

Fast Marker Based C-Arm Pose Estimation

Bernhard Kainz¹, Markus Grabner¹, and Matthias Rüther¹

Institute for Computer Graphics and Vision, Graz University of Technology, Austria.
kainz|grabner|ruether@icg.tugraz.at *

Abstract. To estimate the pose of a C-Arm during interventions therapy we have developed a small sized X-Ray Target including a special set of beads with known locations in 3D space. Since the patient needs to remain in the X-Ray path for all feasible poses of the C-Arm during the intervention, we cannot construct a single marker which is entirely visible in all images. Therefore finding 2D-3D point correspondences is a non-trivial task. The marker pattern has to be chosen in a way such that its projection onto the image plane is unique in a minimal-sized window for all relevant poses of the C-Arm. We use a two dimensional adaption of a linear feedback shift register (LFSR) to generate a two-dimensional pattern with unique sub-patterns in a certain window range. Thereby uniqueness is not achieved by placing unique 2D sub patterns side by side but by the code property itself. The code is designed in a way that any sub window of a minimal size guarantees uniqueness and that even occlusions from medical instruments can be handled. Experiments showed that we were able to estimate the C-Arm's pose from a single image within one second with a precision below one millimeter and one degree.

1 Introduction

Minimally invasive focal therapy is a key requirement in modern medical procedures as for example in our motivating application, prostate cancer screening and subsequent treatment. These techniques depend on accurate monitoring of the operating field. A C-Arm is well suited for this task since it can easily be adjusted to provide the desired view. However, for subsequent measurements it is essential to know the exact location and orientation of the X-Ray device. A constant coordinate frame through biopsy and surgery enables surgeons or surgery robots to exactly localize and remove pathological tissue. This requires exact knowledge of the patient's position, which can be achieved by registration of X-Ray images to a virtual model of the patient (e.g., previously taken volumetric CT-scan). To facilitate such a registration with sufficient performance and precision for real-time clinical applications, the position of the X-Ray device has to be known. One possibility to do this pose estimation is to attach

* We would like to thank Prof. Bartsch from the Urology Clinic of the University Hospital Innsbruck for funding parts of this work and providing the X-Ray and CT data used for the experiments in this paper. We also thank the anonymous reviewers for their valuable comments.

tracking targets to all instruments and to the hip bone of the patient. Then subsequent pose estimation can be performed by common tracking systems. Solutions to this so-called *tracking problem* exist in various commercial setups. However, they are difficult to set up in the operating room due to difficult lighting conditions for optical systems or the presence of metallic components in case of magnetic tracking. Moreover they are quite expensive. Therefore we propose a method to compute the C-Arm pose directly from the image of a target object acquired by the X-Ray device, which is robust, inexpensive, and less sensitive to miscalibration than optical tracking systems.

Usual image based fiducial markers consist of several large black and white areas as for example the well known Augmented Reality Toolkit markers. Using such markers would occlude significant parts of the image, moreover it cannot be guaranteed that any of the markers is entirely visible. Large occluded areas constrain a desired subsequent or online 2D/3D or 2D/2D registration, not to mention the missing image information for the surgeon. Consequently a tracking target for C-Arm pose estimation should provide: *distinct features, maximum contrast, and little occlusion of the patient's anatomy.*

Related Work: Navab et al. [1] proposed code words located side by side, formed by visible beads attached to a ring of acrylic glass to solve the C-Arm tracking problem for cerebral angiography. In contrast to this approach, our motivating clinical application does not allow a field of view containing a lot of beads up to a complete ring target. Therefore one of the problems we treated is that we see only a very small part of the target and consequently no complete code words. Furthermore, we need a code which allows a certain robustness against occlusions caused by medical instruments. Finally we take pelvis images in lithotomy position for which a ring is not feasible since it would be too large and therefore parts of it too close to the X-ray source.

Jain et al. [2] proposed a system to estimate the relative position of the C-Arm between two images from stationary natural features. However, we desire X-Ray registration to a virtual patient model which requires to know the camera position for *each* image relative to a *fixed origin*. Furthermore, the extraction of natural features is a time-consuming procedure. This would not be feasible for online processing of the C-Arm poses.

