

Automatic tracking of implanted fiducial markers in cone beam CT projection images

T E Marchant^{a,b)}, A Skalski^{c)} and B J Matuszewski^{d)}

^{a)} Christie Medical Physics and Engineering, The Christie NHS Foundation Trust, Manchester, M20 4BX, United Kingdom

^{b)} Manchester Academic Health Science Centre, The University of Manchester, Manchester M20 4BX, United Kingdom

^{c)} AGH University of Science and Technology, al. A. Mickiewicza 30, Krakow, 30-059, Poland

^{d)} School of Computing, Engineering and Physical Sciences, University of Central Lancashire, Preston, PR1 2HE, United Kingdom

Abstract

Purpose: This paper describes a novel method for simultaneous intra fraction tracking of multiple fiducial markers. Although the proposed method is generic and can be adopted for number of applications including fluoroscopy based patient position monitoring and gated radiotherapy, the tracking results presented in this paper are specific to tracking fiducial markers in a sequence of cone beam CT projection images.

Methods: The proposed method is accurate and robust thanks to utilizing the mean shift and random sampling principles respectively. The performance of the proposed method was evaluated with qualitative and quantitative methods, using data from 2 pancreatic and 1 prostate cancer patients and a moving phantom. The ground truth, for quantitative evaluation, was calculated based on manual tracking performed by three observers.

Results: The average dispersion of marker position error calculated from the tracking results for pancreas data (6 markers tracked over 640 frames, 3840 marker identifications) was 0.25 mm (at isocentre), compared with an average dispersion for the manual ground truth estimated at 0.22 mm. For prostate data (3 markers tracked over 366 frames, 1098 marker identifications) the average error was 0.34 mm. The estimated tracking error in the pancreas data was < 1 mm (2 pixels) in 97.6% of cases where nearby image clutter was detected and in 100.0% of cases with no nearby image clutter.

Conclusions: The proposed method has accuracy comparable to that of manual tracking and, in combination with the proposed batch post processing, superior robustness. Marker tracking in CBCT projections is useful for a variety of purposes, such as providing data for assessment of intra fraction motion, target tracking during rotational treatment delivery, motion correction of CBCT, and phase sorting for 4D CBCT.

This is an author-created, un-copyedited version of an article accepted for publication in Medical Physics. It is posted here by permission of the AIP for personal use, not for distribution. The definitive publisher-authenticated version is available online at <http://dx.doi.org/10.1118/1.3684959>.

Copyright AIP 2012.

I. INTRODUCTION

Cone beam CT (CBCT) imaging is commonly used in radiotherapy for verification of patient and target position.^{1,2} A three dimensional image is reconstructed from a sequence of projection images acquired as the linear accelerator gantry rotates around the patient. The projection images themselves are often discarded once image reconstruction has taken place, however they contain a wealth of dynamic information about the patient and target position during the acquisition time of the scan. By analyzing the CBCT projection image sequence, this motion data can be obtained at no additional ionizing radiation dose to the patient. Most commonly the scan is acquired prior to delivery of the megavoltage treatment beam, but the advent of rotational IMRT brings the possibility to acquire CBCT data during treatment delivery.^{3,4}

Motion data derived from the projection image sequence can be put to a variety of uses, such as estimation of intra fraction motion,^{5,6} target tracking during rotational treatment delivery,^{7,8} correction of motion artifacts in the CBCT reconstruction,^{9,10} and phase sorting for 4D CBCT reconstruction.¹¹

In some cases it may be possible to directly track a feature of interest within the patient, for example a well defined lung tumor surrounded by low density lung tissue.¹²⁻¹⁴ However, other cases involve target objects that are not clearly visible on kV x-ray projection images. In this case one or more radio-opaque fiducial markers may be implanted to improve visibility, which provide a surrogate for target position.

Extraction of motion data from CBCT projection images requires robust methods to track objects in the image sequence. This remains a challenging task, even when using fiducial markers, for a number of reasons: a high level of noise due to x-ray scatter and the typically low radiation dose delivered during a single projection image acquisition; markers changing shape and size for different projection angles; occlusions and clutter caused by possible presence of foreign objects; markers overlapping with each other or being masked by anatomical structures; large variations of the marker and background intensities with projection angle. Additionally apparent marker displacement in two consecutive images can be quite large as it is a superposition of an intrinsic motion caused, for example, by respiration and an extrinsic motion induced by the sensor rotation.

Few algorithms for tracking of fiducial markers in CBCT projection image sequences have previously been reported in detail. Typically simple template matching or circle fitting methods have been mentioned,^{7,10,15,16,17} which while suitable in some applications, for example in the case of static and/or spherical markers, would not be robust to the range of difficulties described above. Two recently reported algorithms^{18,19} address the problem of changing marker shape with projection angle by creating a 3D marker model from the projected shape and position over a wide range of projection angles. These approaches give robust results but are difficult to implement in real time due to the requirement of the full projection set to identify markers and generate the marker model. Markers of similar shape and orientation could cause confusion and some tracking failures were observed due to overlying objects. There is also a parallel with marker tracking in megavoltage projection imaging where the impact of highly variable visibility and marker shape change with orientation have been addressed using rotating template models²⁰.

In this paper we present a new algorithm for tracking the position of fiducial markers through a sequence of CBCT projection images. The algorithm is designed to track multiple markers by using multi component score functions to select the most likely positions of each marker from a set of generated marker position hypotheses. The algorithm has similarities to that reported by Tang *et al.*²¹ for tracking markers in fluoroscopy image sequences. A number of important distinguishing features of our algorithm for tracking in CBCT projection sequences are: (i) use of prior knowledge of variation in projected marker position with gantry angle to improve robustness of marker identification when tracking multiple markers (ii) use of the mean shift algorithm instead of template matching, which provides higher accuracy due to explicit subpixel accuracy of marker position estimation and dynamic implicit estimation of marker appearance (iii) use of random sampling for hypothesis generation instead of deterministic evaluation of all possible marker

locations in the predefined size window, enabling efficient marker search in a much bigger region, and maintaining track of possibly widely spatially separated positional hypotheses.

II. METHODS AND MATERIALS

II.A Outline of the tracking methodology

The proposed tracking algorithm brings together prior knowledge about the apparent motion of the markers in the CBCT image sequence with a hybrid tracking algorithm combining two well known tracking methodologies namely mean shift^{22,23} and sequential random sampling.²⁴ Both methodologies on their own are well understood and widely applied.^{25,26} Interestingly these methods represent two completely different approaches to optimization of the cost/score functions defined within the tracking framework. Whereas the mean shift performs deterministic optimization, the sequential random sampling is an example of stochastic optimization. Figure 1 shows the outline structure of the proposed tracking algorithm. At initialization the estimated position of all the M markers $\{\hat{p}_1^m; m=1\dots M\}$ are selected manually in the first CB projection image. The positions of the corresponding markers in the subsequent projection images are estimated (tracked) using two interleaving stages of prediction and update. During the prediction stage, N ($i=1..N$) likely positions for the m^{th} marker, $p_{k+1}^{m,i}$, in the current ($k+1$) frame are calculated based on its estimated position \hat{p}_k^m in the previous frame. The prediction step consists of a deterministic propagation model [see Sec. II.D and Eq. (2)] and in some cases stochastic sampling, followed by image driven corrections provided by the mean shift algorithm. The stochastic sampling only takes place when there is considerable clutter in the image around the predicted marker position (see Sec. II.E for more details). In the update stage the N hypothetical marker positions are assessed using a multi component score function sf which for each marker m and each hypothesis i assigns a score value: $\pi_{k+1}^{m,i}$. The hypothesis with the highest score defines the estimated marker position in the $k+1$ image. This is equivalent to the maximum a posteriori probability approach (MAP)²⁷ using sequential Monte Carlo taxonomy.

Due to the presence of multiple markers, clutter and occlusions in the projection images, the posterior probability distribution for p_k^m conditioned on the available measurements could be highly non-Gaussian with multiple important modes. This prevents the use of standard Kalman Filter or Extended Kalman Filter²⁸ based trackers. Indeed the proposed tracker does not make any assumption about the form of the posterior distribution which can be arbitrary in principle.

Although in most cases this scheme provides accurate tracking results, sometimes when markers come very close together or indeed overlap in the projection images the tracking accuracy may drop and the tracker may confuse the identity of the overlapping markers. However, when the markers move apart the tracker is able to recover correct marker identity and resume tracking with normal accuracy due to the use of random sampling and a motion model in the propagation stage.

The rest of this section is organized as follows: in Sec. II.B a brief description of the cone beam CT projection images is given; Sec. II.C explains the motion model for the markers used by the tracker and introduces the ground truth measurements essential for tracker evaluation. The main building blocks of the proposed tracker: propagation, mean shift mode seeking, and measurements are described in Secs. II.D-II.F respectively.

