Markov Random Field-based Layer Separation for Simulated X-Ray Image Sequences

Peter Fischer[1], Thomas Pohl[2], Andreas Maier[1], and Joachim Hornegger[1]

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

Abstract. Motion estimation in X-ray images is a challenging task due to transparently overlapping structures from different depths. We propose to separate an X-ray sequence into a static and a dynamic layer to facilitate motion estimation. The method exploits the idea to use the minimum intensity over time and a spatial smoothness prior for both layers. For numerical optimization, we propose a conditional Markov random field. In experiments on synthetic data, we achieve a root mean squared intensity difference of 36.7 ± 8.4 to the ground truth static layer. In addition, we show qualitative results that demonstrate an improved layer separation compared to state-of-the-art algorithms.

1 Introduction

X-ray images are 2-D projection images formed by accumulated attenuation along a line through a 3-D volume. This leads to a transparency effect that enables physicians to examine the interior of the human body. However, it also means that structures from different depths overlap transparently in the images. In many cases, some of the projected structures are unnecessary or even hinder interpretation and processing of the images. In particular, motion estimation is substantially complicated [1]. Many image registration algorithms are based on intensity similarities. Hence, the estimated motion is dominated by the high-contrast structures. However, the motion of the soft tissue that is investigated in the intervention is required. A separation of X-ray images into independent layers is therefore desired.

In literature, multiple approaches to layer separation in X-ray images have been proposed. Early methods were restricted to rigid motion of the layers and separated the layers by averaging the stabilized X-ray sequences [2]. Preston et al. alternate between non-rigid motion estimation and layer separation [3]. Layer separation is easier for dual-energy X-ray, where additional spectral information is available. In this domain, separation can be performed without motion estimation based on minimizing the mutual information between the layers [4]. Transparent layer separation has also been treated in computer vision. Szeliski et al. iteratively estimate parametric motion and layers, using the minimum over time to extract the static layer from a stabilized sequence [5]. Weiss separates the reflectance from the temporally changing illumination using an independence assumption and the sparsity of natural images in the gradient domain [6].

The contribution of this work is a new method for layer separation. It builds on the idea to use the pixel-wise minimum over time of a X-ray sequence to extract layers. However, in some cases the minimum does not correspond to semantically meaningful images. Prior knowledge, e.g., spatial smoothness and non-negativity, is useful to restrict the layers. We introduce a combined formulation for spatial smoothness of both transparent layers. The model is formulated as a conditional Markov random field (CRF). In the experiments, we show that our method separates X-ray sequences into two motion layers on synthetic data.

2 Materials and Methods

2.1 Layered X-Ray Model

X-ray images are generated by X-ray photons that are attenuated on their path through an imaged volume. We assume monochromatic X-ray. Attenuation is an exponential process, which can be transformed to a linear relationship between image intensities and attenuation using logarithmic processing [7]. We are interested in separating X-ray images into differently moving layers. Therefore, all tissues that undergo a similar motion are summarized into a single layer I_l. The image $I^t \in [0, 255]^{W \times H}$ at time $t \in \{1, \ldots, T\}$ is then computed from the layers as

$$I^t(\boldsymbol{x}) = \sum_{l=0}^{L-1} I_l^t(\boldsymbol{x}) \ , \tag{1}$$

with the image pixel $\boldsymbol{x} \in \mathrm{I\!R}^2$ and the number of layers L. In this work, we limit ourselves to $L = 2$ transparent layers, a static $I_0 = S$ and a dynamic $I_1 = D$ layer. The whole sequence of X-ray images is denoted as $\boldsymbol{I}$.

2.2 Layer Separation using a Conditional Markov Random Field

The basic assumption of our layer separation model is that the layer S is static. A straightforward method to remove a static layer from an X-ray sequence is to compute the pixel-wise minimum over time

$$S^{\min}(\boldsymbol{x}) = \min_t I^t(\boldsymbol{x}) \ , \tag{2}$$

because the dynamic layer can only increase attenuation. This min-composite yields an upper bound on the static layer [5]. Its major problem is that artificial edges are introduced. If an object is larger than the motion it performs in the sequence, a part of the object is assigned to the static layer (Fig. 1). In particular in medical images, moving structures are physically a better explanation than appearing and disappearing structures. To this end, we penalize the creation of artificial edges by introducing a spatial smoothness prior on both layers.

