MARKOV RANDOM FIELD ENERGY MINIMIZATION VIA ITERATED CROSS ENTROPY WITH PARTITION STRATEGY

Jue Wu and Albert C. S. Chung

Bioengineering Program, and Computer Science and Engineering Department,
The Hong Kong University and Science and Technology, Clear Water Bay, Hong Kong.

ABSTRACT

This paper introduces a novel energy minimization method, namely iterated cross entropy with partition strategy (ICEPS), into the Markov random field theory. The solver, which is based on the theory of cross entropy, is general and stochastic. Unlike some popular optimization methods such as belief propagation (BP) and graph cuts (GC), ICEPS makes no assumption on the form of objective functions and thus can be applied to any type of Markov random field (MRF) models. Furthermore, compared with deterministic MRF solvers, it achieves higher performance of finding lower energies because of its stochastic property. We speed up the original cross entropy algorithm by partitioning the MRF site set and assure the effectiveness by iterating the algorithm. In the experiments, we apply ICEPS to two MRF models for medical image segmentation and show the aforementioned advantages of ICEPS over other popular solvers such as iterated conditional modes (ICM) and GC.

Index Terms— Markov random fields, energy minimization, MRF solvers, cross entropy, image segmentation

1. INTRODUCTION

Markov random field (MRF) theory has interested a lot of researchers in computer vision, image processing, pattern recognition, artificial intelligence and so on. The reasons why MRF modeling has so many successful applications in these areas are that it can easily incorporate spatial interaction and convert a problem in image computing into a functional optimization problem. This conversion is usually carried out by means of maximum *a posteriori* (MAP) or identically, energy minimization via Gibbs random fields.

The MRF energy function contains the local evidence (likelihood energy) of each site and the interaction (prior energy) between neighboring sites of MRF. We need to minimize the energy function in order to obtain the resulting MRF configuration. However, it remains an important open problem in MRF theory to optimize the objective function. Due to the large number of pixels in images, the configuration space of MRF in image analysis is huge. Actually, it was proved that obtaining the global optimum of an arbitrary MRF objective function is NP-hard [1]. Therefore, it has been an active research topic to design a "good" solver for MRF models over the past two decades. The goodness of a solver lies in whether it can efficiently find a local optimum which is as "global" as possible (i.e. the lower energy the solver can find, the better the solver is) [2, 1].

The work of Kirkpatrick *et al.* [3], who proposed simulated annealing (SA), is known to be one of the earliest efforts to design solvers for MRF energy minimization. Another pioneering work was done by Besag [4], where the iterated conditional modes (ICM) was presented. After those two methods, quite a few solvers were introduced [2], such as mean field approximation (MFa), relaxation labeling (RL), graduated non-convexity (GNC), etc. More recently two efficient and fairly effective solvers, belief propagation (BP) [5] and graph cuts (GC) [1], have been proposed. These two solvers are now often used for MRF models because they can find "global" optima within a rather large neighborhood while maintaining acceptable time complexity. However, BP and GC are not applicable to all types of objective functions. They obtain their effectiveness at the cost of objective function form restrictions. These limitations may considerably confine the usage of the two popular solvers.

In this paper, we propose a simple stochastic solver for MRF modeling, called iterated cross entropy with partition strategy (ICEPS). This idea is originated from the field of operations research to simulate rare events [6]. This work combines the idea of cross entropy (CE) with MRF theory and applies the method to MRF models that can be used in image segmentation, which is an important topic in image analysis. The ICEPS solver is a general and stochastic optimization method that can be applied to any kind of MRF formulation, unlike BP and GC. Moreover, as a stochastic method, ICEPS tends to find more global optimum than deterministic solvers like BP, GC and ICM. This is supported by a series of experiments on synthetic and clinical image data, which shows ICEPS is able to find lower energy value than GC and ICM. Although stochastic, ICEPS is still efficient because it attaches importance to the high performance part of the samples. In short, the contribution of this paper lies in that the introduction of ICEPS into MRF theory offers a better choice to approximately solve the NP-hard problem of minimizing

MRF energy compared to some of the commonly-used existing solvers in terms of both effectiveness and generalization.

