Journal of Biomedical Informatics 53 (2015) 65–72

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Optimal marker placement in hadrontherapy: Intelligent optimization strategies with augmented Lagrangian pattern search

Cristina Altomare [a,*], Raffaella Guglielmann [b], Marco Riboldi [c,d], Riccardo Bellazzi [a], Guido Baroni [c,d]

[a] Laboratory for Biomedical Informatics "Mario Stefanelli", Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Via Ferrata 1, 27100 Pavia, Italy
[b] Department of Mathematics F. Casorati, University of Pavia, Via Ferrata 1, 27100 Pavia, Italy
[c] Department of Electronics Information and Bioengineering, Politecnico di Milano University, Piazza Leonardo da Vinci 32, 20133 Milano, Italy
[d] Bioengineering Unit, CNAO Foundation, Pavia, Italy

ARTICLE INFO

Article history:
Received 24 May 2014
Accepted 2 September 2014
Available online 8 September 2014

Keywords:
Radiotherapy
Marker placement
Simulated annealing
Pattern search
Constrained optimization

ABSTRACT

Purpose: In high precision photon radiotherapy and in hadrontherapy, it is crucial to minimize the occurrence of geometrical deviations with respect to the treatment plan in each treatment session. To this end, point-based infrared (IR) optical tracking for patient set-up quality assessment is performed. Such tracking depends on external fiducial points placement. The main purpose of our work is to propose a new algorithm based on simulated annealing and augmented Lagrangian pattern search (SAPS), which is able to take into account prior knowledge, such as spatial constraints, during the optimization process.

Material and methods: The SAPS algorithm was tested on data related to head and neck and pelvic cancer patients, and that were fitted with external surface markers for IR optical tracking applied for patient set-up preliminary correction. The integrated algorithm was tested considering optimality measures obtained with Computed Tomography (CT) images (i.e. the ratio between the so-called target registration error and fiducial registration error, TRE/FRE) and assessing the marker spatial distribution. Comparison has been performed with randomly selected marker configuration and with the GETS algorithm (Genetic Evolutionary Taboo Search), also taking into account the presence of organs at risk.

Results: The results obtained with SAPS highlight improvements with respect to the other approaches: (i) TRE/FRE ratio decreases; (ii) marker distribution satisfies both marker visibility and spatial constraints. We have also investigated how the TRE/FRE ratio is influenced by the number of markers, obtaining significant TRE/FRE reduction with respect to the random configurations, when a high number of markers is used.

Conclusions: The SAPS algorithm is a valuable strategy for fiducial configuration optimization in IR optical tracking applied for patient set-up error detection and correction in radiation therapy, showing that taking into account prior knowledge is valuable in this optimization process. Further work will be focused on the computational optimization of the SAPS algorithm toward fast point-of-care applications.

1. Introduction

In high precision photon radiotherapy and in hadrontherapy (HT), the theoretical geometrical selectivity of the treatment (particularly enhanced in charged particle therapy) requires specific technological and methodological efforts to minimize the occurrence and size of geometrical deviations with respect to the treatment plan at each treatment session. In HT, in particular when active scanning beam delivery is applied, inter- and intra-fractional uncertainties may produce severe consequences in dose deposition patterns, thus frustrating highly conformal treatment plans designed to treat deep-seated solid tumor in critical sites.

Beside in-room X-ray imaging and image registration methods, valuable tools for inter- and intra-fractional deviations mitigation consist of infrared (IR) optical tracking and point-based registration [1–4]. Surface tracking methods have also been used in radiotherapy [5–7] but in HT, patients are usually immobilized with a thermoplastic mask that prevents the tracking camera from visualizing the patient's surface directly. In this case, surface-based methods are not suitable for the target registration and this is why we focused on point-based tracking.

Point based IR optical tracking in radiation oncology is based on the real-time detection of the 3D position of a set of external markers placed on patient's skin. Fast iterative estimation of the 6° of

* Corresponding author.
E-mail address: cristina.altomare01@ateneopv.it (C. Altomare).

http://dx.doi.org/10.1016/j.jbi.2014.09.001
1532-0464/

freedom rigid transformation is obtained by minimizing the measured marker displacements with respect to corresponding references coming from the treatment plan dataset obtained with Computed Tomography (CT): this technique allows the compensation of geometrical deviation mainly due to patient set-up errors. The root mean square (RMS) of the distance between corresponding markers, defined as fiducial registration error (FRE) [8] represents the metric for the estimation of the corrective transformation of patient position, under the usual assumption that FRE minimization implies minimizing geometrical deviations affecting the target of the treatment. The distance between the real target position and its corresponding reference position is defined as target registration error (TRE) [9]. TRE cannot be directly calculated, since it is not possible to know the real position of the target. However, statistical predictors can be applied to estimate TRE size as a function of residual FRE and of the geometric distribution of markers on patient surface. Under the hypothesis that optical tracking is efficacious in registering corresponding surface surrogates of the target achieving a minimum FRE, there is the need to identify, by means of appropriate optimization methods, the optimal marker configuration that minimizes the corresponding value of TRE. The final goal is to obtain a higher accuracy in target repositioning [10].

Several authors have faced the issue of optimal marker placement by proposing different optimization methods.

Liu et al. [11] described a floating optimization based on genetic algorithm and obtained a 50% TRE reduction with respect to a random marker configuration. Nevertheless, the methods turned out to be computational expensive, due to the large number of parameters to be optimized and to the continuity of the search space that slows down the execution of the algorithm. Moreover, marker visibility constraints imposed by the optical tracking system (OTS) were not taken into account.

An interesting strategy to deal with complex optimization problems is to resort to methods that are able to incorporate prior knowledge in the search of the best solution, usually by adding a suitable set of constraints. This is typically possible by applying algorithms belonging the AI and to the statistical learning tradition, such as neural networks, genetic algorithms and simulation annealing [12,13].

These strategies lie at the intersection of data analysis and knowledge-based system, an area that is known as "intelligent data analysis" or IDA [14–17]. IDA, since the late 90s, has produced several interesting studies and tools with a noteworthy number of applications in medicine and biology [18–21].

