Multi-layer Deformation Estimation for Fluoroscopic Imaging

J. Samuel Preston[1], Caleb Rottman[1], Arvidas Cheryauka[2], Larry Anderton[2], Ross T. Whitaker[1], and Sarang Joshi[1]

[1] Scientific Computing and Imaging (SCI) Institute, University of Utah
{jsam,crottman,whitaker,sjoshi}@sci.utah.edu
[2] GE Healthcare
{arvi.cheryauka,larry.anderton}@med.ge.com

Abstract. Accurate estimation of motion in fluoroscopic imaging sequences is critical for improved frame interpolation/extrapolation, tracking of surgical instruments, and Digital Subtraction Angiography (DSA). The projection of multiple transparent objects undergoing multiple complicated deformations in 3D onto a single 2D view makes this motion estimation problem quite challenging and ill-suited to existing techniques used in medical image analysis. We propose a novel method for jointly decomposing the observed image into a set of additive layers each associated with its corresponding smooth nonlinear deformation, which together model the non-smooth motion observed in the projection images across several frames. A total variation based regularization penalty is used to incorporate the known structure of the input frames for well posedness of the layer separation problem. We present the use of this model for frame interpolation and artifact reduction in DSA. Results are included from synthetic and real clinical datasets.

1 Introduction

Registration and motion estimation are a mainstay of modern medical image analysis, used for comparison and analysis of structures between patients or across multiple timepoints, motion prediction for treatment planning, and motion compensation for improved reconstruction and denoising, among other applications. Estimating a dense deformation field representing correspondences between image locations is an under-constrained problem, but correspondences modeling image differences due to motion, growth, and inter-subject variability of biological structures has been shown to be well-modeled by smooth deformation fields [12]. Fluoroscopic imaging is an important tool commonly used for diagnosis and interventional procedures. Motion estimation from fluoroscopic imaging is needed for accurate frame interpolation or motion extrapolation. This provides opportunities for lower framerates by reducing the exposure of the patient to ionizing radiation. In addition, Digital Subtraction Angiography (DSA) is a common technique for analyzing the vascular structure for diagnosis as well as interventional procedures [8]. In DSA, a radiographic contrast agent

J.C. Gee et al. (Eds.): IPMI 2013, LNCS 7917, pp. 123–134, 2013.

is injected into a blood vessel, and then a pre-contrast frame (or mask) is subtracted from all the subsequent frames. Ideally, the resulting subtraction would only show the intensity change due to the injected contrast. In practice, normal physiological motion due to breathing and heartbeat as well as other patient motion introduces artifacts in the subtracted images, and so motion estimation between the mask and current frames is needed to suppress these artifacts.

Unlike standard 3D image based deformation estimation problems in which each point in the reference image is assumed to map to a single point in the target image, fluoroscopic imaging techniques generate a projection of a three-dimensional object onto a two-dimensional imaging plane. This results in a motion estimation problem in which smooth motion of the true 3D object projects on to a non-smooth motion in the acquired image, and the intensity of each pixel at one timepoint contributes to the intensity of multiple dispersed locations in another timepoint. A 'true' motion model for this situation would reconstruct the 3D scene and smooth 3D motion field associated with each frame of the imaging sequence. This has been studied in scenes with opaque objects as structure from motion, which is still a difficult and open area of research. As we are primarily interested in estimating realistic deformations and frames as seen from the same viewpoint as the original series, we propose instead to model the motion as a number of additive layers each undergoing a smooth transformation. To achieve our goals, these layers need not represent a segmentation of objects in the scene – they must only separate overlapping objects where contradictory motion violates a smooth-deformation model. The sum of smoothly deforming layers can then accurately describe the motion in the original frames. Our method will jointly estimate these layers and corresponding deformations.

2 Background

While there has been extensive work on layer-based representations of 3D scenes in the computer vision community, the vast majority have assumed opaque objects, segmenting the optical flow field into regions undergoing similar transformations and estimating a set of deformations and unique pixel assignments ([18,19], *etc.*). In cases where transparency or reflections are considered, layer extraction models use only two layers, and even then only in constrained situations such as linear or repetitive motion, or with pairs of images obtained with different compositing functions [13, 16, 17].

Estimation of multiple motions at each point has been studied via the double optical flow constraint proposed in [14] which estimates two motion vectors at each point. Other similar formulations and extensions attempt regularize these fields into consistent motion models or increase the number of motions captured at each point [7, 10, 11, 15]. These formulations do not attempt layer extraction (estimation of the layer intensities) and cannot be used to generate interpolated or motion compensated frames.

In applications to X-ray imaging, Close *et al.* [6] propose an ad hoc hierarchical algorithm for layer extraction in analysis of angiographic stenosis.

