Quantification of Smoothing Requirement for 3D Optic Flow Calculation of Volumetric Images

Alireza Bab-Hadiashar, *Senior Member, IEEE,* Ruwan B. Tennakoon, *Member, IEEE,* and Marleen de Bruijne

***Abstract*— Complexities of dynamic volumetric imaging challenge the available computer vision techniques on a number of different fronts. This paper examines the relationship between the estimation accuracy and required amount of smoothness for a general solution from a robust statistics perspective. We show that a (surprisingly) small amount of local smoothing is required to satisfy both the necessary and sufficient conditions for accurate optic flow estimation. This notion is called "just enough" smoothing, and its proper implementation has a profound effect on the preservation of local information in processing 3D dynamic scans. To demonstrate the effect of "just enough" smoothing, a robust 3D optic flow method with quantized local smoothing is presented, and the effect of local smoothing on the accuracy of motion estimation in dynamic lung CT images is examined using both synthetic and real image sequences with ground truth.**

***Index Terms*— 3D optic flow, 4-D CT, Gaussian smoothing, volumetric images.**

I. Introduction

MOTION estimation is one of the most crucial and well-studied problems of computer vision. The underlying task is very general and has a wide range of applications. In the context of medical imaging, the estimation of motion has received substantial attention. With the advance of 3D dynamic imaging by MR, CT and ultrasound, motion estimation has become important in diagnostics, for instance to assess localized abnormalities in heart wall motion or vessel distensibility, as well as in radiation therapy planning and to compensate for soft tissue motion during image guided interventions.

From the beginning, two distinct approaches to estimation of apparent motion (or optic flow) have emerged. The first approach, described by Horn & Schunck [1], views the estimation as a *global* optimization problem and attempts to find globally smooth warping regimes that relate sequential images to each other. The second approach, presented by Lucas & Kanade [2], views the problem as a *local* correspondence problem.

In both of the above approaches, smoothing plays a crucial role. The optic flow, due to the well-known aperture problem, is ill-posed and cannot be solved for a single data point (a pixel in 2D or a voxel in 3D). Therefore, some degree of smoothing or regularization is always required.

From a historical perspective, the *global* and *local* approaches were conceptually much closer at the beginning (in late 70s and early 80s) than they are now. A typical test image sequence of the time consisted mostly of a few thousands gray level pixels (small digitized TV signal) of usually a single flat moving object (see the Result section of [1]). Therefore, it is important to interpret their assertion that apparent velocity "varies smoothly almost everywhere in the image" [1] in its context which is very different to today's concept of an image sequence (typically containing a large amount of detail including several motions).

It is also important to note here that, although some excellent results have been produced by refining global methods [3]–[5] and the accuracy of optical flow estimation, as measured by Middlebury benchmark [6], has been improving, the effect of local smoothing on the estimation accuracy is yet to be fully understood. To examine the underlying cause of success of those methods, a baseline method similar to Horn & Schunck [1] was used in [7] to study the influence of different choices about how to model an appropriate objective function and its approximation (for computational tractability) and optimization on the overall accuracy. Their comparisons showed that applying a median filter to optical flow estimates in different iterations of those algorithms produced the most significant improvements. Although this is a form of local smoothing, a theory on how much spatial smoothing is required, at a given scale, is yet to be developed. In particular, the above study does not consider the effect of smoothing imposed by Lucas & Kanade [2] type formulation of the optical flow problem. This paper is an attempt to address this important question by examining how much smoothing would be sufficient from a local perspective. In this context, [8] addresses the question of how to merge global and local approaches while we aim to answer the question of where the meeting should take place. In other word, the desire is to study the concept of *just enough* smoothing: The least amount of smoothing that both overcomes the ill-posed nature of this

Manuscript received August 23, 2011; revised December 23, 2012; accepted December 30, 2012. Date of publication February 11, 2013; date of current version March 29, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jose M. Bioucas-Dias.

