

Evaluation of Scan-Line Optimization for 3D Medical Image Registration

Simon Hermann

Department of Computer Science,
Humboldt University of Berlin, Germany

Abstract

Scan-line optimization via cost accumulation has become very popular for stereo estimation in computer vision applications and is often combined with a semi-global cost integration strategy, known as SGM.

This paper introduces this combination as a general and effective optimization technique. It is the first time that this concept is applied to 3D medical image registration.

The presented algorithm, SGM-3D, employs a coarse-to-fine strategy and reduces the search space dimension for consecutive pyramid levels by a fixed linear rate. This allows it to handle large displacements to an extent that is required for clinical applications in high dimensional data.

SGM-3D is evaluated in context of pulmonary motion analysis on the recently extended DIR-lab benchmark that provides ten 4D computed tomography (CT) image data sets, as well as ten challenging 3D CT scan pairs from the COPDgene study archive. Results show that both registration errors as well as run-time performance are very competitive with current state-of-the-art methods.

1. Introduction

Accurate, robust, and run-time efficient non-linear image registration is a requirement for many clinical applications. For example, pulmonary motion estimation in 4D CT data has recently received much attention due to its great potential for breathing induced motion correction in radiation therapy [6]. Figure 1 illustrates the general application context of 3D lung CT registration that is used for evaluation purposes in this paper.

Many methods in this field follow a variational approach and often utilize prior knowledge, such as lung segmentation masks [1, 17, 16, 24, 22, 26], an initial solution from an affine-linear pre-registration [24], a sparse set of landmark pairs for initialization [22], or they incorporate a diffeomorphic motion assumption into the energy model [1, 26] and perform symmetric registration [1, 13]. Results of a recent challenge and study presented in [20] established a comprehensive overview of general state-of-the-art methods. More

details on current¹ state-of-the-art approaches, which serve as methods of comparison in this paper are given in the evaluation Section 4.3.

In contrast to variational registration schemes, discrete methods, which are formulated as an optimal labeling problem, constitute ‘an important new trend in medical image registration’ [10]. Recently published discrete approaches [13, 14, 9, 10] however, are exclusively based on Markov random fields (MRF) and utilize some sort of sparse graph node distribution, either by selecting statistically descriptive image locations or in form of a uniform sampling across the image domain. Those methods are targeted to update control points of a dense transformation model.

An alternative to MRF approaches is scan-line optimization, which was introduced in [25] as a simplified version of scan-line dynamic programming for the task of stereo estimation [21]. It has recently become very popular in combination with a semi-global integration strategy [18], known as semi-global stereo matching (SGM). Due to its computational efficiency and robustness, SGM is nowadays utilized in many industrial applications, such as commercial driver assistance systems [8].

Scan-line optimization is almost exclusively applied for disparity estimation in rectified stereo pairs, i.e. finding an optimal 1D displacement field in 2D images. However, a few methods address dynamic programming for 2D motion estimation, e.g. [23, 28], but the first method that was able to deal with reasonable large 2D displacement fields was proposed in [12] and was then extended for scene flow estimation in [11]. Still, both methods are restricted to deal with displacement vectors of only 25 and 10 pixels, respectively. To overcome even large pixel displacements on high resolution images, a recently published method called *fSGM* [15] embeds scan-line optimization for 2D optical flow estimation into a pyramidal scheme. At the time of publication, *fSGM* ranked second on the KITTI Vision Benchmark Suite² ahead of all submitted variational approaches.

This paper is motivated by the promising results pre-

¹ published within the last 12 months

² <http://www.cvlibs.net/datasets/kitti/>

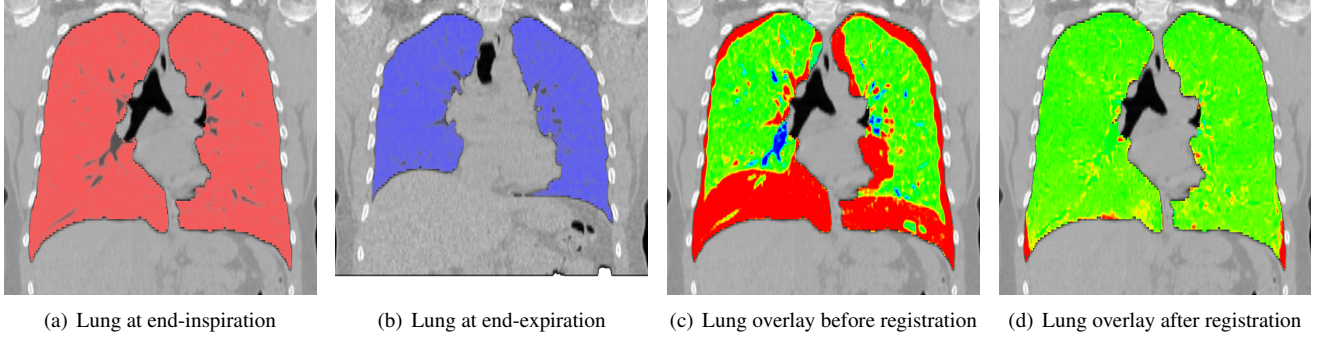


Figure 1. Illustration of 3D lung CT data in context of image registration for pulmonary motion analysis. Images 1(a) and 1(b) show coronal views of lung CT images at end-inspiration and end-expiration. Image 1(c) visualizes intensity differences due to unaligned lung structures when images 1(a) and 1(b) are overlaid. Those structures need to be registered via a non-linear transformation. The resulting 3D vector field is interpreted as the lung motion of an inhalation/exhalation cycle of the patient. Figure 1(d) shows the intensity differences after image registration with SGM-3D.

sented in [15] and introduces *SGM-3D*, a novel discrete optimization algorithm for dense 3D medical image registration. It employs the pyramidal concept of *fSGM*, but couples it with a linear search space reduction strategy. The evaluation in Section 4 shows that it currently ranks amongst the best performing algorithms with respect to both registration accuracy and run-time performance. This highlights the application potential of scan-line optimization for a clinical environment and indicates, that it can successfully be applied to general labeling problems in high dimensional data and is not restricted to stereo estimation.

