

High Accuracy Optical Flow for 3D Medical Image Registration Using the Census Cost Function

Simon Hermann^{1,2} and René Werner³

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. In 2004, Brox et al. described how to minimize an energy functional for dense 2D optical flow estimation that enforces both intensity and gradient constancy.

This paper presents a novel variant of their method, in which the census cost function is utilized in the data term instead of absolute intensity differences. The algorithm is applied to the task of pulmonary motion estimation in 3D computed tomography (CT) image sequences. The performance evaluation is based on DIR-lab benchmark data for lung CT registration. Results show that the presented algorithm can compete with current state-of-the-art methods in regards to both registration accuracy and run-time.

Keywords: Medical image registration, census cost function, pulmonary motion estimation, 4D CT, gradient constancy.

1 Introduction

Accurate motion estimation of anatomical and pathological structures is beneficial in many clinical applications. A typical scenario is radiation therapy of lung cancer patients, where the challenge is to apply a high radiation dose to the tumor while keeping the exposure to surrounding healthy tissue to a minimum. The planning process of the radiation treatment is usually based on one static 3D CT scan, which leads to breathing-induced uncertainties concerning position and shape of the tumor and surrounding structures.

With the introduction of 4D (3D + time) CT imaging protocols, it is possible to acquire a sequence of 3D lung CT images that captures the different breathing phases of the patient. Estimated motion patterns can be incorporated into the treatment planning process and the effects of lung movement on the dose distribution can be analyzed prior to the actual radiation treatment [19]. The reliability of the extracted information, such as dosimetric motion effects, depends to a great extent on the accuracy of the applied registration algorithm that recovers the motion information. In other words, there is a great need for accurate, robust and fast non-linear registration algorithms.

The objective of this paper is to show that the well known algorithm for ‘high accuracy optical flow estimation’, as proposed by Brox et al. [1], yields the potential to fulfill all of these requirements. It minimizes an L_1 energy in combination with a total variation (TV) regularization term. One of the most notable contributions of [1] is to include a gradient constancy assumption into the energy formulation, which is based on the gradient end point error (EPE). This is especially interesting because gradient information have only very recently been successfully exploited for pulmonary motion estimation [15,8]. However, the gradient-based constraints in [15,8] are entirely based on differences in gradient orientations. In other words, they estimate image similarities based on the gradient angular error and not, as presented in [1], on the EPE.

With the exception of [14], we are not aware of any literature that employs and evaluates the EPE in combination with the solution scheme of Brox et al. for the purpose of lung CT image registration. Unfortunately, the paper [14] contains only preliminary evaluation results, based on only one single 4D CT data set. Additionally, Heinrich et al. [7] presented a solution scheme to minimize a TV- L_1 energy similar to [1] – but without enforcing gradient constancy.

Considering application background and related literature, this paper describes two novel contributions.

First, a formulation for arbitrary dimensions of the algorithm following Brox et al. [1] is presented, which includes the census cost function in the data term. This is in line with Müller et al. [13], who introduced the census cost function for coupled convex approaches following the TV- L_1 solution scheme as proposed by Zach et al. [21].

Second, a straight forward 3D implementation of the presented algorithm is evaluated for pulmonary motion estimation in 4D CT data, with a focus on the quantification of the influence of the gradient constancy assumption and the census cost function on the registration accuracy.

The evaluation is based on DIR-lab data for 4D lung CT registration [3,4], a publicly available data base, which is commonly used to benchmark registration errors in the given application context.

The remainder of this paper is structured as follows. Section 2 introduces the energy formulation and the numerical scheme of the method, followed by implementation details of the 3D algorithm. The evaluation is described in Section 3, followed by a discussion in Section 4. Section 5 concludes this paper.

2 General Formulation of High Accuracy Optical Flow

The structure of this section follows Brox et al. [1]. However, the formulation of the energy itself and the resulting equation systems are derived for arbitrary dimensions. Furthermore, we replace the energy part that enforces intensity constancy with an arbitrary residual, similar to Müller et al., who introduced such a term for coupled convex optical flow approaches [13]. The basic notation is taken from [10].

