ELSEVIER

Image and Vision Computing 17 (1999) 37–49

image AND vision COMPUTING

Efficient hybrid search for visual reconstruction problems

Shang-Hong Lai[a], Baba C. Vemuri[b,*]

[a]*Siemens Corporate Research, Princeton, NJ 08540, USA*
[b]*Department of Computer and Information Sciences, University of Florida, Gainesville, FL 326117 USA*

Received 11 March 1997; received in revised form 30 January 1998; accepted 14 February 1998

Abstract

Visual reconstruction refers to extracting stable descriptions from visual data ([1]; A. Blake and A. Zisserman, Visual Reconstruction, MIT Press, Cambridge, MA, 1987). Visual reconstruction problems are commonly formulated in an optimization framework and normally require the optimization of nonconvex functions especially when discontinuity preserving image/shape recovery is the goal. Example problems include, image restoration, surface reconstruction, shape from shading etc. Most existing deterministic methods fail to reach the global optimum and lack the generality to incorporate reasonably complex interactions between boolean valued line process variables used for representing the presence (absence) of discontinuities. Stochastic methods for solving such problems e.g. the simulated annealing algorithm or variants thereof do achieve a global optimum but are plagued by slow convergence rates.

In this paper, we present a new hybrid search algorithm as an efficient solution for achieving the global optimum of the nonconvex function derived from a Markov random field formulation which allows for incorporation of complex interactions between the line process variables to better constraint the line processes. In the hybrid search, for the stochastic part, we develop an informed genetic algorithm (GA) while employing an incomplete Cholesky preconditioned conjugate gradient algorithm ([23]; S.H. Lai and B.C. Vemuri, Robust and efficient algorithms for optical flow computation, in: Proceedings of the International Symposium on Computer Vision, Coral Gables, FL, 1995, pp. 455–460) for the deterministic part. Our informed GA consists of a reproduction operator and an informed mutation operator. The informed mutation operator exploits specific domain knowledge in the search and is accomplished by the Gibbs sampler. Our hybrid search algorithm is highly parallelizable and leads to a globally optimal solution. The performance of our algorithm is demonstrated via experimental results on the sparse data surface reconstruction and the image restoration problem.

Keywords: Visual reconstruction; Hybrid search algorithm; Non-convex optimization; Informed genetic algorithm

1. Introduction

Visual reconstruction refers to generating stable descriptions from visual data [1]. In their seminal work, Blake and Zisserman [1] define, a stable description as any description that is invariant to a class of distortions that characterize the image formation process e.g., optical blurring, sensor noise, rotation and translation in the image plane etc. Example tasks of visual reconstruction include image restoration, surface reconstruction, optical flow computation, shape from shading, stereo matching, the lightness problem [2–8] and others.

Visual reconstruction problems may be formulated either in a deterministic or a probabilistic framework. A popular technique in the deterministic framework is based on the regularization theory [6,9,10] and leads to the minimization of energy functionals. In the probabilistic formulation, a Markov random field (MRF) model is used to characterize the function being estimated and the prior distribution of this MRF model can be identified with the smoothness constraints in the regularization formulation. Both the regularization formulation and the MRF formulation of visual reconstruction problems lead to the minimization of equivalent energy functions [6,10].

Discontinuities in images contain very crucial information for deriving high level image representations. Thus, numerous algorithms have been developed for discontinuity preserving visual reconstruction and there is abundant literature. Early work on visual reconstruction problems reported in literature ignores discontinuities altogether [2,4,11], which yields undesirable smoothing over discontinuity locations. However, later problem formulations allowed for incorporation of prespecified discontinuities [12–15]. An elegant way to treat discontinuities is via the introduction of a set of binary variables called the line

* Corresponding author. Tel: 352-392-1239; Fax: 352-392-1220; e-mail: vemuri@cise.ufl.edu

0262-8856/99/$ - see front matter
PII: S0262-8856(98)00088-2

processes. Each line process variable takes a value of one in the presence of discontinuities and zero otherwise. In the past, line processes have been included in both the deterministic [1,16] and the MRF [17] formulations. The resulting energy function involves real variables (for surface or image) coupled with binary variables (for discontinuities) and is therefore a nonconvex function. Both formulations lead to problems involving the minimization of nonconvex functions.

Numerous algorithms have been proposed to solve this coupled nonconvex minimization problem. They can be categorized into the deterministic methods [1,18,19] and the stochastic methods [17,20]. Most of the deterministic methods can only deal with simple constraints on the line process [1,19,21]; in addition, they can not find the global minimum solution. Stochastic optimization methods for achieving the global optimum solution consist primarily of the simulated annealing (SA) type algorithms. Simulated annealing is a general global optimization algorithm with a slow convergence rate making it impractical for many applications involving large size problems.

The deterministic methods for solving the coupled nonconvex minimization in the visual reconstruction problems have attracted more attention than the stochastic methods, since they usually converge faster and are more practical to use. However, most deterministic methods are restricted to handling only simple constraints on the line processes. For example, the well-known GNC (graduated nonconvex) algorithm proposed by Blake and Zisserman [1] is applicable only for the case when there is no interaction between the individual line process variables, i.e. no smoothness constraint is imposed on the line process. Recently, Bedini et al. [22] generalized the GNC algorithm to allow simple interactions between line process variables. The new closed form line energy contains terms that reduce the probability of occurrence of discontinuities which are parallel and neighboring to each other. However, it is very difficult to generalize the GNC algorithm to include arbitrary line energies. This is because, in general, the procedure for deriving the dual energy to eliminate the line processes in the GNC algorithm will be extremely complicated. Similarly, both the mean field annealing algorithm [19] and the analog network approaches [18,21] can only deal with specific types of constraints on the line process, for e.g. the energy term for constraining the line processes must have a simple closed-form expression. Unfortunately, it is usually very difficult to obtain a simple closed-form expression for the energy of the line process constraints in an MRF model. Therefore, these deterministic methods are in general not applicable for solving the nonconvex minimization problem with a general MRF formulation as described in this paper.

In this paper, we propose a novel hybrid search algorithm as a solution for the general coupled nonconvex minimization problem with an MRF formulation which allows for incorporation of complex interactions between the line process variables to better constraint the line processes. This hybrid search algorithm is a combination of a stochastic and a deterministic search technique. For the stochastic search, we develop a new informed genetic algorithm (GA) whereas for the deterministic search, we employ an incomplete Cholesky preconditioned conjugate gradient algorithm [23]. Our informed GA consists of a reproduction operator and an informed mutation operator which exploits specific domain knowledge in the search and is accomplished by the Gibbs sampler. We present promising experimental results of applying this algorithm to sparse (synthetic) data surface reconstruction with discontinuity detection and the image restoration problems.

The rest of the paper is organized as follows. We briefly review the MRF formulation for the visual reconstruction problems and the resulting nonconvex optimization problems in the next section. A new hybrid search algorithm which comprises of a stochastic search algorithm for the line processes and a deterministic algorithm for each smaller problem (given the line process) is proposed in Section 3. In Section 4, we present a new informed GA as the stochastic search for the line process in our hybrid search algorithm. In Section 5, we briefly present the deterministic algorithm used for solving the quadratic optimization obtained for prescribed discontinuities. Section 6 contains the implementation results for the sparse data surface reconstruction and image restoration problems using the proposed hybrid search algorithm. Finally, we conclude in Section 7.

2. Markov Random Field Formulation

In this section, we will briefly present the probabilistic formulation of a general visual reconstruction problem and use it for solving two example tasks namely, the discontinuity preserving surface reconstruction and the image restoration. The formulation leads to non-convex optimization. We choose the probabilistic formulation due to its many advantages over the deterministic counterpart. For instance, a probabilistic model can provide second (covariance) or higher order statistics of the estimate in addition to the maximum a posteriori estimate. The higher order statistical information is very useful in the context of dynamic vision [24], sensor fusion, and integration of multiple modules [20].