This paper is structured as follows. Section 2 introduces scan-line optimization for image registration in n -dimensional data. Implementation details for the 3D medical case are given in Section 3, followed by a performance evaluation and discussion in Section 4. Section 5 concludes this paper.

2. Scan-Line Optimization

This section gives a formal and general introduction to pyramidal scan-line optimization in context of non-linear motion estimation. This technique is still uncommon for optical flow estimation in computer vision applications and a novelty in 3D medical image registration.

2.1. Basic Notations

Let I_r and I_t be two consecutive frames of an image sequence, defined on an n -dimensional discrete image grid $\Omega \subset \mathbb{N}^n$ with $I : \Omega \rightarrow \mathbb{R}$. I_r and I_t are referred to as *reference* and *target* image, with respective image domains Ω_r and Ω_t . A point $\mathbf{p} \in \Omega$ describes a position (or *pixel*) on the image grid and $I[\mathbf{p}]$ refers to the intensity of an image I at \mathbf{p} . The output is a pixel discrete n -dimensional vector

field

$$\mathbf{u} : \Omega_r \rightarrow \mathbb{Z}^n, \quad \mathbf{p} \mapsto \mathbf{u}[\mathbf{p}] \quad (1)$$

that describes the pixel displacement from I_r to I_t . For the task of 3D medical image registration it is $n = 3$.

2.2. Label Set and Search Space

Let $\mathbb{L} = \{0, \dots, l_{\max} - 1\} \subset \mathbb{N}$ be a set of labels of non-negative integers. Each label $l \in \mathbb{L}$ is associated with one unique displacement vector $\mathbf{d} \in \mathbb{Z}^n$. A unique correspondence between a label l and a displacement vector \mathbf{d} is defined by a bijective discrete mapping

$$\vartheta : \mathbb{L} \rightarrow \mathbb{S}, \quad l \mapsto \vartheta(l) = \mathbf{d} \quad (2)$$

and its inverse mapping

$$\vartheta^{-1} : \mathbb{S} \rightarrow \mathbb{L}, \quad \mathbf{d} \mapsto \vartheta^{-1}(\mathbf{d}) = l \quad (3)$$

where $\mathbb{S} \subset \mathbb{Z}^n$ is defined by a vector $\mathbf{f} \in \mathbb{N}^n$ with positive integer entries f_i as the finite offset domain of all possible displacements with

$$\mathbb{S} = \{\mathbf{d} \in \mathbb{Z}^n : |d_1| \leq f_1 \wedge, \dots, \wedge |d_n| \leq f_n\} \quad (4)$$

and is referred to as *search space*.

2.3. Pixel Matching Cost and Local Matching

To establish correspondences between pixels of reference and target image, a *cost function* ρ is required that defines the matching cost for a pixel pair $(\mathbf{p}, \mathbf{q}) \in \Omega_r \times \Omega_t$.

In practice, a cost function is used to establish a *cost matrix* that holds for each pixel $\mathbf{p} \in \Omega_r$ all matching costs of a finite set of pixels $\mathbf{q} \in \Omega_t$. In this paper, a generic cost matrix is defined as

$$C(\mathbf{p}, l) = \rho(\mathbf{p}, \mathbf{p} + \hat{\mathbf{u}}[\mathbf{p}] + \vartheta(l)) \quad (5)$$

where $\rho : \Omega_r \times \Omega_t \rightarrow \mathbb{R}$ assigns a matching cost between the two pixels $\mathbf{p} \in \Omega_r$ and $\mathbf{p} + \hat{\mathbf{u}}[\mathbf{p}] + \vartheta(l) \in \Omega_t$ that represents their dissimilarity.

A fixed initial motion field $\hat{\mathbf{u}}$ defines for every pixel $\mathbf{p} \in \Omega_r$ the origin in Ω_t around which all matchings costs within the search space \mathbb{S} are calculated via $\vartheta(l)$. An initial motion field $\hat{\mathbf{u}}$ refers in this paper to an upscaled scan-line optimization result from a lower pyramid level, but can of course be of arbitrary origin.

In the following, the notation $C_{\mathbf{p}, \hat{\mathbf{u}}}[l]$ is used for $C(\mathbf{p}, \hat{\mathbf{u}}, l)$ to highlight that a cost matrix $C_{\hat{\mathbf{u}}}$ holds for all pixels $\mathbf{p} \in \Omega_r$ a cost vector $C_{\mathbf{p}, \hat{\mathbf{u}}}$ with l_{\max} matching costs, which depend on a fixed prior solution $\hat{\mathbf{u}}$, and is indexed by labels $l \in \mathbb{L}$.

The displacement field \mathbf{u} that minimizes locally the pixel matching costs over the entire image domain is then defined for every pixel $\mathbf{p} \in \Omega_r$ as

$$\mathbf{u}[\mathbf{p}] = \hat{\mathbf{u}}[\mathbf{p}] + \vartheta\left(\arg \min_{k \in \mathbb{L}} \{C_{\mathbf{p}, \hat{\mathbf{u}}}[k]\}\right) \quad (6)$$

In words, the cost matrix establishes at each pixel $\mathbf{p} \in \Omega_r$ the matching cost for all displacement vectors, which are defined by the sum of an initial solution $\hat{\mathbf{u}}[\mathbf{p}]$ and the vector $\vartheta(l)$. The label $k \in \mathbb{L}$, that indexes the minimum matching cost in $C_{\mathbf{p}, \hat{\mathbf{u}}}$ defines the displacement vector $\vartheta(k)$ that updates the initial solution $\hat{\mathbf{u}}[\mathbf{p}]$.