2.1 Basic Notation

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 $I : \Omega \rightarrow \mathbb{R}$. 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 $\mathbf{u} : \Omega \rightarrow \mathbb{R}^{m \times n}$, with $\mathbf{u} = (u_1, \dots, u_m, 0_{m+1}, \dots, 0_n)$, that describes the displacement from the 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 $n = m = 3$.

2.2 The Energy Functional

The energy to be minimized is the weighted sum

$$E(\mathbf{u}) = E_D(\mathbf{u}) + \alpha E_S(\mathbf{u}) \quad (1)$$

where the data term is defined by

$$E_D(\mathbf{u}) = \int_{\Omega} \Gamma \left(\chi |\rho(\mathbf{x}, \mathbf{u})|^2 + \gamma |\nabla I_1(\mathbf{x} + \mathbf{u}) - \nabla I_0(\mathbf{x})|^2 \right) d\mathbf{x} \quad (2)$$

The term $\rho(\mathbf{x}, \mathbf{u})$ refers to a generic residual with $\rho := \rho_{I_0, I_1} : \Omega \times \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$, which is evaluated between I_0 and I_1 at position \mathbf{x} under consideration of $\mathbf{u}(\mathbf{x})$. The second term enforces gradient constancy by minimizing the end point error of the intensity gradient. Each part of the data term is coupled with a weighting factor (χ and γ). This makes it possible to omit the contribution of the individual data terms in order to evaluate their impact on the registration result separately. The smoothness term of the energy reads

$$E_S(\mathbf{u}) = \int_{\Omega} \Gamma \left(\sum_{b=1}^m |\nabla u_b(\mathbf{x})|^2 \right) d\mathbf{x} \quad (3)$$

The function Γ is given as $\Gamma(s^2) = \sqrt{s^2 + \varepsilon^2}$, with $\varepsilon = 0.0001$, which results in an approximate L_1 energy for both the data and the regularization term.

2.3 Euler-Lagrange Equations

With the set Θ of following abbreviations (for $a = 1, \dots, n$, and $b = 1, \dots, m$)

$$\rho_t := \rho(\mathbf{x}, \mathbf{u}) \quad (4)$$

$$\rho_b := \partial_b \rho(\mathbf{x}, \mathbf{u}) \quad (5)$$

$$I_{at} = I_{ta} := \partial_a I_1(\mathbf{x} + \mathbf{u}) - \partial_a I_0(\mathbf{x}) \quad (6)$$

$$I_{ab} = I_{ba} := \partial_{ab} I_1(\mathbf{x} + \mathbf{u}) \quad (7)$$

we get the m Euler-Lagrange equations for $b = 1, \dots, m$

$$0 \stackrel{!}{=} \mathcal{D} \cdot \left[\chi \rho_b \rho_t + \gamma \sum_{a=1}^n I_{ab} I_{at} \right] - \alpha \operatorname{div} [\mathcal{S} \cdot \nabla u_b(\mathbf{x})] \quad (8)$$

with the constant terms

$$\mathcal{D} = \Gamma' \left(\chi \rho_t^2 + \gamma \sum_{a=1}^n I_{at}^2 \right) \text{ and } \mathcal{S} = \Gamma' \left(\sum_{b=1}^m |\nabla u_b(\mathbf{x})|^2 \right) \quad (9)$$

for all m equations.

2.4 Numerical Approximation

The Euler-Lagrange equations from Eq. (9) are nonlinear in \mathbf{u} due to the non-quadratic penalizers and the non-linearized data constraints [2]. In order to linearize these equations, Brox et al. [1] employ two nested fixed point iterations. The outer fixed point iteration runs over the index k . With the iteration variable \mathbf{u}^k and $\mathbf{u}^0 = \mathbf{0}$, \mathbf{u}^{k+1} is, for $b = 1, \dots, m$, the solution of

$$0 \stackrel{!}{=} \mathcal{D}^{k+1} \cdot \left[\chi \rho_b^k \rho_t^{k+1} + \gamma \sum_{a=1}^n I_{ab}^k I_{at}^{k+1} \right] - \alpha \operatorname{div} [\mathcal{S}^{k+1} \cdot \nabla u_b^{k+1}(\mathbf{x})] \quad (10)$$