In the area of optimal marker placement, an IDA approach has been implemented by Riboldi et al. [22], who combined genetic algorithm (GA) with Taboo search (TS) in a method called Genetic Evolutionary Taboo Search (GETS). They proposed a permutation encoding with the goal to characterize candidate solutions and make the search space discrete. Consequently, the execution time of the algorithm turned out to be reduced with respect to the approach described in [11]. Taboo search allowed the algorithm to reject marker configurations, which would have featured critical visibility for the OTS cameras and to exclude irradiation field areas from the surface available for marker placement. The results obtained on data coming from ten prostate patients showed an average 26.5% reduction of TRE (compared to a random marker configuration), against the 19.4% obtained when a quasi-Newton method was applied. Limitations of the GETS algorithm reside in the fact that possible overlap of markers, commonly occurring when a high marker number is used, is not taken into account. As another example of knowledge-driven approaches, in the frame of image-guided neurosurgery, Shamir et al. [23] described a collaborative framework that allows the surgeon to optimally plan marker location on routine diagnostic images before preoperative imaging, and to select during surgery the fiducial markers and the anatomical landmarks that minimize the target registration error (TRE). The optimal fiducial marker configuration selection can be performed on diagnostic image dataset interactively, by monitoring target selection on a visual Estimated TRE (E-TRE) map, which is automatically updated when the surgeon adds and deletes candidate markers and targets. Data coming from five patients were used and results showed a reduction of the average TRE from 4.7 mm to 3.2 mm.

As a whole, methods previously described exhibit limitations related to a long execution time [11] incomplete constraints about marker placement [22] and the requirement of invasive procedures for the selection of additional anatomical landmarks [23].

The current spread of the clinical centers dedicated to the hadrontherapy and the increasing availability of this therapy worldwide demands new approaches, more accurate and repeatable than the previous ones, with the aim of identifying a standard procedure that ensures the highest precision in tumor localization.

In this paper, we present a novel IDA algorithm that answers to these issues by integrating two different optimization methods: simulated annealing (SA) and pattern search (PS). We have named the algorithm "SAPS". Simulated annealing was selected for its capability of avoiding the entrapment in local minima; pattern search provides reduction of execution time required by SA to converge to a global minimum. Some knowledge-based features of the GETS algorithm [22] have been included in the SAPS algorithm: marker visibility constraints, a priori definition of the surface allowed for marker placement and permutation encoding of candidate solutions. In addition, we have introduced specific constraints that prevent markers overlapping. Table 1 summarizes how the prior knowledge has been converted in specific constraints for the marker placement.

SAPS has been tested on data collected on thirteen head-and-neck and pelvic cancer patients who were treated with proton therapy. Results show that constrained optimization allows us to improve TRE minimization with respect to random fiducial configurations and to those obtained by the GETS algorithm, especially when the number of markers is high. The SAPS algorithm lends itself as a valuable and clinically applicable alternative to improve the accuracy of target localization when IR optical tracking is applied.

2. Methods: data registration techniques and optimization algorithms

2.1. Patient population

The SAPS algorithm was tested on a set of clinical data collected at the National Centre of Oncological Hadrontherapy (CNAO Foundation) in Pavia, Italy [24]. The patient cohort included 6 head and neck and 7 pelvic cancer patients, who were fitted with external

Table 1
The table shows the prior knowledge incorporated into the search, represented by the constraints imposed on the marker placement.

	Prior knowledge	Constraints	Effect
Spatial constraints	Irradiation field areas on the mask surface Marker dimension (cm)	Boundaries for the search space Marker spatial constraint	Avoiding markers placement that affect the treatment Preventing markers overlap
Visual constraints	Marker dimension (pixel)	Marker visibility constraint	Preventing incorrect marker recognition by the OTS system

Table 2
The table shows patient's data as anatomical areas to be treated and the number of markers in the random configuration: elements in the first row identify the head-and-neck patients and elements in the second row identify the pelvic patients.

	Number of markers		
	5	6	7
Head-and-neck	Pat12, Pat13	Pat1, Pat3	Pat2, Pat9
Pelvis	Pat4, Pat8	Pat6, Pat10, Pat11	Pat5, Pat7

surface markers for IR optical tracking applied for patient set-up preliminary correction and continuous monitoring during beam delivery. The number of markers used for each patient is reported in Table 2. Target registration error was calculated for the Planning Target Volume (PTV): PTV is a geometrical concept which takes into consideration the net effect of all the possible geometrical variations and inaccuracies, and it ensures that the prescribed dose is actually absorbed in the clinical target volume (CTV) [25]. We adopt the PTV because it represents the standard model used in treatment planning evaluation and the recommended tool to shape dose distribution. Moreover, in hadrontherapy differences between PTV and CTV dimensions are negligible and they do not influence the calculation of the TRE.

Particle beam treatment planning system (Syngo PT Planning, version VB10, Siemens, Germany) was applied on 3 mm thickness CT slices for anatomical structures delineation (target, organs at risk, skin) and treatment planning. Contours data were saved as DICOM-RT structures.

2.2. Data analysis

Within Matlab (MATLAB R2012a, Mathworks, Inc., Natick, MA, USA), contours were extracted from the DICOM file and specific anatomical landmarks (center of the PTV and OARs) were selected to be included in the procedure for marker configuration optimization. The patient skin surface model, available in the DICOM-RT structure file, was used as the set of spatial position potentially available for marker placement. The model is described by the matrix $S \in R^{N_p \times 3}$ such that $S^T = \{p_1, \ldots, p_{N_p}\}$ where the vector p_j contains the coordinates of the jth point with $j = 1, \ldots, N_p$ and N_p is the total number of points.

From the surface model S, the irradiation fields were removed, as incompatible with marker placement. By knowing the geometry of the optical tracking system set-up in the treatment room, surface areas on which markers would have been invisible to the OTS cameras were identified and excluded from the set of available marker positions.