2.4. Cost Accumulation and Integration

Let $C_{\mathbf{p}, \hat{\mathbf{u}}}$ and $C_{\mathbf{q}, \hat{\mathbf{u}}}$ be two cost vectors of length l_{\max} at adjacent pixels of a cost matrix $C_{\hat{\mathbf{u}}}$ calculated with respect to a cost function ρ . Energy minimization at an individual pixel location, as defined in Equation (6), does not enforce any consistency between adjacent displacement vectors $\mathbf{u}[\mathbf{p}]$ and $\mathbf{u}[\mathbf{q}]$.

Consider a new cost vector $S_{\mathbf{p}, \hat{\mathbf{u}}}$ defined by $S_{\mathbf{p}, \hat{\mathbf{u}}}[l] = C_{\mathbf{p}, \hat{\mathbf{u}}}[l] + C_{\mathbf{q}, \hat{\mathbf{u}}}[l]$, i.e. the sum of the correspondence costs of the current pixel \mathbf{p} and the neighboring pixel \mathbf{q} . Because $S_{\mathbf{p}, \hat{\mathbf{u}}}$ incorporates the matching costs of \mathbf{q} , the likelihood that $S_{\mathbf{p}, \hat{\mathbf{u}}}$ has its cost minimum at the same label as $C_{\mathbf{q}, \hat{\mathbf{u}}}$ is greater than for $C_{\mathbf{p}, \hat{\mathbf{u}}}$. In other words, the solutions for \mathbf{p} and \mathbf{q} are more likely to be consistent. The data costs at \mathbf{q} can therefore be interpreted as regularization costs for the pixel matching costs at \mathbf{p} .

Scan-line optimization via cost accumulation is based on that principle, i.e. data costs of previous pixels along a scan-line are considered to regularize data costs at the current pixel in order to enforce consistent solutions.

The general approach for scan-line optimization recursively defines an *accumulation matrix* $S_{\hat{\mathbf{u}}}$ of pixel matching costs, which are integrated along multiple 1D paths that run across the image domain Ω_r . The vector $S_{\mathbf{p}, \hat{\mathbf{u}}}$ holds the accumulated costs, which are used for regularization at the

subsequent pixel along the path. It is defined for a pixel location \mathbf{p}_i and all labels $l \in \mathbb{L}$ and is based on the scan-line segment $\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_i$, where \mathbf{p}_0 is located at the image border, and $\mathbf{p}_i \in \Omega_r$. The cost at $S_{\mathbf{p}_i, \hat{\mathbf{u}}}[l]$ is then recursively defined for $i = 1, 2, \dots, n$, as

$$S_{\mathbf{p}_i, \hat{\mathbf{u}}}[l] = C_{\mathbf{p}_i, \hat{\mathbf{u}}}[l] + \min_{\mathcal{T}} \left\{ \chi(S_{\mathbf{p}_{i-1}, \hat{\mathbf{u}}}, \phi)[l] - \min_{k \in \mathbb{L}} (S_{\mathbf{p}_{i-1}, \hat{\mathbf{u}}}[k]) \right\} \quad (7)$$

with $S_{\mathbf{p}_0, \hat{\mathbf{u}}}[l] = C_{\mathbf{p}_0, \hat{\mathbf{u}}}[l]$. χ is a generic cost restriction function that reduces the accumulated costs of the previous pixel, based on the spatial relationship between labels inside the search space domain \mathbb{S} .

The function $\phi : \mathbb{Z}^n \times \mathbb{L} \rightarrow \mathbb{L}_{\text{inv}}$ maps corresponding labels of matching costs for the same displacement vector. This is required in case that initial flow vectors differ for two consecutive pixels.

Let $\mathbb{L}_{\text{inv}} = \mathbb{L} \cup \{l_{\text{inv}}\}$, where l_{inv} is a unique label that refers to an invalid displacement. The function ϕ is then defined with Equations (2) and (3) and with $\mathbf{v} \in \mathbb{Z}^n$ as

$$\phi(\mathbf{v}, l) = \begin{cases} \vartheta^{-1}(\mathbf{v} + \vartheta(l)), & \text{if } \mathbf{v} + \vartheta(l) \in \mathbb{S} \\ l_{\text{inv}} & \text{otherwise} \end{cases} \quad (8)$$

where l_{inv} is associated with a high constant matching cost c_{inv} . The function ϕ is applied in Equation (13).

Generally speaking, scan-line accumulation is based on adding up two cost vectors. The regularization costs adjust the data costs at the current pixel, and the data costs update the regularization costs for subsequent processing.

The truncation threshold \mathcal{T} defines the maximum regularization cost. An additional cost restriction function χ imposes cost limits based on the proximity of labels within \mathbb{S} . This limits relative cost differences of spatially close labels. Subtracting the minimum cost value of the last pixel keeps regularization costs at a certain level and ensures that accumulation costs can either increase or decrease, depending on the data cost distribution.

2.4.1 Semi-Global Scan-Line Integration

This paper employs the cost integration strategy of SGM [18], which means that a global minimum is approximated in a semi-global fashion to enforce consistency not only along 1D scan-lines but over the entire (here 3D) image domain Ω_r . In SGM, accumulated costs of multiple scan-lines following different directions Δ are integrated into an *integration matrix*, which is defined for every $\mathbf{p} \in \Omega_r$ as

$$A_{\mathbf{p}, \hat{\mathbf{u}}}[l] = \sum_{\Delta} S_{\mathbf{p}, \hat{\mathbf{u}}}^{\Delta}[l] \quad (9)$$

where the angles between directions Δ are evenly spaced and it is ensured that pixel matching costs are accumulated in opposite directions for symmetry purposes. The final solution is established as in Equation (6) as

$$\mathbf{u}[\mathbf{p}] = \hat{\mathbf{u}}[\mathbf{p}] + \vartheta \left(\arg \min_{k \in \mathbb{L}} \{A_{\mathbf{p}, \hat{\mathbf{u}}}[k]\} \right) \quad (10)$$

In practice, the integration matrix is directly used as accumulation matrix, i.e. $A = S$, and accumulation directions are chosen to run at least along the image axes.