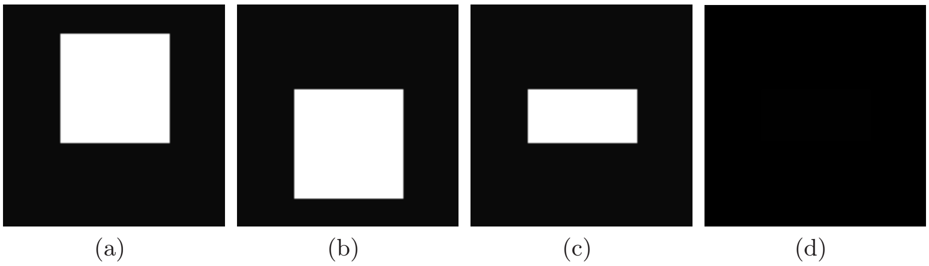

Fig. 1. Visualization of the artificial edge creation problem of the min-composite. The inputs are a white rectangle moving on a black background (a,b). Areas that are covered in all images by the rectangle are assigned to static layer for the min-composite (c). The desired result is a reached by our method (d).

We formulate the layer separation problem in a CRF model. Due to the assumption of a static layer, it is sufficient to represent each static layer pixel with a random variable. The intensity of the dynamic layer can be calculated directly from the static layer and the image

$$D^t(\boldsymbol{x}) = I^t(\boldsymbol{x}) - S(\boldsymbol{x}) \quad . \tag{3}$$

The random variables have discrete labels $z \in \mathcal{Z}$ representing the intensity $z_i = S(\boldsymbol{x}_i)$. The labels of all random variables are denoted as $\boldsymbol{z}$. $\mathcal{Z}$ contains equally distributed intensities in $[0, 255]$ without loss of generality.

In the CRF, we incorporate unary potentials Φ_v for nodes $v \in \mathcal{V}$ and pair-wise potentials Ψ_{ij} for edges $(i,j) \in \mathcal{E}$

$$E(\boldsymbol{z}, \boldsymbol{I}) = \sum_{v \in \mathcal{V}} \Phi_v(z_v, \boldsymbol{I}(\boldsymbol{x}_v)) + \sum_{(i,j) \in \mathcal{E}} \Psi_{ij}(z_i, z_j, \boldsymbol{I}(\boldsymbol{x}_i), \boldsymbol{I}(\boldsymbol{x}_j)) \quad . \tag{4}$$

The unary potential function

$$\Phi_v(z_v, \boldsymbol{I}(\boldsymbol{x}_v)) = \begin{cases} \alpha \min\left\{\beta, \sum_{t=1}^{T} \|z_v - I^t(\boldsymbol{x}_v)\|_1\right\}, & \text{if } z_v \leq I^t(\boldsymbol{x}_v) \,\forall t \\ \infty, & \text{otherwise} \end{cases} \tag{5}$$

with parameters $\alpha, \beta \in \mathbb{R}$ penalizes deviations from the min-composite using a truncated L_1-norm. The unary potential ensures the non-negativity constraint of the X-ray generation model in the dynamic layer. The static layer is non-negative by definition of the label set $\mathcal{Z}$. Φ_v prevents the static layer from being larger than any of the images $I^t(\boldsymbol{x}_v)$ by assigning infinite weight, thus avoiding a negative dynamic layer. Note that the minimum of the unary potential is achieved by the min-composite.

The potential Ψ_{ij} consists of pair-wise terms $(i,j) \in \mathcal{E}$ in a 4-neighborhood

$$\begin{aligned} \Psi_{ij}(z_i, z_j, \boldsymbol{I}(\boldsymbol{x}_i), \boldsymbol{I}(\boldsymbol{x}_j)) = {} & \|z_i - z_j\|_1 + \\ & \sum_{t=1}^{T} \left\|(z_i - z_j) - \left(I^t(\boldsymbol{x}_i) - I^t(\boldsymbol{x}_j)\right)\right\|_1 \quad . \end{aligned} \tag{6}$$

