Virtual Single-frame Subtraction Imaging

Mathias Unberath*†, André Aichert*, Stephan Achenbach‡, and Andreas Maier*†
*Pattern Recognition Laboratory, FAU Erlangen-Nuremberg, Germany
†Graduate School in Advanced Optical Technologies, Erlangen, Germany
‡University Hospital: Cardiology and Angiology, FAU Erlangen-Nuremberg, Germany
Email: mathias.unberath@fau.de

***Abstract*—We outline a generic framework for single-frame, detector-domain material decomposition. The method involves a segmentation and a background estimation step yielding a virtual mask image that can be used for subtraction. In many cases, material decomposition yields non-truncated difference images enabling the use of novel motion estimation methods that exploit epipolar consistency conditions.**
In this work, a pipeline for virtual digital subtraction coronary angiography is presented and evaluated on a phantom data set. The pipeline consists of Hessian-based vessel segmentation followed by background estimation in Fourier domain. Center of mass tracking and a metric based on epipolar consistency conditions is then used to estimate vertical detector translations that serve as a surrogate for respiratory and cardiac motion. When assessing the heart phase, we achieved a correlation of 0.91 between the ground truth ECG and the image-based surrogates. The results encourage further experiments on real data as well as the application for intra-scan motion compensation.

I. Introduction

Robust methods for intra-scan patient motion estimation, such as center-of-mass (CoM) tracking or use of epipolar consistency conditions (ECC), require non-truncated data [1]. Unfortunately, this requirement is hardly ever satisfied for scans of thoracic or abdominal regions. However, the object of interest, such as the coronary arteries, may lie completely in the field of view. In order to increase visibility and enable above-mentioned motion estimation techniques, it would be beneficial to separate the object from the background.
A widely used technique for decomposition is digital subtraction imaging, allowing separation of structures that can be enhanced using some sort of contrast-agent. The most widely known representative of these techniques is digital subtraction angiography (DSA) [2]. Traditionally, the method requires two acquisitions with and without contrast enhancement, that are being referred to as the fill and mask scan, respectively. Requiring two asynchronous scans makes the method susceptible to intra- and inter-scan patient motion [3]. Single-frame material decomposition, however, does not rely on asynchronous imaging of the same scene and allows for lower patient dose. So far, it has been described in the context of energy-resolving X-ray detectors that are not yet part of clinical routine [4]. Nonetheless, single-frame material decomposition may be possible for conventional acquisitions exploiting segmentation and interpolation strategies. Similar methods have been applied in the context of high-intensity object masking [5] and background estimation [6] where they are used for artifact reduction in reconstructions.
We outline a generic framework for virtual single-frame subtraction imaging of spatially sparse structures, such as contrasted vessels or metal implants, to enable the application of algorithms that are restricted to non-truncated data. We present preliminary results of coronary artery motion estimation using CoM-tracking and ECC on phantom data [7], and show that the motion patterns can be used for image-based gating.

II. Material and Methods

The schematic of the generic procedure is shown in Fig. 1. A projection image in which the object of interest is well visible serves as input to the method. The method involves the following steps:

i. Segmentation of the structure yielding a binary mask
ii. Background estimation in the masked regions
iii. Digital subtraction of input and virtual background

Success of the method relies on the two key components segmentation and background estimation, which we will refer to as inpainting. Both elements can be exchanged arbitrarily in the sense that their function is fixed while the specific method can be chosen to best handle the underlying problem.
We describe an exemplary pipeline designed for rotational C-arm CT coronary angiography, elaborating on the methods used for segmentation and inpainting of coronary arteries.

A. Preprocessing and segmentation

The segmentation algorithm described here makes use of derivatives that act like a high-pass filter. To suppress the influence of noise while preserving edges, a bilateral filter is applied to the input images $I'_a,\ a = 1, ..., N$ [8]. The filtered image I_a at position $\mathbf{u} \in \mathbb{R}^2$ is given as

$$I_a(\mathbf{u}) \propto \sum_{\mathbf{u}_i \in \Omega} I'_a(\mathbf{u}_i) \cdot f_{\sigma_r}(\|I'_a(\mathbf{u}_i) - I'_a(\mathbf{u})\|) \cdot f_{\sigma_d}(\|\mathbf{u}_i - \mathbf{u}\|)\,,$$