2.3. Target registration error

The expression of the expected value of TRE, obtained by West et al. [26] is given by:

$$\left\langle \mathrm{TRE}^2(\mathbf{r}) \right\rangle = \frac{\left\langle \mathrm{FRE}^2 \right\rangle}{N_m - 2} \left(1 + \frac{1}{3} \sum_{k=1}^{3} \frac{d_k^2}{f_k^2} \right) \quad (1)$$

where $N_m > 2$ is the number of fiducial points (or fiducials), $\mathbf{r}$ is the vector of target position coordinates measured with respect to the centroid of fiducials, f_k is the RMS distance of the fiducials from the kth principal axis of the fiducials configuration and d_k is the distance of the target from the same axis. The objective function of the iterative optimization procedure was the TRE/FRE ratio, which depends only on the number of fiducials and on their spatial distribution with respect to the target [9,10,26]. According to Riboldi et al. [22], the assumption is that for a given FRE, which results from the intrinsic accuracy in marker 3D localization performed by the OTS, TRE can be minimized if a marker configuration with the minimum TRE/FRE ratio is applied.

2.4. Simulated annealing and pattern search

The SAPS algorithm combines simulated annealing (SA) and pattern search (PS), described in Sections 2.5 and 2.6, respectively. SA initializes the search for the best configuration of the markers and PS refines the results provided by SA. This choice has been made to avoid the entrapment of the optimization process in local minima and, at the same time, to grant convergence in a relatively short time to the global minimum.

Optimal marker configurations found by the SAPS algorithm were compared to the random configuration that was applied for the treatment. The differences provided by the optimal marker configuration were quantified in terms of the TRE/FRE ratio and spatial distribution of markers, also when SAPS and GETS performance were compared. Repeated simulations provided information concerning the repeatability of the optimal solution and the execution time of both algorithms. Moreover, we performed a sensitivity analysis of the algorithm performance as a function of the number of available markers and the inclusion of additional targets (typically the OARs). SAPS algorithm implementation and statistical analysis were performed in Matlab.

2.5. Simulated annealing

Optimal marker placement is a complex combinatorial optimization problem: such problems can be solved by SA, a heuristic algorithm that refers to a stochastic procedure grounded on the analogy with the physical process of heating a material and then slowly lowering the temperature to decrease defects and to minimize the system energy [27,28]. In our case, candidate solutions of SA are the possible configurations of surface markers. As implemented in the GETS algorithm [22] candidate solutions are defined by means of *permutation encoding*: at step i the configuration of N_m points is defined in the following equation

$$\mathbf{x}_i = \left\lfloor \mathrm{ID}_{i,1}, \mathrm{ID}_{i,2}, \ldots\ldots\ldots, \mathrm{ID}_{i,N_m} \right\rfloor \quad (2)$$

where each element $\mathrm{ID}_{i,j}$ represents the ordinal position of the corresponding marker j in the list of points encompassing the skin surface model S. Points of S are ordered in the frontal plane so that consecutive IDs in the list correspond to contiguous points in the model. By adopting this encoding, the search space of solutions becomes discrete and the complexity of the algorithm is significantly reduced.

The initial guess $\mathbf{x}_0$ is randomly chosen. Then, with fixed temperature (T), the algorithm makes a random step in the neighborhood of the current value of the solution ($\mathbf{x}_i$) and a solution $\mathbf{x}_{new}$ is generated by sampling from a *multivariate normal* probability density distribution centered in $\mathbf{x}_i$:

$$\mathbf{x}_{new} = round[mvnrnd(\mathbf{x}_i, \Sigma)] \quad (3)$$

The matrix Σ is a function of T: as the temperature decreases, the algorithm reduces the search space width in order to converge to a minimum. The covariance terms, which represent the correlation between markers, are defined according to the constraint that the markers must not overlap.

A crucial aspect of our approach stands in its capability of including constraints, which can be generated on the basis of the knowledge available on the specific problem.

1. Each configuration must satisfy the *marker visibility constraint*, as implemented in the GETS algorithm [22]: configurations exhibiting an inter-marker distance lower than a given threshold are

rejected, in order to avoid that multiple markers are recognized as a single marker by the OTS. Inter-marker distance is measured on the camera image planes (three cameras are installed in the treatment room), by backprojecting the 3D markers position (estimated at each iteration) through the known OTS camera calibration parameters. The inter-marker distance threshold was fixed to 3 pixels in order to ensure the correct recognition of each marker in all the possible experimental configurations. The constraint is given by:

$$\min_{j,l,k}\{\|\mathbf{T}_k[S(\mathrm{ID}_{i,j})] - \mathbf{T}_k[S(\mathrm{ID}_{i,l})]\|\} \geqslant 3 \text{ pixels con}$$
$$j,l = 1,\ldots,N_m; \quad k = 1,2,3 \tag{4}$$

where $\mathbf{T}_k$ denotes the projection matrix (from physical space to image plane) of the kth camera and $S(\mathrm{ID}_{i,j})$ denotes the row of the matrix S related to the jth marker of the ith configuration.

2. In the SAPS algorithm, we also introduced a *marker spatial constraint*, defined as:

$$\min_{h}\{S(\mathrm{ID}_{i,j};h) - S(\mathrm{ID}_{i,l};h)\} \geqslant 10 \text{ mm con}$$
$$j,l = 1,\ldots,N_m; \quad h = x,y,z \tag{5}$$

This constraint is necessary to avoid marker overlap on the patient's mask. It requires that the minimum distance between the centroids of two neighboring markers is equal to 10 mm in each dimension, since the radius of each marker is 5 mm.

The value of the objective function $f(\mathbf{x}_{new})$, where f is the TRE/FRE, ratio is compared with $f(\mathbf{x}_i)$.The algorithm accepts the new iterate according to the *rule of Metropolis*, with an acceptance probability given by:

$$P_0 = \begin{cases} 1 & \text{if } \Delta f < 0 \\ \exp\left(\frac{-\Delta f}{T}\right) & \text{if } \Delta f \geqslant 0 \end{cases} \tag{6}$$

where $\Delta f = f(\mathbf{x}_{new}) - f(\mathbf{x}_i)$. Let us observe that $\mathbf{x}_{new}$ is accepted also when f increases, but with a probability P_0 depending on T: this feature guarantees that the algorithm is able to explore more widely the space of possible solutions and consequently avoids being trapped in local minima.