A Markov Random Field (MRF) [17] is a generalization of a Markov process from 1D to 2D. Let S be a finite set of $n \times n$ sites, and $G = \{G_{ij},\ (i,j) \in S\}$ be a neighborhood system [17] for S, then a random field $\mathbf{F} = \{\mathbf{F}_{\mathbf{i},\mathbf{j}}, (i,j) \in S\}$ is an MRF with respect to G if and only if the probability distribution is a Gibbs distribution with respect to G [17], i.e.,

$$P(\mathbf{F} = \mathbf{f}) = \frac{1}{Z} e^{-U(\mathbf{f})/T}, \tag{1}$$

where $\mathbf{f} = (f_{1,1},\ldots,f_{n,n})$ is a configuration of $\mathbf{F}$, Z is the normalizing constant, T is a parameter, and the energy

function $U(\mathbf{f})$ is of the form

$$U(\mathbf{f}) = \sum_{C \in C} V_C(\mathbf{f}), \tag{2}$$

The subset $C \subseteq S$ denotes a clique for the neighborhood system and C is the set of all cliques. The function $V_C(\mathbf{f})$ is called the clique potential. Its function value is determined by $\mathbf{f}_{i,j}$ with $(i,j) \in C$.

For the sake of exposition, let's consider the surface reconstruction and the image restoration problems here, but the discussion to follow should apply to most visual reconstruction problems with minor modifications. A coupled MRF model may be constructed for the surface (image) reconstruction (restoration) problem by defining an MRF, **F**, for surface (image) and a boolean MRF, **L**, for the line processes. The prior probability can be written as

$$P(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l}) = \frac{1}{Z} \exp^{-U(\mathbf{f},\mathbf{l})/T}, \tag{3}$$

$$U(\mathbf{f}, \mathbf{l}) = \sum_{C} V_C(\mathbf{f}, \mathbf{l}). \tag{4}$$

The parameter T is assumed to be 1 for simplicity. The prior energy $U(\mathbf{f},\mathbf{l})$ is a smoothness constraint on the surface (image) and the line process.

Let the observed data be denoted by **d**, then the problem to estimate **f** and **l** from **d** can be obtained by the MAP (maximum a posteriori) estimation. Assume the observation model is given by $\mathbf{D} = H(\mathbf{F}) + \mathbf{N}$, where **N** is zero mean white Gaussian random field with variance σ for each variable in **N**. After applying the Bayes rule, we can see that the MAP estimation is equivalent to maximizing $P(\mathbf{D} = \mathbf{d}|\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l})P(\mathbf{F} = \mathbf{f},\mathbf{L} = \mathbf{l})$, and the MAP estimate can be obtained by minimizing the following energy function

$$U(\mathbf{f}, \mathbf{l}|\mathbf{d}) = \frac{1}{2\sigma^2}\|\mathbf{d} - H(\mathbf{f})\|^2 + U(\mathbf{f}, \mathbf{l}) = \frac{1}{2\sigma^2}\|\mathbf{d} - H(\mathbf{f})\|^2 + \sum_{C} V_C(\mathbf{f}|\mathbf{l}) + \sum_{C} V_C(\mathbf{l}). \tag{5}$$

It is possible for both the MRF and deterministic formulations of the above problem to yield the same solution via an appropriate choice of the prior in the MRF formulation [25]. The above energy is a nonconvex function and is a hybrid of real and boolean variables.

All the deterministic optimization methods in literature [26] can only be applied to some restricted energy functions whose line energies contain either no line process interactions or very simple line process interactions. The line energies encode the constraints or prior information we impose on the line process. Bedini et al. [22] used a line energy with simple line process interactions and obtained much better results than those with the simplest line energy, e.g. no interactions between line processes. In an MRF model, the line energy is the summation of line clique potentials. When the neighborhood system in the MRF model is nonempty, there are line process interactions in the line energy. In general, the line clique potentials are accessed through a look-up table and no closed-form expression is available. Thus, the line clique potentials are very crucial to the performance of the MRF-based algorithms [27]. They are regarded as the parameters in an MRF model and can be estimated for different classes of objects [27] to achieve better performance.

Furthermore, most of the deterministic methods with the exception of dynamic programming, can only reach an approximate solution or a locally optimal solution. In fact, the quantization of the real variables in dynamic programming may also be regarded as an approximation. Unlike the deterministic optimization methods, there is no restriction on the clique potentials for the stochastic methods. In addition, simulated annealing as well as the genetic algorithm can converge to a global optimum. The drawback of stochastic optimization methods is that, their computational cost is very high, compared to most of the deterministic methods. However, we will present a novel hybrid search algorithm that has yielded promising and practically feasible results.

3. Proposed solution

In this section, we present the outline of a novel hybrid (stochastic + deterministic) search algorithm as an efficient solution to the coupled (binary-real) nonconvex optimization problem. This hybrid search algorithm consists of an informed genetic algorithm (GA) and an incomplete Cholesky preconditioned conjugate gradient algorithm [23]. The informed genetic algorithm is employed as a global minimizer on the binary line process. Within the GA, for each given line process configuration, the incomplete Cholesky preconditioned conjugate gradient algorithm is used on the surface (image) variables to solve a convex and quadratic minimization problem.

As shown earlier, the probabilistic formulation leads to the minimization of a nonconvex energy function $U(\mathbf{f},\mathbf{l}|\mathbf{d})$, given in Eq. (5). This energy function is a hybrid of the real variables and boolean variables. We know that the GA is well suited for combinatorial optimization problems because of its bit-string representation. This nonconvex minimization problem can be transformed to a combinatorial optimization problem as follows.

The energy function $U(\mathbf{f},\mathbf{l}|\mathbf{d})$ can be rewritten as

$$U(\mathbf{f}, \mathbf{l}|\mathbf{d}) = U(\mathbf{f}|\mathbf{l}, \mathbf{d}) + U(\mathbf{l}), \tag{6}$$

$$U(\mathbf{f}|\mathbf{l}, \mathbf{d}) = \frac{1}{2\sigma^2}\|\mathbf{K}_d\mathbf{f} - \mathbf{d}\|^2 + U(\mathbf{f}|\mathbf{l}), \tag{7}$$

$$U(\mathbf{f}|\mathbf{l}) = \sum_{C} V_C(\mathbf{f}|\mathbf{l}), \tag{8}$$

$$U(\mathbf{l}) = \sum_C V_C(\mathbf{l}). \tag{9}$$

In the above equations, we denote the observation model by $\mathbf{d} = \mathbf{K}_d\mathbf{f} + \mathbf{n}$, where $\mathbf{d} \in \Re^{n^2\times 1}$ is the data vector, the operator $\mathbf{K}_d$ is a matrix when the model is linear, and $\mathbf{n}$ is a zero mean white Gaussian random vector with variance σ for each variable in $\mathbf{n}$. For the four-connected neighborhood system, the conditional clique potential of the real variables $\mathbf{f}$ given the line process, $\mathbf{l}$, $U(\mathbf{f}|\mathbf{l})$, is analogous to the quadratic membrane smoothness energy, i.e.

$$U(\mathbf{f}|\mathbf{l}) = \sum_{i,j} ((f_{i+1,j} - f_{i,j})^2(1 - h_{i,j}) + (f_{i,j+1} - f_{i,j})^2(1 - v_{i,j})), \tag{10}$$

where $\mathbf{l} = [\mathbf{h}\ \mathbf{v}]$.

We can transform the minimization of $U(\mathbf{f},\mathbf{l}|\mathbf{d})$ with respect to $\mathbf{f},\mathbf{l}$ to the minimization of a function of the boolean variables $\mathbf{l}$ as follows.

$$\min_{\mathbf{f},\mathbf{l}} U(\mathbf{f},\mathbf{l}|\mathbf{d}) = \min_{\mathbf{l}}\left\{\min_{\mathbf{f}} U(\mathbf{f},\mathbf{l}|\mathbf{d})\right\} = \min_{\mathbf{l}}\left\{U(\mathbf{l}) + \min_{\mathbf{f}} U(\mathbf{f}|\mathbf{l},\mathbf{d})\right\}. \tag{11}$$

Define the function

$$E(\mathbf{l}) := U(\mathbf{l}) + \min_{\mathbf{f}} U(\mathbf{f}|\mathbf{l},\mathbf{d}). \tag{12}$$

Then the original nonconvex minimization problem is transformed to the combinatorial optimization problem, i.e. minimization of a function of boolean variables. Note that the function $U(\mathbf{f}|\mathbf{l},\mathbf{d})$ is a quadratic and convex function of $\mathbf{f}$ when the operator in the observation model, $\mathbf{K}_d$, is linear. In the surface reconstruction as well as the image restoration problems, one may assume a linear observation model [17,24].

