

Evaluation of a Cardiac Ultrasound Segmentation Algorithm Using a Phantom

Yong Yue¹, Hemant D. Tagare^{1,2}, Ernest L. Madsen³, Gary R. Frank³,
and Maritza A. Hobson³

¹ Department of Diagnostic Radiology, School of Medicine, Yale University, New Haven, CT 06511 USA*

² Department of Biomedical Engineering, Yale University, New Haven, CT 06520

³ Department of Medical Physics, University of Wisconsin-Madison, Madison, WI 53706-1532, USA

Abstract. This paper evaluates the performance of a level set algorithm for segmenting the endocardium in short-axis ultrasound images. The evaluation is carried out using an anthropomorphic ultrasound phantom. Details of the phantom design, including comparison of the ultrasound parameters with in-vitro measurements, are included.

In addition to measuring segmentation accuracy, the effectiveness of the energy minimization scheme is also determined. It is argued that using the phantom along with global minimization algorithms (simulated annealing and random search) makes it possible to assess the minimization strategy.

1 Introduction

This paper has two goals: The first is to comprehensively evaluate a level set algorithm for segmenting the endocardium in B-mode short-axis cardiac ultrasound images. The second is to propose an extension of the standard evaluation methodology. The extension provides an explanation of what limits the ultimate performance of the segmentation algorithm. It does so by providing insight into the relative effectiveness of the energy function and the energy minimization strategy.

The segmentation algorithm we evaluate is a maximum-a-posteriori (M.A.P.) level-set scheme [1] using Gamma probability densities [2] using an energy minimization strategy called *tunneling descent* [3]. Tunneling descent is a deterministic technique derived from gradient descent, but is capable of escaping from poor local minima. Although tunneling descent does not guarantee a global minimum, its performance in practice is superior to gradient descent [3].

1.1 Evaluating and Explaining Accuracy

Most ultrasound segmentation algorithms are evaluated by comparison with manual segmentation [4,5,6,7]. The algorithm is considered validated if its

* This research was supported by the grant R01 HL077810-02 from National Heart, Lung, and Blood Institute.

performance measures are within the variation range of manual segmentation results. Although popular, evaluation with manual segmentation may suffer from observer bias and inter- and intraobserver variability.

Acknowledging the limitations of manual segmentation, some authors have attempted to access the ground truth of the ultrasound image. A number of researchers [8,9] use synthetic images which provide ground truth. Some researchers use physical phantoms [10], but the number of such studies is surprisingly low. This may be due to the difficulty of constructing a physically realistic phantom and the low interaction between the image processing and phantom design communities. Almost all of the reported cardiac-like ultrasound phantoms, are designed for tissue characterization or strain measurements [11,12,13] rather than for assessing segmentation of the boundary of the myocardium.

We evaluate the segmentation algorithm by using an anthropomorphic phantom whose geometry mimics short-axis cardiac geometry. Simulated ribs are also included in the phantom. The phantom provides an operator-free ground truth to assess the accuracy of the algorithm.

Having measured accuracy, we additionally want to know which part of the algorithm is responsible for limiting the accuracy. A segmentation algorithm can be limited in accuracy because either its energy function is a simplified approximation to the “exact” energy function (and hence ultimately not capable of accurate segmentation), or because the energy minimization is stuck at a poor local minimum.

We propose to distinguish between the two by using a global minimization algorithm. The idea is to segment the phantom images first using the algorithm, and then using a global minimizer with the same energy function. If the global minimizer segmentation differs substantially from the ground truth, then we may be suspicious of the energy function because these are the best segmentations that it is capable of. On the other hand, if the segmentations of the global minimizers are close to the ground truth but those of tunneling descent are not, then inaccuracy can be attributed to the optimization strategy.