First order Taylor expansions are used to remove non-linearities in the data constraints. The Taylor expansions read for $a = 1, \dots, n$

$$\rho_t^{k+1} =: \mathcal{T}_t^k \approx \rho_t + \sum_{b=1}^m \rho_b^k du_b^k \quad (11)$$

$$I_{at}^{k+1} =: \mathcal{T}_{at}^k \approx I_{at} + \sum_{b=1}^m I_{ab}^k du_b^k \quad (12)$$

with $\mathbf{u}^{k+1} = \mathbf{u}^k + d\mathbf{u}^k$. The unknown \mathbf{u}^{k+1} is composed of the solution from the previous iteration step \mathbf{u}^k and unknown updates $d\mathbf{u}^k$. Let $\overline{\mathcal{D}}$ refer to the term \mathcal{D} of Eq. (9) with substitution of the respective Taylor approximations, we then get for $b = 1, \dots, m$

$$0 \stackrel{!}{=} \overline{\mathcal{D}}^k \cdot \left[\chi \rho_b^k \mathcal{T}_t^k + \gamma \sum_{a=1}^n I_{ab}^k \mathcal{T}_{at}^k \right] - \alpha \operatorname{div} [\mathcal{S}^k \cdot \nabla (u_b^k(\mathbf{x}) + du_b^k(\mathbf{x}))] \quad (13)$$

In order to remove the remaining non-linearity due to Γ' , an inner fixed point iteration is performed over the unknown flow updates $d\mathbf{u}^k$. Dropping the iteration index k for better readability and introducing a new iteration index l , the final linear system of equations reads for $b = 1, \dots, m$

$$0 \stackrel{!}{=} \overline{\mathcal{D}}^l \cdot \left[\chi \rho_b \mathcal{T}_t^{l+1} + \gamma \sum_{a=1}^n I_{ab} \mathcal{T}_{at}^{l+1} \right] - \alpha \operatorname{div} [\mathcal{S}^l \cdot \nabla (u_b(\mathbf{x}) + du_b^{l+1}(\mathbf{x}))] \quad (14)$$

The discretized linear system of equations is solved by successive over-relaxation (SOR), which is known for its fast convergence and suitability to be implemented on parallel hardware architectures [6,17].

2.5 The Census Cost Function as Data Residual

The census transform was introduced by Zabih and Woodfill [20] in 1994. It assigns a binary signature vector to an image position \mathbf{x} , which 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:

$$C[\mathbf{x}] = \left\{ \mathcal{T}[I(\mathbf{x}) \geq I(\mathbf{y})] \right\}_{\mathbf{y} \in \mathcal{N}_{\mathbf{x}}} \quad (15)$$

where $\mathcal{T}[\cdot]$ 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

$$\rho(\mathbf{x}, \mathbf{u}) = \Xi \{ C_0[\mathbf{x}] \oplus C_1[\mathbf{x} + \mathbf{u}] \} \quad (16)$$

where C_0 and C_1 refer to census signatures of images I_0 and I_1 , respectively. $\Xi\{\cdot\}$ is an operator that counts the 1's of the binary string in the argument. For efficiently counting of 1's in bit strings, we refer to Warren's 'Hackers Delight' [18].

2.6 Implementation Details

Before the input images are processed by the algorithm, their dynamic is mapped into a floating point intensity domain of $[0, 255]$, using 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 $S_0, S_1 : \Omega \rightarrow \{0, 1\}$ as optional input data. In case that such segmentation masks are provided, the input images are cropped accordingly and the evaluation of the data term is omitted where $S_0(\mathbf{x}) = 0$. For the specific application at hand, 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. It additionally speeds up the run-time of the algorithm.

