{SC} = (-1)^{|C|} \sum_{C \subseteq C' \in \mathcal{C}} (-1)^{|C'|} \tag{10}$$

where $\sum_{C \subseteq C' \in \mathcal{C}}$ is performed over the sets C'. A similar decomposition can be obtained for the LCPDF of Eq. (2), as shown by Paget [16], giving,

$$\Pi_s(x_s|\mathbf{x}_{(s)}) = \prod_{C \in \mathcal{C}_s} \left(\Pi_s(x_s|\mathbf{x}_{(s)}^C)\right)^{n_{\mathcal{C}_s C}} \qquad \text{where} \qquad n_{\mathcal{C}_s C} = (-1)^{|C|} \sum_{C \subseteq C' \in \mathcal{C}_s} (-1)^{|C'|} \tag{11}$$

III. STRONG MARKOV RANDOM FIELD THEORY

The decomposition formula of Eq. (11) is tantalizing in the fact that it gives a factorization of the neighborhood probability into clique probabilities, but unfortunately these clique probabilities are still defined over the whole neighborhood. Therefore these clique probabilities are still intractable. What would be more useful is if the factorization of the neighborhood probability could be defined in terms of clique probabilities that were defined just over their individual clique domain. That is, a factorization of the neighborhood probability into lower order functions as in the marginal distributions that can be empirically evaluated from a random field.

It was Moussouris [6] who first proposed that the Markovian system could be simplified by imposing stronger conditions on the LCPDF. The strong-MRF assumes conditional independence between non-neighboring sites for any subset of S. This is a much stronger assumption than is made for a standard MRF which defines a site as being conditionally independent upon its non-neighboring sites given all of the neighboring sites. The difference between the two models can be seen in their mathematical definitions of Eqs. (12) and (13).

MRF condition, Eq. (1)

$$\Pi_s(x_s|x_r, r \neq s) = P(x_s|x_r, r \in \mathcal{N}_s), \qquad \forall\, \mathbf{x} \in \Omega, s \in S \tag{12}$$

Strong MRF condition

$$\Pi_s(x_s|x_r, r \neq s, r \in A \subseteq S) = P(x_s|x_r, r \in \mathcal{N}_s \cap A),\ \forall\, \mathbf{x} \in \Omega, s \in S \tag{13}$$

The strong-MRF condition may be expressed in the form of the following identity. Denote the marginal probability $P(\mathbf{x}_A) = P(x_s, s \in A)$, where $A \subseteq S$. Given two sites $s, t \in S$ for which neither is a neighbor of the other, *i.e.*, $t \notin \mathcal{N}_s \Leftrightarrow s \notin \mathcal{N}_t$, and given $s, t \notin B \subseteq S$, then the strong-MRF condition of Eq. (13) can be expressed as,

$$\begin{aligned} P(x_s|x_t, x_B) &= P(x_s|x_B) \\ \frac{P(\mathbf{x}_{B+s+t})}{P(\mathbf{x}_{B+t})} &= \frac{P(\mathbf{x}_{B+s})}{P(\mathbf{x}_B)}. \end{aligned} \tag{14}$$

The notation $A + s$ is used to denote a set of sites A plus the site s, or alternatively $A - s$ denotes the same set A excluding the site s.

Proposition 1: Given a neighborhood system $\mathcal{N}$, the LCPDF of a strong-MRF may be decomposed as,

$$\log P(x_s|x_r, r \in \mathcal{N}_s) = \sum_{C \in \mathcal{C}_s} \sum_{s \in C' \subseteq C} (-1)^{|C|-|C'|} \log P(x_s|\mathbf{x}_{C'-s}) \tag{15}$$

or,

$$\log P(x_s, x_r, r \in \mathcal{N}_s) = \sum_{C \in \mathcal{C}_s} \sum_{s \in C' \subseteq C} (-1)^{|C|-|C'|} \log P(\mathbf{x}_{C'}) \tag{16}$$

where $\sum_{s \in C' \subseteq C}$ is performed over the sets C'. Through Moussouris's [6] conversion, Eqs. (15) and (16) maybe re-expressed as,

$$P(x_s | x_r, r \in \mathcal{N}_s) = \prod_{C \in \mathcal{C}_s} P(x_s | \mathbf{x}_{C-s})^{n_{\mathcal{C}_s C}}, \tag{17}$$

and,

$$P(x_s, x_r, r \in \mathcal{N}_s) = \prod_{C \in \mathcal{C}_s} P(\mathbf{x}_C)^{n_{\mathcal{C}_s C}}, \tag{18}$$

respectively where,

$$n_{\mathcal{C}_s C} = (-1)^{|C|} \sum_{C \subseteq C' \in \mathcal{C}_s} (-1)^{|C'|}, \tag{19}$$

Proposition 1 is proved via two separate mathematical constructions. The first proof is presented in Appendix I, and is based on the similar proof by Grimmett [14] and Moussouris [6] for the equivalence of a standard MRF and a Gibbs distribution. The second proof is presented in Appendix II, and is based on the ANOVA log-linear construction [11], [12] for testing independence in a distribution. As both mathematical constructions are used to prove Proposition 1, the constructions are equivalent in terms of the strong-MRF.