T represents the control parameter in SA since it affects the region width of solutions investigated by the algorithm. We set the initial temperature T_0 equal to 150 °C and we define T as a variable with an *oscillating behavior* in the initial phase with the final goal to explore a wider region in lower time.

After the initial phase, temperature decreases gradually as the algorithm proceeds, according to a *cooling schedule*: the reannealing interval (L_k) is set equal to 250 points and T decreases according to the following formula

$$T_k = T_{k-1} \cdot \alpha_k \tag{7}$$

where k is the kth iteration and $1 > \alpha > 0$ is a parameter that, together with L_k, determines the rate of T decrease.

The adopted stopping criterion is based on the maximum number of iterations, which is set equal to 200 (the so called stall iteration limit): this strategy guarantees a reasonable computational time. Actually the algorithm could be stopped before the final optimal solution has been reached [28,29].

With the final goal to reduce the computational time, a pattern search strategy was implemented: starting from the solutions found by the simulated annealing, we refined the search in the neighborhood of these solutions, and then we achieved both a further reduction of the TRE/FRE ratio and a lower global execution time.

2.6. Pattern search

Pattern search (PS) is a direct search method, firstly proposed by Lewis and Torczon [30], for nonlinearly constrained optimization as adaption of a bound constrained augmented Lagrangian method, as proposed by Conn et al. [31]. We will refer to it as augmented Lagrangian pattern search (ALPS).

Unlike optimization methods that use information about the gradient and higher derivatives to search for an optimal point, a direct search algorithm scans a set of points belonging to a grid (*mesh*) built around the current point and looks for a point where the objective function is lower than the current value of the function. The mesh is built starting from a set of vectors $\{\mathbf{v}_k\}$ (*pattern*) that define the search directions, and from a scalar (*mesh size*), which defines the distance between points. The pattern can be defined according to two different strategies: in the first one, the pattern is made of $N_m + 1$ vectors corresponding to N_m basis vectors and the vector given by the opposite of the sum of the previous ones; the second strategy considers 2 N_m vectors properly defined. In our implementation, the first approach was selected. At each step the algorithm builds the mesh (M_i) centered around the current point ($\mathbf{x}_i$) according to:

$$M_i = \{\mathbf{x}_i \pm \Delta_i \mathbf{v}_k, k = 1,\ldots,N_m + 1\} \tag{8}$$

and *polls* the points of the mesh by evaluating the objective function in those points. Δ_i is the mesh size, which depends at each step upon the poll, according to the following equation:

$$\Delta_{i+1} = \tau\Delta_i \tag{9}$$

where the initial value of τ is set to 1. If the algorithm finds a point in the mesh where the objective function decreases with respect to the previous step, the poll is called *successful* and that point becomes the center of the new mesh: in this case, the algorithm sets $\tau > 1$ and continues to explore the search space in the new mesh. Otherwise, the poll is called *unsuccessful*: the current point does not change and the search space must be narrowed, i.e. the algorithm sets $\tau < 1$ (actually, it is likely that the algorithm is near the global optimum).

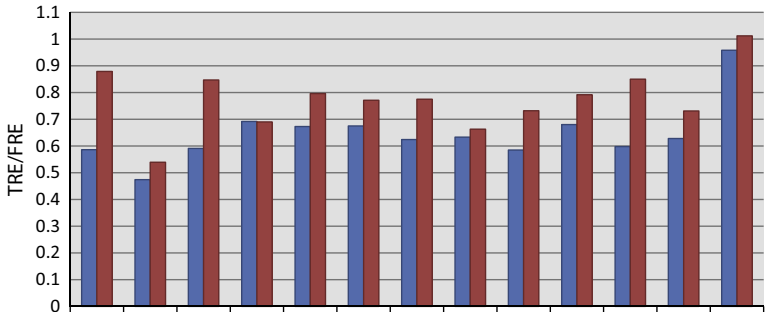
The algorithm proceeds in this way until the stop criterion is met i.e. when the mesh size Δ becomes smaller than mesh tolerance δ [30,32,33]. In the approach we adopted, δ is defined as $(\mathrm{TolMesh}/N_m)^2$ where the default value of *TolMesh* is 1e−6.

The ALPS algorithm, in addition to the classical PS, allows one to deal with the problem of nonlinear constrained optimization: it replaces the explicit knowledge of derivatives with a stopping criterion based on the pattern size in a way that preserves the convergence properties of the augmented Lagrangian method [30]. In general, the m_t *nonlinear constraints* are the following:

$$c_k(\mathbf{x}) \leqslant 0, \; k = 1,\ldots,m \tag{10}$$
$$\mathrm{ceq}_k(\mathbf{x}) = 0, \; k = m+1,\ldots,m_t \tag{11}$$

where $c(\mathbf{x})$ represents m inequality constraints and $\mathrm{ceq}(\mathbf{x})$ represents m equality constraints. In the frame of this approach, we formulate the minimization problem by resorting to strategy well-known in statistical learning, i.e. by combining the objective function, the log-barrier function (for inequality constraints, converted into equality constraints by means of slack variables) and the quadratic penalty function (for equality constraints). The augmented Lagrangian is then defined as

$$\Phi(\mathbf{x},\boldsymbol{\lambda},\mathbf{s},\rho) = f(\mathbf{x}) - \sum_{k=1}^{m}\lambda_k s_k \log(s_k - c_k(\mathbf{x})) + \sum_{k=m+1}^{m_t}\lambda_k \mathrm{ceq}_k(\mathbf{x}) + \frac{\rho}{2}\sum_{k=m+1}^{m_t}\mathrm{ceq}_k(\mathbf{x})^2 \tag{12}$$



	1	2	3	4	5	6	7	8	9	10	11	12	13
Optimal	0.586	0.474	0.591	0.692	0.673	0.675	0.624	0.633	0.585	0.68	0.598	0.628	0.958
Random	0.879	0.539	0.847	0.69	0.796	0.771	0.775	0.663	0.732	0.792	0.85	0.731	1.012

Fig. 1. TRE/FRE values obtained for the optimized and the random configuration are listed. With regard to the optimal configuration (shown in blue) we report the mean calculated on 10 simulations performed by the SAPS algorithm for each patient. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where the components $\lambda_k \geqslant 0$ of the vector $\boldsymbol{\lambda}$ are the Lagrange multiplier estimates; the elements $s_k \geqslant 0$ of the vector $\mathbf{s}$ are shifts (slack variables); $\rho > 0$ is the penalty parameter.