The numerical scheme from section 2.4 is embedded into a coarse-to-fine approach, a common technique to overcome local minima. Image pyramids with arbitrary scale factors and with a fixed number of L pyramid levels are employed, where $\ell = 1$ refers to the finest and $\ell = L$ refers to the coarsest level. The pyramids are configured via target resolutions r_x, r_y, r_z for the coarsest level of each dimension. As a consequence, a scale factor is defined for each dimension based on the number L of pyramid levels and the respective target resolution. In case that the input images exhibit an anisotropic spatial resolution (CT data usually offer high in-plane resolution when compared to the slice thickness; cf. Section 3.1), the following common strategy is followed, see for example [15,12]: In a first step, images are scaled down only along the image axes with a high resolution until the pixel spacing is almost identical for all image dimensions.

```

input : Images  $I_0$  and  $I_1$ 
output: Displacement field  $\mathbf{u}$ 
Initialization of image pyramids based on  $L, r_x, r_y, r_z$ .
Set  $\mathbf{u}^L = \mathbf{0}$ 
for  $\ell \leftarrow L$  to 1 do
  for  $w \leftarrow 1$  to  $\eta$  do
    Set  $\mathbf{d}\mathbf{u}_w^\ell = \mathbf{0}$ 
    Warp  $I_1^\ell \xrightarrow{\mathbf{u}^\ell} I_0^\ell$ , by means of trilinear interpolation
    Initialize  $\Theta^\ell$  according to Eq.(4-7), using central differences
    for  $o \leftarrow 1$  to  $\vartheta$  do
      With  $\mathbf{u}^\ell$ ,  $\mathbf{d}\mathbf{u}_w^\ell$ , and  $\Theta^\ell$  calculate:  $\overline{\mathcal{D}}$  and  $\mathcal{S}$ 
      for  $i \leftarrow 1$  to  $\kappa$  do
        Perform SOR iteration [on Eq.(14)] to update  $\mathbf{d}\mathbf{u}_w^\ell$ 
      end
    end
    Update  $\mathbf{u}^\ell \leftarrow \mathbf{u}^\ell + \mathbf{d}\mathbf{u}_w^\ell$ 
    Smooth  $\mathbf{u}^\ell$  with a Gauss filter  $(\psi, \sigma)$ 
  end
  if  $\ell > 1$  then
    Initialize  $\mathbf{u}^{\ell-1}$  with  $\mathbf{u}^\ell$ 
  else
    Set output:  $\mathbf{u} = \mathbf{u}^1$ 
  end
end

```

Algorithm 1. Structure of the proposed algorithm

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.

Each level consists of a fixed number η of warps, of ϑ outer iterations and of κ SOR iterations. At the end of each warping iteration, the intermediate solution is filtered using a Gauss filter with a fixed window size ψ and a fixed σ . The scheme is summarized in Algorithm 1.

The implementation is C++-based, and the algorithm is executed on an Intel®Core™ i7 Quad-Core Processor with 2.4 GHz. OpenMP is employed to utilize hyper-threading.

3 Experiments

The performance of the algorithm was evaluated on ten thoracic 4D CT data sets, provided by the DIR-lab¹ of the University of Texas M.D. Anderson Cancer Center (Houston, USA) [3,4].

3.1 Image Data

Each 4D CT data set of the DIR-lab data pool consists of a sequence of 3D CT images acquired at ten different breathing phases. The dimension of the sequences varies between $256 \times 256 \times 94$ and $512 \times 512 \times 136$ pixel, the spatial resolution between $0.97 \times 0.97 \times 2.5$ mm³ and $1.16 \times 1.16 \times 2.5$ mm³. To be in line with most of the related publications, we focus on the registration of the end-inspiration (EI, reference image) and end-expiration (EE, target image) scans for evaluation purposes. These are the scan pairs with the largest motion amplitudes.

For each EI and EE scan pair, 300 anatomical landmark pairs within the lungs at prominent bifurcations of the bronchial or vessel trees have been annotated and are provided together with the CT data. The landmark pairs serve as ground truth information for quantitative evaluation of the registration accuracy. Therefore, let $\hat{\mathbf{x}}_{I_0}$ be a landmark position in the reference image, $\hat{\mathbf{x}}_{I_0, \mathbf{u}} := \hat{\mathbf{x}}_{I_0} + \mathbf{u}(\hat{\mathbf{x}}_{I_0})$ its transformed position, and $\hat{\mathbf{x}}_{I_1}$ the true landmark position in the target image. Following the DIR-lab, the registration error for that landmark is then computed as Euclidean distance between $\hat{\mathbf{x}}_{I_1}$ and the center $\hat{\mathbf{x}}_{I_0, \mathbf{u}}^\times$ of the pixel that is closest to $\hat{\mathbf{x}}_{I_0, \mathbf{u}}$. This strategy is referred to as ‘snap-to-pixel’ evaluation and ensures that the comparison of the ‘detected’ landmark positions $\hat{\mathbf{x}}_{I_1}$ (detected by a human observer) and $\hat{\mathbf{x}}_{I_0, \mathbf{u}}^\times$ (registration result) is based on positions defined on the same grid.