Previously, hybrid algorithms [28,29] were proposed that used simulated annealing for the binary line process and an analog network to minimize the quadratic function $U(\mathbf{f}|\mathbf{l},\mathbf{d})$ for each fixed line process. In our hybrid search algorithm, we use an informed genetic algorithm to minimize this new energy function. To compute the new energy function value in the GA, we need to minimize the convex and quadratic function $U(\mathbf{f}|\mathbf{l},\mathbf{d})$ for each line process configuration $\mathbf{l}$. The minimum of a quadratic and convex function can be obtained by solving a linear system with an associated symmetric positive–definite (SPD) matrix. In this paper, we use the incomplete Cholesky preconditioned conjugate gradient algorithm [23] for solving the linear system. Thus, the overall structure of our algorithm is to use the informed GA as a global minimizer of the function $E(\mathbf{l})$ with a fast quadratic convex function minimizer inside the GA to obtain the function value of $E(\mathbf{l})$.

4. Informed genetic algorithm

The GA is an adaptive search algorithm based on principles derived from the dynamics of natural population genetics. It has been successfully applied to combinatorial optimization, function optimization, classifier systems, structure optimization, pattern recognition, and other areas [30,31].

A population of candidate solutions is generated in the beginning as the initial population, denoted by $P(0)$. Then, the GA operators [31], i.e. reproduction, crossover and mutation, are applied to the current population $P(t)$ to generate the descendant population $P(t+1)$. Subsequently, the algorithm is an iterative adaptation from one generation to the next based on the genetic operators and the fitness function. The GA operators are designed to follow the survival of fittest principle. The fitness function in the GA is directly related to the objective function to be minimized or maximized in an optimization problem.

Markov chain models have been employed to analyze the convergence of simple genetic algorithms in [32–34]. Davis [32] used a nonhomogeneous finite Markov chain to model the simple GA (population size is fixed but the mutation and crossover are allowed to vary with the iteration index) and proved that the convergence rate of a simple GA is far superior to that of simulated annealing. Rudolph [33] employed a homogeneous finite Markov chain to model a canonical GA (fixed population size, proportional selection, fixed crossover and mutation probabilities), and proved that a canonical GA never converges to the global optimum regardless of the initialization. But, he also showed that modified versions of canonical GAs with elitist selection (the best individual survives with probability one) asymptotically converges to the global optimum in the probabilistic sense. Because of the global convergence of the elitist selection proved in [33], this strategy is adopted in our informed GA.

In this section, we present a new informed GA which incorporates specific domain knowledge into the genetic operators. In traditional GAs [31–33], only the reproduction operator depends on the fitness function, while the crossover and mutation operators randomly explore the solution space with prespecified probabilities. Totally random search is a very inefficient way to explore the solution space, and it makes the GAs converge very slowly especially when the entire solution space is very large as in visual reconstruction problems. In our informed GA, we use an informed mutation operator which exploits specific domain knowledge in the search, thus making the solution search in the GA very efficient.

The only search procedure in simulated annealing involves the use of the Gibbs sampler [17,35] which is a random perturbation of the current solution and is similar to the mutation operator in the GA. The Gibbs sampler is used to search the solution space based on the Gibbs distribution of the energy function. The search in the Gibbs sampler makes use of the information in the energy function specific to the problem domain, therefore it is an informed search, not a totally random search. We use the Gibbs sampler as an informed mutation operator in our informed GA.

Our informed GA consists of two operators namely, the reproduction operator and an informed mutation operator. The main idea is to design the GA operators to make each configuration of candidate solutions in the generations of the GA to be a sample drawn from the Gibbs distributions of the energy function to be minimized. To achieve this goal, we choose our initial population as a sample from a Gibbs distribution of the energy function $E(\mathrm{l})$, i.e.

$$P_0(\mathbf{l}) = \frac{1}{Z} e^{-E(\mathbf{l})/T}, \tag{13}$$

$$Z = \sum_{\mathbf{l}} e^{-E(\mathbf{l})/T}, \tag{14}$$

where T is a constant. Let the population size be fixed and denoted by a constant M. By using a Gibbs sampler to obtain M samples from the Gibbs distribution given in Eq. (13), we obtain an initial population of size M, denoted by the set $\{\mathbf{l}_1,\mathbf{l}_2,\ldots,\mathbf{l}_M\}$.

In the reproduction operation, we use the Boltzmann weighted selection [36,37] scheme such that the survival probability for the ith individual is given by

$$P_r(\mathbf{l}_i) = \frac{1}{Z'} e^{-E(\mathbf{l}_i)/T}, \tag{15}$$

$$Z' = \sum_{j=1}^{M} e^{-E(\mathbf{l}_j)/T}. \tag{16}$$

For the informed mutation operator, we assign the probabilty of mutating the ith individual $\mathbf{l}_i$ to a line process configuration $\bar{\mathbf{l}}$ to be a Gibbs distribution as follows.

$$P_m(\mathbf{l}_i \rightarrow \bar{\mathbf{l}}) = \frac{1}{Z''(\mathbf{l}_i)} e^{-(E(\bar{\mathbf{l}}) - E(\mathbf{l}_i))/T}, \tag{17}$$

$$Z''(\mathbf{l}_i) = \sum_{\bar{\mathbf{l}}} e^{-(E(\bar{\mathbf{l}}) - E(\mathbf{l}_i))/T}. \tag{18}$$

This mutation transition probability can be simplified as

$$P_m(\mathbf{l}_i \rightarrow \bar{\mathbf{l}}) = \frac{1}{Z} e^{-E(\bar{\mathbf{l}})/T}. \tag{19}$$

The above informed mutation operation can be implemented by using a Gibbs sampler [17,35].

By using the above reproduction and mutation operators with the initial population sampled from the Gibbs distribution given in Eq. (13), we can show that each individual at the nth generation is a sample from the following Gibbs distribution

$$P_n(\mathbf{l}) = \frac{1}{\bar{Z}} e^{-(n+1)E(\mathbf{l})/T}, \tag{20}$$

$$\bar{Z} = \sum_{\mathbf{l}} e^{-(n+1)E(\mathbf{l})/T}. \tag{21}$$

Thus, the asymptotic convergence of this informed GA can be proved directly from Eq. (20) by assuming $n \rightarrow \infty$.

Our informed GA is summarized as follows:

1. Choose an initial population from a Gibbs distribution in Eq. (13);
2. Determine the fitness function value for each individual;
3. Perform a Boltzmann weighted selection;
4. Perform the informed mutation operation by using a Gibbs sampler;
5. Check stopping criterion, if satisfied stop, else go to step 2.

In step 2, we need to compute the fitness function value for each individual line process configuration. To compute the fitness function value $E(\mathbf{l})$ given in Eq. (12), it involves solving a convex and quadratic minimization problem. This is done very efficiently by using the incomplete Cholesky preconditioned conjugate gradient algorithm [23] which is presented in the following section.

5. Incomplete Cholesky preconditioned conjugate gradient

In our hybrid search algorithm, the informed GA is used as a stochastic search for the binary line process only. For each line process configuration $\mathbf{l}$ visited by the GA, we need to compute its objective function value $E(\mathbf{l})$, given in Eq. (12). This computation involves the minimization of a convex and quadratic function $U(\mathbf{f}|\mathbf{l},\mathbf{d})$, which can be obtained by solving a linear system with an associated symmetric positive–definite (SPD) matrix. We designed an incomplete Cholesky preconditioner for use with a conjugate gradient algorithm to solve the large and sparse SPD linear system very efficiently.

From Eqs. (7) and (10), we can obtain the associated SPD matrix $\mathbf{K}$ for the quadratic and convex function $U(\mathbf{f}|\mathbf{l},\mathbf{d})$ as follows.

$$\mathbf{K} = \frac{1}{2\sigma^2}\mathbf{K}_{\mathrm{d}} + \mathbf{K}_{\mathrm{s}}, \tag{22}$$

where $\mathbf{K}_{\mathrm{d}} \in \Re^{N\times N}(N = n^2)$ for example is a diagonal sampling matrix in the surface reconstruction problem and $\mathbf{K}_{\mathrm{s}} \in \Re^{N\times N}$ is the matrix from the quadratic smoothness energy, given in Eq. (10). The matrix $\mathbf{K}_{\mathrm{s}}$ can be formed from the computational molecules [12] at the locations with absence of discontinuities. An example of the matrix $\mathbf{K}_s$, with no discontinuities in the domain is given as follows:

$$\mathbf{K}_{\mathrm{s}} = \begin{bmatrix} \mathbf{D} & -\mathbf{I} & & & \boldsymbol{0} \\ -\mathbf{I} & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & -\mathbf{I} \\ \boldsymbol{0} & & & -\mathbf{I} & \mathbf{D} \end{bmatrix}, \tag{23}$$

$$\mathbf{D} = \begin{bmatrix} 4 & -1 & & & -1 \\ -1 & 4 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 4 & -1 \\ -1 & & & -1 & 4 \end{bmatrix} \in \Re^{n\times n}. \quad (24)$$

Note that the matrix $\mathbf{K}_s$ is a block tridiagonal matrix with the diagonal blocks being tridiagonal and sub-diagonal blocks being diagonal. For the surface reconstruction problem, the diagonal matrix $\mathbf{K}_d$ has 1*s* at the corresponding data location and 0 elsewhere. Therefore, $\mathbf{K}_d$ is an identity matrix for dense data case, while being a sparse diagonal matrix for the sparse data case. In the case of image restoration with additive uncorrelated Gaussian noises, $\mathbf{K}_d$ is simply the identity matrix.

