

Higher-Order Clique Reduction Without Auxiliary Variables

Hiroshi Ishikawa

Department of Computer Science and Engineering

Waseda University

Okubo 3-4-1, Shinjuku, Tokyo, Japan

hfs@waseda.jp

Abstract

We introduce a method to reduce most higher-order terms of Markov Random Fields with binary labels into lower-order ones without introducing any new variables, while keeping the minimizer of the energy unchanged. While the method does not reduce all terms, it can be used with existing techniques that transforms arbitrary terms (by introducing auxiliary variables) and improve the speed. The method eliminates a higher-order term in the polynomial representation of the energy by finding the value assignment to the variables involved that cannot be part of a global minimizer and increasing the potential value only when that particular combination occurs by the exact amount that makes the potential of lower order. We also introduce a faster approximation that forego the guarantee of exact equivalence of minimizer in favor of speed. With experiments on the same field of experts dataset used in previous work, we show that the roof-dual algorithm after the reduction labels significantly more variables and the energy converges more rapidly.

1. Introduction

Recent years have seen an increasing emphasis on higher-order Markov Random Fields in vision. Although their importance was long understood and there were earlier attempts on utilizing them [5, 18, 21], their use [9, 27] and research into minimization of various classes of higher-order MRF energies [1, 6, 8, 11, 12, 17, 20, 24, 25, 26] have intensified significantly in the past few years. While there are useful higher-order MRF energies with specific forms that can be efficiently minimized [15, 16], minimization of general higher-order energy is needed to utilize sophisticated priors, especially if it is learned from data as in the case of Fields of Experts (FoEs) [23].

One approach that can minimize general higher-order MRF energy extends the popular graph-cut methods. In the case of the MRF with unary and pairwise potential, graph

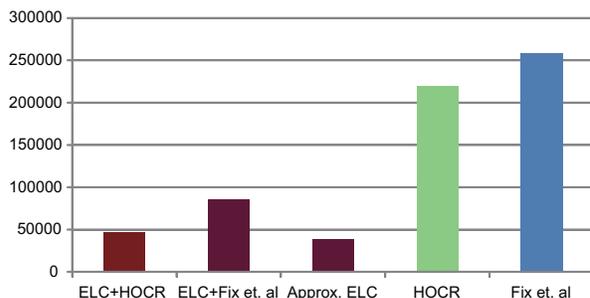


Figure 1. Comparison of the average number of variables in the first-order energy after the reduction. It can be seen that the proposed ELC reduction produces an energy with much less variables. From left to right: ELC reduction followed by the reduction in [10, 11], ELC reduction followed by the reduction in Fix et al.[6], approximate ELC reduction, the reduction in [10, 11], and the reduction in [6]. The numbers are the average over about 180 fusion-move iterations while denoising 10 test images.

cuts can globally minimize submodular energy with binary labels [13]. Even non-submodular energy can be partially solved [3, 2, 14]. Energies with more labels can be approximately minimized by “move-making” algorithms which iteratively update the solution by devising a labeling problem with binary labels and solving it with graph cuts. Popular α -expansion algorithm [4] and more general fusion-move algorithms [19] fall into this category.

To extend these methods to the higher-order case, reduction of higher-order potentials with binary labels into first-order cases have been proposed [6, 7, 8, 11, 13]. The idea is to add new variables to reduce the order of the energy, as discussed in the next section.

In this paper, we introduce a method to reduce the order of the energy *without* any additional variables. We identify the condition that this can be done and propose an algorithm to take advantage of this property. While this condition is not satisfied for all higher-order terms, in experiments we have found that it is satisfied for a significant number of such terms. In fact, in experiments on the same FoE dataset

used in previous work, more than 94% of cubic terms and more than 99% of quartic terms could be reduced without new variables. The remaining higher-order terms can be reduced by prior methods that add new variables. Thus, this method has to be used in combination with one of the existing reductions (but see below for the approximation algorithm).

The experiments also suggest that the resulting energy after the reduction using the present method is easier to minimize than those reduced by previous methods. In the experiments we show that the use of roof-dual algorithm after the reduction also results in significantly more variables labeled than when the previous methods are used alone.

We also present a faster approximation that ignores checking the condition and reduce all the higher-order terms. While this does not have the guarantee that the minimization problem stays the same, used with the fusion-move algorithm it minimizes the energy much faster than the state of the art such as Fix et al. [6].

In the next section, we describe the higher-order MRFs in more detail and discuss prior work in reducing them to first order. In section 3, we introduce the new reduction as well as the faster approximation. In section 4, we describe the experimental validation of the technique and show the results in image denoising. In section 5, we discuss some implication of the algorithms, and conclude in the last section.

