# Combination of Image Registration Algorithms for Patient Alignement in Proton Beam Therapy

Rachid Belaroussi and Guillaume Morel

Institut des Systèmes Intelligents et Robotique,
4 place Jussieu, 75005 Paris, France
{belaroussi,morel}@robot.jussieu.fr
http://www.isir.fr

**Abstract.** We propose a measure of patient alignment in a video by combining different image representations : grey level, edges, and a set of feature points. When patient head is correctly positionned, a reference image with its ellipse is stored as a template of correct alignment. Edges detection results in a second template of the correct head location. Corners inside the ellipse are detected and tracked: a set of $N$ feature points composes a third template. Template matching computes a measure of similarity between a representation of the reference image and a window sliding around the reference location. Similarity with these three models are combined by the product rule. Location of window the most similar to the templates gives the translation $\mathbf{T}$ of the reference model in the image plane. This measure of patient misalignment could avoid X-ray verification of patient alignment, reducing patient dose and duration of treatment sessions.

**Keywords:** Proton beam therapy, expert combination, template matching, feature points, color model, camshift.

## 1 Introduction

Conformal radiotherapy is a recent approach for tumor removal that uses an aperture, a metal block with a hole through which the radiation beam passes. Each patient has a custom-made aperture, the shape of the hole is the approximate shape of the target being treated by the beam. Protontherapy aims energetic ionizing particles (protons) onto the target tumor: protons scatter less easily in the tissue than X-rays and the beam stays focused on the tumor shape without much lateral damage to surrounding tissue. Also, no proton penetrates beyond a depth corresponding to the Bragg peak: dosage to tissue is maximum in the tumor volume and almost null outside. A slight error in tumor positioning can therefore damage surrounding tissue: the required accuracy in patient positionning in the beam line is only 1 mm.

The patient positioning system is made of a patient couch moved by a robot: patient alignement refers to verification of the patient positionning. Before and during a treatment session, patient head is kept still on the couch by a thermoplastic

A. Elmoataz et al. (Eds.): ICISP 2008, LNCS 5099, pp. 183–191, 2008.
© Springer-Verlag Berlin Heidelberg 2008

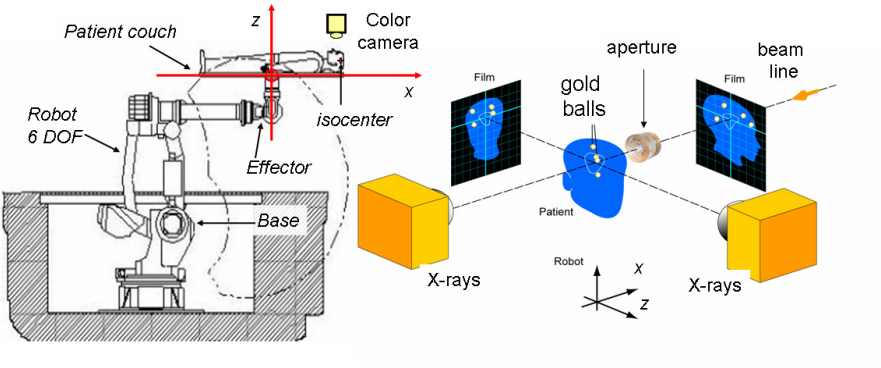


**Fig. 1.** Manipulator robot and patient couch: robot arm places a load from 15 kg to 150 kg at 3,50m from its base. Correctness of patient head alignement is validated using two orthogonal x-rays images to detect fiducial markers implanted in skull.

head mould. To ensure the brain tumor is well positioned in beam line a patient alignement is required before the treatment starts, as well as a patient monitoring during the treatment to stop the proton beam when patient head is moving.

Our goal is to develop a measure of positioning error of patient head, with a color camera. Our approach is non invasive, no marker of the tumour is visible so the object tracked is the head mould, which is a non rigid object. This task requires an approach accurate enough to measure small displacement $\leq 1$ mm, robust to lighting conditions and fast. Our perspective is a visual servoing of a 6 degree of freedom positioner robot : the difference with the template is corrected by a displacement of the robot.

