Toward Global Solution of MAP Image Estimation: Using Common Structure of Local Solutions

Stan Z. Li
School of Electrical and Electronic Engineering
Nanyang Technological University, Singapore 639798
szli@szli.eee.ntu.ac.sg

Abstract. The maximum a posteriori (MAP) principle is often used in image restoration and segmentation to define the optimal solution when both the prior and likelihood distributions are available. MAP estimation is equivalent to minimizing an energy function. It is desirable to find the global minimum. However, the minimization in the MAP image estimation is non-trivial due to the use of contextual constraints between pixels. Steepest descent methods such as ICM quickly finds a local minimum but the solution quality depends much on the initialization. Some initializations are better than others. In this paper, we present an iterative optimization algorithm, called the Comb algorithm, for approximating the global minmum. The Comb maintains a number of best local minima found so far. It uses the Common structure of Best local minima (hence "Comb") to derive new initial configurations. Because the derived configurations contain some structure resembling that of the global minimum, they may provide good starting points for local search to approach the global minimum. Experimental comparisons show that the Comb produces solutions of quality much better than ICM and comparable to simulated annealing.

1 Introduction

Image restoration is to recover a degraded image and segmentation is to partition an image into regions of similar image properties. Efficient restoration and segmentation are very important for numerous image analysis applications. Both problems can be posed generally as one of image estimation where the underlying image or segmentation map is to be estimated from the degraded image. Due to various uncertainties, an optimal solution is sought. A popular optimality criterion is the maximum *a posteriori* (MAP) probability principle in which both the prior distribution of the true image class and the conditional (likelihood) distribution of the data are taken into account. Contextual constraints, *i.e.* constraints between pixels, are important in image analysis. Markov random fields (MRFs) or equivalently Gibbs distributions provide a convenient tool for modeling prior distributions of images which encode contextual constraints. Maximizing the posterior is equivalent to minimizing the energy function in the corresponding Gibbs distribution. The MAP principle and MRF together form the MAP-MRF framework [5, 9].

Minimization methods are an important part of the energy minimization approach. When the pixels of the image to be recovered take discrete values, as is the case dealt with in this paper, the minimization is combinatorial. It is desirable to find the global minimum. However, no practical algorithms guarantee a global minimum. The complication increases due to contextual constraints used in the MAP-MRF image estimation.

Combinatorial optimization methods that are often used in statistical image analysis literature include the iterative conditional modes (ICM) [2] and simulated annealing (SA) [8, 5]. The deterministic ICM uses the *steepest descent* strategy to perform local search. Although it quickly finds a local minimum but the solution quality depends much on the initialization. Some initializations are better than others. An extension to steepest descent is the *multi-start* method: initialize a set of random configurations drawn from a uniform distribution, apply steepest descent independently to everyone configuration, and choose, among all resultant local minima, the one with the lowest energy value as the final solution. The SA is a stochastic algorithm, as opposed to deterministic ones. It is shown that the SA with a slow enough schedule finds a global solution with probability approaching one [5]. But such a slow schedule is impractical in applications and therefore in practice, a quick annealing schedule is usually used with the loss of the guaranteed convergence to the global minimum.

There have been developments in population-based methods such as genetic algorithms (GAs) [6] in recent years. Unlike the above mentioned methods which operate on a single configuration, a population-based method maintains and operates on a population of individuals, *i.e.* a collection of configurations. Two operations are used to produce offspring: crossover and mutation. The resulting offspring update the population according to the fittest survive principle. Heuristics can be incorporated into GAs to constitute hybrid GAs [6]. Combining local (minimization) search with a GA yields a hybrid GA also called hill-climbing GA or memetic algorithm [11, 13]. Applications of GAs in the image and vision area have also taken place; see for example [3, 7, 14].

In this paper, we present a new random search method, called the Comb method, for discrete optimization in MRF-based image restoration and segmentation. The Comb method maintains a number of the best local minima found so far, as a population-based method. It uses the Common structure of Best local minima (hence "Comb") to infer the structure of the global minimum. More specifically, it derives, from two chosen best local minima, one (or two) new initial configuration in such a way: For each pixel location, if the pixel labels (values) of the two local minima are the same (in common), the label is copied to the corresponding location of the new configuration; otherwise, a random label is generated and assigned to it. The configuration thus derived contains about the same percentage of common labels as the two local minima (assuming the two have about the same percentage of common labels). However, because it is no longer a local minimum, it may be improved by using a local minimization procedure to achieve a new local minimum. If the new one has a lower energy value, it updates the existing set of the best ones. This process is repeated until some termination conditions are satisfied.