2. Higher-order MRFs with binary labels and their reduction

Let us denote $\mathbb{B} = \{0, 1\}$. The Markov Random Field energy with labels in \mathbb{B} is a real-valued function on \mathbb{B}^n , which is called a pseudo-Boolean function (PBF). It can be uniquely written as a polynomial:

$$E(x) = E(x_1, \dots, x_n) = \sum_{C \in \mathcal{C}} \alpha_C x_C, \quad (1)$$

where \mathcal{C} is a set of cliques, i.e., a set of subsets of the set $V = \{1, \dots, n\}$, and α_C is a real coefficient, while

$$x_C = \prod_{i \in C} x_i \quad (2)$$

denotes the product of variables x_i with $i \in C$, i.e., a monomial. (We can consider the constant term to be the α_C for $C = \emptyset$, in which case x_C is defined to be 1.)

When the maximum number of elements in the clique is $k + 1$, or equivalently, when the degree of the polynomial is $k + 1$, the energy is said to be of k -th order. When it is quadratic, or of first order, efficient algorithms called graph cuts can be used to minimize it. Especially when all the quadratic terms have negative coefficients, it is called submodular and graph cuts can efficiently find its global minimizer. Even when the function is not submodular, roof-dual

techniques [2, 3, 14], commonly referred to as the QPBO algorithm, can give information on the global minimizer.

To take advantage of these algorithms in the case of higher-order energies, a number of techniques for reducing higher-order terms to first order have been developed.

Until a few years ago, there was only one known method [22] that can reduce general higher-order terms into first order. However, this method is considered impractical, as it uses very large coefficients that makes the resulting energy very hard to minimize.

There were other methods known to apply in the case of terms with negative coefficients [7, 13]. When $a < 0$, the identity

$$ax_1 \dots x_d = \min_{w \in \mathbb{B}} aw \left\{ \sum_i x_i - (d - 1) \right\} \quad (3)$$

allows us to replace $ax_1 \dots x_d$ in a minimization problem with $aw \left\{ \sum_i x_i - (d - 1) \right\}$, making it first order.

This left the case of positive coefficients unsolved for some time, but the gap was closed by a separate reduction [10, 11] for higher-order terms with positive coefficients.

$$ax_1 \dots x_d = a \min_{w_1, \dots, w_{n_d} \in \mathbb{B}} \sum_{i=1}^{n_d} w_i \left(c_{i,d} \left(- \sum_i x_i + 2i \right) - 1 \right) + a \left(\sum_{i < j} x_i x_j \right), \quad (4)$$

where $a > 0$ and

$$n_d = \left\lfloor \frac{d-1}{2} \right\rfloor, \quad c_{i,d} = \begin{cases} 1 & d \text{ is odd and } i = n_d, \\ 2 & \text{otherwise.} \end{cases} \quad (5)$$

Also, [11] pointed out that by flipping variables (using variables $\bar{x} = 1 - x$), reducing the term, and then flipping the variables back, we have many more transformations.

There are also other techniques [6, 8] that improve the minimizeability of resulting first order energy and decrease the number of additional variables. Gallagher et al. [8] proposed using different reduction for different term in the energy, deciding the reduction for each term so that the *number* and the *sum of the coefficients* of non-submodular terms in the reduced energy are minimized. Fix et al. [6] proved a novel identity that turns the coefficient of the highest-order term to negative, allowing the use of the preferable reduction above that only applies to terms with negative coefficients [7, 13]. They also proved a more general formula for reducing multiple higher-order terms at once, with which the coefficients of any number of higher-order terms can be turned negative with only one new variable if they have common variables.

3. Reducing higher order terms without adding variables

What is common among the reduction techniques in the previous section is that new variables are added in order to reduce the higher-order term into first order. This is a problem because more variables translate into more difficult function to minimize. In fact, it is a limiting factor for this line of approach that the number of variables needed to reduce general higher-order potential increases exponentially as the order increases, since such a potential would have an exponential number of monomials with degree > 2 .

As the main contribution of the present paper, we introduce a new reduction of higher-order terms without adding any new variables. We first illustrate the method with an example.

3.1. Raising the local potential

Suppose we have a PBF $\varphi(x, y, z)$ that takes values:

$$\begin{aligned} \varphi(0, 0, 0) &= a & \varphi(1, 0, 0) &= b \\ \varphi(0, 1, 0) &= c & \varphi(1, 1, 0) &= d \\ \varphi(0, 0, 1) &= e & \varphi(1, 0, 1) &= f \\ \varphi(0, 1, 1) &= g & \varphi(1, 1, 1) &= h. \end{aligned} \quad (6)$$

With three variables, the general PBF of this form is cubic. Then, it can be written in the polynomial form:

$$\begin{aligned} \varphi(x, y, z) &= a(1-x)(1-y)(1-z) + bx(1-y)(1-z) \\ &+ c(1-x)y(1-z) + dxy(1-z) + e(1-x)(1-y)z \\ &+ fx(1-y)z + g(1-x)yz + hxyz. \end{aligned} \quad (7)$$

Expanding it, we obtain

$$s = -a + b + c - d + e - f - g + h \quad (8)$$

as the coefficient of the highest-degree term xyz . Since xyz is the only cubic term in the expansion, it follows that if s is zero, the polynomial would be quadratic.

