The Genetic Algorithm-Based Optimization Approach for Gamma Unit Treatment

Sun Fei
Yang Lin
Wang Hong
Donghua University
Shanghai, China

Advisor: Ding Yongsheng

Abstract

The gamma knife is used to treat brain tumors with radiation. The treatment planning process determines where to center the shots, how long to expose them, and what size focusing helmets should be used, to cover the target with sufficient dosage without overdosing normal tissue or surrounding sensitive structures.

We formulate the optimal treatment planning for a gamma-knife unit as a sphere-packing problem and propose a new genetic algorithm (GA)-based optimization approach for it. Considering the physical limitations and biological uncertainties involved, we outline a reasonable, efficient and robust solution.

First, we design a geometry-based heuristic to produce quickly a reasonable configuration of shot sizes, locations, and number. We first generate the skeleton using a 3D-skeleton algorithm. Then, along the skeleton, we use the GA-based shot placement algorithm to find a best location to place a shot. By continuously iterating the algorithm, we obtain the number, sizes, and the locations of all shots. After that, we develop a dose-based optimization method.

Then we implement simulations of our models in Matlab. We did numerous computer simulations, using different shape or size targets, to examine the effectiveness of our model. From the simulation results, we know that the geometry-based heuristic with the GA optimization approach is a useful tool for in the selection of the appropriate number of shots and helmet sizes. Generally, all of the optimized plans for various targets provide full-target coverage with 90% of the prescription isodose.

Moreover, we do sensitivity analysis to our model in the following aspects:

- the sensitive structures;
- at the 30%, 40%, 60%, or 70% isodose level;
the issue of global versus local optimality;
• conformity; and
• the robustness. We also discuss the strengths and limitations of our model.

The results indicate that our approach is sufficiently robust and effective to be used in practice. In future work, we would fit ellipsoids instead of spheres, since some researchers note that the dose is skewed in certain directions.

Introduction

The gamma knife unit delivers ionizing radiation from 201 cobalt-60 sources through a heavy helmet. All beams simultaneously intersect at the isocenter, resulting in an approximately spherical dose distribution at the effective dose levels. Delivering dose is termed a shot, and a shot can be represented as a sphere. Four interchangeable outer collimator helmets with beam channel diameters of 4, 8, 14, and 18 mm are available. For a target volume larger than one shot, multiple shots can be used.

Gamma knife treatment plans are conventionally produced using a manual iterative approach. In each iteration, the planner attempts to determine
• the number of shots,
• the shot sizes,
• the shot locations, and
• the shot exposure times (weights) that would adequately cover the target and spare critical structures.

For large or irregularly shaped treatment volumes, this process is tedious and time-consuming, and the quality of the plan produced often depends on both the patience and the experience of the user. Consequently, a number of researchers have studied techniques for automating the gamma knife treatment planning process [Wu and Bourland 2000a; Shu et al. 1998]. The algorithms that have been tested include simulated annealing [Leichtman et al. 2000; Zhang et al. 2001], mixed integer programming, and nonlinear programming [Ferris et al. 2002; 2003; Shepard et al. 2000; Ferris and Shepard 2000].

The objective is to deliver a homogeneous (uniform) dose of radiation to the tumor (the target) area while avoiding unnecessary damage to the surrounding tissue and organs. Approximating each shot as a sphere [Cho et al. 1998] reduces the problem to one of geometric coverage. Kike [Liu and Tang 1997], we formulate optimal treatment planning as a sphere-packing problem and we propose an algorithm to determine shot locations and sizes.
Assumptions

To account for all physical limitations and biological uncertainties involved in the gamma knife therapy process, we make several assumptions as follows:

A1: The shape of the target is not too irregular, and the target volume is a bounded. As a rule of thumb, the target to be treated should be less than 35 mm in all dimensions. Its three-dimensional (3D) digital image, usually consisting of millions of points, can be obtained from a CT or MRI.

A2: We consider the target volume as a 3D grid of points and divide this grid into two subsets, the subset of points in and out of the target, denoted by $T$ and $N$, respectively.