A. Bab-Hadiashar is with the School of Aerospace, Mechanical and Manufacturing, RMIT University, Melbourne 3001, Australia (e-mail: abh@rmit.edu.au).

R. B. Tennakoon is with the Faculty of Engineering and Industrial Sciences, Swinburne University of Technology, Hawthorn 3122, Australia (e-mail: rtennakoon@swin.edu.au).

M. de Bruijne is with the Department of Computer Science, University of Copenhagen, Copenhagen DK-2100, Denmark, and also with the Departments of Radiology and Medical Informatics, Biomedical Imaging Group, Erasmus MC - University Medical Center, Rotterdam 12306, The Netherlands (e-mail: marleen@diku.dk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2013.2246174

1057-7149/$31.00 © 2013 IEEE

problem and ensures the desired accuracy of the estimation process.

In 2D optical flow estimation, the work of Xu et al. [9] has shown that the imposition of global smoothing can be partial toward global changes and would highlight those changes (motions) at the expense of localized variations. In volumetric images, where there is no 3D to 2D projection, the bias introduced by imposing more than just enough smoothness is particularly undesirable. For example, in 4D chest CT scans the lungs display a complex deformation pattern during the respiratory cycle, with motion boundaries where the lungs slide along the rib cage and lung lobes move relative to each other. The elastic characteristics of airways and large blood vessels differ from those of lung parenchyma, and pathologic tissue such as pulmonary nodules can be expected to deform differently from its surroundings. Global smoothing could lead to errors in the estimate of tumor motion, and thus to inaccuracies in derived treatment plans.

In cases where some prior knowledge does exist, this can simply be included in the motion models. However, the analysis presented here is based on the distribution of residuals near a motion boundary and the result does not depend upon the type of models used.

The concept of *just enough* smoothing has, in the past, been considered in the context of scale space theory and in particular, for finding the appropriate scale to terminate a multiscale hierarchical algorithm [10]. The scale space approach however differs from the robust statistics approach presented in this paper in a fundamental way. In the scale space approach, the discontinuities (motion boundaries) are modeled by transitions toward higher scales in which small discontinuities, except the main motion, are smooth enough to be considered continuous. The problem is then solved at that scale and the solution is then propagated to the finer scales to recover smaller motions. The question of *just enough* smoothing in the above therefore, refers to finding the scale at which the estimation is most reliable.

In robust statistics, the discontinuities are explicitly modeled as separate instances of a single model (or a finite set of plausible models) and the scale of each instance and the shape of the smoothing window are estimated concurrently with the model parameters [11]. In this context, just enough smoothing refers to the size of locality (size of population, bandwidth, etc depending on the type of estimator) where the estimation is based upon. To our knowledge, the relationship between the amount of smoothing and accuracy of the estimation has thus far not been quantified (either by analysis or experiment) and more importantly no specific link between the amount of sufficient smoothing and the accuracy of estimation has yet been established.

Comparison of existing results for calculation of 2D optic flow reveals that for real images with discontinuous flow (like Otte image sequence [12] - in contrast to synthetic images like Yosemite sequence with fairly smooth flow), the local robust approaches perform as well as global ones (for instance, compare the results presented in table 4 of [13] with the ones presented in table 8 of [4] and table 5 of [8]) and the best available result thus far [8] is achieved by modeling the local discontinuities using a robust estimator. Therefore, the question of what should be the extent of local smoothing is also important in fine tuning hybrid (combining local and global) methods.

A. 3D Optic Flow Estimation

In contrast to 2D optic flow calculation, the use of 3D methods, particularly for dynamic CT images, has only started to attract the attention of practitioners in recent years [14]–[16] and its issues and potentials are yet to be fully analyzed. In the biomedical imaging area, 3D optic flow calculation was first used to capture the heart 3D motion using CT [17] and MRI [18] images. The pioneering work of [17] assumes that the image is conserved and incompressible and therefore the velocity field satisfies the divergence-free and the incompressibility constraints. However, the computation of the flow using these two constraints is an ill-posed problem and the solution was found by adding a smoothness term to regularize the penalty function of the weighted sum of the two constraining terms. The velocity field was then calculated by minimizing this penalty function using variational calculus (the minimizing solution generally satisfies the Euler-Lagrange equations) and the solution was found by solving a set of simultaneous coupled elliptic partial differential equations. The differential equations were further discretized resulting in a system of linear equations where the solution is an approximation of the velocity field. This method, in essence, is very similar to some of the contemporary approaches for both calculating the 2D optic flow [3], [8], [19], [20] and energy minimization based image registration [5], [21], [22] techniques.