To solve the above linear system, we use the preconditioned conjugate gradient algorithm [38] with an incomplete Cholesky preconditioner **P** [38,39]. The preconditioner **P** is chosen to be an incomplete Cholesky factorization of the matrix **K**. A good preconditioner can drastically accelerate the convergence rate of the conjugate gradient algorithm. There are two criteria for designing a good preconditioner **P** for the matrix **K**. First, the preconditioner **P** has to be a good approximation to **K** so that the condition number of the preconditioned linear system is dramatically reduced. Secondly, there must exist a very fast numerical method to solve the auxiliary linear system $\mathbf{Pz} = \mathbf{r}$ required in the preconditioned CG algorithm. Taking these two criteria into consideration, a good preconditioner for the above stiffness matrix **K** of the visual reconstruction problem can be obtained via the incomplete Cholesky factorization of **K**.

The standard Cholesky factorization of the sparse matrix **K** 'fills in' entries in the band between nonzero off-diagonals, which means the sparsity structure will be destroyed after the factorization. The idea of incomplete Cholesky factorization is to find an approximate Cholesky factorization of the matrix **K**, i.e. $\mathbf{K} \approx \mathbf{L}\mathbf{L}^T$, such that the lower triangular matrix **L** has the similar sparsity structure. In addition, the product $\mathbf{L}\mathbf{L}^T$ at the locations with nonzero entries in **L** or $\mathbf{L}^T$ still has the same values as those in **K**. Therefore, the preconditioner $\mathbf{P} = \mathbf{L}\mathbf{L}^T$ is a good approximation to **K**. Since the matrix **K** is sparse and well- structured, the matrix **L** is also sparse and well-structured. Thus, the solution to the auxiliary linear system $\mathbf{Pz} = \mathbf{r}$ in the preconditioning step of the preconditioned conjugate gradient algorithm can be obtained via forward and backward substitutions very efficiently.

The matrix **K** in Eq. (22) is a block tridiagonal matrix with the diagonal blocks being tridiagonal and the sub-diagonal blocks being diagonal. The factoring matrix in the incomplete Cholesky factorization has the following structure,

$$\mathbf{L} = \begin{bmatrix} \mathbf{G}^{(1)} & & & \\ \mathbf{H}^{(1)} & \mathbf{G}^{(2)} & & \\ & \ddots & \ddots & \\ & & \mathbf{H}^{(n-1)} & \mathbf{G}^{(n)} \end{bmatrix}, \quad (25)$$

where

$$\mathbf{G}^{(k)} = \begin{bmatrix} \alpha_1^{(k)} & & & \\ \beta_1^{(k)} & \alpha_2^{(k)} & & \\ & \ddots & \ddots & \\ & & \beta_{n-1}^{(k)} & \alpha_n^{(k)} \end{bmatrix} \in \Re^{n\times n}$$

$$H^{(k)} = \begin{bmatrix} \gamma_1^{(k)} & \delta_1^{(k)} & & \\ & \gamma_2^{(k)} & \ddots & \\ & & \ddots & \delta_{n-1}^{(k)} \\ & & & \gamma_n^{(k)} \end{bmatrix} \in \Re^{n\times n},$$

for $k = 1,\ldots,n$. The nonzero entries in the sparse matrix **L** are computed by equating the entries of the product $\mathbf{L}\mathbf{L}^T$ to those in the matrix **K** at the locations with nonzero entries in **L**. The incomplete Cholesky factorization of the matrix **K** [23] can be computed in $O(N)$ operations, where $N(=n^2)$ is the number of discretization points. After the factorization, the preconditioner **P** is chosen as $\mathbf{L}\mathbf{L}^T$, which is close to **K** and has a nice structure. Thus, the preconditioned conjugate gradient algorithm requires $O(N)$ operations in each iteration.

6. Experimental results

In this section, we present the experimental results of applying our hybrid search algorithm to the surface reconstruction and image restoration problems. In our experiments, our hybrid search algorithm consists of the informed GA for the binary line process **l** and the incomplete Cholesky preconditioned conjugate gradient algorithm for determining the fitness function value for a pre-specified line process configuration.

6.1. Surface reconstruction

We implemented our hybrid search algorithm on the sparse data surface reconstruction problem. The sparse data set (see Fig. 1(b)) is obtained from sampling the original surface shown in Fig. 1(a). We discretized this problem on a 32×32 mesh. The sparse data set shown in Fig. 1(b) contains 64 data points.

In our experiments, we use the cliques for the line process used by Geman and Geman [17], shown in Fig. 2(a). The line clique potentials for different line process configurations

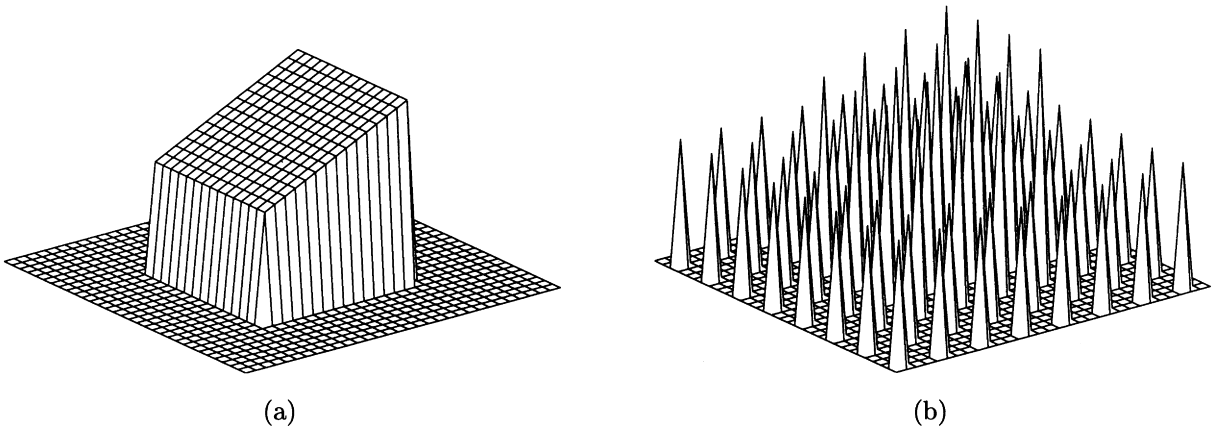

Fig. 1. The sparse data set used for the surface reconstruction with discontinuity detection; (a) original surface, and (b) sparsely sampled data set.

are shown in Fig. 2(b). This choice of the line clique potentials is ad hoc. However, they can be obtained via MRF model parameter estimation [27].

The Gibbs sampler [17,35] is used to implement the informed mutation operator in our informed GA. There are two steps in the Gibbs sampler, i.e. the exploration step (or perturbation step) and acceptance step. Each time after the exploration step, we need to compute the energy function value for the perturbed new line process configuration and then determine the acceptance probability P_{accept}, given by

$$P_{\text{accept}}(\mathbf{l}_{\text{old}} \rightarrow \mathbf{l}_{\text{new}}) = \begin{cases} 1 & \text{when } E(\mathbf{l}_{\text{new}} \leq E(\mathbf{l}_{\text{old}}), \\ e^{-(E(\mathbf{l}_{\text{new}}) - E(\mathbf{l}_{\text{old}}))/T} & \text{when } E(\mathbf{l}_{\text{new}}) > E(\mathbf{l}_{\text{old}}), \end{cases} \quad (26)$$

where $\mathbf{l}_{\text{old}}$ is the line process before the exploration step and $\mathbf{l}_{\text{new}}$ is the one after the exploration step. Computation of the energy function value $E(\mathbf{l})$ given $\mathbf{l}$ involves minimizing a quadratic convex function, which is solved very efficiently via the incomplete Cholesky preconditioned conjugate gradient algorithm. It is necessary to repeat these two steps many times for the Gibbs sampler to converge, therefore the computational cost is extremely high in implementing the Gibbs sampler. In our experiments, we use an approximate Gibbs sampler to reduce the computational cost dramatically. This Gibbs sampler uses the approximation that the minimum solution $\mathbf{u}_{\text{old}}$ obtained by minimizing the function $U(\mathbf{u}|\mathbf{l}_{\text{old}},\mathbf{d})$ is similar to the minimum solution $\mathbf{u}_{\text{new}}$ for the new perturbed line process By using this approximation, the acceptance probability can be obtained by computing the local energy change in E for each (one-site) perturbation.