An example of a strong-MRF that is easy to conceptualize, is the checkerboard pattern. To show that this pattern can be modeled as a strong-MRF, we need to show that Eq. (18) holds. Given the nearest-neighbor neighborhood the empirical estimates of all the marginal and neighborhood probabilities equal either 0 or 0.5. It is easy to see that given these values of the probabilities that Eq. (18) holds. The field is still random, as two possible states of the field exist, however there also exist forbidden states.

IV. Estimation of the strong LCPDF

Bishop *et al.* [11] did not derive the direct estimate Eq. (18) or the general equation for the ANOVA log-linear construction Eq. (33), but they did prove under what conditions Eq. (33) is valid. As the ANOVA log-linear construction is equivalent to the strong-MRF model, we may use the same conditions to determine when the direct estimate of Eq. (18) is valid. Given a set of cliques with non-zero potentials over which Eq. (18) is calculated, the direct estimate is valid when these cliques do not form a loop, see Bishop *et al.* [11] page 76 for details. Therefore Eq. (18) is only valid for the auto-model of the strong-MRF.

$$P(x_s, x_r, r \in \mathcal{N}_s) = \frac{\prod_{r \in \mathcal{N}_s} P(x_s, x_r)}{P(x_s)^{|\mathcal{N}_s|-1}} \tag{20}$$

If higher order marginal distributions are desired, then an iterative proportional fitting technique may be used. Fienberg [12] and Bishop *et al.* [11] described the iterative proportional fitting technique for a distribution defined in three dimensions. A generalized version of the technique is presented by Paget [16]. A problem with the technique is that it is memory intensive and computationally expensive.

Using the marginal distributions of a field to define the strong-MRF allows non-parametric estimation to be applied. This usually implies Parzen density estimation [17]. However in using Eq. (20) to sample $\lambda_s \in \Lambda$, smoothing $P(x_s, x_r)$ along the x_s axis is probably not required or desired. Therefore substituting $P(x_s, x_r) = P(x_r|x_s)P(x_s)$ into Eq. (20) we obtain;

$$P(x_s, x_r, r \in \mathcal{N}_s) = P(x_s) \prod_{r \in \mathcal{N}_s} P(x_r|x_s) \tag{21}$$

To make the Parzen density estimate of $P(x_r|x_s)$ quick to calculate, a box kernel $K(x)$ with a smoothing parameter of h is used, whereby;

$$K(x) = \begin{cases} \alpha + \beta, & |x| \leq h \\ \beta, & \text{else} \end{cases} \qquad \text{where} \qquad \begin{array}{l} \alpha > 0 \\ \beta > 0 \end{array} \tag{22}$$

This means that only sites $r \in S$ for which $|x_r - \lambda_r| \leq h$ for some $\lambda_r \in \Lambda$ are required, and these sites can be pre-listed into an index. Define $n_s = \sum_{s \in S} \delta(x_s - \lambda_s)$ and $n_r = \sum_{s \in S} \delta(x_s - \lambda_s)\gamma(x_r - \lambda_r),\ r \in \mathcal{N}_s$ where $\delta(x) = 1$ for $|x| = 0$ else 0, and $\gamma(x) = 1$ for $|x| \leq h$ else 0. Using the kernel of Eq. (22), the direct estimate of Eq. (21) becomes,

$$P(x_s = \lambda_s, x_r = \lambda_r, r \in \mathcal{N}_s) = \frac{n_s}{|S|} \prod_{r \in \mathcal{N}_s} \frac{1}{Z_r} \left[\frac{n_r}{n_s} \alpha + \beta \right] \tag{23}$$

where $Z_r = \sum_{\lambda_r \in \Lambda} K(x_r - \lambda_r)$ is a constant if $h = 0$. As Eq. (23) implies only sampling from $\lambda_s \in \Lambda$ that are contained in the training field, neighborhood values $x_r,\ r \in \mathcal{N}_s$ will also be found in the training field. Therefore using a smoothing parameter of $h = 0$ is not a problem. This is the same argument as used by Ashikhmin [18].