Except the 3D generalizations of the original version of Lucas & Kanade optic flow method [23], [24], to our knowledge, no other Lucas & Kanade based 3D optic flow method has yet appeared in computer vision literature and the potentials and difficulties associated with using such methods on dynamic 3D data such as 4D CT are yet to be explored. Lucas & Kanade optic flow based method, in contrast to variational based methods, is of particular interest here because it allows the effect of smoothing to be directly controlled and measured.

B. Outline

In this paper, we first examine the theoretical relationship between the estimation accuracy and required amount of smoothness for a general solution from a robust statistics perspective. The analysis leads to a guideline for the sufficient amount of smoothing for the 3D optic flow estimation.

A robust 3D optic flow in which the imposition of smoothing can be locally quantized is then devised to test the proposed hypothesis of smoothing requirement. The hypothesis was then tested using a geometrically realistic synthetic CT image sequence of the breathing lung and five cases of real 4D CT lung images with extensive set of expert annotated land marks. We further examined, both quantitatively and qualitatively, the suitability of imposing the "just enough" smoothing on real 4D CT scans in calculating motions near fissures (borders of lobes) and show the deteriorating effect of

unnecessary smoothing on the estimation of motion particularly in those areas.

II. How Much Smoothing is "Just Enough"?

To answer the above question, we first need to establish an explicit relationship between the smoothing requirement and the estimation accuracy. To ascertain this relationship, we first consider the 3D optic flow estimation problem. The optic flow constraint in 3D is generally written as [17]:

$$I_x u + I_y v + I_z w + I_t = 0 \tag{1}$$

where I_x, I_y and I_z are the spatial and I_t is the temporal derivatives of the image brightness function I and u, v and w are unknown components of the local flow along x, y and z axes respectively. The imposition of this constraint implies that for every voxel, there is only one equation for three unknowns and therefore it is not possible to solve this without adding extra assumption that is generally referred to as smoothing.

The simplest form of smoothing is imposed by assuming the flow to be constant in a local neighborhood and therefore, the above equation is turned into a system of linear equations that can be solved for the unknown components of the flow. In this context, the question of how much smoothing is imposed directly relates to size of this neighborhood and the number of voxels included in the calculation of velocity components. We use this scenario as the basis of our analysis.

To measure the effect of smoothing on accuracy, we then need to examine how the estimation is performed. We already know that in absence of noise one needs to apply the smoothing assumption to only three voxels to be able to calculate the flow in 3D (this is the necessary condition). However, noise is always present and therefore substantially more voxels need to be included. *Extra smoothing assumption is therefore required solely for the sake of accuracy*. More importantly, as the motion boundaries and unmodeled data are unavoidable, some form of outlier rejection (robust statistics) is also required to ensure the accuracy. It's worth noting here that assuming other models of motion (for example affine) only affects the amount of necessary smoothing (would be 12 voxels for full 3D affine). Regardless of the chosen model, one would need to include significantly more points than the necessary number in order to obtain an accurate estimate and therefore the sufficiency requirement is not directly affected by the type of motion model.

