A Robust Probabilistic Model for Motion Layer Separation in X-ray Fluoroscopy

Peter Fischer[1(✉)], Thomas Pohl[2], Thomas Köhler[1], Andreas Maier[1], and Joachim Hornegger[1]

[1] Pattern Recognition Lab and Erlangen Graduate School in Advanced Optical Technologies (SAOT), FAU Erlangen-Nürnberg, Erlangen, Germany
peter.fischer@fau.de
[2] Siemens Healthcare, Forchheim, Germany

Abstract. Fluoroscopic images are characterized by a transparent projection of 3-D structures from all depths to 2-D. Differently moving structures, for example due to breathing and heartbeat, can be described approximately using independently moving 2-D layers. Separating the fluoroscopic images into the motion layers is desirable to facilitate interpretation and diagnosis. Given the motion of each layer, it is state of the art to compute the layer separation by minimizing a least-squares objective function. However, due to high noise levels and inaccurate motion estimates, the results are not satisfactory in X-ray images.

In this work, we propose a probabilistic model for motion layer separation. In this model, we analyze various data terms and regularization terms theoretically and experimentally. We show that a robust penalty function is required in the data term to deal with noise and shortcomings of the image formation model. For the regularization term, we propose to enforce smoothness of the layers using bilateral total variation. On synthetic data, the mean squared error between the estimated layers and the ground truth is improved by 18 % compared to the state of the art. In addition, we show qualitative improvements on real X-ray data.

1 Introduction

Minimally-invasive interventions are often guided by fluoroscopic X-ray imaging. X-ray imaging offers good temporal and spatial resolution and high contrast of interventional devices and bones. However, the soft-tissue contrast is low and the patient and the physician are exposed to ionizing radiation. In addition to the low soft-tissue contrast, the loss of 3-D information due to the transparent projection to 2-D complicates interpretation of the fluoroscopic images. To simplify the analysis, fluoroscopic images can be decomposed into independently moving layers. Each layer contains similarly moving structures, leading to the separation of background structures like bones from moving soft-tissue like the heart or the liver. In addition, other post-processing algorithms like segmentation or frame interpolation can benefit from the motion layer separation. Another clinically relevant post-processing application is digital subtraction angiography (DSA). DSA is performed by subtracting a reference frame. However, if there is too

S. Ourselin et al. (Eds.): IPMI 2015, LNCS 9123, pp. 288–299, 2015.
DOI: 10.1007/978-3-319-19992-4_22

much motion, the selection of an appropriate reference frame is difficult. In particular for coronary arteries, complex respiratory and cardiac motion complicate traditional DSA and make motion layer separation a good alternative [17].

In the literature, multiple approaches to layer separation have been investigated. Layer separation is sometimes combined with motion estimation, but we limit ourselves to layer separation in this work. Close et al. estimate rigid motion of each layer in a region of interest [3]. The layers are computed by stabilizing the sequence w.r.t. the layer motion and subsequent averaging. Preston et al. jointly estimate motions and layers using a coarse-to-fine variational framework [10], but the results are not physically meaningful motions or layers. In [14], an iterative scheme for motion and layer estimation is used. For layer separation, a constrained least-squares optimization problem is solved. Weiss estimates a static layer from a transparent image sequence exploiting the sparsity of images in the gradient domain [16]. Zhang et al. assume the motions as given and solve a constrained least-squares problem for estimating the layers [17].

So far, regularization has rarely been applied to aid layer separation. Exception are [10], where a layer gradient penalty is introduced, and [16], where the objective function implicitly favors smooth layers. In other areas of image processing, regularization is widely used. Inverse problems in image processing, often formulated to minimize an energy function, benefit from regularization, for example denoising [11], image registration [7], and super-resolution [4]. Total variation is a popular, edge-preserving regularization that was originally introduced for denoising [11]. Super resolution is conceptually similar to layer separation and is often formulated as a probabilistic model with robust regularization, e.g., bilateral total variation [4].

In this paper, we introduce a novel probabilistic model for layer separation in transparent image sequences. As likelihood and prior in the Bayesian model, we propose to use a robust data term and edge-preserving regularization. In particular, a non-convex data term is used that is robust w.r.t. noise, errors in the image formation model, and errors in the motion estimates. Furthermore, we theoretically analyze different spatial regularization terms for layer separation. Inference in the Bayesian model leads to maximum a posteriori estimation of the layers, as opposed to the previously used maximum likelihood. In the experiments, we extensively compare possible data and regularization terms. We show that layer separation can benefit from our robust approach.

2 Materials and Methods

2.1 Image Formation Model

In this paper, we are interested in separating X-ray images $I^t \in \mathbb{R}^{W \times H}$, $t \in \{1, \dots, T\}$ into different motion layers L_l, where each layer may undergo independent non-rigid 2-D motion $\boldsymbol{v}_l^t$. A motion layer can roughly be assigned to each source of motion, e.g., breathing, heartbeat, and background.

