registration, matching, optimization, simulated annealing
similarity measure, non-deterministic methods
cooling schedule, tomographic data

Zdzisław KRÓL*

AN ADAPTIVE STOCHASTIC OPTIMIZATION METHOD FOR MEDICAL REGISTRATION PROBLEM

This paper presents a methodology that addresses important issues concerned with the optimization of the misregistration measures in the volumetric medical data registration problem. Our registration framework uses robust simulated annealing method to handle multiple local minima of the cost function. Our efforts have been centred on obtaining a reliable, efficient and generally applicable method to solve such optimization problems. This has been accomplished through developing an adaptive cooling schedule for the simulated annealing method. The proposed method is very reliable for the estimation of the global minimum in the optimization of objective functions with highly differentiated search space landscapes. We present the detailed description of the method as well as discussion of its advantages and disadvantages.

1. INTRODUCTION

Alignment of two or more complementary tomographic datasets constitutes one of the research mainstreams in the area of medical data processing and is essential for many diagnostic and therapeutic procedures in the modern medicine. The main goal of the registration is to find the global minimum of a continuous real-valued misregistration measure $C : S \to \Re$ (or the cost function in the optimization nomenclature), where S is the search space. The position of the global minimum determines the anatomically optimal matching transformation relating both data volumes. Let us assume that one of the datasets called *object* is geometrically transformed in the registration process to fit the second dataset called *model.* The space where we have to look for a solution of the registration problem is extremely large, even if the datasets to register are of the same modality and they differ slightly. The total number of possibilities for the position, orientation, size and shape of the *object* determine the dimension of the search space. In this work the matching transformation is constrained to rigid motion and only three translations t_x, t_y, t_z and three rotations r_x, r_y, r_z must be determined. Let us simplify the problem to a combinatorial version where the values of the transformation parameters are discrete. For a $N \times N \times N$ *model* the number of possibilities for the position of the centroid of the *object* in the discrete

* Modeling Group, CAESAR - Centre of Advanced European Studies And Research, Ludwig-Erhard-Allee 2, 53175 Bonn, Germany. E-mail: krol@caesar.de

model volume is $O(N^3)$. Since the *model* is discrete, we assume that the number of possible angle-triples for the rotations in 3D is $O(N^3)$. Thus, the size of the search space in the simplest case (rigid motions) is $O(N^6)$. For a data volume of size $256 \times 256 \times 256$, the search space has a size of order 10^{14}. An exhaustive search of this space to find an optimal match between the *object* and the *model* would be practically unacceptable. The cost functions broadly used today in the medical image processing are non-linear and they could have multiple minima. The topography of the search space is characterized by high diversity of cost function landscapes (see Figure 1). In general the similarity measure is a function of continuous variables. Thus, the problem can be classified as multivariate, continuous, nonlinear and constrained optimization problem. By this the estimation of the global optimum demands a robust optimization method. In the context of image and range data processing the simulated annealing has been already used to solve similar registration problems [11]. This work is devoted to the development of a reliable, efficient, and generally applicable optimization framework based on the simulated annealing method.

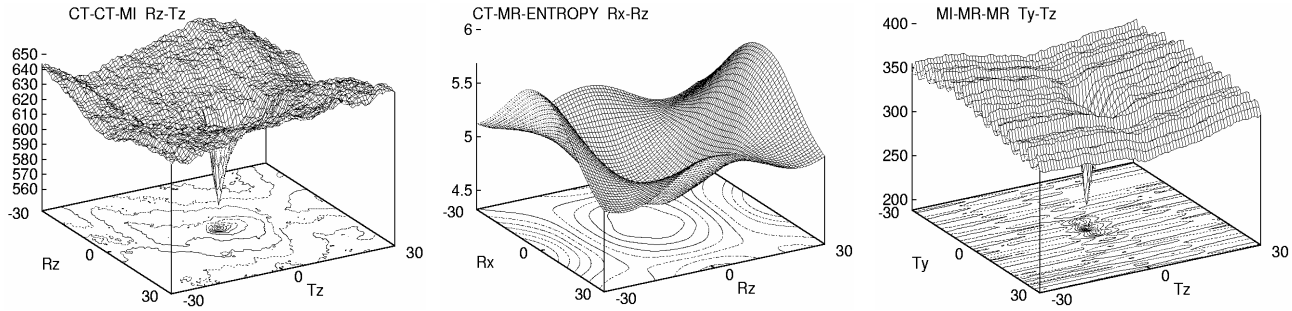

Fig. 1. Examples of different search space landscapes in the intra-modality and inter-modality medical registration problem. The misregistration graphs show the behaviour of the objective function along different pairs of principal axes for various similarity measures: mutual information (left and right) and information entropy (middle).

