Data Driven Mean-Shift Belief Propagation For non-Gaussian MRFs

Minwoo Park[1], Somesh Kashyap[2], Robert T Collins[1] and Yanxi Liu[1,2]
Department of Computer Science and Engineering[1]
Department of Electrical and Engineering[2]
The Pennsylvania State University
{mipark,rcollins,yanxi}@cse.psu.edu, {suk203}@psu.edu

Abstract

We introduce a novel data-driven mean-shift belief propagation (DDMSBP) method for non-Gaussian MRFs, which often arise in computer vision applications. With the aid of scale space theory, optimization of non-Gaussian, multimodal MRF models using DDMSBP becomes less sensitive to local maxima. This is a significant improvement over standard BP inference, and extends the range of methods that are computationally tractable. In particular, when pair-wise potentials are Gaussians, the time complexity of DDMSBP becomes bilinear in the numbers of states and nodes in the MRF. Experimental results from simulation and non-rigid deformable neuroimage registration demonstrate that our method is faster and more accurate than state-of-the-art inference algorithms.

1. Introduction

Probabilistic graphical models, a marriage of graph theory and probability theory, has become a popular area of research in vision and machine learning in recent years [1, 11, 29, 21]. Probabilistic graphical models are graphs in which nodes represent random variables and edges represent probabilistic relationships. A graphical model compactly specifies the factored form of a joint probability distribution over all variables, allowing inferences to be efficiently performed by marginalization. Probabilistic graphical models are especially popular in the computer vision community because many common vision problems can be mapped naturally into a graphical model framework [3, 8, 21, 11, 29, 22, 15, 28].

A major computational stumbling block in graphical models is the fact that summing over the joint probability distribution is exponential with respect to the number of nodes in a graphical model. Many researchers have proposed various alternatives to find approximate solutions instead [1]. When the graph is sparsely connected, belief propagation (BP) algorithms can turn an exponential inference computation into one that is linear with respect to the number of nodes in the graph. However belief propagation is only applicable when the variables in the nodes are discrete-valued or jointly represented by a single multivariate Gaussian distribution. This limitation rules out its use in many computer vision applications.

In this paper, we propose a novel inference algorithm called DDMSBP that is less sensitive to local maxima for non-Gaussian MRFs. In addition, when pair-wise potentials are Gaussians, the time complexity of the algorithm becomes O(LG)[1], and convergence is guaranteed as long as the dynamic MRF structure maintained during DDMSBP has a diagonally dominant inverse covariance. Besides being applicable to continuous MRFs, the proposed method is also applicable to discrete MRFs where the number of labels is too large to perform standard discrete belief propagation.

To validate the performance and applicability of the proposed algorithm, we apply the method to both continuous MRFs and a deformable neuroimage registration problem formulated as a discrete MRF. Results show the proposed method is faster and more accurate than other state-of-the-art algorithms [21, 13].

2. Related Work

Improving speed of inference on a graphical model is an active area of research. For discrete domains, Ramanan and Forsyth [23] improved the speed by static state space pruning and Coughlan and Shen [3] improved the speed by dynamic pruning and adding of states. For special case MRFs with convex potential functions, Felzenszwalb and Huttenlocher [5] reduced the computation time of belief propagation by several orders of magnitude using min convolution, bipartite graphs and multi-grid methods. More recently, Komodakis and Tziritas [13] proposed an optimization method using duality theory of linear programming to provide an alternative, more general view of state-of-the-art techniques

[1] L: number of labels or samples, G: number of nodes in the graph.

like α-expansion. In contrast to α-expansion, their derived algorithms generate solutions with guaranteed optimality properties for a much wider class of problems. However, none of these methods is suitable for high dimensional continuous spaces.

For continuous cases, variational inference and expectation propagation (EP) has been proven to be effective [1]. However, as noted by T. Minka [20], EP can lead to a poor approximation when potentials in the graphical model are complicated (multimodal).

Several versions of continuous BP for non-Gaussian graphical models have been proposed [29, 11, 8]. By representing arbitrary density functions using particles, nonparametric belief propagation (NBP) can approximate BP inference for a continuous hidden variable space. However, the standard NBP algorithm is slow due to the sampling process on the products of samples [8]. To address this issue, a variant of NBP was proposed using sequential density estimation and mode propagation [8]. Later, Park et al [22, 21] proposed a more efficient mean-shift belief propagation (MSBP) method. They use mean-shift to perform nonparametric mode-seeking on implicitly defined belief surfaces generated within the belief propagation framework. Since MSBP only needs to compute a local grid of samples, it can achieve better efficiency. However, MSBP can suffer from local maxima if the mean-shift bandwidth is not set properly. Another limitation of all current BP algorithms is the uncertainty about their convergence when applied to a non-Gaussian MRF.

Recent work on smoothing-based optimization [14] and mean-shift with a variable bandwidth [2] has shown that it is possible to find a global maximum of a highly multimodal function. These results are based on scale space theory [16], which shows that local optima disappear very fast for most functions with increasing variance of the Gaussian blur kernel. Although it is not always possible to trace back from the global maximum of a blurred function to the original global maximum, it is typically possible to trace back to a significant local maximum [14].

3. Data Driven Mean-Shift Belief Propagation

We develop two theorems that connect MSBP [22, 21], smoothing based optimization [14], and Gaussian belief propagation for inference of a non-Gaussian MRF. Since the proposed method adapts gradually to the given data during the MSBP procedures, we call this method data driven mean-shift belief propagation (DDMSBP).

3.1. Background

Consider smoothing a function $f(x)$ by convolving with a zero-mean Gaussian blur kernel with covariance matrix $\sigma^2 I$. The value of this smoothed function evaluated at location μ is equivalent to $F(\mu,\sigma) = \int \mathcal{N}(\mathbf{t};\mu,\sigma) f(\mathbf{t})\mathbf{dt}$ where $\mathcal{N}(\mathbf{t};\mu,\sigma)$ is a multidimensional Gaussian with mean μ and covariance matrix $\sigma^2 I$.