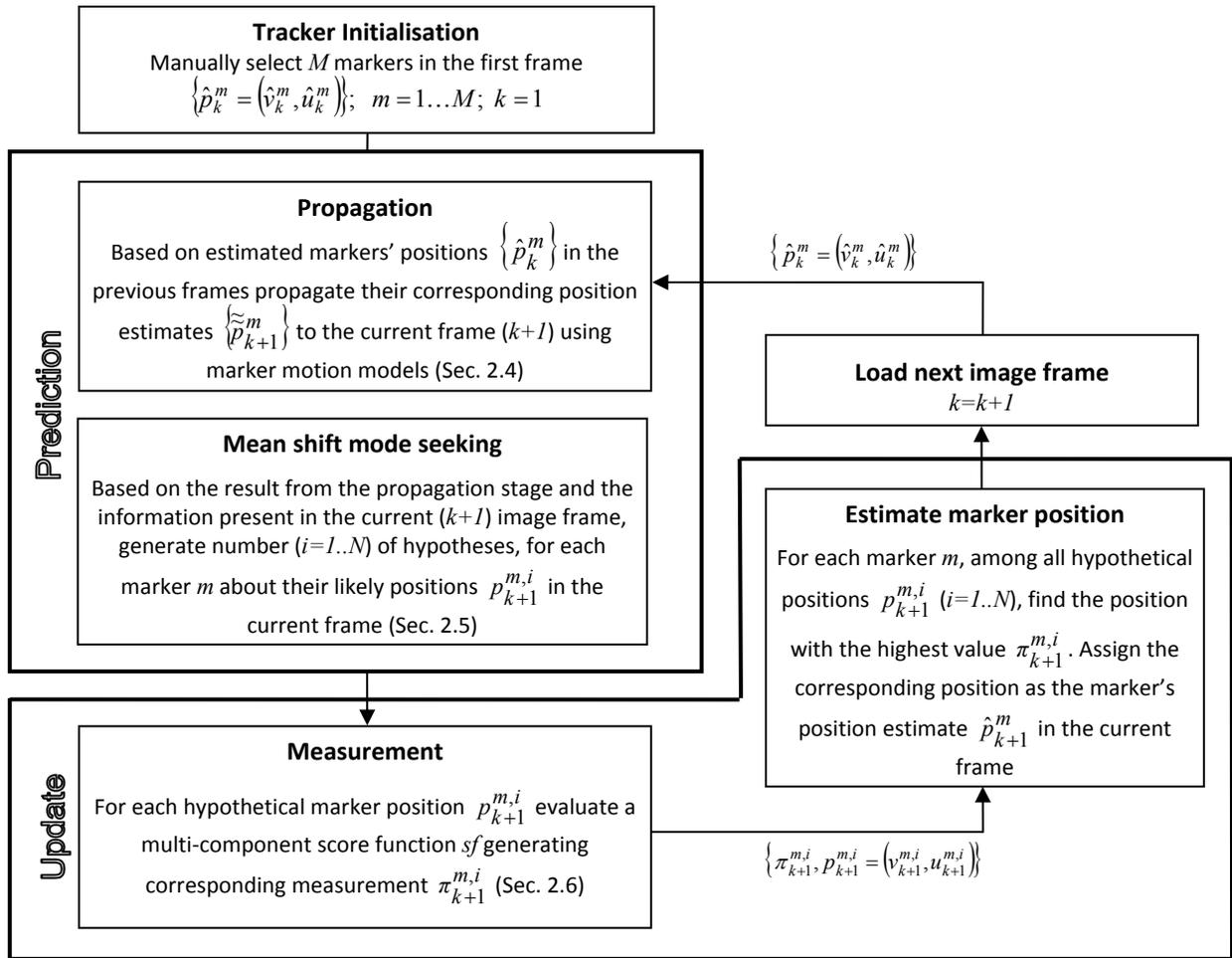


FIG. 1. A block diagram of the proposed tracking algorithm (see relevant sections for symbol definitions).

II.B Cone beam CT projection images

The method was tested using projections from CBCT images of two pancreatic cancer patients, each with 6 gold fiducial markers (1 mm x 10 mm) placed into the pancreas at surgery prior to radiotherapy, one prostate patient, with 3 gold markers (1 mm x 5 mm) implanted into the prostate, and also a breast phantom placed on a motorized platform. The CBCT projection images were acquired using Elekta Synergy (XVI, Elekta, Crawley, UK). This system has a kV imager fixed to the rotating gantry, mounted orthogonally to the MV treatment beam of a radiotherapy linear accelerator. Projection images were captured over 360° of rotation using a 512×512 matrix with square pixels of size $s = 0.8$ mm at the detector. For the pancreas images approximately 640 projections were acquired with frame rate 5.5 Hz at 120 kV, 10 mA, 40 ms per projection giving a dose of approximately 9mGy per scan (2/3 peripheral + 1/3 central) within the imaged FOV. For the prostate image 366 projections were acquired at 130 kV, 40 mA, 30 ms per projection giving 20mGy within the imaged FOV. Figure 2 shows sample projection images of the same patient acquired for different gantry rotation angles, illustrating changes in signal level and contrast.

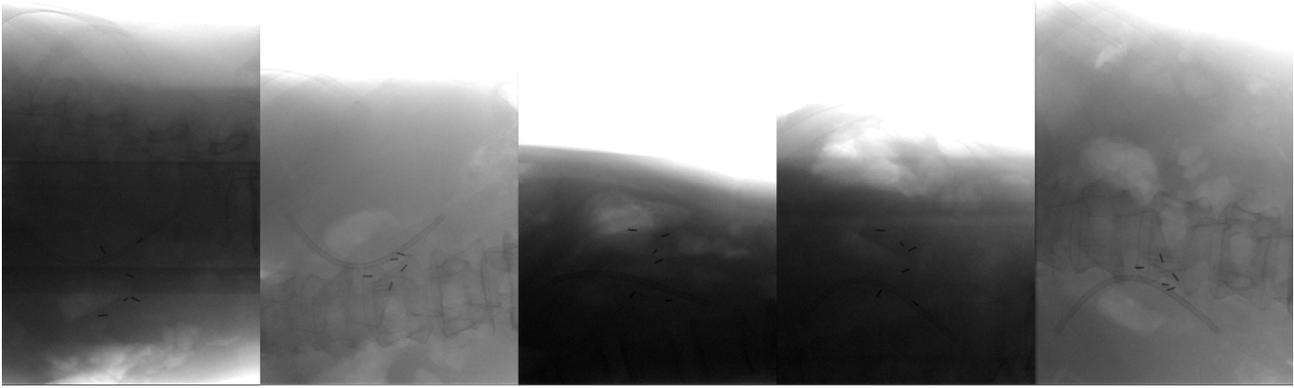


FIG. 2. Sample of CBCT projection images acquired for the same patient at different gantry rotation angles, illustrating changes in signal level and contrast.

II.C Marker motion in the projection images

Assuming that the position of a marker in 3D space is fixed, its apparent motion in the projection images as a function of gantry angle is given by:⁵

$$u(\theta) = \frac{SDD}{s} \times \frac{\sin(\theta)x + \cos(\theta)y}{SAD - (\cos(\theta)x - \sin(\theta)y)} + o_u \quad v(\theta) = \frac{SDD}{s} \times \frac{z}{SAD - (\cos(\theta)x - \sin(\theta)y)} + o_v, \quad (1)$$

where $u(\theta)$ and $v(\theta)$ are respectively row and column coordinates of the marker in the projection image acquired at gantry angle θ , (x,y,z) are 3D coordinates of the marker, o_u and o_v represent the position of the principal point in the image coordinates; θ is the known gantry rotation angle, s is the pixel size, SDD and SAD are the known distances from the source to the detector and gantry rotation axis respectively. The detector is assumed to have no tilt and moves in a circular orbit. For convenience it is assumed that the rotation axis is aligned with the z axis of the world coordinate frame and parallel to the sensor horizontal axis (v coordinate). For the Cone Beam imaging system used to acquire images shown in this paper $SDD=1536$ mm and $SAD=1000$ mm. The apparent 2D motion of the marker in the sequence of projection images is a superposition of the motion induced by the rotating gantry and the motion due to 3D marker movement. In the data shown in this paper the marker displacement in 3D space is attributed mainly to respiration.

This paper deals specifically with the detection of the marker positions in each projection image with coordinates estimated with respect to the image coordinate system. When using these marker positions to estimate marker location within the patient it may also be necessary to consider other sources of error e.g. machine sag.²⁹

To provide means for algorithm validation, ground truth data was gathered for one of the pancreatic cancer patient image sequences and for the prostate image sequence. The ground truth data describes the position of the marker midpoint for each marker in each projection image. It was calculated as an arithmetic mean from three manual measurements taken by three different observers. Two of these observers were asked to select a midpoint on the markers whereas the third was asked to select the marker's end points and the midpoint was calculated from the selected endpoints. The second approach although more time consuming gives slightly more stable results as marker endpoints are easier to define in the image than a mid marker point. Note that the marker mid-point in the projection images does not correspond exactly to the same 3D point for all projections since the portion of the marker closer to the source will be magnified to a greater degree than the portion of the marker further from the source. However the resulting uncertainty in marker position is neglected since it is far smaller than the inter-observer variations in estimating the ground truth positions.

The apparent marker motion in the projection images due to gantry rotation and actual marker displacement in 3D space can be separated by fitting functions $u(\theta)$ and $v(\theta)$ given in (1) to

the observed marker positions $p_k^m = (v_k^m, u_k^m)$ in the projection images, with k and m indexes identifying projection image and marker respectively. This has been implemented using the Levenberg-Marquardt method to optimize a nonlinear least squares cost function. This gives an estimate of the mean position of each marker in 3D space $(\bar{x}^i, \bar{y}^i, \bar{z}^i)$. When the mean position of each marker is estimated for the first time the initialization used for the optimization is $x = 0, y = 0, z = 0$. For each subsequent estimation the initial position is chosen as the previous solution. The trajectories $u^i(\theta)$ and $v^i(\theta)$ for the first marker ($i=1$) are shown in Fig. 3. From this it can be concluded the apparent motions in the u and v directions are considerably different. As seen from this figure, changes in the u coordinate of the projected marker position are dominated by the motion of the gantry. Most of the marker position variability is well explained by the model of motion induced by gantry rotation with a maximum error of 4 pixels. The variation of the v coordinate is mainly due to respiration as can be deduced from the semi-periodicity of the oscillations in the v direction with an average period of 3 seconds. The contribution of the gantry motion model $v(\theta)$ to the overall motion in that direction is very small. This can be explained in part by the fact that the v direction coincides with the gantry rotation axis (z axis of the world coordinate system), whereas the u axis is perpendicular to the rotation axis and as such this coordinate is more affected by the gantry rotation.