A3: Four interchangeable outer collimator helmets with beam channel diameters $w = \{4, 8, 14, 18\}$ mm are available for irradiating different size volumes. We use $(x_s, y_s, z_s)$ to denote the coordinates of the center location of the shot and $t_{s,w}$ to denote the time (weight) that each shot is exposed. The total dose delivered is a linear function of $t_{s,w}$. For a target volume larger than one shot, multiple shots can be used to cover the entire target. There is a bound $n$ on the number of shots, with typically $n \leq 15$.

A4: Neurosurgeons commonly use isodose curves as a means of judging the quality of a treatment plan; they may require that the entire target is surrounded by an isodose line of $x\%$, e.g., 30–70%. We use an isodose line of 50%, which means that the 50% line must surround the target.

A5: The dose cloud is approximated as a spherically symmetric distribution by averaging the profiles along $x$, $y$, and $z$ axes. Other effects are ignored.

A6 The total dose deposited in the target and critical organ should be more than a fraction $P$ of the total dose delivered; typically, $25\% \leq P \leq 40\%$.

Optimization Models

Analysis of the Problem

The goal of radiosurgery is to deplete tumor cells while preserving normal structures. An optimal treatment plan is designed to:

R1: match specified isodose contours to the target volumes;
R2: match specified dose-volume constraints of the target and critical organ;
R3: constrain dose to specified normal tissue points below tolerance doses;
R4: minimize the integral dose to the entire volume of normal tissues or organs;
R5: minimize the dose gradient across the target volume; and
R6: minimize the maximum dose to critical volumes.

It also is constrained to
C1: prohibit shots from protruding outside the target,
C2: prohibit shots from overlapping (to avoid hot spots),
C3: cover the target volume with effective dosage as much as possible (at least 90% of the target volume must be covered by shots), and
C4: use as few shots as possible.

We design the optimal treatment plan in two steps.

- We use a geometry-based heuristic to produce a reasonable configuration of shot number, sizes and locations.
- We use a dose-based optimization to produce the final treatment plan.

**Geometry-Based Heuristic for Sphere-Packing**

We model each shot as a sphere, and we use the medial axis transform (known as the skeleton) of the target volume to guide placement of the shots. The skeleton is frequently used in shape analysis and other related areas [Wu et al. 1996; Wu and Bourland 2000b; Zhou et al. 1998]. We use the skeleton just to find good locations of the shots quickly. The heuristic is in three stages:

- We generate the skeleton using a 3D skeleton algorithm.
- We place shots and choose their sizes along the skeleton to maximize a measure of our objective; this process is done by a genetic algorithm (GA)-based shot placement approach.
- After the number of focusing helmets to be included in the treatment plan is decided, the planning produces a list of the possible helmet combinations and a suggested number of shots to use.

**Skeleton Generation**

We adopt a 3D skeleton algorithm that follows similar procedures to Ferris et al. [2002]. We use a morphologic thinning approach [Wu 2000] to create the skeleton, as opposed to the Euclidean-distance technique. The first step in the skeleton generation is to compute the contour map containing distance information from the point to a nearest target boundary. Then, based on the contour map, several known skeleton extraction methods [Ferris et al. 2002; Wu et al. 1996; Wu and Bourland 2000b; Zhou et al. 1998; Wu 2000] can be used. Since the method in Ferris et al. [2002] is simple and fast, we use it.
Genetic Algorithm-Based Shot Placement

We restrict our attention to points on the skeleton. We start from a special type of skeleton point, an endpoint (Figure 1): A point in the skeleton is an endpoint if it has only one neighbor in the skeleton.

Figure 1. Examples of endpoints.

Starting from an endpoint, we look for the best point to place a shot and determine the shot size by using GA [Goldberg 1989; Mann et al. 1997]. In the GA-based shot placement algorithm, we must solve the following problems:

The encoding method. In general, bit-string (0s and 1s) encoding is the most common method adopted by GA researchers because of its simplicity and tractability. However, in this case, if we directly encode the point coordinates \((x_s, y_s, z_s)\) into a bitstring, crossover and mutation generate some points that are not in the skeleton. To solve this problem, we build a table of correspondence between the point coordinates and the point number (1 to \(M\)); instead of encoding the point, we encode the point number. We select \(m\) points from all points of the skeleton to form a population; a single point is a chromosome.