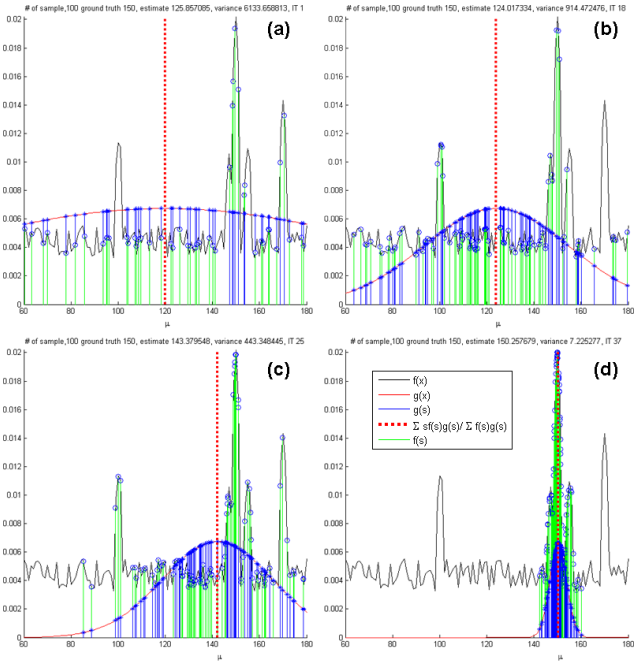


Figure 1: Progress of smoothing based optimization is shown from (a) ∼(d), where the dotted red line is the sample mean computed by Eq. (1), the solid red line is the unnormalized Gaussian proposal function $\mathcal{N}(x;\mu^{(t)},\sigma^{(t)})$, the black solid line is the original objective function $f(x)$, the solid blue line is a sample drawn from the Gaussian proposal function $\mathcal{N}(x;\mu^{(t)},\sigma^{(t)})$, and the green solid line is the original objective function at sample location s, $f(s)$.

Leordeanu and Hebert [14] define a sequence of mean and variance pairs $(\mu^{(t)},\sigma^{(t)})$ where

$$\mu^{(t+1)} = \frac{\int x\mathcal{N}(x;\mu^{(t)},\sigma^{(t)})f(x)dx}{\int \mathcal{N}(x;\mu^{(t)},\sigma^{(t)})f(x)dx} \tag{1}$$

$$\sigma^{(t+1)} = \sqrt{\frac{\int (1/n\sum_{i=1}^{n}(x_i^k-\mu_i^{(t)})^2)\mathcal{N}(x;\mu^{(t)},\sigma^{(t)})f(x)dx}{\int \mathcal{N}(x;\mu^{(t)},\sigma^{(t)})f(x)dx}} \tag{2}$$

and prove that the following inequalities hold:

$$F(\mu^{(t+1)},\sigma^{(t)}) \geq F(\mu^{(t)},\sigma^{(t)}) \tag{3}$$

$$F(\mu^{(t)},\sigma^{(t+1)}) \geq F(\mu^{(t)},\sigma^{(t)}) \tag{4}$$

Fig. 1 shows sample progress of this smoothing-based optimization procedure [14] on an arbitrary function f(x). As can be seen, the estimate of the global mode is approaching the true global mode with increasing confidence expressed by inverse variance of $\mathcal{N}(x;\mu^{(t)},\sigma^{(t)})$ as the

smoothing-based algorithm iterates. Although the estimation of global mode and confidence are wrong at first, the estimation error decreases and confidence increases with more iterations.

3.2. Theorems for DDMSBP

We generalize the result of Leordeanu and Hebert [14] to apply to belief propagation in a non-Gaussian MRF. With the aid of scale space theory, optimization of non-Gaussian, multimodal MRF models become less sensitive to local maxima. This is a significant improvement over standard BP inference.

Theorem 1: Consider the joint probability in an MRF:

$$p(\mathbf{x}) = k \prod_{ij} \psi(x_i, x_j) \prod_i \phi(x_i) \tag{5}$$

where k, $\mathbf{x}$, ϕ, and ψ are a normalizing constant, N dimensional column vector $\mathbf{x} = [x_1 x_2 \cdots x_N]^T$, unary potential function, and pair-wise potential function respectively. To estimate

$$\hat{\mathbf{x}} = arg_{\mathbf{x}} \max\, p(\mathbf{x}) \tag{6}$$

belief propagation can be applied repeatedly to a dynamic MRF[2] defined by unary potential functions $\phi(x_i)\mathcal{N}(x_i; \mu_i^{(t)}, \sigma^{(t)})$ where $\mathcal{N}(x_i; \mu_i^{(t)}, \sigma^{(t)})$ is a Gaussian distribution with mean $\mu_i^{(t)}$ and standard deviation $\sigma^{(t)}$. Let $p(x_i)^{(t)}$ be the marginal density for node i at iteration t. Then the updated mean and standard deviation for the next iteration are given as $\mu^{(t+1)} = \int x_i p(x_i)^{(t)} dx_i$ and $\sigma^{(t+1)} = \left[\frac{1}{N}\sum_{i=1}^{N} \int (x_i - \mu_i^{(t)})^2 p(x)^{(t)} dx_i\right]^{1/2}$

Proof : We start by multiplying $p(\mathbf{x})$ by the multivariate Gaussian $\mathcal{N}(\mathbf{x}; \mu^{(t)}, \sigma^{(t)} I)$

$$\begin{gathered}\mathcal{N}(\mathbf{x}; \mu^{(t)}, \sigma^{(t)} I) p(\mathbf{x}) = \\ k\mathcal{N}(\mathbf{x}; \mu^{(t)}, \sigma^{(t)} I) \textstyle\prod_{ij} \psi(x_i, x_j) \prod_i \phi(x_i).\end{gathered} \tag{7}$$