2 Method

We use a plate mounted below the operating table which is prepared with a regular grid of holes on both sides. The plate is made of acrylic glass to provide sufficient X-Ray transparency. The holes form a grid with 5 mm \times 25 mm on one side of the plate. On the other side the same grid was used but shifted by 12.5 mm. Overall we have 78 \times 18 available positions in our target. The challenge is to equip these holes with beads in a suitable pattern such that the reoccurring two-dimensional maximum of sub-pattern's cross correlation is minimal due to their code distance [3]. In other words, to provide unique and maximally robust

2D patterns of a certain size $u \times w$, we seek to maximize

$$\sum_{x=0}^{u-1} \sum_{y=0}^{w-1} v_{x,y}^{(i)} \oplus v_{x,y}^{(j)}, \quad i \neq j, \quad u \geq 3, \quad w \geq 1, \quad (1)$$

where $v_{x,y}^{(i)}$ denotes the value at position (x, y) of the i -th sub-pattern out of the list of N possible sub-patterns of size $u \times w$ extracted from the complete 2D pattern (and likewise $v_{x,y}^{(j)}$). The operator \oplus defines the binary XOR, evaluating to 0 or 1. We require $u \geq 3$ since for robust detection of the pattern the distance between two beads has to be either 5 mm or 10 mm (i.e., there must not be more than one “empty” position in succession). Therefore a binary “1” is represented by a single bead, and a binary “0” by an empty space followed by a bead.

One-dimensional linear feedback shift registers (LFSR) [3] provide a certain code unambiguity for a given length. In our case, we are interested in the minimum size of a 2D window such that the pattern inside the window does not appear anywhere else in the entire code. This form exactly implies the already formulated coding condition that the maximum of the in-between cross correlations is minimal. A Gold Code LFSR [4] provides this behavior for one dimension by connecting selected bits from two feedback shift registers to the last produced code part with exclusive or (XOR) gates to the input of the next sequence. The period of the code sequence depends on the register length N and the feedback pattern. Its maximum is $2^N - 1$. To produce a Gold Code from an LFSR, the XOR connected bits have to be connected according to two principal polynomials of the same order N .

Experiments showed that a first prototype with five lines of one-dimensional six bit LFSR code works well for C-Arm pose estimation with lateral rotation. However, a six bit code is restricted to a maximum of 63 unique code positions, and there are only six different code sequences of this length which can be produced by a six bit LFSR, hence severely limiting the search space in the perpendicular direction (i.e., craniocaudal rotation).

We therefore investigate methods to develop a target supporting also craniocaudal rotations of the C-Arm. In the craniocaudal case, the image window size is in some positions too small to provide unambiguous patterns. We developed a variant of the two-dimensional LFSR as proposed by Chen et al. [5] with an optimal binary sequence as defined by Gold [4]. Since we are not interested in minimizing the number of register stages for the 2D-LFSR as proposed by Chen et al., we can use the idea of that algorithm directly by XOR cross connection of two 1D-LFSR Gold-Codes with different principal polynomials of the same order and $N \times M$ register stages. Considering the length of the sides of our target, the next best period size is $2^6 - 1 = 63$ for the longer side.

The C-Arm has to be calibrated once with the acrylic glass target to determine its intrinsic parameters. Once these parameters are known, the C-Arm’s orientation (rotation and translation) in 3D space can be determined. Camera calibration is the usual way to determine the focal length, pixel dimensions, and optical center intrinsic parameters of a camera by means of computer vision. This requires to take one picture of a large known 3D target or several

pictures of a planar calibration target (e.g., a checkerboard). Since we already have a target with features in two parallel planes, we have chosen the second approach. However, this step is crucial for a subsequent one-image pose estimation, and it models the attributes of the system for the whole acquisition matrix. Camera calibration is a thoroughly researched topic in computer vision and photogrammetry, and several fully automatic methods have been proposed [6]. The methods mainly differ in the underlying algebraic model of the projection process, the method to determine the model parameters, and the calibration normal used. Using planar normals (i.e., all reference points are coplanar [7]) is state of the art in the computer vision community. Several images of the normal are acquired by the unknown camera, each time from a different viewpoint. From this information it is possible to determine the unknown parameters of the underlying projection model. There is no definite rule as to how many calibration images are needed, since this strongly depends on the used camera. In most cases, 20 – 30 images are sufficient for good calibration results [8]. The operator is not directly exposed to radiation during these calibration steps since she or he only has to move the plate by a few degrees between the activations. Concerning the underlying projection model, the classical camera model of central perspective projection, nonlinear radial lens distortion, and tangential distortion has been employed in our experiments. This amounts to 7 intrinsic camera parameters: one for focal length, two for principal point, two radial distortion coefficients and two tangential distortion coefficients. While many alternative models have been proposed for the thorough description of C-Arms and X-Ray cameras, e.g. [9], this standard is sufficient since model imprecisions are far below the accuracy of the “virtual” reference target which was created from a CT-scan of the real target.