This formulation assumes linear transformation of layers, and the sequential layer estimation and removal seems error prone. Chen *et al.* [4] propose a method for two layer extraction with one layer being static and the availability of dual-energy X-rays, which is a far more constrained situation than we wish to model. Auvray *et al.* [1] attempt motion compensation for denoising of fluoroscopic sequences using the three-frame motion constraint of [10]. Although this formulation is derived in terms of translational motion, it is used to solve for nonlinear motion. Also, although multiple motion layers are estimated, only two motions may occur at any point. Further, only motions and regions of influence are modeled, not the actual layer intensities. This greatly constrains the type of prediction or compensation for which this technique can be used.

3 Methods

Focusing on data from fluoroscopic imaging provides both simplifications and complications when compared to working with video sequences with transparency and reflections, a case often studied in the literature. With the exception of objects completely attenuating the imaging signal (a case we will ignore), we can assume that all objects are 'translucent', such that we do not have to consider occlusions. Also, unlike a video sequence in which a panning of the camera to follow a moving object may create one 'layer' whose extent is much larger than the area captured by a single frame, we assume the object being imaged stays mostly within the field of view, with relatively small portions moving in and out of frame. However, unlike much of the work on video analysis, we will not assume that layer motions can be described by a homographies or even low-order polynomial parameterizations. Further, we will not assume a 'dominant' motion between frames, and will therefore avoid a hierarchical motion decomposition.

3.1 Frame Generation Model

We are interested in modeling a time series of fluoroscopic imaging frames as a number of superimposed layers each undergoing a smooth deformation. For a scene with multiple objects, we will model the X-ray intensity at the detector as $\beta \exp(-\sum_i \mu_i d_i)$, where μ_i is the attenuation coefficient, d_i is the thickness of the ith object, and β is a constant representing the maximum transmission value. Assuming a log-transformed image (although in practice clinical data may have some approximation applied), our model subtracts the contribution of each layer from M, a maximum observed image intensity. This produces layers where zero values are be interpreted as the absence of an object.

Assume we are given a series of F frames acquired at evenly spaced time intervals, and that we wish to represent this series with N layers. The frames will be indexed by time, where the first frame occurs at time t_0 and frame F occurs at time t_T, with $\mathrm{T} = \mathrm{F} - 1$. These input frames will be labeled $\mathrm{I}_0(\mathbf{x}) \ldots \mathrm{I}_\mathrm{T}(\mathbf{x})$. We will model each layer $\{\mathrm{L}_n\}_{n=0\ldots\mathrm{N}-1}$ as undergoing its own

smooth deformation $\phi^n(\mathbf{x}, \mathrm{t}), \mathrm{t} \in [\mathrm{t}_0, \mathrm{t}_\mathrm{T}]$ using the standard large deformation model [3, 5], where $\phi(\mathbf{x}, \mathrm{t})$ is defined via the time-varying velocity field $\mathbf{v}^n(\mathbf{x}, \mathrm{t})$, $\phi(\mathbf{x}, \mathrm{t}) = \phi(\mathbf{x}, \mathrm{t}_0) + \int_{\mathrm{t}_0}^{\mathrm{t}} \mathbf{v}(\phi(\mathbf{x}, s), s)\, \mathrm{d}s$, where $\phi(\mathbf{x}, \mathrm{t}_0) = \mathbf{x}$. Smoothness of the deformation ϕ is enforced by penalizing $\int_{\mathrm{t}_0}^{\mathrm{t}_\mathrm{T}} \|\mathbf{v}(\mathbf{x}, \mathrm{t})\|_\mathrm{V}^2\, \mathrm{dt}$, where $\|\mathbf{v}\|_\mathrm{V}^2 =< \mathcal{L}\mathbf{v}, \mathbf{v} >_2$ for smooth velocity field $\mathbf{v}$, and $\mathcal{L}$ is a differential operator penalizing non-smoothness. Under reasonable smoothness assumptions the deformation $\phi(\mathbf{x}, \mathrm{t})$ is guaranteed to be diffeomorphic, guaranteeing the existence of $\phi^{-1}(\mathbf{x}, \mathrm{t})$. Following [3] we will use the notation $\phi_{t,s}(\mathbf{x}) = \phi(\phi^{-1}(\mathbf{x}, \mathrm{t}_t), \mathrm{t}_s)$ for the deformation moving the point $\mathbf{x}$ at time t_t to its location at time t_s. Going forward we will also drop explicit mention of the spatial parameter $\mathbf{x}$ for notational convenience. The model for frame t is then

$$\mathrm{I}_t = M - \sum_{n=0}^{\mathrm{N}-1} \mathrm{L}_t^n + \mathcal{N}(0, \sigma^2), \tag{1}$$

where $\mathrm{L}_t^n = \mathrm{L}^n \circ \phi_{t,0}^n$ and $\mathcal{N}(0, \sigma^2)$ represents the corruption of the ideal image by additive zero-mean Gaussian noise. Although this is not technically correct for our log-transformed photon count model, we assume the process is sufficiently close to a Gaussian model.