In our spatially discrete formulation, the images and layers are vectorized to $\boldsymbol{I}^t, \boldsymbol{L}_l \in \mathbb{R}^{WH}$. The transformation of a layer by its motion and subsequent interpolation is modeled in the system matrix $\boldsymbol{W}_l^t \in \mathbb{R}^{WH \times WH}$ [14]

$$\boldsymbol{I}^t = \sum_{l=1}^{N} \boldsymbol{W}_l^t \boldsymbol{L}_l + \boldsymbol{\epsilon}^t, \tag{1}$$

where we introduce $\boldsymbol{\epsilon}^t$ to account for model errors and observation noise. N is the number of layers in the image sequence. This model is justified by the log-linearity of Lambert-Beer's law applied to X-ray attenuation. In $\boldsymbol{W}_l^t$, we use bilinear interpolation, but the method generalizes to other interpolation or point spread functions. Boundary treatment for image pixels moving outside of the spatial support of the layers is to take the nearest layer pixel. Alternatively, the layer support can be increased to cover all motions in the current sequence [15]. For all images and layers, the joint forward model is used

$$\boldsymbol{I} = \boldsymbol{W}\boldsymbol{L} + \boldsymbol{\epsilon}, \tag{2}$$

where $\boldsymbol{I} = \left(\boldsymbol{I}^{1\intercal}, \ldots, \boldsymbol{I}^{T\intercal}\right)^\intercal$, $\boldsymbol{L} = \left(\boldsymbol{L}_1{}^\intercal, \ldots, \boldsymbol{L}_N{}^\intercal\right)^\intercal$, and $\boldsymbol{\epsilon} = \left(\boldsymbol{\epsilon}^{1\intercal}, \ldots, \boldsymbol{\epsilon}^{T\intercal}\right)^\intercal$. The system matrix $\boldsymbol{W} = \left(\boldsymbol{W}^{1\intercal}, \ldots, \boldsymbol{W}^{T\intercal}\right)^\intercal$ is composed of matrices $\boldsymbol{W}^t = \left(\boldsymbol{W}_1^t, \ldots, \boldsymbol{W}_N^t\right)$ to transform all layers to a certain point in time.

2.2 Probabilistic Approach to Layer Separation

The goal of layer separation is to find the layers $\boldsymbol{L}$ given the images $\boldsymbol{I}$ and the motions encoded in $\boldsymbol{W}$. From a Bayesian point of view, the observed images $\boldsymbol{I}$, the noise $\boldsymbol{\epsilon}$, and the layers $\boldsymbol{L}$ are random variables. Assuming conditionally independent observed images, the posterior probability of the layers given the images $p(\boldsymbol{L}|\boldsymbol{I})$ is given by

$$p(\boldsymbol{L}|\boldsymbol{I}) = \frac{p(\boldsymbol{L})\, p(\boldsymbol{I}|\boldsymbol{L})}{p(\boldsymbol{I})} = \frac{p(\boldsymbol{L}) \prod_{t=1}^{T} p(\boldsymbol{I}_t|\boldsymbol{L})}{p(\boldsymbol{I})}, \tag{3}$$

with the prior probability for the layers $p(\boldsymbol{L})$ and the likelihood $p(\boldsymbol{I}_t|\boldsymbol{L})$ for each image given the layers. Common priors in image processing are defined on local neighborhoods, such that Eq. (3) corresponds to a Markov random field. The maximum a posterior (MAP) estimate

$$\hat{\boldsymbol{L}} = \underset{\boldsymbol{L}}{\operatorname{argmax}}\, p(\boldsymbol{L}) \prod_{t=1}^{T} p(\boldsymbol{I}_t|\boldsymbol{L}) \tag{4}$$

yields the statistically optimal layers for the given model and input images. In previous work, no probabilistic motivation [3] or maximum likelihood (ML) estimation was often used [14,17], implicitly assuming a uniform prior $p(\boldsymbol{L})$.

By applying the logarithm and negating, the probabilistic formulation can be equivalently regarded as an energy. Assuming positive values, it is possible to write prior $p(\boldsymbol{L})$ and likelihood $p(\boldsymbol{I}_t|\boldsymbol{L})$ as $p(\boldsymbol{L}) = \frac{1}{Z_R} \exp\left(-\lambda R(L)\right)$ and $p(\boldsymbol{I}_t|\boldsymbol{L}) = \frac{1}{Z_D} \exp\left(-D(\boldsymbol{I}_t, \boldsymbol{L})\right)$, where Z_R, Z_D are partition functions to normalize the probabilities. Consequently, MAP inference as in Eq. (4) turns into energy minimization

$$\hat{\boldsymbol{L}} = \underset{\boldsymbol{L}}{\operatorname{argmin}}\, \lambda R\left(\boldsymbol{L}\right) + \sum_{t=1}^{T} D\left(\boldsymbol{I}_t, \boldsymbol{L}\right), \tag{5}$$

where $D\left(\boldsymbol{I}_t, \boldsymbol{L}\right)$ is the data term, $R\left(\boldsymbol{L}\right)$ the regularization, and $\lambda \in \mathbb{R}_0^+$ the regularization weight. In the following sections, we concretize $D\left(\boldsymbol{I}_t, \boldsymbol{L}\right)$ and $R\left(\boldsymbol{L}\right)$.