Experiments: A total of 152 calibration images were acquired with different rotation angles and inclination angles. We use a *Siemens Siremobil Compact L* for our experiments. The calibration method proposed in [7] has been adapted to the setup at hand. We get reference coordinates of all beads with a certain measurement error from a CT-scan of the target. Scanning this target with the same scanner as used for the patients cancels out possible inaccuracies of the CT-scanner for later image based registration tasks. The actual C-Arm pose estimation can subsequently be performed by placing the unique pattern target below the patient. A resulting C-Arm X-Ray image is shown in Figure 1(a). Efficient post-processing is accomplished by the following steps. First the images are undistorted with the previously determined fourth order polynomial approximating the camera distortion (Figure 1(b)). To segment the beads, we make use of the fact that they appear in the image at almost constant size, so they can easily be detected and separated from the background by hardware accelerated variational filtering methods as proposed by Pock et al. [10]. Then a *random sample consensus* (RANSAC) algorithm is applied to find the vanishing point of the pattern lines, and subsequently the lines themselves. Correction of the perspective line distortion leads to an orthographic view of the displayed beads (Figure 1(c)). Correspondences between them and the complete target are es-

tablished through a comparison of the observed sub-pattern with the complete known pattern. Due to the uniqueness of sub-patterns, each bead can exactly be located on the target. To avoid distraction by the visible beads, an optional hardware accelerated edge preserving structural inpainting algorithm [11] can be applied. Since the statistical texture properties of the neighborhood of the occluded regions (in particular noise) are not taken into account, the calculated areas are still identifiable in case of doubt. Nevertheless, an average image area of less than two percent is occluded by the beads.

The X-Ray images will not necessarily show a complete sub-pattern of the whole target. Some beads may be occluded by medical instruments, e.g., an ultrasound sensor. Therefore we use a binary XOR comparison with all sub-patterns from the original target and choose the position which produces the smallest error. The resulting point correspondences are used to estimate position and orientation of the calibrated X-Ray camera relative to the target. To cope with the presence of outliers and occlusion, we employ the *direct linear transformation* [12] (DLT) method within a RANSAC framework to get an initial pose estimation. This pose estimation is further refined on the inlier points by nonlinear optimization with an iterative steepest descent optimization method [12].

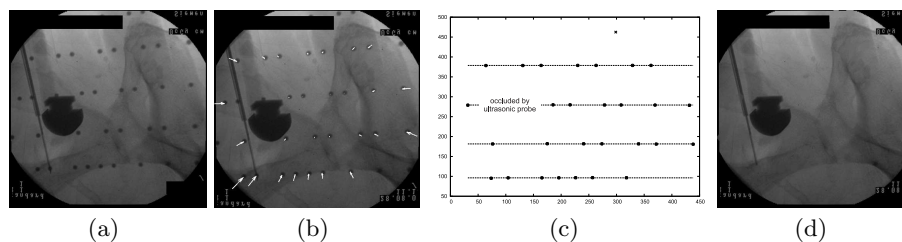


Fig. 1. A processing example for an X-Ray images taken by a *Siemens Siremobil Compact L* including our pose estimation target. The images were taken during a prostate biopsy. Therefore the biopsy needle and the head of the ultrasound probe are also visible. Figure (a) is an example with a quite narrow field of view, showing only a small part of the target, but still sufficient to reconstruct the C-Arm’s pose. (b) illustrates the distortion correction process with a 4th order polynomial. (c) shows the detected beads after correction and denoising based on variational methods. We are using a ROF-Chambolle model with an L1 norm in this case [10]. Line detection is done with a RANSAC based algorithm and subsequent perspective equalization. (d) shows an optional hardware accelerated inpainting [11] step to remove the beads from the resulting image.

3 Results

X-Ray images were taken in the range of $\pm 20^\circ$ craniocaudal and $\pm 30^\circ$ lateral rotation using a *Siemens Siremobil Compact L* during biopsy interventions. Hence the C-Arm pose was not optimized in any way, and our method was tested with an uncontrolled surgical scenario. The mean reprojection error during pose estimation settled down between **0.67 pixels** and **0.89 pixels** for different datasets.