3.2 Layer Gradient Penalty

Our method jointly estimates both the layers and the deformations. Even with the smoothness constraint on the deformations imposed by the large deformation model this is an extremely underconstrained problem. In order to formulate a well posed estimation problem, some assumptions regarding the properties of the layers must be made. We wish to separate overlapping objects in the input frames into different layers in our model. Reducing the number of edges in layer images will help with this goal, as the overlapping of transparent objects introduces an edge which will appear in multiple layers if proper separation has not been achieved. A natural choice would be to impose a total variation penalty on the layer images. Such a penalty encourages sparsity of gradients within an image, and more importantly for our application, encourages sparsity of gradients across the layer images. Even with the 'smooth layers' constraint, the formulation permits ambiguous solutions for even simple motion as shown in Figure 1. This shows a small object moving to the right between two timepoints. Solution (a) is the 'expected' solution, however we see that solution (b) may actually be the optimal solution given the tradeoff between deformation[1] and gradient penalties. In order to improve this situation we note that other information is available which can improve the solution. Consistent motion across

[1] A pure translation will not be penalized by our smoothness penalty, however in realistic 2D scenes a pure translation would be uncommon, so for purposes of this example we can associate increased size of deformation with increasing smoothness penalty.

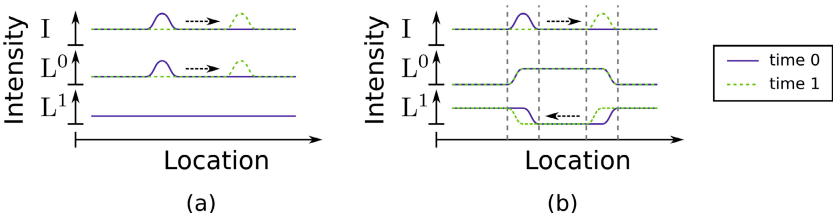

Fig. 1. A 1D example of two possible solutions accounting for the movement of a small object with two layers. Note that a larger deformation is required in case (a), and more edges are required in case (b).

multiple frames can remove ambiguity. We can also use the frame data to improve the layer smoothness penalty. We know that if there is an edge in a frame, that edge should exist in at least one layer. Further, if some location contains no edge in a frame, no edge should exist at that location in any layer at that time. Observe that this is violated in Figure 1 (b). However, as our goal is to separate objects into different layers, we do not want to force every edge into every layer. Consider the following penalty

$$\int_{\Omega} \|\nabla \mathrm{L}^n\|_2 - \nabla \mathrm{L}^n \cdot \mathbf{n}_0 \, \mathrm{d}\mathbf{x}, \tag{2}$$

where $\mathbf{n}_0$ represents the normals of the pseudo-attenuation frame $M - \mathrm{I}_0$ taken with a 'cutoff' regularization based on parameter ϵ:

$$\mathbf{n}_0 = \begin{cases} -\frac{\nabla \mathrm{I}_0}{\|\nabla \mathrm{I}_0\|_2} & \text{if } \|\nabla \mathrm{I}_0\|_2 \geq \epsilon, \\ 0 & \text{if } \|\nabla \mathrm{I}_0\|_2 < \epsilon. \end{cases} \tag{3}$$

This looks like a standard TV penalty on the layer L^n except where an edge occurs in the frame I_0. Here the penalty is eliminated if the gradient in the layer aligns (in the same direction) with the gradient of the frame, and is doubled if they align in opposite directions. Also note that this value is always nonnegative, as $\nabla \mathrm{L}^n \cdot \mathbf{n}_0$ takes its maximum value when $\mathbf{n}_0 = \frac{\nabla \mathrm{L}^n}{\|\nabla \mathrm{L}^n\|_2}$, resulting in zero penalty. It is also zero if $\nabla \mathrm{L}^n$ is zero, meaning there is no penalty for a layer *not* representing an edge in I_0. Of course, noise in the frame will cause incorrect estimates of the normals. It will be important for our results to have nonzero normals only where true edges in the frame exist. To ensure this, we will calculate normals from $\bar{\mathrm{I}}_0$, a denoised version of frame I_0. As we desire sparse gradients, a standard total variation based denoising method is employed. We note that a penalty very similar to (2) is proposed in [9] for the purposes of denoising, where the normals used are estimates the 'true' normals of the image being denoised, and are approximated by the normals of the TV-denoised image.