Having formulated the problem in this setting, for a given level of noise, the amount of required smoothing now directly relates to the number of included voxels and in turn, the number of included voxels directly relates to the accuracy of estimation that is expressed by a measure called *finite sample bias*. Following [25], the finite sample bias of a robust estimator is defined as:

$$\lambda(n; \Theta^\star, H) \equiv \left| E\left[\hat{\sigma}_n^2; \Theta^\star, H\right] - \sigma^2 \right| / \sigma^2 \tag{2}$$

where σ is the true scale and $E\left[\hat{\sigma}_n^2; \Theta^\star, H\right]$ is the statistical mean of the square of an estimated scale for a given hypothesized fit $\Theta^\star$ and a specific data distribution H. The arguments n, $\Theta^\star$ and H are to show that this is a scale invariant definition of bias and the above measure only depends on the number of data samples, the hypothesized fit and the data population, but not on the scale σ.

Using the above definition we propose an approach that generates a straightforward relationship between the required smoothing and the estimation accuracy by which, the minimum *sufficient* smoothing requirement can be evaluated based on the desired level of accuracy. The analysis here does not include the relationship between the estimation bias and the level of noise as the latter is fixed in a given set of data and interested readers are referred to [25] for detailed discussion of that relationship.

For the sake of quantifying the amount of "just enough" smoothing, we also need to choose the estimator. Based on the analysis of [26], we implemented the MSSE (Modified Selective Statistical Estimator) [27], as the most consistent[1] robust estimator, for the estimation task. The analysis presented in [25] shows that all estimators including MSSE are heavily biased when the distance between the two parallel structures (e.g. two very similar motions in a selected region), μ, is less than 5σ. In this scenario, structures in data are too close to be separated from each other and existence of one structure heavily bias the estimation of the other structure. The analysis also showed that the estimation accuracy would not improve by increasing the amount of data. However, for $\mu \geq 5\sigma$, the MSSE finite sample bias is always less than 20% and the finite sample bias does not change with the inlier percentage or the distance between the two structures (μ). Importantly, the finite sample bias of the MSSE as well as a number of other robust estimators analyzed in [25] do not significantly change as the number of data samples increases beyond a relatively small sample size (of around 100). As such, the minimum amount of required smoothing would not be significantly different if any other high breakdown estimators were used instead.

We also assume that the smoothing support is a cube (as the data is discretized in Cartesian grid) centered on the voxel of interest. In absence of any prior information, the cube is to be constructed symmetrically around the target. Having specified the estimation parameters, we are now able to hypothesize the minimum size requirement of the Gaussian window by analyzing the finite sample bias of the estimator in cases where there are multiple motions in the region of interest. Since we use locally constant velocity model, the motion boundary is modeled by a step edge in the velocity space.

The finite sample bias of various robust estimators, including MSSE, for segmenting a step edge structure, as the worst-case scenario, have already been analyzed [25]. The regression framework used in [25], as shown in section IV-D, is identical to the flow estimation presented here. Part of the results presented in fig. 9 of [25] are reproduced here and shown in Fig. 1. The above results show that for cases where the estimator is consistent, less than 100 data points are required to minimize the effect of finite sample bias of the robust estimator. Therefore, a cube with sides as little as 5 or 7 voxels should provide the minimum required smoothing even

[1] An estimator is said to be consistent if its estimated value approaches the true value as the number of data approaches infinity.

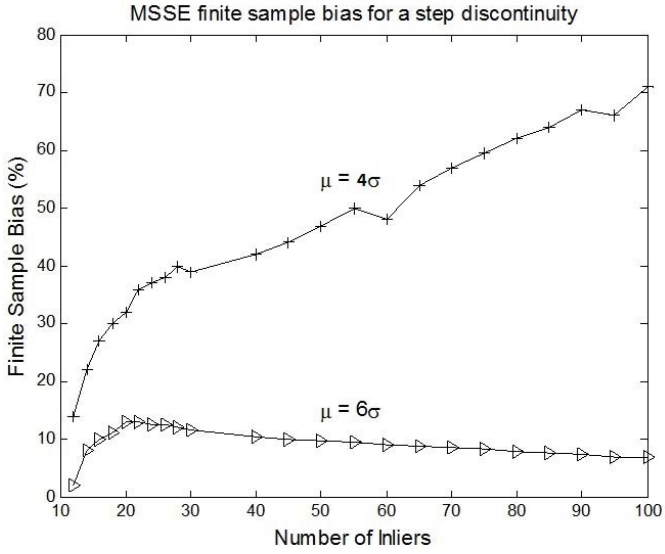

Fig. 1. Finite sample bias of MSSE in the segmentation of a step edge discontinuity with height of either 4 or 6 times the scale of noise in the measured data with 30% inliers. It is important to note here that as shown in [25], for $\mu > 5\sigma$, the amount of finite sample bias is not affected by either the μ or the inlier ratio.