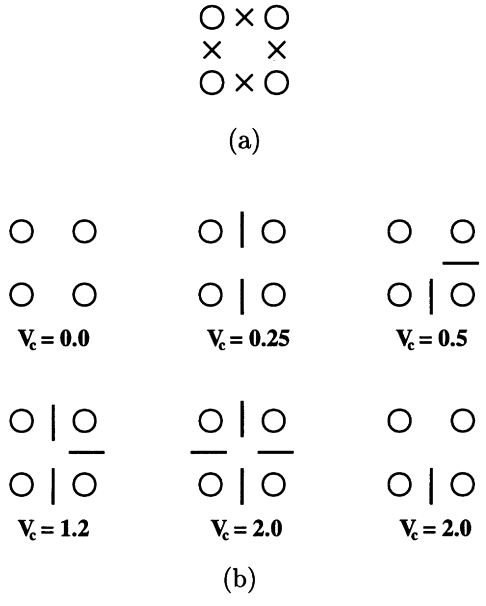

Fig. 2. (a) Clique for the line process (× , line process; ○, nodal variable for surface); and (b) clique potentials for different process configurations.

In our experiments, the population size in the informed GA is arbitrarily chosen to be 50. By using our hybrid search algorithm on the sparse data set shown in Fig. 1(b), the converged solution, which means the best visited solution remains the same for many generations, was found after 3 generations of the informed GA. Fig. 3(a) shows the reconstructed surface and the detected discontinuities are given in Fig. 3(b). We can see that this solution is very close to the true solution.

In another experiment, we added noise to the sparse data in Fig. 1(b) to test our hybrid search algorithm. The noise for each data point i is a uniform random variable ranging in the interval $[-0.2\mathbf{d}_i, 0.2\mathbf{d}_i]$, where $\mathbf{d}_i$ is the data value sampled from the original shape shown in Fig. 1(a). Fig. 4(a) shows the noisy sparse data set. Our hybrid search algorithm found the converged solution after five generations of the informed GA. The reconstructed surface and detected discontinuities are shown in Fig. 4(b) and (c), respectively. As evident from this experiment, our hybrid search algorithm has obtained a very accurate solution for sparse data surface reconstruction problems with high noise.

With regards to computational effort, in the above surface reconstruction experiments, our incomplete Cholesky preconditioned conjugate gradient algorithm usually takes less than 0.03 s on a multi-user 185 MHz SUN Ultra-Spare

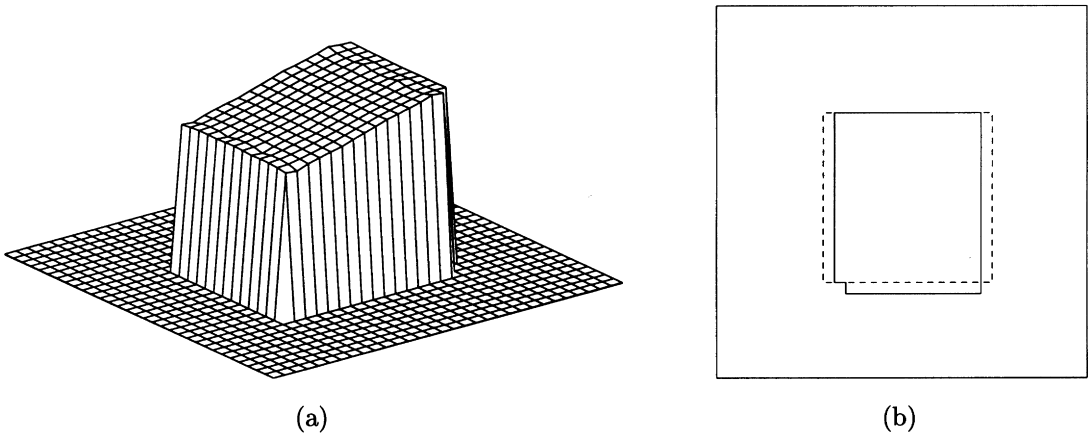

Fig. 3. Surface reconstruction example: (a) the reconstructed surface; and (b) the recovered discontinuity map (solid line) obtained after three generations of the informed GA and imposed on the true discontinuity map (dashed line).

workstation to solve a convex quadratic minimization problem. The population size was set to 50 in our hybrid search algorithm. In the experiments, three generations of the hybrid search algorithm were completed in less than 4.7 s. For 5 generations of the same, it took less than 7.5 s. The structure of our hybrid search algorithm is primarily a genetic algorithm, which is easily parallelizable [40]. A parallel implementation will make our algorithm a very useful and efficient technique for solving hard nonconvex optimization problems such as the one described here.

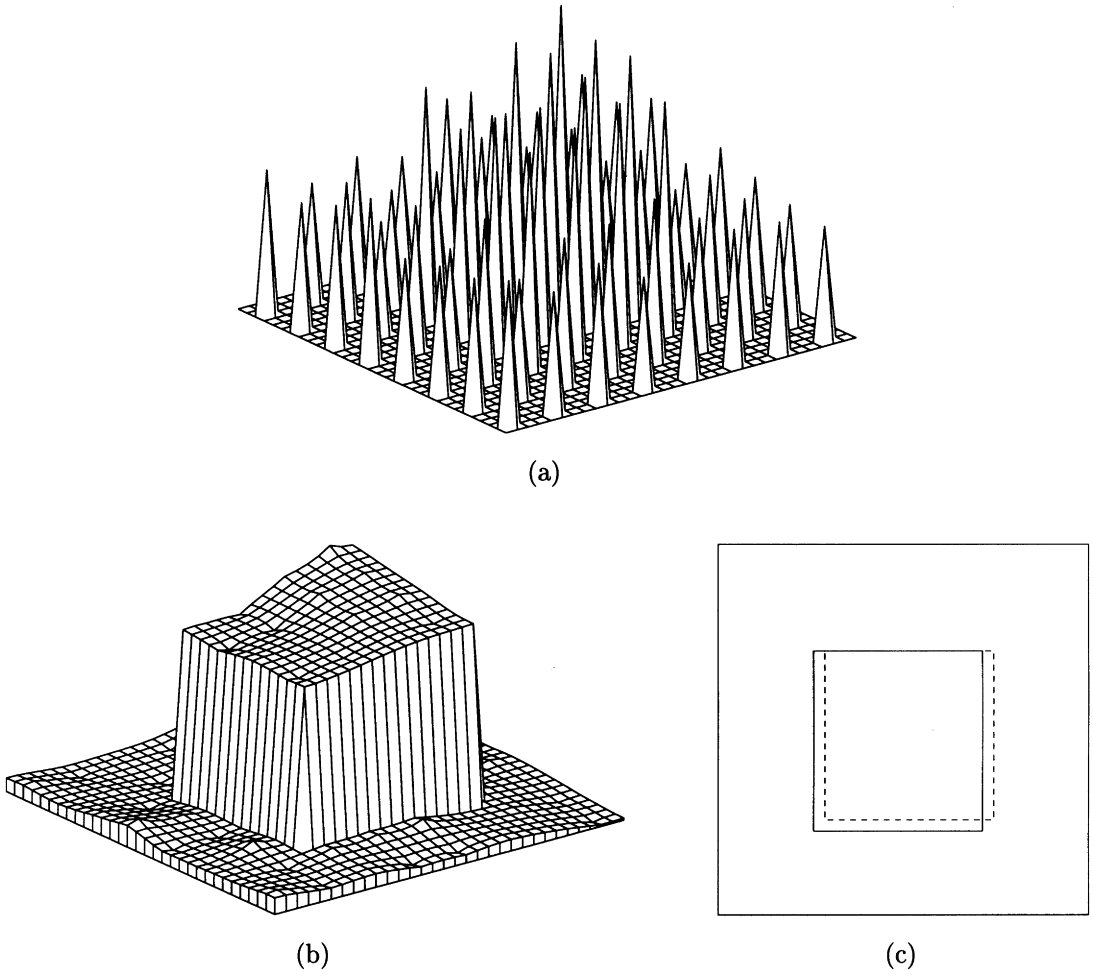

Fig. 4. Surface reconstruction with noise example: (a) the noisy sparse data set; (b) the reconstucted surface; and (c) the recover discontinuity map (solid line) obtained after five generations of the informed GA and imposed in the true discontinuity map (dashed line).