Suppose we define a new function $\varphi'(x, y, z)$ by using (7) but substituting $a' = a + s$ for a . Then the coefficient of xyz in $\varphi'(x, y, z)$ is

$$s' = -a' + b + c - d + e - f - g + h \quad (9)$$

$$= -(a + s) + b + c - d + e - f - g + h \quad (10)$$

$$= -a + b + c - d + e - f - g + h - s \quad (11)$$

$$= s - s = 0. \quad (12)$$

Thus, we have reduced the cubic polynomial $\varphi(x, y, z)$ into a quadratic function $\varphi'(x, y, z)$ without adding any variables.

However, these are not the same functions. Therefore we need to ascertain that this replacement does not affect

the outcome of the minimization problem. Suppose that $\varphi(x, y, z)$ is a potential in a larger energy function. Then, x, y, z are part of the many variables in the problem.

The idea is that if a is large and s is positive, and thus the minimizer of the whole energy does not include $(x, y, z) = (0, 0, 0)$ that gives $\varphi(x, y, z) = a$, then adding $s > 0$ and thus increasing the value a which is not a part of the global minimum does not change the outcome of the minimization. And of course, it does not have to be a that we raise the value. Any of the values a, d, f , and g of $\varphi(x, y, z)$, which appear with negative signs in (8) would do, as long as the assignment of 0 or 1 to the variables that gives rise to the value is not a part of the global minimizer.

Similarly, if $s < 0$, we add $-s$ to one of b, c, e , and h .

Of course, this does not guarantee that the actual global minimizer does not take that local configuration, or increasing it does not change the minimizer. Thus, we investigate the condition that does guarantee it.

3.2. Excludable local configuration

Suppose we have the energy (1):

$$E(x) = E(x_1, \dots, x_n) = \sum_{C \in \mathcal{C}} \alpha_C x_C,$$

where C denotes a subset (clique) of $V = \{1, \dots, n\}$ and

$$x_C = \prod_{i \in C} x_i.$$

For any $C \subset V$, separate the energy into two parts:

$$E(x) = E_C(x) + E_{\overline{C}}(x) \quad (13)$$

with

$$E_C(x) = \sum_{D \in \mathcal{C}, C \cap D \neq \emptyset} \alpha_D x_D, \quad (14)$$

$$E_{\overline{C}}(x) = \sum_{D \in \mathcal{C}, C \cap D = \emptyset} \alpha_D x_D. \quad (15)$$

Note that $E_{\overline{C}}(x)$ does not contain any variable $x_i, i \in C$.

For $x \in \mathbb{B}^n$ and $C \subset V$, let us denote the vector of variables indexed by C by

$$x|_C = (x_i)_{i \in C}. \quad (16)$$

Also, let \mathbb{B}^C be the set of all values that $x|_C$ can take.

Definition 1. For a clique $C \subset V$, a value assignment $u \in \mathbb{B}^C$ is said to be an *excludable local configuration* (ELC) on C if there exists another assignment $v \in \mathbb{B}^C$ such that

$$\max_{x \in \mathbb{B}^n, x|_C = v} E_C(x) < \min_{x \in \mathbb{B}^n, x|_C = u} E_C(x). \quad (17)$$

At the core of the polynomial $E_C(x)$, there is the part that consists entirely of variables with indices in C . That part does not change value if $x|_C$ is fixed to u or v . However, the value of $E_C(x)$ can change, since there are monomials in $E_C(x)$ that depend on variables with indices outside of C . Excludable local configuration $u \in \mathbb{B}^C$ and the accompanying $v \in \mathbb{B}^C$ have non-overlapping value range in spite of that variability. That is, $E_C(x)$ with $x|_C = u$ is always larger than $E_C(x)$ with $x|_C = v$, so that the global minimizer would always take v on C rather than taking u . Formally, we have the following theorem.

Theorem 1. *If $u \in \mathbb{B}^C$ is an ELC, then*

$$\min_{x \in \mathbb{B}^n} E(x) < \min_{x \in \mathbb{B}^n, x|_C = u} E(x). \quad (18)$$

In other words, no minimizer x of the energy $E(x)$ takes the local configuration u on C .

Proof. For any $x \in \mathbb{B}^n$ and $v \in \mathbb{B}^C$, let us denote by $x^{(v)}$ the vector made by “overwriting” x by v :

$$(x^{(v)})_i = \begin{cases} v_i, & (i \in C) \\ x_i, & (i \notin C) \end{cases} \quad (19)$$

Then for any $x \in \mathbb{B}^n$ and $v \in \mathbb{B}^C$, we have

$$E_{\overline{C}}(x^{(v)}) = E_{\overline{C}}(x) \quad (20)$$

since $E_{\overline{C}}(x)$ does not contain any variable x_i with $i \in C$. Also for any $v \in \mathbb{B}^C$, we have

$$\min_{x \in \mathbb{B}^n, x|_C = v} E_{\overline{C}}(x) = \min_{x \in \mathbb{B}^n} E_{\overline{C}}(x) \quad (21)$$

since any minimizer x of $E_{\overline{C}}(x)$ can be overwritten by v .