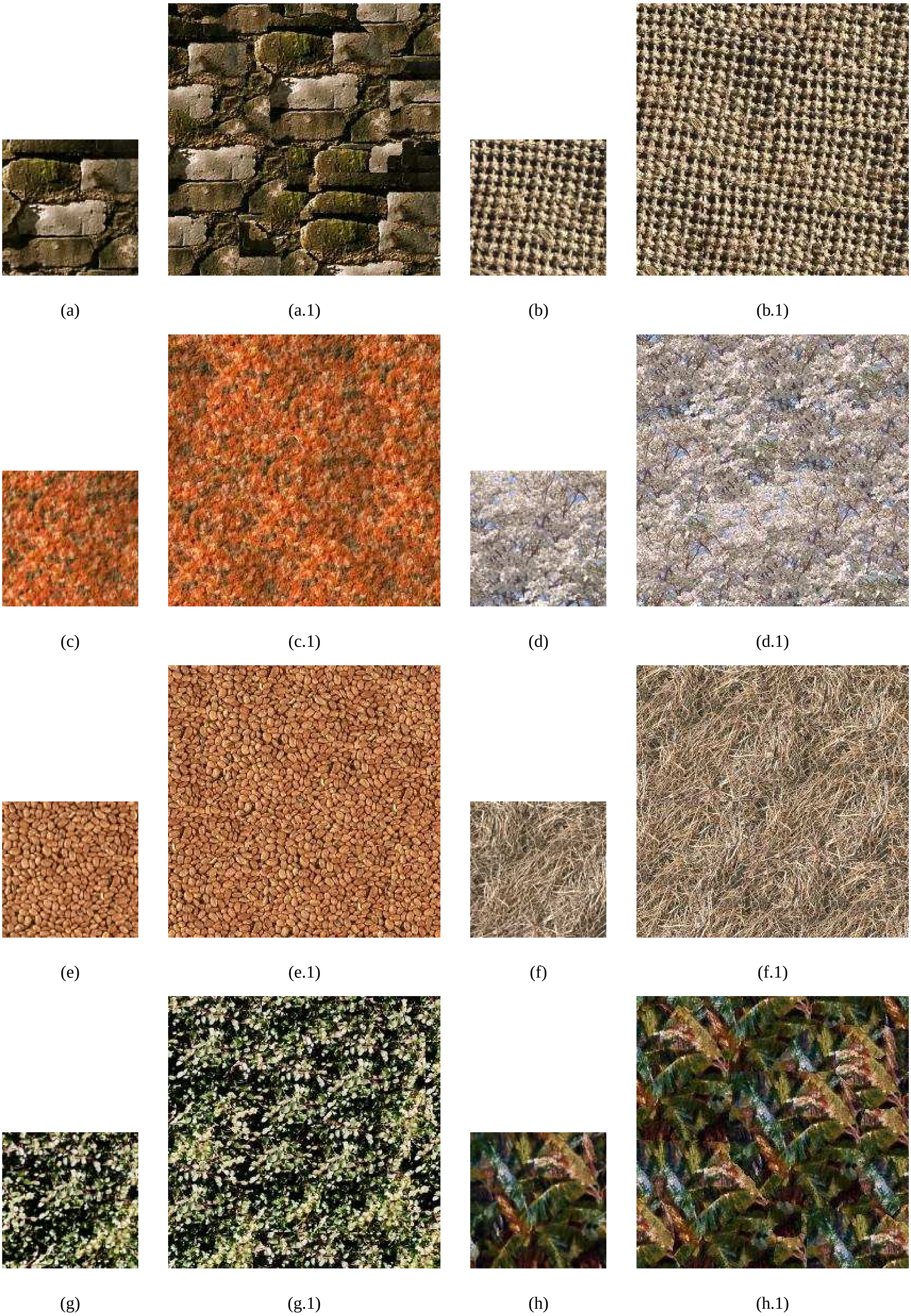

Fig. 1. VisTex textures: (a-h) Original 128×128 pixel image; (a.1-h.1) synthesized 256×256 pixel image by the non-parametric strong-MRF model using the direct estimate with a 3×3 neighborhood.

V. Synthesis

To synthesize a texture we need to sample from the LCPDF $P(x_s|x_r, r \in \mathcal{N}_s)$ to update a single site of the synthetic texture. This is repeated iteratively over the synthetic texture until the field stabilizes. As we can use the direct estimate technique to obtain a non-parametric estimate of the LCPDF, the synthesis algorithm by Paget and Longstaff [10] was used.