The above analysis was repeated for all the markers and it can be concluded that the results shown for the first marker are representative for all the other markers.

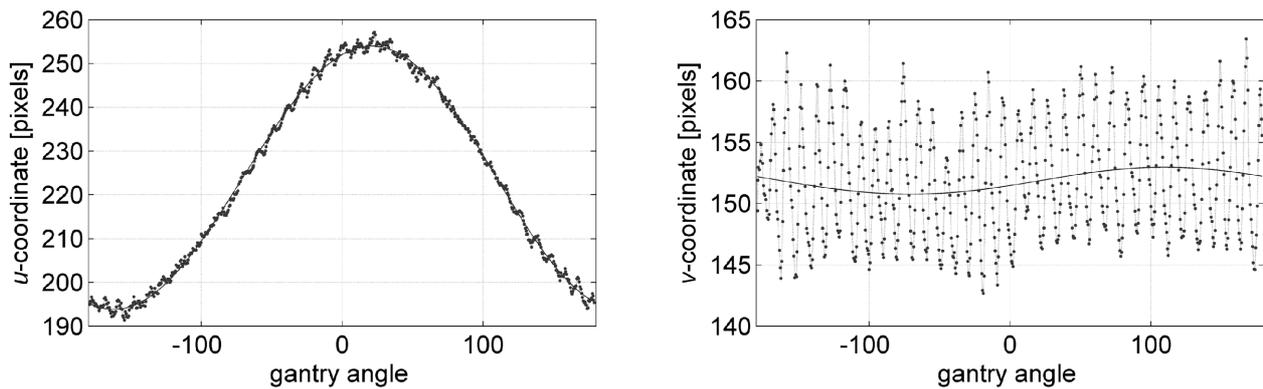


FIG. 3. $u_i(\theta)$ and $v_i(\theta)$ trajectories, shown as solid lines, obtained for the first marker and ground truth data shown as dots.

II.D Propagation

As mentioned in Sec. II.A most tracking algorithms can be divided into two general intertwined stages namely prediction and update or measurement. In the proposed algorithm the prediction stage is further divided into propagation and mode seeking stages. The propagation stage aims at finding an initial likely position of the marker in the current ($k+1$) projection image. This initial prediction is given as an output from a deterministic propagation function taking as arguments estimated positions of the marker in the previous images, up to the k th image. The subsequent mode seeking stage (see Sec. II.E) aims to improve this prediction by searching for the modes in the image intensity pattern which are in agreement with the current estimate of the marker intensity pattern. This mode seeking stage uses randomized search methodology, contrary to the propagation stage which is entirely deterministic.

The simplest, and probably the most often practically used, propagation model is the first order predictor where the marker position estimated in the previous image (k) is used as the initial position prediction in the current image ($k+1$). As the first order propagation model is very simplistic it is expected that the actual marker position may be considerably different. This effectively means that the search undertaken in the next stage needs to be performed in a relatively

big neighborhood of the predicted position to make sure that the true marker position can be recovered. This of course has a negative effect on execution time but also increases the probability of the tracker being distracted by clutter present in the images. Such a simple propagation algorithm can still be justified in cases where the markers are well separated and there is little or no clutter in the neighborhood of the prediction so marker association errors are unlikely. If there is clutter, including other markers, in the neighborhood of the m^{th} marker a better prediction should be used if possible.

To improve the robustness of marker position prediction the prior knowledge of its motion pattern is utilized. As shown in the previous section the marker's u coordinate is strongly dependant on the gantry rotation angle, and therefore its prediction in the next image can be obtained from the function given in Eq. (1) using the best estimate of the marker mean 3D position calculated as described above from the reliable marker position estimates available so far. Marker position estimates where there are no other proximate markers and little nearby clutter are considered reliable. We estimate clutter using local image statistics in a small window (21 x 21 pixels) around the estimated marker position. We measure the percentage of pixels $\delta_{clutter}$ which differ from the marker intensity by less than σ_{int} . For images normalized into the range 0-1000, we used $\sigma_{int}=20$. A marker was judged to be affected by low contrast clutter if $\delta_{clutter}>10\%$.

Unfortunately marker v coordinate is not well correlated with gantry rotation angle. However, in cases where each marker is attached to the same anatomical organ of interest, marker motion in the v direction is strongly correlated with the v motion of the other markers. The estimated displacement of marker m , used to predict its v position in the $k+1$ image, is therefore calculated as the mean of the displacements calculated for all the other reliably identified markers.

It is not always possible or advisable to use these prediction methods. It is required that reliable estimates of the marker mean 3D positions are available for prediction of the u coordinate, and reliable position estimates for some other markers are available for prediction of the v coordinate. In the experiments reported in the results section the marker mean 3D position was deemed as an acceptable estimate if more than 15 previous marker positions are available. Subsequently if the estimate of the marker position in the next image is deemed reliable this new estimate is used in re-estimation of the marker mean 3D position for use in the subsequent image frames.

The final propagation formulas for \tilde{u}_{k+1}^m and \tilde{v}_{k+1}^m , the predicted u and v positions of marker m in the current frame ($k+1$) respectively, are given as follows:

$$\tilde{u}_{k+1}^m = \begin{cases} \text{use first-order predictor } (\hat{u}_k^m) & \text{if there is no clutter or the gantry rotation model } u^m(\theta_{k+1}) \text{ cannot be used} \\ \text{use gantry rotation model } \hat{u}^m(\theta_{k+1}) & \text{elsewhere} \end{cases} \quad (2)$$

$$\tilde{v}_{k+1}^m = \begin{cases} \text{use first-order predictor } (\hat{v}_k^m) & \text{if there is no clutter or other reliable markers are not available} \\ \text{use other reliable markers in image } k+1 & \text{elsewhere} \end{cases}$$

II.E Mean shift mode seeking

The propagation described in the previous section gives an initial position of the marker in the next image. This is a rather rough approximation as no information from this new image is used in this process. To utilize information present in the new image, the proposed tracking approach, employs the mean shift technique,²⁵ to search for the modes (local peaks) in the intensity pattern which are the most likely representation of the projected markers. This process can be understood as estimating modes of a probability density function. Mean shift starts at the initial position calculated in the propagation step and subsequently iteratively improves the estimated position by climbing towards the local peak of the intensity distribution pattern (for convenience the image intensity has been inverted with markers represented by high value, bright pixels). This is achieved by calculating the weighted first moment of the pattern with a kernel representing the expected size and orientation of the marker. A 2D Gaussian kernel is used, whose size and orientation can be

selected to represent spherical or elongated fiducial markers of various designs and viewed at different orientations. A more complex kernel could be adopted if it was required to track more complex objects. The estimated marker size and orientation in the projection image, encoded by the Gaussian kernel is updated in each image as long as there is no clutter in the proximity of marker m . The update is performed in a way previously proposed by Bradski.³⁰ Figure 4 illustrates the Gaussian kernel estimated for example markers. For isolated markers [with no clutter or other proximate markers, shown in Fig. 4(a) and 4(b)] the size of the kernel is increased as to improve convergence properties of the mean shift algorithm. Figure 4(c) shows a case with other proximate markers. In this case the sizes of the kernels are not increased so as to improve spatial resolution of the mean shift algorithm. The kernel size and orientation are not re-estimated until the markers separate, and therefore some small errors in kernel size and orientation can be seen in Fig. 4(c).

In the case when there is considerable clutter in the image, including presence of other proximate markers, the procedure needs to be modified as the mean shift algorithm only guarantees convergence to a local maximum. To increase the probability of convergence to the correct mode in this case, the result of the propagation step is used to define the centre of a window in which N random positions $\tilde{p}_k^{m,i}$ ($i=1\dots N$) are drawn and passed to the mean shift algorithm as the initial positions. The positions in the window are selected with a probability defined by the pixel intensity. This approach enables the method to estimate the positions of all local modes. The detected modes are subsequently passed to the measurement stage where one of them is selected as the best estimate as described in Sec. II.F.

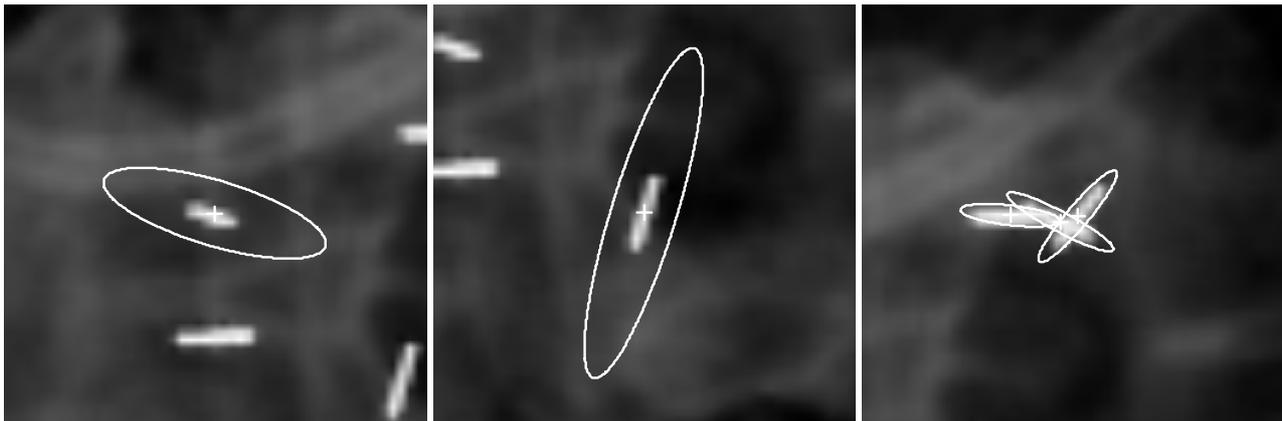


FIG. 4. Gaussian kernels, represented by 0.5 isocontours, estimated for different markers using the mean shift algorithm: (a) and (b) show kernels for two markers without clutter which are well separated from other markers, the estimated Gaussian kernel position together with kernel parameters describe marker position, size and orientation; (c) shows three overlapping markers with corresponding Gaussian kernels.