The ALPS algorithm minimizes a sequence of subproblems defined as

$$\min \Phi^{(j)}(\mathbf{x}, \boldsymbol{\lambda}^j, \mathbf{s}^j, \rho^j), \quad j = 1, 2 \ldots \tag{13}$$

Each subproblem has a fixed value of λ, s, and ρ and it is solved providing a solution with a required accuracy and satisfying the constraints. Then the Lagrangian multipliers and the penalty parameter are properly updated [30,31,34,35]. Let us note that in our case, we considered only inequality constraints (the visibility and spatial constraints defined in Section 2.5): such constraints make the algorithm rather time-consuming.

3. Results

3.1. Optimal versus random configuration

First, we evaluated the performance of the SAPS algorithm in terms of the TRE/FRE ratio, calculated for each patient and for the optimized and the random configuration: results are reported in Fig. 1. Let us note that optimal configurations are characterized by TRE/FRE values lower than the random ones, except in the case of the pelvic patient 4. The repeatability of the solutions provided by the SAPS algorithm is highlighted by the low value of the standard deviations varying from 0.001 to 0.007 (values are not reported in the figure).

Differences between patients may be due to several factors, such as the number of markers, their position with respect to the target and the width of the available surface for their placement. The latter factor is important because, with a larger surface model, the algorithm finds configurations where markers are more distant from each other leading to a lower TRE/FRE ratio, as also described in [26]. For example, this occurs in patients 2 and 8 where the number of points in the surface model is higher than the ones in the other patients.

3.2. Result dependence on the number of markers

The number of markers (N_m) significantly affects the TRE/FRE ratio; in order to investigate this issue more in detail, we calculated a number of optimized solutions with N_m varying from five to ten. The upper limit was chosen by considering that the placement of more than ten markers would give raise to a too crowded marker configuration, especially on head and neck district. Results for patient 1 are shown in Fig. 2 where we can observe that the TRE/FRE ratio decreases almost exponentially when the number of markers increases. The Kruskal–Wallis test[1] confirms a significant difference among the TRE/FRE values obtained with different values of N_m (p-value <<0.001). Post-hoc comparisons reveal that statistical significance of TRE/FRE reduction is obtained when at least eight markers are used.

3.3. SAPS versus GETS algorithm

The performances of SAPS have been compared with those of GETS algorithm both in terms of TRE/FRE values and average execution time. Let us remind a crucial feature of SAPS: unlike GETS, a marker spatial constraint given in Eq. (5) is implemented in order to prevent markers overlapping in the optimal configuration.

Firstly we compared mean and standard deviation of TRE/FRE values obtained from ten simulations for each value of N_m (between five and ten), where the SAPS algorithm has been initialized with N_s random marker configurations for each simulation. Fig. 2 reports as an exemplifying case the results obtained on patient 1.

Results show that the SAPS algorithm is able to find marker configurations featuring a lower TRE/FRE ratio for each value of N_m. In addition, standard deviations for SAPS algorithm have been found to be lower than those obtained with GETS, with a range between 0.002 and 0.010 for SAPS and between 0.009 and 0.014 for GETS. Although the improvement is not major, SAPS performed more effectively and more reproducibly with respect to GETS. Similar results have been found for the other patients, as shown in Fig. 3.

A multi-way analysis of variance was performed in order to investigate statistically significant differences between the two algorithms. Factors were patients, the number of markers and the two algorithms. Results confirm the trend depicted in Fig. 2: even if the mean values of the TRE/FRE are comparable, the lower standard deviations superimposed on the results obtained by SAPS cause a significant difference between the two methods ($p < 0.001$).

Moreover we investigated the average execution time for the two algorithms. Fig. 4 reports the values obtained on patient 1 for a simulation in which $N_s = 10$ candidate solutions have been explored.

Firstly, let us note that both the algorithms take a longer time when the number of markers increases which means that a high

[1] We have applied the Kruskal–Wallis test for all patients, not only for patient 1. Each group considered in the test is identified by the TRE/FRE values calculated on ten simulations performed by the SAPS algorithm with a given value of N_m.

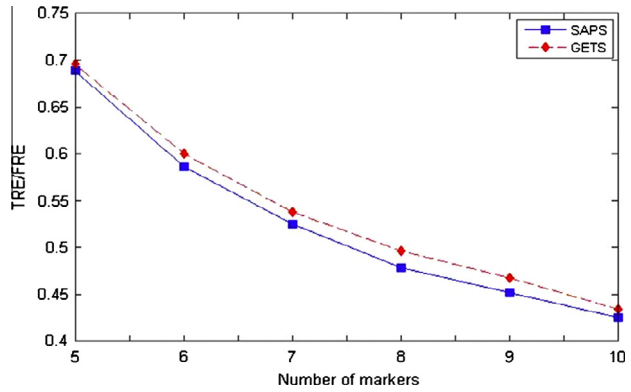
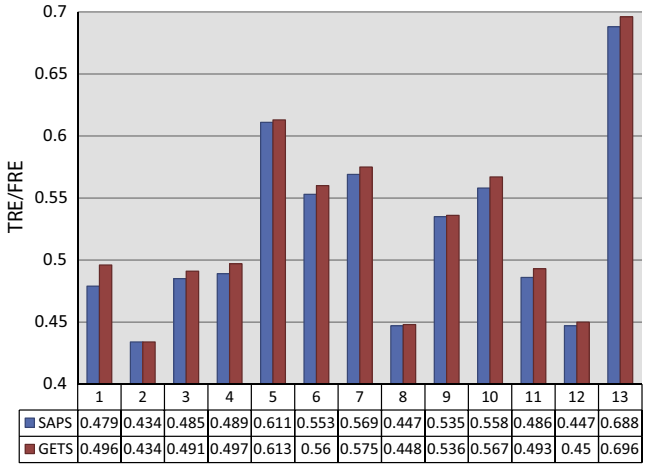