The undeformed layers are estimated at time t_0, and therefore equation (2) only makes sense for the normals of frame I_0. We do not wish to bias our solution to the configuration of objects observed at t_0, so we propose a version incorporating all frames

$$\sum_{t=0}^{\mathrm{T}} \int_{\Omega} \|\nabla \mathrm{L}^n\|_2 - \nabla \mathrm{L}^n \cdot \tilde{\mathbf{n}}_t \, \mathrm{d}\mathbf{x}, \tag{4}$$

where $\tilde{\mathbf{n}}_t$ represents the normals of the denoised version of $M - \mathrm{I}_t$ taken after deforming it to time t_0, once again with cutoff ϵ:

$$\tilde{\mathbf{n}}_t = \begin{cases} -\frac{\nabla(\bar{\mathrm{I}}_t \circ \phi_{t,0})}{\|\nabla(\bar{\mathrm{I}}_t \circ \phi_{t,0})\|_2} & \text{if } \|\nabla(\bar{\mathrm{I}}_t \circ \phi_{t,0})\|_2 \geq \epsilon, \\ 0 & \text{if } \|\nabla(\bar{\mathrm{I}}_t \circ \phi_{t,0})\|_2 < \epsilon. \end{cases} . \tag{5}$$

As standard numerical solutions of TV denoising do not result in perfectly uniform image regions, the parameter ϵ is chosen to ignore very small gradients, in our case approximately one percent of the input image intensity range.

3.3 Energy Formulation

We will formulate the estimation as an energy minimization problem. Given the current constraints, the energy will have the form

$$E_{\text{total}} = E_{\text{data}} + \lambda_{\mathrm{L}} E_{\text{layer}} + \lambda_{\mathbf{v}} E_{\text{def}}, \tag{6}$$

where E_{data} is the data attachment term, E_{layer} is the layer gradient penalty, and E_{def} enforces the deformation smoothness constraints. The constants λ_{L} and $\lambda_{\mathbf{v}}$ control the tradeoff between the goals pursued by the different terms. We can now explicitly state the penalty terms. The data term will come directly from the frame generation equation (1)

$$E_{\text{data}} = \sum_{t=0}^{\mathrm{T}} \|\hat{\mathrm{I}}_t - \mathrm{I}_t\|_2^2, \tag{7}$$

where $\hat{\mathrm{I}}_t$ is the estimated frame at time t_t; $\hat{\mathrm{I}}_t = M - \sum_{n=0}^{\mathrm{N}-1} \mathrm{L}_t^n$. The layer gradient penalty, as outlined above, is

$$E_{\text{layer}} = \sum_{n=0}^{\mathrm{N}-1} \sum_{t=0}^{\mathrm{T}} \left\| \|\nabla \mathrm{L}^n\|_2 - \nabla \mathrm{L}^n \cdot \tilde{\mathbf{n}}_t \right\|_1, \tag{8}$$

and the deformation smoothness, again as discussed above, is

$$E_{\text{def}} = \sum_{n=0}^{\mathrm{N}-1} \int_{\mathrm{t}_0}^{\mathrm{t}_\mathrm{T}} \left\| \mathbf{v}_t^n \right\|_{\mathrm{V}}^2 \mathrm{d}t, \tag{9}$$

where $\|\cdot\|_{\mathrm{V}}$ is dependent on our choice of $\mathcal{L}$. In this work we will use $\mathcal{L} = \nabla^2$, the Laplacian operator. Note that we also should normalize for the number of frames, but for brevity we have absorbed this in our constants λ_{L} and $\lambda_{\mathbf{v}}$. Our problem is then formulated as

$$\underset{\{\mathrm{L}^n\}_n, \{\mathbf{v}_t^n\}_{n,t}}{\arg\min} \; E_{\text{total}} \left(\{\mathrm{I}_t\}_t, \{\mathrm{L}^n\}_n, \{\mathbf{v}_t^n\}_{n,t} \right). \tag{10}$$

3.4 Residual Layers

Although the smooth deformation of layers described above can adequately describe the motion caused by breathing, heartbeat, and other deformations of 3D anatomy, there are some important cases for fluoroscopic imaging which violate this model. Specifically, the introduction of new objects such as tools during interventional procedures or contrast agents during angiography are not handled by our model. In order to accommodate these cases, additional layers with specific properties meant to model these situations are introduced.

We will look at the case of a radiographic contrast medium introduced into the vascular system to increase the contrast between blood vessels and surrounding tissue, thereby exposing vessel blockages, leaks, and abnormalities . The contrast enters the frame as a large dark object, and spreads rapidly from frame to frame. In order to model the contrast, we introduce a 'residual' layer which accounts for inter-frame changes not well modeled by a deformation. Based on the observed properties of contrast, we model it as a smooth contiguous object. The layer should also be sparse, containing mostly zero values. We therefore estimate a layer at each timepoint which is not subject to deformation, and impose a TV penalty and L^1 penalty on this layer. The model for our estimated frame is then $\hat{\mathrm{I}} = M - \sum_{n=0}^{\mathrm{N}-1} \mathrm{L}_t^n - \mathrm{b}_t$, where b_t is the residual layer at t_t, $t \in \{0 \dots \mathrm{T}\}$, and we introduce a new term to the energy minimization

$$E_{\mathrm{res}} = \lambda_{\mathrm{TV}} \sum_{t=0}^{\mathrm{T}} \left\| \|\nabla \mathrm{b}_t\|_2 \right\|_1 + \lambda_{L^1} \sum_{t=0}^{\mathrm{T}} \left\| \mathrm{b}_t \right\|_1. \tag{11}$$

again accounting for normalization over the number of frames in the constants λ_{TV} and λ_{L^1}.