II.F Measurements

The propagation stage and the mean shift mode seeking stage, described in the two previous sections, provide hypotheses for the estimated marker position. The number of these hypotheses depends on the local image complexity. This may include image clutter, and/or presence of other markers being close or indeed overlapping with the marker for which position is estimated. For complex configurations the mode seeking stage can generate tens of hypothetical positions for each marker, or just a single one if there is no image clutter or other adjacent markers. In the case of multiple marker position hypotheses each predicted position is given a score, based on a number of criteria, assessing how well the given prediction describes the prior knowledge about the marker, improving accuracy and helping to recover from possible track losses. The score function adopted in this paper is defined as:

$$\pi_k^{m,i} = sf(p_k^{m,i}) = \alpha_{Ms} S_{Ms}(p_k^{m,i}) + \alpha_{Mu} S_{Mu}(p_k^{m,i}) + \alpha_{Mv} S_{Mv}(p_k^{m,i}) + \alpha_{Ds} S_{Ds}(p_k^{m,i}) \quad (3)$$

where: $p_k^{m,i} = (v_k^{m,i}, u_k^{m,i})$ is i th position prediction for m th marker in the k th image; $s_{Ms}(p_k^{m,i})$, $s_{Mu}(p_k^{m,i})$, $s_{Mv}(p_k^{m,i})$, and $s_{Ds}(p_k^{m,i})$ are respectively mode similarity measure, u and v coordinate prediction measures, and distance measure to the closest marker; α_{Ms} , α_{Mu} , α_{Mv} , and α_{Ds} are weighting factors assigned to each component of the score function. The position prediction for each marker having the maximum score function value is selected as the final estimate of that marker's position.

Mode similarity measure is directly obtained from the mean shift mode seeking stage. It indicates how strongly the intensity pattern around the predicted position (mode location) reflects the expected marker shape and orientation as encoded by the Gaussian kernel.

u coordinate prediction measure penalizes (assigns low score values to) predicted marker positions for which the error between $u_{k+1}^{m,i}$ and the corresponding model prediction $u^m(\theta_k)$ differs significantly from what is expected. Based on the analysis in Sec. II.D $s_{Mu}(p_k^{m,i})$ is modeled as having a Gaussian distribution with mean value of $\hat{u}^m(\theta_k)$ calculated from Eq. (1) using all reliable position estimates \hat{u}_n^m for all $n=1\dots k-1$, and the dispersion σ_u set as one of the method design parameters.

v coordinate prediction measure weights the v coordinate of the predicted marker position $v_{k+1}^{m,i}$ with respect to the corresponding coordinate \hat{v}_k^m obtained for the previous frame, assuming the v motion model introduced in Sec. II.D [Eq. (2)]. This component of the score function is only used if there is at least a single marker from which the predicted v position can be calculated. The prediction error $v_{k+1}^{m,i} - \hat{v}_k^m$ is modeled using a Gaussian distribution with a mean value given as an average v motion between frames k and $k+1$ calculated for all reliable markers and dispersions σ_v , set as a method design parameter.

Distance measure to the closest marker is used to “encourage” separation of the overlapping markers. Without this component, on some occasions when markers are crossing each other, the estimated positions for two or more markers can all be assigned to the marker which is better defined in the image. This component has only local influence as the u direction prediction quality measure would dominate when the markers are becoming significantly separated in u direction, and therefore this is only used when the minimum distance between any of the predicted positions and the position of another marker is smaller than a predefined threshold value. Predicted positions with greater separation from the other markers are assigned higher scores.

The corresponding weighting parameters α_{Ms} , α_{Mu} , α_{Mv} , α_{Ds} are problem dependent and are selected so that $\alpha_{Ms} + \alpha_{Mu} + \alpha_{Mv} + \alpha_{Ds} = 1$.

III. EXPERIMENTAL RESULTS

The performance of the proposed method has been tested quantitatively against manually generated ground truth data for one pancreas and one prostate image sequence. Further qualitative assessment was made using a second pancreas patient image sequence and a phantom image sequence. The images were acquired using the system described in Sec. II.B, with the ground truth data generated as described in Sec. II.C. The values of the design parameters were selected based on the characteristics of the data and required performance. In the mode seeking stage the design parameters include size of the resampling window, number of samples, and the scaling factor σ_g

controlling the size of the Gaussian kernel in the mean shift algorithm. The size of the sampling window should be big enough to encompass possible errors between the true and predicted marker position. In our experiments the window size was set to 13×13 . The number of random samples should reflect the complexity of the image, for more complex images with significant clutter and closely clustered markers more samples should be used. For the results shown the number of

random samples was set to 20. The size of the kernel in the mean shift algorithm, as explained in Fig. 4, controls the resolution of the mode detection. For the results shown $\sigma_g = 0.6$ when markers are clustered closely together and $\sigma_g = 2$ otherwise. In the measurement stage, based on the analysis of the ground truth data, the values of the dispersion parameters: σ_u , σ_v for the u and v prediction measures were chosen as $\sigma_u = 2$, $\sigma_v = 3$ if only one other marker is available for prediction of v position, and $\sigma_v = 0.7$ if 2 or more markers are available. The weights: α_{Ms} , α_{Mu} , α_{Mv} , α_{Ds} defining contribution of the different components to the score function, should be chosen based on the available prior knowledge. For example when markers are not too close to each other α_{Ds} can be set to zero, or when there is no strong correlation between markers' motion in v direction the α_{Mv} should have a small value. In principle when markers are well defined in the projection images (with no image clutter or marker clustering) all weights but α_{Ms} could be set to zero. For the experiment reported in this paper: $\alpha_{Ms} = 0.4$, $\alpha_{Mu} = 0.25$, $\alpha_{Mv} = 0.25$, and $\alpha_{Ds} = 0.1$. The proposed algorithm has been tested with different values of the design parameters. Based on the results it can be concluded that the performance of the proposed method is not particularly sensitive to changes of any of the design parameters. In all the experiments it was deemed that markers are close when the distance between them was below 12 pixels.

Figures 5-7 show typical effects on the tracker performance when some of the score function's components are not used. The u coordinate prediction measure effectively imposing constraint on the motion in the u direction, derived from the estimates of the mean 3D marker positions seems to have the strongest regularizing effect on the tracker. As seen in Fig. 5 when this constraint is not used the tracker may loose track of a marker, due to clutter or presence of other proximate markers and may never recover from it.

The *mode similarity measure* and the v coordinate prediction measure components have more local regularization effects. Both of them are effective in the presence of other proximate markers and image clutter but contrary to the u coordinate prediction measure if these are not used the tracker is still capable of recovering from tracking errors when the marker separates from image clutter or other markers (see Fig. 6 around gantry angle 100 degrees and Fig. 7 around gantry angles -70 and 100 degrees).

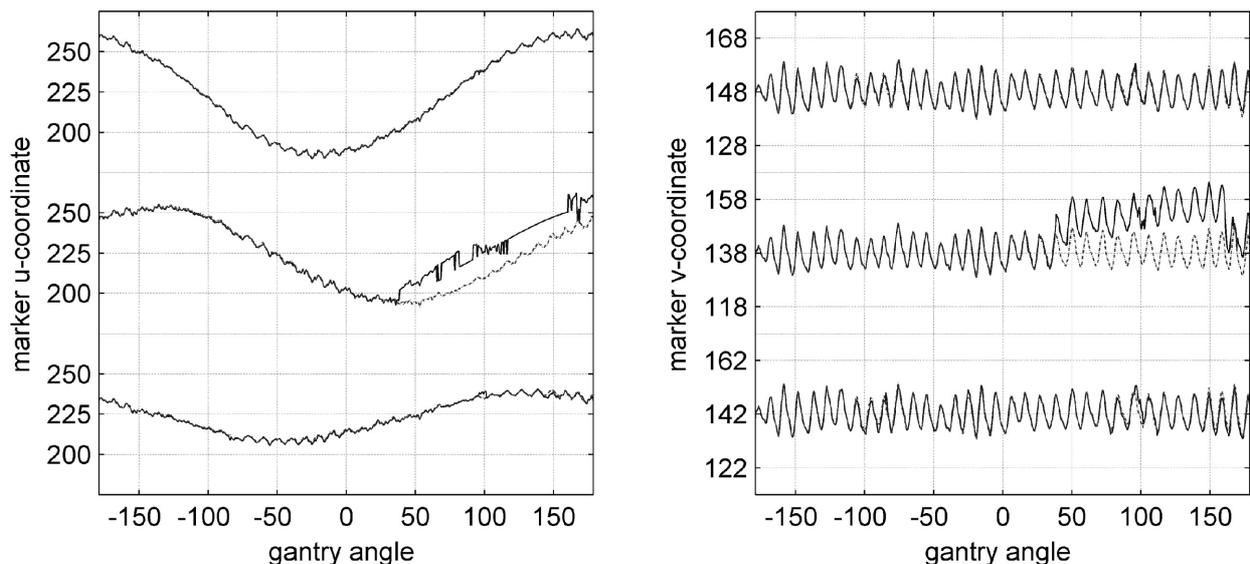


FIG. 5. Example tracking results when u -coordinate prediction measure, s_{Mu} , is not included in the score function. Dotted line represents ground truth and the solid line represents tracking results.