2 Cardiac Segmentation Using Tunneling Descent

Due to lack of space, we only give a brief sketch of the segmentation algorithm here. The reader is encouraged to refer to [3] for details. The energy function J of the algorithm is the negative posterior likelihood of the image gray levels:

$$J(\phi, \beta_0, \beta_1) = \lambda_1 \left\{ \int (1 - H(\phi)) \log p_1(I | \alpha_1, \beta_1) dA - \int H(\phi) p_0(I | \alpha_0, \beta_0) dA \right\} \\ - \lambda_1 \log p_\beta(\beta_0, \beta_1) + \lambda_2 \int \delta_0(\phi) |\nabla \phi| dA,$$

where, ϕ is the level set function, H is Heaviside function, and δ_0 is the Dirac delta function. Further, $p_0(I | \alpha_0, \beta_0)$ and $p_1(I | \alpha_1, \beta_1)$ are Gamma distributions modeling the gray levels of ventricular cavity and myocardial tissue [2] and α_0, α_1 are the shape parameters and β_0, β_1 are the scale parameters of the Gamma

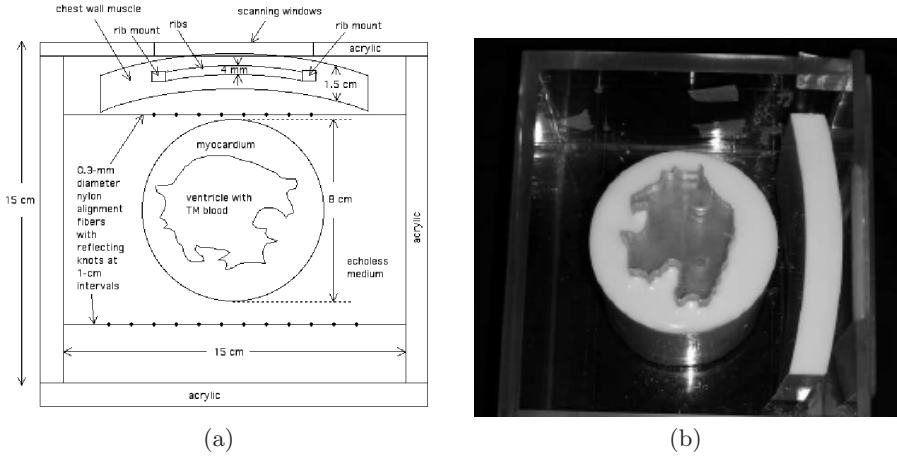


Fig. 1. Ultrasound Phantom. (a) A view of the two-dimensional phantom along the direction of translational invariance. The phantom is scanned through the scanning window at the top with the scan plane parallel to the plane of the figure. (b) The mold in the process of being cast.

distribution. The shape parameters are assumed to be known, while the scale parameters are estimated from the data, and the term $p_{\beta}(\beta_0, \beta_1)$ is the prior on β s. Finally, λ_1 is the external energy parameter, and λ_2 is the internal energy parameter.

Since α_0, α_1 are known, the task of the segmentation algorithm is find the ϕ, β_0, β_1 that minimize J . Minimization with respect to the β s is carried using standard gradient descent while minimization with respect to ϕ is carried out using a technique called *Tunneling Descent* (TD). Reference [3] contains a detailed description of this method including an evaluation of the sensitivity to parameters and initial conditions.

3 Ultrasound Phantom

We designed the ultrasound phantom shown in Fig. 1 for the short-axis section of the left ventricle at end diastole. Fig. 1(a) shows the design and Fig. 1(b) shows the tissue-mimicking (TM) myocardium part of the phantom during casting.

The phantom is essentially two-dimensional; it only models the short-axis cross section of the ventricle rather than the entire three dimensional shape. The phantom has two simulated ribs and two pairs of parallel fibers for aligning the scan plane of the ultrasound transducer. Each alignment fiber is a 0.3 mm diameter nylon fiber with a simple knot at 1 cm intervals along its length.

The phantom has four components: 1) a TM chest wall with simulated ribs embedded in it; 2) the TM myocardium; 3) blood; 4) a low echo TM background. All of these components are made from water, agarose, ultrafiltered whole bovine

Table 1. Ultrasonic properties of the soft-tissue mimicking materials in the phantom

TM Material	Speed (m/s)	Attenuation Coeff.		Backscatter Coeff.	
		α_0 (dB cm ⁻¹ MHz ⁻ⁿ)	n	η_0 (Sr ⁻¹ cm ⁻¹ MHz ^{-m})	m
Chest Wall	1541	0.580	1.08	1.79×10^{-6} *	3.67 *
Myocardium	1543	0.435	1.12	1.79×10^{-6}	3.67
Blood	1534	0.215	1.18	1.34×10^{-7}	3.65
Low echo background	1541	0.393	1.12	4.48×10^{-7} *	3.67 *

*Values deduced based on concentrations of glass bead scatterers relative to the same type of beads (same diameter distribution) contained in the TM myocardium.

milk, n-propanol, glass bead scatterers (7 micron diameter in blood and 22 micron diameter in the rest of the phantom [14]). Ultrasonic properties of the materials are given in Table 1. The propagation speeds and attenuation coefficients were measured by the method of [15], and the backscatter coefficients by [16].