A X-rays alignment procedure is used for patient setup verification : two X-rays images of orthogonal sections of the head are taken to validate the positionning as shown by figure 1. Fiducial markers (golden balls) implanted in the patient skull are detected by an operator, and a correction is sent to the positionner. The positionning is X-ray checked again until convergence. A similar approach of correspondance between current fiducial markers orthogonal images and reference is found in [12], and can be applied in radio-surgery [13]. A refinement step can be provided by an infrared stereovision camera POLARIS that returns 3D locations of five reflecting ball implanted on the plastic mould [6]. Other patient alignment systems use image registration techniques based on skull image, or external markers detected in 3D with several infrared stereovision cameras [9], [8]. For the treatment of mobile tumour in soft tissues like liver, such as infrared stereovision markers are used to monitor respiration phase, with electromagnetic markers implanted close to the tumour [11]. A model of tumour mobility versus surface displacement is built to compensate motion due to respiration phase. AlignRT [10] proposes a 3D surface reconstruction of a part of the body, using a structured lighting and a stereovision camera. It is used in patient respiration phase monitoring to command a 4 dof robot, and is tested in proton therapy for brain tumour at Massachusetts General Hospital. For prostate radiotherapy

targeting, a comparative study between ultrasound and x-ray imaging of markers is proposed in [7].

Our work aims at replacing that last validation step by a visual servoing of the patient positioning system, using a color camera. In this study, we investigate the use of template matching over different image representations to estimate the translation $\mathbf{T} = [T_x \quad T_y \quad T_z]^T$ vector required to alignate the current couch position to the correct patient positioning

Section 2 presents our procedure of patient alignment, head mould being modeled as an ellipse used for template building and matching. Section 3 describes models based on various representations of head mould (grey vel and image gradient), and section 4 experts combination is formuled. Experimental results are discussed in section 5, and section 6 gives conclusion and perspectives of this work.

## 2 Patient Alignment System Overview

In a first session, an image $I^*_{grey}$ of patient head is acquired when the patient positioning i known as correct: head location is represented by an ellipse with the state $\mathbf{s}^*$, estimated by CamShift. The template matching is implemented in a region close to the location of reference ellipse $\mathbf{s}^*$: the positioner translation error is supposed to be less than 1 cm.

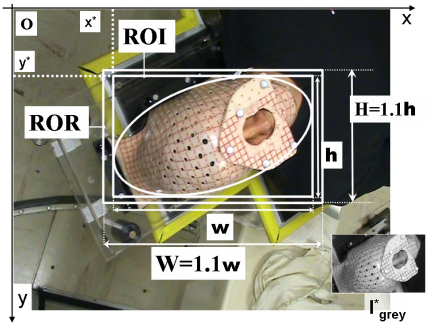


**Fig. 2.** Region of interest (ROI) is the rectangle bounding the template ellipse: region of research ROR of the matching window is set around the ROI

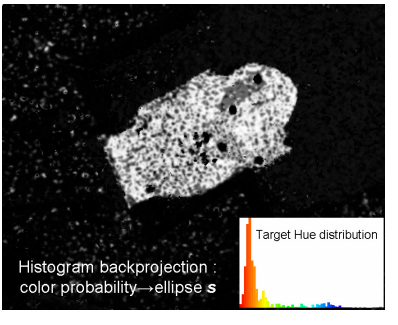


**Fig. 3.** Backprojection of head mould color model (a 64 bins histogram in the Hue channel of HSV): second moment order are computed to estimate ellipse

The initialisation stage stores the template image $I^*_{grey}$ and the binary image of the ellipse $\ell^*$ resulting from the CamShift [1]. $\ell^*$ bounding box is used to define the region of interest of the frame stored in $I^*_{grey}$:

$$\mathbf{ROI}^* = [x^*_{roi} \quad y^*_{roi} \quad w \quad h]$$

with $(x^*, y^*)$ top-left corner coordinate, $w$ is width and $h$ is height of $\mathbf{ROI}^*$. A Canny edges detection of $I^*_{grey}$ is computed: a second template is made off the

binary image $I^*_{edges}$ of $I^*_{grey}$ contours. All pixels of $I^*_{edges}$ outside of $\ell^*$ are set to 0. A set $\mathbf{E}^*$ of $N = 200$ 2D feature points of $I^*_{grey}$ are detected:

$$\mathbf{E}^* = \{\mathbf{x_k^*} = (x_k^*, y_k^*)\}_{k=\{0\ldots N-1\}}$$

The test stage starts with CamShift target initialisation, that can be done manually. Patient head is tracked by Camshift which results in an ellipse state estimation $\mathbf{s}(t)$ in the image at time $t$. Ellipse image $\ell(t)$ is used to mask pixels of edges $I_{edges}(t)$ and intensity $I_{grey}(t)$ images outside $\ell(t)$. Corners are detected at time $t = 0$ of target initialisation: $N$ feature points are searched in $\ell(0)$ ellipse. These points are further tracked at time $t > 0$ resulting in a set of $N$ points $\mathbf{E}(t)$:

$$\mathbf{E}(t) = \{\mathbf{x_k(t)} = (x_k(t), y_k(t))\}_{k=\{0\ldots N-1\}}$$

Distance between this set and $\mathbf{E}^*$ is used as a similarity measure. Results of template matching with the three models are combined to evaluate the translation error $\mathbf{T}(t)$. The time parameter is leaved when possible in the following sections, as the process applied to the current image except for feature points tracking.

## 3 Basic Models Description

### 3.1 Appearance Based Model

Template matching is realized by sliding a $w$x$h$ window and computing the Euclidian distance between $I^*_{grey}$ and overlapped $I_{grey}(t)$ region. It is computed on a $W$x$H$ region of research **ROR** centered on **ROI**$^*$, 10% wider and larger:

$$\mathbf{ROR} = [x_{ror} \quad y_{ror} \quad W \quad H]$$

The resulting array is named `GreyMap` and have a size of $(W-w+1)$x$(H-h+1)$, where each element correspond to a window in image at time $t$:

$$GreyMap(x_0, y_0) = a \sum_{x,y} \left(I_{grey}(x + x_0, y + y_0) - I^*_{grey}(x, y)\right)^2 + b$$

(a,b) are used to normalize `GreyMap` : it is inversed so that a high value represents a window close to the template, and its range is linearly scaled onto $[0 \quad 1]$.

In figure 4, `GreyMap` is a 44x26 array: 1144 $w$x$h$ window locations are evaluted, in the neighborhood of **ROI**$^*$. To evaluate accuracy of the approach, diagonal of **ROR** is made of about 560 pixels and represents 45 cm: image resolution is then $\approx 12$ pixels per cm. Our approach is then able to estimate a translation error with an accuracy of 0.8 mm. In the case of figure 4, patient is exposed with a diagonal incidence to the beam line: translation error range is $T_x \in [0 \quad 17.6\,mm]$ along the x-axis and $T_y \in [0 \quad 10.4\,mm]$ along y-axis.

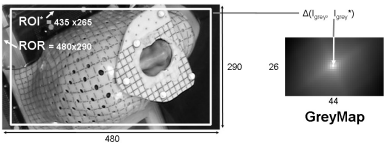


**Fig. 4.** Template Matching in the grey level image: GreyMap elements are correlation between reference and overlapped region

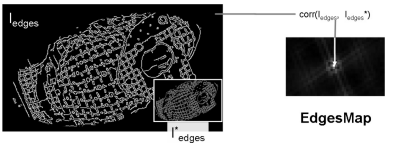


**Fig. 5.** EdgesMap is the result of correlation between template edges and edges image

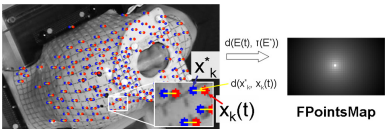


**Fig. 6.** FPointsMap is the tempate matching between reference feature points and set of points in the current image, using the Euclidian distance

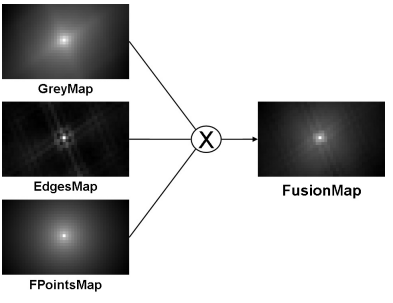


**Fig. 7.** FusionMap maximum correspond to window of current image the most similar with the reference models

### 3.2 Geometrical Model

A Canny detector is used to estimate edge pixels in the image : it results in a binary image $I_{edges}$. Correlation between $I^*_{edges}$ is calculated over the search region **ROR** resulting in `EdgesMap` array, of same size as `GreyMap`. It is a classical approach of image registration [14]: assumption is made of an affine transform between intensity of the template image and a window of the region of research.