Let $\tilde{x} \in \mathbb{B}^n$ be a minimizer of $E_{\overline{C}}$. Then for any $v \in \mathbb{B}^C$,

$$\min_{x \in \mathbb{B}^n} E_{\overline{C}}(x) = E_{\overline{C}}(\tilde{x}) = E_{\overline{C}}(\tilde{x}^{(v)}). \quad (22)$$

Finally, since u is an ELC on C , there exists a $v \in \mathbb{B}^C$ satisfying (17).

Putting these together, we have

$$\begin{aligned} \min_{x \in \mathbb{B}^n, x|_C = u} E(x) &= \min_{x \in \mathbb{B}^n, x|_C = u} (E_C(x) + E_{\overline{C}}(x)) \\ &\geq \min_{x \in \mathbb{B}^n, x|_C = u} E_C(x) + \min_{x \in \mathbb{B}^n, x|_C = u} E_{\overline{C}}(x) \\ &> \max_{x \in \mathbb{B}^n, x|_C = v} E_C(x) + \min_{x \in \mathbb{B}^n} E_{\overline{C}}(x) \\ &\geq E_C(\tilde{x}^{(v)}) + E_{\overline{C}}(\tilde{x}^{(v)}) \\ &= E(\tilde{x}^{(v)}) \\ &\geq \min_{x \in \mathbb{B}^n} E(x). \end{aligned}$$

□

Assume that we would like to eliminate a higher-order monomial $\alpha_C x_C$ in the energy. If we can find an ELC $u \in \mathbb{B}^C$, then we can raise the value of the energy by α_C only when $x|_C = u$. By the theorem, no minimizer of the energy takes u on C , hence raising the value in that case does not affect the minimization problem. The only caveat is that we must take care so that u has the right parity, so that raising the value would correctly cancel the coefficient of x_C .

Definition 2. We say $u \in \mathbb{B}^C$ has an even (odd) parity if there are even (odd) number of $i \in C$ with $u_i = 1$.

3.3. Our algorithm

More precisely, our algorithm is as follows.

- 1) To eliminate a higher-order monomial $\alpha_C x_C$ in the energy, do the following.
- 2) If $\alpha_C < 0$, find an ELC with odd parity if $|C|$ is odd and even parity if $|C|$ is even; if $\alpha_C > 0$, find an ELC with the opposite parity.
- 3) If such an ELC u is found, add the polynomial

$$\psi(x) = |\alpha_C| \prod_{i \in C} \{u_i x_i + (1 - u_i)(1 - x_i)\}. \quad (23)$$

to the energy and obtain the new energy

$$E'(x) = E(x) + \psi(x). \quad (24)$$

- a) The new energy has the value

$$E'(x) = \begin{cases} E(x) + |\alpha_C| & (x|_C = u) \\ E(x) & (x|_C \neq u). \end{cases} \quad (25)$$

Because u is an ELC, any minimizer \tilde{x} of the original energy $E(x)$ has $\tilde{x}|_C \neq u$. Thus $E(x)$ and $E'(x)$ have exactly the same minimizers.

- b) The added polynomial $\psi(x)$ has the highest-degree monomial $-\alpha_C x_C$ and contain no other monomial with degree $\geq |C|$. Therefore, the monomial $\alpha_C x_C$ in $E(x)$ is eliminated in $E'(x)$ and there is one less monomial of degree $|C|$.

- 4) If such an ELC u is not found, use one of the known algorithms to reduce the higher-order monomial.

3.4. Remarks

It is important that the condition (17) for ELC can be checked relatively locally. That is, the polynomial $E_C(x)$ contains those monomials that have at least one variable in C and in a typical problem in vision, such a polynomial is fairly localized. In general, however, it is possible that $E_C(x) = E(x)$.

It is inevitable that in some cases no ELC exists. After all, the variability of $E_C(x)$ with fixed $x|_C$ has no general limit; thus it is possible that no matter how we chose a pair of fixed vectors for $x|_C$, the value ranges of $E_C(x)$ for the pair always overlap. However, in our experiments, we found more than 94% of cubic monomials and 99% of quartic monomials were eliminated.

3.5. Faster approximation

An incompleteness in the algorithm described above is that it does not specify how to find the ELC. Our algorithm in the experiments takes the brute force approach; and a problem in the current implementation is the time it takes to find the ELC. For a fixed $x|_C$, we can delete the monomials in $E_C(x)$ that contains the variables fixed to 0 and eliminate the variables fixed to 1, thereby shrinking the polynomial considerably. Then we have to find a pair of the ELC u and its accompanying v ; which we currently implement as an exhaustive search. On the other hand, this part of the algorithm can be easily parallelized, unlike the max-flow algorithm that constitutes the main graph-cut algorithm.

Here, we introduce an approximation motivated by the same intuition as the above algorithm. It does not have the theoretical guarantee that the global minima are unmoved, but in practice it makes the minimization much faster. That is, we simply find the maximizer of the higher-order potential among the value-assignments with the right parity.

