

Detecting Patient Motion in Projection Space for Cone-beam Computed Tomography

Wolfgang Wein and Alexander Ladikos

White Lion Technologies AG, Munich, Germany

Abstract. Cone-beam X-Ray systems strictly depend on the imaged object being stationary over the entire acquisition process. Even slight patient motion can affect the quality of the final 3D reconstruction. It would be desirable to be able to discover and model patient motion right from the actual projection images, in order to take it into account during reconstruction. However, while the source-detector arrangement is rotating around the patient, it is difficult to separate this motion from the additional patient motion. We present a novel similarity metric for successive X-Ray projections, which is able to distinguish between the expected rotational and additional sources of motion. We quantitatively evaluate it on simulated dental cone-beam X-Rays and qualitatively demonstrate its performance on real patient data.

1 Introduction

As opposed to rapidly spinning Computed Tomography (CT) devices, 3D cone-beam X-Ray systems take much longer (10-30 seconds) to acquire enough projection images for a 3D reconstruction. Both unexpected motion of the device itself (i.e. deviation from its calibrated circular path) as well as patient motion have negative effects on the reconstruction, if not modeled correctly. Depending on the organ being imaged, the latter can be categorized into respiratory movement, cardiac motion, as well as direct movement of the patient.

If additional information, such as an ECG sensor, is available which allows to separate motion phases, a binned reconstruction is possible, whereafter further motion refinement is possible [5]. In the absence of such information, both image and motion can be simultaneously reconstructed using an iterative framework [2]. While such approaches are computationally demanding, they potentially are able to recover small internal anatomic movement. However, convergence is not guaranteed if large-scale, non-periodic patient motion is present, as an initial reconstruction has to use either zero motion or a prior motion model.

Estimating motion in projection image space is difficult because the source-detector motion itself causes change. On top of that, the amount of change varies with the viewing angle and anatomy traversed by the X-Rays. In [6], an optical flow estimation between projection images and known reference images is used for a rough estimation. In [1] various standard similarity measures are evaluated with respect to their capability of distinguishing motion directly from successive X-Ray projections. Regardless of the specific measure used, such an approach can

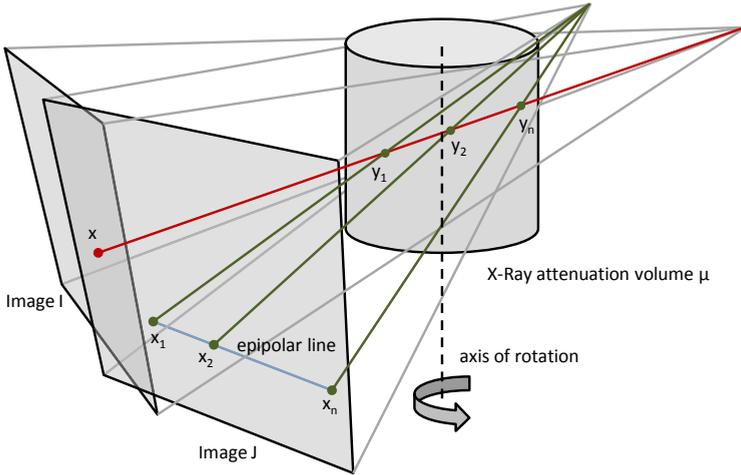


Fig. 1. Geometry of two successive X-Ray images during a cone-beam acquisition. For better illustration, the angle increment, and therefore the length of the epipolar line segment, has been increased.

only work if the amount of image change due to patient motion is significantly larger than the change induced by the device angle increment.

In the following, we present our approach for detecting motion in projection space. A new similarity measure is developed, which uses the projection geometry of two successive X-Ray images for estimating to which region in the second image a certain location in the first image can contribute. Inconsistent structural change can therefore be singled out and used to assess patient motion.