Fig. 1 presents 8 VisTex textures [19] which are synthesized from the non-parametric strong-MRF auto-model using the direct estimate with a 3×3 neighborhood. The visual similarity of the synthesized textures compared to the originals, demonstrates that these textures can be modeled from just second order statistics.

The synthesis algorithm was performed on 166 VisTex textures [19]. On a Sun-Blade 100: 500 MHz, 256 MBytes RAM, the average synthesis time for a 256×256 pixel image was 48 minutes with a standard deviation of 51 minutes. The synthesis results of Fig. 1, and the synthesis times are given for just 2 iterations over the whole image. More synthesis results plus source code is provided in [20].

VI. Conclusion

Low order statistical models are better at presenting more stochastic versions of synthetic texture, and are better at classification [21], [22], [23]. The strong-MRF model is not a general model and in this particular implementation will only produce visually pleasing results for certain textures. Generally, as the theory implicitly indicates, this model will only work for stationary homogeneous textures with limited long range correlations. However from the synthetic textures presented, it is clear that a surprising variety of textures can be modeled from just second order statistics. This gives credence to using second order models for classification.

For supervised classification, the strong-MRF model has not yet been shown to perform better than the standard second order models like the fractal, Gabor, GLCM, or Gaussian MRF models [21]. For texture synthesis, other non-parametric MRF models are more pliable to a wider variety of textures [7], [10], [9].

The advantage of the strong-MRF model is that it can be used to acquire a non-parametric model of any statistical order directly from any stationary homogeneous random field. The non-parametric strong-MRF model is not constrained by the functionality of arbitrary predefined potential functions. These advantages make it an excellent candidate for the application of texture recognition in images that contain other textures of unknown origin [22]. However, it is hoped that this model will light the way to finding the optimal model that will give realistic realizations of a texture while at the same time being used for its classification.

APPENDIX I
PROOF 1 OF PROPOSITION 1

This proof relies on the Möbius inversion theorem [15] Eq. (7), and is based on Grimmett's [14] and Moussouris's [6] equivalence proof for a standard MRF and a Gibbs distribution. For any sets $A, B \subseteq S$, define the potentials V_B of the strong-MRF such that;

$$\log P(\mathbf{x}_A) = \sum_{B \subseteq A} V_B(\mathbf{x}_B), \tag{24}$$

From the Möbius inversion theorem [15] Eq. (7), then,

$$V_B(\mathbf{x}_B) = \sum_{C \subseteq B} (-1)^{|B|-|C|} \log P(\mathbf{x}_C), \qquad \forall\ \mathbf{x} \in \Omega. \tag{25}$$

Moreover, for any element $s \in B$,

$$\begin{aligned} V_B(\mathbf{x}_B) &= \sum_{s \in C \subseteq C} (-1)^{|B|-|C|} \log P(\mathbf{x}_C) + \sum_{s \notin C \subseteq B} (-1)^{|B|-|C|} \log P(\mathbf{x}_C) \\ &= \sum_{s \in C \subseteq B} (-1)^{|B|-|C|} (\log P(\mathbf{x}_C) - \log P(\mathbf{x}_{C-s})) \\ &= \sum_{s \in C \subseteq B} (-1)^{|B|-|C|} \log P(x_s | \mathbf{x}_{C-s}). \end{aligned} \tag{26}$$

Given that $\mathbf{x}$ is defined on a strong-MRF with respect to $\mathcal{N}$, then V is a strong $\mathcal{N}$-potential iff $V_B(\mathbf{x}_B) = 0\ \forall\ B \notin \mathcal{C}$. Choose $B \notin \mathcal{C}$, then $\exists s, t \in B$ such that $t \notin \mathcal{N}_s \Leftrightarrow s \notin \mathcal{N}_t$.

$$\begin{aligned} V_B(\mathbf{x}_B) &= \sum_{C \subseteq B} (-1)^{|B|-|C|} \log P(\mathbf{x}_C) \\ &= \sum_{C \subseteq B-s-t} (-1)^{|B|-|C|} \log P(\mathbf{x}_C) + \sum_{C \subseteq B-s-t} (-1)^{|B|-|C+s|} \log P(\mathbf{x}_{C+s}) + \\ &\quad \sum_{C \subseteq B-s-t} (-1)^{|B|-|C+t|} \log P(\mathbf{x}_{C+t}) + \sum_{C \subseteq B-s-t} (-1)^{|B|-|C+s+t|} \log P(\mathbf{x}_{C+s+t}) \\ &= \sum_{C \subseteq B-s-t} (-1)^{|B|-|C|} \log \left[\frac{P(\mathbf{x}_C) P(\mathbf{x}_{C+s+t})}{P(\mathbf{x}_{C+s}) P(\mathbf{x}_{C+t})} \right] \\ &= 0. \end{aligned} \tag{27}$$

In obtaining Eq. (27) the strong-MRF identity of Eq. (14) is used. Therefore a strong-MRF may be expressed with respect to the $\mathcal{N}$-potentials of Eq. (25) or Eq. (26). However given the fact that here C could be the null set, Eq. (27) suggests that the autocorrelation function could be used to determine the neighborhood size.