Approximation Algorithm.

- 1) To eliminate a higher-order monomial $\alpha_C x_C$ in the energy, do the following.
- 2) Collect the monomials in the energy that consists of variables in C and make a polynomial:

$$\phi(x|_C) = \sum_{D \subset C} \alpha_D x_D. \quad (26)$$

- 3) If $\alpha_C < 0$, find the maximizer u of $\phi(u)$ among the u 's with odd parity if $|C|$ is odd and even parity if $|C|$ is even. If $\alpha_C > 0$, find the maximizer u of $\phi(u)$ with the opposite parity.
- 4) Add the polynomial

$$\psi(x|_C) = |\alpha_C| \prod_{i \in C} \{u_i x_i + (1 - u_i)(1 - x_i)\}. \quad (27)$$

to the energy.

This approximation eliminates the monomial with the highest degree in $\phi(x|_C)$ and reduces its degree by one; as well as raising $\phi(u)$. Since $\phi(u)$ is the maximum among the half of possible values of ϕ , it is unlikely that u is part of a minimizer of the entire energy. But there is no guarantee. On the other hand, it proved to speed up the minimization in the experiments.

4. Experimental results

Ever since the performance of the clique reduction algorithm in [10, 11] was compared with that of earlier belief propagation algorithms [18, 21], some authors [6, 12] who proposed different algorithms for minimizing higher-order energies have also compared the performance of their algorithms using the same benchmark on image denoising using the FoE model. Therefore, we used the same benchmark to compare with these state-of-the-art methods.

The FoE model represents the prior probability of an image as the product of several student-T distributions:

$$p(I) \propto \prod_C \prod_{i=1}^K \left(1 + \frac{1}{2}(J_i \cdot I_C)^2\right)^{-\alpha_i}. \quad (28)$$

where C runs over the set of all 2×2 patches in the image, and J_i is an 2×2 filter. The parameters J_i and α_i are learned from a database of natural images.

For the details of the particular FoE energy and the fusion move formulation used in the benchmark, see [11]. The algorithms compared here have the same basic structures and minimize the same third-order multiple-label energy. They are iterative algorithms in which a proposal configuration is generated in each iteration, and fused with the current configuration by minimizing a third-order energy with binary labels. This third-order energy is converted into first-order energy by various reduction methods. We compared the following reductions:

- [ELC+HOCR] the proposed reduction using the ELC, followed by the reduction in [10, 11] to reduce the higher-order terms that do not have any ELC with the correct parity,
- [ELC+Fix et al.] the proposed reduction followed by the method by Fix et al. [6], to reduce the leftover terms,
- [ELC Approx.] the approximation described in 3.5, which reduce all the higher-order terms,
- [HOCR] the reduction in [10, 11], and
- [Fix et al.] the method by Fix et al. [6], which is the state of the art.

The resulting energy of the reduction is minimized using the QPBO algorithm, which may or may not label each variable. The ratio of those labeled is an important measure of the quality of the reduction.

4.1. Results

Figure 1 compares the average number of variables in the first-order energy after the reduction. It can be seen that the proposed ELC reduction produces an energy with much

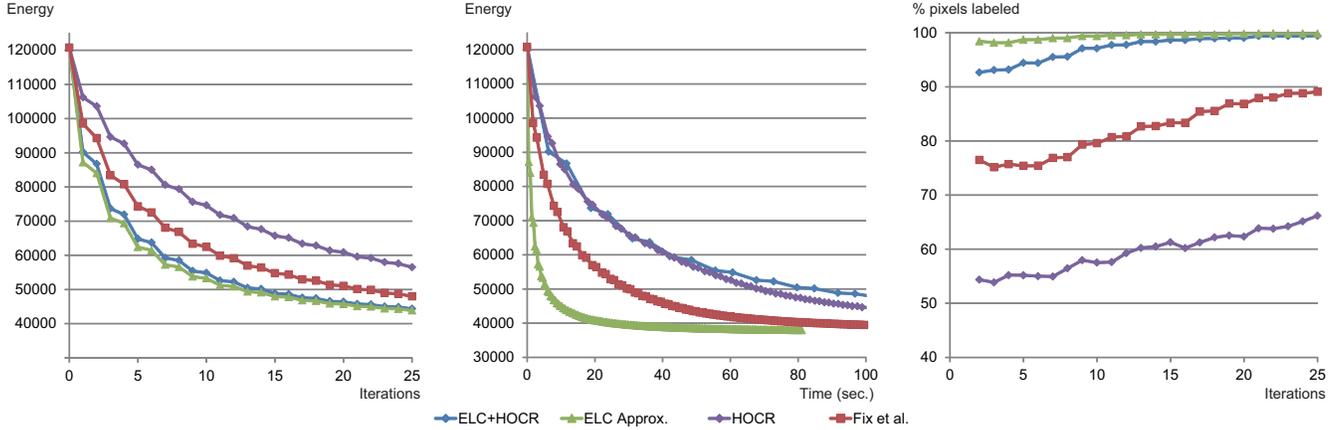


Figure 2. Energy vs. iteration (left), energy vs. time (middle), and the percentage of pixels labeled (right). ‘ELC+HOOCR’ is the ELC reduction followed by the reduction in [10, 11]; ‘ELC Approx.’ is the approximation algorithm we propose in 3.5; ‘HOOCR’ is the reduction in [10, 11]; and ‘Fix et al.’ is the reduction in Fix et al. [6]. The ELC reduction followed by the reduction in Fix et al. [6] was also tested, but is omitted here because it is almost identical with ‘ELC+HOOCR’.