3. Implementation Details

The particular choices for cost function ρ and cost restriction function χ of Equations (5) and (7) are specified for the 3D medical image registration context.

3.1. The Census Cost Function

The census cost function is employed to establish pixel matching costs and is based on the census transform, which was introduced in [29]. The census transform is a binary representation of all intensity differences between a reference pixel and its immediate neighbourhood. It is based on ordering statistics; it encodes the spatial relationship between the considered pixels. A binary signature vector $\psi(\mathbf{x})$ is assigned to an image position \mathbf{x} and is calculated based on the ordinal characteristic of $I[\mathbf{x}]$ in relation to intensities within a defined neighbourhood $N_{\mathbf{x}}$ of \mathbf{x} . It is generated as follows:

$$\psi(\mathbf{x}) = \left(\Phi[I(\mathbf{x}) \geq I(\mathbf{y})] \right)_{\mathbf{y} \in N_{\mathbf{x}}} \in \{0, 1\}^{|N_{\mathbf{x}}|} \quad (11)$$

where $\Phi[\cdot]$ returns 1 if true, and 0 otherwise. The matching cost ρ is the Hamming distance of two signature vectors which are $\{0, 1\}$ -sequences of length $|N_{\mathbf{x}}|$. This distance is calculated as

$$\rho_{\text{census}}(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^{|N_{\mathbf{x}}|} \left(\psi_r(\mathbf{a}) \oplus \psi_t(\mathbf{b}) \right)_i \quad (12)$$

where \oplus refers to an ‘exclusive or’ operation (XOR) between two binary signatures. The final matching cost is the sum of all 1’s in the resulting signature vector. Its components are indexed by the subscript i .

3.2. Truncated Linear Regularization

A truncated linear cost restriction function is chosen for SGM-3D. The following efficient implementation is based on a distance transformation concept and was proposed by Felzenszwalb and Huttenlocher [7].

The cost vector $\chi(S_{\mathbf{p}_{i-1}, \hat{\mathbf{u}}}, \phi)[l]$ of Equation (7) is calculated for every $l \in \mathbb{L}$ in a forward and a backward pass as

follows:

$$S_{\mathbf{p}, \hat{\mathbf{u}}}^{\text{fw}}[\vartheta^{-1}((d_1, d_2, d_3))] = \min \begin{cases} S_{\mathbf{p}, \hat{\mathbf{u}}}[\phi(\hat{\mathbf{u}}[\mathbf{p}_i] - \hat{\mathbf{u}}[\mathbf{p}_{i-1}], \vartheta^{-1}((d_1, d_2, d_3)))] \\ S_{\mathbf{p}, \hat{\mathbf{u}}}^{\text{fw}}[\vartheta^{-1}((d_1 - 1, d_2, d_3))] + \mathcal{T}_{\lambda} \\ S_{\mathbf{p}, \hat{\mathbf{u}}}^{\text{fw}}[\vartheta^{-1}((d_1, d_2 - 1, d_3))] + \mathcal{T}_{\lambda} \\ S_{\mathbf{p}, \hat{\mathbf{u}}}^{\text{fw}}[\vartheta^{-1}((d_1, d_2, d_3 - 1))] + \mathcal{T}_{\lambda} \end{cases} \quad (13)$$

where d_i run from $-f_i + 1, \dots, f_i$, for $i = 1, 2, 3$. The backward pass is defined as:

$$S_{\mathbf{p}, \hat{\mathbf{u}}}^{\text{bw}}[\vartheta^{-1}((d_1, d_2, d_3))] = \min \begin{cases} S_{\mathbf{p}, \hat{\mathbf{u}}}^{\text{fw}}[\vartheta^{-1}((d_1, d_2, d_3))] \\ S_{\mathbf{p}, \hat{\mathbf{u}}}^{\text{bw}}[\vartheta^{-1}((d_1 + 1, d_2, d_3))] + \mathcal{T}_{\lambda} \\ S_{\mathbf{p}, \hat{\mathbf{u}}}^{\text{bw}}[\vartheta^{-1}((d_1, d_2 + 1, d_3))] + \mathcal{T}_{\lambda} \\ S_{\mathbf{p}, \hat{\mathbf{u}}}^{\text{bw}}[\vartheta^{-1}((d_1, d_2, d_3 + 1))] + \mathcal{T}_{\lambda} \end{cases} \quad (14)$$

where d_i run backwards from $f_i - 1, \dots, -f_i$. The implementation of $S_{\mathbf{p}_i, \hat{\mathbf{u}}}[l]$ of Equation (7) is

$$S_{\mathbf{p}_i, \hat{\mathbf{u}}}[l] = C_{\mathbf{p}_i, \hat{\mathbf{u}}}[l] + \min \begin{cases} S_{\mathbf{p}_{i-1}, \hat{\mathbf{u}}}^{\text{bw}}[l] - \min_{k \in \mathbb{L}} (S_{\mathbf{p}_{i-1}, \hat{\mathbf{u}}}[k]) \\ \mathcal{T}_{\lambda} \mathcal{T}_{\tau} \end{cases} \quad (15)$$

where \mathcal{T}_{λ} represents the slope of the linear restriction function χ and the factor \mathcal{T}_{τ} defines the upper limit of the regularization costs.

4. Performance Evaluation

The performance of SGM-3D is evaluated for all publicly available thoracic CT data sets, provided as benchmark data by the DIR-lab³ of The University of Texas M.D. Anderson Cancer Center (Houston, USA).