Fig. 2. The TRE/FRE ratio calculated for patient 1 with N_m between 5 and 10. Blue square and red diamond represents the mean of the TRE/FRE values obtained on ten simulations performed respectively by the SAPS and the GETS algorithm. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



	1	2	3	4	5	6	7	8	9	10	11	12	13
SAPS	0.479	0.434	0.485	0.489	0.611	0.553	0.569	0.447	0.535	0.558	0.486	0.447	0.688
GETS	0.496	0.434	0.491	0.497	0.613	0.56	0.575	0.448	0.536	0.567	0.493	0.45	0.696

Fig. 3. TRE/FRE values of optimal configurations obtained by SAPS (shown in blue) and GETS (shown in red) are listed. We report the mean calculated on 10 simulations performed by the two algorithms for each patient, with a number of markers equal to 8. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

number of markers increases the time complexity of both. The difference between them is in the average value of their execution time. They differ by one order of magnitude: this feature can be easily explained by the fact that during each iteration of the optimization process, GETS explores a population of N_s candidate solutions in parallel (due to the implicit parallelism of GA), while SAPS performs the optimization by exploring all the candidate solutions sequentially.

Considering that time taken by SAPS is about N_s times that one taken by GETS and that the presence of spatial constraints in SAPS makes it more time-consuming, the execution time of the two algorithms is similar indeed. Of course, it would be useful parallelizing the SAPS algorithm in order to obtain a better running time with respect to GETS.

3.4. Inclusion of OARs

A set of simulations has been performed by adding as additional targets the anatomical structures identified as organs at risk. Their sparing (keeping the dose delivered in these volumes below specific threshold) is a strict requirement to minimize treatment toxicity. Differences, in terms of average TRE/FRE ratio with respect to the PTV-only condition (where only the PTV as target structure is considered) turned out to be consistently negligible (<0.1). The Kruskal–Wallis ANOVA equivalent test applied on the three conditions (PTV-only, PTV + OARs and random configuration) revealed overall significant differences in the TRE/FRE ratio ($p = 0.0035$). Post-hoc comparisons confirmed significantly higher TRE/FRE values for the random configuration along with no significant differences between the PTV-only and PTV + OARs conditions.

3.5. Graphical representation of an optimal configuration

In order to provide evidence of the size of fiducials displacement for TRE/FRE optimization, an exemplifying case obtained for patient 11 is reported in Fig. 5. Contours of skin, target and OARs are represented along with surface fiducials, with white markers belonging to the random configuration and red markers to the optimized configuration.

From the geometrical perspective, the optimization process tends to displace markers on the boundary points of the skin surface model, as a way to maximize the inter-marker distances and give raise to a configuration that surrounds the target, thus well interpreting Eq. (1).

4. Discussion

In this paper, we proposed a method for the optimization of the configuration of surface fiducials for patient set-up verification and correction through point-based IR optical tracking. By using the contouring information from the planning data coming from 13 patients undergoing particle beam therapy in cranial and extracranial area, we demonstrated that patient-specific knowledge-based fiducials placement optimization leads to significant reduction of the TRE/FRE ratio. This shows that, for a given residual error on surface fiducials after point-based registration, accuracy in target localization can be significantly increased with respect to a random placement of fiducials on patient surface.

The optimization problem was faced by means of a new IDA optimization strategy, which couples simulated annealing for the initial research of markers' configurations and pattern search for solution refinement. The strategy was named SAPS and was compared with previously reported fiducials optimization strategy based on evolutionary approach, the GETS algorithm proposed in [22].

The reported results showed that optimal marker configurations consistently give rise to significant lower TRE/FRE values with respect to the random configurations, which are obtained by not negligible displacement of fiducials position. The optimization process tends to align markers in correspondence of the target and to maximize the inter-marker distances. This general result fulfils the cost function parameters (see Eq. (1)) and it is in line with the guidelines reported in [26], namely to avoid collinear placement of markers, to keep markers as far apart as possible, to place markers in such a way that they surround the target and to use as many markers as possible. With specific reference to the optimal number of markers to be used, we found that, although in the frame of a consistent improvement with respect to random configurations, statistically significance of the SAPS optimization was obtained when a high number of markers is used. Improvements were obtained despite the restriction in the available surface where marker positions could be selected, with specific constraints in terms of marker visibility from the optical tracking system TVC configuration and the need to avoid marker placement within the irradiations fields, due to potential beam range perturbation.

The comparison of the SAPS algorithm with the previously reported evolutionary marker optimization algorithm GETS highlighted two noteworthy results. First, optimal solutions provided

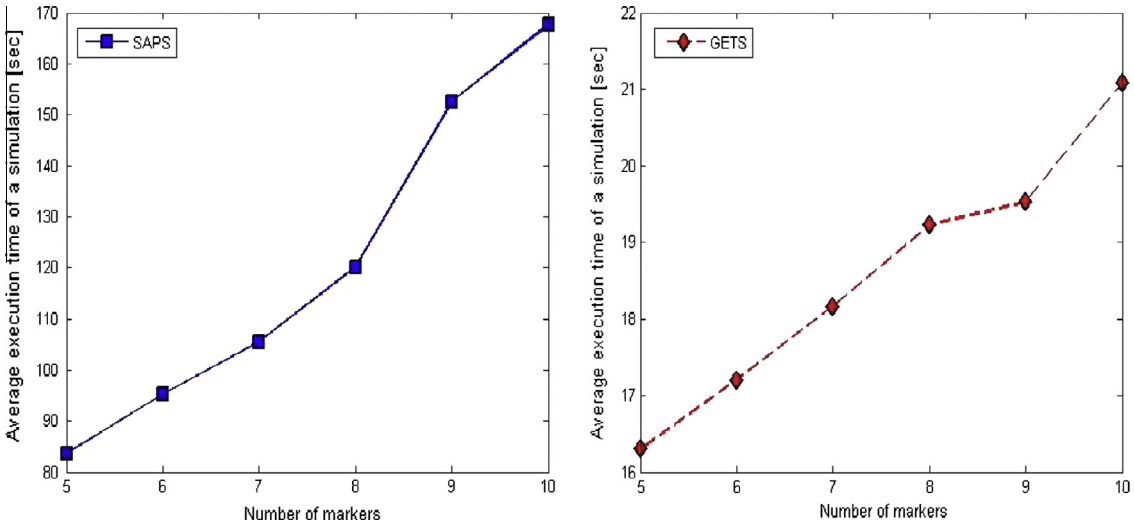


Fig. 4. The average execution time of a simulation calculated for patient 1 with N_m between 5 and 10 and with $N_s = 10$. Blue square and red diamond represents the values obtained respectively by the SAPS and the GETS algorithm. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