3.2 Algorithm Configuration and Evaluation Setup

For the evaluation, the following default parametrization of the algorithm is used. The census transform is calculated for a $5 \times 5 \times 5$ neighborhood on isotropic and $5 \times 5 \times 3$ on anisotropic levels. The motivation for reducing the kernel in z -direction on anisotropic levels is to compensate for the coarser spatial resolution along the z -axis of the DIR-lab data. Values calculated by ρ are scaled to the domain $[0, 21] \in \mathbb{R}$. Furthermore, the number of warps is set to $\eta = 7$, $\vartheta = 2$ outer iterations are employed and the equation system is solved with $\kappa = 4$ SOR iterations. The weights that control the contribution of each data term are set to $\chi = 1$ (for the census cost function) and $\gamma = 8.5$ (for the EPE). In order to exclude one of the energy terms from the registration, the respective weight is set to zero. The regularization term is weighted by $\alpha = 1.25$. The image pyramids are initialized with 16 levels and a target resolution of $r_x = 16, r_y = 16, r_z = 16$,

¹ <http://www.dir-lab.com>

Table 1. Published landmark-based registration error values for the 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 masked registration and combined use of census-based and EPE data term.

	Our approach	CCW+ [5] ¹	HSS+ [11] ²	HW [10]	RHK+ [15]	w / o registration
01	0.80 (0.92)	0.85 (1.00)	0.98 (1.00)	0.78 (0.92)	0.78 (0.91)	3.89 (2.78)
02	0.77 (0.92)	0.74 (0.99)	0.83 (1.02)	0.78 (0.92)	0.74 (0.87)	4.34 (3.90)
03	0.92 (1.10)	0.93 (1.07)	1.08 (1.15)	0.93 (1.09)	0.94 (1.07)	6.94 (4.05)
04	1.22 (1.24)	1.33 (1.51)	1.45 (1.53)	1.24 (1.30)	1.26 (1.26)	9.83 (4.86)
05	1.21 (1.47)	1.14 (1.25)	1.55 (1.75)	1.22 (1.43)	1.22 (1.48)	7.48 (5.51)
06	0.90 (1.00)	1.04 (1.05)	1.52 (1.28)	0.94 (0.99)	0.97 (1.03)	10.89 (6.97)
07	0.98 (1.01)	1.03 (1.01)	1.29 (1.22)	1.01 (0.96)	0.91 (1.00)	11.03 (7.43)
08	1.16 (1.45)	1.11 (1.18)	1.75 (2.40)	1.11 (1.28)	1.07 (1.24)	14.99 (9.01)
09	1.00 (0.97)	1.04 (1.00)	1.22 (1.07)	0.98 (1.00)	1.03 (1.01)	7.92 (3.98)
10	0.99 (1.28)	1.05 (1.10)	1.47 (1.68)	0.94 (1.03)	0.98 (1.10)	7.30 (6.35)
$\bar{\mathcal{O}}_{\text{err}}$	0.99 (1.13)	1.03 (1.12)	1.31 (1.41)	0.99 (1.09)	0.99 (1.10)	8.46 (6.58)
$\bar{\mathcal{O}}_{\text{time}}$	46 s	192 s ⁴	55/73 min ³	110 s	104 s	-

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].

Gauss smoothing of the flow field at the end of each warp iteration is performed with $(\psi, \sigma) = (5, 1)$.

The methodological approach for the evaluations is as follows. First, a baseline for an optimal performance is established when both data terms are employed. In order to quantify the individual impact of each data term, the algorithm is then evaluated once with $\chi = 0$ (EPE only), once with $\gamma = 0$ (census only). All setups are evaluated with and without lung segmentation masks supporting the registration process (cf. Section 2.6).