As Eq. (25) and Eq. (26) have been shown to be $\mathcal{N}$-potentials, we can now use them to prove Proposition 1. Consider a site $s \in S$, then the strong-LCPDF may be expressed as;

$$\begin{aligned}
P(x_s|x_r, r \in \mathcal{N}_s) &= \frac{P(\mathbf{x}_S)}{P(\mathbf{x}_{S-s})} = \exp\left[\sum_{C\in\mathcal{C}} V_C(\mathbf{x}_C) - \sum_{s\notin C\in\mathcal{C}} V_C(\mathbf{x}_C)\right] \\
&= \exp\left[\sum_{s\in C\in\mathcal{C}} V_C(\mathbf{x}_C)\right] = \exp\left[\sum_{C\in\mathcal{C}_s} V_C(\mathbf{x}_C)\right] \\
\log P(x_s|x_r, r \in \mathcal{N}_s) &= \sum_{C\in\mathcal{C}_s} \sum_{s\in C'\subseteq C} (-1)^{|C|-|C'|} \log P(x_s|\mathbf{x}_{C'-s}) \qquad (28)
\end{aligned}$$

Where in the above equation, the $\mathcal{N}$-potential Eq. (26) is used since all the cliques $C \in \mathcal{C}_s$ contain the site s. This proves the first part of Proposition 1, Eq. (15). The second part of Proposition 1, Eq. (16) is proved by applying Möbius inversion theorem [15] Eq. (8) to Eq. (25) for the set of sites $\mathcal{N}_s + s \subset S$;

$$\begin{aligned}
\log P(x_s, x_r, r \in \mathcal{N}_s) &= \sum_{C\subseteq\mathcal{N}_s+s} \sum_{C'\subseteq C} (-1)^{|C|-|C'|} \log P(\mathbf{x}_{C'}) \\
&= \sum_{C\subseteq\mathcal{N}_s} \left\{ \sum_{C'\subseteq C+s} (-1)^{|C+s|-|C'|} \log P(\mathbf{x}_{C'}) + \sum_{C'\subseteq C} (-1)^{|C|-|C'|} \log P(\mathbf{x}_{C'}) \right\} \\
&= \sum_{C\subseteq\mathcal{N}_s} \left\{ \sum_{C'\subseteq C} (-1)^{|C+s|-|C'+s|} \log P(\mathbf{x}_{C'+s}) + \sum_{C'\subseteq C} (-1)^{|C+s|-|C'|} \log P(\mathbf{x}_{C'}) + \right. \\
&\qquad \left. \sum_{C'\subseteq C} (-1)^{|C|-|C'|} \log P(\mathbf{x}_{C'}) \right\} \\
&= \sum_{C\subseteq\mathcal{N}_s} \sum_{C'\subseteq C} (-1)^{|C+s|-|C'+s|} \log P(\mathbf{x}_{C'+s}) \\
&= \sum_{C\subseteq\mathcal{C}_s} \sum_{s\in C'\subseteq C} (-1)^{|C|-|C'|} \log P(\mathbf{x}_{C'}) \qquad (29)
\end{aligned}$$

Finally, to obtain Eq. (17) and Eq. (18) of Proposition 1 we may observe that both Eq. (28) and Eq. (29) have the correct Möbius set decomposition with respect to the set $\mathcal{N}_s$. Even though the site s is included in the decomposition, it is included in all cliques and therefore does not compromise the Möbius decomposition over the set $\mathcal{N}_s$. Therefore the Moussouris [6] conversion can be applied to both Eq. (28) and Eq. (29) over the set of sites $\mathcal{N}_s$ to obtain Eqs. (17) and (18) respectively.