Since $\mathcal{N}(\mathbf{x}; \mu^{(t)}, \sigma^{(t)} I)$ has a diagonal covariance matrix, $\mathcal{N}(\mathbf{x}; \mu^{(t)}, \sigma^{(t)} I)$ can be decomposed to a product of univariate Gaussians

$$\mathcal{N}(\mathbf{x}; \mu^{(t)}, \sigma^{(t)} I) = \prod_i \mathcal{N}(x_i; \mu_i^{(t)}, \sigma^{(t)}) \tag{8}$$

and $p(\mathbf{x})\mathcal{N}(\mathbf{x}; \mu^{(t)}, \sigma^{(t)} I)$ can be rewritten as:

$$\begin{gathered}p(\mathbf{x})\mathcal{N}(\mathbf{x}; \mu^{(t)}, \sigma^{(t)} I) = \\ k \textstyle\prod_{ij} \psi(x_i, x_j) \prod_i \phi(x_i) \mathcal{N}(x_i; \mu_i^{(t)}, \sigma^{(t)})\end{gathered} \tag{9}$$

[2]A dynamic MRF is an MRF varying over time. In our case the MRF is varying over iterations (t).

To compute the mean and standard deviation of each variable x_i, a marginal density (belief), $p(x_i)^{(t)} = k' b(x_i)^{(t)}$ where k' is a normalizing constant, is first computed by the BP algorithm on an MRF with unary potential function given as $\phi(x_i)\mathcal{N}(x_i; \mu_i^{(t)}, \sigma^{(t)})$. The updated mean and standard deviation at node i, $i = 1 \sim N$, are then given by:

$$\mu_i^{(t+1)} = k' \int_{x_i} x_i b(x_i)^{(t)} dx_i \tag{10}$$

$$\sigma_i^{(t+1)} = \sqrt{k' \int_{x_i} (x_i - \mu_i^{(t)})^2 b(x_i)^{(t)} dx_i} \tag{11}$$

respectively.

Therefore, the mean of all nodes in column vector notation $\mu^{(t+1)}$ is given as:

$$\mu^{(t+1)} = \begin{bmatrix} \mu_1^{(t+1)} \\ \mu_2^{(t+1)} \\ \vdots \\ \mu_N^{(t+1)} \end{bmatrix} = \begin{bmatrix} k' \int_{x_1} x_1 b(x_1)^{(t)} dx_1 \\ k' \int_{x_2} x_2 b(x_2)^{(t)} dx_2 \\ \vdots \\ k' \int_{x_N} x_N b(x_N)^{(t)} dx_N \end{bmatrix} \tag{12}$$

Since the marginal density for node i, $p(x_i)^{(t)} = k' b(x_i)^{(t)}$ is defined as:

$$\begin{aligned} k' b(x_i)^{(t)} &= \frac{\int\int \ldots \int_{x_{k,k\neq i}} p(\mathbf{x}) \mathcal{N}(\mathbf{x}; \mu^{(t)}, \sigma^{(t)}) dx_1 \ldots dx_k}{\int\int \ldots \int_{x_k} p(\mathbf{x}) \mathcal{N}(\mathbf{x}; \mu^{(t)}, \sigma^{(t)}) dx_1 \ldots dx_k} \\ &= \frac{\int\int \ldots \int_{x_{k,k\neq i}} p(\mathbf{x}) \mathcal{N}(\mathbf{x}; \mu^{(t)}, \sigma^{(t)}) dx_1 \ldots dx_k}{\int_{\mathbf{x}} p(\mathbf{x}) \mathcal{N}(\mathbf{x}; \mu^{(t)}, \sigma^{(t)}) \mathbf{x}} \end{aligned} \tag{13}$$

for unnormalized $p(\mathbf{x})\mathcal{N}(\mathbf{x}; \mu^{(t)}, \sigma^{(t)})$, the denominator of Eq. (13) is the same for all nodes, $i = 1 \cdots N$. Then Eq. (12) can be summarized as:

$$\mu^{(t+1)} = \frac{\int_{\mathbf{x}} \mathbf{x} p(\mathbf{x}) \mathcal{N}(\mathbf{x}; \mu^{(t)}, \sigma^{(t)}) \mathbf{dx}}{\int_{\mathbf{x}} p(\mathbf{x}) \mathcal{N}(\mathbf{x}; \mu^{(t)}, \sigma^{(t)}) \mathbf{dx}} \tag{14}$$

It follows from Eqs. (1) and (3) that $\mu^{(t+1)} - \mu^{(t)}$ is in a gradient ascent direction of the smoothed function:

$$P(\mathbf{x}, \sigma) = \int_{\mathbf{t}} \mathcal{N}(\mathbf{x} - \mathbf{t}; 0, \sigma^2 I) p(\mathbf{t}) \mathbf{dt} \tag{15}$$

and that

$$P(\mu^{(t+1)}, \sigma^{(t)}) \geq P(\mu^{(t)}, \sigma^{(t)}). \tag{16}$$

The sigma update rule can be shown similarly. This theorem explains the good experimental results of mean-shift belief propagation (MSBP) presented by Park et al. [22, 21] where a constant blur kernel was used and no theoretical connection was given between BP and non-parametric mode seeking. Based on Theorem 1, we can seek the MAP estimate $\hat{\mathbf{x}}$ by applying BP repeatedly on a dynamic MRF with unary

potential functions $\phi(x_i)\mathcal{N}(x_i;\mu_i^{(t)},\sigma^{(t)})$ updated according to Eq. (14) at each iteration (t).

Theorem 1 states that optimization on an MRF only requires computing mean and variance of a smoothed marginal density function, and that the mean and variance updates perform non-parametric gradient ascent in a scale-space representation of the joint probability. The motivation of Theorem 2, below, is that we may approximate the unary potential function of the dynamic MRF by a Gaussian and use the closed form Gaussian BP rule [31] for each update. Theorem 2 also provides a confidence measure on the approximation error during each iteration.

