Coronary Tree Extraction Using Motion Layer Separation

Wei Zhang[1], Haibin Ling[2], Shaohua Zhou[1] and Dorin Comaniciu[1]

[1]Siemens Corporate Research,755 College Road E. Princeton, NJ 08540, USA
[2] Dept. Computer Science, Temple University, Philadelphia, PA 19122, USA

Abstract. Fluoroscopic images contain useful information that is difficult to comprehend due to the collapse of the 3D information into 2D space. Extracting the informative layers and analyzing them separately could significantly improve the task of understanding the image content. Traditional Digital Subtraction Angiography (DSA) is not applicable for coronary angiography because of heart beat and breathing motion. In this work, we propose a layer extraction method for separating transparent motion layers in fluoroscopic image sequences, so that coronary tree can be better visualized.. The method is based on the fact that different anatomical structures possess different motion patterns, *e.g.*, heart is beating fast, while lung is breathing slower. A multiscale implementation is used to further improve the efficiency and accuracy. We show that the proposed approach helps to enhance the visibility of the vessel tree, both visually and quantitatively.

1 Introduction

Coronary angiography is an important enhancement method for the analysis of coronary images and therefore has many clinician applications for cardiac related diseases. However, the task is very challenging because images contain overlaying structures besides blood vessel, *e.g.* Figure 1. Moreover, radiation is kept low for patient's health and thus blood vessels often have poor visibility. *Digital subtraction angiography* (DSA) [7] has been widely used in interventional radiology for enhancing the vessel structure. However, it meets difficulties when applied to the coronary images due to the complicated motion induced by heart beating and breathing.

In this paper we propose a new framework to solve the above problem using motion layer separation. The intuition is based on the fact that different anatomical structure have different motion patterns. For example, lungs usually move in a slow and simple way, heart beating causes much more complicated vessel motion, while bones and spines usually remain static. Consequently, it is natural to model an image sequence as a superposition of different motion layers and then use motion separation to extract layers we are interested in. The basic idea is illustrated in Figure 1. The proposed approach is tested on both real and synthetic sequences and demonstrated promising results.

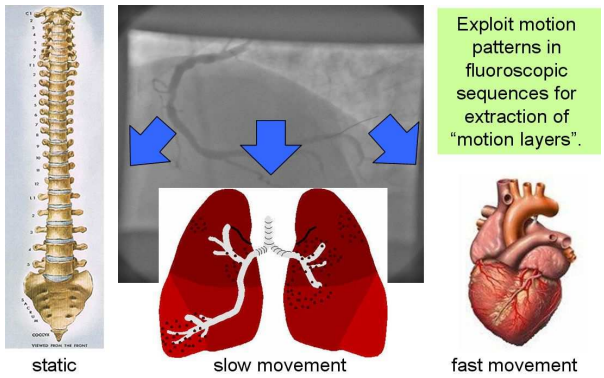


Fig. 1. Different anatomical structures in the scene have different motion patterns.

There are several key ingredients in the proposed approach. First, the proposed coronary extraction approach naturally uses motion layer separation. By doing this, it not only avoids the challenging and error prone mask seeking procedure used in traditional DSA approaches, but also achieves robustness by easily integrating neighborhood information. Second, a multiscale scheme is adopted. In addition to achieving better efficiency and accuracy, the multiscale framework naturally maps different layers to different scales. Third, a dense motion field is established by using *thin plate spline* (TPS) [4]. This enables us to handle complicated vessel motions. Fourth, a quantitative measurement consistent to human perception is introduced for evaluating visibility enhancement.