where Ω is a local neighborhood, and $f_\sigma(x) \propto \exp(^{-x^2}/_{2\sigma^2})$ is a Gaussian function defining neighborhood weights in spatial and intensity domain. In the above equation, we omitted the normalization factor $\mathcal{N}(\mathbf{u})$ for more compact notation.
Subsequently, the projection images are segmented using a combination of morphological and Hessian-based filters [9]. Coronary arteries manifest as bright, small tubular structures on a darker, slowly varying background. A circular top-hat filter yields high responses for bright structures smaller than its radius R and can, therefore, be used to remove all structures larger than the structuring element [10]. Generally, in the

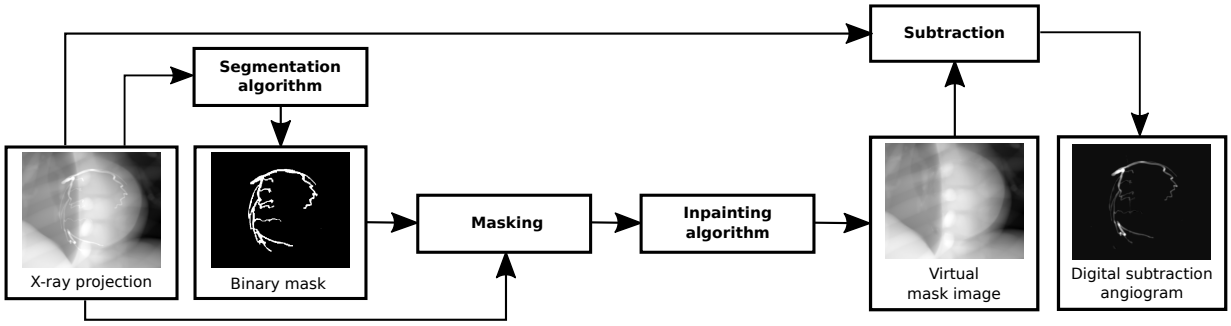

Fig. 1. Schematic of the proposed method. Having obtained a binary segmentation of the object of interest, the background is estimated using an inpainting method. The virtual mask image can then be subtracted from the original projection yielding the difference image.

top-hat filtered image I_a^{TH} responses from non-tubular structures are not sufficiently suppressed. Therefore, we include a vesselness-filtered version of the image I_a^{V}. The filter uses the Eigenvalues of the Hessian $\lambda_{1,2} = \lambda_{1,2}(\mathbf{u}, s)$ at position $\mathbf{u}$ and scale s such that $|\lambda_1| \geq |\lambda_2|$. They enable the definition of physically meaningful measures, the blobness $R_b = {}^{\lambda_2}/_{\lambda_1}$, and the structureness $\mathcal{S} = \sqrt{\lambda_1^2 + \lambda_2^2}$. Then if $\lambda_1 > 0$, the vesselness reads

$$V(\mathbf{u}) = \max_s \ \exp(-\alpha R_b) * (1 - \exp(-\beta \mathcal{S}))\,, \qquad (1)$$

where α, and β are constants [11]. The enhanced images I_a^{E} are then obtained by combination of vesselness and top-hat filtered images I_a^{V} and I_a^{TH}, respectively, yielding

$$I_a^{\mathrm{E}}(\mathbf{u}) = \begin{cases} I_a^{\mathrm{V}}(\mathbf{u}) & \text{if } I_a^{\mathrm{V}}(\mathbf{u}) > t_{\mathrm{V}} \wedge I_a^{\mathrm{TH}}(\mathbf{u}) > 0 \\ 0 & \text{else} \end{cases}, \qquad (2)$$

where t_{V} is an empirically determined threshold. A binary segmentation mask $\mathcal{W}_a$ is then calculated from the enhanced image I_a^{E} using hysteresis thresholding. The mask will be used as a binary defect window. Therefore, $\mathcal{W}_a(\mathbf{u}) = 1$ if $\mathbf{u}$ belongs to the background and 0 otherwise.