4 Finding the Global Minimum

Recall that we intend to use global minimizers to understand what limits segmentation accuracy. In fact, we use two global minimizers: simulated annealing and random search.

4.1 Simulated Annealing (SA)

Simulated annealing (**SA**) [17,18] is a well-known minimizer which works by making random moves on the energy landscape. To minimize $J(\phi, \beta_0, \beta_1)$ the random moves correspond to perturbing the level set according to $\phi^{[n]} = \phi^{[n-1]} + \Delta\phi$. Downhill perturbations are always accepted; uphill perturbations are accepted with a probability that depends on the algorithm’s “temperature” which varies according to a “schedule”. We use three different temperature schedules. They all start from an initial temperature T_1 and reach a final temperature T_N in a fixed number of iterations, N . The first schedule is a linear schedule, called **SA TS1**, $T_n = T_1 - (n-1)\frac{T_1-T_N}{N-1}$, where N is the total number of iterations. The second is a slow decay nonlinear schedule, called **SA TS2**, $T_n = a(T_1 - T_N)(1 - \arctan(\frac{n-1}{N-1}b - c)) + T_N$, where a, b, c are parameters controlling decay speed, $a = 1/(1 + \arctan(c))$ and $b = 1.5574 + c$. The third is a fast decay nonlinear schedule, called **SA TS3**, $T_n = T_1(\frac{T_N}{T_1})^{(n-1)/(N-1)}$.

4.2 Random Search (RS)

This is a heuristic algorithm developed by us. The basic idea is to maintain a set of m feasible energy minimizers which is updated at every iteration such that the new set has better minimizers than the previous set. Denoting this set

Table 2. Random research

Initialize $A^{[1]} = \{\phi_i | \phi_i = \phi^{[1]} + \Delta\phi_i, i = 1, \dots, m\}$ where $\phi^{[1]}$ is a given initial level set
 For $n=2:N$ do:
 $B^{[n]} = A^{[n-1]} \cup \{\tilde{\phi} | \tilde{\phi} = \phi + \Delta\phi, \phi \in A^{[n-1]}\}$
 For every $\phi \in B^{[n]}$ calculate $\beta_0, \beta_1 = \arg \min_{\beta_0, \beta_1} J(\phi, \beta_0, \beta_1)$
 Sort all $\phi \in B^{[n]}$ according to their energy $J(\phi, \hat{\beta}_0, \hat{\beta}_1)$.
 Set $A^{[n]}$ to the m smallest energy ϕ 's in $B^{[n-1]}$.
 End do
 Return the smallest energy ϕ in $A^{[N]}$.

in the n^{th} iteration as $A^{[n]}$, the recipe for creating $A^{[n]}$ is as follows: all level sets in $A^{[n-1]}$ are randomly perturbed and a new set $B^{[n]}$ is created according to $B^{[n]} = A^{[n-1]} \cup \{\tilde{\phi} | \tilde{\phi} = \phi + \Delta\phi, \phi \in A^{[n-1]}\}$, where $\Delta\phi$ are the random perturbations. $B^{[n]}$ has $2m$ level sets as its elements. These are sorted according to energy and the m lowest energy level sets form $A^{[n]}$. After N iterations, the lowest energy curve in $A^{[N]}$ is taken as the global minimizer. The algorithm is initialized by setting to $A^{[1]}$ to an initial level set plus $m-1$ random perturbations of it. The pseudo-code for the algorithm is listed in Table 2.

4.3 Initialization, Random Perturbations etc.

SA and RS are both initialized at the tunneling descent segmentation. For all SA schedules, the initial and final temperatures are $T_1 = 1000$, $T_N = 1$ with a total of $N = 2000$ iterations. For SA TS2, the parameters are $a = 0.4213$, $b = 6.5574$, and $c = 5.0$. For RS, the size of the feasible solution set $A^{[n]}$ is set to $m = 100$, and the total iterations to $N = 1000$.