Performance evaluation. The key to the GA-based approach is the fitness function. Ideally, we would like to place shots that cover the entire region without overdosing within (or outside) of the target. Overdosing occurs outside the target if we choose a shot size that is too large for the current location, and hence the shot protrudes from the target. Overdosing occurs within the target if we place two shots too close together for their chosen sizes.

Before defining a fitness function, we give some definitions:

- **Fraction**: A target part that is not large enough to be destroyed by the smallest shot without any harm to the surrounding normal tissue.
- **Span**: The minimum distance between the current location and the endpoint at which we started.
- **Radius**: The approximate Euclidean distance to the target boundary.

We would like to ensure that the span, the radius, and the shot size \(w\) are as close as possible. Therefore, we choose a fitness function that is the sum of the squared differences between these three quantities. The fitness function can ensure that the generating fraction is the smallest after every shot is placed on the target [Ferris et al. 2002]:

\[
\text{Fit} = \phi_{s,r}(x, y, z) + \phi_{s,w}(x, y, z) + \phi_{r,w}(x, y, z),
\]
where
\[
\phi_{s,r}(x, y, z) = \left[ \text{span}(x, y, z) - \text{radius}(x, y, z) \right]^2,
\]
\[
\phi_{s,w}(x, y, z) = \left[ \text{span}(x, y, z) - w \right]^2,
\]
\[
\phi_{r,w}(x, y, z) = \left[ \text{radius}(x, y, z) - w \right]^2.
\]

- Equation (1) ensures that we pack the target volume as well as possible, that is, the current span between shots should be close to the distance to the closest target boundary.
- Equation (2) is used to choose a helmet size that fits the skeleton best for the current location.
- Equation (3) favors a location that is the appropriate distance from the target boundary for the current shot size.

**Genetic operators.** Based on the encoding method, we develop the genetic operators in the GA: crossover and mutation.

- Crossover/recombination is a process of exchanging genetic information. We adopt one-point crossover operation; the crossover points are randomly set.
- Mutation operation. Any change in a gene is called a *mutation*; we use point mutation.

We propose the following GA-based shot placement algorithm:

1. Find a skeleton and all its endpoints. Take one of the endpoints as a starting point.
2. Randomly search all the points in the skeleton using the GA to find the location and size of the best shots as follows:
   (a) Generate randomly \( m (= 100) \) points from all the points (e.g., \( M = 1000 \)) in the skeleton. The \( m \) points are the chromosomes. Set the crossover rate \( p_c = .95 \), the mutation rate \( p_m = .05 \), the desired Fit function, and the number \( n_g \) of generations.
   (b) Calculate the Fit of all the \( m \) points in the skeleton.
   (c) If the algorithm has run \( n_g \) steps, or if one of the \( m \) points satisfies the desired Fit, the GA stops (at this time, the best shot is chosen); else encode \( m \) points into the bit strings.
   (d) To do crossover and mutation operation to the \( m \) bitstrings, go to 2b.
3. Considering the rest of the target in whole as a new target, repeat Steps 1–2 until the rest of the target are fractions (at this time, all best shots are found).
Dose-Based Optimization

After we obtain the number, sizes and the locations of shots, we develop a dose-based optimization method.