Theorem 2: Consider the joint probability in a dynamic MRF:

$$p(\mathbf{x}) = k\prod_{ij}\psi(x_i,x_j)\prod_i \phi(x_i)\mathcal{N}(x_i;\mu_i^{(t)},\sigma^{(t)}) \quad (17)$$

where the belief propagation message update rule is:

$$m_{ik}(x_k) = \int_{x_i}\phi(x_i)\mathcal{N}(x_i;\mu_i^{(t)},\sigma^{(t)})q(x_i,x_k)dx_i \quad (18)$$

with $q(x_i,x_k)$ given by:

$$q(x_i,x_k) = \psi(x_i,x_k)\prod_{m\in\mathbb{N}(i)\setminus k} m_{mi}(x_i). \quad (19)$$

Consider approximating the message given by Eq. (18) with

$$\hat{m}_{ik}(x_k) = \int_{x_i}\mathcal{N}(x_i;\hat{\mu},\hat{\sigma})q(x_i,x_k)dx_i \quad (20)$$

where $\mathcal{N}(x_i;\hat{\mu},\hat{\sigma})$ is a Gaussian distribution with mean $\hat{\mu}_i$ and variance $\hat{\sigma}_i^2$ defined as

$$\hat{\mu}_i = \int_{x_i} x_i\phi(x_i)\mathcal{N}(x_i;\mu_i^{(t)},\sigma^{(t)})dx_i \quad (21)$$

and

$$\hat{\sigma}_i^2 = \int (x_i-\hat{\mu})^2\phi(x_i)\mathcal{N}(x_i;\mu_i^{(t)},\sigma^{(t)})dx_i. \quad (22)$$

When $\sigma^{(t)} \to 0$ the approximation error, $|m_{ik}(x_k) - \hat{m}_{ik}(x_k)|$ becomes 0, and the estimated marginal means and variances over unary potential functions given by $\phi(x_i)\mathcal{N}(x_i;\mu_i,\sigma)$ and $\mathcal{N}(x_i;\hat{\mu},\hat{\sigma})$ become identical.

Proof : Without loss of generality, let the non-Gaussian multimodal function $\phi(x_i)$ be represented as a mixture of Gaussians

$$\phi(x_i) = \sum_{m=1}^{\infty} w_m\mathcal{N}(x_i;\mu_m,\sigma_m) \quad (23)$$

where the w_m are mixing coefficients that sum to 1, $\sum_{m=1}^{\infty} w_m = 1$. Since the product of Gaussians is also a Gaussian, Eq. (18) becomes :

$$m_{ik}(x_k) = \int_{x_i}\sum_{m=1}^{\infty} w_m\mathcal{N}(x_i;\mu_{mi},\sigma_{mi})q(x_i,x_k)dx_i \quad (24)$$

where μ_{mi} and σ_{mi} are:

$$\mu_{mi} = \frac{\mu_m{\sigma^{(t)}}^2 + \mu_i^{(t)}\sigma_m^2}{{\sigma^{(t)}}^2+\sigma_m^2}, \sigma_{mi} = \sqrt{\frac{\sigma_m^2{\sigma^{(t)}}^2}{{\sigma^{(t)}}^2+\sigma_m^2}} \quad (25)$$

Similarly, the mean $\hat{\mu}_i$ and variance $\hat{\sigma}_i^2$ are computed as:

$$\begin{aligned}\hat{\mu}_i &= \textstyle\int_{x_i} x_i\phi(x_i)\mathcal{N}(x_i;\mu_i^{(t)},\sigma^{(t)})dx_i \\ &= \textstyle\sum_{m=1}^{\infty} w_m\int_{x_i} x_i\mathcal{N}(x_i;\mu_{mi},\sigma_{mi})dx_i \\ &= \textstyle\sum_{m=1}^{\infty} w_m\mu_{mi}\end{aligned} \quad (26)$$

and

$$\begin{aligned}\hat{\sigma}_i^2 &= \textstyle\int(x_i-\hat{\mu})^2\phi(x_i)\mathcal{N}(x_i;\mu_i^{(t)},\sigma^{(t)})dx_i \\ &= \textstyle\sum_{m=1}^{\infty} w_m\int(x_i-\hat{\mu})^2\mathcal{N}(x_i;\mu_{mi},\sigma_{mi})dx_i \\ &= \textstyle\sum_{m=1}^{\infty} w_m\int\{x_i-(\hat{\mu}-\mu_{mi})\}\}^2\mathcal{N}(x_i;0,\sigma_{mi})dx_i \\ &= \textstyle\sum_{m=1}^{\infty} w_m\left\{\sigma_{mi}^2+(\hat{\mu}-\mu_{mi})^2\right\}\end{aligned} \quad (27)$$

respectively.

Now as $\sigma^{(t)} \to 0$, we find that $\mu_{mi} \to \mu_i$, $\sigma_{mi} \to 0$, $\hat{\mu} \to \mu_i$, and $\hat{\sigma} \to 0$ from Eq. (25) $\sim$ (27). Therefore $m_{ik}(x_k)$ becomes:

$$\begin{aligned}m_{ik}(x_k) &= \textstyle\sum_{m=1}^{\infty} w_m\int_{x_i}\mathcal{N}(x_i;\mu_i,0)q(x_i,x_k)dx_i \\ m_{ik}(x_k) &= q(\mu_i,x_k)\textstyle\sum_{m=1}^{\infty} w_m = q(\mu_i,x_k)\end{aligned} \quad (28)$$

where we use the equalities $\mathcal{N}(\mu_i,0) = \delta(x-\mu_i)$[3] and $\sum_{m=1}^{\infty} w_m = 1$. Similarly, the approximated message becomes:

$$\hat{m}_{ik}(x_k) = \int_{x_i}\mathcal{N}(x_i;\mu_i,0)q(x_i,x_k)dx_i = q(\mu_i,x_k). \quad (29)$$

Therefore, the approximated messages and original message become identical, and so do the beliefs.