2 Related Work

Traditional motion estimation with adaptation [3] does not model the transparency issue specifically. It does not work for general transparent motion.
Processing in Fourier space As opposed to the traditional spatial domain approaches, Shizawa and Mase [11] reports that the 3D Fourier transform of a transparent sequence whose transparent layers are in constant translation is made of different planes corresponding to different layers. However, this approach has a major drawback: it assumes a constant transparent motion over a significant number of frames [8].
Explicit transparency modeling This framework assumes constant motion over three successive frames for two-transparent-layer case. The constraints become cumbersome when we are dealing with more transparent layers. Many classical motion estimation methods have been adapted to the transparency case substituting the brightness assumption with some constraint equations. The popular methods that fall under this category are : block-matching techniques and random fields [12], regularization [13], multi-resolution [1] etc.
Parametric models - The Chicken and Egg problem The detection of multiple motions can be addressed as a segmentation problem. However, the

Algorithm 1 Multiscale Coronary Extraction

/*Motion layer separtion at low resolution of 128×128*/
Data initialization and motion field preparation at low resolution
Solve equation (4) for three layers: background, lung, and vessel
/*Motion layer separtion at middle resolution of 256×256*/
Upsample from low resolution and remove background layer
Data initialization and motion field preparation at middle resolution
Prepare motion field at middle resolution
Solve equation (4) containing two layers for lung, and vessel
/*Final result at high resolution of 512×512*/
Upsample from middle resolution and subtract lung layer at high resolution

optical flow-field segmentation problem is coupled with the estimation of the flow itself which is a chicken-and-egg problem. If the flow were accurately given everywhere then we can find the motion boundaries. These methods try to solve the segmentation and estimations iteratively. The popular frameworks proposed in the literature are the expectation-maximization framework [14], shift-and-subtract strategy [2], temporal integration to blur the uninterested regions [6].
Repetitive motion When the dynamics in one layer is assumed to be repetitive, global-to-local space-time alignment can be used to extract the other layer which can have arbitrary motion [9].

3 Our Approach

In this section we describe the proposed approach. We take a hierarchical framework for both efficiency and effectiveness. In this framework, we first subtract the background layer at low resolution. Then the lung and vessel layers are separated at middle resolution. Finally, the vessel layer is refined at the finest scale. A summary of the approach is given in Algorithm 1.

3.1 Problem Formulation

Without loss of generality, we assume that our problem is to extract layers from image sequences which consist of N layers overlapping with uniform transparency, i.e., the contribution of each layer is $1/N$. For X-ray image formation, an exponential relationship exists between the incident and transmitted photon fluence [10]. After passing through multiple layers of materials, the final amount of photon fluence received by the detector which forms the X-ray image is:

$$N_x = N_0 e^{\sum \mu_i x_i}, \tag{1}$$

where μ_i is attenuation coefficient of the i_{th} layer and x_i is its thickness.

Because of the exponential form of the X-Ray image formation, X-ray images are usually manipulated in the logarithmic space. In the logarithmic image space,

the image can be written as a linear combination of the layers:

$$I_m = \sum_{l=1}^{N} T_l^m L_l, \tag{2}$$

where I_m is the m^{th} observed image, L_l is the l^{th} unknown layer. T_l^m is the motion field for the l^{th} layer that maps L_l to I_m. Refer to Sec. 3.3 for how the motion field is constructed. I_m and L_l have the same size. Denote W and H as their width and height, respectively. L_l has $W \times H$ pixels with unknown values. Our goal is to find images of different layers by minimizing the reconstruction error. For achieving robustness to noise, we use $M(M > N)$ images to find a least square solution to the following equation

$$\min \sum_{m=1}^{M} \| \sum_{l=1}^{N} T_l^m L_l - I_m \|^2 \quad . \tag{3}$$

To solve (3), it is transformed into a least square optimization form. We omit the details because of the page limit.

$$\arg\min_x \|Cx - d\|^2 \ , \tag{4}$$

where $C \in \mathbb{R}^{K \times J}$ is a matrix that is constructed based on the motion field, $x \in \mathbb{R}^J$ is a vector that is the concatenation of all the unknown pixel values of all N layers, $d \in \mathbb{R}^K$ is a vector which is the concatenation of all the (known) pixel values of all the M images. $K = W \times H \times M$ and $J = W \times H \times N$.

In practice, the number of images M should be relatively big to tolerate small shape deformation in one image. On the other hand, it should not be too big because shape deformation would accumulate with more images, which would violate the assumption that sum of all layers equal to the observed image. In our experiments, we empirically found that $M = 9$ produces the best result. This coincides with the fact that the number of frames per heart cycle is roughly 9. The workflow is the following: for each image I_m in the sequence, we use images in a local time window $I_{m-4}, I_{m-3}, \ldots, I_{m+4}$ and the corresponding motion fields to construct C and d in Equation 4. Once the optimization converges to a solution x, we re-arrange x and obtain all the transparent layers in that image. After all the images have been processed, a video sequence is obtained for each layer.