3.5 Discretization and Solution

Since the time interval between frames is arbitrary in our formulation, we choose unit temporal spacing, $\mathrm{t}_0 = 0, \mathrm{t}_1 = 1, \dots, \mathrm{t}_{\mathrm{T}} = \mathrm{T}$. We expect small deformations between subsequent frames, so our discretization of a time-varying velocity field $\mathbf{v}(\mathbf{x}, \mathrm{t})$ will match the frame times such that there is one piecewise-constant (in time) velocity field $\mathbf{v}_t(\mathbf{x})$ corresponding to each frame I_t, $t \in \{0 \dots \mathrm{T}-1\}$. Euler integration in time will be used for generating ϕ from v, and bilinear interpolation is used for deformation of images and composition of deformations. For reverse-time integration, the small-deformation approximation $v^{-1} = -v$ will be used.

In order to optimize the layer gradient penalty (8), we use a primal-dual strategy based on [20]. This choice is based on properties of the regularized primal variation as $\nabla \mathrm{L}_t^n \to 0$ with $\nabla \mathrm{I}_t \neq 0$, which could force some portion of $\nabla \mathrm{I}_t$ into each layer. Noting that $\|\nabla \mathrm{L}^n\|_2 = \nabla \mathrm{L}^n \cdot \frac{\nabla \mathrm{L}^n}{\|\nabla \mathrm{L}^n\|_2}$ and choosing a dual variable $\mathbf{p}^n s.t. \|\mathbf{p}^n\|_2 \leq 1$ approximating $\frac{\nabla \mathrm{L}^n}{\|\nabla \mathrm{L}^n\|_2}$ for $n \in \{0 \dots \mathrm{N}-1\}$, we rewrite (8) as

$$E_{\mathrm{layer}} = \sum_{n=0}^{\mathrm{N}-1} \sum_{t=0}^{\mathrm{T}} \left\| \nabla \mathrm{L}^n \cdot \left(\mathbf{p}^n - \tilde{\mathbf{n}}_t \right) \right\|_1. \tag{12}$$

This transforms the minimization (10) in to a max/min problem. When including a residual layer in our formulation, we also introduce a dual variable $\mathbf{q}_t$ for each residual image b_t in order to solve the total variation penalty. The L^1 penalty is solved using a 'shrinkage'-based [2] L^1 update on b_t.

A multiscale algorithm is used to ensure correct deformations are found even in cases of large movements of small structures such as vessels. At each scale level $\bar{\mathrm{I}}_t$ is computed from the appropriately downsampled version of I_t, and $\tilde{\mathbf{n}}_t$ is then calculated from $\bar{\mathrm{I}}_t$ via equation (5). The algorithm iteratively updates each $\{\mathrm{L}^n\}_n$, $\{\mathbf{p}^n\}_n$, and $\{\mathbf{v}_t^n\}_{n,t}$ (and $\{\mathbf{q}_t\}_t$ and $\{\mathrm{b}_t\}_t$ if using residual layers), taking appropriate gradient descent steps on the primal variables, and gradient ascent steps on the dual variables followed by reprojection onto their constraints, repeating until convergence.

4 Results

We first present results on a synthetic dataset meant to approximate a fluoroscopic image sequence of a contrast-enhanced vessel structure (see the first column of Figure 2). This is included to highlight characteristics of solutions this algorithm produces. Results are then presented on the clinical angiography dataset the synthetic data was meant to approximate (see first column of Figure 3). Finally results are presented for the DSA application using the deformation with residual model on a clinical dataset (Figure 5 (a) and (b)). Values for the λ constants were determined experimentally. The synthetic and clinical data frames are 256×256 and 512×512 pixels, respectively. Optimization was run on a NVIDIA Tesla C1060. Each gradient descent iteration on the clinical data at the finest scale level takes approximately 310 ms. Results presented were run for 4000 iterations at each scale level to ensure convergence.

The synthetic data is generated from four layers; a static 'rib', a slowly moving 'diaphragm', and two 'vessel' layers representing a vessel tree undergoing a nonlinear deformation with branches overlapping from the imaged viewpoint. Although the data is generated from four layers, experimental results reveal that three layers are sufficient to capture the independent motion of different structures, and additional layers do not improve the result.