2.3 Data Term

The data term describes how deviations from the image formation model are penalized. From a probabilistic point of view, it corresponds to an assumption on the observation noise $\boldsymbol{\epsilon}^t$. The classic choice of a least-squares data term

$$D_{L_2}\left(\boldsymbol{I}_t, \boldsymbol{L}\right) = \left\|\boldsymbol{I}^t - \boldsymbol{W}^t \boldsymbol{L}\right\|_2^2 \tag{6}$$

corresponds to a Gaussian noise model, which has been used in most of the prior work [10,14,17] and is a fitting model for images with good photon statistics [9]. This model is easy to optimize by solving a sparse linear system of equations. Its major drawback is the sensitivity to outliers, i.e., a few erroneous measurements lead to artifacts in the estimated layers. However, outliers are very common in X-ray layer separation, for example due to errors in motion estimation, which is challenging in X-ray without knowing the layers (Sect. 2.6). Another important source of outliers is the simplified image formation model (Sect. 2.1). Many effects occurring in X-ray images are not captured by this model, e.g., foreshortening and out-of-plane motion.

The least absolute deviation corresponds to a Laplacian noise model

$$D_{L_1}\left(\boldsymbol{I}_t, \boldsymbol{L}\right) = \left\|\boldsymbol{I}^t - \boldsymbol{W}^t \boldsymbol{L}\right\|_1, \tag{7}$$

which is more robust to outliers and still a convex function. In contrast to Eq. (6), it is not smooth due to the non-differentiability at 0. Therefore, a smooth approximation to the L_1-norm is helpful for gradient-based optimization schemes, e.g., the Charbonnier function $\|z\|_1 \approx \phi(z) = \sqrt{z^2 + \tau^2} - \tau$, for $\tau > 0$ [13].

A non-convex data term can be derived using a generalization of the Charbonnier function $\phi_\alpha(z) = \left(z^2 + \tau^2\right)^\alpha - \tau^{2\alpha}$ [13]. $\phi(z)$ is equivalent to $\phi_{0.5}(z)$ and z^2, as used in D_{L_2}, is equivalent to $\phi_1(z)$. Then, the general data term is

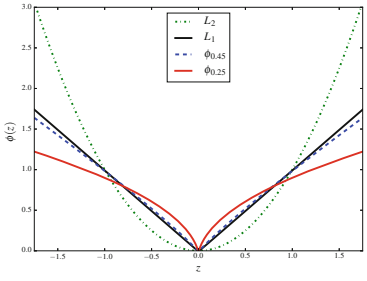


Fig. 1. Behavior of different penalty functions (best viewed in color) (Color figure online).

$$D_{\text{Charb.}}(\boldsymbol{I}_t, \boldsymbol{L}) = \sum_{k=1}^{WH} \phi_\alpha \left(\left[\boldsymbol{I}^t - \boldsymbol{W}^t \boldsymbol{L} \right]_k \right). \tag{8}$$

$[\boldsymbol{x}]_k$ extracts the k-th component of $\boldsymbol{x}$. Using the generalized Charbonnier function, the value of α can be tuned to fit the statistics of the observation noise. τ is only required for numerical reasons and set to 0.01. The penalty functions are visualized in Fig. 1. It is evident that L_1 and L_2 are convex penalties, and that large deviations are penalized less by $\phi_\alpha(z)$ with smaller values of α.

2.4 Regularization Term

Common priors in image processing favor smoothness of the images. The most basic prior is based on Tikhonov regularization and penalizes high gradients

$$R_{L_2}(\boldsymbol{L}) = \sum_{l=1}^{N} \|\boldsymbol{\nabla} \boldsymbol{L}_l\|_2^2, \tag{9}$$

where $\boldsymbol{\nabla}$ is a matrix computing the spatial derivatives for each layer. As image gradients in natural images are heavy-tailed, Eq. (9) leads to oversmoothed images. For layer separation, the L_2 regularization term is particularly counterproductive. Assume a certain gradient at an image location has to be represented somehow by the layers. The L_2-norm gives the lowest penalty if all layers contribute equally to the image gradient. However, this corresponds to a separation into two equal layers.