3.2 Multiscale Framework

The least squares problem (4) is very large in scale. For example, to estimate 3 layers for a sequence with image size 512×512, the number of unknowns is 786432. On the other hand, the matrix C is very sparse with most of its entries to be zero. So we can resort to the iterative optimization technique to solve the problem. Notice that the value of unknowns can not be negative, we are actually trying to solve a constrained optimization problem. In the current implementation, the solution is found using a trust region method based on the

interior-reflective Newton method. In each iteration, an approximate solution is obtained using the method of preconditioned conjugate gradients.

For the coronary angiography, we could assume that there are 3 layers in the image: the static background like rib, the slow moving lung and fast beating vessels. To reduce the complexity of the problem, we perform the estimation in a pyramid. For the image of size 512×512, we build additional two levels of size 256×256 and 128×128. Assuming that the background remain unchanged through out the sequence, we estimate it first using the lowest resolution. Following the workflow mentioned above, we could get a sequence of estimated background images. Most of them are quite similar but not necessarily the same. There might be few erroneous estimates because of the correlation in motions of different layers. Nevertheless, the median of of all those estimation gives a reliable estimate of the overall background. Then the estimated background is upsampled to 256×256 and we subtract the background from each image for estimating the remaining two moving layers (vessel tree and diaphragm). As Figure2 illustrated, only the moving objects are retained after subtracting the background and they look more evident than in original image.

Now the new (foreground) sequence contains only two transparent layers. We follow the same procedure to estimate the diaphragm layer and vessel layer in the middle resolution (256×256). After reducing the number of layers, the number of unknowns in Equation 4 is now $256 \times 256 \times 2 = 131072$. Then we upsample the estimated diaphragm layer to 512×512 and subtract it from the foreground sequence, thus obtain the final vessel layer at the original resolution. By doing the whole process in a pyramid, we keep pushing details into the vessel layer because this is really what the clinician want. In addition, we avoid the big problem of trying to optimize 786432 unknowns all together. This not only brings the huge advantage in efficiency, but also helps to avoid some local maxima for the direct optimization as mentioned above. The idea is to takes advantage of the fact that the background layer is invariant across time and slow moving layer is relatively invariant compared to vessel, so that a large part of the search space can be eliminated.

3.3 Constructing Motion Field

The motion of vessels is very complicated, global transformation like affine transformation [1] is therefore insufficient to model it. Instead, we use a dense motion field to represent the vessel motion. Specifically, for any location with coordinates (x, y) at layer L_l, the motion field at this particular coordinates $T_l^m(x, y)$ maps (x, y) to a new position (x', y') in image I_m.

An efficient way to construct the motion field is through the *thin plate spline* (TPS) [4] interpolation. Given two point sets with correspondence between them, TPS finds a nonlinear warping by minimizing a second order "bending energy". In our task, the TSP warping has the following formula

$$\begin{cases} x' = a_x x + b_x y + c_x + \sum_{j=1}^{n} d_{x,j} U((x, y) - (x_j, y_j)) \\ y' = a_y x + b_y y + c_y + \sum_{j=1}^{n} d_{y,j} U((x, y) - (x_j, y_j)) \end{cases}, \tag{5}$$

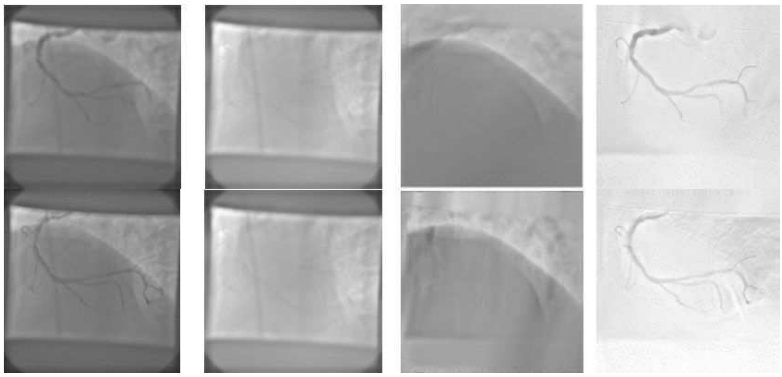


Fig. 2. Vessel tree separation. From left to right: input fluoroscopic image, background, diaphragm, and vessel tree.