APPENDIX II
PROOF 2 OF PROPOSITION 1

This proof is based on the ANOVA log-linear construction [11], [12] for testing independence in a distribution. In ANOVA-type notation, the probability $P(\mathbf{x}_A)$ is decomposed into its marginal distributions in terms of the *general log-linear construction* [11]:

$$\log P(\mathbf{x}_A) = \sum_{B\subseteq A} U_B(\mathbf{x}_B). \tag{30}$$

From the Möbius inversion theorem [15] Eq. (7), we have:

$$U_B(\mathbf{x}_B) = \sum_{C\subseteq B} (-1)^{|B|-|C|} \log P(\mathbf{x}_C) \tag{31}$$

giving,

$$\log P(\mathbf{x}_A) = \sum_{B\subseteq A}\sum_{C\subseteq B} (-1)^{|B|-|C|} \log P(\mathbf{x}_C) \tag{32}$$

This is the general formula for the ANOVA log-linear construction. The summation is performed over all sets $B \subseteq A$ for which the potential function $U_B(\mathbf{x}_B) \neq 0$. Moussouris's [6] conversion gives,

$$P(\mathbf{x}_A) = \prod_{C\subseteq A} P(\mathbf{x}_C)^{n_{AC}}, \qquad n_{AC} = (-1)^{|C|} \sum_{C\subseteq B\subseteq A} (-1)^{|B|} \tag{33}$$

Although this formula was proposed by Moussouris for the strong-MRF, the formula can also be applied to the study of ANOVA for contingency tables. It is not known whether this has been made apparent to the contingency tables community.

In the ANOVA log-linear construction, $U_\emptyset$ is the grand mean of the logarithmic probabilities $\log P(\mathbf{y}_A),\ \mathbf{y}_A \in \Omega_A$:

$$U_\emptyset(\mathbf{x}_\emptyset) = \frac{1}{|\Omega_A|} \sum_{\mathbf{y}_A\in\Omega_A} \log P(\mathbf{y}_A). \tag{34}$$

The rest of the potential functions $U_B(\mathbf{x}_B), B \subseteq A$, represent successive deviations from the grand mean $U_\emptyset$ such that,

$$\sum_{C\subseteq B} U_C(\mathbf{x}_C) = \frac{1}{|\Omega_{A-B}|} \sum_{\mathbf{y}_{A-B}\in\Omega_{A-B}} \log P(\mathbf{x}_B\mathbf{y}_{A-B}). \tag{35}$$

The equivalence between the ANOVA log-linear construction and the strong-MRF model is proved by proving that U is a strong $\mathcal{N}$-potential. Given that $\mathbf{x}$ is defined on a strong-MRF with respect to $\mathcal{N}$, then U is a strong $\mathcal{N}$-potential if $U_B(\mathbf{x}_B) = 0 \ \forall \ B \notin \mathcal{C}$. Choose $B \notin \mathcal{C}$, then $\exists s, t \in B$ such that $t \notin \mathcal{N}_s \Leftrightarrow s \notin \mathcal{N}_t$. From Eq. (35),