less variables: the new algorithms result in about fifth of the variables as the state of the art. This results in more efficient energy minimization: Figure 2 shows the plot of the energy against iteration (left), energy against time (middle), and the percentage of pixels labeled (right).

The main observation is that the approximation algorithm is superior to the state of the art (Fix et al. [6]) in all respect, at least in this problem. Using the approximation, more than 98% of the pixels are labeled from the beginning, compared to about 75% with Fix et al. As a result, the approximation algorithm decreases the energy more in each iteration. It also takes less time for each iteration than Fix et al., so that the plot of the energy with time shows even larger difference between the proposed approximation algorithm and Fix et al. [6].

Figure 3 shows the denoising examples. Each row shows, from left to right, the noise-added image, the denoised image using HOOCR [10, 11], the denoised image using Fix et al. [6], and the denoised image using the approximation algorithm proposed in 3.5. Each denoising was done only for 10 seconds on a machine with Intel Core i7-4770K CPU with clock speed 3.5GHz. The resulting energy values are shown below each image. Our result obtains consistently lower energy in the same amount of time.

4.2. Approximation quality

We also empirically evaluated the quality of the approximation by the algorithm in 3.5.

For each cubic or quartic term, the algorithm takes the polynomial ϕ that consists of all the monomials in the energy that contain only the variables in the term, and then finds the local configuration u that maximizes $\phi(u)$, and uses u instead of the ELC, i.e., raises the energy value by a

constant only when u is a part of the global configuration. We found on average u is in fact an ELC for 88% of cubic terms and 97% of quartic terms, which means that the energy was correctly reduced for these terms even in the approximation algorithm.

Even if u is not an ELC, it is still possible that raising $\phi(u)$ does not affect the minimization problem, i.e., that u is not a part of a global minimizer. To see the extent that this is the case, for each term that the maximizer configuration u is not an ELC, we checked if it is a part of a global minimizer. We used the Fix et al. reduction for the original energy and minimized the reduced energy using the QPBO (roof dual) algorithm. Because of the persistency, when variables are labeled, they are part of a global minimizer. About 65% of the cubic terms and 58% of the quartic terms for which u is not an ELC had all their variables labeled. For very few (0.02% for cubic terms and 0.003% for quartic terms) among them, u was in fact a part of the global minimizer.

In general, quartic terms seem more likely to be reduced correctly; this is to be expected since there are more possible configurations of variable assignments for quartic terms than for cubic terms (2^4 vs. 2^3), and thus less probability that the chosen maximizer configuration turns out to be a part of a global minimizer. Thus, for terms of even higher order, we can expect the approximation algorithm to work even better.

Comparing pixel by pixel, about 83.3% of them are labeled both using the approximation algorithm and using the Fix et al. algorithm. Among them, only 0.00012% were labeled differently.