We determine a functional form for the dose delivered at a point \((i, j, k)\) from the shot centered at \((x_s, y_s, z_s)\). The complete dose distribution can be calculated as a sum of contributions from all of the individual shots of radiation:

\[
D(i, j, k) = \sum_{(s, w) \in S \times W} t_{s, w} D_{s, w}(x_s, y_s, z_s, i, j, k),
\]

where \(D_{s, w}(x_s, y_s, z_s, i, j, k)\) is the dose delivered to \((i, j, k)\) by the shot of size \(w\) centered at \((x_s, y_s, z_s)\) with a delivery duration of \(t_{s, w}\). Since \(D_{s, w}\) is a complicated (nonconvex) function, we approximate it by

\[
D_{s, w}(x_s, y_s, z_s, i, j, k) = \sum_{i=1}^{2} \lambda_i \left( 1 - \int_{-\infty}^{x} \frac{1}{\sqrt{\pi}} e^{-x^2} dx \right),
\]

where \(x = (t - r_i)/\sigma_i\) and \(\lambda_i, \gamma_i, \text{ and } \sigma_i\) are coefficients [Ferris and Shepard 2000].

To meet the requirement of matching specified isodose contours to target volume at the 50% isodose line, the optimization formulation should impose a constraint on the 50% isodose line that must surround the target. We impose strict lower and upper bounds on the dose allowed in the target, namely, for all \((i, j, k) \in T\), the dose \(D(i, j, k)\) satisfies

\[
0.5 \leq D(i, j, k) \leq 2.
\]

To meet the requirement (R2) to match specified dose-volume constraints of the target and critical organ, based on assumption (A3) (which sets out the size of the beam channel diameters), no more than \(n\) shots are to be used; so in each card (any section of the target) we have

\[
\text{card}\left[\{(s, w) \in S \times W \mid t_{s, w} > 0\}\right] \leq n.
\]

The value of tolerance doses of the normal tissue points is \(q\), \(D(i, j, k) < q\) for all \((i, j, k)\). The number of shots \(n\) is no more than 15, so the tolerance doses of a specified normal tissue point should be \(q = 15/201 = 7.46\%\), or

\[
0 \leq D(i, j, k) < q = 7.46\%,
\]

for all \((i, j, k) \in N\).

To meet requirement (R3) (keep the does at normal tissue below a certain level), based on assumption (A6) (which sets the does levels), the tolerance dose radio of the total dose deposited in the target and critical organ to the total dose delivered by a plan is

\[
\frac{\sum_{(i,j,k) \in T} D(i, j, k)}{\sum_{(i,j,k) \in T \cup N} D(i, j, k)} \geq P, \quad P \in [.25, .40].
\]
We wish to satisfy constraint (C3) (at least 90% of the target volume must be covered). We set

\[ V = \text{the total volume of target}; \]

\[ V_s = \text{the total effective dosage volume of the target whose dose value at the point is more than 0.5}; \]

\[ f = \text{the effective dosage rate, which satisfies the inequality} \]

\[ 90% \leq f = \frac{V_s}{V} \leq 100%. \] (9)

The exposure time of each shot \( t_{s,w} \) should be nonnegative:

\[ t_{s,w} \geq 0. \] (10)

We introduce a binary variable \( \delta_{s,w} \) that indicates whether shot \( s \) uses width \( w \) or not, i.e.,

\[ \delta_{s,w} = \begin{cases} 1, & \text{if shot } s \text{ uses width } w; \\ 0, & \text{otherwise.} \end{cases} \]

Moreover, we have the constraints (C1) (no shots protrude outside the target), (C2) (shots do not overlap), and (C4) (as few shots as necessary).

Given all these constraints (5–(10), and based on the requirement (R4) (minimize dose to normal tissue), the goal is to minimize the dose outside of the target.

Also, to meet the requirement (R5) (minimize the dose gradient across the target volume), the treatment plan needs to be both conformal and homogeneous. It is easy to specify homogeneity in the models simply by imposing lower and upper bounds on the dose delivered to points in the target \( T \). Typically, however, the imposition of rigid bounds leads to plans that are overly homogeneous and not conformal enough—that is, they provide too much dose outside the target. To overcome this problem, the notion of underdose (UD) is suggested in Ferris and Shepard [2000]. UD measures how much the delivered dose is below the prescribed dose on the target points. In our models, we either constrain UD to be less than a prespecified value or attempt to minimize the total UD.