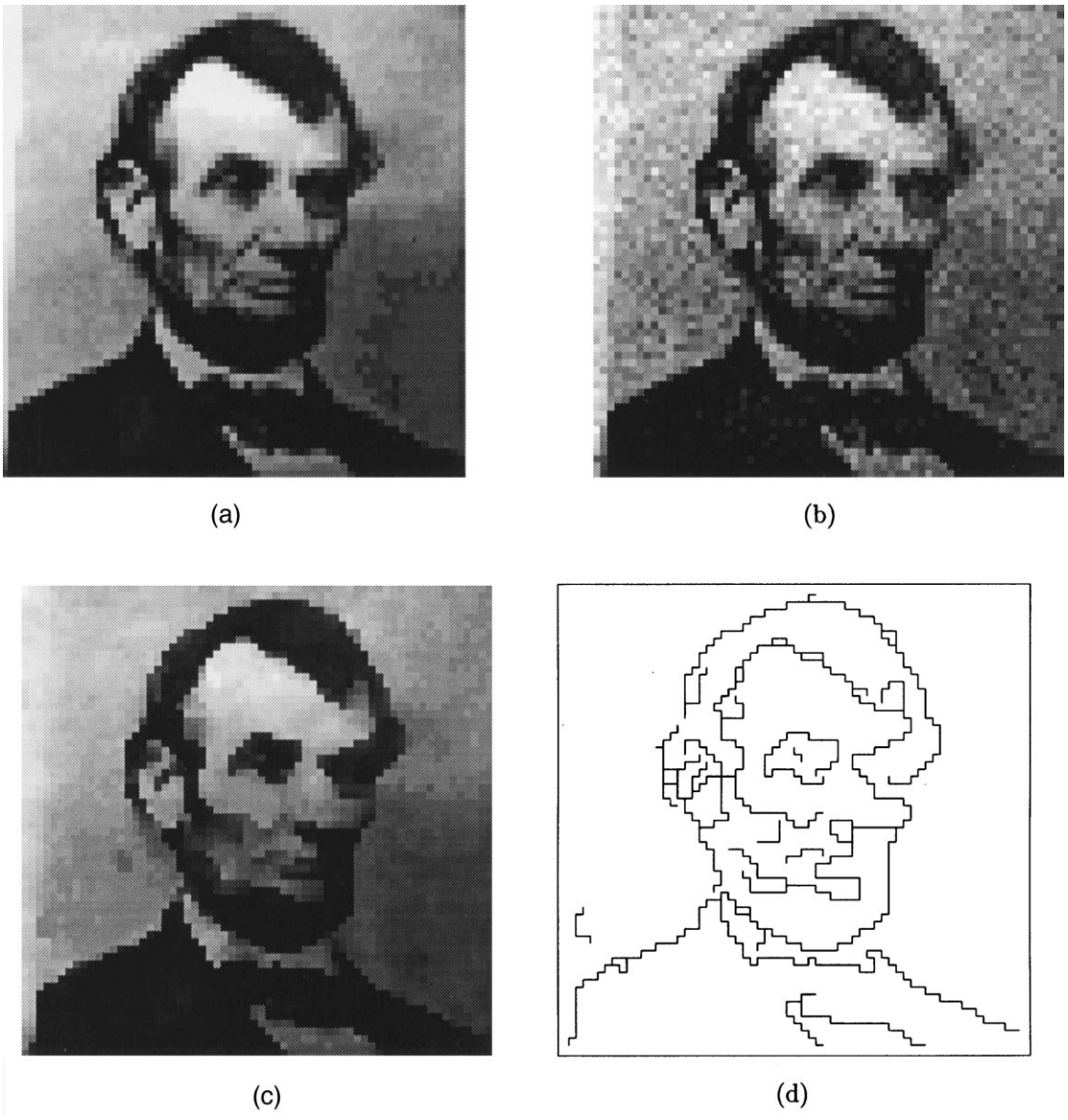

Fig. 5. (a) The original Lincoln image; (b) the blurred Lincoln image with additive uncorrelated Gaussian noise (SNR = 19 db); (c) the restored image; and (d) the recovered discontinuity map after six generations of our informed GA.

6.2. Image restoration

For the image restoration experiments, the population size is chosen to be 25. A second-order neighborhood system in the MRF model is used to better constrain the line process configuration. In the first image restoration example, we apply our hybrid search algorithm on the 64 $\times$ 64 blurred Lincoln image with the additive uncorrelated Gaussian noise. The original Lincoln image and the noise-corrupted image with SNR 19 db are shown in Fig. 5(a) and (b), respectively. In this experiment, the population size of our informed GA was chosen to be 25. For this example, the incomplete Cholesky preconditioned conjugate gradient algorithm converges in about four iterations for each convex and quadratic minimization problem. The restored image and the recovered discontinuity map after six generations of the informed GA are shown in Fig. 5(c) and (d), respectively. We can see that the noise was greatly reduced and the edges were preserved in the restored image. For six generations the informed GA, it took about 12 s on a multi-user 185 MHz SUN Ultra-Sparc workstation.

In the next image restoration experiment, we increased the variance of the Gaussian noise added to the Lincoln image. This noise-corrupted Lincoln image (with SNR 15 db) is shown in Fig. 6(a). The population size of the GA is also chosen to be 25. After six generations of our informed GA, we obtained the restored image and the recovered discontinuity map as shown in Fig. 6(b) and (c). The execution time for six generations of our algorithm on this example was about 12 seconds on a 185 MHz SUN Ultra-Sparc workstation.