The results in Figure 2 are a representative example of the solutions found by our algorithm. While the results are not correct from a segmentation perspective, in a given region any objects displaying contradictory motion are separated. Note that the diaphragm object has been removed from the upper portion of L^0, allowing the vessel to cross the diaphragm boundary. While we employ a total variation based regularization on the layers in order to help separate objects and estimate coincident motions, we do not typically want the highly denoised results shown in columns (b) and (c). In fact, we see areas at the tips of the vessel structure where fine detail has been obliterated by the de-noising properties of the estimation. If the set of estimated deformations are consistent with the motion of the imaged objects, it is possible to re-estimate only the layer intensities given these fixed deformations. In this case very little regularization

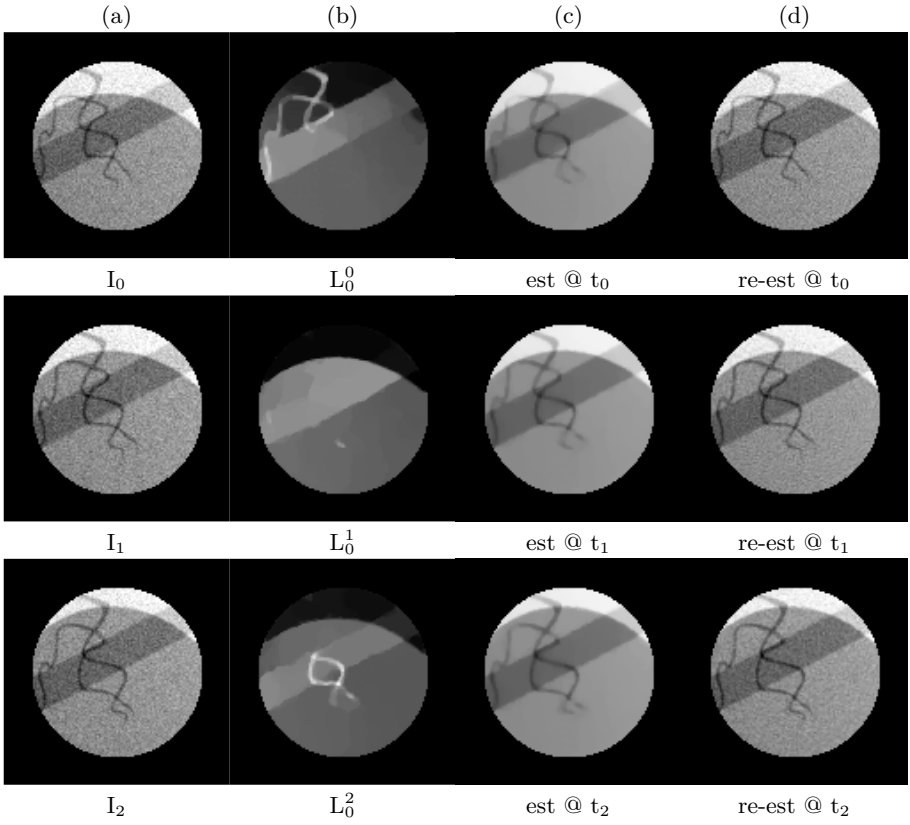

Fig. 2. Results of layer and deformation estimation on synthetic dataset. Column (a) shows the input frames. Column (b) shows each layer at t_0. Column (c) shows the estimated frame at each timepoint. Column (d) show the reconstruction with re-estimated layers.

is necessary in the intensity estimation, and the resulting layers preserve the noise texture and much of the fine detail of the original images, as shown in column (d). By temporal interpolation we can use our layer model to generate intermediate frames, as shown in Row (a) of Figure 4 where the re-estimated layers have been used.

Figure 3 shows results on a clinical angiography dataset. A Three-layer model has again been chosen based on experimental results. Note the separation of the most prominent vessel from the diaphragm. Once again, we show initial denoised layers as well as re-estimated layers. Row (b) of Figure 4 shows an interpolated frame between times t_0 and t_1. Note the crossing of vessels in the upper portion of the image.

Figure 5 shows the results of using a residual layer and two deforming layers to estimate the motion and contrast between two frames of an angiographic