Corollary 1 : For an MRF defined by non-Gaussian multimodal unary potential functions and Gaussian pair-wise potential functions[4], a problem given by:

$$\hat{\mathbf{x}} = arg_{\mathbf{x}}\max\, p(\mathbf{x}) \quad (30)$$

[3] $\delta(x)$ is a Dirac delta function

[4] We note that non-Gaussian multimodal distributions as unary potential functions are prevalent while pair-wise potential functions tend to be Gaussians in computer vision problems, e.g. regularized deformation fields, MRF models for lattice structures, and so on.

may be solved by an iterative sequence of closed-form solutions when $\sigma^{(t)} \rightarrow 0$.

Corollary 2 : For the same problem, Eq. (30), the iterative closed form solution is guaranteed to converge as long as the dynamic MRF during each iteration has diagonally dominant Σ_{ij}^{-1}. Proof: This is so because the dynamic MRF at any iteration (t) is a Gaussian MRF, by definition. The convergence of the Gaussian MRF for a diagonally dominant Σ_{ij}^{-1} is proven by the work of Weiss and Freeman [31].

3.3. DDMSBP Algorithm

When all the messages and potential functions are Gaussian for a dynamic MRF at iteration (t), they can be expressed by their means and covariances. Therefore, we may perform BP in bilinear time with respect to L and G (Theorem 1 and 2). Letting the number of iterations required for the "while" loop and Gaussian BP be k_w and k_b respectively (Fig. 2), the time complexity of the DDMSBP is $O(Gk_wk_b + LGk_w) \simeq O(LG)$ where k_w and k_b are constant. Furthermore, the proposed algorithm can be implemented on a parallel processor using up to G threads without loss of accuracy. Then time complexity becomes $O(L)$.

1 Start with initial values of $\mu^{(0)}$ and $\sigma^{(0)}$, set $t = 0$
2 while $\sigma^{(t)} \leq \epsilon$
3 **Draw samples** $s_1, s_2, ..., s_m$ **from the normal distribution** $\mathcal{N}(x_i; \mu_i^{(t)}, \sigma^{(t)})$
4 **Compute weighted means for every node Set:** $\hat{\mu}_i = \frac{\sum_{j=1}^{m} s_j \phi_i(s_j)}{\sum_{j=1}^{m} \phi_i(s_j)}$
5 **Compute weighted variances for every node Set:** $\hat{\sigma}_i = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \frac{\sum_{j=1}^{m} (s_j - \hat{\mu}_j)^2 \phi_i(s_j)}{\sum_{j=1}^{m} \phi_i(s_j)}}$
6 **Build Gaussian MRF using computed** $\hat{\mu}_i$ **and** $\hat{\sigma}_i$
7 **Run Gaussian BP**
8 **Set:** $\mu^{(t+1)}$ to mean estimate of Gaussian BP
9 **Set:** $\sigma^{(t+1)}$ to the mean of all variance estimates of Gaussian BP
10 $t \leftarrow t + 1$
11 end

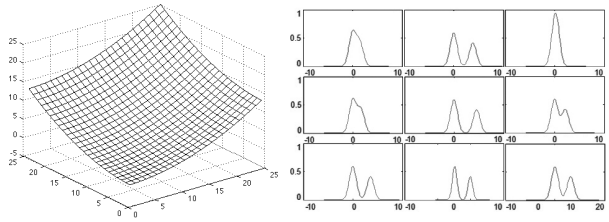
Figure 2: Pseudo code for DDMSBP algorithm: Lines 4 and 5 are for the approximation of $\phi(x_i)\mathcal{N}(x_i; \mu_i^{(t)}, \sigma^{(t)})$by $\mathcal{N}(x_i; \hat{\mu}_i, \hat{\sigma}_i)$ to define the dynamic MRF at iteration (t).

4. Validation of DDMSBP

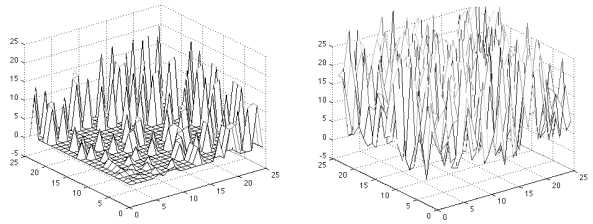
To validate our approach, we demonstrate applications of DDMSBP to an interpolation problem on a continuous MRF, and to 3D deformable registration using a discrete MRF.

4.1. Interpolation Problem on Continuous MRF

To test the performance of the proposed method on a continuous MRF, we adopt the approach of Park et al [21] who modified the simulation performed by Weiss and Freeman [31] to use non-Gaussian unary potentials. Likewise, we also run NBP [27, 11, 29], simulated annealing using Markov Chain Monte Carlo (MCMC) moves [6, 24], MSBP, EP, and the proposed method on the same problem to perform inference on a 25×25 grid with the same parameter settings (Fig. 3). Approximating the joint MAP by the max-marginal estimate using NBP is a common practice in problems defined on graphs since there is no max-product BP for continuous multimodal MRFs [10]. We use L=200 samples for DDMSBP, NBP, and simulated annealing, epsilon=0.01 for DDMSBP and MSBP, and 11 bins for MSBP for the first experiment.