2 Methods

2.1 Geometric Modeling

Let us consider the geometric relationship of two successive cone-beam X-Ray projection images I and J as illustrated in figure 1. Let the projection matrices of images I and J be $\mathbf{P}_I = \mathbf{K}[\mathbf{I}|\mathbf{0}]$ and $\mathbf{P}_J = \mathbf{K}[\mathbf{R}|\mathbf{t}]$ with \mathbf{K} the internal camera parameters and $[\mathbf{R}|\mathbf{t}]$ the pose increment of image J with respect to image I . If the image intensities represent linear X-Ray attenuation (as opposed to detector intensity), the value of a single pixel $I(\mathbf{x})$ at location \mathbf{x} can be described as a line integral

$$I(\mathbf{x}) = \int_{\mathbf{y}_1}^{\mathbf{y}_n} \mu(\mathbf{r}) d\mathbf{r} \quad (1)$$

where $\mu : \mathcal{R}^3 \rightarrow \mathcal{R}$ is the volume of X-Ray attenuation, and $\mathbf{y}_1, \mathbf{y}_n$ the start and end position along the ray between source and detector position \mathbf{x} . As illustrated in figure 1 the attenuation values $\mu(\mathbf{r})$ along this ray also contribute

to pixel intensities in the neighboring projection image J . The affected pixels are located on a segment of the epipolar line of \mathbf{x} in J . We can discretize and rewrite equation 1 as

$$I(\mathbf{x}) = \sum_{k=1}^n \mu(\mathbf{y}_k) + \mu_\epsilon \quad (2)$$

Here, \mathbf{y}_k are sampling points along the projection ray going through \mathbf{x} and μ_ϵ is the residual X-Ray attenuation not found within these sampling points. The attenuation values $\mu(\mathbf{y}_k)$ contribute each with a particular fraction to the pixels $J(\mathbf{x}_k)$:

$$\mu(\mathbf{y}_k) = w_k J(\mathbf{x}_k) \quad (3)$$

It is therefore valid to describe the relationship of a pixel $I(\mathbf{x})$ with the neighboring image as follows:

$$I(\mathbf{x}) = \sum_{k=1}^n w_k J(\mathbf{x}_k) + w_{n+1} \quad (4)$$

where \mathbf{x}_k are a number of chosen discrete locations along the epipolar line in J . Essentially, the value of pixel \mathbf{x} in I is composed of a linear combination of some of the X-Ray attenuation values its ray shares with the ones in the neighboring image's locations \mathbf{x}_k , plus a constant value for attenuation values outside the epipolar line segment.

If we restrict the locations of \mathbf{x}_k such that the ray segment $|\mathbf{y}_1 - \mathbf{y}_n|$ corresponds to the cylindrical core reconstruction area, the epipolar line segment in J should span from a few to maximum a dozen pixels (depending on image resolution, angle increment and reconstruction size). More precisely the start and end point of the epipolar line segment are given by

$$\begin{aligned} \mathbf{x}_1 &= \mathbf{K}\mathbf{R}\mathbf{K}^{-1}\mathbf{x} + \frac{\mathbf{K}\mathbf{t}}{Z_{start}} \\ \mathbf{x}_n &= \mathbf{K}\mathbf{R}\mathbf{K}^{-1}\mathbf{x} + \frac{\mathbf{K}\mathbf{t}}{Z_{end}} \end{aligned} \quad (5)$$

where Z_{start} and Z_{end} are the start and end depths along the projection ray through \mathbf{x} (for a derivation see [3] chapter 9). We linearly interpolate the samples in between as $\mathbf{x}_k = \mathbf{x}_1 + \frac{k-1}{n-1}(\mathbf{x}_n - \mathbf{x}_1)$.

The weights w_k basically describe at which depths in reconstruction space corresponding structures occur, and, consequently, in which horizontal direction and how fast they move in different "layers" in projection space. If they were known, one could assess how well the images satisfy equation 4.

2.2 Projection Similarity

Image similarity metrics which allow local variation of unknown parameters in the assumed intensity relationship, exhibit superior performance in certain X-Ray image registration problems. One particular example is local cross-correlation, where brightness and contrast may vary, since they are implicitly computed,

within the neighborhood of every image pixel [4]. It is also possible to implicitly recover unknown parameters of a linear combination of two image components within a similarity formulation [7].

