TV-L_1-Based 3D Medical Image Registration with the Census Cost Function

Simon Hermann[1,2] and René Werner[3]

[1] Department of Computer Science, The University of Auckland, New Zealand
[2] Department of Computer Science, Humboldt University of Berlin, Germany
[3] Department of Computational Neuroscience, University Medical Center Hamburg-Eppendorf, Germany

Abstract. A recent trend in computer vision is to combine the census cost function with a TV-L_1 energy minimization scheme. Although this combination is known for its robust performance in computer vision applications, it has not been introduced to 3D medical image registration yet. Addressing pulmonary motion estimation in 4D (3D+t) CT images, we propose incorporating the census cost function into a 3D implementation of the 'duality-based approach for realtime TV-L_1 optical flow' for the task of lung CT registration. The performance of the proposed algorithm is evaluated on the DIR-lab benchmark and compared to state-of-the-art approaches in this field. Results highlight the potential of the census cost function for accurate pulmonary motion estimation in particular, and 3D medical image registration in general.

Keywords: Medical image registration, census transform, pulmonary motion estimation, 4D CT.

1 Introduction

This paper is motivated by the clinical need for fast and accurate registration of 3D computed tomography (CT) images. Recent advances in computer vision resulted in strong performing 2D optical flow estimation algorithms, which include the census cost function into a TV-L_1 energy minimization scheme. This paper demonstrates their potential for the task of pulmonary motion analysis. Figure 1 illustrates the application context.

The applied registration method is a 3D variant of the 'duality-based approach for realtime TV-L_1 optical flow', which was published in 2007 by Zach et al. [27]. They introduced a coupled convex approximation of a TV-L_1 energy functional, defined by an L_1 data term and a total variation (TV) regularization term. Both terms are alternately minimized within an iterative scheme. The decoupling of data and smoothness term, combined with a point-wise gradient-descent step to minimize the data term residual, makes this algorithm very applicable to be implemented on parallel hardware architectures. For the sake of simplicity, we refer to this particular minimization scheme as TV-L_1. However, there are other concepts to minimize the same energy; see for example Brox et al. [2].

R. Klette, M. Rivera, and S. Satoh (Eds.): PSIVT 2013, LNCS 8333, pp. 149–161, 2014.

Fig. 1. Illustration of the task of 3D lung CT registration for pulmonary motion analysis. Left and middle: Coronal views of lung CT images at end-expiration (EE) and end-inspiration (EI). Right: Overlay of the EE (blue) and EI (orange) lung structures. These structures are to be aligned by a physiologically plausible, non-linear transformation; the corresponding displacements are interpreted as the motion field between EE and EI.

The census cost function is based on the census transform, which was introduced in 1994 by Zabih and Woodfill [26]. The census transform is a binary representation of all intensity differences at a pixel in relation to its immediate pixel neighborhood. It is based on ordering statistics and encodes in addition the spatial relationship between the considered pixels. The census cost function, which is the Hamming distance of two binary representations, is often applied for correspondence analysis in discrete methods. Those methods usually combine a data cost accumulation scheme with a cost limitation function to provide piecewise smooth and consistent solutions. An established example of such a combination is semi-global matching (SGM), as introduced by Hirschmüller [11] for stereo analysis.

In 2011, Müller et al. [14] discussed the formulation of an L_1 data term for arbitrary non-linear residuals in coupled energy minimization schemes such as TV-L_1. They highlighted the performance gain for dense 2D optical flow estimation that results when intensity-based data terms are replaced by the census cost function. For the performance of correspondence detection methods using census data terms we refer to the KITTI Vision Benchmark Suite.[1]

The work presented in this paper may be considered as an extension of the method proposed by Pock et al. [16], who applied the intensity-based TV-L_1 scheme [27] to lung CT registration. The introduction of the census cost function, following Müller et al. [14], represents a novel contribution to 3D medical image registration in general, and to pulmonary motion estimation in 4D CT image data in particular.

The paper is structured as follows: Section 2 presents the clinical background and informs about the state-of-the-art in 3D pulmonary medical image registration. Section 3 outlines TV-L_1 optical flow estimation, the applied numerical scheme, and the incorporation of the census data term. Experiments and results are presented in Section 4. Section 5 concludes.

[1] `http://www.cvlibs.net/datasets/kitti/`

2 Image Registration for Pulmonary Motion Estimation

Since the adaption of cardic imaging protocols for time-resolved thoracic and abdominal CT in the last decade, 4D CT scanners and images are making their way into an increasing number of medical facilities [21]. Here, the 3D CT images of a 4D CT data set are assumed to sample the patient's breathing cycle. A typical application scenario of 4D CT data is the estimation of breathing-induced effects on a planned dose distribution (dose accumulation) during radiotherapy treatment planning of thoracic and abdominal tumors [7]. Dose accumulation aims at understanding the interaction between organ and tumor motion and the process of dose delivery for a specific treatment modality.