This paper is an extension of our previous work [7] and the current work introduces two significant modifications to considerably reduce the time complexities and to keep the effectiveness of the original pure CE algorithm.

2. THE OPTIMIZATION METHOD

Owing to its simplicity and effectiveness, the cross entropy (CE) has many successful applications in operations research and machine learning [8]. In this section, we extend the original CE method to a simple stochastic solver for MRF modeling, namely, iterated cross entropy with partition strategy (ICEPS). Consider the following general energy minimization problem in MRF modeling. Let $\mathcal{F}$ be the configuration space of MRF, F, and f is one configuration of F. The energy minimization of MRFs is formulated by

$$f^* = \arg\min_{f \in \mathcal{F}} E(f), \tag{1}$$

where $E(\cdot)$ is the energy function to be minimized and f^* is the desired F configuration. ICEPS method associates an estimation problem with the optimization problem (Eq. 1). We suppose $p(\cdot; \vec{v})$ is a family of discrete probability density functions (pdf) on $\mathcal{F}$ and $\vec{v}$ is its parameter. Let us estimate the following probability

$$\mathcal{P}_v(E(F) \leq e) = \sum_f I_{\{E(f) \leq e\}} p(f; \vec{v}), \tag{2}$$

where $I_{\{event\}}$ is an indicator function, which is equal to 1 when the event is true otherwise 0, $\mathcal{P}_v$ is the probability measure and F is a vector of configurations that has pdf $p(\cdot; \vec{v})$. This equation shows a way to estimate the probability of the situation that the energy of one configuration in a sequence of configurations is not larger than a threshold e. If $e = e^* = \min_{f \in \mathcal{F}} E(f)$ and $p(\cdot; \vec{v})$ is a uniform density on $\mathcal{F}$, Eq. 1 and Eq. 2 are connected. Note that $\mathcal{P}_v(E(F) \leq e^*)$ is typically $1/|\mathcal{F}|$, which is very small. This is similar to the situation of rare event simulation. Thus, following [8] we can borrow the idea of CE from rare event simulation to construct a multi-level optimization approach for MRF energy, where we generate a sequence of levels $e_1, e_2, \ldots, e_T$ and parameter vectors $v_1, v_2, \ldots, v_T$ such that e_T is close to the optimal e^* and v_T is the density that assigns high probability mass to the configuration which corresponds to a low energy.

Suppose m is the size of the label space of the MRF model and there are n sites altogether. We partition the n sites into l subsets $S_1, \ldots, S_l$. The ICEPS solver for MRF labeling can be described as follows.

———————————————————————

ICEPS Algorithm for MRF energy minimization

———————————————————————

1. Set $p = 1$. Iterate steps 2 to 4 until $p = P$. P is a predefined positive integer.

2. For each site subset S_d, $d = 1, \ldots, l$, repeat the following steps from 3.1 to 3.5. $|S_d| = r$.

3.1. Set level $t = 1$ and the initial parameter vector $\vec{v}_0 = \{\vec{v}_{0,1}, \ldots, \vec{v}_{0,r}\}$. Each $\vec{v}_{t,i} = \{v_{t,i}^1, \ldots, v_{t,i}^m\}$ is a vector with m elements for site i.

3.2. Generate a sequence of samples $F_1, \ldots, F_N$ ($F = \{f_1, \ldots, f_r\}$ is one MRF configuration) from the density $p(\cdot; \vec{v}_{t-1})$ and compute the energy $E_i(F_i)$ for every $i \in \{1, \ldots, N\}$. N is the number of samples.

3.3. Sort all the $E_i(F_i)$ in a non-increasing order to $\{E_1, \ldots, E_N\}$. Then pick $e_t = E_{\lceil(1-\rho)N\rceil}$.