We extend such an approach to create a similarity measure which assesses how well equation 4 locally holds. We use the following variables to specify the situation for a neighborhood Ω of size m surrounding a pixel \mathbf{x}

$$\mathbf{i}_x = \begin{pmatrix} i_{11} \\ \vdots \\ i_{m1} \end{pmatrix}; \mathbf{J}_x = \begin{pmatrix} j_{11} & \dots & j_{1n} & 1 \\ \vdots & & \vdots & \vdots \\ j_{m1} & \dots & j_{mn} & 1 \end{pmatrix}; \mathbf{w}_x = \begin{pmatrix} w_1 \\ \vdots \\ w_{n+1} \end{pmatrix} \tag{6}$$

where \mathbf{i}_x is a linearization of the pixel intensities $\{I(\mathbf{x})|\mathbf{x} \in \Omega\}$ and row l of \mathbf{J}_x contains n samples taken from the epipolar line corresponding to location i_l and a constant entry. We then interpret equation 4 as a linear regression problem

$$\mathbf{i}_x = \mathbf{J}_x \mathbf{w}_x \tag{7}$$

which we solve using ordinary least squares (OLS):

$$\mathbf{w}_x = (\mathbf{J}_x^T \mathbf{J}_x)^{-1} \mathbf{J}_x^T \mathbf{i}_x \tag{8}$$

The local similarity around pixel \mathbf{x} then writes as

$$S_x(I, J) = 1 - \frac{|\mathbf{i}_x - \mathbf{J}_x \mathbf{w}_x|^2}{Var(\mathbf{i}_x)} \tag{9}$$

The global image similarity is composed weighted by the local variance, in order to suppress regions without structural appearance:

$$S(I, J) = \frac{\sum_{\mathbf{x}} (Var(\mathbf{i}_x) - |\mathbf{i}_x - \mathbf{J}_x \mathbf{w}_x|^2)}{\sum_{\mathbf{x}} Var(\mathbf{i}_x)} = 1 - \frac{\sum_{\mathbf{x}} |\mathbf{i}_x - \mathbf{J}_x \mathbf{w}_x|^2}{\sum_{\mathbf{x}} Var(\mathbf{i}_x)} \tag{10}$$

Rather than computing vector \mathbf{i}_x and matrix \mathbf{J}_x for every pixel (and hence re-evaluating all neighborhood pixels repeatedly), we directly establish all required sums to solve equation 8 using an efficient recursive filtering approach.

2.3 Motion Detection

Depending on the parameters m and n which discretize the local X-Ray depth properties, the similarity measure value will slightly vary with changing image content even in absence of patient motion. On the other hand, it is dependant on the projection matrices \mathbf{P}_I and \mathbf{P}_J . If we incorporate a rigid transformation matrix $\mathbf{T}(\mathbf{p}) = [\Delta \mathbf{R} | \Delta \mathbf{t}]$, parameterized over \mathbf{p} , the start and end point of the epipolar line segment of point $I(x)$ become

$$\begin{aligned} \mathbf{x}_s &= \mathbf{K}(\mathbf{R}\Delta\mathbf{R})\mathbf{K}^{-1}\mathbf{x} + \frac{\mathbf{K}(\mathbf{R}\Delta\mathbf{t} + \mathbf{t})}{Z_{start}} \\ \mathbf{x}_e &= \mathbf{K}(\mathbf{R}\Delta\mathbf{R})\mathbf{K}^{-1}\mathbf{x} + \frac{\mathbf{K}(\mathbf{R}\Delta\mathbf{t} + \mathbf{t})}{Z_{end}} \end{aligned} \tag{11}$$