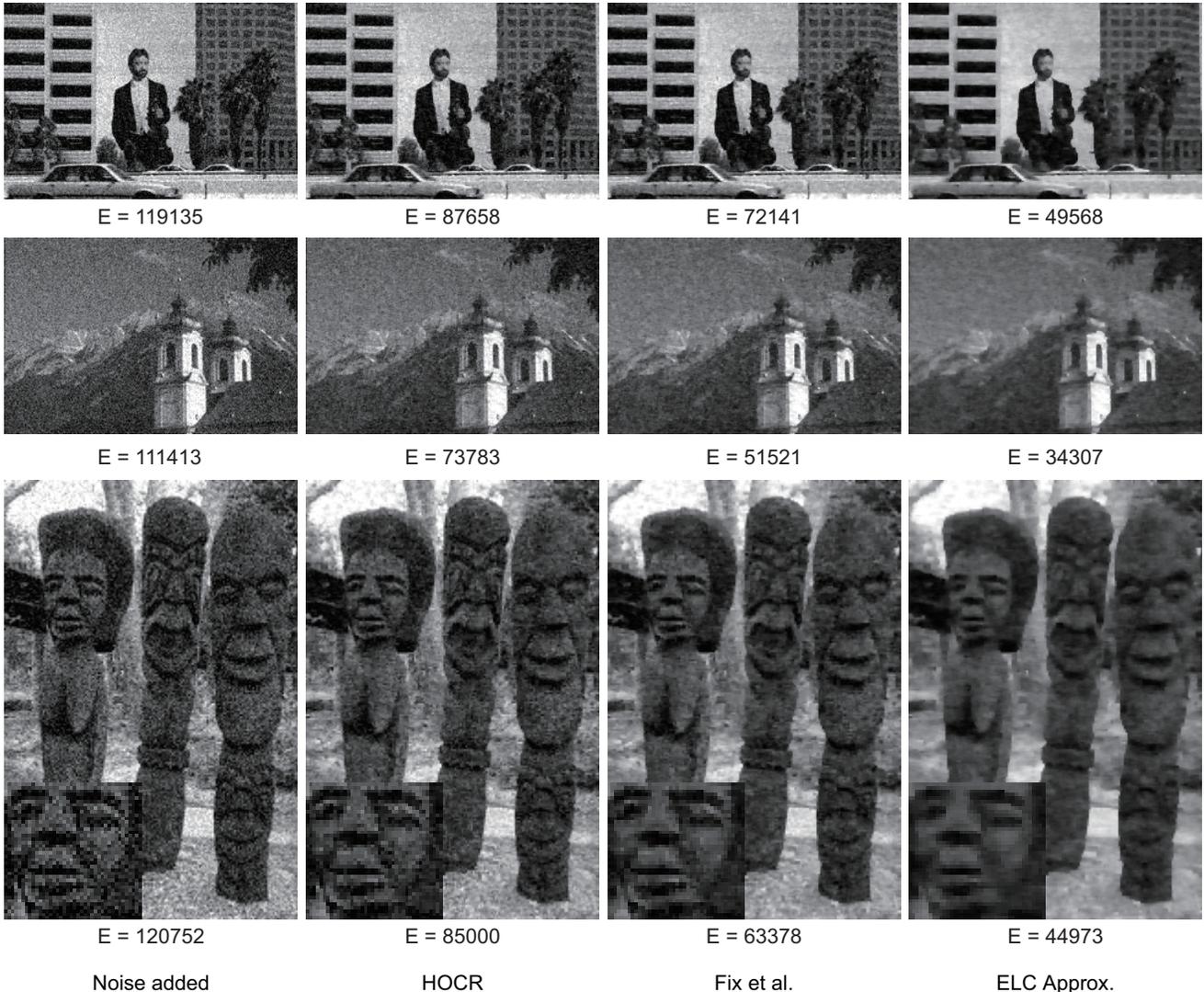


Figure 3. Image denoising examples. From the left, the noise-added image, denoised using HOCR [10, 11], Fix et al. [6], and the approximation algorithm proposed in 3.5. To show the difference in speed, the denoising was done for 10 seconds in each case.

5. Discussion

There have been some objections to the logic of our argument in this paper. On one hand, we argue that higher-order energies are essential for inference problems in vision; on the other, we just approximate the higher-order energy by a lower-order one and minimize it. This may seem to indicate that higher-order energies are mostly not necessary.

However, it is not that simple. The actual energy that represents the inference problem is often, as in the case of the FoE example, not directly reducible to lower order by the present method or others, since it has more labels than two. The binary energy in each iteration only arises in the process of fusion, after some proposal is chosen. So we can only reduce the resulting binary energy after there is

some interaction between the current configuration and the proposal; and it does not mean that the original multi-label energy can always be of first order.

Also, the multi-label energy itself is a combination of different factors such as the prior and the likelihood. In the case of FoE, the higher-order part is the prior that is learned from a database of natural images. The important point here is that the learned prior contains the information that is useful for inference based on *any* given natural image. Thus it is entirely possible that, even if the combination of the prior and the given image somehow produces an instance of the inference problem that can be approximated by a first-order energy, the prior itself must be of higher-order to maintain the information on the distribution of all the natural images.

6. Conclusion

In this paper, we introduce a method to reduce the order of potentials in an MRF energy without adding any new variables. For that, we raise a value of the potential only when its variables take a particular configuration that satisfies a certain condition so that the highest-order monomial in the polynomial representation of the potential vanishes. We investigate the condition so that this modification of the potential can be done without affecting the minimizer of the energy. While this condition is not always satisfied, in experiments on the same FoE dataset used in previous work, more than 94% of cubic terms and more than 99% of quartic terms satisfied the condition and thus could be reduced without adding new variables. We also introduce an approximation algorithm that raise the value without checking the condition, which turns out to be more efficient than the state of the art in experiments.

Acknowledgment. This work was partially supported by the JSPS Grant-in-Aid for Scientific Research (B) #24300075 and the JSPS Grant-in-Aid for Challenging Exploratory Research #25540075.