It is usually performed on a pixel-by-pixel basis and requires non-linear 3D registration to estimate the displacement fields between different frames of the 3D CT image sequence [13,24].

The reliability of the extracted information, which is applied, for example, to dose accumulation, is directly linked to the accuracy of the registration [15]. This motivation led to great interest in 3D medical image registration of lung CT data over the past years. Although there is a great diversity of registration approaches in this field (see, for example, [1,7,15]), from a methodological perspective it is noticeable that the majority of these approaches realizes a minimization of common intensity-based distance measures or data terms, which are usually formulated as squared L_2 norms. Here, we refer to the EMPIRE10[2] study (EMPIRE = Evaluation of Methods for Pulmonary Image REgistration 2010) as a recent and large comparison benchmark. Considering the top-ranked algorithms as current state-of-the-art in this field, it is interesting to see that six of the top-ten approaches used variants of SSD (sum of squared intensity distances) as data terms during the initial study phase [15].

Thus, in contrast to computer vision applications, the use of an L_1 data term is not common for pulmonary motion estimation in 4D CT data. The same is true for the application of TV-based regularization. Nevertheless, some work exists in this direction: For example, three of the EMPIRE10 participants proposed using SAD (sum of absolute intensity differences) as distance measure, with some recent publications on lung registration also following this idea (e.g. [8]). As mentioned before, Pock et al. and, in addition, Urschler et al. proposed applying the 'duality-based approach for realtime TV-L_1 optical flow' of Zach et al. [27] for 3D lung CT registration [22,16]. However, the published registration errors are considerably larger when compared to state-of-the-art algorithms for lung registration.

To our best knowledge, the integration of a census cost function as penalty term for TV-L_1 has not yet been evaluated in the given context.

[2] `http://empire10.isi.uu.nl`

3 Formulation of Duality-Based TV-L_1 Optical Flow

This section provides a formulation of the duality-based TV-L_1 optical flow energy optimization. It follows the notation of Zach et al. and Pock et al. [27,16]. The integration of an arbitrary data term is taken from Müller et al. [14].

3.1 Basic Notations

The input of the algorithm are two single channel images I_0 and I_1, defined on an n-dimensional image domain $\Omega \subset \mathbb{R}^n$ with $I : \Omega \longrightarrow \mathbb{R}$. A point $\mathbf{x} \in \Omega$ describes a position within the image domain and $I(\mathbf{x})$ refers to the intensity of an image I at $\mathbf{x}$.

The output is an n-dimensional vector field $\mathbf{u} : \Omega \longrightarrow \mathbb{R}^{m \wr n}$, with $\mathbf{u} = (u_1, ..., u_m, 0_{m+1}, ..., 0_n)$, that describes the displacement from the base or reference image I_0 to the match or target image I_1. Here, m refers to the degrees of freedom of the elements of the vector field. A typical example for $m < n$ is the case of stereo estimation for rectified image pairs, where $n = 2$ but $m = 1$. For the task of 3D medical image registration we have $n = m = 3$.

3.2 The Energy Functional

The approach for recovering the displacement field $\mathbf{u}$ is based on minimizing the energy functional

$$E(\mathbf{u}) = \int_\Omega \sum_{b=1}^{m} \left|\nabla u_b(\mathbf{x})\right| + \lambda \left|\rho(\mathbf{x}, \mathbf{u}(\mathbf{x}))\right| \, \mathrm{d}\mathbf{x} \tag{1}$$

where $\rho(\mathbf{x}, \mathbf{u})$ refers to a generic residual with $\rho := \rho_{I_0,I_1} : \Omega \times \mathbb{R}^{m \wr n} \longrightarrow \mathbb{R}$, which is evaluated between I_0 and I_1 at position $\mathbf{x}$ under consideration of $\mathbf{u}(\mathbf{x})$. To minimize this energy, the following convex approximation of Eq. (1) was introduced by Zach et al [27]:

$$E(\mathbf{u}, \mathbf{v}) = \int_\Omega \sum_{b=1}^{m} \left[\left|\nabla u_b(\mathbf{x})\right| + \frac{1}{2\theta}(u_b(\mathbf{x}) - v_b(\mathbf{x}))^2\right] + \lambda \left|\rho(\mathbf{x}, \mathbf{v}(\mathbf{x}))\right| \, \mathrm{d}\mathbf{x} \tag{2}$$

with $\mathbf{v} : \Omega \longrightarrow \mathbb{R}^{m \wr n}$. This can be efficiently solved by applying the following iterative scheme:

(1) For every b and fixed v_b, minimize over u_b:

$$\arg\min_{u_b} \left\{ \int_\Omega \left|\nabla u_b(\mathbf{x})\right| + \frac{1}{2\theta}(u_b(\mathbf{x}) - v_b(\mathbf{x}))^2 \, d\mathbf{x} \right\} \tag{3}$$