2. SIMULATED ANNEALING

The simulated annealing method has proved in many fields of industry and technology to be a promising tool for solving complex non-linear optimization problems. Simulated annealing is a probabilistic hill-climbing technique, which is based on the annealing/cooling process of metals. In contrast to the deterministic hill-climbing procedure deteriorations of the cost function can also be accepted here. Starting at an arbitrary point in the search space the algorithm repeatedly generates a new solution by making small transformations to the current solution (see Figure 2). If an improvement is reached, the new parameter configuration is accepted. In case of deterioration, acceptance is probabilistic, depending on the cost function value and on the actual parameter t_k called temperature (the meaning of the parameter r_k will be explained in the next section). The number of intermediate configurations (inner loop) and the decrease of the temperature α_k are controlled according

to a cooling schedule. The function *equilibrium* determines the length of a single run at the constant temperature level. An extensive study of the method can be found in [1].

$v_0 := [r_x, r_y, r_z, t_x, t_y, t_z]$;
(* initial temperature and radius: *)
$t_0 := t_{start}$; $r_0 := r_{start}$; $k := 0$;
REPEAT
  REPEAT
    choose random neighbor state v_1;
    $\Delta C := C(v_1) - C(v_0)$;
    IF $\Delta C < 0$ **THEN** $v_0 := v_1$; (* downhill move *)
    ELSE
      IF $random(0,1) < \exp\left(\frac{-\Delta C}{t_k}\right)$
      THEN $v_0 := v_1$; (* uphill move *)
  UNTIL *equilibrium()*;
  $t_{k+1} := \alpha_k \cdot t_k$; $r_{k+1} := \beta_k \cdot r_k$;
  $k := k + 1$
UNTIL *stop-criterion()*; (* termination *)

Fig. 2. Procedure Simulated Annealing

The analogy between an optimization problem and the problem of determining the lowest-energy ground state of a physical system with many interacting atoms was first observed by Kirkpatrick, Gelatt and Vecchi [2] and Černy [3]. In particular, the atomic states at any temperature t satisfy the Boltzmann's distribution, namely, the probability that a system is in a given state r is given by $e^{\frac{-E(r)}{k_b t}}$, where $E(r)$ is the energy associated with state r, and k_b is the Boltzmann's constant. According to the thermodynamical analogy the acceptance probability of a new move in the simulated annealing method is given by the following formula:

$$p_k = \begin{cases} 1 & \text{if } \Delta C \le 0 \\ e^{\frac{-\Delta C}{t_k}} & \text{otherwise} \end{cases}$$

When ΔC is negative or zero, then the new move is always accepted. Cost-increasing moves are accepted with a non-zero probability, which decreases gradually as the algorithm continues its execution. High temperature values lead to the acceptation of large deteriorations in cost function value, at low temperatures only improvements are accepted. Note that in the simulated annealing procedure, the Boltzmann's constant is combined with the temperature and we shall use the term *temperature* to refer to their product. Temperature can also be viewed as nothing but a control parameter for the optimization procedure.

When simulated annealing is applied to continuous problems, as in our case, the choosing of a random neighbour state is more complicated compared with combinatorial problems. One idea is to determine first a search direction. Continuous simulated annealing would choose a point on the unit hypersphere at random about the point v_0 that gives the search direction. The algorithm would then choose a random length ($\leq r_k$) to step in that direction.

2.1. COOLING SCHEDULES

The cooling schedule consists of a sequence of monotonically decreasing temperatures and the condition of equilibrium at each temperature. In the building of simulated annealing algorithms the following parameters must be specified to define a cooling strategy:

- the initial temperature and radius,
- the temperature decrement (rule for decreasing temperature),
- the equilibrium condition (the duration of search at each temperature),
- the stopping or convergence criterion (final value of temperature or radius).

There has been much research on practical implementation of simulated annealing for solving many hard optimization problems such as VLSI circuits design or image analysis. The majority of these implementations use problem specific cooling schedules. Azencott [4] divides the cooling schedules presented in the literature into four classes:

- Logarithmic cooling schedules: $t_k = d/\log(1+k)$ (a very slow cooling ratio).