$$EdgesMap(x_0, y_0) = \sum_{x,y} I^*_{edges}(x, y) I_{edges}(x + x_0, y + y_0)$$

In figure 5, `EdgesMap` array is illustrated: pixels intensity is proportional to the correlation between the $wxh$ block it represents and the template edges. It appears that this model is quite sensitive to translation in the image plane because of the grid drawn on the head mould. Global maxima represents the window that best matches the template, but we can also see some local maximum are regularly distributed in `EdgesMap`: they are due to the grid squares periodicity.

### 3.3 Structural Model

Ensemble $\mathbf{E}(t)$ of features point is detected and tracked during the sequence:

- feature points detection [3] computes correlation matrix of horizontal and vertical image gradient: pixels with the higher eigenvalues are detected.
- at time $t > 0$, optical flow of feature points is computed using iterative Lucas-Kanade [2] method in pyramids: each point $\mathbf{x_k(t)} \in \mathbf{E}(t)$ is calculated knowing $\mathbf{x_k(t-1)}$, $I_{grey}(t)$ and $I_{grey}(t-1)$.

The distance between $\mathbf{E}(t)$ and the reference set $\mathbf{E}^*$ is computed by sliding the template over the $\mathbf{ROR}(t)$ rectangle :

$$FPointsMap(x_0, y_0) = \frac{}{\frac{1}{N}\sqrt{\sum_{k=0}^{N-1}\left((x_k^* - x_{roi}^*) - (x_k - x_{ror}) + x_0\right)^2 + \left((y_k^* - y_{roi}^*) - (y_k - y_{ror}) + y_0\right)^2}}$$

where $(x_0, y_0)$ are coordinates of the top left corner of the $w$x$h$ region overlapped by the feature points template. Array `FPointsMap` size is also $(W-w+1)$x$(H-h+1)$, but the more a window is similar to the template, the lower the distance to the template is. It is inversed and linearly scaled onto $[0 \quad 1]$ as we did for `GreyMap` array. In figure 6, `FPointsMap` array values are video-inversed: the window closest to the template is represented by a white pixel, while a darker intensity represents a window less likely to match the template $\mathbf{E}^*$.

## 4 Models Combination

### 4.1 Translation Error in the Horizontal Plane

We computed three $(W - w + 1)$x$(H - h + 1)$ arrays storing the results of comparison between reference models and windows of the current image. In our experiments, the camera is placed above the patient couch so that the image plane is horizontal . Each of these map can be used to estimate the translation error $\mathbf{T_{hor}} = [T_x \quad T_y]^T$ in the image plane. An arbitration is necessary when the sources are in conflict: a data fusion methods is applied.

A recent study of information fusion applications and methods can be found in [15]. Goal of information fusion depends on the application : dimension reduction, classification, robustness to imperfect sources. Several architectures of data fusion are proposed in literature, and can be divided into three types: parallel, serial and hybrid. Level of data fusion refers to the degree of abstraction of inputs and outputs. At signal level, data are combined without transformation: it is the smallest abstraction level. At an intermediate level, features (dimension, area, compacity, mean ...) are extracted before combination. The highest level is the fusion of decisions or experts combination: this level is adapted to our issue.

Different methods of fusion could be used like weighted average, majority voting, maximum rule, fuzzy logic or product rule, as we only have one training sample for each image representation. The product rule is a quite efficient [4]

approach for combining experts output, it is close to intuition and more easy to interpret:

$$FusionMap = GreyMap \cdot EdgesMap \cdot FPointsMap$$

An advantage of the product rule is that the issue of comeasurability of sources is not raised. A weighted average requires normalization of physically different types of measurements : grey level, contours and feature points. A connexionist approach is not applicable: no image database is available and mould are custom-made anyway.