B. Background estimation: Spectral inpainting

Masking the object of interest can be expressed as a multiplication of the projection with the defect window in spatial domain, yielding a defective image $G_a(\mathbf{u}) = I_a(\mathbf{u}) \cdot \mathcal{W}_a(\mathbf{u})$. Inpainting of the defective image is equivalent to estimating the measurement of defect detector pixels, where spatially extensive regions do not carry information. Many estimation techniques in spatial domain, such as Thin-Plate-Spline interpolations, exist [5]. However, due to their locality they only perform well for sufficiently small defects and change the noise characteristic, which may lead to an unnatural appearance of the image [12].

Frequency domain methods, such as spectral interpolation, have been shown to work well for large defect areas and noisy structures [5][12]. We seek to estimate the undistorted background $\mathcal{B}_a$ from the observation $G_a(\mathbf{u}) = \mathcal{B}_a(\mathbf{u}) \cdot \mathcal{W}_a(\mathbf{u})$. In frequency domain the relation is expressed in terms of convolution $\mathcal{G}_a = \mathcal{B}_a \star \mathcal{W}_a$, where lowercase letters denote the Fourier transform of its uppercase equivalent. The method tries to iteratively deconvolve the unknown background spectrum $\mathcal{B}_a$ and the window spectrum $\mathcal{W}_a$, utilizing the symmetry property of the Fourier transform of real valued signals [12].

To preserve the locality of image appearance, spectral interpolation is performed consecutively on patches rather than on the whole projection at once. Patch-based processing, however, implies weighting with a rectangular window that decreases the dynamic range and sensitivity due to spectral leakage [13]. To mitigate such effects, an apodization window is applied to the patches before converting to Fourier domain.

Finally, the estimated background image is then subtracted from the projection, yielding the virtual digital subtraction angiogram $D_a(\mathbf{u}) = I_a(\mathbf{u}) - \mathcal{B}_a(\mathbf{u})$.

C. Intra-scan motion and its implications

The subtraction images D_a are, in contrast to the original projections I_a not truncated and can, therefore, serve as an input to CoM- and ECC-based motion estimation algorithms. The CoM $\mathbf{c}_a$ in frame a is calculated from the pixel intensities:

$$\mathbf{c}_a = \frac{1}{\sum_{\mathbf{u} \in D_a} D_a(\mathbf{u})} \sum_{\mathbf{u} \in D_a} D_a(\mathbf{u}) \cdot \mathbf{u}. \qquad (3)$$

The method yields 2D shifts, corresponding to motion in the plane orthogonal to the viewing direction.

Use of ECC relies on the epipolar geometry between views a_1 and a_2 with projection matrices P_{a_1} and P_{a_2}, respectively. Assuming a parallel acquisition geometry, integration over corresponding epipolar lines $\mathbf{l}_{a_1}(\kappa)$ and $\mathbf{l}_{a_2}(\kappa)$ gives two redundant ways of calculating the integral over the epipolar plane $\mathbf{E}(\kappa)$ through the object, where κ defines the angle of the epipolar plane [1]. There exists a pencil of such planes around the baseline each with a different angle κ, allowing for the definition of consistency conditions. For cone-beam geometry the relationship outlined above has to be modified using Grangeat's theorem, yielding

$$\frac{\mathrm{d}}{\mathrm{d}t} \rho_{a_1}(\mathbf{l}_{a_1}) - \frac{\mathrm{d}}{\mathrm{d}t} \rho_{a_2}(\mathbf{l}_{a_2}) \approx 0, \qquad (4)$$

where $\frac{\mathrm{d}}{\mathrm{d}t}\rho_a(\mathbf{l})$ is the derivative of the integral over line $\mathbf{l}_a$ in image I_a, and t is the distance of the line to the image origin. Requiring mutual consistency among all images $I_a,\ a = 1, ..., N$ at multiple lines $\mathbf{l}$ allows the definition of a

metric that can be optimized to estimate motion parameters responsible for the inconsistencies [1].