To better preserve edges in the layers, the total variation (TV) regularization

$$R_{\text{TV}}(\boldsymbol{L}) = \sum_{l=1}^{N} \|\boldsymbol{\nabla} \boldsymbol{L}_l\|_1 \tag{10}$$

is useful [11], which again leads to a convex optimization problem. In contrast to the L_2-norm, the L_1-norm does neither hinder nor enforce layer separation. Sparse solutions, i.e., an image gradient is represented by a single layer, have the same energy as equal gradients in all layers.

In super-resolution, bilateral total variation (BTV) is a popular regularizer [4]. It generalizes TV regularization to include a wider spatial support of $2P+1$ pixels in each dimension, can lead to better edge preservation, and is convex. BTV is defined as

$$R_{\text{BTV}}(\boldsymbol{L}) = \sum_{l=1}^{N} \sum_{m=-P}^{P} \sum_{n=-P}^{P} \beta^{|m|+|n|} \left\| \boldsymbol{L}_l - \boldsymbol{S}_v^m \boldsymbol{S}_h^n \boldsymbol{L}_l \right\|_1, \tag{11}$$

where $0 \leq \beta \leq 1$ is a spatial weighting factor and $\boldsymbol{S}_v^m$ ($\boldsymbol{S}_h^n$) corresponds to vertical (horizontal) shifts of the layer $\boldsymbol{L}_l$ by m (n) pixels.

All the aforementioned regularization terms are spatially independent. Additional information for layer regularization can be gained from the images, i.e.,

the regularization term can be generalized to $R(\boldsymbol{I}, \boldsymbol{L})$. For example, the image gradient offers information about the desired position and direction of the layer gradients. Preston et al. use this to define the regularization term

$$R_{\text{Pres.}}(\boldsymbol{I}, \boldsymbol{L}) = \sum_{t=1}^{T} \sum_{l=1}^{N} \sum_{k=1}^{WH} \left(\left\| \nabla \left[\boldsymbol{W}_l^t \boldsymbol{L}_l \right]_k \right\|_2 - \left(\nabla \left[\boldsymbol{W}_l^t \boldsymbol{L}_l \right]_k \right)^\intercal \boldsymbol{n}_k^t \right) \tag{12}$$

to remove the penalty if the layer gradient is in the same direction as the image gradient [10], which is computed using ∇. The image gradient is thresholded

$$\boldsymbol{n}_k^t = \begin{cases} \frac{\nabla [\boldsymbol{I}^t]_k}{\left\| \nabla [\boldsymbol{I}^t]_k \right\|_2} & \text{if } \left\| \nabla [\boldsymbol{I}^t]_k \right\|_2 > \delta \\ 0 & \text{else} \end{cases}, \tag{13}$$

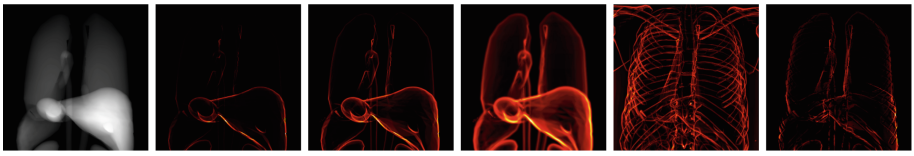
such that small gradients caused by noise do not influence the regularization. Other than that, the magnitude of the image gradient is not important. Consequently, at a single position the gradients of multiple layers can point in the same direction without increasing the energy. An advantage of this regularization term is that layer gradients with magnitude 0 always lead to 0 energy. In this sense, it is a TV regularization that is switched off if the layer gradient points in the same direction as the image gradient.

Inspired by [16], we define another regularization term that uses image gradient information. Assuming sparsity of layer gradients, it is likely that an observed image gradient comes from a single layer. Therefore, the magnitude of the layer gradient should be the same as the image gradient, as in the regularization term

$$R_{\text{Weiss}}(\boldsymbol{I}, \boldsymbol{L}) = \sum_{t=1}^{T} \sum_{l=1}^{N} \left\| \boldsymbol{\nabla} \boldsymbol{W}_l^t \boldsymbol{L}_l - \boldsymbol{\nabla} \boldsymbol{I}^t \right\|_1. \tag{14}$$

For the layer that explains the corresponding image gradient, 0 energy is incurred. However, the remaining layers all create an energy of $\|\boldsymbol{\nabla} \boldsymbol{I}\|_1$. The minimum value of this regularization term is not attained for a layer without gradients, as in TV or L_2-regularization. Instead, it is attained when the layer gradient is equal to the median of the image gradients over time [16], where the layer motion is compensated in the image. As the image gradient is sparse and the layer motions are independent, the median is often close to 0. For the previously described regularization terms, the L_1-norm can be replaced by the generalized Charbonnier penalty ϕ_α to enforce sparsity even more.