3.4. Use the samples $F_1, \ldots, F_N$ to update $\vec{v}_t$ by

$$v_{t,i}^j = \frac{\sum_{k=1}^N I_{\{E_k(F_k) \leq e_t\}} I_{\{F_{ki}=j\}}}{\sum_{k=1}^N I_{\{E_k(F_k) \leq e_t\}}}, \tag{3}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, m$.

3.5. If e_t remains unchanged for several iterations, go to step 4; else, set $t = t + 1$ and go to step 3.1.

4. $p = p + 1$.

5. The combination of $E_N(F_N)$ for each site subset at the P-th iteration is the estimated minimal MRF energy. The corresponding configuration is embodied by the parameter vector $\vec{v}_T$ for every subset, where each element $v_{T,i}^j$ assigns most probability mass to a preferable label among m labels for site i.

———————————————————————

Intuitively, among randomly-generated samples $\{F_1, \ldots, F_N\}$, the ICEPS algorithm prefers those high performance ones whose energy values are lower than a threshold e_t. It calculates the occurrence of the labeling of each site in these high performance samples and uses this statistical information to guide the next inner iteration. This method is named Cross Entropy because the optimization is actually a procedure that minimizes the cross entropy (a.k.a. Kullback-Leibler distance) between the optimal importance sampling density and the target density [8]. This algorithm is bound to terminate after a finite number of iterations because in the end the probability density v_t converges to 0 or 1 according to Proposition 4.2 in [8].

Function $p(\cdot; \vec{v})$ can be any kind of pdf but the independent m-point distribution is usually enough. This means the probability that the site i is set to label j at iteration t is proportional to $v_{t,i}^j$ (Note that $\sum_{j=1}^m v_{t,i}^j = 1$). Besides, three parameters need to be pre-defined, ρ, N and P. Usually, ρ is a small value between 1% and 10%. When the number

of sites, r, is large, we tend to choose larger ρ. Regarding the sample size N, we set $N = cr$, where c is a constant and often between 1 and 10. We usually keep the total number of outer iterations P less than 10 because the algorithm generally reaches a quite low energy after a few iterations. The ICEPS solver is flexible because we can strike a desirable balance between effectiveness and efficiency by tuning these three parameters, ρ, N and P. Usually the larger ρ (less than 1), N or P is, the lower energy we can achieve and it takes more time to run ICEPS. From the algorithm above, we can see that ICEPS only needs to evaluate the energy function so unlike GC and BP, ICEPS is general and can be applied to any type of MRF energy function.

There are some explanations on the above ICEPS algorithm. First, the division of the site set is implemented by partitioning an image into several sub-images. It is preferred to have some overlapping between a pair of subsets which are adjacent in the image. We can adjust the size of overlapping according to the size of neighborhood system of MRF models. The larger the neighborhood system is, the wider the overlapping region can be. The motivation to use such a partition strategy is to reduce the time complexity of the original CE algorithm. This strategy is justified by the MRF theory. The most significant characteristic of Markov random field is Markovianity, which means the status of one site is only dependent on its neighborhood. This feature allows us to use the "divide and conquer" strategy to obtain lower computational complexity. Second, we need to iterate the ICEPS algorithm in order to allow long-range interaction which may span more than one subset. The initialization of each iteration should be based on the previous one. In the previous iteration, we get an intermediate result on the labeling and we can assign a large weight to the site according to the intermediate result for the initialization. For example, if in the previous result site i is set to label j, in current iteration, the initialization can lean to that label, e.g. $v_{t,i}^j = 1/2$ and $v_{t,i}^i = 0.5/(m-1)$, for $i \neq j$. Thus, the link between two iterations in succession is established.

3. EXPERIMENTAL RESULTS

In this section, we apply the three solvers ICEPS, ICM and GC to two MRF models (Potts model and a boundary model as the prior MRF energy functions) for brain image segmentation and compare their performance. The three solvers are all used for the simple Potts model but only ICEPS and ICM are used for the complex compound model because GC is not applicable for it. Please refer to [9] for more details about the boundary model due to limited space. We adopt the simple but effective Gaussian function as the likelihood MRF function. Also notice that another popular solver, belief propagation (BP), is not proper for that model, either. The comparison of BP with cross entropy as solvers for simple MRF models can be found in the previous work [7]. In both MRF

models, the parameters for the likelihood energy are specified manually by selecting region of interest (ROI) and calculating the means and standard deviations since we know the ground truth. These parameters are kept the same in the comparison. The implementation environment is Microsoft Visual C++ .NET 2003 on a computer with 2.26GHz CPU and 1 GB memory.