$$\begin{aligned}
\sum_{C \subseteq B} U_C(\mathbf{x}_C) &= \frac{1}{|\Omega_{A-B}|} \sum_{\mathbf{y}_{A-B} \in \Omega_{A-B}} \log P(\mathbf{x}_B \mathbf{y}_{A-B}) \\
U_B(\mathbf{x}_B) &= \frac{1}{|\Omega_{A-B}|} \sum_{\mathbf{y}_{A-B} \in \Omega_{A-B}} \log P(\mathbf{x}_B \mathbf{y}_{A-B}) - \sum_{C \subset B} U_C(\mathbf{x}_C) \\
&= \frac{1}{|\Omega_{A-B}|} \sum_{\mathbf{y}_{A-B} \in \Omega_{A-B}} \log P(\mathbf{x}_B \mathbf{y}_{A-B}) - \sum_{C \subseteq B-s} U_C(\mathbf{x}_C) - \\
&\quad \sum_{C \subseteq B-t} U_C(\mathbf{x}_C) + \sum_{C \subseteq B-s-t} U_C(\mathbf{x}_C) - \sum_{C \subset B-s-t} U_{C+s+t}(\mathbf{x}_{C+s+t}) \\
\sum_{C \subseteq B-s-t} U_{C+s+t}(\mathbf{x}_{C+s+t}) &= \frac{1}{|\Omega_{A-B}|} \sum_{\mathbf{y}_{A-B} \in \Omega_{A-B}} \log P(\mathbf{x}_B \mathbf{y}_{A-B}) - \sum_{C \subseteq B-s} U_C(\mathbf{x}_C) - \\
&\quad \sum_{C \subseteq B-t} U_C(\mathbf{x}_C) + \sum_{C \subseteq B-s-t} U_C(\mathbf{x}_C) \\
&= \frac{1}{|\Omega_{A-B}|} \sum_{\mathbf{y}_{A-B} \in \Omega_{A-B}} \log P(\mathbf{x}_B \mathbf{y}_{A-B}) - \\
&\quad \frac{1}{|\Omega_{A-B+s}|} \sum_{\mathbf{y}_{A-B+s} \in \Omega_{A-B+s}} \log P(\mathbf{x}_{B-s} \mathbf{y}_{A-B+s}) - \\
&\quad \frac{1}{|\Omega_{A-B+t}|} \sum_{\mathbf{y}_{A-B+t} \in \Omega_{A-B+t}} \log P(\mathbf{x}_{B-t} \mathbf{y}_{A-B+t}) + \\
&\quad \frac{1}{|\Omega_{A-B+s+t}|} \sum_{\mathbf{y}_{A-B+s+t} \in \Omega_{A-B+s+t}} \log P(\mathbf{x}_{B-s-t} \mathbf{y}_{A-B+s+t}) \\
&= \frac{1}{|\Omega_{A-B}|} \frac{1}{|\Omega_{s+t}|} \sum_{\mathbf{y}_{A-B+s+t} \in \Omega_{A-B+s+t}} \log P(\mathbf{x}_B \mathbf{y}_{A-B}) - \\
&\quad \frac{1}{|\Omega_{A-B+s}|} \frac{1}{|\Omega_t|} \sum_{\mathbf{y}_{A-B+s+t} \in \Omega_{A-B+s+t}} \log P(\mathbf{x}_{B-s} \mathbf{y}_{A-B+s}) - \\
&\quad \frac{1}{|\Omega_{A-B+t}|} \frac{1}{|\Omega_s|} \sum_{\mathbf{y}_{A-B+s+t} \in \Omega_{A-B+s+t}} \log P(\mathbf{x}_{B-t} \mathbf{y}_{A-B+t}) + \\
&\quad \frac{1}{|\Omega_{A-B+s+t}|} \sum_{\mathbf{y}_{A-B+s+t} \in \Omega_{A-B+s+t}} \log P(\mathbf{x}_{B-s-t} \mathbf{y}_{A-B+s+t})
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{|\Omega_{A-B+s+t}|} \sum_{\mathbf{y}_{A-B+s+t} \in \Omega_{A-B+s+t}} \log \left[\frac{P(\mathbf{x}_B \mathbf{y}_{A-B}) P(\mathbf{x}_{B-s-t} \mathbf{y}_{A-B+s+t})}{P(\mathbf{x}_{B-s} \mathbf{y}_{A-B+s}) P(\mathbf{x}_{B-t} \mathbf{y}_{A-B+t})} \right] \\
&= \frac{1}{|\Omega_{A-B+s+t}|} \sum_{\mathbf{y}_{A-B+s+t} \in \Omega_{A-B+s+t}} \log \left[\frac{P(x_s | \mathbf{x}_{B-s} \mathbf{y}_{A-B}) P(y_s | \mathbf{x}_{B-s-t} \mathbf{y}_{A-B+t})}{P(y_s | \mathbf{x}_{B-s} \mathbf{y}_{A-B}) P(x_s | \mathbf{x}_{B-s-t} \mathbf{y}_{A-B+t})} \right] \\
&= 0
\end{aligned} \tag{36}$$

To obtain Eq. (36), the strong-MRF identity Eq. (14) was used. For $B = \emptyset$ we have from Eq. (36), $U_{s+t}(\mathbf{x}_{s+t}) = 0$. Assume $U_B(\mathbf{x}_B) = 0$ for all $|B| < n$, then from Eq. (36) for $|B| = n$ we have $U_B(\mathbf{x}_B) = 0$. By the principle of mathematical induction, $U_B(\mathbf{x}_B) = 0 \ \forall \ B \notin \mathcal{C}$.

Updating Eq. (32) for a strong-MRF, the ANOVA log-linear construction may now be rewritten as,

$$\log P(x_s, x_r, r \in \mathcal{N}_s) = \sum_{\substack{C \subseteq \mathcal{N}_s + s,\\ C \in \mathcal{C}}} \sum_{C' \subset C} (-1)^{|C|-|C'|} \log P(\mathbf{x}_{C'}) \tag{37}$$

As in the derivation for Eq. (29), Eq. (37) can be re-expressed as,

$$\log P(x_s, x_r, r \in \mathcal{N}_s) = \sum_{C \subseteq \mathcal{C}_s} \sum_{s \in C' \subseteq C} (-1)^{|C|-|C'|} \log P(\mathbf{x}_{C'}) \tag{38}$$

Therefore via the ANOVA log-linear construction, Eq. (16) of Proposition 1 is proved. The rest of Proposition 1 is subsequently proved via the same derivations as used in the first proof of Proposition 1.