In practical application, rather than calculating the dose at every point, it is easy to estimate accurately the total dose delivered by a plan based solely on the \( t_{s,w} \) variables and other precalculated constants. An upper bound is also placed on the dose to the target. Given these constraints, the optimizer seeks to minimize the total underdosage in the target. A point is considered to be underdosed if it receives less than the prescribed isodose \( \theta \), which for the example formulation is assumed to be 1. We actually use the optimization process to model UD, which is constrained to be

\[ UD(i,j,k)m = \max[0, 1 - D(i,j,k)] \]
at every point in the target. We can implement this construct using linear constraints

\[
\begin{align*}
\theta & \leq UD(i, j, k) + D(i, j, k), \\
0 & \leq UD(i, j, k)
\end{align*}
\]

for all \((i, j, k) \in T\).

Our second minimization problem is

**Objective:** \[ \min \sum_{(i,j,k) \in N} UD(i, j, k) \]

subject to the same constraints (5)–(10) as earlier plus (11)–(12).

To meet the requirement (R6) (minimize maximum dose to critical volumes), we have the additional optimization problem

**Objective:** \[ \min \sum_{(i,j,k) \in N} D(i, j, k) \text{ for all } (i, j, k) \in T \text{ for which } \delta_{s,w} = 0 \]

subject to the same constraints (5)–(10) as earlier.

All of the formulations are based on the assumption that the neurosurgeon can determine a priori a realistic upper bound \(n\) on the number of shots needed. Several issues need to be resolved to create models that are practical, implementable, and solvable (in a reasonable time frame). Two main approaches are proposed in the literature [Ferris et al. 2002; 2003; Shepard et al. 2000; Ferris and Shepard 2000], namely mixed integer programming and nonlinear programming, to optimize simultaneously all of the variables.

**Simulation Results and Model Testing**

We developed an optimization package to implement the algorithms of our models in Matlab and perform numerous computer simulations using targets of different shapes and sizes.

To examine its correctness, we plot a dose-volume histogram for the four different helmets using (4), as shown in Figure 2. The histogram depicts the fraction of the volume that receives a particular dose for the target volume. The fit is best for the small shots and decreases slightly in accuracy for the larger ones. The lines show the fraction of the target and critical organ that receives a particular dosage.

Generally speaking, the shape of the target is not too irregular, so we choose five typical shapes of the targets in different sizes. In Figure 3a, we illustrate the maximum section of a typical bean-shaped target, whose maximum dimension is 35 mm. Using the skeleton generation algorithm, we get the corresponding skeleton shown in Figure 3b. Then, we apply the GA-based shot placement algorithm, resulting in three shots for the target: one 14 mm helmet and two
Figure 2. Dose–volume histograms for four different helmets.

Figure 3a. The maximum section of the target.

Figure 3b. The skeleton.

Figure 3c. The locations and sizes of the helmets, in 2D.
8 mm helmets. The locations and sizes of the helmets in 2D are indicated in Figure 3c, while 3D shot placements are shown in Figure 4.

For this target, we also plot six different isodose lines: 30%, 40%, 50%, 60%, 70%, and 100% (Figure 5). The thick (red) line is the target outline, and the thin (black) line is the isodose line. In Figure 5c, the 50% isodose line covers all the points of the target, while in Figure 5f for the 100% isodose line, no point of the shots exceeds the boundary of the target. We also present the 3D shot placements for four other target shapes in Figure 6.

The optimized plans for all of the five shapes of the targets are shown in Table 1, together with the minimum target doses and the percentage coverages. From all of the results, we know that the geometry-based heuristic with the GA optimization approach is a useful tool for assisting in the selection of the appropriate number of shots and helmet sizes. Also, they indicate that our model exceeds the predefined quality of the treatment planning.

### Sensitivity Analysis

- **Can the model be applied to sensitive structures?** Yes, by applying more dose constraints, such as an upper bound on either the mean dose or the maximum dose to the sensitive structures

- **Can we treat the tumor at an isodose level other than 50%?** In Figure 5, with a lower isodose line, the dose outside of the target volume decreases rapidly, resulting in a reduction in the integral dose to normal tissue. With a higher
Figure 5. The specified isodose lines of different values: 30%, 40% 50%; 60%, 70%, 100%.