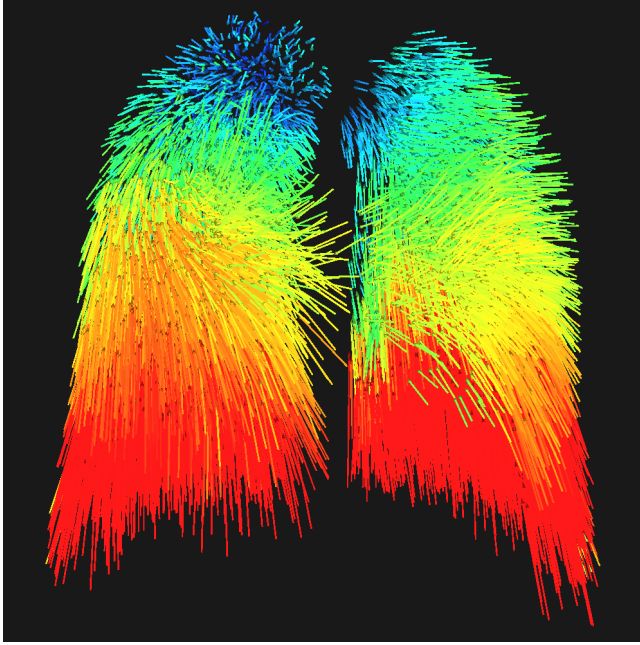
4.1. Dir-lab Benchmark Data

The DIR-lab provides currently two sets of benchmark data. The first set contains ten thoracic 4D CT images consisting of ten 3D CT scans, which are used for the treatment planning process of thoracic tumors at the Anderson Cancer Center of The University of Texas [5, 3]. The dimensions of the 4D data range from $256 \times 256 \times 94$ and $512 \times 512 \times 136$ pixels, with a resolution between $0.97 \times 0.97 \times 2.5 \text{ mm}^3$ and $1.16 \times 1.16 \times 2.5 \text{ mm}^3$.

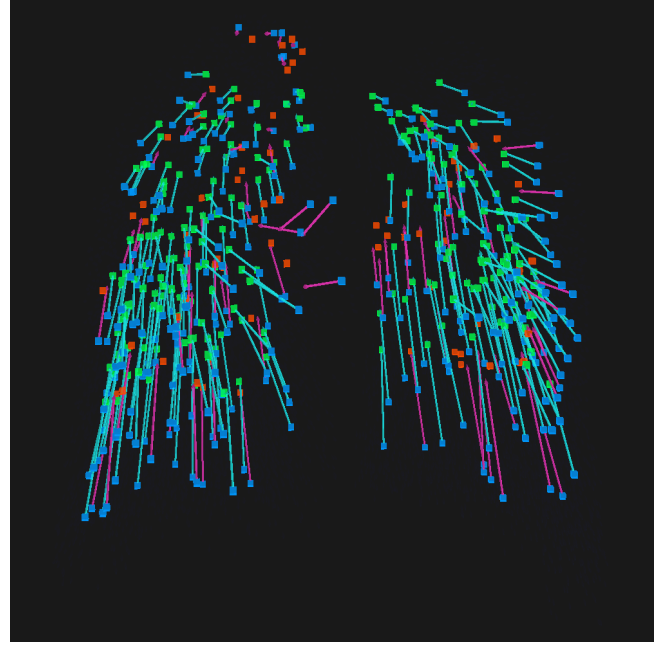
The second set originates from the National Heart Lung Blood Institute COPDgene study archive⁴ and contains ten

³ <http://www.dir-lab.com>

⁴ <http://www.copdgene.org>



(a) Visualization of the estimated 3D motion field



(b) Visualization of the 300 registered landmarks

Figure 2. Image 2(a) visualizes the 3D vector field obtained by SGM-3D for the COPD data set 04, which is the data set with the worst performance for SGM-3D. The color is defined by the hue color scale and ranges from red to blue, indicating large (≥ 42 mm) to small motion amplitudes. Image 2(b) shows the corresponding set of landmark pairs. Blue refers to a landmark of the reference image. The corresponding landmark in the target image is either colored in green or red, depending on whether the registration error is above (red) or below (green) an error threshold of 2.5mm.

inspiratory/expiratory breath-hold 3D CT image pairs [4]. The dimension of the image pairs are $512 \times 512 \times 102/135$ with a resolution between $0.586 \times 0.586 \times 2.5$ mm³ and $0.742 \times 0.742 \times 2.5$ mm³.

Each scan pair of the COPDgene study (in the following referred to as *COPD data*) and each end-inspiration/-expiration pair of the *4D CT data* sets comes with 300 publicly available anatomical landmark pairs to be used for quantitative evaluations. These landmarks were manually annotated by medical experts and are usually located at prominent bifurcations of the bronchial or vessel trees. The registration error is calculated as the Euclidean distance between a landmark pair in world coordinates. Figure 2 shows an example of a registration result and the corresponding set of landmark pairs.

However, there are two slightly different evaluation approaches in current literature. The first approach adds the corresponding 3D displacement vector to the 3D position of the reference landmark and calculates the Euclidean distance to the position of the target landmark. This straightforward approach is in the following referred to as *direct evaluation*. The DIR-lab shifts the translated reference landmark to the closest pixel center on the image grid before calculating the Euclidean distance. This *snap-to-pixel*

evaluation is based on the argument that human observers select discrete pixel locations in image pairs when identifying the landmark sets. This paper follows snap-to-pixel evaluation for the 4D CT data and direct evaluation for the COPD data. The reason for this is based on the fact that the best published results in current literature follow these approaches on respective data sets.