where $\{(x_j, y_j)\}_{j=1}^{n}$ is the sparse anchoring point set; $a_i, b_i, c_i, d_{i,j}$ for $i = \{x, y\}, j = 1..n$ are warping parameters that are estimated from sparse motion vectors; and $U(.)$ is a Radial Basis Function. We use manually selected control points (e.g. junction points on vessels) to get sparse motion vectors since this work focuses on motion layer separation. For automatic estimation of sparse motion vectors, We have developed techniques for automatic detection [15] and tracking [16] of thin curved structures, which could be used for vessel detection and tracking in low resolution images. The detection of junction points can be done using the Marginal Space Learning Framework [17].

4 Experimental Results

4.1 Experiments on Real Sequences

We tested the proposed approach on two real sequences, one for vessel enhancement and another for guidewire enhancement. Figure 2 shows the extracted three layers for one image. As can be seen, static objects like bones are retained in the background layer. Slow moving diaphragm is extracted in another layer. The vessel tree we are interested in lies in its exclusive layer and is more visible than that in the original fluoroscopic image. We also applied the method to separating the guidewire and lung in another sequence and the result is shown in Figure 3. Please refer to supplemental videos for more results.

4.2 Evaluation of Visibility Enhancement

While we can see the visibility improvement qualitatively, it's desired to quantitatively analyze the enhancement. A synthetic sequence is shown in Figure 4(a). The measurement we use is based on the Jeffries-Matusita distance, which measures the separability between two classes, vessel and background. The higher

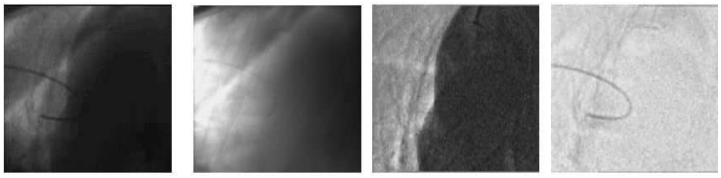
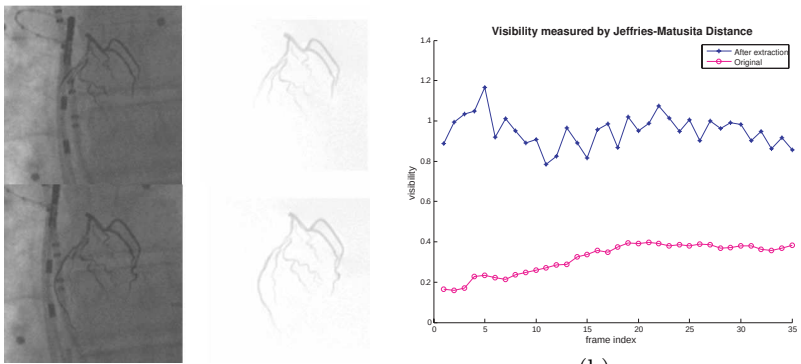


Fig. 3. Guidewire separation. From left to right: original fluoroscopic image, background, lung, and guidewire. The guide wire is hard to see before separation.

the measure, the better the separability.

$$J_{vb} = 2(1 - e^{-B_{vb}}), \qquad B_{vb} = \frac{(\mu_v - \mu_b)^2}{4(\sigma_v^2 + \sigma_b^2 + \epsilon)} + \frac{1}{2} \ln \frac{\sigma_v^2 + \sigma_b^2}{2\sigma_v \sigma_b}, \tag{6}$$

where μ_v and μ_b are means of vessel and background, respectively and σ is the standard deviation. Note there is a term ϵ in the the formulation, which is added to to make it consistent with the human perception. We have experimentally set $\epsilon = 100$ according to the input of multiple experts. For the synthetic image in Figure 4(a), $J_{vb} = 0.31$, while the extracted result in the right has $J_{vb} = 0.89$. The evaluation of the whole sequence is summarized in Figure 4(b). We have carried out extensive experiments with different noise adding to the synthetic image. They all exhibited clearly quantitatively enhancement.