(2) For fixed $\mathbf{u}$, minimize over $\mathbf{v}$:

$$\arg\min_{\mathbf{v}} \left\{ \int_\Omega \sum_{b=1}^{m} \frac{1}{2\theta}(u_b(\mathbf{x}) - v_b(\mathbf{x}))^2 + \lambda \left|\tilde{\rho}(\mathbf{x}, \mathbf{v})\right| \, d\mathbf{x} \right\} \tag{4}$$

where the residual ρ is linearized around the fixed value $\mathbf{u}$ using the Taylor expansion [14]:

$$\rho(\mathbf{x}, \mathbf{v}(\mathbf{x})) \approx \tilde{\rho}(\mathbf{x}, \mathbf{v}(\mathbf{x})) = \rho(\mathbf{x}, \mathbf{u}(\mathbf{x})) - (\mathbf{u} - \mathbf{v})\nabla\rho(\mathbf{x}, \mathbf{u}) \tag{5}$$

3.3 The Numerical Scheme

The minimization problem defined by Eq. (3) is the TV energy functional as introduced by Rudin et al. [17]. It is solved by the efficient numerical scheme proposed by Chambolle [6]. With $\mathbf{p} \in \mathbb{R}^n$, and omitting the dependency on $\mathbf{x}$ for better readability, the solution of Eq. (3) is for $b = 1, .., m$:

$$u_b = v_b - \theta\,\mathbf{div}\,\mathbf{p}_b \tag{6}$$

with

$$\nabla(\theta\,\mathbf{div}\,\mathbf{p}_b - v_b) = |\nabla(\theta\,\mathbf{div}\,\mathbf{p}_b - v_b)|\mathbf{p}_b \tag{7}$$

To solve this equation, the following iterative fixed point scheme is applied:

$$\mathbf{p}_b^{k+1} = \frac{\mathbf{p}_b^k + \tau\nabla(\mathbf{div}\,\mathbf{p}_b^k - \frac{1}{\theta}v_b)}{1 + \tau|\nabla(\mathbf{div}\,\mathbf{p}_b^k - \frac{1}{\theta}v_b)|} \tag{8}$$

with $\mathbf{p}_b^0 = \mathbf{0}$ and τ as a time step.

The minimization problem of Eq. (4) can be solved by applying the thresholding scheme described in [27]:

$$\mathbf{v} = \mathbf{u} + \begin{cases} \lambda\theta\nabla\tilde{\rho} & \text{if } \tilde{\rho} < -\lambda\theta|\nabla\tilde{\rho}|^2 \\ -\lambda\theta\nabla\tilde{\rho} & \text{if } \tilde{\rho} > \lambda\theta|\nabla\tilde{\rho}|^2 \\ -\frac{\tilde{\rho}\nabla\tilde{\rho}}{|\nabla\tilde{\rho}|^2} & \text{if } |\tilde{\rho}| \leq \lambda\theta|\nabla\tilde{\rho}|^2 \end{cases} \tag{9}$$

3.4 The Census Cost Function as Data Residual

The census transform, as introduced by Zabih and Woodfill [26], assigns a binary signature vector to an image position $\mathbf{x}$. The signature is calculated based on the ordinal characteristic of $I(\mathbf{x})$ in relation to intensities within a defined neighborhood $\mathcal{N}_\mathbf{x}$ of $\mathbf{x}$. The binary signature vector $C[\mathbf{x}]$ at $\mathbf{x}$ is generated as follows:

$$C[\mathbf{x}] = \Big\{\Upsilon\big[I(\mathbf{x}) \geq I(\mathbf{y})\big]\Big\}_{\mathbf{y}\in\mathcal{N}_\mathbf{x}} \tag{10}$$

where $\Upsilon[\cdot]$ returns 1 if true, and 0 otherwise. The residual $\rho(\mathbf{x}, \mathbf{u})$ is the Hamming distance of two signature vectors. Formally, we write

$$\rho(\mathbf{x}, \mathbf{u}) = \mathit{\Gamma}\big\{C_0[\mathbf{x}] \oplus C_1[\mathbf{x} + \mathbf{u}]\big\} \tag{11}$$

where C_0 and C_1 refer to census signatures of the images I_0 and I_1. $\mathit{\Gamma}\{\cdot\}$ is an operator that counts the 1's of the binary string in the argument. For efficiently counting 1's in bit strings, we refer to Warren's 'Hackers Delight' [23].

3.5 Implementation Details

Before the input images are given to the algorithm, their dynamic is reduced to an intensity domain of $[0, 1]$, with a scale factor that is based on the maximum intensity of both images. In order to restrict the focus during the registration process on a specific region of interest, the algorithm takes binary lung segmentation masks $S_0, S_1 : \Omega \to \{0, 1\}$ as optional input data. In case that segmentation masks are provided, the input images are cropped accordingly and the evaluation of the data term is omitted where $S_0(\mathbf{x}) = 0$. This provides a way of coping with motion discontinuities near the lung borders, which are the consequence of the lungs sliding along the inner part of the chest wall, and additionally speeds up the run-time of the algorithm.