4.2. Algorithm Configuration

SGM-3D is a coarse-to-fine approach that employs Gauss pyramids with a fixed number of three pyramid levels, where $\ell = 1$ refers to the finest level. The displacement field at the coarsest level is initialized with zero, and the result of each level is used as initial solution for the scan-line optimization at the next higher image resolution.

The search space \mathbb{S} , as defined in Equation (4), represents a 3D cube with a dimension that is set to $(2 \cdot f_m + 1)^3$, with $f_1 = f_2 = f_3 = f_m$. In order to cope with the memory complexity, the search space is linearly decreasing for each pyramid level with $f_m = 3 \cdot \ell$.

To restrict the registration process to lung regions only, binary lung segmentation masks are generated to crop the input images and to set image intensities outside the lungs to zero. This is a common strategy to deal with strong motion

4D CT #	w / o registr.	observer error	SGM-3D [new approach]	cEPE [16]	cTVL1 [17]	NLR [24]	SWE [26]
01	3.89 (2.78)	0.85 (1.24)	0.76 (0.92)	0.80 (0.92)	0.78 (0.92)	0.78 (0.91)	0.87 (0.93)
02	4.34 (3.90)	0.70 (0.99)	0.72 (0.87)	0.77 (0.92)	0.78 (0.92)	0.74 (0.87)	0.84 (0.95)
03	6.94 (4.05)	0.77 (1.01)	0.94 (1.07)	0.92 (1.10)	0.93 (1.09)	0.94 (1.07)	1.02 (1.13)
04	9.83 (4.86)	1.13 (1.27)	1.24 (1.26)	1.22 (1.24)	1.24 (1.30)	1.26 (1.26)	1.35 (1.27)
05	7.48 (5.51)	0.92 (1.16)	1.15 (1.42)	1.21 (1.47)	1.22 (1.43)	1.22 (1.48)	1.39 (1.47)
06	10.89 (6.97)	0.97 (1.38)	0.90 (0.98)	0.90 (1.00)	0.94 (0.99)	0.97 (1.03)	1.25 (1.14)
07	11.03 (7.43)	0.81 (1.32)	0.89 (0.95)	0.98 (1.01)	1.01 (0.96)	0.91 (1.00)	1.19 (1.12)
08	14.99 (9.01)	1.03 (2.19)	1.13 (1.40)	1.16 (1.45)	1.11 (1.28)	1.07 (1.24)	2.55 (3.70)
09	7.92 (3.98)	0.75 (1.09)	0.91 (0.93)	1.00 (0.97)	0.98 (1.00)	1.03 (1.01)	1.23 (1.16)
10	7.30 (6.35)	0.86 (1.45)	0.83 (0.92)	0.99 (1.28)	0.94 (1.03)	0.98 (1.10)	1.15 (1.25)
\emptyset_{err}	8.46 (6.58)	0.88 (1.31)	0.95 (1.07)	0.99 (1.13)	0.99 (1.09)	0.99 (1.10)	1.29 (1.41)
\emptyset_{time}	–	–	98 s	46 s	110 s	104 s	64 min

Table 1. A list of the currently lowest published landmark-based mean registration errors after registration of the end-inspiration to end-expiration images of the DIR-lab 4D CT data sets. Standard deviations are given in small brackets. All values are in mm, following the snap-to-pixel evaluation.

discontinuities that often occur near lung borders.

The data resolution specifications in Section 4.1 highlight that CT data usually features a high spatial resolution along the x- and y-axis but often has a significantly lower resolution along the z-axis. To deal with the anisotropic nature of the data, a census neighborhood of $5 \times 5 \times 3$ is used to reduce the spatial impact of the data term along the z-axis. Additionally, the COPD data is scaled down by 50% along the x- and y-axis as pixel resolution differences are particularly high for this data set.

The result of each level is first filtered by a $3 \times 3 \times 3$ median filter, followed by a $5 \times 5 \times 5$ Gauss filter with $\sigma = 1.0$.

The parameters that define the truncated linear cost regularization are set to $\mathcal{T}_\lambda = 14.8$ and $\mathcal{T}_\tau = 14.8$ and six scan-line directions along the image axes were chosen.

4.3. Result Tables and Methods of Comparison

Tables 1 and 2 list mean registration errors of SGM-3D for each data set of the 4D CT benchmark and the COPD benchmark, along with some of the lowest reported error values in current literature. The minimum error value for each data set is highlighted with bold letters and a gray background. In case of two identical mean values, the minimum is identified by the standard deviation, which is given in brackets behind each error value.

The mean landmark distances, which are obtained without registration, are listed in the first column and give an indication of the mean motion amplitudes for each data set. The larger those values, the bigger the assumed lung deformation and the harder the registration task.

The observer error in the second column of the tables is provided by the DIR-lab website⁵ and describes the mean landmark deviation that resulted from multiple annotations of a defined subset of landmarks by three different human observers. The registration goal is to get close or below this error value.

In the following, the methods of comparison are described. *cTVL1* [17] and *cEPE* [16] refer to methods, which minimize an L_1 energy and employ total variation regularization following the numerical schemes of Zach et al. [30] (*cTVL1*) and Brox et al. [2] (*cEPE*). Both methods utilize the census cost function in the data term. The error values of *SWE* [26] are taken from [17]. It is a diffeomorphic variational scheme that minimizes a normalized variant of the sum of squared intensity differences and employs diffusion regularization. The method currently ranks amongst the top ten methods on the EMPIRE10 website⁶, which started off as a pulmonary image registration challenge in 2010 [20], and continues as another benchmark for performance evaluation in that field.