The convergence curves for our informed GA on the

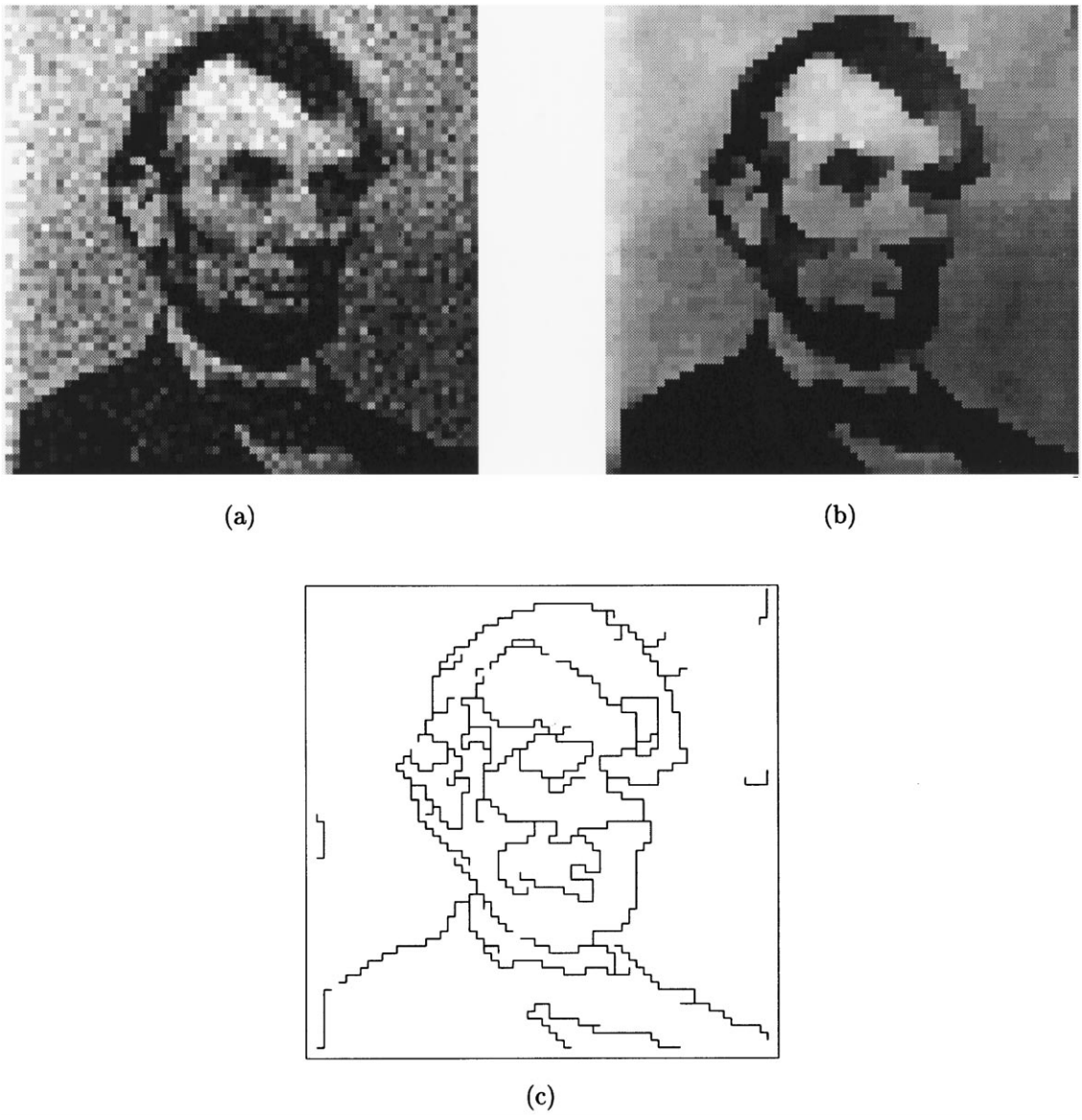

Fig. 6. (a) The blurred Lincoln image with white Gaussian noise (SNR = 15 db); (b) the restored image; and (c) the restored discontinuity map after six generations of our informed GA.

above two image restoration experiments with SNR being 19 and 15 db are depicted in Fig. 7(a) and (b), respectively. These convergence curves are the energy function values versus the number of generations in the informed GA. As shown in the figure, it is evident that our algorithm converges very fast and reasonably good solutions can be obtained in just a few generations.

The third restoration example is on a larger image of size (128 × 128). The image consists of a mechanical part shown in Fig. 8(a). The image corrupted by additive Gaussian noise (SNR = 20 db) is shown in Fig. 8(b). The population size is set to 25 as before. After 6 generations of our informed GA, the restored image and the recovered discontinuity map are shown in the Fig. 8(c) and (d), respectively. In this experiment, each generation requires solving 25 linear systems with matrices of size 16384×16384 ($= 128^2 \times 128^2$), which took less than 9 seconds execution time by using the incomplete Cholesky preconditioned conjugate gradient algorithm on a 185 MHz SUN Ultra-Sparc workstation. In this example, 6 generations of our informed GA was accomplished within 57 s.

7. Discussions and conclusions

A new hybrid search algorithm was presented in this paper to solve the coupled (binary-real) nonconvex optimization problem. Our hybrid search algorithm consists of a novel informed GA and the incomplete Cholesky preconditioned conjugate gradient algorithm [23]. The informed GA is used as a stochastic global minimizer for a new energy function consisting of the binary line process

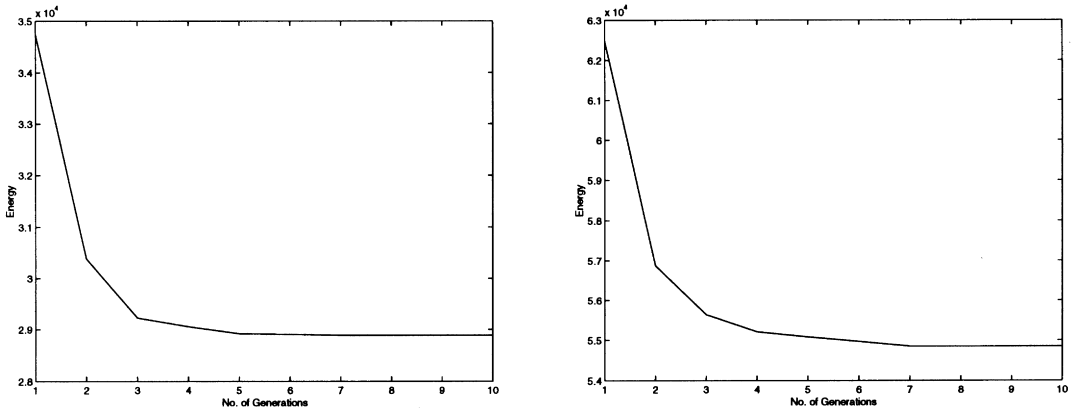

Fig. 7. The convergence curves of our hybrid search algorithm on the image restoration experiments with: (a) SNR-19 db; and (b) SNR = 15db.

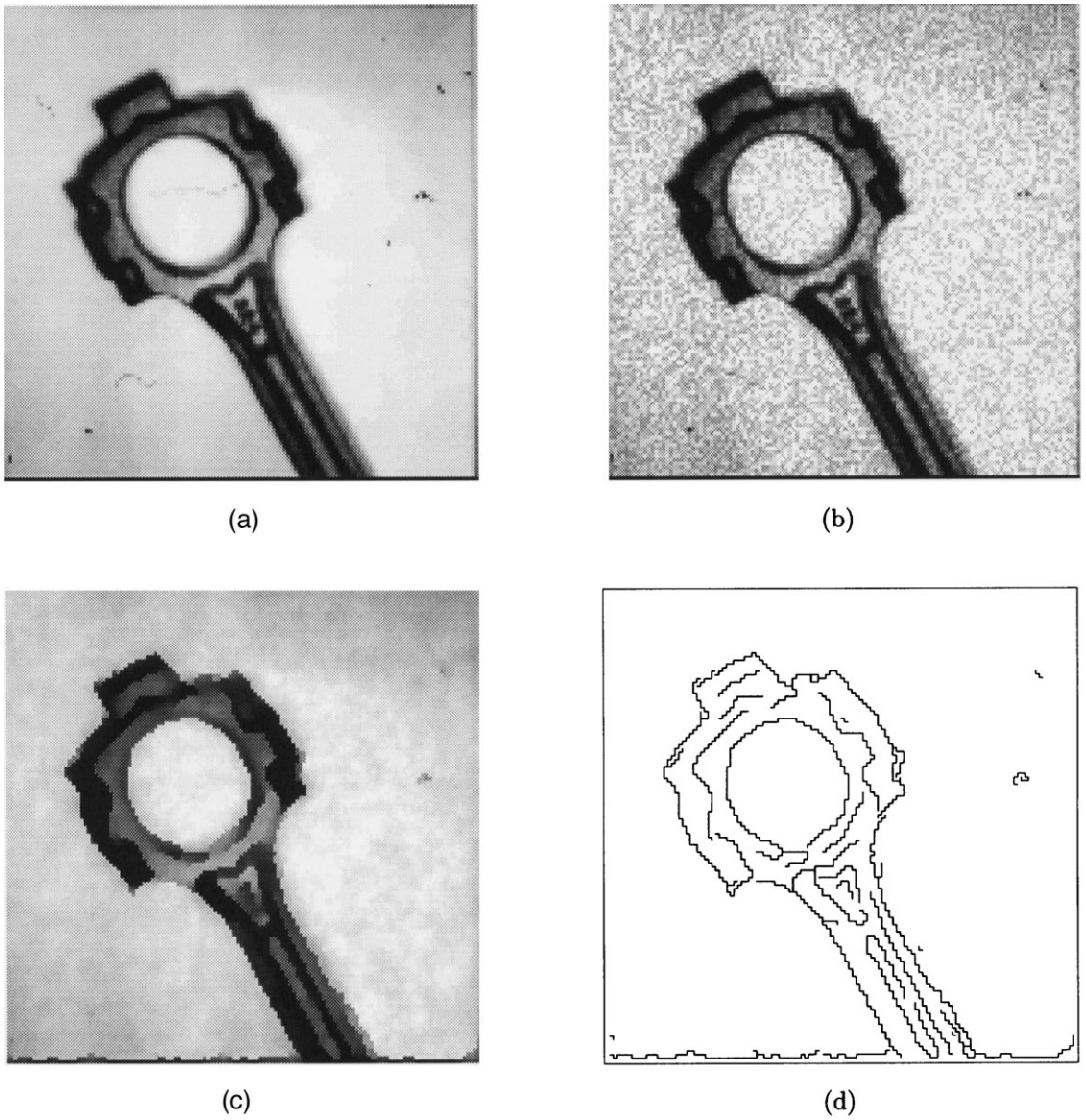

Fig. 8. (a) The original mechanical part image; (b) image with additive white Gaussian noise (SNR-20 db); (c) the restored image; and (d) the recovered discontinuity map after six generations of our informed GA.

variables only. Inside the GA, an incomplete Cholesky preconditioned conjugate gradient is used to determine the new energy function values for each line process configuration visited in the GA. The performance of this hybrid search algorithm is demonstrated via experiments on the sparse data surface reconstruction and image restoration problems. Since the GA is a highly parallelizable algorithm, we can improve the efficiency of this hybrid search algorithm by porting it to a multi-processor workstation.