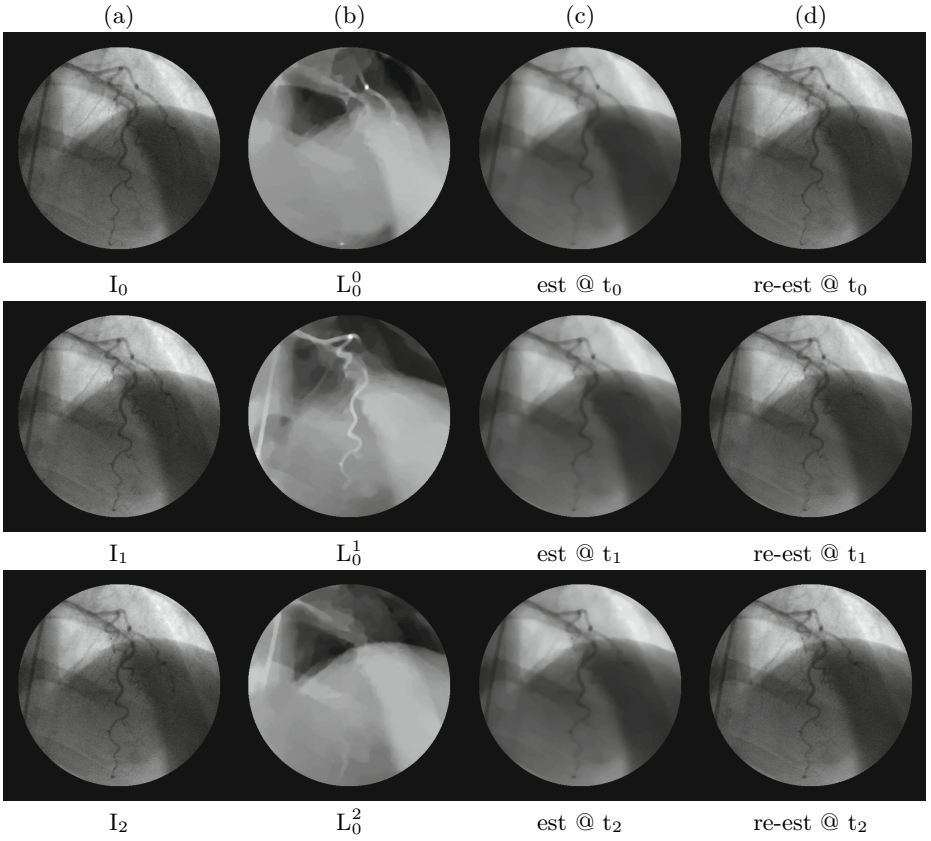

Fig. 3. Results of layer and deformation estimation on clinical dataset. Column (a) shows the input frames. Column (b) shows each layer at t_0. Column (c) shows the estimated frame at each timepoint. Column (d) show the reconstruction with re-estimated layers.

sequence to diagnose stent placement for treatment of an abdominal aortic aneurysm, and compares against static subtraction and elastic registration.

In this paper we proposed a model of dynamic X-ray images that consists of a set of superimposed, smoothly deforming layers, that combine additively to describe the spatio-temporal behavior of projected 3D motion. We also described an estimate procedure and demonstrated the feasibility of this technique for motion modeling in fluoroscopic imaging. We observed that this process should not be considered as a conventional segmentation problem; the estimated layers need not be a segmentation of physical objects in the scene in order to accurately represent the observed sequence. The estimation of such a layered model is inherently difficult. It is under constrained and there are many feasible solutions, even for simple examples. To address this challenge we propose regularizing the problem with a set of both general and application-specific penalties or models. For this we have introduced a novel penalty of gradients in the layers which forces layer edges to align with those in the observations. Also, we have introduced a

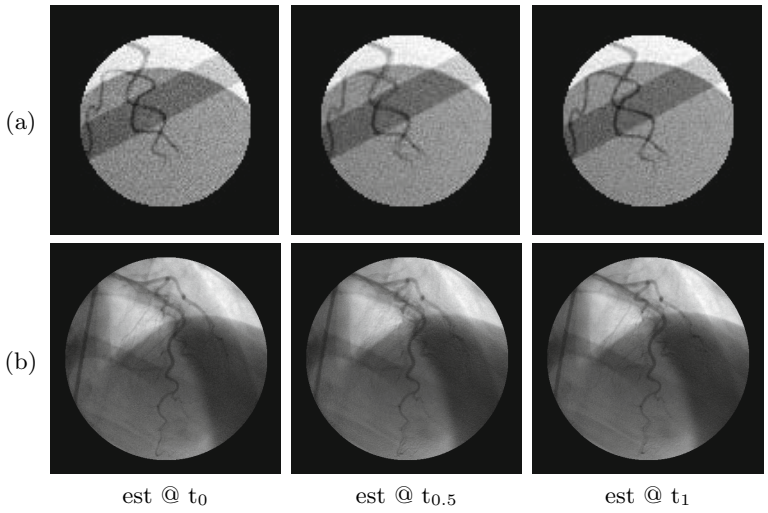

Fig. 4. Frame interpolation using estimated layers and deformations. Row (a) shows results on synthetic data from Figure 2, and row (b) on clinical data from Figure 3. The center frame is estimated between t_0 and t_1.