This value is calculated through cross-validation, randomly taking 75% of the detected inliers to calculate the intrinsic camera parameters and the remaining 25% to calculate the error. For a subsequent 2D/3D registration, ideally two images acquired at a relative angle of 90 degrees are required. Consequently, images at angles of 180 degrees are not considered. Due to mechanical restrictions during (prostate) investigations, we only examined the limited operational range of $\pm 45^\circ$. Visible beads are masked out for subsequent registration or processed with an optional inpainting algorithm [11] for a proper visual result. To evaluate the precision of the C-Arm pose estimation, we add Gaussian noise to the detected positions in the undistorted X-Ray images with a mean of a quarter of the beads' radius considering the calculated reprojection error. A subsequent principal components analysis results in a variance maximum of **0.33 mm** for the X-Ray source translation and **0.15°** for the X-Ray source rotation on the principal axis. We are aware of the adverse effects of gravitation and the earth magnetic field on calibration consistency. However, since we operate the C-arm at a small number of repeatable poses, it is easy to compute separate calibrations for each of them. Interior effects like pincushion/barrel distortion are sufficiently compensated by the radial distortion model, as the reprojection error suggests. Furthermore, Jain [2] noted that a faulty calibration has only little influence at least on relative poses.

To evaluate our code in terms of unambiguity and robustness against occlusions, we compare each sub-pattern of the same size with all other sub-patterns. An array of XOR gates indicates the difference of two sub patterns and can therefore be used as an error measurement. To evaluate the robustness against occlusion, we artificially increase the number of bit-inversions until the first detection failure occurs. A bit-inversion corresponds to the occlusion of a bead in a sub-pattern or a false positive detection due to noise. Our experiments show that **19 to 24 positions** are sufficient to guarantee uniqueness assuming that the complete pattern is optimized for the size of the target and that a binary "0" takes two positions. However, two neighboring positions on adjacent lines can both be empty. Therefore, the required unique sub-pattern size depends more on the number of visible lines than on the number of visible positions within a line. 19 positions would allow for only two occlusions according to Figure 2. Nevertheless, in the clinical setup we detect 50 positions and more, even for disadvantageous views (e.g., Figure 1(a)). This allows for 13 and more occluded positions (an ultrasound probe, for example, usually occludes only 3 to 7 positions). Consequently, unambiguity is getting better the more lines are visible. Figure 2 illustrates the growth of the required window size for an increasing number of occlusions. Above a sub-pattern size of 60 positions, half of the positions can be occluded before a detection failure occurs.

Runtime: The runtime for a complete pose estimation is the sum of the runtimes for image undistortion, line detection, segmentation of beads, localization of the sub-pattern on the target, and camera pose estimation from the known point correspondences. C-Arm camera calibration and analysis of the target's CT-scan has to be done only once and can therefore be performed offline.

The RANSAC based algorithms (line detection and camera pose estimation) are the critical parts and have to be considered for runtime measurements. The variational methods algorithm for sphere detection is hardware accelerated (i.e., executed on a graphics processing unit) and hence not comparable to CPU executed tasks. Its runtime is in the range of 10 ms. On a Linux PC (Intel Core2 Duo 2.13 GHz CPU, 2 GB RAM, NVidia Geforce 8800 GTX) the mean runtime for line detection and calculation of the point correspondences is **0.94 seconds**, and **0.08 seconds** for subsequent pose estimation. Consequently, we are able to reconstruct the pose of a calibrated C-Arm within one second from one image displaying a quite small part of our target. An optional hardware accelerated inpainting step as shown in Figure 1(d) takes 0.15 seconds for satisfactory results with 2000 iterations.

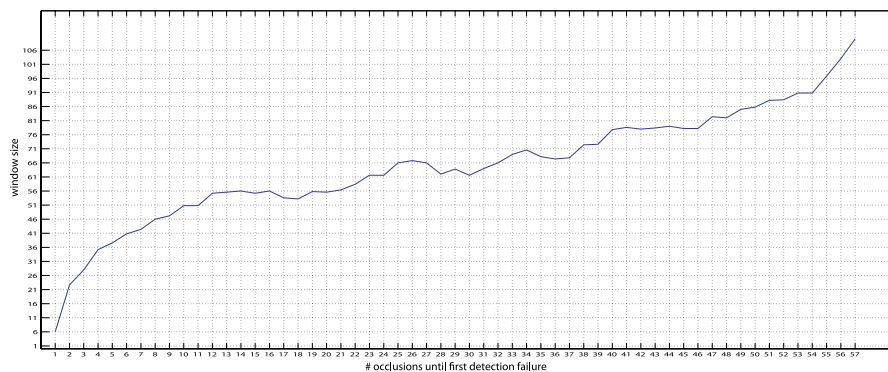


Fig. 2. The curve shows the minimal window size containing a unique sub-pattern as a function of the number of occluded positions. Above 60 visible positions, half of the sub-pattern’s positions can be occluded until a detection failure occurs.