The resulting Comb algorithm is equivalent to a GA hybridized with steepest descent, in which the Comb initialization works like a uniform crossover operator with a uniform selection. There have been various interpretations for the crossover operation. The intuitive motivation of encouraging the common structure provides a new perspective for interpreting the crossover operation in GA.

Experiment results in both image restoration and segmentation are provided to compare the Comb method with the ICM, HCF [4] and SA. The results show that the Comb yields better quality solutions than the ICM and HCF and comparable to SA.

The rest of the paper is organized as follows: Section 2 describes the Comb method for MAP-MRF image restoration and segmentation. Sections 3 discusses relationship between the Comb and GA. Section 4 presents the experimental comparisons.

2 The Comb Method for MAP-MRF Image Estimation

In this section, we formulate the energy function for MAP-MRF Image Restoration and Segmentation, define local energy minima, and describe the Comb method for obtaining good local minima.

2.1 MAP-MRF Image Restoration and Segmentation

Let $\mathcal{S} = \{1, \ldots, m\}$ index the set of sites corresponding to image pixels and denote the underlying image as $f = \{f_1, f_2, \ldots, f_m\} = \{f_i \mid i \in \mathcal{S}\}$. In our image estimation problem, every pixel can take on a discrete value in the label set $\mathcal{L} = \{1, \ldots, M\}$, *i.e.* $f_i \in \mathcal{L}$. Therefore, f is a configuration in the solution space $\mathbb{F} = \mathcal{L}^m$.

The spatial relationship of the sites, each of which is indexed by a single number in $\mathcal{S}$, is determined by a neighborhood system $\mathcal{N} = \{\mathcal{N}_i \mid i \in \mathcal{S}\}$ where $\mathcal{N}_i$ is the set of sites neighboring i. A single site or a set of neighboring sites form a clique denoted by c. In this paper, only up to pair-site cliques defined on the 8-neighborhood system are considered.

The type of the underlying image f can be blob-like regions or a texture pattern. Different types are due to different ways of interaction between pixels, *i.e.* due to different contextual interactions. Such contextual interactions can be modeled as MRFs or Gibbs distributions of the form $P(f) = Z^{-1} \times \mathrm{e}^{-\sum_{c \in \mathcal{C}} V_c(f)}$ where $V_c(f)$ is the potential function for clique c, $\mathcal{C}$ is the set of all cliques, and Z is the normalizing constant.

Among various MRFs, the multi-level logistic (MLL) model is a simple yet powerful mechanism for encoding a large class of spatial patterns such as textured or non-textured images. In MLL, the pair-site clique potentials take the form: $V_2(f_i, f_{i'}) = \beta_c$ if sites on clique $\{i, i'\} = c$ have the same label or $V_2(f_i, f_{i'}) = -\beta_c$ otherwise where β_c is the parameter for type-c cliques; while the single site potentials is defined by $V_1(f_i) = \alpha_I$ where α_I is the potential for the label $I = f_i$.

When the true pixel values are contaminated by identical independently distributed (i.i.d.) Gaussian noise, the observed data, or the likelihood model, is $d_i = f_i + e_i$ where $e_i \sim N(0, \sigma^2)$ is the zero mean Gaussian distribution with standard deviation σ. With these prior and likelihood models, the energy in the posterior distribution $P(f \mid d) \propto \mathrm{e}^{-E(f)}$ is

$$E(f) = \sum_{i \in \mathcal{S}} (f_i - d_i)^2/[2\sigma^2] + \sum_{\{i\} \in \mathcal{C}} V_1(f_i) + \sum_{\{i,i'\} \in \mathcal{C}} V_2(f_i, f_{i'}) \tag{1}$$

The MAP estimate for the restoration or segmentation is defined as

$$f^* = \arg\min_{f \in \mathbb{F}} E(f) \tag{2}$$

The above minimization is combinatorial because the minimization of $E(f)$ is performed in the discrete space $\mathbb{F}$.

In this paper, we assume that the objective function, *i.e.* the energy function $E(f)$, has been fully defined, which means that the MRF and noise parameters α, β and σ are known; we concentrate on the problem of minimizing $E(f)$.