The level set random perturbation in both algorithms is taken to be a weighted sum of Gaussian radial basis functions. Thus, $\Delta\phi = \sum_j^{N_{rbf}} \omega_j \phi_j$, where N_{rbf} is the total number of basis functions, ω_j is the amplitude of the RBF, and $\phi_j(x) = \exp(-\frac{\|x - \mu_j\|^2}{\sigma_j^2})$, where μ_j and σ_j are the center and the standard deviation. The parameters $\omega_j, \mu_j, \sigma_j$ are generated randomly. The variances σ_j are uniformly chosen from the range $\sigma_j = 6 \sim 15$. By informal experimentation we determined that this range of σ_j gave perturbations of the speckle size in the image. We also determined that $N_{rbf} = 100$ gave sufficient basis functions for a rich perturbation.

4.4 The Best Optimizer

Typically global minimizers give the global minimum after infinite iterations. In practice, they have to be terminated after a finite number of iterations and they do not all find the same minimum at termination. For any image, we call the global minimizer that has found the lowest energy, the *best minimizer* (b.o.).

Thus, $b.o. \in \{RS, SA\ TS1, SA\ TS2, SA\ TS3\}$. We use the best minimizer results for comparison with tunneling descent.

5 Experiments and Discussion

The phantom was imaged by a Phillips SONOS 7500 ultrasound system at 3.0, 5.5 and 7.5 MHz settings. The speckle size at these frequencies were 11, 9, and 8 pixels respectively. The 3.0 and 7.5 MHz images are shown in left column of Fig. 2.

All images were processed in the following way: First, the boundary of the endocardium in the image was segmented using tunneling descent. For each image, tunneling descent was deployed with three different internal energy parameters ($\lambda_2 = 40, 60, 80$), and a fixed external energy parameter ($\lambda_1 = 500$). Thus, for each image, nine segmentation experiments were performed and indexed as experiments 1 through 9 (see Table 3). After the tunneling descent segmentation, the global minimizers RS, SA TS1, TS2, and TS3 were used and the best optimizer found for each image.

The middle column of Fig. 2 shows the tunneling descent segmentation overlaid with the best optimizer segmentation. The right column shows the best optimizer segmentation overlaid with the phantom ground truth (GT).

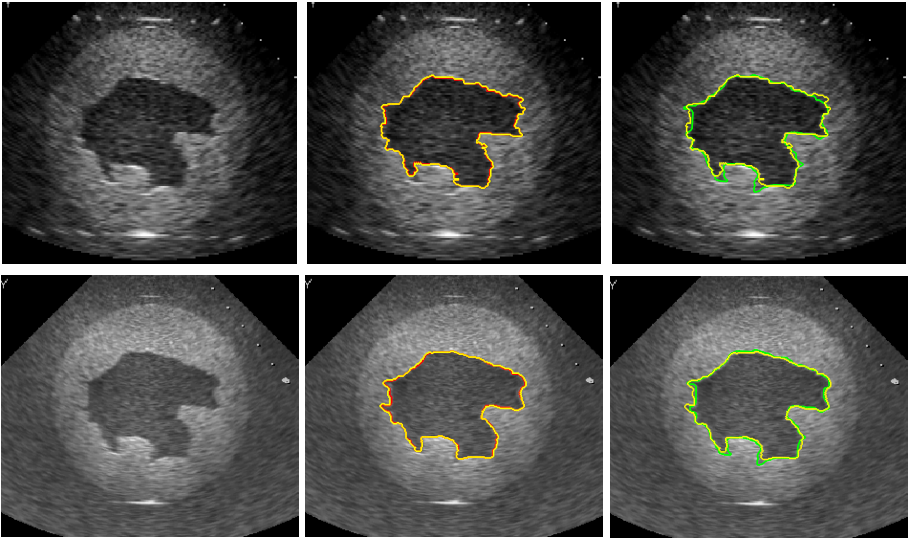


Fig. 2. Left column: ultrasound images of the phantom, acquired at 3.0 MHz(top), and 7.5 MHz (bottom). Middle column: segmentations of the best optimizers (yellow) overlaid on the segmentations of the tunneling descent (red) for the experiment 2 and and 8 (see Table 3). Right column: the segmentations of the best optimizers (yellow) overlaid on the ground truth (green).