Figure 2 shows the effects of the different regularizers. R_{L_2} focuses on large gradients in the layer, leading to oversmoothing. R_{TV} is more robust, i.e., the relative penalty on large gradients is reduced compared to R_{L_2}. R_{BTV} is a smoothed version of R_{TV}, because the spatial shifts cover a wider area. R_{Weiss} has no penalty for the gradients of Fig. 2a. However, a penalty must be paid for image gradients that are not explained by this layer, which could lead to worse separation. $R_{\text{Pres.}}$ is identical to R_{TV}, except that the TV penalty is switched off if there is an image gradient. Due to their dependence on the layer motions, $R_{\text{Pres.}}$ and R_{Weiss} have artifacts for inexact motion estimates.



(a) Layer (b) R_{L_2} (c) R_{TV} (d) R_{BTV} (e) R_{Weiss} (f) $R_{\mathrm{Pres.}}$

Fig. 2. Penalty of the ground truth layer (a) for different regularization terms. Dark corresponds to low and bright to high penalty (best viewed in color) (Color figure online).

2.5 Numerical Optimization

The layer estimation problem is processed in a coarse-to-fine pyramid. This ensures that an approximate solution is found quickly on low-resolution images and greatly reduces computation time. In contrast to [17], we estimate all layers on all resolutions. Thus, the coarse-to-fine pyramid is mainly used for speeding up the convergence. In addition, it helps to avoid local minima for the non-convex energy terms involving the generalized Charbonnier penalty.

The optimization method on each level is limited-memory Broyden-Fletcher-Goldfarb-Shanno with bound constraints (L-BFGS-B). This method requires smooth gradients, so all L_1-norms are approximated by the Charbonnier function. For some combinations of data terms and regularization terms, specialized solvers exist that are much faster. For example, a L_2 data term with L_2 regularization can be solved in closed form using the pseudo-inverse, and L_2 data term with TV regularization can be optimized using a split-Bregman solver [5]. However, as we prefer generality over runtime in this work, we always use L-BFGS-B. Optimization is run until convergence on each level of the pyramid. Boundary conditions are enforced for the layers that can be derived from the additive image formation model, e.g., non-negativity [14].

2.6 Sources of Layer Motions

An important prerequisite for our approach to layer extraction from fluoroscopic images is the motion of each layer. By itself, this is a challenging problem. However, there are several applications where this is feasible.

The first application is joint layer and motion estimation. The layers and their motions are assumed to be unknown and jointly estimated from a fluoroscopic sequence. This can be optimized using an alternation scheme, with the two subtasks of motion estimation given the layers and layer estimation given the motions. For the latter, the proposed method of this paper is applicable. In particular for this application, it can not be presupposed that the given layer motions are accurate. Consequently, robust methods are mandatory.

The second application is post-processing of fluoroscopic sequences. Separated motion layers are useful for improved interpretation of the image content. Dense motion of the background can be computed using robust parametric

registration methods [1]. More complex motion patterns require more effort. A possibility is tracking of control points, devices [6] or anatomical curves [2]. To get a dense motion field from the tracking results, interpolation methods like thin-plate splines (TPS) can be used. In post-processing, there is enough time to accurately perform these tasks.

2.7 Experiments

Synthetic data is used for quantitative analysis and real X-ray data for qualitative results. The synthetic data is created by independently projecting different organs of the XCAT phantom to 2-D [8,12]. The resulting layers are transformed using 2-D motion fields. The 2-D motion is created by TPS interpolation of manual control point motions. In total, we use two datasets, each with $N = 2$ layers and $T = 10$ images of size $W = H = 250$ and with a dynamic range of $[0, 1]$. On the synthetic data, we simulate different types of errors. First, we add measurement noise in the form of Gaussian and Laplacian noise to the image intensities ($\sigma_{\text{Gauss}} = \sigma_{\text{Laplace}} = 0.01$). Second, we simulate registration and model errors by smoothing the ground truth motion field and randomly disturbing it by adding Gaussian noise to the motion vectors ($\sigma_{\text{motion}} = 1.5$ px). In addition, two images of the sequence are translated randomly ($\sigma_{\text{trans}} = 4.0$ px). For each of the datasets, 10 instances with random errors are created. As the error measure for ground truth comparison, we use a modified version of the mean squared error (MSE). As a uniform intensity offset can not be determined using layer separation, the means of the layers are subtracted before computing the MSE.

The real X-ray data consists of a sequence of 10 images of $W = 670$, $H = 1000$. The required layer motions are extracted from the images manually. In each image, the motion of ∼25 control points is annotated. The motion of these control points is converted to a dense motion field using TPS interpolation.