Table 1.
Optimized plans for five targets.

<table>
<thead>
<tr>
<th>Target (figure)</th>
<th>Maximum section width (mm)</th>
<th>Helmet sizes (mm)</th>
<th>Number of shots</th>
<th>Minimum target dose</th>
<th>Coverage (by isodose) 50%</th>
<th>Coverage (by isodose) 80%</th>
<th>Coverage (by isodose) 100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>35</td>
<td>18 4 1 5</td>
<td>0.51</td>
<td></td>
<td>100% 97% 90%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8a</td>
<td>26</td>
<td>8 4 4 3</td>
<td>0.52</td>
<td></td>
<td>100% 97% 88%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8b</td>
<td>20</td>
<td>14 8 4 1</td>
<td>0.52</td>
<td></td>
<td>100% 96% 92%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8c</td>
<td>10</td>
<td>4 6 1 1</td>
<td>0.45</td>
<td></td>
<td>99% 82% 69%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8d</td>
<td>8</td>
<td>4 2</td>
<td>0.67</td>
<td></td>
<td>100% 97% 57%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 6. Shot placement in four targets.

• The outstanding question from an optimization viewpoint is global vs. local optimality. How about our model? First, the GA-based shot placement algorithm can ensure that the generating fraction is the smallest one after every shot is placed on the target. For the whole target, it also minimizes the sum of all fractions.

• When we have several comparably optimization schemes for shot placement, how do we choose the best one? For example, for a 10 mm-diameter sphere target, we have two comparably optimization schemes, as shown in Table 1. The first places one 8-mm shot and the second places six 4-mm shots. If we consider the treatment merely as a sphere-packing problem, the better choice is the first one. However, in practical treatment, we should consider the diffuse regions where no shot is irradiated. Under the first plan, in these regions the sum of the dose value at a point is more than 1, resulting in increasing the total effective dosage; so we adopt the second plan.

isodose line, the isodose line cannot cover all the points of the target. For the nodular regions or sensitive organs, the higher isodose coverage level should be specified.
• Will there be any points outside the tumor whose dose value is greater than 1? In Figure 7, though the shots have not protruded outside the target (constraint (C1)), some points outside the tumor overdose. This occurs due to the very irregular shape of the target, which is not avoidable. In this case, we should choose how to choose an optimization planning under some constraints.

![Figure 7. The 100% isodose of the target in Figure 6a.](image)

• How about the robustness of our model? For the many cases optimized thus far, high-quality dose distributions have been obtained in all cases.

**Strengths and Limitations**

**Strengths**

• Our optimization-based automated approach generates more-uniform and better treatment plans in less time than is currently used.

• The geometry-based approach is based on skeletonization ideas from computational graphics, which can speed the process of shot placement.

• The GA-based shot placement algorithm can guide the planner in selecting the number of shots of radiation and the appropriate collimator helmet sizes, it can quickly place a shot, and it can ensure global and local optimality simultaneously.

• The model parameters can be tuned to improve solution speed and robustness.

• The graphical interface is an intuitive way to demonstrate the isodose curve and the treatment effects of the planning.

**Limitations**

• The skeleton is a key factor for the effectiveness of the algorithm; we should seek better methods to determine it.
• Whether our model can handle very irregular targets needs to be examined.

• We use the function in Ferris and Shepard [2000] to approximate the dose calculation. Other methods of dose calculation should be examined.

• There is no guarantee that there is not a better treatment plan. Some more-intelligent algorithms, such as a neural network-based dynamic programming algorithm, could be considered.

• We have not tested our model on actual patient data.

Conclusions

From the simulation results, we know that the geometry-based heuristic with the GA optimization approach is a useful tool in the selection of the appropriate number of shots and helmet sizes. Our approach is sufficiently robust and effective to be used in practice.

In future work, we may fit ellipsoids instead of spheres, since some researchers have commented that the dose is skewed in certain directions.

References