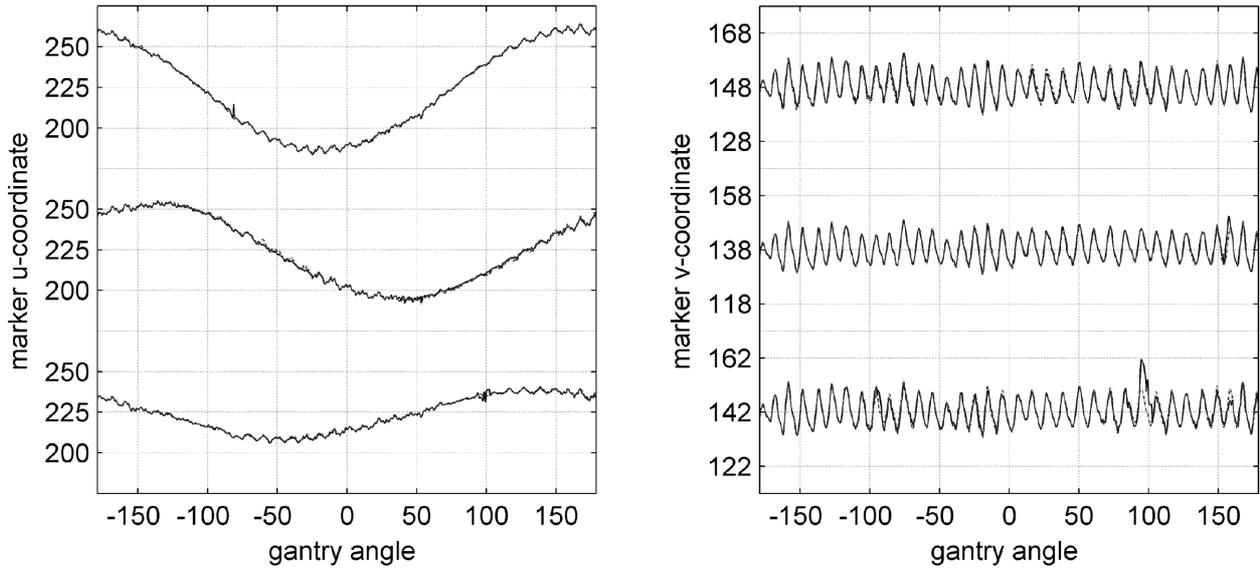


FIG. 6. Example tracking results when v direction prediction measure, s_{Mv} , is not included in the score function.

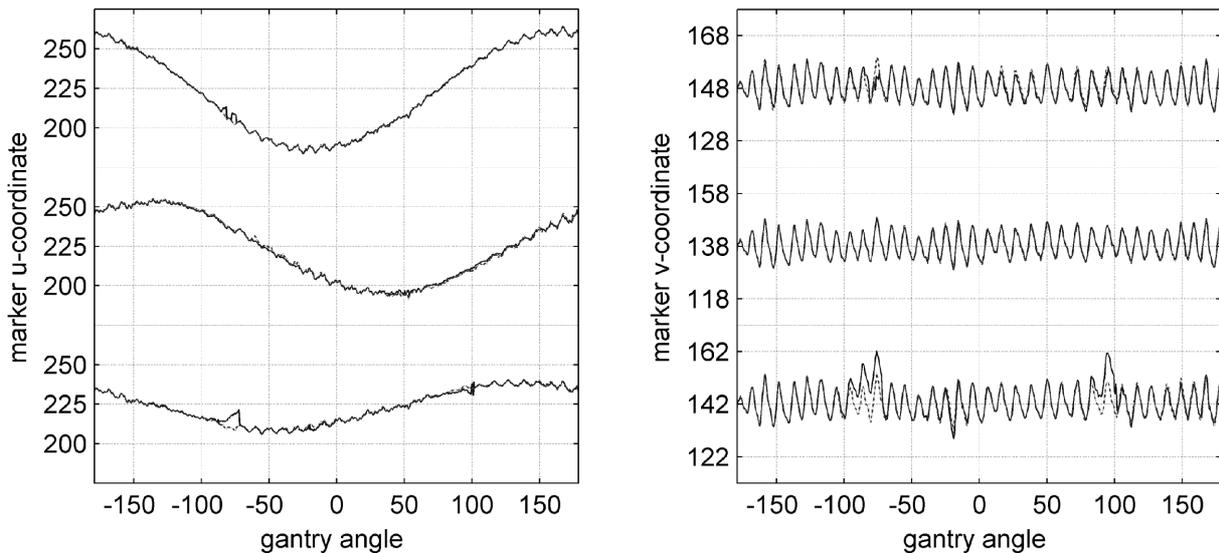


FIG. 7. Example tracking results when the mode similarity measure, s_{Ms} , is not included in the score function.

The working of the algorithm is illustrated in Fig. 8 with four images extracted from the CBCT projection sequence showing behavior of the tracker for clustered markers. In these images the ground truth marker positions are shown by circles and the corresponding estimated marker positions by crosses, with the results from the mean shift mode seeking stage represented by squares. The sequence demonstrates an example where the identities of some markers are wrongly estimated. The sequence from left to right shows (a) two of the markers getting closer until (b) complete overlap, (c) subsequently markers separate but the identity of the markers is wrongly estimated (mid-gray cross overlap with light-gray circle), (d) eventually the tracker recovers when the markers move further apart due to dominant contribution from the u coordinate prediction measure, s_{Mu} , in the score function. In the presented example the problem caused by the overlapping markers had been made even worse by the fact that the two markers have the same orientation and therefore response from the tracker (mode similarity measures) using the corresponding mean shift kernels returned for the two markers could be very similar.

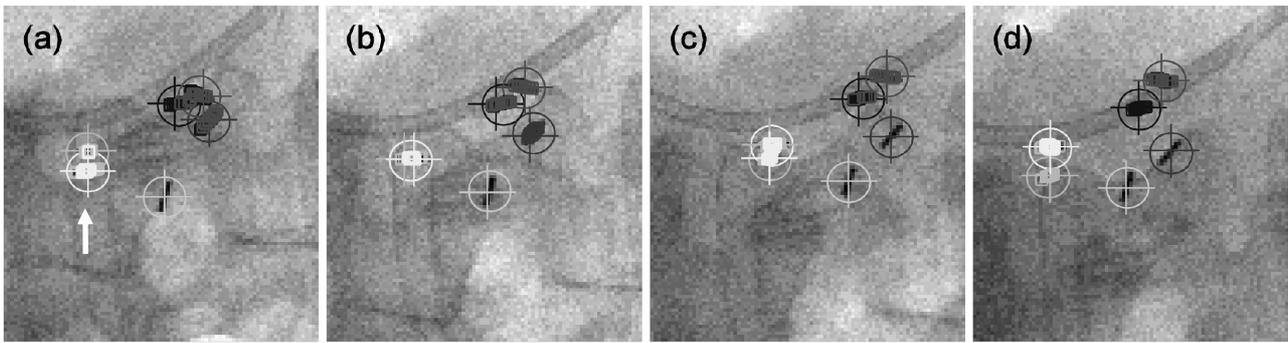


FIG. 8. Images (a)-(d) show tracking results for overlapping markers: (a) two of the markers are getting closer (indicated by white arrow); (b) the two markers are overlapping; (c) the two markers are separated but their identity is wrongly estimated; (d) when the two markers move further apart the tracker is able to recover correct identities.

Figure 9 shows pancreas marker tracking results as a function of gantry angle for markers 2, and 6. To better visualize the accuracy of the tracker Fig. 10 shows the difference (black line) between estimated and ground truth position for the u and v direction for markers 2 and 6. For reference standard deviation limits $\pm 3\sigma$ for the ground truth are also shown as gray lines. The ground truth standard deviation changes significantly from one image to another as it is estimated from only three measurements per marker. To reduce this effect the dispersion shown in the figure has been filtered by a moving average filter. As seen most of the error is within or close to $\pm 3\sigma$ of the ground truth apart from a few spikes caused by temporary errors in marker identity estimation (for the gantry rotation angle around -75°). It is evident that the proposed tracker is able to estimate position of the fiducial markers accurately.

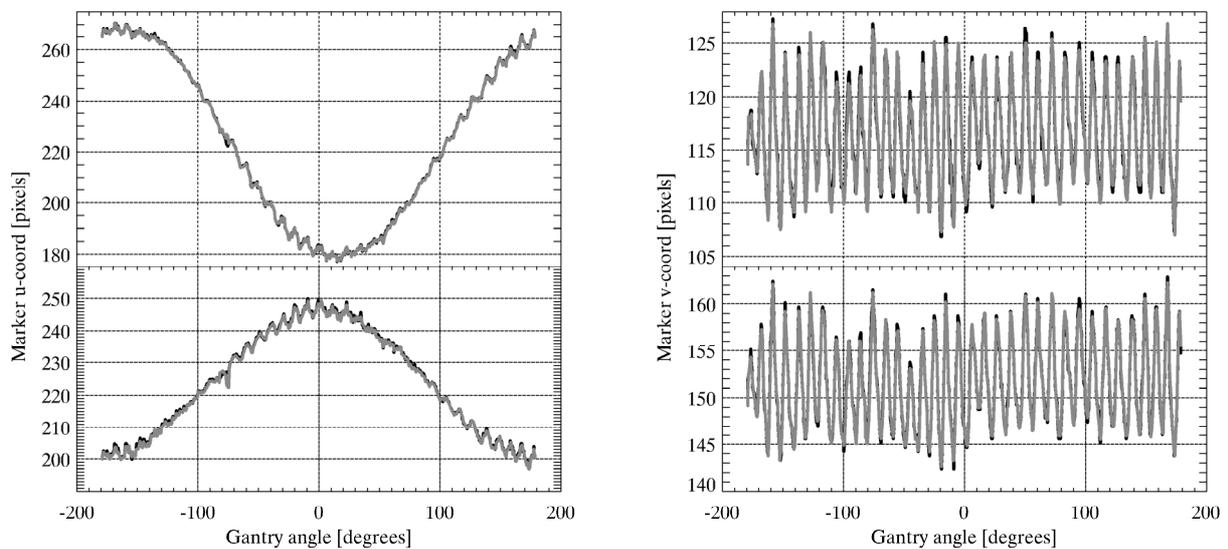


FIG. 9. Tracking results for markers 6 and 2. Black line shows ground truth and the gray line estimated markers position.