in cases where the target voxel is in the vicinity of a motion boundary and only a fraction of voxels actually belong to the motion of interest. The minimum requirement in 3D data is therefore delightfully small (a CT scan typically has more than 10 million voxels). This means that very localized changes should be directly observable. We examine this hypothesis both in terms of the average accuracy using synthetic and real data (with known ground truth) and its implication for segmenting lung motion based on known anatomical features in real 4D CT data.

III. Estimation of 3D Optic Flow

To examine the full implications of the smoothing requirements in the context of robust optic flow calculation, we present a straightforward optic flow method. In this method, we estimate the 3D flow by assuming that the majority of voxels in a local cube can be explained by velocity perturbations around a constant value (inlier group). Voxels with motions that cannot be explained by the above model (based on the MSSE criteria) would be considered an outlier and would not be included in the estimation process. The velocity of the inlier group is calculated using the Least Square method.

We have implemented the above regularization approach and solved the estimation (MSSE [27]) step using random sampling. The spatial and temporal derivatives are calculated as prescribed by [23] using either Simoncelli's [28] or Gaussian derivative masks.

The implementation of MSSE, as described in Algorithm 1, is very straight forward and only involves taking N number of random samples of 3 voxels from within the support volume. The 3D flow is then calculated for each 3-tuple by solving the system of three linear equations for the three unknown components of the flow at that point. The flow vectors for all samples are then used to calculate square residuals (algebraic distance between voxels' optic flow constraints and the calculated flow of a given 3-tuple) of all voxels inside the support window. The sample that has the minimum sorted square residuals at the K^{th} order index of those residuals is selected as the best estimate (in out experiments we set $K = 0.5$ which gives the median). By starting from the K^{th} residual of this best estimate, the point where the condition: $|r_{i+1}| < T\sigma_i$ based on $\sigma_i^2 = \Sigma_{j=1}^{i} r_j^2/(i-3)$ is no longer true is found and voxels up to this index are considered as inliers. In the above, r is the residual and is given by: $r^2 = (I_x u + I_y v + I_z w + I_t)^2$ where i is the sorted index, T is a constant threshold and numbers around 1.5-2.5 is usually used to indicate an inclusion of around 93−99% of inliers based on a normal distribution for noise. The least squares solution of all the inlier voxels is considered as the final estimate. Since the final scale estimate is calculated by using least squares and including all inliers, the variation of T has little effect on the final results [27].

Algorithm 1 The Step-by-Step Algorithm for MSSE

Inputs: Spatial and temporal derivatives, number of repetitive epochs N

1: Repeat step 2-6 for N times:
2: Choose an elemental subset(3-tuple) by random sampling;
3: Compute the Corresponding velocity vector using

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = - \begin{bmatrix} I_{x1} & I_{y1} & I_{z1} \\ I_{x2} & I_{y2} & I_{z2} \\ I_{x3} & I_{y3} & I_{z3} \end{bmatrix}^{-1} \begin{bmatrix} I_{t1} \\ I_{t2} \\ I_{t3} \end{bmatrix}$$

4: Calculate the Square Residuals $r^2 = (I_x u + I_y v + I_z w + I_t)^2$;
5: Sort the square residuals in ascending order;
6: Find the sample with the least K^{th} Square Residuals;
7: Recalculate the square residuals using the velocity vector with the least median square residual and sort them;
8: Find the first point starting from the median where $|r_{i+1}| > 2.5\sigma_i$, the data up to this point are considered inliers;
9: Calculate the final velocity vector using all the inlier points.