2.2 Local Minima

The minimal solution defined in (2) is meant to be the global one, *i.e.* the one with the lowest possible energy value. Finding the global solution of a combinatorial optimization problem is usually intractable, and therefore a local solution, hopefully a good one, is found instead.

A local minimum f is said with respect to its neighborhood $\mathcal{N}(f)$ composed of all the configurations neighboring f. We may define the following "*k-neighborhood* of a configuration f^*

$$\mathcal{N}^k(f^*) = \{x \in \mathbb{F} \mid x \text{ differs from } f^* \text{ by at most } k \text{ labels}\} \tag{3}$$

For example, assuming that $\mathcal{L} = \{a, b, c\}$ is the set of allowable label for every f_i $(i = 1, \ldots, 5)$, then $\{a, b, a, a, a\}$ is a 1-neighbor of $\{a, a, a, a, a\}$ and $\{b, a, a, a, c\}$ is a 2-neighbor of it. Note that $\{a, b, a, a, a\}$ is also a 2-neighbor of $\{a, a, a, a, a\}$ (due to the phrase "at most" in the definition) but not vice versa.

A point f^* is a *local minimum* of E with respect to $\mathcal{N}$ if

$$E(f^*) < E(f) \quad \forall f \in \mathcal{N}(f^*) \tag{4}$$

A local minimum is also the *global minimum* if the neighborhood is defined to include all the other configurations, $\mathcal{N}(f^*) = \mathbb{F} - \{f^*\}$. Local solutions are often defined with respect to the 1-neighborhood, $\mathcal{N}^1$, because finding such a local solution is computationally more economical than using higher k-neighborhood.

```
initialize f = (f_1, ..., f_m);
repeat
      changed=0;
      for each i ∈ S in turn
            f_i^new = arg min_{f_i ∈ L} E(f_i | f − {f_i});
            if f_i^new ≠ f_i then
                  f_i = f_i^new, changed=1;
until (changed==0);
return f as f*;
```

Fig. 1. A steepest descent algorithm for finding a local minimum with respect to $\mathcal{N}^1$.

The *steepest descent* is a fast algorithm for finding a local solution. Let $E(f_i \mid f - \{f_i\})$ be the local energy of label f_i given all the other labels. Such an algorithm iteratively minimizes $E(f_i \mid f - \{f_i\})$ for each $i \in \mathcal{S}$. Fig.1 gives a sequential version of it. The "coding method" [1] may be incorporated for a parallel implementation.

Although fast, steepest descent has two major drawbacks: First, once reaching a local minimum, it can no more improve it. This is unlike the SA in which occasional jump from a lower energy configuration to a higher one is allowed so that improvements are still possible. Second, the local solution found by the steepest descent depends very much on the initial configuration. A good initialization scheme is always desired. A question is how to derive a good initial configuration. The Comb provides an answer.

2.3 The Comb Method

The Comb method maintains a number of N best local minima found so far, denoted $F = \{f^{[1]}, \ldots, f^{[N]}\}$, as a population-based method. In every iteration, it derives a new initial configuration from F, and perform steepest descent using the derived initial configurations. If the found local minimum is better than an existing one in F, it replaces it.

Ideally, we desire that all local minima in F should converge towards the global minimum $f^{[global]}$, in which case, there must be

$$f_i^{[n]} = f_i^{[global]} \qquad 1 \le n \le N \tag{5}$$

for all $i \in \mathcal{S}$. We call $f_i^{[global]}$ the *minimal label* at i. To achieve (5), all the labels at i, $\{f_i^{[n]} \mid \forall n\}$, should finally converge to the minimal label $f_i^{[global]}$. The Comb is aimed to perform towards this objective.

The following heuristic is the basis for deriving new initial configurations: Although $f^{[1]}, \ldots, f^{[N]}$ are local minima, they share some structure with the