Although recommended from theoretical point of view, it is unsatisfactory from a practical point of view. It requires too much computation time.

- Exponential or geometrical cooling schedules: $t_k = t_{start} \times \alpha^k$

These cooling schedules are characterized by fixed initial temperature t_{start}, a constant temperature decrement α (α < 1 and very close to 1) and a constant number of moves at all temperature plateaus. In Figure 3 the evolution of the optimal cost function values and temperature/radius plateaus in an exponential cooling schedule has been shown. A major disadvantage of this strategy is that the results are sensitive to the choice of the parameter values representing t_{start} and α.

- Suitably adjusted exponential cooling schedules: The designed annealing schedule exploits special properties of the cost function.
- Adaptive cooling schedules: Dynamic selections of t_k according to progressively discovered features of the cost function landscape.

In the next section we will describe an adaptive cooling schedule, which has been implemented in our registration software framework.

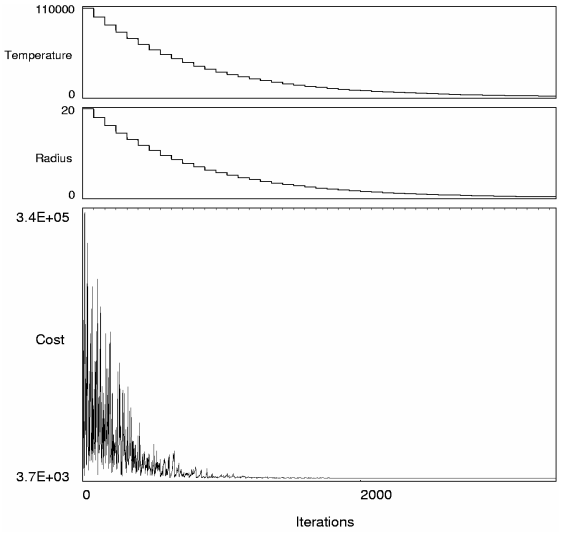

Fig. 3. The exponential cooling schedule of the standard simulated annealing.

3. ADAPTIVE COOLING SCHEDULE

For a cooling schedule to be problem independent, the parameters used in the four conditions should be determined by the algorithm itself and should not have any predefined values. A number of self-adaptive schedules that decrease the temperature in dependence on the average cost function value and its variance has been proposed in the literature (see for example [1], [5] or [6]). Majority of the presented cooling schedules has been formulated for combinatorial optimization problems. Adapting such cooling schedules to the continuous optimization problems is not easy.

In our optimization problem the feasible region M is defined as follows:

$$M = \left\{ v \in \Re^6 : r_{x\min} \le r_x \le r_{x\max}, \ldots, t_{z\min} \le t_z \le t_{z\max} \right\} \qquad (1)$$

Solutions, obtained by the simulated annealing, are relatively insensitive to the choice of the starting transformation vector v_{start}. Crucial is the definition of the feasible set M. If the anatomically optimal transformation vector v_{opt} belongs to M and v_{opt} is there a global minimum, then the described below adaptive cooling schedule will with high probability lead to achieve a near-optimal solution of the given registration problem. In our preliminary work we have observed, that to achieve optima of the same quality, the exponential cooling schedules need 20 to 50 % more computation time than the adaptive cooling schedule presented below. The plots in Figure 4 show convergence of the cost function values, radius/temperature plateaus and the equilibrium monitoring in an adaptive cooling schedule.

3.1. INITIAL TEMPERATURE AND RADIUS

The selection of initial temperature is very important. If we select too high starting temperature, then the program will spend a lot of time in the beginning of the process without progress of any kind. If we select too low initial temperature, then the quality of the final solutions will deteriorate. The starting temperature t_{start} is calculated according to the following formula [5]:

$$t_{start} = c \cdot \sigma$$

where c = const and σ is the standard deviation of the cost distribution. To determine σ, an initial evaluation of l randomly generated moves on the feasible set M is performed ($200 \leq l \leq 1000$ depending on the complexity of the cost function). Varying the value of c corresponds to increasing or reducing the acceptance ratio at the beginning of the annealing process. The starting temperature t_{start} has to be sufficiently high to allow uphill moves away from local minima. In our implementation initial experiments showed that $c = 2$ is a good choice.