IV. Experimental Results

To investigate the effect of the local smoothing on the accuracy of optic flow we created a sequence of synthetic 3D images having a variety of known motions. The sequence is designed to mimic typical changes in lung CT sequences with several objects having different motions and irregularly shaped motion boundaries. The geometry of lungs are generated from the segmentation of a real human lung CT image and the textures are created by superimposing three 3D sinusoidal patterns similar to the ones used in [29].

In this sequence, the lungs have affine motions while the background has constant 3D motion. A small stationary column, between two lungs, has also been included to simulate motionless parts of a real image. A sample 2D slice (axial view) and its associated flow field are shown in Fig. 2. Although the texture in this sequence is synthetic, the geometry and types of motions are quite realistic and exhibits similar issues encountered in real images including issues associated with existence of quantization noise and deterioration of the

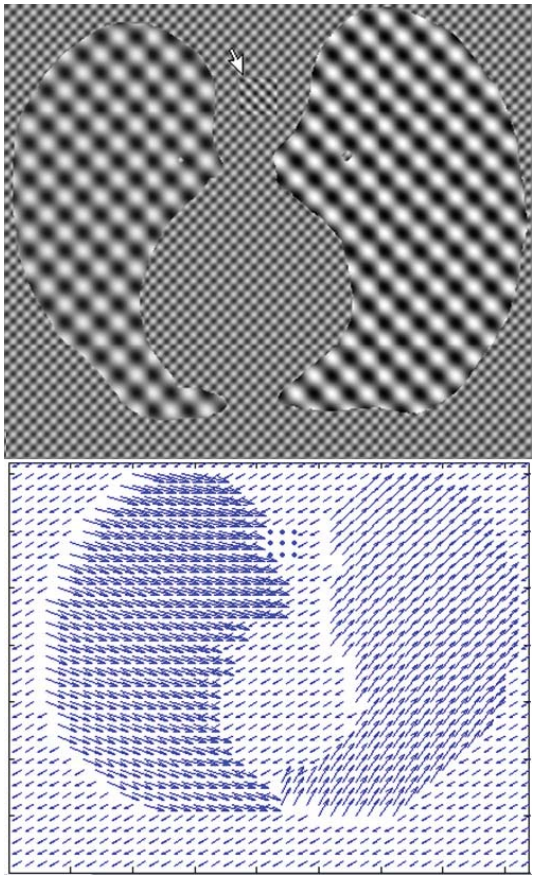

Fig. 2. Sample of axial view of the sinusoidal image (top) and its flow field (bottom). The white arrow identifies the stationary column area.

estimation of derivatives near the motion boundaries. The lungs are however not segmented into different lobes (with different motions) because the analysis is based on using a spatially small area and therefore having different segments does not improve the relevance of the experiment.

For error measurement, we have extended and used the Fleet and Jepson [30] angular measure of error. In the extension of this scheme to 3D, the flow at every voxel is represented by a 4D vector of its Cartesian components in a homogeneous coordinate system $(u,\ v,\ w,\ 1)$ and the error is measured as an angular deviation of the calculated flow from the true velocity. The error is therefore calculated by finding the inverse cosine of the dot product of two vectors in the above 4D homogeneous coordinate system. A detailed analysis of this measure is provided in [29] and will not be repeated here.

A. Effect of Smoothing on Accuracy

To demonstrate the effect of the size of the Gaussian window on the estimation accuracy, we have varied the size from $3 \times 3 \times 3$ to $11 \times 11 \times 11$ voxels. The results shown in Table I are in full agreement with our predictions based on the finite sample bias (see the last part of section 2). The accuracy is significantly enhanced when the size of window increases from $3 \times 3 \times 3$ to $5 \times 5 \times 5$ (which is only around 0.00028% of whole data). However any more increases in the amount of smoothing results in only small changes to the final accuracy.