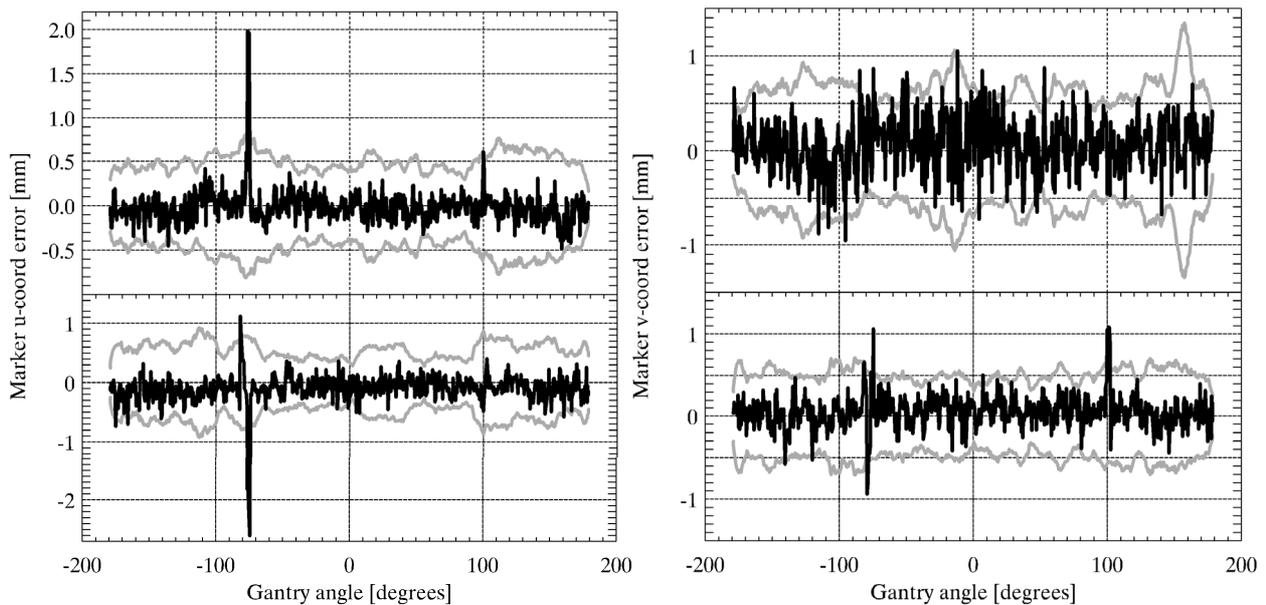


FIG. 10. Error between estimated marker position and the corresponding ground truth shown as dark lines for pancreas markers 6 and 2. Standard deviations limits $\pm 3\sigma$ for the ground truth are shown as gray lines.

Figure 11 shows sample images from the prostate image sequence, illustrating good and poor quality data. Figure 12 shows prostate marker tracking results as a function of gantry angle for markers 1 and 2. To better visualize the accuracy of the tracker Fig. 13 shows the difference (black line) between estimated and ground truth position for the u and v direction for markers 1 and 2. Again most of the error is within or close to $\pm 3\sigma$ of the ground truth. A few larger errors occur, mostly close to gantry angle 0° or $\pm 180^\circ$. These gantry angles correspond to lateral projections where the signal to noise ratio is greatly reduced making it difficult, even for the human observers, to accurately identify the marker positions (see example low quality projection image in Fig. 11 and note the increase in standard deviation of manual ground truth around these angles in Fig. 13).

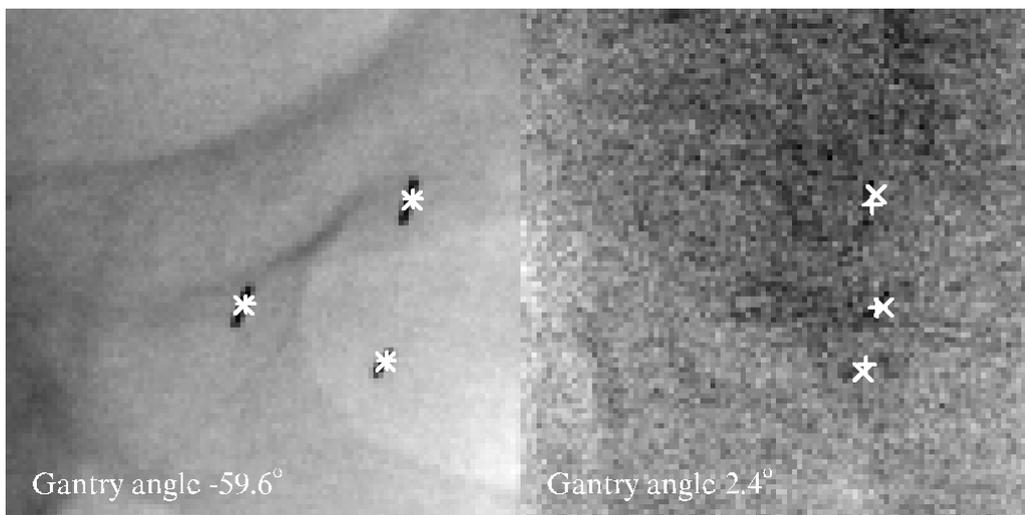


FIG. 11. Examples of the tracking results obtained for the prostate seed data. Cross symbols ('+') show position of ground truth and X's show marker position estimated by the tracker.

Automatic tracking of implanted fiducial markers in cone beam CT projection images

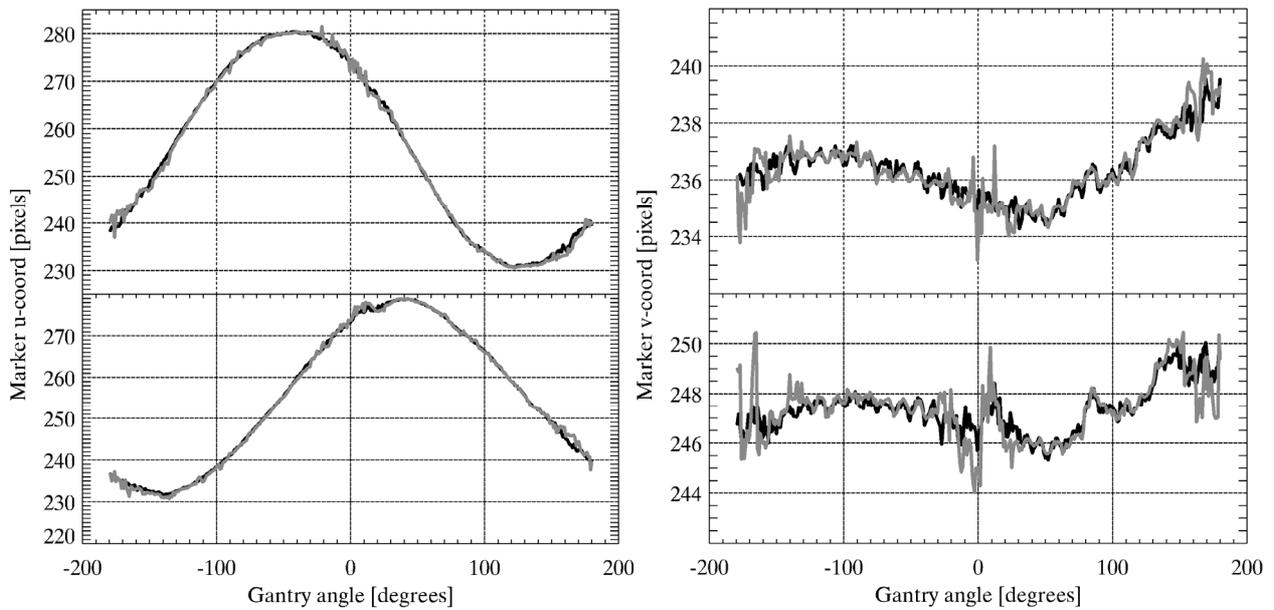


FIG. 12. Tracking results for prostate markers 1 and 2. Black line shows ground truth and the gray line estimated markers position.