(a) Surface $z(x, y)$ (b) Samples of unary potentials



(c) Given measurement of $z(x, y)$ (d) Initialization

Figure 3: (a) Surface $z(x, y) = (x^2 + y^2)/50$ (b) Sample non-Gaussian unary potentials out of 25×25 functions (c) Actual measurement where only 20% of the nodes are measurable (d) Random initialization of pixel labels at the 2D grid points

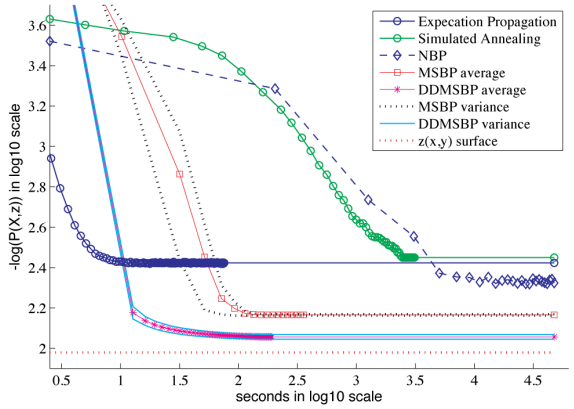
The joint probability used in the simulation is the same as in [21] and is given as:

$$P(\mathbf{X}, z) = k \prod_{ij} e^{-\beta(\mathbf{x_i} - \mathbf{x_j})^2} \prod_{i} \phi(\mathbf{x_i}, z_i), \qquad (31)$$

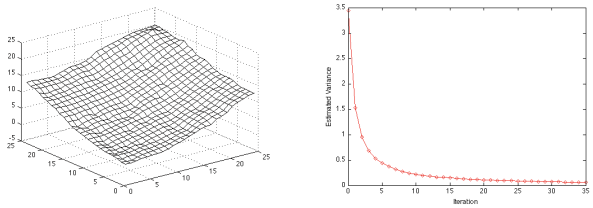
where k is a normalization constant and $\beta = 0.1$ if nodes $\mathbf{x_i}, \mathbf{x_j}$ are neighbors in a 4 connected graph and 0 otherwise. The multimodal unary potential is

$$\phi(\mathbf{x_i}, z_i) = e^{-\alpha_i(\mathbf{x_i}-z_i)^2} + 0.7e^{-\alpha_i(\mathbf{x_i}-z_i-r_i)^2}, \quad (32)$$

where α_i is randomly selected to be 10^{-6} or 1 with probability of 0.8 and 0.2 respectively and r_i is randomly chosen between $[0, 5]$ (Fig. 3b). When $\alpha_i = 1$, we set measurement z_i to be a sample from the surface $z(x, y) = (x^2 + y^2)/50$ and $z_i = 0$ otherwise. As can be seen in Fig. 3c only 20% of the nodes are observable and the rest of the nodes have a weak prior of zero. Note that z_i is a measurement at node i, one of the 2D grid points expressed by (x, y).



(a) Accuracy vs time trade off by $-log(P(\mathbf{x}, z))$



(b) Left:Estimate surface, Right: Smoothing bandwidth at each iteration

Figure 4: (a) Accuracy vs time trade off by $-log(P(\mathbf{X}, z))$ (b) The results of estimation using the proposed method (Left) and the smoothing bandwidth at sample iterations (Right) are shown.

As can be seen in Fig. 4a, EP converges fast but with less accuracy in terms of MAP criterion than NBP, MSBP, and DDMSBP. This lack of accuracy is expected when EP is used for multimodal unary potentials. The proposed method converges several orders of magnitude faster than the other methods (except EP) while achieving better accuracy in terms of MAP criterion. We ran both MSBP and the proposed DDMSBP method 100 times to measure the bounds plotted in Fig. 4a. Fig. 4b shows progress of variance (bandwidth) smoothing parameter updates during a sample run.

4.2. 3D Deformable Registration via discrete MRF

Neuroimage registration has been a fundamental research topic for many years in medical image analysis, where establishing correspondence between two sample images is crucial for many applications, including neuroimage classification [26], computer aided diagnosis [17, 18], statistical quantification of human brains and neuroimage segmentation [30]. While traditional registration techniques such as affine registration can account for differences in position and shape, they are not sufficient for modeling local deformations. Many deformable registration techniques have been proposed [12]. One of the major limitations of these algorithms is their execution time, which can take hours to complete. Recently Glocker et al. and Shekhovtsov et al. [7, 25] proposed a fast deformable registration algorithm using an MRF-based model. In this section we compare our DDMSBP algorithm with [7] on a volumetric deformable registration problem.

4.2.1 Registration Method using DDMSBP

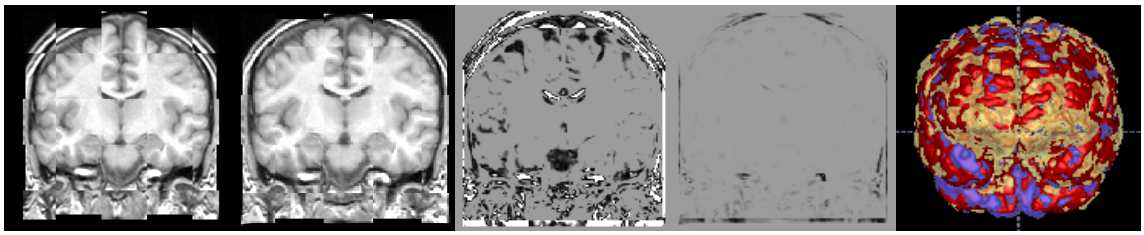
We follow the method developed by Glocker et al.[7] to validate the proposed algorithm. We use four resolution levels, starting with 40 mm control point spacing which is then refined to 20, 10 and finally 5 mm, resulting in a grid size varying from $7 \times 7 \times 6$ at the lowest resolution to $-49 \times 49 \times 39$ at the highest.