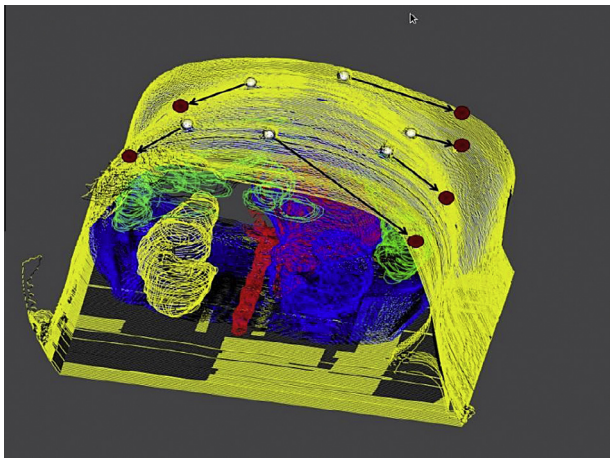


Fig. 5. 3D representation of the patient's skin (in yellow) with the tumor target and OARs inside, for patient 11. White points are the markers of the random configuration, while the red points are the markers of the optimal configuration. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

by SAPS featured consistently lower TRE/FRE ratios, thus behaving better than GETS; second, the SAPS algorithm turned out to be much more time-consuming with respect to GETS. These results are explained by the different computational approaches of the two algorithms, respectively based on simulated annealing and genetic algorithms. GA is characterized by implicit parallelism and can exploit massively parallel architectures, thus optimization computational costs, but may suffer from disruption and poor convergence properties. As a matter of fact, GA always accepts new solutions even if they score less than solutions belonging to previous generations. This leads to loss of good scoring solutions, thus potentially preventing optimal performance. Conversely, SA features good convergences properties and is in some sense immune to disruption, but parallel computational approaches are not easily implementable [36]. Therefore, better results were expected from SAPS, even if at the price of higher computational cost.

In order to investigate the repeatability in the optimal solutions we performed multiple runs with the two algorithms: the obtained results highlighted that the SAPS algorithm produced solutions with lower standard deviations superimposed to the TRE/FRE ratios and a lower variability in spatial distribution of markers with respect to the GETS algorithm. The resulting lower variability in SAPS solutions compared to GETS can be explained by the evolutionary approach underlying the GETS and indicates that optimal solutions provided by SAPS may be closer to a global optimum. A further study could concern the introduction of *ad hoc* initial marker configurations with the final goal to evaluate and improve both the effectiveness and the accuracy of the algorithm.

Interestingly, no significant difference in the pattern of optimal marker configurations was obtained when also OARs were included in the optimization process. This was true both in cranial and extra-cranial sites (pelvis). Such result suggests a limited space of solutions especially in presence of the described visibility constraints, even if a higher number of cases is required to conclude that the optimization of fiducials accounting for one single target may ensure consistent localization of organs at risk.

5. Conclusions

The SAPS algorithm lends itself as a valuable strategy for fiducial configuration optimization in IR optical tracking applied for patient set-up error detection and correction in radiation therapy. Moreover, target registration error can be reduced to approximately half of the residual fiducial registration error when marker configuration is optimized, thus granting consistently submillimetric target localization in presence of FRE around 1 mm, which is an acceptable number including instrumental intrinsic accuracy and marker manual repositioning accuracy. Of course, we are aware of the fact that this conclusion is valid only if one assumes that external fiducials represent reliable surrogate of the target. Any inter- and intra-fractional relative motion of internal target with respect to surface fiducials, widely documented in literature [37–39] cannot be interpreted by any optimization algorithm of external fiducials. Nevertheless, we believe that in the frame of a combined use of in-room imaging and IR optical tracking, external fiducials optimization may lead to improvement in quality control of irradiation geometry, thus representing a good practice in radiotherapy clinical routine.

Future improvements concern the SAPS algorithm optimization leading to a point-of-care utilization, for providing health practitioners with an immediate feedback describing TRE/FRE values as a function of changes in marker positioning. Thanks to our approach,

it is possible to envisage the automated definition of the SAPS constraints on the basis of the anatomical features of the specific patient. Computational optimization can be obtained by parallelizing the algorithm according to the techniques described in [40]. In particular, asynchronous SA would ensure significant improvements, by exploiting processors that exchange information asynchronously to bring the system toward a global minimum, thus leading to a decision support system that can be used in clinical practice.

One of the limitations of our study is the small patient cohort that includes only 13 patients. Let us point out that patients treated in the CNAO center are few since hadrontherapy is a specific technique applied only for those cancers where conventional radiotherapy does not provide significant advantages, such as for radio-resistant tumors and for those located close to organs at risk. However, the spread of cases in terms of geometry of thermoplastic masks is sufficient to draw general conclusions about the capability of our approach to improve the state of art, i.e. semi-random placements of markers, as well as previously presented approaches. Anyway, we plan to widen the cohort in the future work in order to further validate our results.

References

[1] Soete G, Van de Steene J, Verellen D, et al. Initial clinical experience with infrared-reflecting skin markers in the positioning of patients treated by conformal radiotherapy for prostate cancer. Int J Radiat Oncol Biol Phys 2002;2:694–8.

[2] Baroni G, Ferrigno G, Orecchia R, Pedotti A. Real-time 3-D motion analysis for patient positioning verification. Radiother Oncol 2000;54:21–7.

[3] Weiss E, Vorwerk H, Richter S, Hess CF. Interfractional and intrafractional accuracy during radiotherapy of gynecologic carcinomas: a comprehensive evaluation using the ExacTrac system. Int J Radiat Oncol Biol Phys 2003;56:69–79.