The efficiency of our hybrid search algorithm can be further improved by generalizing the informed GA as a multi-scale stochastic optimization method. To achieve this, it is necessary to construct a multi-scale representation for the line process field. Heitz et al. [41] constructed a multi-scale representation for a label field and used a multi-scale relaxation algorithm to minimize the energy function for some early vision problems. It is possible to generalize this construction for the line process field and apply a multi-scale informed GA to make the hybrid search algorithm more efficient. Our future research efforts will focus on this very generalization.

Acknowledgements

Supported in part by the NSF grant ECS-9210648 and the Whitaker Foundation.

References

[1] A. Blake, A. Zisserman, Visual Reconstruction. The MIT Press, Cambridge, MA, 1987.
[2] W.E.L. Grimson, An implementation of computational theory for visual surface interpolation, Comput. Vis., Graph., Imag. Process. 6 (22) (1983) 3669.
[3] K. Ikeuchi, B.K.P. Horn, Numerical shape from shading and occluding boundaries, Artificial Intell. 17 (1981) 141–184.
[4] B.K.P. Horn, B.G. Schunk, Determining optical flow, Artificial Intell. 17 (1981) 185–203.
[5] B.K.P. Horn, Height from shading, Int. J. Comput. Vis. 5 (1) (1990) 174–208.
[6] T. Poggio, V. Torre, C. Koch, Computational vision and regularization theory, Nature 317 (1985) 314–319.
[7] M. Bertero, T.A. Poggio, V. Torre, Ill-posed problems in early vision, Proc. IEEE 76 (1988) 869–889.
[8] T. Simchony, R. Chellappa, M. Shao, Direct analytical methods for solving Poisson equations in computer vision problems, IEEE Trans. Patt. Anal. Mach. Intell., 12 (1990) 435–446.
[9] A.N. Tikhonov, V.Y. Arsenin, Solutions of Ill-posed Problems, Winston, Washington, DC, 1977.
[10] D. Terzopoulos, Regularization of inverse visual problems involving discontinuities, IEEE Trans. Patt. Anal. Mach. Intell. 8 (4) (1986) 413–424.
[11] T. Poggio, V. Torre, Ill-posed problems and regularization analysis in early vision, Technical Report AI Memo 773, MIT, A.I. LAB. Cambridge, MA, 1984.
[12] D. Terzopoulos, The computation of visible representations, IEEE Trans. Patt. Anal. Mach. Intell. 10 (4) (1988) 417–438.
[13] R. Szeliski, Fast surface interpolation using hierarchical basis functions, IEEE Trans. Patt. Anal. Mach. Intell. 12 (6) (1990) 513–528.
[14] S.H. Lai, B.C. Vemuri, An $O(N)$ iterative solution to the Poisson equations in low-level vision problems, in Proc. IEEE Conf. Comput. Vis. Patt. Recog., Seattle, WA, June 1994, pp. 9–14.
[15] S.H. Lai, B.C. Vemuri, Physically-based adaptive preconditioning for early vision, IEEE Trans. Patt. Anal. Mach. Intell. 19 (1997) 594–607.
[16] D. Mumford, J. Shah, Boundary detection by minimizing functionals, in: Proc. IEEE Conf. Comput. Vis. Patt. Recog., San Francisco, CA, June 1985, pp. 22–26.
[17] S. Geman, D. Geman, Stochastic relaximation, Gibbs distributions, and the Bayesian restoration of images, IEEE Trans. Patt. Anal. Mach. Intell. 6 (6) (1984) 721–741.
[18] G. Koch, J. Marroquin, A. Yuille, Analog 'neuronal' networks in early vision, Proc. Natl Acad. Sci. USA 83 (1986) 4263–4267.
[19] D. Geiger, F. Girosi, Parallel and deterministic algorithms from MRF'S: surface reconstruction, IEEE Trans. Patt. Anal. Mach. Intell. 13 (5) (1991) 401–412.
[20] J. Marroquin, S. Mitter, T. Poggio, Probabilistic solution of ill-posed problems in computational vision, J. Am. Stat. Assoc. 82 (397) (1987) 76–89.
[21] J.G. Harris, C. Koch, E. Staats, J. Luo, Analog hardware for detecting discontinuities in early vision, Int. J. Comput. Vis. 4 (1990) 211–223.
[22] I. Gerace, A. Tonazzini, A deterministic algorithm for reconstructing images with interacting discontinuities, CVGIP: Graph. Models Imag. Process. 56 (1994) 109–123.
[23] S.H. Lai and B.C. Vemuri, Robust and efficient algorithms for optical flow computation, in Proc. Int. Symp. Comput. Vis., Coral Gables, FL, 1995, pp. 455–460.
[24] R. Szeliski, Bayesian Modeling of Uncertainty in Low-level Vision, Kluwer, Boston, 1989.
[25] G. Wahba, Spline Models for Observational Data, Society for Industrial and Applied Mathematics, Philadelphia, 1990.
[26] A.A. Amini, T.E. Weymouth, R.C. Jain, Using dynamic programming for solving variational problems in vision, IEEE Trans. Patt. Anal. Mach. Intell. 12 (1990) 855–867.
[27] S.G. Nadabar, A.K. Jain, Parameter estimation in MRF line process models, in: Proc. IEEE Conf. Comput. Vis. Patt. Recog., Champaign, IL, June 1992, pp. 528–533.
[28] J.L. Marroquin, Probabilistic solution of inverse problems. Ph.D. dissertation, Massachusetts Institute of Technology, 1985.
[29] L. Bedini, A. Tonazzini, Image restoration preserving discontinuities: the Bayesian approach and neural networks, Imag. Vis. Comput. 10 (2) (1992) 108–118.
[30] J.J. Grefenstette, Genetic algorithms and their applications, in: A. Kent, J.G. Williams (Eds.), The Encyclopedia of Computer science and Technology, vol. 21, Marcel Dekker, New York, 1990, pp. 1391–52.
[31] D.E. Goldberg, Genetic Algorithm in Search, Optimization and Machine Learning, Addison-Wesley, Reading, MA, 1989.
[32] T. Davis, Toward an extrapolation of the simulated annealing convergence theory onto the simple genetic algorithm. Ph.D. dissertation, University of Florida, 1991.
[33] G. Rudolph, Convergence analysis of canonical genetic algorithms, IEEE Trans. Neural Networks 5 (1) (1994) 96–101.
[34] J. Suzuki, A Markov chain analysis on simple genetic algorithm, IEEE Trans. Syst, Man Cyber. 25 (4) (1995) 655–659.
[35] G. Winkler, Image Analysis, Random Fields and Dynamic Monte Carlo Methods, Springer, Berlin, 1995.
[36] D.E. Goldberg, K. Deb, A comparative analysis of selection schemes used in genetic algorithms, Technical Report TCGA Report No. 90007, University of Alabama, 1990.
[37] M. de la Maza, B. Tidor, Boltzmann weighted selection improves performance of genetic algorithms, Technical Report AI Memo 1345, MIT, A.I. LAB., Cambridge, MA, 1991.
[38] G.H. Golub, C.F. Van Loan, Matrix Computations, 2nd ed., The Johns Hopkins University Press, Baltimore, 1989.

[39] J.A. Meijerink, H.A. van der Vorst, An iterative solution method for linear systems of which the coefficient matrix is a symmetric **M**-matrix, Math. Comput. 31 (37) (1977) 148–162.
[40] R. Shonkwiler, E. Van Vleck, Parallel speed-up of Monte Carlo methods for global optimization, J. Complex. 10 (1994) 64–95.
[41] F. Heitz, P. Perez, P. Bouthemy, Multiscale minimization of global energy functions in some visual recovery problems, CVGIP: Imag. Understand. 59 (1) (1994) 125–134.