3.3 Results

The identified optimal baseline configuration, described in the previous section 3.2, leads to registration errors that are currently amongst the lowest published values on the DIR-lab benchmark. The results are listed in Table 1, along with a few of the best reported error values that can be found in recent literature. Minimum error values are marked by a gray background and bold letters. In case of two identical mean values, the minimum is determined by the standard deviation, found in brackets behind each error value. The rightmost column of Table 1 refers to the registration error when no registration was performed, i.e.

$\mathbf{u} = \mathbf{0}$. Those values give an indication of the mean motion amplitudes of the data sets. All those values correspond to the ‘snap-to-pixel evaluation’ concept and with lung masks supporting the registration.

It is, however, important to mention that although the application of binary lung masks leads to significantly better registration results, it also leads to invalid motion patterns outside the specified regions of interest. Although this is not acceptable in several clinical applications [16], error values for unmasked registration are only rarely found in literature, and are therefore not included in Table 1.

The results of the comparative evaluation of the impact of the individual data terms on the registration accuracy are summarized in Table 2. A few examples of the registration results are visualized in Figure 1. If we consider the registration results obtained with binary lung masks, we see that the algorithm with the isolated EPE data term outperforms the configuration with the isolated census data term – the combination of both data terms, however, performs better than any of them individually.

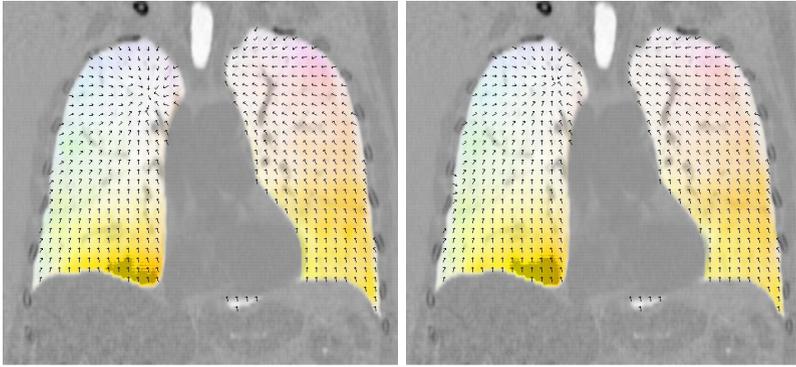
This observation is in partial disagreement to registration results obtained without lung segmentation masks. For this particular evaluation setup, only 60% of the error minima are scored by the combined data term configuration. Furthermore, the isolated census data term is often superior over the EPE data term. This is especially apparent for CT data that features large motion amplitudes (see w/o registration columns of Table 1), such as data set 08. The behavior seems to be reversed for data with only small motion amplitudes (data sets 01-03).

4 Discussion

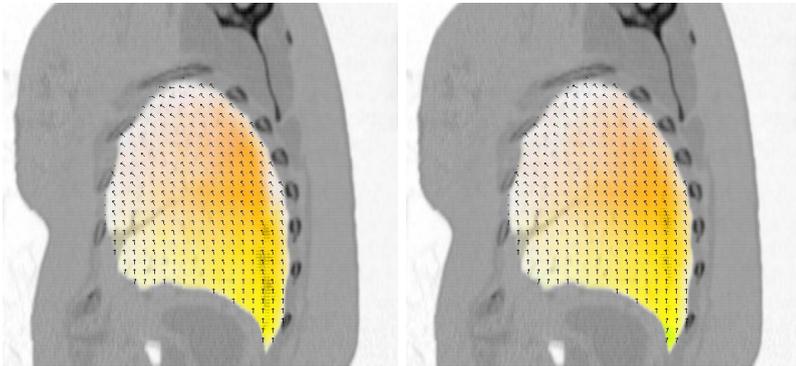
Comparison to State-of-the-art Approaches. If we compare the registration errors listed in Table 1, we see that the presented algorithm can compete with current state-of-the-art methods in the field of pulmonary motion estimation in 4D CT data. With an average run-time of 46 seconds per data set, we may consider the presented approach to be of interest for actual clinical applications that require close to real-time performance.