The registration is performed by optimizing joint density over an MRF defined by:

$$P(\mathbf{X}, z) = k \prod_{ij} e^{-\beta(\mathbf{x_i}-\mathbf{x_j})^2} \prod_i \phi(\mathbf{p_i} + \mathbf{x_i}, z_i), \quad (33)$$

where $\phi(\mathbf{p_i} + \mathbf{x_i}, z_i) = e^{-\alpha(1-(z_i+1)/2)^2}$, α is a weight for similarity, z_i is a normalized cross correlation (NCC) computed between a volume at control point $\mathbf{p_i}$ in a fixed image over volumes at $\mathbf{p_i} + \mathbf{x_i}$ in a warped image where $\mathbf{x_i}$ corresponds to deformation vector, and β is a regularization weight set to 0.005 as suggested by Glocker et al.[7]. For $\mathbf{x_i}$, we use sparse sampling along the main axis as in [7]. The voxel size at each grid point is a cube of size equal to the grid spacing.

Although the proposed method is intended for continuous MRFs, the method does not assume smoothness of the unary potential $\phi(x_i, z_i)$. Hence, DDMSBP can be applied to a discrete MRF as well. Sparse sample labels and associated observations $\phi(\mathbf{p_i} + \mathbf{x_i}, z_i)$ can be generalized to form a mixture of Gaussians $\sum_{m=1}^{N} \phi(\mathbf{p_m} + \mathbf{x_m}, z_m)\mathcal{N}(\mu_m, \sigma_m)$ where each Gaussian component has very small σ_m. We therefore can apply Eqs. (26) and (27) to perform the proposed DDMSBP algorithm on discrete data instead of performing lines 4 and 5 in the pseudo code given by Fig. 2.



(a) Checkerboard before (b) Checkerboard after (c) Difference before (d) Difference after (e) Volume overlap

Figure 5: Sample results of registration for one subject are shown. Checkerboard and difference image before and after registration clearly show the performance of DDMSBP-based deformable registration. (e) shows the volume overlap of gray matter between the manual segmentation by experts and the one found by our algorithm. The red color shows True Positive (TP) regions, the yellow color shows False Positive (FP) regions, and the blue color shows False Negative (FN) regions. ITK-SNAP was used to generate this figure [32]

4.2.2 Inter Subject Brain Registration

We use the Internet Brain Segmentation Repository (IBSR) structural brain MR dataset provided by the Center for Morphometric Analysis at Massachusetts General Hospital[5]. This dataset can be considered as a gold standard for registration as all the images in the dataset have been manually labeled by experts. The T1-weighted images in this dataset have been positionally normalized into the Talairach orientation (rotation only) and have a resolution of $0.9375 \times 0.9375 \times 1.5$ mm (image dimensions of $256 \times 256 \times 128$). To validate our method, we select one of the 8 images as the reference (ISBR_01_ana) and deformably register the remaining seven to this reference. The recovered deformation field is then applied to the segment labels. The accuracy of the method is quantitatively shown by reporting three measures derived from the volume overlap of the manual segmentation and the segmentation found by our algorithm (Fig. 6). The measures are DICE coefficient, sensitivity and specificity [4]. In the literature, a DICE score larger than 0.7 is considered good even for the segmentation of large objects (e.g. brain tissues) [19].

4.2.3 Experiment

All the experiments are performed on an OS X 10.5.8 machine with Intel Core 2 Duo 2.4Ghz 4GB DDR3. Since the scope of this paper is limited to the inference algorithm, we use the standard implementation for data computation and BSpline warping in the ITK library [9]. We compare the inference engine of the proposed DDMSBP method to a state-of-the-art method [13] for inference on a discrete MRF, called FastPD[6]. We use $L(= N) = 31$ and $epsilon = 0.01$ for DDMSBP and FastPD. As can be seen in Fig. 6 the proposed method produces better results in terms of DICE, sensitivity and specificity, and it is also two times faster than FastPD for the inference computation (2.60 sec vs. 5.62 sec for 3D)[7].

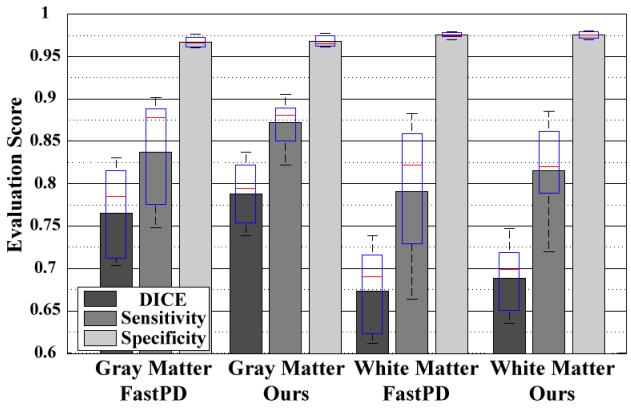


Figure 6: We compare the proposed method to FastPD in terms of DICE, sensitivity, and specificity.

5. Conclusion

We have presented a new BP algorithm called DDMSBP along with theoretical results demonstrating that DDMSBP is less sensitive to local maxima for non-Gaussian MRFs. In particular, when pair-wise potentials are Gaussians, our method has a time complexity of $O(LG)$, and guarantees convergence as long as the dynamic MRF constructed has a diagonally dominant inverse covariance. Application of the proposed method to numerical simulation on a continuous MRF and non-rigid deformable 3D brain registration on a discrete MRF confirm that our method is faster and more accurate than state-of-the-art algorithms in both domains.

[5]http:// www.cma.mgh.harvard.edu/ibsr/

[6]http://www.csd.uoc.gr/∼komod/FastPD/

[7]The score we measured for FastPD [7] is 6.11% lower than the one reported by Glocker et al. [7]. However, running that algorithm ourselves, as in the experiments here, is the only way to rule out other factors and get a fair comparison between the two inference engines.