If the bounds on the transformation parameters $r_x,...,t_z$ are predefined (see Eq. 1), then the initial radius is given by:

$$r_{start} = 0.5 \times max\{r_{x\max} - r_{x\min},...,t_{z\max} - t_{z\min}\}$$

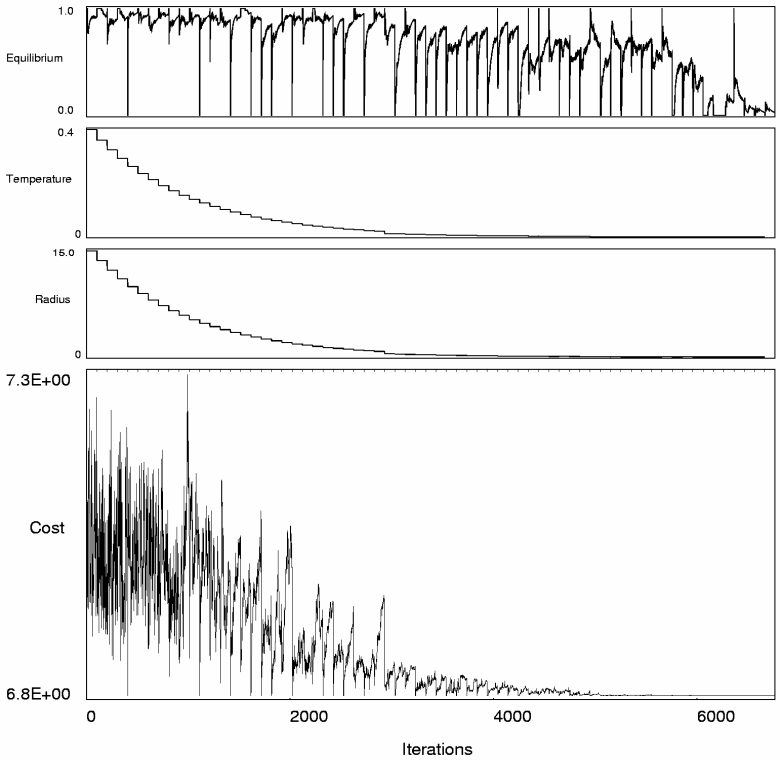

Fig. 4. Adaptive cooling schedule.

3.2. THE TEMPERATURE DECREMENT

Approaches have been proposed which use σ ($\overline{C}$) in the determination of the next temperature decrement [7]. The advantage of these approaches is that the temperature is dynamically controlled by the annealing process itself and is therefore applicable to a wide variety of problems. Huang et al. [5] has proposed an adaptive cooling schedule that has achieved good empirical performance. In this schedule the starting temperature t_{start} and the temperature decrement α_k are chosen according to the mean value $\overline{C}$ and standard deviation σ of the cost function *C*:

$$t_{k+1} = t_k \cdot \exp\left(\frac{-\lambda t_k}{\sigma}\right) \qquad (2)$$

where σ is a standard deviation of the cost distribution and the parameter λ controls the speed of temperature decrease. In our cooling schedule $\lambda = 0.8$.

In some instances of the registration problem the decrementing of temperature can cause large changes in the standard deviation σ (for the subsequent temperature plateaus). To avoid dependencies of the decrement ratio α_k on such changes, Otten and van Ginneken [6] have proposed a smoothing technique. The smoothed σ_s is evaluated according to:

$$\sigma_s(t_{k+1}) = (1-\omega)\cdot\sigma(t_{k+1}) + \omega\cdot\sigma_s(t_k)\cdot\frac{t_{k+1}}{t_k}$$

where the parameter ω controls the smoothing level.

3.3. EQUILIBRIUM IDENTIFICATION

At high temperatures the system is far from thermal equilibrium. It does not tend toward minimum free energy and maximum specific entropy but amplifies certain fluctuations and evolves toward a new dynamic regime that is radically different from stationary states at or near equilibrium. In Figure 5 (bottom) we can observe the evolution of the cost function value (accepted moves only) at high temperatures.

An efficient approach to check whether or not equilibrium is established is an empirical estimator: $\hat{S} = A/(A+R)$, where A is the number of accepted states and R is number of rejected states [8]. The successive values of the acceptance ratio $\hat{S}$ are monitored during the whole annealing process (see Figure 5 (top)). If it fluctuates only by a small amount (about 3 % or less) then equilibrium has been reached. After reaching the equilibrium the temperature parameter *t* is decreased according to Equation 2.