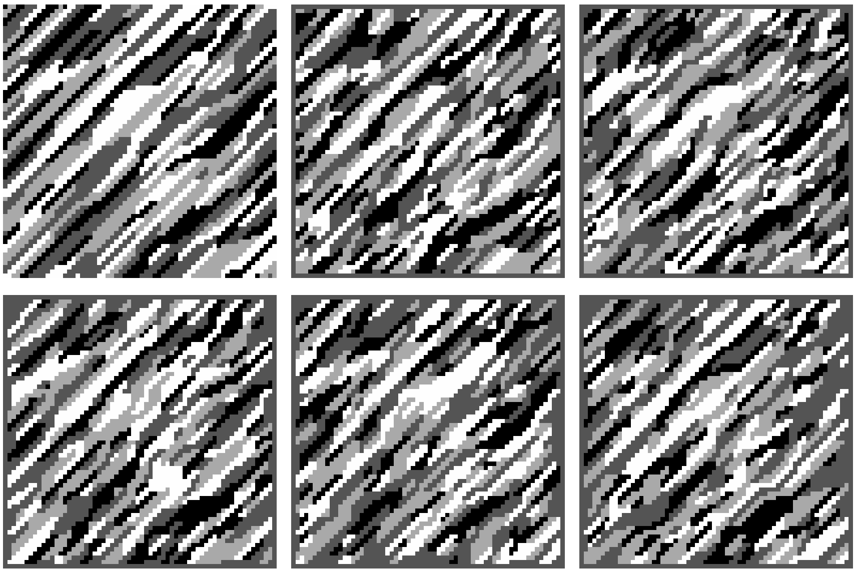


Fig. 2. The global minimum (upper-left) and 5 local minima. The local minima share some structure with the global minimum.

global minimum $f^{[global]}$. More specifically, some local minima $f^{[n]}$ have the minimal label $f_i^{[n]} = f_i^{[global]}$ for some $i \in \mathcal{S}$. Fig.2 shows the (approximate) global minimum for an MAP-MRF restoration problem and some local minima found by using the multi-start method with initially random configurations. A statistic over a number of $N = 10$ local minima is made to see how many minimal labels they have. Table 1 shows the statistic in terms of the percentile of the sites (pixels) $i \in \mathcal{S}$ at which at least k local minima $f^{[n]}$ have the minimal label $f_i^{[n]} = f_i^{[global]}$.

The Comb initialization is aimed to derive configurations having a substantial number of minimal labels so as to improve F towards the objective of (5). Although a configuration with a larger number of minimal labels does not necessarily have a lower energy value, we hope that it serves as a good initial configuration for a steepest descent algorithm.

k	0	1	2	3	4	5	6	7	8	9	10
%	100	98	95	87	75	60	43	28	16	7	2

Table 1. Percentile (rounded-up to integers) of the sites (pixels) $i \in \mathcal{S}$ at which *at least* k local minima $f^{[n]}$ have the same label $f_i^{[n]} = f_i^{[global]}$ as the global minimum $f^{[global]}$.

```
comb_initialization(f^[a], f^[b], F)
begin
        for all i ∈ S:
            if (f_i^[a] == f_i^[b] && rand[0,1] < 1 − τ)
                then f_i^[0] = f_i^[a];
            else
                f_i^[0] = rand(L);
end
```

```
Comb_Algorithm
begin
        initialize the set F = {f^[1], ..., f^[N]};
        do {
            random_selection(f^[a], f^[b], F);
            comb_initialization(f^[0], f^[a], f^[b]);
            steepest_descent(f^*, f^[0]);
            update(F, f^*);
        } until (termination condition satisfied);
        return(arg min_{f∈F} E(f));
end
```

Fig. 3. The Comb algorithm with a standard Comb-initialization scheme.

The Comb algorithm is described in Fig.3. The initialization at the beginning of the Comb algorithm is done according to a uniform distribution as the multi-start method. This is followed by iterations of four steps. Firstly, two local minima in F, $f^{[a]}$ and $f^{[b]}$, $(a \neq b)$ are randomly selected according to a uniform distribution. Secondly, a new initial configuration $f^{[0]}$ is derived from $f^{[a]}$ and $f^{[b]}$ using the standard Comb initialization, which will be explained shortly. Thirdly, steepest descent is applied to $f^{[0]}$ to produce a local minimum f^*. Then, the worst solution in F is replaced by f^* if the former has a higher energy than the latter, $\max_{f\in F} E(f) > E(f^*)$. The termination condition may be that either all configurations in F are the same or a certain number of local minima have been performed. The algorithm returns the best local minimum in F, *i.e.* the one having the lowest energy.