Table 3. Evaluation of the segmentation using the ultrasound images of the phantom

Exp	3.0 MHz			5.5 MHz			7.5 MHz		
	1	2	3	4	5	6	7	8	9
λ_2^*	40	60	80	40	60	80	40	60	80
HD (GT,TD)**	10.512	10.784	10.736	15.497	15.046	14.965	9.067	9.071	8.843
MAD (GT,TD)	1.872	1.823	1.864	1.912	1.811	1.775	1.562	1.482	1.389
HD (GT,b.o.)	11.436	9.785	10.722	14.593	13.668	14.681	9.639	8.963	8.501
MAD (GT,b.o.)	2.185	1.983	2.073	1.490	1.675	1.617	1.379	1.396	1.287
ΔE (TD,b.o.)	-0.016	-0.015	-0.014	-0.015	-0.022	-0.0173	-0.013	-0.021	-0.019
HD (TD,b.o.)	5.099	3.162	4.000	5.385	3.000	2.2361	2.000	2.236	1.000
MAD (TD,b.o.)	0.718	0.557	0.464	0.928	0.554	0.549	0.601	0.512	0.415

* $\lambda_1 = 500$ for all experiments. ** HD and MAD are measured in pixels.

The accuracy of the segmentations was evaluated by using the ground truth of the phantom. Two criteria were used for comparison: (1) The Hausdorff Distance (HD) between the segmentation and the ground truth $HD(C_1, C_2) = \max_{a \in C_1} \min_{b \in C_2} \|a - b\|$, and (2) the Mean Absolute Distance (MAD): $MAD(C_1, C_2) = 0.5 * [\frac{1}{n} \sum_{i=1}^n d(a_i, C_2) + \frac{1}{m} \sum_{i=1}^m d(b_i, C_1)]$, with $d(a_i, C_2) = \min_j \|b_j - a_i\|$. In these formulae, C_1, C_2 are the two curves being compared.

The top four rows of Table 3 show the accuracy of tunneling descent and the best optimizer in comparison to ground truth. Since the speckle size in these images is $8 \sim 11$ pixels, in all cases, tunneling descent and the best optimizer segmentations have an accuracy of one to two times the speckle size.

Next, recall from the discussion in section 1 that we are interested in not only evaluating accuracy, but also in understanding what limits the accuracy. We argued in section 1 that this could be done by comparing tunneling, the best optimizer, and the ground truth. We carried out this comparison in two ways. First, we asked whether the best optimizer minimum was significantly lower energy than the tunneling descent minimum. We did this by calculating $\Delta E(TD, b.o.) = (J(\phi_2, \beta_{0,2}, \beta_{1,2}) - J(\phi_1, \beta_{0,1}, \beta_{1,1})) / |J(\phi_2, \beta_{0,2}, \beta_{1,2})|$, where $J(\phi_2, \beta_{0,2}, \beta_{1,2})$ is the b.o. minimum energy and $J(\phi_1, \beta_{0,1}, \beta_{1,1})$ is the TD minimum energy. The fifth row of Table 3 shows ΔE for the nine segmentations. The numbers show that the b.o. minimum energy is smaller than the TD minimum energy by only about 2%.

Second, we compared the geometry of the segmentations found by tunneling descent and the best optimizer by calculating the Hausdorff distance and the mean absolute distance between them. The results are shown in rows six and seven of Table 3. These rows show that Hausdorff distances are less than 6 pixels and the mean absolute distances are less than 1 pixel.

It is clear from rows 5-7 that for all practical purposes, the energy minimum found by tunneling descent is as good as the minimum found by the best optimizer. In turn, this suggests that it is the energy function rather than the optimization strategy that limits the segmentation accuracy. Support for this conclusion also comes from examining the right column of Fig. 2. The figure

shows very vividly that the most significant difference between the b.o. and ground truth is near sharp corners. Apparently, the energy function prevents the segmentation from entering into these sharp corners. If higher accuracy is required, the energy function will have to be modified so that some sharp corners are permitted while still constraining the rest of the segmentation curve to be smooth.