The method named *NLR* currently ranks fourth on EMPIRE10. It is described in [24] as a variational approach, which is based on second order regularization and a data term that penalizes deviations of gradient orientations. It employs affine linear transformation prior to non-linear registration. The method *LMP* [22] can be considered as a straight forward extension to *NLR*. It generates a pre-registration based on a sparse set of thin plate splines

⁵ <http://www.dir-lab.com/ReferenceData.html>

⁶ <http://empire10.isi.uu.nl>

COPD #	w / o registr.	observer error	SGM-3D [new approach]	NLR [24, 22]	gsyn [22, 27, 1]	LMP [22]	TPS [22]
01	26.33 (11.44)	0.65 (0.73)	1.22 (2.73)	1.39 (1.40)	1.21 (1.36)	1.26 (1.23)	3.69 (3.79)
02	21.79 (6.47)	0.70 (0.99)	2.48 (3.79)	2.36 (2.79)	3.01 (4.46)	2.02 (2.29)	4.10 (3.60)
03	12.64 (6.40)	0.58 (0.87)	1.01 (0.93)	1.18 (0.81)	1.24 (1.08)	1.14 (0.70)	1.56 (1.07)
04	29.58 (12.95)	0.71 (0.96)	2.42 (3.56)	1.57 (1.39)	1.38 (1.14)	1.62 (1.60)	4.39 (3.89)
05	30.08 (13.36)	0.65 (0.87)	1.93 (3.24)	1.44 (1.14)	1.31 (1.19)	1.47 (1.26)	3.63 (3.31)
06	28.46 (9.17)	1.06 (2.38)	1.45 (2.42)	2.08 (2.92)	1.49 (2.25)	1.38 (1.46)	2.69 (2.94)
07	21.60 (7.74)	0.65 (0.78)	1.05 (1.43)	1.18 (1.13)	1.24 (1.24)	1.22 (1.30)	2.25 (2.18)
08	26.46 (13.24)	0.96 (3.07)	1.16 (1.79)	1.65 (1.98)	2.09 (3.32)	1.63 (2.16)	3.47 (3.76)
09	14.86 (9.82)	1.01 (2.54)	0.81 (0.67)	1.13 (1.09)	1.18 (1.25)	1.12 (1.14)	2.20 (2.34)
10	21.81 (10.51)	0.87 (1.65)	1.28 (1.29)	1.44 (1.23)	1.63 (2.05)	1.45 (1.31)	4.00 (3.16)
\emptyset_{err}	23.36 (10.11)	0.82 (1.54)	1.48 (2.19)	1.54 (1.59)	1.58 (1.93)	1.43 (1.45)	3.20 (3.00)

Table 2. A list of the currently lowest published mean landmark-based registration errors after registration of the breath-hold pairs of the DIR-lab COPD data sets. Standard deviations are given in small brackets. All values are in mm, following the direct evaluation. The values for comparison are taken from the study [22].

following [19] and uses those pre-calculated landmarks as additional constraint for the non-linear registration step. Error values for *TPS* refer to the result without non-linear registration.

gsyn [27] refers to a diffeomorphic variational method that uses cross-correlation and regularization via Gauss smoothing within a symmetric registration framework. It ranks first on EMPIRE10 since the initial challenge phase in 2010. Comparative values for this method were generated by Polzin et al. for the study in [22], using the publicly available ANTs⁷ library with the algorithm configuration reported for EMPIRE10 in [27]. It is so far the only work with published result values for the challenging COPD benchmark data.

4.4. Discussion for 4D CT Data

Comparing mean registration errors of Table 1, we see that there are no big differences between the listed methods, not in terms of registration accuracy and, with the exception of SWE, not in terms of run-time performance.

Considering that the majority of these methods are only 0.1 mm away from the mean observer error, the challenge of this benchmark may be considered as close to being solved. However, it is an important reference benchmark in current literature and therefore included in this paper. It will stay relevant as a baseline benchmark, and it is safe to assume that registration results for the COPD benchmark will soon become the next challenge for performance evaluation in pulmonary image registration.

⁷ <http://stnava.github.io/ANTs/>

4.5. Discussion for COPD Data

Initial mean registration errors, which are listed in the first column of Tables 1 and 2, indicate that displacement vectors of the COPD benchmark are on average three times larger than those of the 4D CT benchmark. The COPD data is therefore significantly more challenging than the data of the current 4D CT reference benchmark.

All listed methods perform very good in general, with the exception of *TPS*, which was listed in [22] only as a baseline result. *SGM-3D* in particular yields very good performance on most COPD data sets. However, registration errors for data sets 04 and 05, which exhibit the largest motion amplitudes, are significantly higher compared to all other variational based methods. This is likely to be due to the resulting larger deformations of the lung structures. Variational methods based on image warping have in this case a natural advantage over matching based methods. This argument is supported by the fact that *gsyn*, which employs a symmetric registration approach, outperforms *NLR* and *LMP* in those particular cases.

5. Conclusions

This paper presented SGM-3D, the first method based on scan-line optimization that was successfully applied to pulmonary motion estimation in context of 3D medical image registration. SGM-3D is competitive with current state-of-the-art methods, even on the challenging and only recently released COPD benchmark. This result indicates further, that scan-line optimization has the potential to be applied to general labeling problems in high dimensional data.

Acknowledgements. I thank Dr. René Werner for introducing me to the exciting field of 3D medical image registration.