The central part of the Comb method is the derivation of new initial configurations. The Comb is aimed to derive $f^{[0]}$ in such a way that $f^{[0]}$ contains as many minimal labels as possible. Because the minimal labels are not known *a priori*, the Comb attempts to use the common structure, or *common labels*, of $f^{[a]}$ and $f^{[b]}$ to infer the minimal labels. We say that f_i^{comm} is a common label of $f^{[a]}$ and $f^{[b]}$ if $f_i^{comm} = f_i^{[a]} = f_i^{[b]}$. The Comb makes a hypothesis that f_i^{comm}

is a minimal label if $f_i^{[a]} = f_i^{[b]}$. The Comb initialization schemes are illustrated as follows:

1. The *basic Comb* initialization. For each $i \in \mathcal{S}$, if $f_i^{[a]}$ and $f_i^{[b]}$ are identical, then set $f_i^{[0]} = f_i^{[a]}$; otherwise set $f_i^{[0]} = rand(\mathcal{L})$ which is a label randomly drawn from $\mathcal{L}$. The use of the common label in the initial configuration encourages the enlargement of the common structure in the local minimum to be found subsequently.
2. The *standard Comb* initialization (the one in Fig.3). The basic Comb initialization is accepted with a probability $1 - \tau$ where $0 < \tau < 1$ (in Fig.3, $rand[0, 1]$ is a uniformly distributed number in $[0, 1]$). The probabilistic acceptance of common labels diversifies the search and prevents F from converging to a local minimum too soon.

Then, how many minimal labels are retained in $f^{[0]}$ as the result of inferring minimal labels using common labels? In supervised tests where the (near) global minimum is known, we find that the percentages of minimal labels in $f^{[0]}$ is usually only slightly (about 1.0-2.0%) lower than those in $f^{[a]}$ and $f^{[b]}$. That is, the number of minimal labels retained in $f^{[0]}$ is about the same as those in $f^{[a]}$ and $f^{[b]}$. Given this and that $f^{[0]}$ is no longer a local minimum as $f^{[a]}$ and $f^{[b]}$, there is room to improve $f^{[0]}$ using a subsequent local minimization. This makes it possible to yield a better local minimum from $f^{[0]}$.

There are two parameters in the Comb algorithm, N and τ. $N = 10$ seems a good choice as the size of F for producing good quality solutions. Roughly speaking, the solution quality increases, *i.e.* the minimized energy value decreases, as N increases from 2 to 10. After that, the solution quality remains about the same (probabilistically) for greater N values. Empirically, when $\tau = 0$, all $f \in F$ converge sooner or later to a unique configuration and a smaller N makes such convergence quicker. The value of 0.01 is empirically a good choice for τ.

3 Relationship with GA

The Comb algorithm corresponds to a hybrid GA. Fig.4 describes a GA-like version of the Comb. The standard Comb initialization is effectively the same as a crossover operation followed by a mutation operation, the major and minor operations in genetic algorithms (GA) [6]. More exactly,

- the basic Comb corresponds to uniform crossover and
- the probability acceptance in the standard Comb corresponds to mutation.

In GA, two offspring, $f_i^{[01]}$ and $f_i^{[02]}$, are produced as the result of crossover. In the uniform crossover, either of the following two settings are accepted with equal probability:

(i) $f^{[01]} = f_i^{[a]}$ and $f^{[02]} = f_i^{[b]}$,
(ii) $f^{[01]} = f_i^{[b]}$ and $f^{[02]} = f_i^{[a]}$.

```
GA_comb_initialization(f^[a], f^[b], F)
begin
        for each i ∈ S
            /* uniform crossover */
            if (f_i^[a] == f_i^[b])
                then f_i^[01] = f_i^[02] = f_i^[a];
            else if (rand[0,1] < 0.5)
                f_i^[01] = f_i^[a] and f_i^[02] = f_i^[b];
            else
                f_i^[01] = f_i^[b] and f_i^[02] = f_i^[a];

            /* mutation */
            if (rand[0,1] < τ)
                f_i^[01] = rand(L);
            if (rand[0,1] < τ)
                f_i^[02] = rand(L);
end
```

```
GA_Comb_Algorithm
begin
        initialize F = {f^[1], ..., f^[N]};
        do {
            random_selection(f^[a], f^[b], F);
            GA_comb_initialization(f^[01], f^[02], f^[a], f^[b]);
            steepest_descent(f^*1, f^[01]);
            steepest_descent(f^*2, f^[02]);
            update(F, f^*1, f^*2);
        } until (termination condition satisfied);
        return(arg min_{f∈F} E(f));
end
```

Fig. 4. A GA-like Comb algorithm.