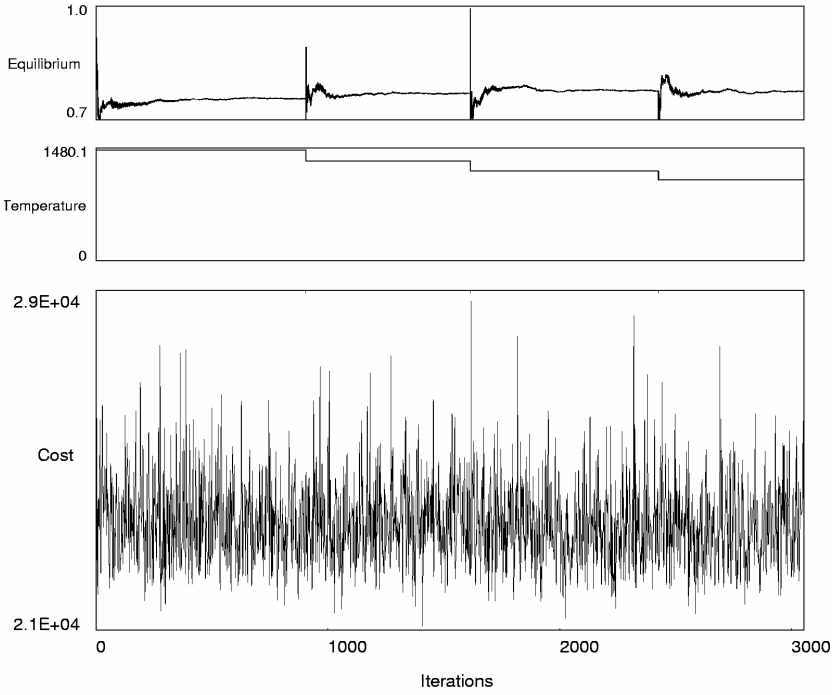

Fig. 5. Equilibrium estimation.

3.4. STOPPING CRITERION

The decision of when to terminate the search may be according to some maximum number of moves, the achievement of a desired accuracy level or some other method. In our cooling schedule both parameters temperature and radius have been made mutually dependent. Thus the amount of search space sampled effectively decreases with the temperature. The desired stopping condition is when the sought global optimum of the cost function is reached at the temperature t_{stop}. The final temperature t_{stop} should be sufficiently low to ensure a high quality global optimum. Additionally, we assume that the terminating criterion is fulfilled, if the acceptance ratio is less than 0.1. A lower bound of 0.001 was placed on the final radius t_{stop} of the hypersphere in the search space. Thus the temperature t_{stop} is relatively easy to determine:

$$t_{stop} = \frac{t_{start}}{r_{start}} \cdot r_{stop}$$

Additionally, we have included two further stopping conditions. The first one set an upper bound on the total number of iteration in the optimization process. If this number is exceeded, the optimization process is terminated. The second one was introduced to prevent a situation when the required optimum was found at high temperatures and there is no improvement in the objective function value at the lower temperatures. The system is assumed to be frozen if no improvements are made throughout the 5–8 successive const temperature plateaus. This criterion can be refined by requiring that the acceptance ratio $\hat{S}$

is smaller than a given value e_r (e_r = 0.2). The search can be also terminated prematurely, before convergence to correct results occurs, if the user finishes the process using the abort button.

4. RESULTS AND DISCUSSION

The described above adaptive cooling schedule for the simulated annealing method has been applied in more than 30 registration cases. Different similarity measures (mutual information, surface similarity, information entropy of the scatter-plot, cross-correlation) have been optimized in the intra-modality as well as in the inter-modality context. The proposed adaptive approach proved to be more robust than the conventional exponential or geometrical cooling schedules. In the majority of cases where the standard exponential cooling schedules have get stuck on local optima, the proposed adaptive method delivered anatomically correct global optima. The main problem with the simulated annealing method is its long execution time. In general, the time required to perform the optimization calculations is a function of the data size, corresponding features complexity, the required accuracy and the computer speed. The typical running times for smaller data sets (from 128x128x50 to 256x256x50) were between 5 minutes and 4 hours depending on the cost function type and the required precision. For the larger data sets (from 256x256x50 to 512x512x70) this time has been increased to 5-20 hours (SGI Octane2, 360/600 MHz CPU, 1 GB RAM). It should be mentioned, however, that both simulated annealing algorithm and cost function computations were not accelerated or parallelized.