To find the parameters for each method, we perform grid search. For $D_{\text{Cha.}}$, α is searched in $\{0.25, 0.3, 0.35, 0.4, 0.45, 0.5\}$. For the regularizers, λ is searched in $\{0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1.0, 5\}$ while α is fixed. The threshold for the gradient magnitude δ in $R_{\text{Pres.}}$ is set to 0.01 [10]. The parameters of R_{BTV} are searched in $\beta = \{0.5, 0.7, 0.9\}$ and $P = \{3, 5\}$. For each experiment, 10 different random instances with the same error and noise type are used as training data. We use forward differences to approximate spatial derivatives $\boldsymbol{\nabla}$. The coarse-to-fine pyramid is implemented with a downsampling factor of 0.5 and 6 levels.

3 Results

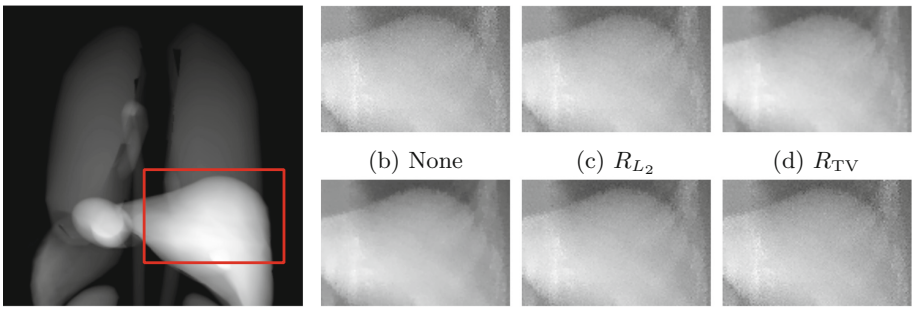
3.1 Analysis of Data Terms

To analyze the behavior of different data terms, we apply ML estimation without regularization for each of them. For $D_{\text{Cha.}}$, the grid search yields $\alpha = 0.25$, $\beta = 0.5$, and $P = 5$. The MSE decreases with increasing robustness of the data

term, see Table 1. Qualitatively, the errors in D_{L_2} correspond to artifacts at positions of wrong motion. Note that D_{L_2} is the common data term in the state of the art [10,14,17]. Using robust data terms, these artifacts in the layers are removed.

Table 1. MSE ($\cdot 10^{-3}$) for different data terms on synthetic test data (mean $\pm$ std). Grid search determined $\alpha = 0.25, \beta = 0.5, P = 5$ for $D_{\text{Cha.}}$.

	D_{L_2}	D_{L_1}	$D_{\text{Cha.}}$
MSE	8.9 ± 5.6	8.5 ± 5.5	8.3 ± 5.4



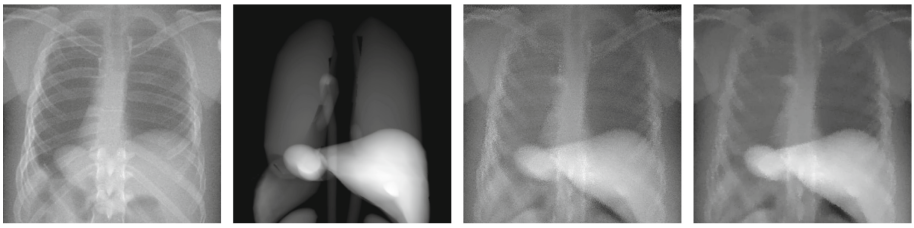
(a) Ground truth (b) None (c) R_{L_2} (d) R_{TV} (e) R_{BTV} (f) $R_{\text{Pres.}}$ (g) R_{Weiss}

Fig. 3. View of a region of interest (red) of a layer extracted using different regularization terms. $D_{\text{Cha.}}$ is used in all cases (Colour figure online).

3.2 Analysis of Regularization Terms

We investigate all combinations of $D_{\text{Cha.}}$ with $\alpha = 0.25$ and the introduced regularization terms. The respective regularization weights are listed in Table 2, together with the experimental results. All regularization methods improve the MSE compared to ML estimation. The image-driven regularizers $R_{\text{Pres.}}$ and R_{Weiss} have only a small effect, as training assigned low weights. This means that using higher weights for these regularizers deteriorates the results. With an MSE of $7.3 \cdot 10^{-3}$, R_{BTV} is the best regularizer in our experiments. R_{TV} is second, as it also preserves edges. Since R_{BTV} is a generalization of R_{TV}, it is more flexible. R_{BTV} has the highest runtime as multiple finite differences are evaluated. R_{Weiss} and $R_{Pres.}$ are slow as well, since they must be computed for each point in time.

A qualitative impression of the effect of the regularization is given in Fig. 3 for a region of interest. The robust data term already removed most of the outliers. The main difference between the regularization terms the denoising performance, including edge preservation. In Fig. 4, we highlight the the difference between the state of the art and the proposed robust probabilistic model. D_{L_2} has blurred edges and a high-noise level, while our method is closer to the ground truth.