So, if $f_i^{[a]} = f_i^{[b]}$, there must be $f_i^{[01]} = f_i^{[02]} = f_i^{[a]}$, just as in the Comb initialization. In effect, the crossover encourages common labels because common labels are copied to the new initial configurations while non-common labels are subject to the swap. Although this about the *uniform* crossover, it should be noted that even the simplest *one-point* crossover also works in a way that encourages common labels.

The discussion suggests that the essence of both the Comb and GA is captured by the use of the common structure of local minima. This is supported by the fact that the original Comb and the GA-like Comb yield comparable results: In the GA-like Comb algorithm (Fig.4), when $f_i^{[a]} \neq f_i^{[b]}$, $f_i^{[01]}$ and $f_i^{[02]}$ inherit the values of $f_i^{[a]}$ and $f_i^{[b]}$, as does a crossover operator. However, setting $f_i^{[01]}$ and $f_i^{[02]}$ to a random label $rand(\mathcal{L})$, *i.e.* not necessarily inheriting $f_i^{[a]}$ and $f_i^{[b]}$, leads to comparable results, as long as the common labels are copied to $f_i^{[0]}$ when $f_i^{[a]} = f_i^{[b]}$. Moreover, whether to derive one initial configuration $f^{[0]}$ or two initial configurations $f^{[01]}$ and $f^{[02]}$ makes little difference; both schemes yield comparable results. In summary, the Comb and the GA-like Comb produce comparable results; and this suggests that retaining common labels is important and provides an interpretation for the crossover operation in GA.

The Comb produces much better results than the multi-start method. The former uses the best local minima to get an initial estimate whereas the latter uses but random initialization. Therefore, Comb has a much higher efficiency of descending to good local minima.

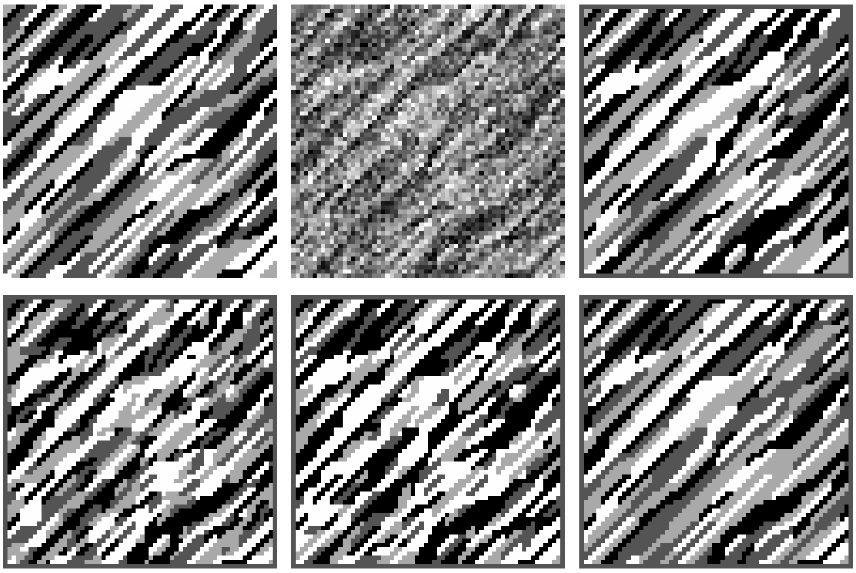
4 Experimental Results

In the following, we present two experiments, one for MAP-MRF image restoration and the other for segmentation, to compare the performance of the following algorithms: (1) the Comb; (2) the ICM [2]; (3) the HCF (the parallel version) [4]; (4) the SA with the Metropolis sampler [8]. For the Comb, the parameters are $N = 10$ and $\tau = 0.01$. The implementation of SA is based on a procedure given in [12]. The schedules for the SA are set to $T^{(t+1)} \leftarrow 0.999T^{(t)}$ with $T^{(0)} = 10^4$. The initial configuration for the ICM, HCF and SA is taken as the ML estimate whereas those in F for the Comb is entirely random; the difference will be explained later. The termination condition for the Comb is that all configurations in F are the same or that a number of 10000 new local minima have been generated.

Image	σ	α_I	β_1	β_2	β_3	β_4
No.1	1	0	-1	-1	-1	1
No.2	1	0	-2	-2	1	1
No.3	1	0	1	1	1	1

Table 2. The MRF parameters (α and β) and noise parameter σ for generating the three images.