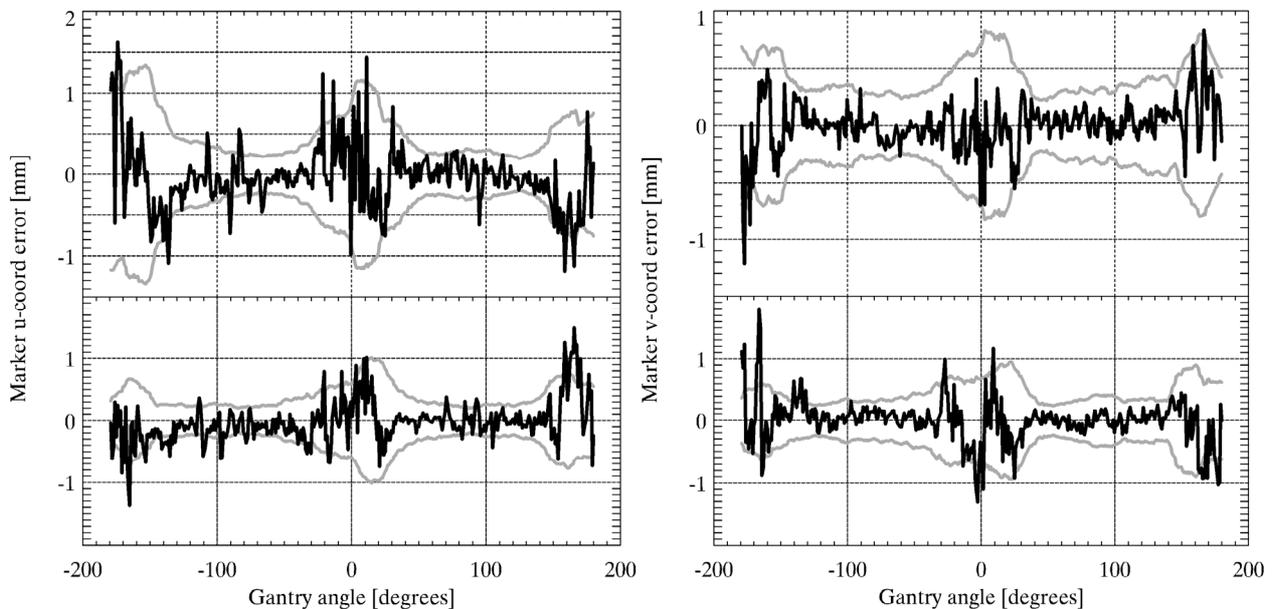


FIG. 13. Error between estimated marker position and the corresponding ground truth shown as dark lines for prostate markers 1 and 2. Standard deviations limits $\pm 3\sigma$ for the ground truth are shown as gray lines.

The overall performance of the tracker is summarized in Tables I and II. The ground truth statistics given in Table I do not reflect the actual performance of the human observers as the human observers had made a number of gross errors mainly due to misidentifying corresponding markers in the sequence. These gross errors were subsequently detected and corrected as this data was collected mainly for creating the ground truth. As is evident from the presented statistics, the tracking error is comparable with the ground truth error. This means that the method performs as well as a human observer, indeed it is actually performing better as the outliers have been eliminated from the manually selected data.

TABLE I. Pancreas marker tracking results statistics, calculated as a difference between estimated and ground truth marker position: dispersion in the u and v directions and maximum Euclidian distance. For comparison the table contains also corresponding statistics for the ground truth. The results after applying post-processing³¹ (see discussion) are also included for comparison. Results are quoted in mm when scaled to the isocentre plane.

Marker	Before post processing (mm)			After post processing (mm)			Ground truth (mm)		
	σ_u	σ_v	max	σ_u	σ_v	max	σ_u	σ_v	max
1	0.22	0.20	1.26	0.20	0.20	1.25	0.22	0.22	1.26
2	0.25	0.21	2.82	0.17	0.20	1.16	0.22	0.20	0.75
3	0.15	0.29	1.30	0.15	0.28	1.30	0.22	0.23	0.92
4	0.33	0.22	2.16	0.27	0.21	1.15	0.25	0.20	0.93
5	0.18	0.27	1.06	0.17	0.28	1.06	0.19	0.26	1.12
6	0.20	0.29	1.56	0.14	0.30	1.06	0.19	0.25	0.73

TABLE II. Prostate marker tracking results statistics, calculated as a difference between estimated and ground truth marker position: dispersion in the u and v directions and maximum Euclidian distance. Results are quoted in mm when scaled to the isocentre plane.

Marker	Tracking results (mm)		
	σ_u	σ_v	max
1	0.39	0.22	1.68
2	0.35	0.36	2.09
3	0.34	0.35	1.74

The robustness of the tracker was also characterized using the fraction of identified marker positions that were within 1 mm (2 pixels) of the estimated ground truth. For the pancreas image this was quantified separately for markers with and without nearby clutter (estimated as described in section II.D). Markers with no nearby clutter were detected successfully in 100.0% of cases (2060 marker locations). Markers with nearby clutter were detected successfully in 97.6% of cases (1737 of 1780 marker locations). The combined success rate in the pancreas image for cases with and without clutter was 98.9%. In the prostate image, markers with no nearby clutter were detected successfully in 98.3% of cases (677 of 689 marker locations). Markers with nearby clutter were detected successfully in 86.8% of cases (355 of 409 marker locations). The combined success rate for the prostate image was 94.0%.

To further investigate the performance of the proposed tracker, qualitative results without ground truth data, were obtained for 378 projection images of a motorized breast phantom and another sequence of 640 projection images obtained for a different pancreatic cancer patient. Figure 14 shows example images from the motorized phantom data. It is evident from Fig. 14 that the tracker is capable of dealing with significant contrast changes, as well as significant clutter due to wires (in projection #360), occluding surface (in projection #141), as well as honeycomb internal structure of the phantom (in projections #269 and #285). Tracking results from the second pancreas patient dataset showed similar performance to the first, including examples where the tracker copes with overlapping markers and severe clutter/noise in the data.

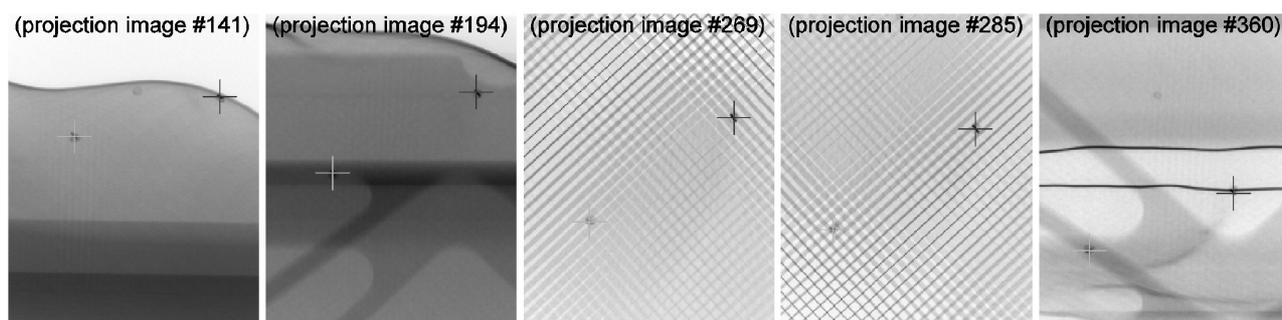


FIG. 14. Examples of the tracking results obtained for the phantom data.

IV. DISCUSSION

The experimental results demonstrate that the proposed method achieves comparable or better (when gross errors from the manual tracking are included in the analysis) tracking accuracy as was reached by human observers. The proposed algorithm is designed for, and can achieve, real time tracking at a frame acquisition rate of 5.5 Hz as used by the CBCT image acquisition system. Although not all potential application areas require real time performance (tracking is not always necessary at the rate of the data acquisition) it still significantly simplifies and reduces the cost of the procedure. It took human observers between three and five hours to perform marker selection in the whole image sequence used in the experiments, and this time does not include gross error detection, which was necessary as all three human observers made a number of such errors. The notable accuracy and robustness of the proposed tracker was obtained thanks to two main factors, namely: (i) the use of the mean shift algorithm (accuracy) in conjunction with random sampling methodology (robustness); (ii) selection of constraints which reflect the characteristics of the CBCT projection image sequences. Whereas the latter is application specific the former is more universal as it does not depend on the application area.

The success rate of any tracking algorithm is highly dependent on the quality of the input data and the definition used for tracking success. The evaluation data that we used is particularly challenging as it contained six identical markers close to the target volume (pancreas images) or very noisy projection data (prostate images). The large number of markers means there is a higher potential for overlapping or occlusion of markers compared to input data featuring fewer markers. Indeed, almost half of the tracked marker positions in our test data were judged to have clutter nearby. We calculated robustness of the algorithm separately for cases with and without nearby clutter. This may be useful in predicting the success of the algorithm for other data sets with different levels of clutter. The overall success rate of the algorithm for the pancreas test image was 98.9% (defined as tracking error of less than 1 mm compared to estimated ground truth). Based on a similar criterion, Poulsen et al.¹⁸ quoted a success rate of 99.9% for their CBCT marker tracking algorithm, although it is difficult to compare these results directly due to differences in the evaluation data (different anatomical site, different image resolution, different number and type of markers, different dose per projection). The success rate by this measure for our prostate marker test image was somewhat lower at 94.0%, however the dispersion of the error from ground truth was still small at less than 0.5mm. The reduction in performance was mainly caused by very low image contrast and high levels of noise in lateral projections of the prostate image. Although the accuracy of the tracker worsens slightly when tracking markers in low quality images, it should be noted that the tracker is able to resume normal accuracy when the quality of the images improves. What is remarkable is that the proposed tracker is able to achieve this autonomously in the on-line mode. The previously proposed tracking methods^{18,19} operate in off-line mode to provide similar robust performance.