D. Surrogate signal extraction

Respiratory and cardiac motion both are assumed to manifest in a global translation in head-foot direction with low and high frequency, respectively [14]. Therefore, we will restrict the motion model to 1D translations in vertical detector coordinate direction, that we will refer to as v-shifts. The shifts $\mathbf{v} = (v_1, ..., v_N)$ are transformed to Fourier domain applying a Hann window for apodization [13]. Subsequently, the signal is separated into two components $\mathbf{v}_{\text{resp}}$ and $\mathbf{v}_{\text{card}}$ containing frequencies below and above a certain threshold f_{sep}, respectively. Then, the heart rate f_{ecg} is extracted as the dominant frequency in the power spectral density of $\mathbf{v}_{\text{card}}$. Using f_{ecg} as the heart rate, a normalized cardiac time is calculated for each image a as $t_{a,\text{card}} = \frac{aTf_{\text{card}}}{N} \pmod 1$, where T is the scan duration.

E. Data and Experiments

The proposed methods were evaluated on Cavarev, an XCAT-based phantom data set exhibiting respiratory and cardiac motion [7], [15]. The data set consists of 133 projections acquired over 5 s. The radius for top-hat filtering was 4 pixels, the vesselness threshold $t_{\text{V}} = 0.5$. The segmentation accuracy was assessed using the Dice score. The patch size used for spectral interpolation was 160×160, 100 iterations were used. A Blackman window was applied to each patch. We calculated v-shifts using CoM- and ECC-based methods. As no mask scan is included in the Cavarev data set, the inpainting and subtraction algorithm could only be evaluated implicitly using the results of both motion estimation techniques. The threshold for respiratory and cardiac signal separation was chosen as $f_{\text{sep}} = 1\,\text{Hz}$. We demonstrate the correlation of the respiratory signal $\mathbf{v}_{\text{resp}}$ with the ground-truth qualitatively, but present quantitative results for the normalized cardiac time, i.e. Pearson's R.

III. Results and Discussion

Representative results of the virtual single-frame subtraction pipeline are shown in Fig. 2. The segmentation algorithm described in Sec. II-A yielded a Dice score of 0.98 ± 0.14 with respect to a manual segmentation. Nevertheless, segmentation errors were present (see Fig. 2b) indicating that a more complex segmentation algorithm may be necessary to handle real data. The virtual mask and difference image are shown in Fig. 2c and 2d, respectively. While quantitative evaluation of the background estimation is not possible with Cavarev, the visual results suggest that spectral interpolation is able to satisfactorily estimate the background for narrow structures such as the coronary arteries. Larger scale structures, however, may require different inpainting algorithms up to a point where reliable background estimation becomes impossible as too much information is omitted from the image. Segmentation errors are not as prominent in the resulting difference image or not visible at all, indicating that background estimation may hamper artifact propagation if the erroneously masked region is well explained by the remaining image. This observation is encouraging, as the subsequent motion estimation step then has to deal with fewer artifacts which may lead to a more robust estimation.

The v-shifts $\mathbf{v}$ obtained from CoM- and ECC-based methods are shown in Fig. 3a. Both shifts have a similar range on the detector of $37.64\,\text{mm}$ and $39.24\,\text{mm}$ for CoM and ECC, respectively. The offset, both in Fig. 3a and 3b, was adjusted for better visualization. Shifts obtained using the ECC-based method appear smoother than the ones extracted using CoM tracking. An explanation for this behavior could be that CoM calculation is more susceptible to remaining artifacts as it directly uses image intensities of every image independently. The ECC-based method on the other hand requires integration over epipolar lines and bundle optimization [1], potentially allowing more robust estimations. In the same figure it can be observed that the extracted low frequency signal $\mathbf{v}_{\text{resp}}$ is in good agreement with the ground truth respiratory phase. An attempt to directly assess the breathing frequency using the power spectral density did not yield meaningful results. We believe this shortcoming to be related to the low amount of observed breathing cycles (fewer than 1.5).

The high frequency signal $\mathbf{v}_{\text{card}}$, resulting normalized time $\mathbf{t}_{\text{card}}$, and the ground truth cardiac time are shown in Fig. 3b. The periodic signals $\mathbf{v}_{\text{card}}$ visually correlate well with the ground truth and support the assumption of global head-foot motion during contraction at least for this phantom study. Albeit different in spatial and frequency domain, both signals exhibit the same dominant frequency of $1.41\,\text{Hz}$ yielding the same normalized times $\mathbf{t}_{\text{card}}$ and, therefore, Pearson R of 0.91. Hence, the shifts could be used as surrogate for the heart phase allowing image-based gating. This may be beneficial for gated reconstructions as the surrogate is derived from the actual motion state rather than the electrophysiological excitation.