The product rule [5] assumes that the representations used are conditionally statistically independent. It can be a severe rule as a single expert can inhibit an interpretation by outputting a small similarity for it. However integration of new expert is simplified in this combinatory architecture; this modularity is desirable for further development. The translation error is estimated using `FusionMap` maximum location:

$$(x_{max}, y_{max}) = \max_{x,y} FusionMap(x, y)$$
$$\Rightarrow T_x = x^*_{roi} - (x_{ror} + x_{max}) \quad T_y = y^*_{roi} - (y_{ror} + y_{max})$$

### 4.2 Couch Altitude Estimation

Translation error in a vertical plane can be determined using the same approach with a camera which axis is orthogonal to the z-axis. This way, two cameras could do the same task as X-ray tubes used for patient position validation (see figure 1), and translation $\mathbf{T_{vert}} = [T_x \quad T_z]^T$ is estimated. However, in this study our test sequences are acquired with a single camera placed above the couch: we will suppose couch altitude constant, and the z-axis translation error is not further investigated in this paper.

## 5 Patient Alignement Performances

As mentionned earlier, our approach is able to estimate a translation error of 0.8 mm. The computation time is **180 ms per frame** on a Pentium Celeron @1.2GHz, 240Mo RAM, with no particular optimization. To process a 720x576 image, 50 ms are required by the CamShift and 130 ms for the image processing (edges detection, pyramidal images for optical flow), template matching and fusion algorithms with a 435x265 head size.

We tested a sequence of 1500 images with a specific sceniaro: patient couch is correctly positioned, and still. Patient is also still in the beginning of the sequence, then moves his head more or less. This scenario validate our approach as a misalignement due to head motion can be detected. To quantify sensitivity of our estimate of translation error, we computed the amount of motion pixels in the intersection of ellipses $\ell^*$ and $\ell(t)$. Motion detection is realized by frame differencing with the template grey level image. Temporal gradient of the images sequence is computed in $\mathbf{ROI}^*$ rectangle and binarized: pixels with module

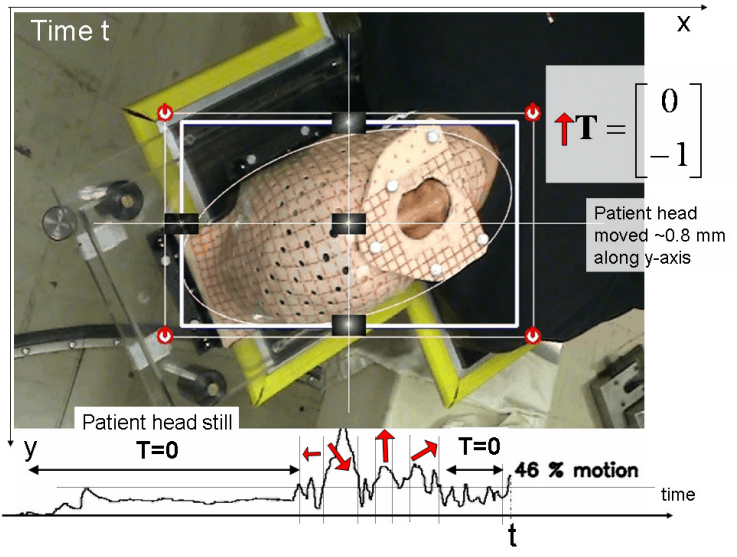


**Fig. 8.** Edges and feature points of templates (in blue) and in image at time t (in red): matching edges and point are drawn in white. Proportion of motion pixels in region of interest versus time during the test sequence: translation error is estimated as null when patient head is still, and is drawn with a red arrow when non zero.

higher than an experimentally defined threshold are classified as in motion in $D(t)$ image. Ratio between number of motion pixels and number of edges pixels is used to express the motion quantity in percent. When patient head moves more than 25%, a non zero translation vector is proposed by our algorithm.

In figure 8, an illustration is provided with at time t a motion of 46% and a translation error estimated of 1 pixel along the y-axis. At this moment, patient head is moving inside its mould, and the caused mould displacement is evaluated at $T_x = 0.8$ mm. It is worth noticing that when patient head is still, the translation error estimated is null, which is desirable too.