A common image prior is to penalize gradients, e.g., using the L_1-norm to promote sparsity. Eq. 6 jointly encodes smoothness of both layers. This is straightforward for the static layer using $\|z_i - z_j\|_1$. It needs to be added only once, because the pixel values of the static layer are perfectly statistically dependent $p\left(\boldsymbol{S}(\boldsymbol{x})\right) = p\left(S(\boldsymbol{x})\right)$ over time. For the dynamic layer, we reformulate $\left\|D^t\left(\boldsymbol{x}_i\right) - D^t\left(\boldsymbol{x}_j\right)\right\|_1$ using Eq. 3, thus removing the need to directly model D^t. Assuming statistical independence of the gradients over time $p\left(\boldsymbol{D}(\boldsymbol{x})\right) = \prod_{t=1}^{T} p\left(D^t(\boldsymbol{x})\right)$, different time steps can be combined by summation in the energy. With the assumption of independence of the static and the dynamic layer $p\left(\boldsymbol{S}(\boldsymbol{x}), \boldsymbol{D}(\boldsymbol{x})\right) = p\left(\boldsymbol{S}(\boldsymbol{x})\right) \cdot p\left(\boldsymbol{D}(\boldsymbol{x})\right)$, the individual layer contributions can be added (Eq. 6). Note that the minimum of the pair-wise potential is achieved by the median gradient over time [6].

To perform the layer separation in a new X-ray sequence, the maximum a posteriori (MAP) estimate of the CRF model is obtained by

$$\boldsymbol{z}^* = \underset{\boldsymbol{z}}{\operatorname{argmin}}\, E\left(\boldsymbol{z}, \boldsymbol{I}\right) \quad . \tag{7}$$

This yields the statistically optimal layer under this model given the input sequence. For inference, sequential tree-reweighted message passing (TRWS) [8] in the OpenGM framework is used [9].

2.3 Experiments

In the experiments, we compare the proposed algorithm to the min-composite [5] and Weiss method [6]. Min-composite and Weiss method do not have any parameters. The parameters of our method were set empirically to $\alpha = 0.1, \beta = 10$. TRWS optimization is run for 40 iterations, using $\|\mathcal{Z}\| = 256$ labels.

As experimental data, we use four simulated X-ray sequences. Simulated images resemble real X-ray images, but ground truth is still available. They are created by adding two independent layers, where one is static and the other one dynamic. Layers are created by segmenting 3-D volumes and projecting the parts independently to 2-D. The 3-D volumes are created by clinical CT scanners or simulations using CONRAD [10]. The dynamic layer is transformed with artificial motions, which are interpolated from manually specified control point motions using thin-plate splines. The error is computed as the root mean squared difference (RMSD) of the image intensities. Before the RMSD, we subtract the mean from the compared layers, because it cannot be uniquely determined and is not relevant for motion estimation.

3 Results

The RMSD error for the min-composite is 42.1 ± 8.4, for the Weiss method 42.2 ± 6.9, and for our method 36.7 ± 8.4 (mean ± standard deviation). These results indicate a better performance of our method compared to the others.

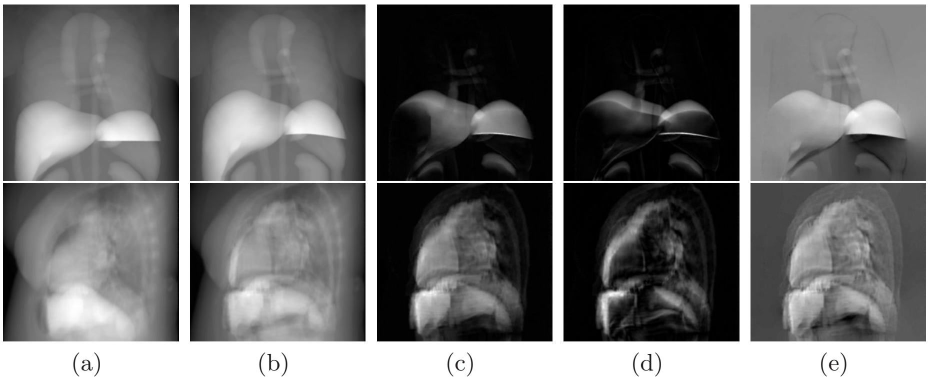

Fig. 2. Qualitative results on simulated X-ray sequences are shown, one per row: two images of the input sequence (*a*,*b*), a dynamic layer extracted using our (*c*), min-composite (*d*), and Weiss (*e*) method. Contrast is enhanced for better display.