References

- [1] C. Arora, S. Banerjee, P. Kalra, and S. N. Maheshwari. “Generic Cuts: An Efficient Algorithm for Optimal Inference in Higher Order MRF-MAP.” In *ECCV2012*, V:17–30, 2012. 1
- [2] E. Boros, P. L. Hammer, R. Sun, and G. Tavares, “A max-flow approach to improved lower bounds for quadratic unconstrained binary optimization (QUBO),” *Discrete Optimization*, vol. 5, issue 2, pp. 501–529, 2008. 1, 2
- [3] E. Boros, P. L. Hammer, and G. Tavares. “Preprocessing of unconstrained quadratic binary optimization.” Technical Report RRR 10-2006, RUTCOR Research Report, Apr 2006. 1, 2
- [4] Y. Boykov, O. Veksler, R. Zabih. “Fast Approximate Energy Minimization via Graph Cuts.” *IEEE Trans. PAMI* **23**:1222–1239, 2001. 1
- [5] D. Cremers and L. Grady. “Statistical Priors for Efficient Combinatorial Optimization Via Graph Cuts.” In *ECCV2006*, III:263–274, 2006. 1
- [6] A. Fix, A. Gruber, E. Boros, and R. Zabih. “A Graph Cut Algorithm for Higher-order Markov Random Fields.” In *ICCV2011*, pp. 1020–1027, 2011. 1, 2, 5, 6, 7
- [7] D. Freedman and P. Drineas. “Energy minimization via graph cuts: Settling what is possible.” In *CVPR2005* II:939–946, 2005. 1, 2
- [8] A. C. Gallagher, D. Batra, and D. Parikh, “Inference for Order Reduction in Markov Random Fields.” In *CVPR2011*, pp. 1857–1864, 2011. 1, 2
- [9] B. Glocker, H. Heibel, N. Navab, P. Kohli, and C. Rother. “Triangleflow: Optical flow with triangulation-based higher-order likelihoods.” In *ECCV2010*, III:272–285, 2010. 1
- [10] H. Ishikawa. “Higher-Order Clique Reduction in Binary Graph Cut.” In *CVPR2009*, pp. 2993–3000, 2009. 1, 2, 5, 6, 7
- [11] H. Ishikawa. “Transformation of General Binary MRF Minimization to the First Order Case.” *IEEE Trans. Patt. Anal. Mach. Intell.*, **33**(6):1234–1249, 2011. 1, 2, 5, 6, 7
- [12] F. Kahl and P. Strandmark. “Generalized Roof Duality.” *Discrete Applied Mathematics*, **160**(16–17):2419–2434, 2012. 1, 5
- [13] V. Kolmogorov and R. Zabih. What Energy Functions Can Be Minimized via Graph Cuts? *IEEE Trans. PAMI* **26**(2):147–159, 2004. 1, 2
- [14] V. Kolmogorov and C. Rother. Minimizing Non-submodular Functions with Graph Cuts - A Review. *IEEE Trans. PAMI* **29**(7):1274–1279, 2007. 1, 2
- [15] P. Kohli, M. P. Kumar and P. H. S. Torr. “P3 & Beyond: Move Making Algorithms for Solving Higher Order Functions.” *IEEE Trans. PAMI*, **31**(9):1645–1656, 2009. 1
- [16] P. Kohli, L. Ladický and P. H. S. Torr. “Robust Higher Order Potentials for Enforcing Label Consistency.” *Int. J. Computer Vision*, **82**(3):302–324, 2009. 1
- [17] N. Komodakis and N. Paragios, “Beyond Pairwise Energies: Efficient Optimization for Higher-order MRFs.” In *CVPR2009*, pp. 2985–2992, 2009. 1
- [18] X. Lan, S. Roth, D. P. Huttenlocher, and M. J. Black. “Efficient belief propagation with learned higher-order Markov random fields.” In *ECCV2006*, II:269–282, 2006. 1, 5
- [19] V. Lempitsky, C. Rother, and A. Blake. “LogCut - efficient graph cut optimization for Markov Random Fields.” In *ICCV2007*. 1
- [20] A. Osokin and D. Vetrov. “Submodular Relaxation for MRFs with High-Order Potentials.” In *ECCV2012*, III:305–314, 2012. 1
- [21] B. Potetz. “Efficient belief propagation for vision using linear constraint nodes.” In *CVPR2007*. 1, 5
- [22] I. G. Rosenberg. “Reduction of bivalent maximization to the quadratic case.” *Cahiers du Centre d’Etudes de Recherche Operationnelle* 17. pp. 71–74, 1975. 2
- [23] S. Roth and M. J. Black. “Fields of Experts.” *Int. J. Computer Vision*, **82**(2):205–229, 2009. 1
- [24] S. Ramalingam, P. Kohli, K. Alahari, and P. H. S. Torr. “Exact inference in multi-label CRFs with higher order cliques.” In *CVPR2008*. 1
- [25] C. Rother, P. Kohli, W. Feng, and J. Jia, “Minimizing Sparse Higher Order Energy Functions of Discrete Variables,” In *CVPR2009*, pp. 1382–1389, 2009. 1
- [26] T. Windheuser, H. Ishikawa, and D. Cremers. “Generalized Roof Duality for Multi-Label Optimization: Optimal Lower Bounds and Persistency.” In *ECCV2012*, VI: 400–413, 2012. 1
- [27] O. J. Woodford, P. H. S. Torr, I. D. Reid, and A. W. Fitzgibbon. “Global stereo reconstruction under second-order smoothness priors.” *IEEE Trans. PAMI*, **31**(12):2115–2128, 2009. 1