ACKNOWLEDGMENT

The author would first like to thank Christine Graffigne for help in checking the mathematics. The author would also like to thank the reviewers for their immensely constructive critic of this paper. Although the review process was quite laborious, the comments supplied by the reviewers were very insightful and helpful in constructing a more definitive paper. Unfortunately not all of the suggestions or comments attributed to the reviewers could be included into the short paper.

REFERENCES

[1] J. E. Besag, "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society, series B*, vol. 36, pp. 192–326, 1974.

[2] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.

[3] F. Spitzer, "Markov random fields and Gibbs ensembles," *American Mathematical Monthly*, vol. 78, pp. 142–154, 1971.

[4] G. L. Gimel'farb, "Texture modeling by multiple pairwise pixel interactions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 11, pp. 1110–1114, Nov. 1996.

[5] L. Seymour, "Parameter estimation and model selection in image analysis using Gibbs–Markov random fields," Ph.D. dissertation, The University of North Carolina, Chapel Hill, 1993.

[6] J. Moussouris, "Gibbs and Markov random systems with constraints," *Journal of Statistical Physics*, vol. 10, no. 1, pp. 11–33, 1974.

[7] A. Efros and T. Leung, "Texture synthesis by non-parametric sampling," in *International Conference on Computer Vision*, vol. 2, Sept. 1999, pp. 1033–1038.

[8] P. Harrison, "A non-hierarchical procedure for re-synthesis of complex textures," in *The 9-th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision 2001*, University of West Bohemia, Campus Bory, Plzen, Czech Republic, February 2001, http://www.csse.monash.edu.au/~pfh/resynthesizer.

[9] L.-Y. Wei and M. Levoy, "Fast texture synthesis using tree-structured vector quantization," in *SIGGRAPH 2000, 27th International Conference on Computer Graphics and Interactive Techniques*, 2000, pp. 479–488.

[10] R. Paget and D. Longstaff, "Texture synthesis via a noncausal nonparametric multiscale Markov random field," *IEEE Transactions on Image Processing*, vol. 7, no. 6, pp. 925–931, June 1998, http://www.vision.ee.ethz.ch/~rpaget.

[11] Y. M. M. Bishop, S. E. Fienberg, and P. W. Holland, *Discrete Multivariate Analysis: Theory and Practice*. Cambridge: MIT Press, 1975.

[12] S. E. Fienberg, *The Analysis of Cross–Classified Categorical Data*. MIT Press, 1981, vol. Second Edition.

[13] D. Geman, "Random fields and inverse problems in imaging," in *Lecture Notes in Mathematics*. Springer–Verlag, 1991, vol. 1427, pp. 113–193.

[14] G. R. Grimmett, "A theorem about random fields," *Bulletin of the London Mathematical Society*, vol. 5, pp. 81–84, 1973.

[15] G. C. Rota, "On the foundations of combinatorial theory," *Zeitschrift Fur Wahrscheinlichkeitstheorie Und Verwandte Gebietee*, vol. 2, pp. 340–368, 1964.

[16] R. Paget, "Nonparametric Markov random field models for natural texture images," Ph.D. dissertation, University of Queensland, St Lucia, QLD Australia, Dec. 1999.

[17] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. Stanford Reseach Institute, Menlo Park, California: John Wiley, 1973.

[18] M. Ashikhmin, "Synthesizing natural textures," in *The proceedings of 2001 ACM Symposium on Interactive 3D Graphics*, Research Triangle Park, NorthCarolina, March 2001, pp. 217–226, http://www.cs.utah.edu/~michael/ts.

[19] Vision and Modeling Group, "Vistex texture," http://www-white.media.mit.edu/vismod/imagery/VisionTexture/vistex.html, 1995.

[20] R. Paget, "Texture synthesis and analysis," http://www.vision.ee.ethz.ch/~rpaget, 2002.

[21] R. Paget and D. Longstaff, "Nonparametric markov random field model analysis of the meastex test suite," in *International Conference on Pattern Recognition*, vol. 3. Barcelona, Spain: IAPR, Sept. 2000, pp. 939–942.

[22] ——, "Open–ended texture classification for terrain mapping," in *International Conference on Image Processing*, vol. 3. Vancouver, Canada: IEEE Signal Processing Society, Sept. 2000, pp. 584–587.

[23] S. C. Zhu, Y. Wu, and D. Mumford, "FRAME: filters, random fields, rnd minimax entropy towards a unified theory for texture modeling," *Proceedings 1996 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 686–693, 1996.