(a) Input (b) Ground truth (c) D_{L_2} (d) $D_{\text{Cha.}} + R_{\text{BTV}}$

Fig. 4. An image of the input sequence (a), a ground truth layer (b), and the corresponding layer extracted with the state of the art (c) and our method (d).

Table 2. Value of regularization weight λ found using grid search on training data, MSE ($\cdot 10^{-3}$), and runtime [s] on test data (mean $\pm$ std).

	-	R_{L_2}	R_{TV}	R_{BTV}	$R_{\text{Pres.}}$	R_{Weiss}
λ	-	1.0	0.5	0.1	0.05	0.0001
MSE	8.3 ± 5.4	8.1 ± 5.5	7.8 ± 5.3	7.3 ± 5.7	8.1 ± 5.0	8.3 ± 5.4
Runtime	55.8 ± 9.2	50.0 ± 8.1	42.5 ± 9.6	203 ± 114	116 ± 18.4	117 ± 25.5

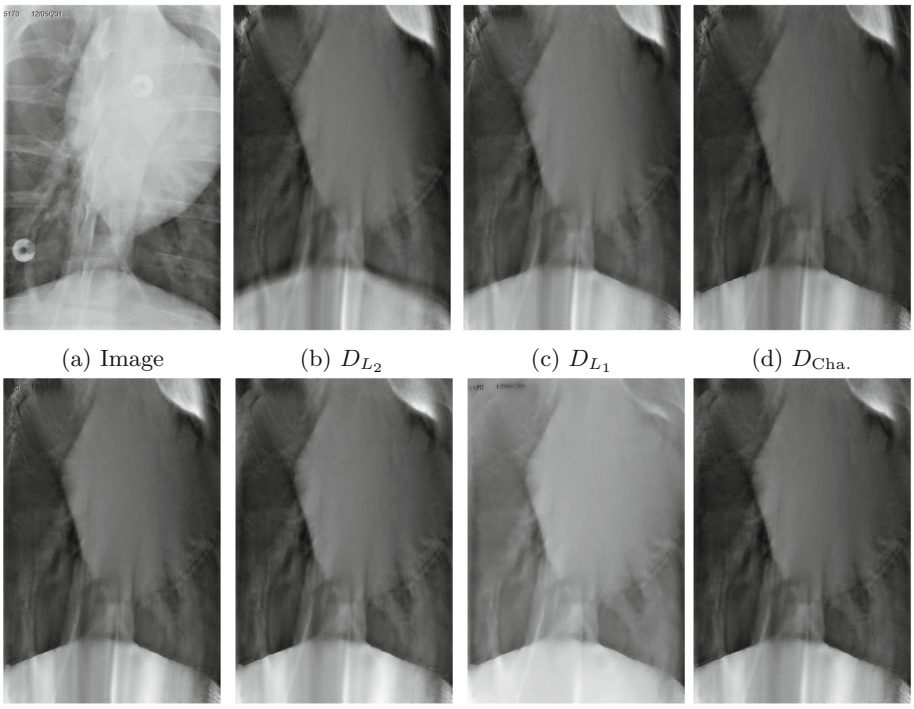
3.3 Real X-ray Data

For the real X-ray data, the same parameters as in Table 2 are used. The experiments on real data have to deal with many sources of error. The manual labeled motion is inaccurate, because it is only based on a few sparse control points. In addition, the layered image formation model is not fulfilled here. A reconstructed layer from an X-ray sequence containing soft tissue motion of the heart, lung and diaphragm is shown in Fig. 5. Ribs, spine, and skin markers are static and should be removed from the shown layer. The state-of-the-art D_{L_2} data term without regularization creates artifacts and smooths edges (bottom left). D_{L_1} and $D_{\text{Cha.}}$ are able to suppress most of the artifacts. High noise levels are visible for all data terms (top left).

All regularizers help to reduce this noise. As $R_{\text{Pres.}}$ and R_{Weiss} have similar results, only the former is shown. Both do not sufficiently suppress noise. R_{TV} and R_{BTV} smooth the noise and preserve the edges, for example near the diaphragm and the heart shadow. In contrast, R_{L_2} slightly blurs edges and does not suppress noise. R_{BTV} is best at reducing streak artifacts (bottom middle).

4 Conclusions and Outlook

In this paper, a Bayesian probabilistic model for layer separation in transparency was presented. As this model is only a rough approximation of the real X-ray image generation process, it has to tolerate many outliers. To this end, we introduce robust data terms and robust regularization for motion layer separation in



(a) Image (b) D_{L_2} (c) D_{L_1} (d) $D_{\text{Cha.}}$

(e) $D_{\text{Cha.}} + R_{L_2}$ (f) $D_{\text{Cha.}} + R_{\text{TV}}$ (g) $D_{\text{Cha.}} + R_{\text{BTV}}$ (h) $D_{\text{Cha.}} + R_{\text{Pres.}}$

Fig. 5. Layer extracted from a real X-ray sequence using different combinations of data and regularization term (contrast enhanced for display).

fluoroscopy. A slowly increasing penalty function like the generalized Charbonnier is crucial in the data term. Furthermore, we showed that robust regularization like BTV yields semantically better separation. Image-driven regularization did not improve upon BTV, but might help in joint motion and layer estimation [10].