4 Discussion

We showed that we are able to estimate the pose of a C-Arm during biopsies or surgeries within one second with a precision of 0.15° and 0.33 mm in the presence of Gaussian noise. Our approach is based on a unique pattern X-Ray target which can be placed below the operating table during intervention. Hence our technique does not interfere with the clinical work volume. The challenge for building such a target is to find a code with properties of a Gold-Code in two dimensions. Codes which fulfill the properties of a Gold-Code, namely a minimization of the maximum cross correlation between sub-patterns, are hard to find and can, as far as we know, only be validated experimentally. A compromise between reliable line detection and binary coding had to be found for the placement of the beads on the target. We decided to represent a binary “1” as a position equipped with a bead and a binary “0” as empty space followed by a bead. This unbalanced code size impairs the Gold-Code conditions for the two dimensional case, hence we had to optimize the generating polynomials by hand. However,

our pattern performed better in terms of unambiguity compared to a random pattern (including our distance constraint) because we optimized the maximum code lengths to the size of our X-Ray target in contrast to a common random number generator. Even though a common random number generator is based on the similar principles, the code period length is much longer for them. This makes the size of an unambiguous sub-pattern quite large since the maximum code distance is only given for a whole period.

An overall systematic error can be obtained by testing a complete setup with different C-Arm types. Furthermore we are looking forward to generate a reference transformation matrix with different external tracking systems. Another improvement of our target would be to optimize the code for only one half of the target. From the location of the vanishing point in the image plane we can determine if the C-Arm has been turned to the right or to the left. A smaller grid for the pattern would indeed improve the matching accuracy for the displayed sub-pattern, but more beads also denote more occlusion in the image or more artificial information in case of inpainting. This could be a problem for further image registration or for the surgeon.

References

1. Navab, N., Bani-Hashemi, A., Mitschke, M., Fox, A., Holdsworth, D., Fahrig, R., Graumann, R.: Dynamic geometrical calibration for 3-D cerebral angiography. *SPIE Medical Imaging* **2708 / 361** (April 1996) 361–370
2. Jain, A., Fichtinger, G.: C-Arm Tracking and Reconstruction Without an External Tracker. In: *Proc. MICCAI '06, Part I. Volume 4190 of LNCS.* (2006) 494–502
3. Golomb, S.W.: *Shift Register Sequences.* Aegean Park Press, Laguna Hills, CA, USA (1981)
4. Gold, R.: Optimal binary sequences for spread spectrum multiplexing. *IEEE Trans. Inform. Theory* **13**(4) (1967) 619–621
5. Chen, C.I.H.: Synthesis of configurable linear feedback shifter registers for detecting random-pattern-resistant faults. In: *Proc. ISSS '01, ACM* (2001) 203–208
6. Clarke, T., Fryer, J.: The development of camera calibration methods and models. *Photogrammetric Record* **16**(91) (April 1998) 51–66
7. Zhang, Z.: Flexible camera calibration by viewing a plane from unknown orientations. In: *Proc. ICCV '99.* (1999)
8. Sun, W., Cooperstock, R.: An empirical evaluation of factors influencing camera calibration accuracy using three publicly available techniques. *Machine Vision and Applications* **17**(1) (2006) 51 – 67
9. Livyatan, H., Yaniv, Z., Joskowicz, L.: Gradient-Based 2D/3D Rigid Registration of Fluoroscopic X-Ray to CT. *IEEE Trans. Medical Imaging* **22**(11) (2003) 1395–1406
10. Pock, T., Pock, M., Bischof, H.: Algorithmic differentiation: Application to variational problems in computer vision. *IEEE Trans. Pattern Analysis and Machine Intelligence* **29**(7) (July 2007) 1180–1193
11. Bertalmio, M., Sapiro, G., Caselles, V., Ballester, C.: Image inpainting. In: *Proc. SIGGRAPH '00, ACM* (2000) 417–424
12. Hartley, R., Zisserman, A.: *Multiple View Geometry in Computer Vision.* Second edn. Cambridge University Press (2004)