References

[1] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006. 1, 2

[2] S. Chunhua, M. J. Brooks, and A. van den Hengel. Fast global kernel density mode seeking with application to localization and tracking. In *ICCV 2005*. 2

[3] J. Coughlan and S. Huiying. Shape matching with belief propagation: Using dynamic quantization to accomodate occlusion and clutter. In *CVPR Workshop*, 2004. 1

[4] L. R. Dice. Measures of the Amount of Ecologic Association Between Species. *Ecology*, 26(3):297–302, 1945. 7

[5] P. F. Felzenszwalb. Efficient Belief Propagation for Early Vision. *IJCV*, 70(1):41–54, 2006. 1

[6] R. Frost. Simulated Annealing Tools for Matlab v1.03. 5

[7] B. Glocker, N. Komodakis, G. Tziritas, N. Navab, and N. Paragios. Dense image registration through MRFs and efficient linear programming. *Medical Image Analysis*, 12(6):731–741. 6, 7

[8] T. X. Han, N. Huazhong, and T. S. Huang. Efficient nonparametric belief propagation with application to articulated body tracking. In *CVPR*. 1, 2

[9] L. Ibanez, W. Schroeder, L. Ng, and J. Cates. *The ITK Software Guide*. Kitware, Inc. ISBN 1-930934-15-7, http://www.itk.org/ItkSoftwareGuide.pdf, second edition, 2005. 7

[10] A. Ihler, J. Fisher, J. W. F. Iii, A. Willsky, and R. Moses. Nonparametric belief propagation for self-calibration in sensor networks. In *In Proceedings of the Third International Symposium on Information Processing in Sensor Networks*, pages 225–233, 2004. 5

[11] M. Isard. PAMPAS: real-valued graphical models for computer vision. In *CVPR*. 1, 2, 5

[12] A. Klein, J. Andersson, B. A. Ardekani, J. Ashburner, B. Avants, M.-C. Chiang, G. E. Christensen, D. L. Collins, J. Gee, and P. Hellier. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration., 2009. 6

[13] N. Komodakis and G. Tziritas. Approximate labeling via graph-cuts based on linear programming. *PAMI*, 29(8):1436–1453, 2007. 1, 7

[14] M. Leordeanu and M. Hebert. Smoothing-based Optimization. In *CVPR*, pages 1–8, 2008. 2, 3

[15] W. C. Lin and Y. Liu. A Lattice-based MRF Model for Dynamic Near-regular Texture Tracking. *PAMI*, pages 777–792, 2007. 1

[16] T. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer, December 1993. 2

[17] Y. Liu, L. Teverovskiy, O. Carmichael, R. Kikinis, M. Shenton, C. Carter, V. Stenger, S. Davis, H. Aizenstein, J. Becker, O. Lopez, and C. Meltzer. Discriminative MR Image Feature Analysis for Automatic Schizophrenia and Alzheimer's Disease Classification. In *Proceedings of the 7th International Conference on MedicalImage Computing and Computer Aided Intervention (MICCAI '04)*, pages 393 – 401, October 2004. 6

[18] Y. Liu, L. Teverovskiy, O. Lopez, H. Aizenstein, C. Meltzer, and J. Becker. Discovery of Biomarkers for Alzheimer's Disease Prediction from Structural MR Images. In *2007 IEEE International Symposium on Biomedical Imaging*, April 2007. 6

[19] B. C. M., C. L., T. Butz, C. O., and J. Thiran. Comparison and validation of tissue modelization and statistical classification methods in T1-weighted mr brain images. *IEEE Trans. Med. Imag.*, 24. 7

[20] T. Minka. Expectation propagation for approximate bayesian inference. In *Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence*, pages 362–369, 2001. 2

[21] M. Park, K. Brocklehurst, R. Collins, and Y. Liu. Deformed Lattice Detection in Real-World Images. *PAMI*, 2009. 1, 2, 3, 5

[22] M. Park, Y. Liu, and R. T. Collins. Efficient Mean Shift Belief Propagation for Vision Tracking. In *CVPR*, pages 1–8, Anchorage, Alaska, 2008. 1, 2, 3

[23] D. Ramanan and D. A. Forsyth. Finding and tracking people from the bottom up. In *CVPR*, volume 2, pages 467–474, 2003. 1

[24] P. Salamon, P. Sibani, and R. Frost. *Facts, Conjectures, and Improvements for Simulated Annealing*. SIAM Monographs on Mathematical Modeling and Computation. Society for Industrial and Applied Mathematics, 2002. 5

[25] A. Shekhovtsov, I. Kovtun, and V. l. Hlav. Efficient MRF deformation model for non-rigid image matching. *Computer Vision and Image Understanding*, 112(1):91–99, 2008. 6

[26] S. V. Shinkareva, H. C. Ombao, B. P. Sutton, A. Mohanty, and G. A. Miller. Classification of functional brain images with a spatio-temporal dissimilarity map. *NeuroImage*, 33(1):63–71, 2006. 6

[27] L. Sigal. Pampas / Non-parametric Belief Propagation Toolbox for Matlab v0.1, 2005. 5

[28] L. Sigal, M. Isard, B. H. Sigelman, and M. J. Black. Attractive people: Assembling loose-limbed models using nonparametric belief propagation. In *NIPS*, 2003. 1

[29] E. B. Sudderth, A. T. Ihler, W. T. Freeman, and A. S. Willsky. Nonparametric belief propagation. In *CVPR, 2003*, volume 1, pages 605–612. 1, 2, 5

[30] L. Teverovskiy, O. T. Carmichael, H. J. Aizenstein, N. Lazar, and Y. Liu. Feature-based vs. intensity-based brain image registration: Comprehensive comparison using mutual information. In *Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium on*, pages 576–579, 2007. 6

[31] Y. Weiss and W. Freeman. Correctness of Belief Propagation in Gaussian Graphical Models of Arbitrary Topology. *Neural computation*, 13(10):2173, 2001. 0899-7667. 4, 5

[32] P. A. Yushkevich, J. Piven, H. Cody Hazlett, R. Gimpel Smith, S. Ho, J. C. Gee, and G. Gerig. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage*, 31(3):1116–1128, 2006. 7