References

- [1] B. B. Avants, C. L. Epstein, M. Grossmann, and J. C. Gee. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1):26–41, 2008.
- [2] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High Accuracy Optical Flow Estimation Based on a Theory for Warping. In *Proc. European Conf. on Computer Vision (ECCV)*, volume 3024 of *LNCS*, pages 25–36, Berlin, Heidelberg, 2004. Springer-Verlag.
- [3] E. Castillo, R. Castillo, J. Martinez, et al.. Four-dimensional deformable image registration using trajectory modeling. *Physics in Medicine and Biology*, 55(1):305–327, 2009.
- [4] R. Castillo, E. Castillo, D. Fuentes, et al.. A reference dataset for deformable image registration spatial accuracy evaluation using the COPDgene study archive. *Physics in Medicine and Biology*, 58(9):2861–2877, 2013.
- [5] R. Castillo, E. Castillo, R. Guerra, et al.. A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Physics in Medicine and Biology*, 54(7):1849–1870, 2009.
- [6] J. Ehrhardt and C. Lorenz. *4D Modeling and Estimation of Respiratory Motion for Radiation Therapy*. Springer-Verlag Berlin Heidelberg, 2013.
- [7] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient Belief Propagation for Early Vision. *Int. J. of Computer Vision*, 70(1):41–54, 2006.
- [8] S. Gehrig, F. Eberli, and T. Meyer. A Real-Time Low-Power Stereo Vision Engine Using Semi-Global Matching. In *Computer Vision Systems*, volume 5815 of *LNCS*, pages 134–143. Springer-Verlag Berlin Heidelberg, 2009.
- [9] B. Glocker, N. Komodakis, G. Tziritas, et al.. Dense image registration through MRFs and efficient linear programming. *Medical Image Analysis*, 12(6):731–741, 2008.
- [10] B. Glocker, A. Sotiras, N. Komodakis, and N. Paragios. Deformable Medical Image Registration: Setting the State of the Art with Discrete Methods. *Annual Review of Biomedical Engineering*, 13:219–244, 2011.
- [11] M. Gong. Real-time joint disparity and disparity flow estimation on programmable graphics hardware. *Computer Vision and Image Understanding*, 113(1):90–100, 2009.
- [12] M. Gong and Y.-H. Yang. Estimate Large Motions Using the Reliability-Based Motion Estimation Algorithm. *Int. J. of Computer Vision*, 68(3):319–330, 2006.
- [13] M. P. Heinrich, M. Jenkinson, M. Brady, and J. A. Schnabel. Mrf-based deformable registration and ventilation estimation of lung ct. *IEEE Transactions on Medical Imaging*, 32(7):1239–1248, 2013.
- [14] M. P. Heinrich, M. Jenkinson, B. W. Papiez, et al.. Edge- and Detail-Preserving Sparse Image Representations for Deformable Registration of Chest MRI and CT Volumes. In *Information Processing in Medical Imaging*, volume 7917 of *LNCS*, pages 463–474. Springer-Verlag Berlin Heidelberg, 2013.
- [15] S. Hermann and R. Klette. Hierarchical Scan-Line Dynamic Programming for Optical Flow Using Semi-Global Matching. In *ACCV 2012 Workshops*, volume 7729 of *LNCS*, pages 556–567, 2013.
- [16] S. Hermann and R. Werner. High Accuracy Optical Flow for 3D Medical Image Registration using the Census Cost Function. In *6th Pacific-Rim Symposium on Image and Video Technology*, 2013.
- [17] S. Hermann and R. Werner. TV-L1-based 3D Medical Image Registration with the Census Cost Function. In *6th Pacific-Rim Symposium on Image and Video Technology*, 2013.
- [18] H. Hirschmüller. Stereo Processing by Semiglobal Matching and Mutual Information. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 30(2):328–341, February 2008.
- [19] J. Modersitzki. *FAIR: Flexible Algorithms for Image Registration*, volume 6. SIAM, 2009.
- [20] K. Murphy, B. van Ginneken, J. Reinhardt, et al.. Evaluation of Registration Methods on Thoracic CT: The EMPIRE10 Challenge. *IEEE Transactions on Medical Imaging*, 30(11):1901–1920, 2011.
- [21] Y. Ohta and T. Kanade. STEREO BY TWO-LEVEL DYNAMIC PROGRAMMING. In *9th Int. Joint Conf. on Artificial Intelligence*, pages 1120–1126, 1985.
- [22] T. Polzin, J. Rühaak, R. Werner, J. Strehlow, S. Heldmann, H. Handels, and J. Modersitzki. Combining Automatic Landmark Detection and Variational Methods for Lung CT Registration. In *Fifth International Workshop on Pulmonary Image Analysis*, 2013.
- [23] G. M. Quénot. Computation of Optical Flow using Dynamic Programming. In *IAPR Workshop on Machine Vision Applications*, pages 249–252, 1996.
- [24] J. Rühaak, S. Heldmann, T. Kipshagen, and B. Fischer. Highly Accurate Fast Lung CT Registration. In *SPIE Medical Imaging: Image Processing*, volume 8669, pages 86690Y–86690Y–9, 2013.
- [25] D. Scharstein and R. Szeliski. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *Int. J. of Computer Vision*, 47(1-3):7–42, 2002.
- [26] A. Schmidt-Richberg, J. Ehrhardt, R. Werner, and H. Handels. Diffeomorphic Diffusion Registration of Lung CT Images. In *Medical Image Analysis for the Clinic: A Grand Challenge, MICCAI 2010*, pages 55–62, 2010.
- [27] G. Song, N. Tustison, B. B. Avants, and J. C. Gee. Lung ct image registration using diffeomorphic transformation models. In *Medical Image Analysis for the Clinic: A Grand Challenge, MICCAI 2010*, pages 23–32, 2010.
- [28] C. Sun. Fast optical flow using 3d shortest path techniques. *Image and Vision Computing*, 20(13):981–991, 2002.
- [29] R. Zabih and J. Woodfill. Non-parametric Local Transforms for Computing Visual Correspondence. In *Proc. European Conf. on Computer Vision (ECCV)*, volume 801 of *LNCS*, pages 151–158, Berlin, Heidelberg, 1994. Springer-Verlag.
- [30] C. Zach, T. Pock, and H. Bischof. A Duality Based Approach for Realtime TV-L1 Optical Flow. In *Proc. DAGM (Pattern Recognition)*, volume 4713 of *LNCS*, pages 214–223, Berlin Heidelberg, 2007. Springer Verlag.