For the future, there are several areas for possible improvements. The image formation model can be extended to better model true X-ray physics, e.g., scattering. Joint motion and layer estimation with anatomically plausible layers would greatly enhance the practical usefulness. Another issue is runtime. Although the coarse-to-fine approach considerably reduces runtime, the current configuration of the optimizer requires up to a minute for computing the layers. The runtime can be improved using preconditioning or specialized solvers [5].

Acknowledgments. The authors gratefully acknowledge funding by Siemens Healthcare and of the Erlangen Graduate School in Advanced Optical Technologies (SAOT) by the German Research Foundation (DFG) in the framework of the German excellence initiative. The concepts and information presented in this paper are based on research and are not commercially available.

References

1. Black, M.J., Anandan, P.: The robust estimation of multiple motions: parametric and piecewise-smooth flow fields. Comput. Vis. Image Underst. **63**(1), 75–104 (1996)
2. Cao, Y., Wang, P.: An adaptive method of tracking anatomical curves in X-ray sequences. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 173–180. Springer, Heidelberg (2012)
3. Close, R.A., Abbey, C.K., Morioka, C.A., Whiting, J.S.: Accuracy assessment of layer decomposition using simulated angiographic image sequences. IEEE Trans. Med. Imaging **20**(10), 990–998 (2001)
4. Farsiu, S., Robinson, M.D., Elad, M., Milanfar, P.: Fast and robust multiframe super resolution. IEEE Trans. Image Process. **13**(10), 1327–1344 (2004)
5. Goldstein, T., Osher, S.: The split Bregman method for L1-regularized problems. SIAM J. Imaging Sci. **2**(2), 323–343 (2009)
6. Heibel, H., Glocker, B., Groher, M., Pfister, M., Navab, N.: Interventional tool tracking using discrete optimization. IEEE Trans. Med. Imaging **32**(3), 544–555 (2013)
7. Hermosillo, G., Chefd'Hotel, C., Faugeras, O.: Variational methods for multimodal image matching. Int. J. Comput. Vision **50**(3), 329–343 (2002)
8. Maier, A., Hofmann, H., Berger, M., Fischer, P., Schwemmer, C., Wu, H., Müller, K., Hornegger, J., Choi, J.H., Riess, C., Keil, A., Fahrig, R.: CONRAD-a software framework for cone- beam imaging in radiology. Med. Phys. **40**(11) (2013)
9. Manhart, M., Kowarschik, M., Fieselmann, A., Deuerling-Zheng, Y., Royalty, K., Maier, A., Hornegger, J.: Dynamic iterative reconstruction for interventional 4-D c-arm CT perfusion imaging. IEEE Trans. Med. Imaging **32**(7), 1336–1348 (2013)
10. Preston, J.S., Rottman, C., Cheryauka, A., Anderton, L., Whitaker, R.T., Joshi, S.: Multi-layer deformation estimation for fluoroscopic imaging. In: Gee, J.C., Joshi, S., Pohl, K.M., Wells, W.M., Zöllei, L. (eds.) IPMI 2013. LNCS, vol. 7917, pp. 123–134. Springer, Heidelberg (2013)
11. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Phys. D **60**(14), 259–268 (1992)
12. Segars, W., Mahesh, M., Beck, T., Frey, E., Tsui, B.: Realistic CT simulation using the 4D XCAT phantom. Med. Phys. **35**(8), 3800–3808 (2008)
13. Sun, D., Roth, S., Black, M.J.: A quantitative analysis of current practices in optical flow estimation and the principles behind them. Int. J. Comput. Vision **106**(2), 115–137 (2014)
14. Szeliski, R., Avidan, S., Anandan, P.: Layer extraction from multiple images containing reflections and transparency. In: CVPR, vol. 1, pp. 246–253. IEEE (2000)
15. Tipping, M.E., Bishop, C.M.: Bayesian image super-resolution. In: Becker, S., Thrun, S., Obermayer, K. (eds.) Advances in Neural Information Processing Systems, vol. 15, pp. 1303–1310. MIT Press, Cambridge (2003)
16. Weiss, Y.: Deriving intrinsic images from image sequences. In: ICCV, vol. 2, pp. 68–75. IEEE (2001)
17. Zhang, W., Ling, H., Prummer, S., Zhou, K.S., Ostermeier, M., Comaniciu, D.: Coronary tree extraction using motion layer separation. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 116–123. Springer, Heidelberg (2009)