The first set of experiments is for MAP-MRF restoration performed on three synthetic images of $M = 4$ gray levels, shown in Figs.5-7. The original have the label set $\mathcal{L} = \{1, 2, 3, 4\}$, and the pixel gray values also in $\{1, 2, 3, 4\}$. Table 2 gives the clique potential parameters α_I and $\beta_1, \cdots, \beta_4$ for generating the three types of textures and the standard deviation σ of the Gaussian noise.



a	b	c
d	e	f

Fig. 5. Restoration of image 1. (a) Original clean image. (b) Observed noisy image (input data). (c) Comb solution. (d) ICM solution. (e) HCF solution. (f) SA solution.

The second experiment compares the algorithms in performing MAP-MRF segmentation on the Lenna image of size 256×240 into a tri-level segmentation map. The results are illustrated in Fig.8. The input image is the original Lenna image corrupted by the i.i.d. Gaussian noise with standard deviation 10. The observation model is assumed to be the Gaussian distribution superimposed on the mean values of 40, 125 and 200 for the three-level segmentation. An isometric MRF prior is used with the four β parameters being $(-1, -1, -1, -1)$.

Table 3 compares the quality of restoration and segmentation solutions in terms of the minimized energy values. We can see that the Comb outperforms the ICM and the HCF and is comparable to SA. A subjective evaluation of the resultant images would also agree to the objective numerical comparison. The quality of the Comb solutions is generally also better than that produced by using the continuous method of augmented Lagrange-Hopfield developed in [10].

The Comb as a random search method needs many iterations to converge, the number being increasing as τ decreasing. All the Comb solutions with $\tau = 0.01$ are obtained when the limit of generating 10000 local minima is reached. This is about 1000 times more than the fast converging ICM and HCF. Nonetheless, the Comb takes about 1/20 of the computational effort needed by the SA.

The Comb algorithm does not rely on initial configurations at the very beginning of the algorithm to achieve better solutions. The ML initialization has

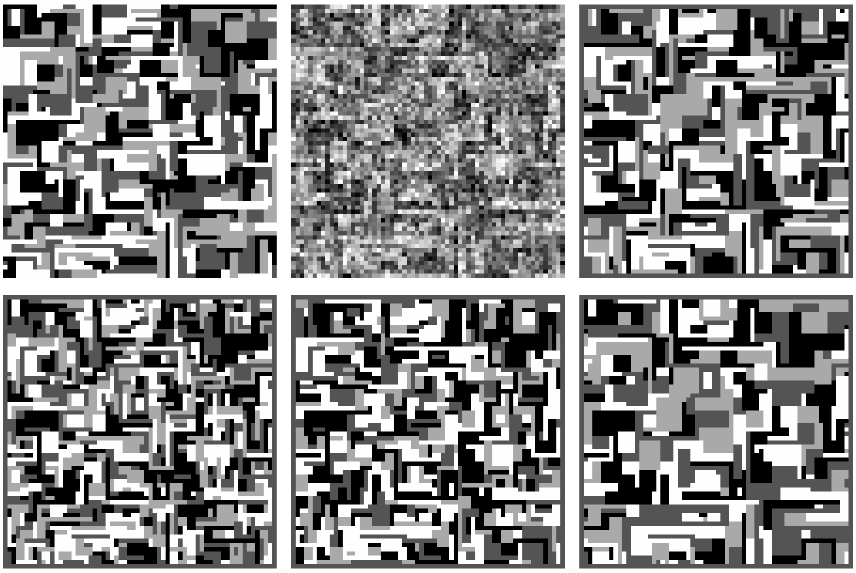


Fig. 6. Restoration of image 2. Legends same as Fig.5.

a significantly effect on improving the ICM, HCF and SA results. But for the Comb, the ML and random initializations make little difference. This insensitivity is probably due to that the Comb operates on multiple, rather than a single, configurations.

	Comb	ICM	HCF	SA
No.1	-12057	-10003	-10269	-11988
No.2	-10944	-8675	-9650	-11396
No.3	-27511	-25881	-26629	-27526
Lenna	-175647	-171806	-167167	-173301

Table 3. The minimized energy values for the restoration of images 1–3 and the segmentation of the Lenna image.