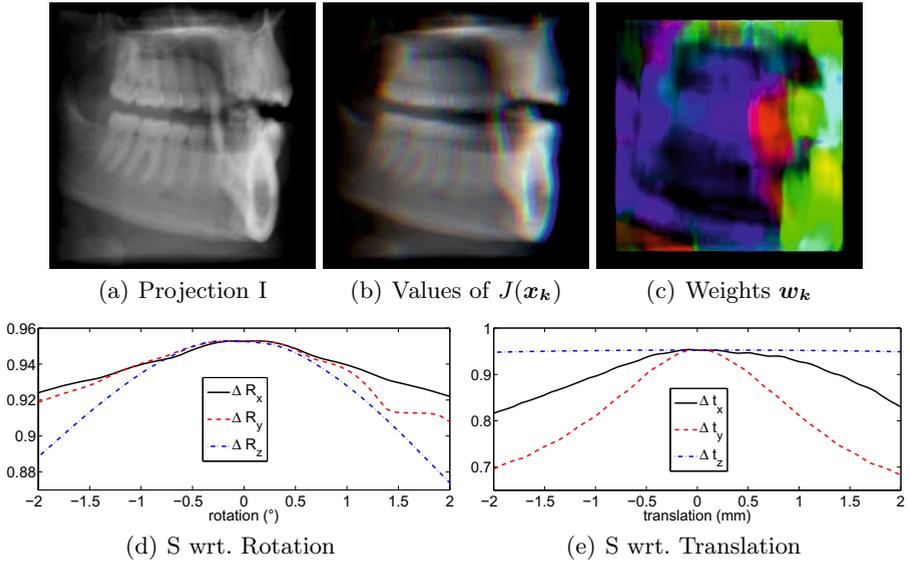


Fig. 2. Illustration of similarity measure properties from a pair of DRRs (please refer to the colored electronic version of this manuscript)

Even though it is just defined on the 2D projection images, the measure S will change with respect to all 3D transformation parameters, because $\mathbf{T}(\mathbf{p})$ directly affects the pixel locations \mathbf{x}_k in equation 4. If there is no patient motion the similarity measure should have a maximum at $\mathbf{T}(\mathbf{0}) = \mathbf{I}$, i.e. $\frac{\partial S}{\partial \mathbf{p}}|_{\mathbf{p}=\mathbf{0}}$ should be zero. Due to the local least-squares solutions inherent to the measure, analytic gradients are difficult to derive. We therefore use a numeric difference operator in order to assess if $S(I, J, \mathbf{p})$ is at a maximum. More precisely, we sum over all forward and backward differences which are positive:

$$S' = \sum_{i=1}^{12} \frac{1}{2} (\text{sgn}(\Delta S_i) + 1) \Delta S_i \quad \text{with } \Delta S_i = S(\Delta_i) - S(\mathbf{0}) \quad (12)$$

3 Evaluation

3.1 Analysis of the Similarity Value

We use a high-fidelity dental cone beam reconstruction with 512^3 voxels and compute digitally reconstructed radiographs (DRRs) from it. Their quality is insignificantly lower than the original projection images, however they enable us to simulate arbitrary motion of the patient's head in known conditions. Figure 2 visualizes the similarity measure. A parameter selection of $n = 3$ and $m = 21^2$ was used. The three values $J(\mathbf{x}_i)$, as well as the weights w_i , are encoded in the red, green and blue color channels of figure 2(b). The distribution

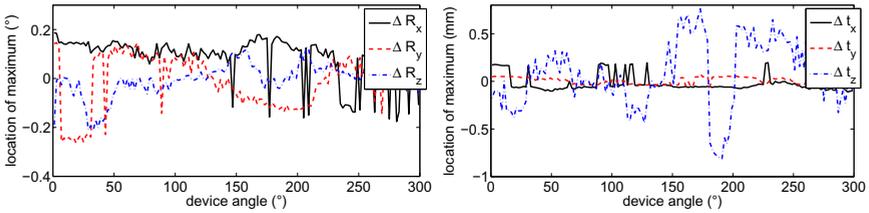


Fig. 3. Location of local maxima for the parameters of $\Delta\mathbf{R}$ and $\Delta\mathbf{t}$