The numerical scheme from section 3.3 is embedded into a coarse-to-fine approach, a common technique to overcome local minima. We employ Gauss pyramids with L levels, where $\ell = 1$ refers to the finest and $\ell = L$ refers to the coarsest level. If the input images exhibit an anisotropic spatial resolution (CT data usually offer high in-plane resolution when compared to the slice thickness; cf. Section 4.1), we follow a common strategy, see for example [18,12]: In a first step, we scale down only along the image axes with a high resolution until the pixel spacing is almost identical for all image dimensions. The following downscaling steps are then performed along all image axes and are subsequently referred to as 'isotropic levels'.

At the coarsest level, the displacement field is initialized with zero and the result of each level is propagated to the next finer level. At the beginning of each level, the dual variables $\mathbf{p}_b$ with $b = \{1, 2, 3\}$ are initialized with zero. Each level consists of a fixed number of image warps η and a fixed number υ of iterations, in which the alternating numerical scheme is applied. After each iteration on isotropic levels, the intermediate solution is first filtered by a median filter to remove potential outliers (see [25]), followed by a Gauss filter with a fixed window size ω and a fixed σ. The scheme is summarized in Algorithm 1.

Our TV-L_1 implementation is C++ based, and the algorithm is executed on an Intel®Core$^{\text{TM}}$ i7 Quad-Core Processor with 2.4 GHz. We use OpenMP to utilize hyper-threading. No calculation was, however, outsourced to a GPU. This should be noted because one of the main features of the applied algorithm is that it is especially well suited to be implemented on GPUs, which could result in an additional run-time improvement over our CPU-based implementation.

4 Experiments

The performance of the algorithm was evaluated on ten thoracic 4D CT data sets, provided by the DIR-lab[3] of the University of Texas M.D. Anderson Cancer Center (Houston, USA) [3,4].

[3] http://www.dir-lab.com

input : images I_0 and I_1
ouput: displacement field $\mathbf{u}$

initialization of image pyramids
set $\mathbf{u}^L = \mathbf{0}$

for $\ell \leftarrow L$ **to** 1 **do**
  for $b \leftarrow 1$ **to** 3 **do**
    set $\mathbf{p}_b^\ell = \mathbf{0}$
  end
  for $w \leftarrow 1$ **to** η **do**
    warp $I_1^\ell \xrightarrow{\mathbf{u}_w^\ell} I_0^\ell$, by means of trilinear interpolation
    for $i \leftarrow 1$ **to** υ **do**
      solve Eq. (9)
      for $b \leftarrow 1$ **to** 3 **do**
        solve Eq. (8)
      end
      if ℓ *is isotropic level* **then**
        filter $\mathbf{u}_w^\ell$ with a $3 \times 3 \times 3$ median filter
      end
      smooth $\mathbf{u}_w^\ell$ with a Gauss filter (ω, σ)
    end
  end
  if $\ell > 1$ **then**
    initialize $\mathbf{u}^{\ell-1}$ with $\mathbf{u}^\ell$
  else
    return current solution $\mathbf{u}^1$
  end
end

Algorithm 1. Census-based TV-L$_1$ algorithm

4.1 Image Data and Evaluation

Each 4D CT image sequence of the DIR-lab data pool consists of 3D CT images for ten different breathing phases. The dimension of the sequences varies between 256×256×94 and 512×512×136 pixel, the spatial resolution between 0.97×0.97×2.5 mm^3 and 1.16×1.16×2.5 mm^3. To be in line with most of the related publications, we focussed on the registration of the end-inspiration (EI, base image) and end-expiration (EE, target image) scans for evaluation purposes. These are the scan pairs with the largest motion amplitudes. The corresponding lung segmentations are generated using basic image processing operations with manual correction, followed by a level set-based refinement; cf. [20] for details.

For each EI and EE scan pair, 300 anatomical landmark pairs within the lungs (prominent bifurcations of the bronchial or vessel trees) have been annotated and are provided by the DIR-lab. The landmark pairs serve as ground truth information for quantitative evaluation of the registration accuracy, for which

the Euclidean distances between the landmark positions in the target image and the positions of the transformed reference landmarks are computed.

The resulting landmark-based registration errors are compared with corresponding numbers that can be found in current literature and on the DIR-lab website. These numbers can, however, be computed in two different ways. Let $\hat{\mathbf{x}}_{I_0}$ be a landmark position in the reference image, $\hat{\mathbf{x}}_{I_0} + \mathbf{u}(\hat{\mathbf{x}}_{I_0})$ its transformed position, and $\hat{\mathbf{x}}_{I_1}$ the true landmark position in the target image. One way is to directly compute $\|\hat{\mathbf{x}}_{I_0} + \mathbf{u}(\hat{\mathbf{x}}_{I_0}) - \hat{\mathbf{x}}_{I_1}\|_2$. Instead, the DIR-lab proposes moving the transformed reference landmark towards the next pixel center before calculating its distance to $\hat{\mathbf{x}}_{I_1}$. This 'snap-to-pixel evaluation' is justified by arguing that human observers only select integer pixel locations in image pairs when identifying landmark sets. Working on the same integer grid would therefore increase consistency when comparing registration results with landmark positions as defined by the human observers. In this paper, we follow the DIR-lab argumentation and evaluation strategy. We also list the intra-observer errors as given on the DIR-lab website. For a landmark $\hat{\mathbf{x}}_{I_0}$, this value describes the mean distance between 'true' positions $\hat{\mathbf{x}}_{I_1}$ if they are detected by a human observer in several runs. In other words: The final goal is to end up with registration errors in the order of and below these observer errors.