[4] Desplanques M, Tagaste B, Fontana G, Pella A, Riboldi M, Fattori G, et al. A comparative study between the imaging system and the optical tracking system in proton therapy at CNAO. J Radiat Res 2013;54:129–35.

[5] Riboldi M, Baroni G, Orecchia R, Pedotti A. Enhanced surface registration techniques for patient positioning control in breast cancer radiotherapy. Technol Cancer Res Treat 2004;3:51–8.

[6] Schöffel PJ, Harms W, Sroka-Perez G, Schlegel W, Karger CP. Accuracy of a commercial optical 3D surface imaging system for realignment of patients for radiotherapy of the thorax. Phys Med Biol 2007;52:3949.

[7] Krengli M, Gaiano S, Mones E, Ballarè A, Beldì D, Bolchini C, et al. Reproducibility of patient setup by surface image registration system in conformal radiotherapy of prostate cancer. Radiat Oncol 2009;4:9.

[8] Maurer Jr CR, Wang MY, Galloway Jr RL, Maciunas RJ, Allen GS. Registration of head volume images using implantable fiducial markers. IEEE Trans Med Imaging 1997;16:447–62.

[9] Fitzpatrick JM, West JB, Maurer Jr CR. Predicting error in rigid body, point-based registration. IEEE Trans Med Imaging 1998;17:694–702.

[10] Fitzpatrick JM, West JB. The distribution of target registration error in rigid-body point-based registration. IEEE Trans Med Imaging 2001;20:917–27.

[11] Liu H, Yu Y, Schell MC, O'Dell WG, Ruo R, Okunieff P. Optical marker placement in photogrammetry patient positioning system. Med Phys 2003;30:103–10.

[12] Korürek M, Nizam A. A new arrhythmia clustering technique based on ant colony optimization. J Biomed Inform 2008;41:874–81.

[13] Alexandridis A, Chondrodima E. A medical diagnostic tool based on radial basis function classifiers and evolutionary simulated annealing. J Biomed Inform 2014;49:61–72. http://dx.doi.org/10.1016/j.jbi.2014.03.008.

[14] Bellazzi R, Zupan B. Intelligent data analysis – special issue. Methods Inf Med 2001;40:362–4.

[15] Holmes JH, Peek N. Intelligent data analysis in biomedicine. J Biomed Inform 2007;40:605–8.

[16] Zupan B, Holmes JH, Bellazzi R. Knowledge-based data analysis and interpretation. Artif Intell Med 2006;37:163–5.

[17] Peek N, Swift S. Intelligent data analysis for knowledge discovery, patient monitoring and quality assessment. Methods Inf Med 2012;51:318–22.

[18] Paterson A, Ashtari M, Ribé D, Stenbeck G, Tucker A. Intelligent data analysis to model and understand live cell time-lapse sequences. Methods Inf Med 2012;51:332–40.

[19] Ochs MF. Knowledge-based data analysis comes of age. Brief Bioinform 2010;11:30–9.

[20] Hirsch M, Swift S, Liu X. Optimal search space for clustering gene expression data via consensus. J Comput Biol 2007;14:1327–41.

[21] Dinevski D, Povalej P, Kravos M. Intelligent data analysis for the diagnosis of alcohol dependence syndrome. J Int Med Res 2011;39:988–1000.

[22] Riboldi M, Baroni G, Spadea MF, Tagaste B, Garibaldi C, Cambria R, et al. Genetic evolutionary taboo search for optimal marker placement in infrared patient setup. Phys Med Biol 2007;52:5815–30.

[23] Shamir RR, Joskowicz L, Shoshan Y. Optimal landmarks selection and fiducial marker placement for minimal target registration error in image-guided neurosurgery. Proc SPIE – Med l Imaging 2009:7261.

[24] Rossi A. The status of CNAO. Eur Phys J Plus 2011;126:1–39.

[25] International commission of radiation units and measurements (ICRU). Prescribing, recording, and reporting photon beam therapy. ICRU 50; 1993.

[26] West J, Fitzpatrick JM, Toms SA, Maurer CR, Maciunas RJ. Fiducial point placement and the accuracy of point-based, rigid body registration. Neurosurgery 2001;48:810–6.

[27] Casotto A, Romeo F, Sangiovanni-Vincentelli A. A parallel simulated annealing algorithm for the placement of macro-cells. IEEE Trans Comput – Aided Des 1987;6:838–47.

[28] Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by simulated annealing. Science 1983;220:671–80.

[29] Cerny V. Thermodynamical approach to the traveling salesman problem: an efficient simulation algorithm. J Opt Theory Appl 1985;45:41–5.

[30] Lewis RM, Torczon V. A globally convergent augmented lagrangian pattern search algorithm for optimization with general constraints and simple bounds. SIAM J Opt 2002;12:1075–89.

[31] Conn AR, Gould NIM, Toint PL. A globally convergent augmented lagrangian algorithm for optimization with general constraints and simple bounds. SIAM J Num Anal 1991;28:545–72.

[32] Torczon V. On the convergence of pattern search algorithms. SIAM J Opt 1997;7:1–25.

[33] Audet C, Dennis JE. Analysis of generalized pattern searches. SIAM J Opt 2003;13:889–903.

[34] Conn AR, Gould NIM, Toint PL. A globally convergent augmented lagrangian barrier algorithm for optimization with general inequality constraints and simple bounds. Math Comp 1997;66:261–88.

[35] Nocedal J, Wright SJ. Numerical Optimization. New York: Springer; 2000.

[36] Chen H, Flann NS. Parallel simulated annealing and genetic algorithms: a space of hybrid methods. Lect Notes Comp Sci 1994;866:428–38.

[37] Bert C, Durante M. Motion in radiotherapy: particle therapy. Phys Med Biol 2012;56:R113–44.

[38] Riboldi M, Orecchia R, Baroni G. Real-time tumour tracking in particle therapy: technological developments and future perspectives. Lancet Oncol 2012;13:e383–91.

[39] Langen KM, Jones DT. Organ motion and its management. Int J Radiat Oncol Biol Phys 2001;50:265–78.

[40] Greening DR. Parallel simulated annealing techniques. Phys D: Nonlinear Phenom 1990;42:293–306.