(a) (b)

Fig. 4. (a) Synthetic image with moving background (left), extraction result (right). (b) Quantitative evaluation of visibility enhancement.

5 Conclusion

In this paper, we propose a framework for separating transparent layers from fluoroscopic image sequences, so that coronary tree can be better visualized. It utilizes motion information to decompose an image into different layers. The proposed approach does not require any pre-selected mask, thus avoid the difficulties of the traditional DSA. The motion layer extraction is proceeded in a hierarchical fashion to achieve both efficiency and robustness. The proposed approach is tested on both real and synthetic sequences, and promising results are observed. Our ongoing work focus on automatic motion vector detection and learning-based tracking of vessel structures.

References

1. V. Auvray, J. Liénard, and P. Bouthemy, "Multiresolution parametric estimation of transparent motions and denoising of fluoroscopic images," *MICCAI*, 2005.
2. J. R. Bergen, P. J. Burt, R. Hingorani, and W. Peleg, "A three-frame algorithm for estimating two component image motion.," *IEEE PAMI*, 14(9):886-895, 1992.
3. M. Black and P. Anandan, "The robust estimation of multiple mo- tions:parametric and piecewise-smooth fields," *CVIU*, 63(1):75-104, 1996.
4. F. Bookstein. "Principal Warps: Thin-Plate-Splines and Decomposition of Deformations", *IEEE Trans. PAMI*, 11(6):567-585, 1989.
5. M. Irani and S. Peleg, "Motion analysis for image enhancement: Resolution, occlusion, and transparency," 1993.
6. M. Irani, B. Rousso, and S. Peleg, "computing occluding and transparent motions," *IJCV*, 12(1):5-16, 1994.
7. E. Meijering, W. Niessen, and M. Viergever, "Retrospective Motion Correction in Digital Subtraction Angiography: A Review". *IEEE. Med. Imaging*, 18(18),1999.
8. M. Pingault and D. Pellerin, "Motion estimation of transparent objects in the frequency domain," *J. of Signal Processing*, 84(4):709-719, 2004.
9. B. Sarel and M. Irani, "Separating transparent layers through layer information exchange," *ECCV*, 4:328-341, 2004.
10. J. A. Seibert and J. M. Boone, "X-ray imaging physics for nuclear medicine technologists 2: X-ray interactions and image formation," *J. Nuclear Med. Tech.*, 2005.
11. M. Shizawa and K. Mase, "Principle of superposition: A common computational framework for analysis of multiple motion," 1991, 164-172.
12. I. Stuke, T. Aach, E. Barth, and C. Mota, "Estimation of multiple motions using block-matching and markov random fields," 2004.
13. J. Toro, F. Ownes, and R. Medina, "Multiple motion estimation and segmentation in transparency," *ICASSP*, 2000.
14. Y.Weiss and E. Adelson, "A unified mixture framework for motion segmentation: Incorporating spatial coherence and estimating the number of models," *CVPR*, 1996.
15. A. Barbu, V. Athitsos, B. Georgescu and D. Comaniciu, "Hierarchical Learning of Curves: Application to Guidewire Localization in Fluoroscopy," *CVPR*, 2007.
16. P. Wang, Y. Zhu, T. Chen, W. Zhang, S. Zhou and D. Comaniciu, "Robust Guidewire Tracking in Fluoroscopy," *CVPR*, 2009.
17. Y. Zheng, A. Barbu, B. Georgescu, M. Scheuering and D. Comaniciu. "Fast Automatic Heart Chamber Segmentation from 3D CT Data Using Marginal Space Learning and Steerable Features," *ICCV*, 2007
18. T. Nakaura, S. Ooishi, S. Tayama, and H. Ogawa, "A New Technique for Coronary Digital Subtraction Angiography", *Scientific Meeting Japan. Circul. Society*, 2006.