Furthermore, we applied the common intensity-based registration approach of Schmidt-Richberg et al. [19] to the DIR-lab data, which is one of the top-ten-ranked approaches of the EMPIRE10 benchmark. This allows us to present some qualitative comparison of the estimated motion fields in addition to the quantitative landmark-based evaluation.

4.2 Algorithm Configuration

In this paper, we used the following registration parameters:

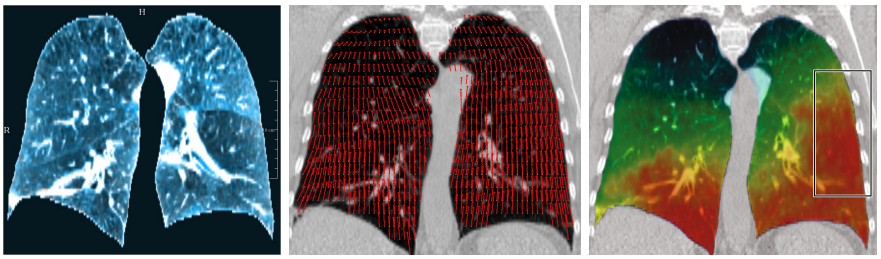
Census-Based TV-$\mathbf{L}_1$ Registration: Census evaluated in a $5 \times 5 \times 5$ neighborhood on isotropic and $5 \times 5 \times 3$ neighborhood on anisotropic levels; image pyramids with 5 levels; $\eta = 32$ and $\upsilon = 2$ for each level, Gauss smoothing of the flow with $(\omega, \sigma) = (5 \text{ pixel}, 1.0 \text{ pixel})$; $\lambda = 30$, $\tau = 0.25$, $\theta = 0.1$.

Intensity-Based TV-$\mathbf{L}_1$ Registration: Same parameters as for the census-based registration, except for the number of warps per level (now: $\eta = 128$) and the relative influence of the residual (now: $\lambda = 150$).

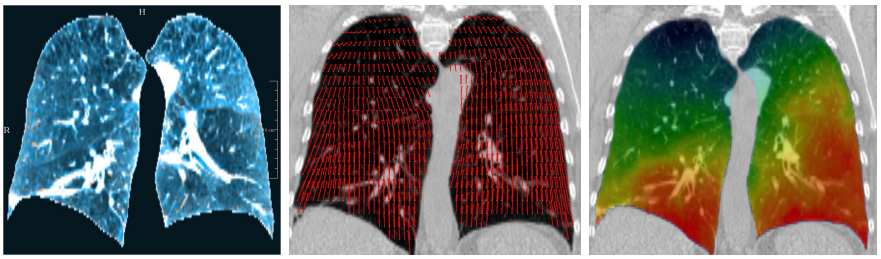
Registration Approach of Schmidt-Richberg et al.: default parameter setting as reported in the paper [19].

4.3 Results

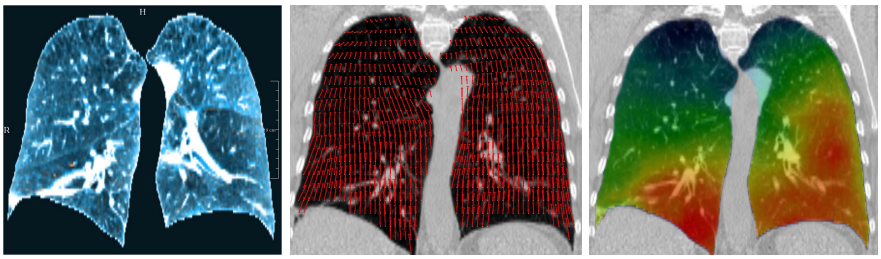
Intensity (SAD)- vs. Census-Based TV-$\mathbf{L}_1$ Registration. The registration results and estimated motion fields of data set 08 (maximal landmark displacement from all data sets) are illustrated in Fig. 2. The figure demonstrates that both the TV-L_1-based registration with common intensity-based data term



(a) TV-L_1-based registration with census cost function



(b) TV-L_1-based registration with common intensity-based data term (SAD)



(c) Intensity-based registration with approach of Schmidt-Richberg et al. [19]

Fig. 2. Illustration of registration results (here: DIR-lab case 08, registration masked by lung segmentation data). From left to right: Overlay of lung structures of the reference (orange) and the warped target image (blue), estimated motion vectors between EI and EE, and color-coded visualization of the motion field magnitude inside the lungs (blue: <2 mm motion; red: >25 mm). Rows from top to bottom: (a) census-based TV-L_1 registration, (b) standard intensity-based TV-L_1 registration, and (c) the registration approach following [19].