3.1. Synthetic Images

We obtained a series of BrainWeb T1-weighted MRI data with different levels of noise from a public simulated brain database [10]. The image size is 217×181. It is a multi-class segmentation problem to divide human brains into different tissues like white matter (WM), gray matter (GM), cerebrospinal fluid (CSF), fat, skull, glial matter, connective, etc. In our experiments, we focus on the former three tissues because they are of major interest, have complicated structures and occupy around 70% of the whole brain volume. As such, we segment the data into four classes {WM, GM, CSF, others}. We remove all the other irrelevant tissue (e.g. skull, fat, etc) from the images by setting them the same as background. The results of minimal energy values and segmentation errors found by the three solvers using the Potts model as formulation are shown in Tab. 1. The results of minimal energy values and segmentation errors found by the two solvers using the second compound MRFs model as formulation are shown in Tab. 2. All the parameters and initialization are set to the same for the comparisons. We repeat the ICEPS algorithm for 5 times and calculate the means and standard deviations of the energy values because it is a stochastic solver. We show its mean performance and standard deviation.

noise level	ICM		GC		ICEPS	
	minimum	error	minimum	error	minimum	error
3%	25123	2.38%	24993	2.37%	**24938 ± 18**	**2.32% ± 0.03%**
5%	27828	2.96%	27694	2.93%	**27605 ± 2**	**2.91% ± 0.01%**
7%	29553	3.27%	29426	3.26%	**29145 ± 2**	**3.01% ± 0.01%**
9%	33739	4.30%	33436	4.21%	**32753 ± 23**	**3.43% ± 0.05%**

Table 1. *Results of the Potts MRF model using three different solvers for multi-class segmentation on BrainWeb data. The lowest energy and error rate of each case is highlighted in bold.*

dataset	ICM		ICEPS	
	minimum	error	minimum	error
3%	20094	2.37%	**15537 ± 3**	**1.88% ± 0.02%**
5%	22749	3.26%	**17955 ± 8**	**2.79% ± 0.05%**
7%	24576	3.52%	**19160 ± 30**	**3.43% ± 0.01%**
9%	28327	4.85%	**25463 ± 69**	**4.75% ± 0.04%**

Table 2. *Results of the compound MRF model using two different solvers for multi-class segmentation on BrainWeb data. The lower energy and error rate of each case is highlighted in bold.*

3.2. Real Medical Images

The real clinical datasets are obtained from the Internet Brain Segmentation Repository (IBSR) [11], which provides real magnetic resonance cerebral image data. The data we use for experiments are T1-weighted brain scan from a male subject and each slice is 256×256 large. The segmentation performed here has three classes {white matter, gray matter, others}. We again test the Potts MRF model on 4 datasets (slices) and calculate the energy values and segmentation errors obtained by the three solvers, which are shown in Tab. 3. The minimal energy reached by the two solvers ICM and ICEPS using the second compound MRFs model as formulation are shown in Tab. 4. Again we repeat the ICEPS algorithm for 5 times and calculate the means and standard deviations considering its stochastic property.

dataset	ICM		GC		ICEPS	
	minimum	error	minimum	error	minimum	error
1	19487	3.03%	19315	3.01%	**19283 ± 9**	**2.91% ± 0.02%**
2	25176	3.26%	23719	3.19%	**23687 ± 6**	**3.08% ± 0.01%**
3	19938	3.10%	19560	3.06%	**19523 ± 6**	**2.92% ± 0.01%**
4	20396	2.49%	20061	2.46%	**20020 ± 3**	**2.29% ± 0.03%**