Separation of low and high frequency components of $\mathbf{v}$ by thresholding in Fourier domain worked well overall. However, the separation does not seem optimal everywhere. For the signals obtained using the CoM and the ECC the effect can be observed at small and large projection numbers, respectively. Although the effect was subtle and did not affect the heart rate estimation in Fourier domain, it may become bothersome when using the shifts for motion compensated reconstructions. In such cases, however, the signal decomposition could be performed in a reversed order if an ECG was acquired simultaneously. Then frequencies related to the heart beat could easily be omitted allowing for accurate respiratory motion estimation.

IV. Conclusion

We discussed a generic pipeline for virtual single-frame subtraction imaging enabling detector domain material decomposition. We introduced a representative pipeline targeted

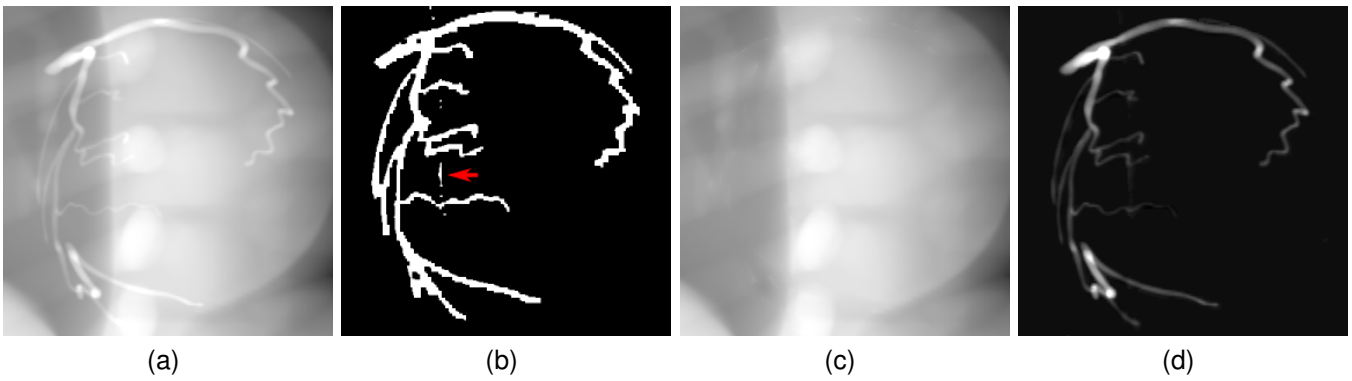

Fig. 2. Crop of the original projections (not shown here) to the region of interest. The original projection, the segmentation mask, the inpainted image, and the DSA image are shown in Fig. 2a, 2b, 2c, and 2d, respectively.

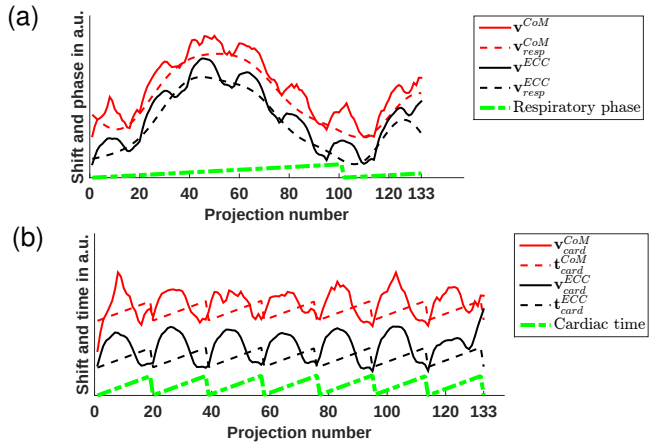

Fig. 3. 1D motion estimation of CoM- and ECC-based methods (solid lines) and the low frequency respiratory signal (dashed lines) are shown in Fig. 3a together with the ground-truth respiratory phase. Fig. 3b shows the high frequency signal, the corresponding normalized time, and the ground-truth.

at coronary arteries and demonstrated its capabilities and limitations in a phantom study. We argued that material decomposition may void truncation for the object of interest allowing for motion estimation techniques based on the CoM or ECC. We demonstrated the applicability of such methods and showed that v-shifts correlate well with both, the respiratory and the cardiac phase. In future work we will evaluate the method on real patient data including but not limited to interventional coronary angiography. We see applications in metal artifact reduction but most importantly in image-based motion compensation. A natural next step would be the extension of the motion model. Estimating 3D translations may allow for respiratory motion compensation, which is hardly feasible using vertical detector coordinate shifts only.