(i.e. SAD) and the proposed variant that minimizes the census cost function yield a visually good alignment of inner lung structures (see Fig. 1 for the displacement prior registration). The estimated motion fields also look similar for the two variants, with small differences being visible, e.g., near the outer left lung border (see highlighted area), and the field calculated by means of the census data term exhibits in general slightly larger motion amplitudes.

However, these small differences lead for the particular data set and the given algorithm configuration to a reduction of the landmark-based registration error from 1.51 mm to 1.11 mm when applying the census cost function instead of

Table 1. Mean landmark distances and corresponding standard deviations before and after registration (snap-to-pixel evaluation) and the intra-observer error for repeated landmark identification for the individual data sets values as specified by the DIR-lab. All values in mm.

			w/o lung masks		with lung masks	
#	No reg.	Obs. error	Intensities	Census	Intensities	Census
01	3.89 (2.78)	0.85 (1.24)	1.57 (1.29)	0.79 (0.93)	1.29 (1.11)	**0.78 (0.92)**
02	4.34 (3.90)	0.70 (0.99)	0.76 (0.98)	0.80 (0.96)	**0.73 (0.92)**	0.78 (0.92)
03	6.94 (4.05)	0.77 (1.01)	2.37 (2.33)	1.02 (1.18)	1.24 (1.19)	**0.93 (1.09)**
04	9.83 (4.86)	1.13 (1.27)	1.73 (1.95)	1.23 (1.27)	1.31 (1.29)	**1.24 (1.30)**
05	7.48 (5.51)	0.92 (1.16)	1.99 (2.30)	1.27 (1.52)	1.44 (1.61)	**1.22 (1.43)**
06	10.89 (6.97)	0.97 (1.38)	1.79 (1.61)	1.09 (1.35)	1.38 (1.20)	**0.94 (0.99)**
07	11.03 (7.43)	0.81 (1.32)	3.67 (4.59)	1.87 (3.06)	1.36 (1.22)	**1.01 (0.96)**
08	14.99 (9.01)	1.03 (2.19)	7.09 (8.56)	3.01 (5.16)	1.51 (1.71)	**1.11 (1.28)**
09	7.92 (3.98)	0.75 (1.09)	2.06 (1.70)	1.11 (1.24)	1.16 (1.03)	**0.98 (1.00)**
10	7.30 (6.35)	0.86 (1.45)	2.01 (2.66)	1.17 (1.87)	1.32 (1.51)	**0.94 (1.03)**
$Ø_{err}$	8.46 (6.58)	0.88 (1.31)	2.50 (2.80)	1.34 (1.85)	1.27 (1.28)	**0.99 (1.09)**

intensity differences (registration with lung masks). This trend can be observed for most of the other data sets as well; corresponding results can be found in Table 1. The table also illustrates the advantage of lung segmentation information for the registration process, with especially high differences for the intensity-based registration (mean error without lung masks: 2.50 mm; with lung masks: 1.27 mm).

Comparison to State-of-the-Art Approaches. Corresponding error values for the approach of Schmidt-Richberg et al. [19] and related numbers of other state-of-the-art algorithms for pulmonary motion estimation are summarized in Table 2. Minimum values are highlighted by a gray background and with bold letters. In case of two identical mean values, the standard deviation (found in brackets behind each error value) determines the minimum.

In contrast to published data on TV-L_1-based image registration [16,22], we achieved registration errors in the order of the state-of-the-art approaches already for our implemented SAD-based TV-L_1 variant. However, these values are only obtained if lung segmentation masks are used during the registration process. This common pre-processing step has not been described in [16,22] and may be the reason for the comparably weak performance reported in these papers.

Replacing the intensity-based data term with the census cost function significantly reduces registration errors. The resulting values of our census-based TV-L_1 algorithm are competitive to the lowest registration errors that are currently published on the DIR-lab benchmark. This holds for both registration with (cf. Table 2) and without employing lung segmentation data (see, e.g., [8] with a reported average registration error of 1.43 mm). It should be mentioned

Table 2. Landmark-based registration errors for the approach following [19] and corresponding published numbers after registration of the end-inspiration and -expiration images of the DIR-lab 4D CT data sets; values given in mm (snap-to-pixel evaluation). 'Our approach' refers to the masked census-based TV-L$_1$ registration.