along the epipolar line can nicely be seen by the horizontal split into color components, in particular at vertical structures. The computed weights w_i in figure 2(c) essentially represent a rough depth reconstruction from the two projections. Structures in the front of the volume which are moving to the right in the projections are blue. Structures in the center of the volume or tangential to the rotation are green; structures in the back (moving to the left) are red. The similarity measure S in equation 10 was evaluated with respect to all parameters in $\Delta\mathbf{R}$ and $\Delta\mathbf{t}$, figure 2(d) and (e) depict the corresponding curves for the image pair shown. This evaluation was then repeated for all projection pairs of a simulated cone-beam acquisition without any patient motion (300° rotation, 2° angle increment). Figure 3 shows the location of the similarity measure maximum for all frames (it should be zero). The mean deviation and variance (σ) are:

	$\Delta\mathbf{R}_x$	$\Delta\mathbf{R}_y$	$\Delta\mathbf{R}_z$	$\Delta\mathbf{t}_x$	$\Delta\mathbf{t}_y$	$\Delta\mathbf{t}_z$
mean	0.0892	-0.0024	-0.0040	-0.0214	0.0060	0.0754
σ	0.0870	0.1173	0.0655	0.0879	0.0351	0.3542

All parameters vary within a very small scale. The variation of the translation $\Delta\mathbf{t}_z$ towards the X-Ray source is highest, which is expected since it causes very little change in the projections (compare fig. 2(e)).

3.2 Recovering Motion

A cone-beam acquisition was simulated which includes a gentle nod of the patient's head with an amplitude of 10° and 7° in the rotational parameters $\Delta\mathbf{R}_y$ and $\Delta\mathbf{R}_z$. Here, the first parameter corresponds to a relative change of the cone-beam rotation speed, therefore the combined rotation slows down first. In this case, a standard similarity measure will yield higher similarity, as can be seen in figure 4(a) for the normalized cross-correlation (NCC) similarity. Our detection S' accurately reflects the amount of motion (the plot shows $S' + 0.99$ to fit on the same scale). S' is zero throughout the sequence without motion (hence not included in the plot). Figure 4(b) shows the derivative of our simulated motion (i.e. the motion per frame pair) along with the result of a full non-linear optimization over S' , individually for each image pair. Even though the incremental patient motion between frames has a maximum amplitude of only $\approx 0.75^\circ$, our method can approximately recover it with an average deviation of $0.093 \pm 0.122^\circ$ in $\Delta\mathbf{R}_y$ and $0.046 \pm 0.059^\circ$ in $\Delta\mathbf{R}_z$.