## 6 Conclusion and Perspectives

We presented three models of a reference based on various image representations: one based on grey level, two others based on image spatial gradient (edges and feature points). We proposed to aggregate these heterogenous sources for a measure patient alignment error in a color image. An interesting point of the approach is that, even if the grid drawn on the head mould makes the edges matching more efficient, it is not modality specific and it could be applied to patient alignment in radiotherapy or tomotherapy for example, with or without mould. Our first experiments of the proposed models fusion are very encouraging, our algorithm is fast and have a resolution of less than 1 mm. Alignement accuracy requires a ground truth to be estimated: this is to be done with an

infrared stereovision camera. Also, our next step is to implement a visual servoing of our 6 dof robot positionner based on our measurement.

## References


1. Bradsky, G.: Computer Vision Face Tracking For Use in a Perceptual User Interface. Intel Technology Journal (1998)
2. Bouguet, J.-Y.: Pyramidal Implementation of the Lucas Kanade Feature Tracker: Description of the algorithm. Technical report, Intel Corporation Microprocessor Research Labs (2000)
3. Shi, J., Tomasi, C.: Good Features to Track. In: Conference on Computer Vision and Pattern Recognition (1994)
4. Milgram, M., Belaroussi, R., Prevost, L.: Multi-stage Combination of Geometric and Colorimetric Detectors for Eyes Localization. In: 13th International Conference Image Analysis and Processing, pp. 1010–1017 (2005)
5. Kittler, J., Hatef, M., Duin, R., Matas, J.: On Combining Classifiers. Transactions on Pattern Analysis and Machine Intelligence 20(3), 226–239 (1998)
6. Pinault, S., Morel, G., Ferrand, R., Auger, M., Mabit, C.: Using an external registration system for daily patient repositioning in protontherapy. In: International Conference on Intelligent Robots and Systems (2007)
7. Fuller, D.C., Thomas, C.R., Schwartz, S., Golden, N., Ting, J., Wong, A., Erdogmus, D., Scarbrough, T.J.: Method comparison of ultrasound and kilovoltage x-ray fiducial marker imaging for prostate radiotherapy targeting. Phys. Med. Biol. 51(19), 4981–4993 (2006)
8. Baronia, G., Ferrignoa, G., Orecchia, R., Pedotti, A.: Real-time three-dimensional motion analysis for patient positioning verification. Radiotherapy and Oncology 54 (2000)
9. de Kock, E.A., Muller, N., Maartens, D., van der Merwe, J., Muller, D., van Rooyen, R., van der Merwe, A., Eksteen, J., von Hoesslin, N., Wagener, D., Hough, J.: Integrating an industrial robot and multi-camera computer vision systems into a patient positioning system for high-precision radiotherapy. In: ISR Symposium (2004)
10. Harms, W., Schoffel, P.J., Sroka-Perez, G., Schlegel, W., Karger, C.P.: Accuracy of a commercial optical 3D surface imaging system for rea- lignment of patients for radiotherapy of the thorax. Phys. Med. Biol. 52(5), 3949–3963 (2007)
11. Tanga, J., Dieterichb, S., Clearya, K.: Respiratory Motion Tracking of Skin and Liver in Swine for CyberKnife Motion Compensation. In: SPIE Medical Imaging (2004)
12. Verellen, D., Soete, G., Linthout, N., Van Acker, S., De Roover, P., Van de Steene, J., Vinh-Hung, V., Storme, G.: Quality assurance of a system for improved target localization and patient setup that combines real time infrared tracking and stereoscopic Xray imaging. Radiotherapy and Oncology 67(5), 129–141 (2003)
13. Gerszten, P.C., Ozhasoglu, C., Burton, S.A., Welch, W.C., Vogel, W.J., Atkins, B.A., Kalnicki, S.: CyberKnife frameless single fraction ste- reotactic radiosurgery for tumors of the sacrum. Neurosurg. 15(2) (2003)
14. Brown, L.G.: A survey of image registration techniques. ACM Computing Surveys 24(4), 325–376 (1992)
15. Valet, L., Mauris, G., Bolon, P.: A Statistical Overview of recent literature in Information Fusion. IEEE Magazine on Aeronautics and Electronics Systems 16(3), 7–14 (2001)