Comparison of Census and Gradient Data Term. The strong registration performance of a combination of census-based and EPE data term when using lung masks indicates that both terms provide, in general, complementary information when applied to pulmonary motion estimation. The interpretation of registration results obtained without lung masks is, however, more difficult.

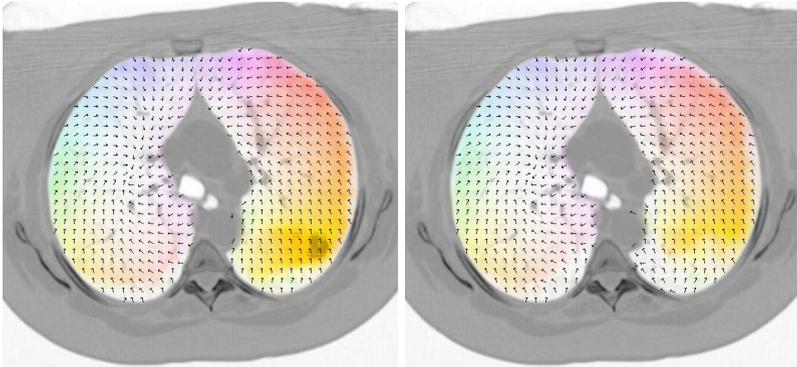
We assume that the EPE lacks robustness at strong motion discontinuities that occur at the transition of inner and outer lung structures (cf. Section 2.6), which is likely to have a major impact on registration errors of landmarks which are located close to lung borders. Providing lung segmentation masks is one way



(a) Coronal views of masked registration



(b) Sagittal views of masked registration



(c) Axial view of masked registration

Fig. 1. Illustration of registration results for the DIR-lab case 08, performing a masked registration. Left: Results of census-based registration. Right: Results of gradient-based registration; estimated motion between EI and EE; the color-coded visualization uses the hue color channel to encode orientation (also highlighted by normalized arrows) and the saturation channel to encode the motion magnitude. Rows from top to bottom: 2D views of the registration results in (a) coronal, (b) sagittal, and (c) axial views.

Table 2. Mean landmark distances and corresponding standard deviations after registration with the proposed algorithm variants. All values in mm.

#	with lung masks			w/o lung masks		
	Census	EPE	combined	Census	EPE	combined
01	0.87 (0.94)	0.76 (0.91)	0.80 (0.92)	0.86 (0.97)	0.79 (0.88)	0.81 (0.91)
02	1.00 (1.01)	0.79 (0.92)	0.77 (0.92)	0.96 (1.06)	0.81 (0.94)	0.80 (0.94)
03	1.14 (1.13)	0.94 (1.09)	0.92 (1.10)	1.19 (1.11)	0.94 (1.10)	0.93 (1.10)
04	1.42 (1.33)	1.25 (1.24)	1.22 (1.24)	1.49 (1.37)	1.26 (1.29)	1.25 (1.27)
05	1.38 (1.42)	1.25 (1.48)	1.21 (1.47)	1.33 (1.51)	1.42 (1.82)	1.28 (1.58)
06	1.25 (1.06)	0.92 (0.96)	0.90 (1.00)	1.36 (1.23)	1.43 (2.20)	1.19 (1.50)
07	1.29 (1.06)	1.00 (1.05)	0.98 (1.01)	1.74 (2.68)	3.64 (5.85)	3.03 (5.26)
08	1.34 (1.55)	1.55 (2.80)	1.16 (1.45)	2.62 (4.93)	5.39 (7.64)	3.52 (5.94)
09	1.23 (1.11)	1.01 (1.00)	1.00 (0.97)	1.26 (1.28)	1.20 (1.27)	1.16 (1.23)
10	1.19 (1.19)	1.01 (1.31)	0.99 (1.28)	1.29 (1.82)	1.35 (2.24)	1.52 (2.17)
\emptyset_{err}	1.21 (1.18)	1.05 (1.28)	0.99 (1.13)	1.41 (1.79)	1.82 (2.52)	1.52 (2.17)

to cope with the effects (although resulting motion fields are only valid inside the lungs; cf. Section 3.3). In case that no lung masks are applied, we assume that the EPE responds to the strong intensity gradients at the lung borders and the resulting high energy is propagated into the lungs by the regularization term.