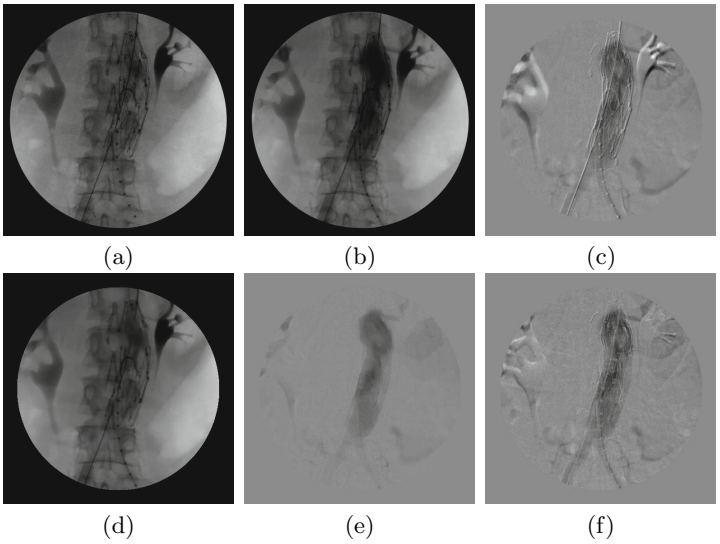

Fig. 5. DSA Results. (a) and (b) show initial (mask) frame and current frame, respectively. (c) show static subtraction. (d) is the estimated current frame (from layered registration) without residual layer. (e) shows DSA using frame (d). (f) shows subtraction using simple elastic registration between frames.

layer with a dynamic contrast model, and shown its effectiveness in correcting the motion artifacts in digital subtraction angiography.

The authors would like to acknowledge the support from GE Healthcare, which made this explorational research possible.

References

1. Auvray, V., Bouthemy, P., Liénard, J.: Joint motion estimation and layer segmentation in transparent image sequenc: application to noise reduction in x-ray image sequences. EURASIP J. Adv. Signal Process. 2009, 19:1–19:21 (2009)
2. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM Journal on Imaging Sciences 2(1), 183–202 (2009)
3. Beg, M.F., Miller, M.I., Trouv, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. IJCV 61(2), 139–157 (2005)
4. Chen, Y., Chang, T., Zhou, C., Fang, T.: Gradient domain layer separation under independent motion. In: ICCV, pp. 694–701. IEEE (2009)
5. Christensen, G.E., Rabbitt, R.D., Miller, M.I.: Deformable templates using large deformation kinematics. Transactions on Image Processing 5, 1435–1447 (1996)
6. Close, R., et al.: Accuracy assessment of layer decomposition using simulated angiographic image sequences. Transactions on Medical Imaging 20(10), 990–998 (2001)
7. Darrell, T., Simonecelli, E.: Nulling filters and the separation of transparent motions. In: Computer Vision and Pattern Recognition, pp. 738–739. IEEE (1993)
8. Jasper, J.: Role of digital subtraction fluoroscopic imaging in detecting intravascular injections. Pain Physician 6(3), 369–372 (2003)
9. Osher, S., et al.: An iterative regularization method for total variation-based image restoration. Multiscale Modeling & Simulation 4(2), 460–489 (2005)
10. Pingault, M., et al.: A robust multiscale b-spline function decomposition for estimating motion transparency. Tran. on Image Processing 12(11), 1416–1426 (2003)
11. Ramírez-Manzanares, A., Rivera, M., Kornprobst, P., Lauze, F.: A variational approach for multi-valued velocity field estimation in transparent sequences. In: Sgallari, F., Murli, A., Paragios, N. (eds.) SSVM 2007. LNCS, vol. 4485, pp. 227–238. Springer, Heidelberg (2007)
12. Rueckert, D., Aljabar, P., Heckemann, R.A., Hajnal, J.V., Hammers, A.: Diffeomorphic registration using B-splines. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006, Part II. LNCS, vol. 4191, pp. 702–709. Springer, Heidelberg (2006)
13. Sarel, B., Irani, M.: Separating transparent layers through layer information exchange. In: Pajdla, T., Matas, J. (eds.) ECCV 2004. LNCS, vol. 3024, pp. 328–341. Springer, Heidelberg (2004)
14. Shizawa, M., Mase, K.: Simultaneous multiple optical flow estimation. In: International Conference on Pattern Recognition, vol. 1, pp. 274–278. IEEE (1990)
15. Stuke, I., et al.: Estimation of multiple motions using block matching and markov random fields. In: Proceedings of SPIE, vol. 5308, pp. 486–496 (2004)
16. Szeliski, R., Avidan, S., Anandan, P.: Layer extraction from multiple images containing reflections and transparency. In: CVPR, vol. 1, pp. 246–253. IEEE (2000)
17. Toro, J., Owens, F., Medina, R.: Using known motion fields for image separation in transparency. Pattern Recognition Letters 24(1), 597–605 (2003)
18. Wang, J., Adelson, E.: Representing moving images with layers. Transactions on Image Processing 3(5), 625–638 (1994)
19. Xiao, J., Shah, M.: Motion layer extraction in the presence of occlusion using graph cuts. Pattern Analysis and Machine Intelligence 27(10), 1644–1659 (2005)
20. Zhu, M., Chan, T.: An efficient primal-dual hybrid gradient algorithm for total variation image restoration. Tech. rep., CAM UCLA Tech. Rep. 08-34 (2008)