Table 3. *Results of the Potts MRF model using three different solvers for multi-class segmentation on IBSR data. The lowest energy and error rate of each case is highlighted in bold.*

dataset	ICM		ICEPS	
	minimum	error	minimum	error
1	17039	3.06%	**14511 ± 18**	**2.85% ± 0.01%**
2	21936	3.31%	**19206 ± 10**	**2.92% ± 0.02%**
3	17758	3.11%	**15264 ± 41**	**2.74% ± 0.01%**
4	18130	2.56%	**15580 ± 2**	**2.22% ± 0.01%**

Table 4. *Results of the compound MRF model using two different solvers for multi-class segmentation on IBSR data. The lower energy and error rate of each case is highlighted in bold.*

From the above experimental results we can find both ICM and ICEPS are general solvers but ICM is inclined to be stuck in very local minima (fairly high energy). GC improves this problem a lot by finding more global optimum (fairly low energy) but it has the limitation of applicability. Not only can ICEPS find low energy (even lower than GC) but also can be applied to arbitrary energy functions (e.g. the complicated boundary model above). The effectiveness of ICEPS is shown in the quantitative comparison of segmentation error, which is calculated by dividing the number of misclassified pixels by the number of total pixels. It can be found that ICEPS is able to work better both for simple and complex models than a commonly-used optimization method and a state-of-the-art one.

The running time of the ICM algorithm is less than ten seconds. GC takes a little longer than ICM. ICEPS finishes the job within five minutes. The computation time of SA is expected to be a few hours. However, since the architecture of ICEPS algorithm is inherently parallel and its steps are all simple, it has large potential to speed up by parallel processing or algorithm optimization.

4. CONCLUSION

In this paper, we propose a new solver, namely iterated cross entropy with partition strategy (ICEPS), for Markov random field modeling, which is an important tool not only in image segmentation but in medical image analysis as well. The proposed solver is shown to be general because it is applicable to arbitrary objective functions. Thanks to its stochastic property, ICEPS is capable of finding lower energy than some commonly-used or state-of-the-art solvers. The contribution of ICEPS lies in that it is appropriate for any type of MRF models especially those complex models that GC and BP cannot be used. Moreover, it is able to find very low energy and thus achieve better segmentation results.

5. REFERENCES

[1] Y.Boykov, O.Veksler, and R.Zabin, "Fast approximate energy minimization via graph cuts," *IEEE PAMI*, vol. 23, pp. 1222–1239, 2001.

[2] S.Z.Li, Ed., *Markov Random Field Modeling in Image Analysis*, Springer-Verlag Tokyo, 2001.

[3] S.Kirkpatrick, D.Gellatt, and M.Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671–680, 1983.

[4] J.Besag, "On the statistical analysis of dirty pictures," *Journ. of the Roy. Stat. Soc.*, vol. Series B 48, pp. 259–302, 1986.

[5] J.S.Yedidia, W.T.Freeman, and Y.Weiss, "Understanding belief propagation and its generalizations," *Technical Report, Mitsubishi Electric Research*, vol. TR-2001-22, 2001.

[6] R.Rubinstein, "Optimization of computer simulation models with rare events," *European Journal of Operations Research*, vol. 99, pp. 89–112, 1997.

[7] J.Wu and A.C.S.Chung, "Cross entropy: A new solver for markov random field modeling and applications to medical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention*, 2005, vol. 2, pp. 229 – 237.

[8] R.Rubinstein and D.Kroese, Eds., *The Cross-Entropy Method: A Unified Approach to Cominatorial Optimization, Monte-Carlo Simulation and Machine Learning*, Springer-Verlag, New York, 2004.

[9] J.Wu and A.C.S.Chung, "A segmentation method using compound markov random fields based on a general boundary model," in *International Conference on Image Processing*, 2005, vol. 2, pp. 1182 – 1185.

[10] BrainWeb, "http://www.bic.mni.mcgill.ca/brainweb/," .

[11] The Internet Brain Segmentation Repository, "http://www.cma.mgh.harvard.edu/ibsr/," .