5 Conclusions

The Comb attempts to derive good initial configurations from the best local minima found so far in order to achieve better solutions. To do so, it uses the common structure of the local minima to infer the label values in the global

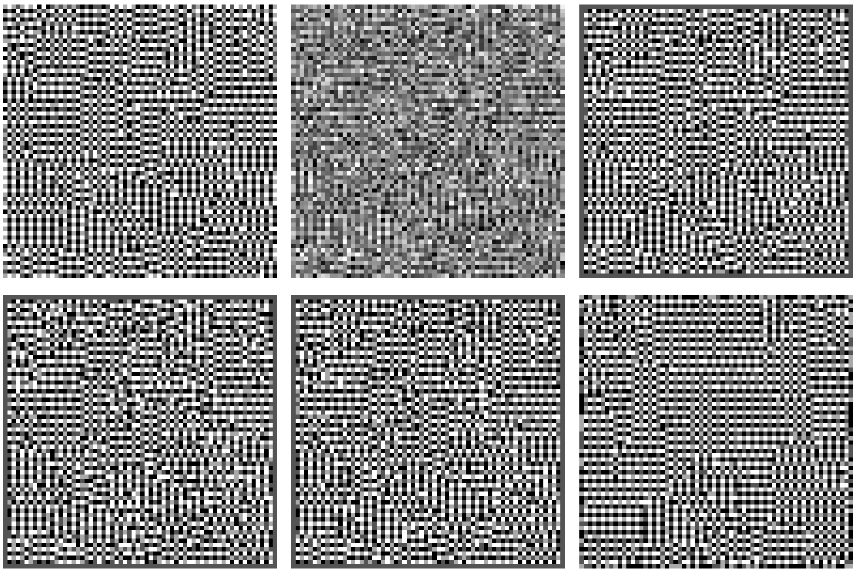


Fig. 7. Restoration of image 3. Legends same as Fig.5.

minimum. An initial configuration thus derived has about the same number of minimal labels as the two local minima from which it is derived. However, it is no longer a local minimum and thus its quality can be improved by a subsequent local minimization. This makes it possible to yield a better local minimum and thus increments the solution quality step by step. The comparisons shows that the Comb produces better results than the ICM and HCF though at higher computational costs and comparable results to the SA at lower costs. This suggests that the Comb can provide a good alternative to the well-known global minimizer SA.

References

1. J. Besag. "Spatial interaction and the statistical analysis of lattice systems" (with discussions). *Journal of the Royal Statistical Society, Series B*, 36:192–236, 1974.
2. J. Besag. "On the statistical analysis of dirty pictures" (with discussions). *Journal of the Royal Statistical Society, Series B*, 48:259–302, 1986.
3. B. Bhanu, S. Lee, and J. Ming. "Adaptive image segmentation using a genetic algorithm". *IEEE Transactions on Systems, Man and Cybernetics*, 25:1543–1567, 1995.
4. P. B. Chou, P. R. Cooper, M. J. Swain, C. M. Brown, and L. E. Wixson. "Probabilistic network inference for cooperative high and low level vision". In R. Chellappa and A. Jain, editors, *Markov Random Fields: Theory and Applications*, pages 211–243, Boston, 1993. Academic Press.

a	b	c
d	e	f

Fig. 8. Segmentation of the Lenna image. (a) The noisy Lenna image. (b) ML configuration. (c) Comb solution. (d) ICM solution. (e) HCF solution. (f) SA solution.

5. S. Geman and D. Geman. "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, November 1984.
6. D. E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, 1989.
7. Y. Huang, K. P. abd X. Zhuang, and J. Cavanaugh. "Optic flow field segmentation and motion estimation using a robust genetic partitioning algorithm". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:1177–1190, 1995.
8. S. Kirkpatrick, C. D. Gellatt, and M. P. Vecchi. "Optimization by simulated annealing". *Science*, 220:671–680, 1983.
9. S. Z. Li. *Markov Random Field Modeling in Computer Vision*. Springer-Verlag, New York, 1995.
10. S. Z. Li. "MAP image restoration and segmentation by constrained optimization". *IEEE Transactions on Image Processing*, page accepted for publication, 1997.
11. P. Moscato. "On evolution, search, optimization, genetic algorithms and martial arts: Towards memetic algorithms". C3P Report 826, Caltech Concurrent Computation Program, 1989.
12. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in C*. Cambridge University Press, 2 edition, 1988.
13. N. J. Radcliffe and P. D. Surry. Formal memetic algorithms. In T. Fogarty, editor, *Evolutionary Computing: AISB Workshop*, Lecture Notes in Computer Science, pages 1–14. Springer-Verlag, 1994.
14. G. Roth and M. D. Levine. "Geometric primitive extraction using a genetic algorithm". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16:901–905, 1994.