Acknowledgment

The authors gratefully acknowledge funding of DFG MA 4898/3-1 and the Erlangen Graduate School in Advanced Optical Technologies (SAOT) by the German Research Foundation (DFG) in the framework of the German excellence initiative.

References

[1] A. Aichert, M. Berger, J. Wang, N. Maass, A. Doerfler, J. Hornegger, and A. Maier, "Epipolar Consistency in Transmission Imaging," *Medical Imaging, IEEE Transactions on*, vol. 34, no. 10, pp. 1–15, 2015.

[2] W. R. Brody, "Digital subtraction angiography," *Nuclear Science, IEEE Transactions on*, vol. 29, no. 3, pp. 1176–1180, 1982.

[3] M. Nejati and H. Pourghassem, "Multiresolution image registration in digital X-Ray angiography with intensity variation modeling," *Journal of Medical Systems*, vol. 38, no. 2, pp. 1–10, 2014.

[4] X. Wang, D. Meier, K. Taguchi, D. J. Wagenaar, B. E. Patt, and E. C. Frey, "Material separation in X-ray CT with energy resolved photon-counting detectors," *Medical Physics*, vol. 38, no. 3, pp. 1534–1546, 2011.

[5] M. Berger, C. Forman, C. Schwemmer, J. H. Choi, K. Müller, A. Maier, J. Hornegger, and R. Fahrig, "Automatic Removal of Externally Attached Fiducial Markers in Cone Beam C-arm CT," in *Bildverarbeitung für die Medizin 2014*, 2014, pp. 168–173.

[6] C. Blondel, R. Vaillant, G. Malandain, and N. Ayache, "3D tomographic reconstruction of coronary arteries using a precomputed 4D motion field," *Physics in Medicine and Biology*, vol. 49, no. 11, p. 2197, 2004.

[7] C. Rohkohl, G. Lauritsch, A. Keil, and J. Hornegger, "CAVAREV - An Open Platform for Evaluating 3D and 4D Cardiac Vasculature Reconstruction," *Physics in Medicine and Biology*, vol. 55, no. 10, pp. 2905–2915, 2010.

[8] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Computer Vision. Sixth International Conference on*. IEEE, 1998, pp. 839–846.

[9] M. Unberath, S. Achenbach, R. Fahrig, and A. Maier, "Exhaustive Graph Cut-based Vasculature Reconstruction," in *Biomedical Imaging, 2016 IEEE 13th International Symposium on*. IEEE, 2016, to appear.

[10] C. Schwemmer, C. Rohkohl, G. Lauritsch, K. Müller, and J. Hornegger, "Residual motion compensation in ECG-gated interventional cardiac vasculature reconstruction." *Physics in Medicine and Biology*, vol. 58, no. 11, pp. 3717–37, 2013.

[11] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *MICCAI 1998: 1st International Conference, Proceedings of the*. Springer, 1998, pp. 130—-137.

[12] A. Aach and V. Metzler, "Defect interpolation in digital radiography - how object-oriented transform coding helps," in *SPIE Medical Imaging*, 2001, pp. 824–835.

[13] F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proceedings of the IEEE*, vol. 66, no. 1, pp. 51–83, 1978.

[14] C. Blondel, G. Malandain, R. Vaillant, and N. Ayache, "Reconstruction of Coronary Arteries from a Single rotational X-ray projection sequence." *IEEE Transactions on Medical Imaging*, vol. 25, no. 5, pp. 653–63, 2006.

[15] A. Maier, H. Hofmann, C. Schwemmer, J. Hornegger, A. Keil, and R. Fahrig, "Fast Simulation of X-ray Projections of Spline-based Surfaces using an Append Buffer," *Physics in Medicine and Biology*, vol. 57, no. 19, pp. 6193–6210, 2012.