6 Conclusion

In this paper, we evaluated the accuracy of a segmentation algorithm with a phantom study. We also argued that it is important to understand the origin of any segmentation inaccuracy, and that the origin could be understood by using phantoms with global minimizers. The experiments show that TD is accurate to one to two times the speckle size, and the TD finds a minimum that is very close to the global minimum. Furthermore, the residual inaccuracy of the segmentation can be attributed to the energy function rather than the minimization scheme.

References

1. Osher, S., Paragios, N.: *Geometric Level Set Methods in Imaging, Vision, and Graphics*. Springer, New York (2003)
2. Tao, Z., Tagare, H., Beaty, J.: Evaluation of four probability distribution models for speckle in clinical cardiac ultrasound images. *IEEE Trans. on Medical Imaging* 25, 1483–1491 (2006)
3. Tao, Z., Tagare, H.: Tunneling descent for m.a.p. active contours in ultrasound segmentation. *Med. Image Anal.* 11, 266–281 (2007)
4. Angelini, E.D., Laine, A.F., Takuma, S., Holmes, J.W., Homma, S.: Lv volume quantification via spatiotemporal analysis of real-time 3-d echocardiography. *IEEE Trans. Med. Imag.* 20(6), 457–469 (2001)
5. Lin, N., Yu, W.C., Duncan, J.S.: Combinative multi-scale level set framework for echocardiographic image segmentation. *Med. Image Anal.* 7, 529–537 (2003)
6. Sarti, A., Corsi, C., Mazzini, E., Lamberti, C.: Maximum likelihood segmentation of ultrasound images with rayleigh distribution. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 52, 947–960 (2005)
7. Dydenko, I., Jamal, F., Bernard, O., Dhooge, J., Magnin, I., Friboulet, D.: A level set framework with a shape and motion prior for segmentation and region tracking in echocardiography. *Med. Image Anal.* 10, 162–177 (2006)
8. Boukerroui, D., Basset, O., Baskurt, A., Gimenez, G.: A multiparametric and multiresolution segmentation algorithm of 3-d ultrasonic data. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 48, 64–77 (2001)
9. Jardim, S., Figueiredo, M.: Segmentation of fetal ultrasound images. *Ultrasound Med. Biol.* 31, 243–250 (2005)
10. Xiao, G.F., Brady, M., Noble, J.A., Zhang, Y.Y.: Segmentation of ultrasound b-mode images with intensity inhomogeneity correction. *IEEE Trans. Med. Imag.* 21, 48–57 (2002)
11. Mottley, J.G., Miller, J.G.: Anisotropy of the ultrasonic attenuation in soft tissue: Measurement in vitro. *J. Acoust. Soc. Am.* 88(3), 1203–1210 (1990)

12. Langeland, S., D'Hooge, J., Claessens, T., Claus, P., Verdonck, P., Suetens, P., Sutherland, G.R., Bijmens, B.: R.f.-based two-dimensional cardiac strain estimation: a validation study in tissue-mimicking phantom. *I.E.E.E. Trans. Ultra. Ferro. Freq. Contrl.* 51(11), 1537–1546 (2004)
13. Erpelding, T.N., Hollman, K.W., O'Donnell, M.: Bubble-based acoustic radiation force elasticity imaging. *I.E.E.E. Trans. Ultra. Ferro. Freq. Contrl.* 52(6), 971–979 (2005)
14. Madsen, E.L., Frank, G.R., Dong, F.: Liquid or solid ultrasonically tissue-mimicking materials with very low scatter. *Ultrasound in Med. Biol.* 24, 535–542 (1998)
15. Wear, K.A., Stiles, T.A., Frank, G.R., Madsen, E.L.: Interlaboratory comparison of ultrasonic backscatter coefficient measurements from 2 to 9 MHz. *J. Ultrasound in Med.*, 1235–1250 (2005)
16. Chen, J., Zagzebski, J., Madsen, E.L.: Tests of backscatter coefficient measurement using broadband pulses. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control* 40, 603–607 (1993)
17. Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E.: Equations of state calculations by fast computational machine. *Journal of Chemical Physics* 21, 1087–1091 (1953)
18. Kirkpatrick, S., Gelatt, C., Vecchi, M.: Optimization by simulated annealing. *Science* 220, 671–680 (1983)