Thus, in areas where the dominating intensity gradients of the lung borders and the motion patterns of the poorly contrasted inner lung structures are not aligned, larger registration errors occur for the EPE. It appears, that the ordinal character of the census cost function leads to a relatively robust performance, because it compensates, to a certain degree, the misbalance of the high intensity contrast at the lung borders and the low contrast of inner lung structures. This leads to very low registration errors for census-based unmasked registration, which is comparable to those reported by Heinrich et al. [9]. They employ a discrete method without lung segmentation information, that results in a mean error of 1.43 mm for the DIR-lab data.

To support this interpretation, we additionally consider the algorithms ALMI and STORM, which are described in another work by Heinrich et al. [8]. Both represent gradient orientation-based registration methods, which are also evaluated on DIR-lab data without the use of lung segmentation masks. Their results are very similar to the EPE-based registration results, i.e. we see that both methods perform excellent on data sets with small displacements, while they tend to fail when large motion amplitudes are present. We therefore assume that our gradient-based registration results, without the application of lung segmentation masks, are state-of-the-art – and consequently conclude that census-based registration must be very robust for this particular task.

5 Conclusions

This paper shows that the presented 3D variant of the ‘high accuracy optical flow’ algorithm is as accurate and fast as the currently best performing state-of-the-art methods. Its robust behavior that results from the inclusion of the census cost function was demonstrated based on registration results of unsegmented CT scans. We therefore conclude that this algorithm has the potential to meet the requirements for real-world clinical application.

Acknowledgments. The authors thank Dr. Gisela Klette for thoroughly proof reading this paper.

References

1. Brox, T., Bruhn, A., Papenberger, 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)
2. Brox, T.: From Pixels to Regions: Partial Differential Equations in Image Analysis. PhD thesis, Saarland University (2005)
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. Gwosdek, P., Bruhn, A., Weickert, J.: High Performance Parallel Optical Flow Algorithms on the Sony Playstation 3. In: Proc. Vision, Modeling, and Visualization - VMV, pp. 253–262 (2008)
7. Heinrich, M.P., Jenkinson, M., Brady, M., Schnabel, J.A.: Discontinuity preserving regularisation for variational optical-flow registration using the modified L_p norm. In: *Med. Image Anal. Clinic: A Grand Challenge*, pp. 185–194 (2010)
8. Heinrich, M.P., Jenkinson, M., Gleeson, F.V., Brady, M., Schnabel, J.A.: Deformable multimodal registration with gradient orientation based on structure tensors. *Annals of the BMVA*, 1–11 (2011)
9. 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)
10. Hermann, S., Werner, R.: TV- L_1 -based 3D Medical Image Registration with the Census Cost Function. In: Klette, R., Rivera, M., Satoh, S. (eds.) PSIVT 2013. LNCS, vol. 8333, pp. 149–161. Springer, Heidelberg (2014)
11. 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)
12. Kabus, S., Lorenz, C.: Fast Elastic Image Registration. *Medical Image Analysis for the Clinic: A Grand Challenge*, 81–89 (2010)

13. 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)
14. Negahdar, N., Amini, A.A.: A 3D Optical Flow Technique based on Mass Conservation for Deformable Motion Estimation from 4-D CT Images of the Lung. In: SPIE Medical Imaging 2012: Biomedical Applications in Molecular, Structural, and Functional Imaging. SPIE, vol. 8317, p. 83171F (2012)
15. 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)
16. 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)
17. Sundaram, N., Brox, T., Keutzer, K.: Dense Point Trajectories by GPU-Accelerated Large Displacement Optical Flow. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part I. LNCS, vol. 6311, pp. 438–451. Springer, Heidelberg (2010)
18. Warren, H.S.: *Hacker's Delight*, pp. 65–72. Addison-Wesley Longman, Amsterdam (2002)
19. 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 Effects. *Z. Med. Phys.* 22, 109–122 (2012)
20. 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)
21. 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)