Two X-ray sequences from different views including the spine, diaphragm, ribs, heart, and lungs are depicted in Fig. 2. The images are already preprocessed to fit the additive model. The sequence in the first row is created using CONRAD. Note that the static structures are removed from the dynamic layer in all cases. The main difference is how well the soft tissue is preserved. There, the problem of artificial edges is clearly visible in the min-composite. In the second sequence created from a 3-D CT, more structure is present in the soft tissue. Nevertheless, the same problems occur in the min-composite. Our approach and Weiss method perform similarly well. Both have problems with inconsistent gradient estimates, e.g., visible in the liver in the first sequence. The main differences is that Weiss method does not ensure non-negativity of the dynamic layer, which corresponds to physically impossible negative attenuations. Non-negativity of the static layer can be achieved by simple postprocessing. Another difference is an offset of the mean intensity, which cannot be uniquely determined from gradient information alone, but is irrelevant for subsequent motion estimation.

The runtime of the method depends linearly on the number of pixels. For images of size $W = H = 256$ and a sequence of length $T = 50$, the runtime is about 200 s.

4 Discussion

We propose a novel approach to separate an X-ray image sequence into static and dynamic layers. This intermediate representation can facilitate further processing. In particular, soft tissue motion estimation would not be possible without it due to overlapping structures from different depths. The separation is based on the min-composite, which is only an upper bound on the static layer. Our method adds a smoothness term to suppress artificial edges in either layer.

The current runtime of the method is not yet sufficient for clinical use. However, real-time performance is not feasible by design, as a whole image sequence is postprocessed. The goal can only be to reduce the latency to a minimum.

In future work, the method needs to be transferred to and tested on clinical X-ray data. The validity of a static layer is questionable for clinical X-ray images. Some structures, e.g., ribs, move slightly, although from an application point of view they should be in the static layer. Additionally, patient body motion is possible. Consequently, the robustness of the method to these challenges needs to be evaluated. In the future, the use of the dynamic layer for motion estimation should be investigated. Furthermore, substantial speed ups of the method are possible for example using an inference method that is amenable to parallelization and a GPU implementation.

Acknowledgement. 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. Klüppel M, Wang J, Bernecker D, Fischer P, Hornegger J. On Feature Tracking in X-Ray Images. Procs BVM. 2014; p. 132–137.
2. Close RA, Abbey CK, Morioka CA, Whiting JS. Accuracy assessment of layer decomposition using simulated angiographic image sequences. IEEE Trans Med Imaging. 2001;20(10):990–998.
3. Preston JS, Rottman C, Cheryauka A, Anderton L, Whitaker R, Joshi S. Multilayer Deformation Estimation for Fluoroscopic Imaging. In: Inf Process Med Imaging. vol. 7917 of Lect Notes Comput Sci. Springer; 2013. p. 123–134.
4. Chen Y, Chang TC, Zhou C, Fang T. Gradient domain layer separation under independent motion. In: Proc. IEEE Int Conf Comput Vis. IEEE; 2009. p. 694–701.
5. Szeliski R, Avidan S, Anandan P. Layer Extraction from Multiple Images Containing Reflections and Transparency. In: Proc. IEEE Comput Soc Conf Comput Vis Pattern Recognit. vol. 1. IEEE; 2000. p. 246–253.
6. Weiss Y. Deriving intrinsic images from image sequences. In: Proc. IEEE Int Conf Comput Vis. vol. 2. IEEE; 2001. p. 68–75.
7. Buzug TM. Computed tomography: from photon statistics to modern cone-beam CT. Springer; 2008.
8. Kolmogorov V. Convergent tree-reweighted message passing for energy minimization. IEEE Trans Pattern Anal Mach Intell. 2006;28(10):1568–1583.
9. Andres B, Beier T, Kappes JH. OpenGM: A C++ library for discrete graphical models. arXiv. 2012;1206.0111:1–5.
10. Maier A, Hofmann H, Berger M, Fischer P, Schwemmer C, Wu H, et al. CONRAD - A software framework for cone-beam imaging in radiology. Med Phys. 2013;40(11):111914–1 – 111914–8.