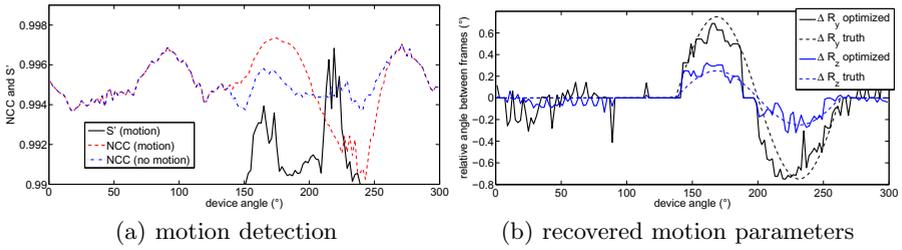


Fig. 4. Result of the motion detection experiment

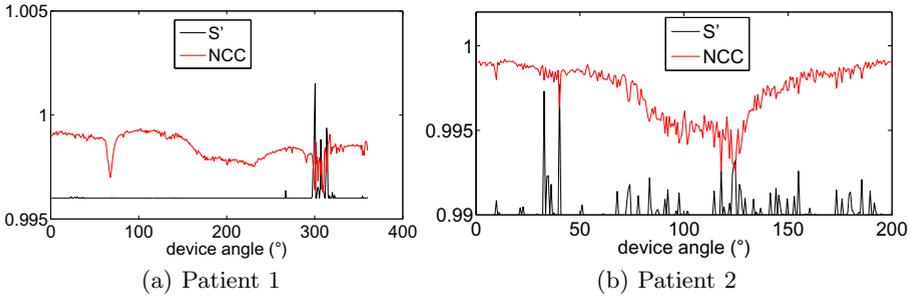


Fig. 5. Motion detection on clinical projection data from two patients

3.3 Parameter Selection

The parameters of our system to be defined up-front, namely the local neighborhood size m and the number of motion layers n , have to be chosen in order to optimize the similarity robustness for a particular clinical application or anatomy. While a systematic evaluation is beyond the scope of this manuscript, we have generally obtained good results for $m = 21^2 \dots 31^2$ and $n = 3 \dots 7$ on various anatomy, given a projection image size of 512^2 pixels. Those parameters are also intuitively defined; in dental applications $n = 3$ is sufficient since the structures are dominated by two layers of teeth moving in opposite direction (see figure 2). The maximum reasonable value of n is where $|\mathbf{x}_{i+1} - \mathbf{x}_i|$ is roughly the size of an individual image pixel (see equation 11).

3.4 Real Projection Data

We have computed the motion detection measure S' on the original projection images of two orthodontic cone-beam acquisitions. Patient 1 exerts a strong head movement towards the end of the sequence. Patient 2 is trembling several times during the entire acquisition. Figure 5 depicts the detection value S' for all frame pairs, and the NCC measure for comparison. The value S' only deviates from zero where patient motion is clearly visible.

4 Conclusion

As opposed to disparity computation from stereo cameras, a pair of X-Ray images yields correspondence at multiple depths. We were able to approximately recover this information as a number of discrete weights, through a local least-squares estimation for every pixel's neighborhood. This in turn allowed us to derive an optimal similarity formulation in projection space, which is sensitive to 3D object changes beyond the expected relation of the projections themselves. We have shown on dental cone-beam data that this measure accurately detects global rigid patient motion. Besides, we could estimate the intra-frame motion parameters with an accuracy of $< 0.1^\circ$ and $< 0.4\text{mm}$, respectively.

We believe that this formulation can help future reconstruction algorithms to better take patient motion into account. Equation 12 can detect motion right during the acquisition and suggest if the reconstruction needs to consider it in the first place. While our actual motion estimates from X-Ray image pairs alone would yield drift applied over a sequence, they can bootstrap more powerful iterative methods which simultaneously estimate a volume and motion parameters. The computed depth weights w_i are a by-product of our formulation, and could be exploited to accelerate the back-projection step during a reconstruction. Extensions to use more than two projections, as well as non-linear motion models, are straightforward.

References

1. Ens, S., Ulrici, J., Hell, E., Buzug, T.: Automatic detection of patient motion in cone-beam computed tomography. In: 2010 IEEE International Symposium on Biomedical Imaging, pp. 1257–1260 (2010)
2. Hansis, E., Schomberg, H., Erhard, K., Dössel, O., Grass, M.: Four-dimensional cardiac reconstruction from rotational x-ray sequences - first results for 4D coronary angiography. In: SPIE Medical Imaging 2009 Conference, vol. 7258 (2009)
3. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press, Cambridge (2004)
4. Khamene, A., Bloch, P., Wein, W., Svatos, M., Sauer, F.: Automatic registration of portal images and volumetric CT for patient positioning in radiation therapy. Medical Image Analysis, 96–112 (February 2006)
5. Rohkohl, C., Lauritsch, G., Biller, L., Hornegger, J.: ECG-gated interventional cardiac reconstruction for non-periodic motion. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 151–158. Springer, Heidelberg (2010)
6. Schretter, C., Pilatus, F., Rose, G., Aach, T., Bertram, M.: Optical flow motion estimation for approximate motion compensation in cone-beam CT. In: Proceedings of the 10th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, pp. 327–330 (2009)
7. Wein, W., Brunke, S., Khamene, A., Callstrom, M., Navab, N.: Automatic CT-ultrasound registration for diagnostic imaging and image-guided intervention. Medical Image Analysis 12, 577–585 (2008)