For some applications, for example motion corrected CBCT volume reconstruction⁹, real time tracking is not necessary. In this case the results can be further improved by post processing of the marker tracks. A post processing algorithm has been developed³¹ where all the hypothetical marker positions, generated for all the markers and all the images, are fed to a function which assigns a score to all possible track configurations supported by the hypotheses generated by the tracker. The configuration with the highest score defines final estimate of the markers' position. This has the effect of correcting any temporary misassignments of marker identity close to instances of marker overlap, thereby further improving the robustness and accuracy of the tracker. Results of applying this post-processing³¹ to our evaluation data are shown in Table I. The overall "success" rate is increased to 99.5%.

Although the described algorithm has been tested on CBCT projection images, it can be adapted to other tracking applications e.g. in fluoroscopy. The required modifications will be mainly restricted to the design of specific constraints whereas the mean shift with random sampling would remain unchanged. Further modifications of the algorithm are possible by including more involved

propagation formulae. For the given application this was not necessary as the strong constraints introduced via the score function compensated for the simplicity of the predictor.

This paper does not include discussion on image preprocessing and tracking failure detection. The latter is essential for most real time tracking applications. This can be simply implemented by monitoring one or all components of the score function falling below some threshold and initializing a tracker recovery procedure; in the simplest case asking for operator's assistance (e.g. Tang *et al.*²¹). In general the selection of appropriate image preprocessing is strongly data dependant. For the results shown in this paper a simple intensity adjustment was used to improve image contrast.

V. CONCLUSIONS

The paper describes in detail an automatic fiducial marker tracking algorithm developed for tracking markers in a sequence of CBCT projection images. This enables marker motion apparent in the projection images to be separated into components due to gantry rotation and true 3D marker motion. The proposed method uses a combination of the mean shift algorithm, sequential random sampling and custom designed constraints. The initial tests performed show that the accuracy and robustness of the proposed method is superior to the results obtained by human observers. The methodology for designing constraints from the prior knowledge about the experiment was explained. Although the method was developed for CBCT projection images it can be simply adapted to other tracking applications e.g. fluoroscopy. To the best knowledge of the authors this is the first time an automatic robust tracking method has been described in detail for CBCT projection images and indeed the first time mean shift and sequential random sampling methods have been used in the context of fiducial markers tracking.

ACKNOWLEDGMENTS

This work was supported in part by the UK Engineering and Physical Sciences Research Council (EPSRC) grant No. EP/D077540/1.

REFERENCES

- ¹ D.A. Jaffray, J.H. Siewerdsen, J.W. Wong, and A.A. Martinez, "Flat-panel cone-beam computed tomography for image-guided radiation therapy," *Int. J. Radiat. Oncol. Biol. Phys.* **53**, 1337-1349 (2002).
- ² C.A. McBain, A.M. Henry, J. Sykes, A. Amer, T. Marchant, C.J. Moore, J. Davies, J. Stratford, C. McCarthy, B. Porritt, P. Williams, V.S. Khoo, and P. Price, "X-ray volumetric imaging in image-guided radiotherapy: The new standard in on-treatment imaging," *Int. J. Radiat. Oncol. Biol. Phys.* **64**, 625-634 (2006).
- ³ K. Nakagawa, A. Haga, K. Shiraishi, H. Yamashita, H. Igaki, A. Terahara, K. Ohtomo, S. Saegusa, T. Shiraki, T. Oritate, and K. Yoda, "First clinical cone-beam CT imaging during volumetric modulated arc therapy," *Radiother. Oncol.* **90**, 422-3 (2009).
- ⁴ T. Marchant, C. Rowbottom, and C. Moore, "Image Quality of On-board Cone-beam CT Acquired during VMAT Delivery," (abstract) *Int. J. Radiat. Oncol. Biol. Phys.* **75**, S97-S98 (2009).
- ⁵ T.E. Marchant, A.M. Amer, and C.J. Moore, "Measurement of inter and intra fraction organ motion in radiotherapy using cone beam CT projection images," *Phys. Med. Biol.* **53**, 1097-1098 (2008).
- ⁶ P.R. Poulsen, B. Cho, and P.J. Keall, "A method to estimate mean position, motion magnitude, motion correlation, and trajectory of a tumor from cone-beam CT projections for image-guided radiotherapy," *Int. J. Radiat. Oncol. Biol. Phys.* **72**, 1587-1596 (2008).
- ⁷ P.R. Poulsen, B. Cho, A. Sawant, and P.J. Keall, "Implementation of a New Method for Dynamic Multileaf Collimator Tracking of Prostate Motion in Arc Radiotherapy Using a Single kV Imager," *Int. J. Radiat. Oncol. Biol. Phys.* **76**, 914-923 (2010).
- ⁸ W. Liu, R.D. Wiersma, W. Mao, G. Luxton, and L. Xing, "Real-time 3D internal marker tracking during arc radiotherapy by the use of combined MV-kV imaging," *Phys. Med. Biol.* **53**, 7197-213 (2008).
- ⁹ T.E. Marchant, G.J. Price, B.J. Matuszewski, and C.J. Moore, "Reduction of motion artefacts in on-board cone beam CT by warping of projection images," *Brit. J. Radiol.* **84**, 251-264 (2011).
- ¹⁰ I. Ali, N. Alsbou, T. Herman, and S. Ahmad, "An algorithm to extract three-dimensional motion by marker tracking in the kV projections from an on-board imager: four-dimensional cone-beam CT and tumour tracking implications," *J. Appl. Clin. Med. Phys.* **12**, 223-238 (2011).
- ¹¹ J.J. Sonke, L. Zijp, P. Reameijer, and M. van Herk, "Respiratory correlated cone beam CT," *Med. Phys.* **32**, 1176-86 (2005).
- ¹² G. Hugo, J. Liang, and D. Yan, "Marker-free lung tumor trajectory estimation from a cone beam CT sonogram," *Phys. Med. Biol.* **55**, 2637-50 (2010).

- ¹³Y. Cui, J.G. Dy, G.C. Sharp, B. Alexander, and S.B. Jiang, "Robust fluoroscopic respiratory gating for lung cancer radiotherapy without implanted fiducial markers," *Phys. Med. Biol.* **52**, 741-55 (2007).
- ¹⁴R. Li, J.H. Lewis, L.I. Cerviño, and S.B. Jiang, "A feasibility study of markerless fluoroscopic gating for lung cancer radiotherapy using 4DCT templates," *Phys. Med. Biol.* **54**, N489-N500 (2009).
- ¹⁵N.M. Hamming, M.J. Daly, J.C. Irish, and J.H. Siewerdsen, "Automatic image-to-world registration based on x-ray projections in cone-beam CT-guided interventions," *Med. Phys.* **36**, 1800-12 (2009).
- ¹⁶S. Wiesner and Z. Yaniv, "Respiratory signal generation for retrospective gating of cone-beam CT images," *Proc. SPIE* **6918**, 691817 (2008).
- ¹⁷T. Li, L. Xing, P. Munro, C. McGuinness, M. Chao, Y. Yang, B. Loo, and A. Koong, "Four-dimensional cone-beam computed tomography using an on-board imager," *Med. Phys.* **33**, 3825-3833 (2006).
- ¹⁸P.R. Poulsen, W. Fledelius, P.J. Keall, E. Weiss, J. Lu, E. Brackbill, and G.D. Hugo, "A method for robust segmentation of arbitrarily shaped radiopaque structures in cone-beam CT projections," *Med. Phys.* **38**, 2151-2156 (2011).
- ¹⁹W. Fledelius, E. Worm, U.V. Elstrom, J.B. Petersen, C. Grau, M. Hoyer, and P.R. Poulsen, "Robust automatic segmentation of multiple implanted cylindrical gold fiducial markers in cone-beam CT projections," *Med. Phys.* **38**, 6351-6361 (2011).
- ²⁰J.R. Sykes, P. Whitehurst, and C.J. Moore, "Automatic detection, with confidence, of implanted radiographic seeds at megavoltage energies using an amorphous Silicon imager," *SPIE Medical Imaging 2003: Visualization, Image-Guided Procedures, and Display*, San Diego, CA, USA, February (2003).
- ²¹X. Tang, G.C. Sharp, and S.B. Jiang, "Fluoroscopic tracking of multiple implanted fiducial markers using multiple object tracking," *Phys. Med. Biol.* **52**, 4081-98 (2007).
- ²²Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Trans. Pattern Analysis and Machine Intelligence* **17**, 790-9 (1995).
- ²³D. Comaniciu and P. Meer, "Mean Shift: A robust approach towards feature space analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence* **24**, 603-19 (2002).
- ²⁴A. Doucet, N. Freitas, and N. Gordon, "Sequential Monte Carlo Methods in Practice," Springer, New York (2001).
- ²⁵D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Trans. Pattern Analysis and Machine Intelligence* **25**, 564-77 (2003).
- ²⁶S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for on-line non-linear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing* **50**, 174-188 (2002).
- ²⁷M. Bishop, "Pattern Recognition and Machine Learning," Springer, New York (2006).
- ²⁸B. Ristic, S. Arulampalam, and N. Gordon, "Beyond the Kalman Filter, Particle Filters for Tracking Applications," Artech House Publishers, London (2004).
- ²⁹M.B. Sharpe, D.J. Moseley, T.G. Purdie, M. Islam, J.H. Siewerdsen, and D.A. Jaffray, "The stability of mechanical calibration for a kV cone beam computed tomography system integrated with linear accelerator," *Med. Phys.* **33**, 136-144 (2006).
- ³⁰G.R. Bradski, "Computer vision face tracking for use in a perceptual user interface," *Intel Technology Journal* **2**, Q2 (1998).
- ³¹B. Matuszewski, T. Marchant, and A. Skalski, "Marker tracks post-processing for accurate fiducial marker position estimation in cone beam CT projection images," *International Conference on Computer Vision Theory and Applications (VISAPP 2010)*, Angers, France, 17-21 May (2010).