The proposed method is very reliable for the estimation of the global minimum in the optimization of objective functions with highly differentiated search space landscapes. However, the simulated annealing is not entirely satisfactory for actual registration problems, since it does not appear to take full advantage of the constraints inherent in the problems. An obvious drawback of this method is that it does not use any local information (such as the derivatives). There is no guidance in the search process.

During the investigation it has been observed that the decrementing of the radius of the hypersphere in the search space should be done non-isotropically. So, that we obtain 6D hyper-ellipsoids in the succeeding steps. The differentiation between the radii for the translations and rotations plays the most important role. In the adaptive cooling schedules the results appear to be relatively insensitive to the parameter choices. This is important when the simulated annealing algorithm is to be used in such different cost function landscapes as in Figure 1. Development of a set of statistical measures, which would give some indication of the *ruggedness* of a landscape, could be helpful in the choice of suitable search strategy.

There are many alternative algorithms to attack such complex nonlinear optimization problems. Davidon-Fletcher-Powell or Levenberg-Marquardt [9] are examples of the deterministic methods. A drawback of these techniques is that, they are susceptible to local minima. As a non-deterministic alternative to simulated annealing, the genetic algorithms [10] can be used. This approach is a search procedure inspired by the mechanisms of

biological evolution that emulates natural genetic operators: mutation, crossover. The algorithm evaluates multiple points in the search space simultaneously and thus has the potential to converge to the global optimum. Probably the most important disadvantage of genetic algorithms is their strong dependence on a large set of parameters (e.g., size of the population, number of generations, probabilities for applying the random operators, rate of generational reproduction, etc.) that have to be experimentally tuned for a given problem domain.

5. CONCLUSIONS

We found that the simulated annealing is a reliable optimization method. In the medical data registration problem, where the optimization is done over highly variable cost function landscapes, the method delivers good estimation of the optimal solution. The advantage of using a self-adapting cooling schedule is that the temperature decrease is guided by the search process itself, thus the parameterization of the algorithm is minimized. The main disadvantage of the proposed method is the long computation time. This drawback could be removed by using parallel simulated annealing algorithms.

BIBLIOGRAPHY

[1] VAN LAARHOVEN P.J.M., AARTS E.H.J., *Simulated Annealing: Theory and Applications*. Series: Mathematics and its applications. D. Reidel Publishing Company, Dordrecht, 1987.

[2] KIRKPATRICK S., GELATT JR C.D., VECCHI M.P., Optimization by simulated annealing, *Science*, 220 (4568), pp. 671-680, 1983.

[3] ČERNY V., Thermodynamical approach to the travelling salesman problem: An efficient simulation algorithm, Journal of Optimisation Theory and Applications, Vol. 45, pp. 41-51, 1985.

[4] AZENCOTT R., *Simulated annealing- parallelization techniques*, John Wiley and Sons, Inc., New York, 1992.

[5] HUANG M.D., ROMEO F., SANGIOVANNI-VINCENTELLI A.L., An efficient general cooling schedule for simulated annealing, In *IEEE International Conference on Computer Aided Design*, pp. 381-384, 1986.

[6] OTTEN R., VAN GINNEKEN L., *The Annealing Algorithm*, Kluwer Academic Publishers, 1989.

[7] SECHEN C., *VLSI Placement and Global Routing Using Simulated Annealing*, Series in Engineering and Computer Science, 54. Kluwer Academic Publishers, 1988.

[8] RODRIGUES R., BATEL ANJO A., *On Simulating Thermodynamics, Applied Simulated Annealing*, Ed. by René V. V. Vidal, Lecture Notes in Economics and Mathematical Systems, Vol. 396, pp. 45-60, Springer-Verlag, 1993.

[9] PRESS W.H., FLANNERY B.P., TEUKOLSKY S.A., VETTERLING W.T., Numerical Recipes in C, Cambridge University Press, second edition, 1992.

[10] MICHALEWICZ Z., *Genetic algorithms + data structures = evolution programs (3rd ed.),* Springer-Verlag, London, 1996.

[11] LUCK J., LITTLE C., HOFF W., Registration of Range Data Using a Hybrid Simulated Annealing and Iterative Closest Point Algorithm, Proc. of IEEE International Conference on Robotics and Automation, San Francisco, April 24-28, 2000.