	Our approach	**CCW+ [5]**[1]	**HSS+ [10]**[2]	**HW [9]**	**RHK+ [18]**	**SEW+ [19]**
01	0.78 (0.92)	0.85 (1.00)	0.98 (1.00)	0.80 (0.92)	**0.78 (0.91)**	0.87 (0.93)
02	0.78 (0.92)	0.74 (0.99)	0.83 (1.02)	0.77 (0.92)	**0.74 (0.87)**	0.84 (0.95)
03	0.93 (1.09)	0.93 (1.07)	1.08 (1.15)	**0.92 (1.10)**	0.94 (1.07)	1.02 (1.13)
04	1.24 (1.30)	1.33 (1.51)	1.45 (1.53)	**1.22 (1.24)**	1.26 (1.26)	1.35 (1.27)
05	1.22 (1.43)	**1.14 (1.25)**	1.55 (1.75)	1.21 (1.47)	1.22 (1.48)	1.39 (1.47)
06	0.94 (0.99)	1.04 (1.05)	1.52 (1.28)	**0.90 (1.00)**	0.97 (1.03)	1.25 (1.14)
07	1.01 (0.96)	1.03 (1.01)	1.29 (1.22)	0.98 (1.01)	**0.91 (1.00)**	1.19 (1.12)
08	1.11 (1.28)	1.11 (1.18)	1.75 (2.40)	1.16 (1.45)	**1.07 (1.24)**	2.55 (3.70)
09	**0.98 (1.00)**	1.04 (1.00)	1.22 (1.07)	1.00 (0.97)	1.03 (1.01)	1.23 (1.16)
10	**0.94 (1.03)**	1.05 (1.10)	1.47 (1.68)	0.99 (1.28)	0.98 (1.10)	1.15 (1.25)
$\emptyset_{\text{err}}$	**0.99 (1.09)**	1.03 (1.12)	1.31 (1.41)	0.99 (1.13)	0.99 (1.10)	1.29 (1.41)
$\emptyset_{\text{time}}$	110 s	192 s[4]	55/73 min[3]	**46 s**	104 s	64 min

[1] Ranked 1st on DIR-lab homepage; please note that the 300 freely available landmark correspondences, which are used throughout our paper, are only a subset of the landmarks used for evaluation by the DIR-lab.
[2] Ranked 2nd on DIR-lab homepage; see above note on the used landmark sets.
[3] Times for balanced flow; format: cases 01-05 / 06-10 (after additional preprocessing).
[4] Sum of mean times of grid search and filtering as given in table 5 of [5].

that employing lung masks in the registration process leads to invalid motion fields outside the specified region of interest. This is not acceptable for several applications, such as dose accumulation for abdominal structures during radiotherapy planning for lung tumors. Therefore, error values for registration without segmentation masks are as well of interest in the given application context.

The reported gain in performance due to the introduction of the census data term into TV-L$_1$-based medical image registration is also reflected by the values listed in Table 2 for the approach [9]. In [9], we describe an extension of the 'high accuracy optical flow' formulation of Brox et al. [2] for 3D medical image registration in combination with the census cost function; we refer to the corresponding paper for methodical details.

5 Conclusions

This paper reports about a 3D implementation of the 'duality-based approach for realtime TV-L $_1$ optical flow' in combination with the census cost function. This combination constitutes a novel contribution to 3D medical image registration in general and pulmonary motion estimation in 4D CT images in particular. The algorithm was evaluated on the DIR-lab benchmark for lung CT registration and

results demonstrate competitive performance compared to current state-of-the-art methods. Registration accuracy, computation times, as well as the robustness regarding registration without segmentation masks highlight the potential of the census cost function for medical image registration tasks.

Acknowledgments. The authors thank Dr. Clemens Rabe for a helpful discussion on the linearization of the census data residual, Prof. Dr. Reinhard Klette for helpful comments on early drafts and Dr. Gisela Klette for thoroughly proof reading this paper.

References

1. Brock, K.K.: On behalf of the Deformable Registration Accuracy Consortium: Results of a multi-institution deformable registration accuracy study (MIDRAS). Int. J. Radiat. Oncol. Biol. Phys. 76, 583–596 (2010)
2. Brox, T., Bruhn, A., Papenberg, N., Weickert, J.: High Accuracy Optical Flow Estimation Based on a Theory for Warping. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3024, pp. 25–36. Springer, Heidelberg (2004)
3. Castillo, R., Castillo, E., Guerra, R., Johnson, V.E., McPhail, T., et al.: A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. Phys. Med. Biol. 54, 1849–1870 (2009)
4. Castillo, R., Castillo, E., Martinez, J., Guerrero, T.: Ventilation from four-dimensional computed tomography: density versus Jacobian methods. Phys. Med. Biol. 55, 4661–4685 (2010)
5. Castillo, E., Castillo, R., White, B., Rojo, J., Guerrero, T.: Least median of squares filtering of locally optimal point matches for compressible flow image registration. Phys. Med. Biol. 57, 4827–4833 (2012)
6. Chambolle, A.: An Algorithm for Total Variation Minimization and Applications. J. Mathematical Imaging Vision 20, 89–97 (2004)
7. Ehrhardt, J., Lorenz, C.: 4D Modeling and Estimation of Respiratory Motion for Radiation Therapy. Springer, Berlin (2013)
8. Heinrich, M.P., Jenkinson, M., Brady, M., Schnabel, J.A.: MRF-based deformable registration and ventilation estimation of lung CT. IEEE Trans. Medical Imaging 32, 1239–1248 (2013)
9. Hermann, S., Werner, R.: High Accuracy Optical Flow for 3D Medical Image Registration using the Census Cost Function. In: Klette, R., Rivera, M., Satoh, S. (eds.) PSIVT 2013. LNCS, vol. 8333, pp. 23–35. Springer, Heidelberg (2013)
10. Hoog, A.C.B., Singh, T., Singla, P., Podgorsak, M.: Evaluation of advanced Lukas-Kanade optical flow on thoracic 4D-CT. J. Clin. Monit. Comput. 27, 433–441 (2013)
11. Hirschmüller, H.: Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information. In: IEEE Conf. Computer Vision Pattern Recognition, pp. 807–814 (2005)
12. Kabus, S., Lorenz, C.: Fast Elastic Image Registration. Medical Image Analysis for the Clinic: A Grand Challenge, 81–89 (2010)
13. Li, H., Li, Y., Zhang, X., Li, X., Liu, W., et al.: Dynamically accumulated dose and 4D accumulated dose for moving tumors. Med. Phys. 39, 7359–7367 (2012)

14. Müller, T., Rabe, C., Rannacher, J., Franke, U., Mester, R.: Illumination-Robust Dense Optical Flow Using Census Signatures. In: Mester, R., Felsberg, M. (eds.) DAGM 2011. LNCS, vol. 6835, pp. 236–245. Springer, Heidelberg (2011)
15. Murphy, K., van Ginneken, B., Reinhardt, J., Kabus, S., Ding, K., et al.: Evaluation of registration methods on thoracic CT: The EMPIRE10 Challenge. IEEE Trans. Medical Imaging 30, 1901–1920 (2011)
16. Pock, T., Urschler, M., Zach, C., Beichel, R.R., Bischof, H.: A Duality Based Algorithm for TV-L1-Optical-Flow Image Registration. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 511–518. Springer, Heidelberg (2007)
17. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D 60, 259–268 (1992)
18. Rühaak, J., Heldmann, S., Kipshagen, T., Fischer, B.: Highly Accurate Fast lung CT Registration. In: SPIE Medical Imaging: Image Processing. SPIE, vol. 8669, p. 86690Y (2013)
19. Schmidt-Richberg, A., Ehrhardt, J., Werner, R., Handels, H.: Diffeomorphic Diffusion Registration of Lung CT Images Medical Image Analysis for the Clinic: A Grand Challenge, pp. 55–62 (2010)
20. Schmidt-Richberg, A., Werner, R., Handels, H., Ehrhardt, J.: Estimation of slipping organ motion by registration with direction-dependent regularization. Med. Image Anal. 16, 150–159 (2012)
21. Simpson, D.R., Lawson, J.D., Nath, S.K., Rose, B.S., Mundt, A.J., Mell, L.K.: Utilization of advanced imaging technologies for target delineation in radiation oncology. J. Am. Coll. Radiol. 6, 876–883 (2009)
22. Urschler, M., Werlberger, M., Scheurer, E., Bischof, H.: Robust Optical Flow Based Deformable Registration of Thoracic CT Images. In: Medical Image Analysis for the Clinic: A Grand Challenge, pp. 195–204 (2010)
23. Warren, H.S.: Hacker's Delight, pp. 65–72. Addison-Wesley Longman, New York (2002)
24. Werner, R., Ehrhardt, J., Schmidt-Richberg, A., Albers, D., Frenzel, T., et al.: Towards accurate dose accumulation for Step-&-Shoot IMRT: Impact of weighting schemes and temporal image resolution on the estimation of dosimetric motion Eeffects. Z. Med. Phys. 22, 109–122 (2012)
25. Wedel, A., Pock, T., Zach, C., Bischof, H., Cremers, D.: An Improved Algorithm for TV-L1 Optical Flow. In: Cremers, D., Rosenhahn, B., Yuille, A.L., Schmidt, F.R. (eds.) Statistical and Geometrical Approaches to Visual Motion Analysis. LNCS, vol. 5604, pp. 23–45. Springer, Heidelberg (2009)
26. Zabih, R., Woodfill, J.: Non-parametric Local Transforms for Computing Visual Correspondence. In: Eklundh, J.-O. (ed.) ECCV 1994. LNCS, vol. 801, pp. 151–158. Springer, Heidelberg (1994)
27. Zach, C., Pock, T., Bischof, H.: A Duality Based Approach for Realtime TV-L1 Optical Flow. In: Hamprecht, F.A., Schnörr, C., Jähne, B. (eds.) DAGM 2007. LNCS, vol. 4713, pp. 214–223. Springer, Heidelberg (2007)