TABLE I
ESTIMATION ACCURACY OF CALCULATING 3D OPTIC FLOW USING DIFFERENT SIZES OF SMOOTHING WINDOWS

Size of Smoothing Window	Average Error (degrees)	Standard Deviation (degrees)
$3 \times 3 \times 3$	4.49	16.04
$5 \times 5 \times 5$	2.74	11.10
$7 \times 7 \times 7$	2.07	8.88
$9 \times 9 \times 9$	1.84	8.24
$11 \times 11 \times 11$	1.75	8.12

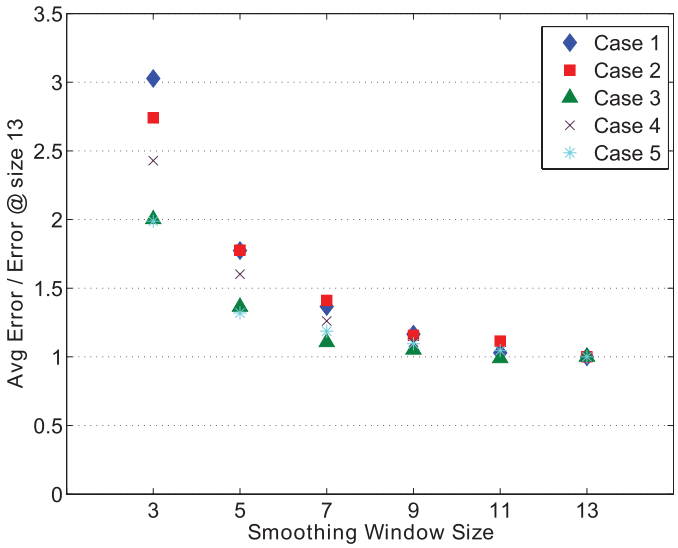

Fig. 3. Change in average error when the smoothing window size is varied, $\sigma = 2.0$.

B. Optic Flow Estimation Using Real 4-D CT Images

Both cardiac and respiratory motions severely affect the quality of lung CT images. Four dimensional CT images are developed to facilitate the analysis of respiratory motion by using spirometer signals to align and synchronize data acquired during different breathing cycles. The data however contains a large amount of noise owing to low-dose protocols used in dynamic CT imaging as well as various types of artifacts due to the fact that lung deformation is not exactly the same in each respiratory cycle (4D CT is typically acquired during several cycles) and cardiac motion is out of sync with breathing cycle. However, 4D CT data is increasingly used for tracking lung motion and helps clinicians minimize and better target radiation in oncology treatments.

Currently there are two commonly used and openly available thoracic 4D CT data sets with ground truth in terms of landmark motions [31], [32] that can be used to validate the accuracy of dense flow calculations. In our experiments, we used both datasets to show real data applications of the proposed theory. The data provided by the "University of Texas M.D. Anderson Cancer Center" [31] contains 300 manually identified landmark points per case. The results of nine deformable registration algorithms using these data have also been provided in [31], [33], [34]. To provide context, results of similar approaches have also been included here. The Dataset provided by "Léon Bérard Cancer Center & CREATIS

TABLE II

AVERAGE ERROR OF CALCULATING 3D OPTIC FLOW USING DIFFERENT SIZES OF SMOOTHING WINDOW FOR REAL 4-D CT DATA

Gaussian Window Size	Case 1 Av Error (SE)[a]	Case 2 Av Error (SE)	Case 3 Av Error (SE)	Case 4 Av Error (SE)	Case 5 Av Error (SE)
Average Disp	4.01 (2.91)	4.65 (4.09)	9.42 (4.81)	6.73 (4.21)	7.10 (5.14)
Maximum Disp	12.65	17.8	21.0	18.46	24.78
3 × 3 × 3	5.028 (0.416)	4.496 (0.371)	6.383 (0.374)	9.193 (0.450)	9.018 (0.853)
5 × 5 × 5	2.945 (0.221)	2.916 (0.192)	4.349 (0.205)	6.062 (0.257)	5.986 (0.381)
7 × 7 × 7	2.269 (0.138)	2.315 (0.141)	3.528 (0.176)	4.773 (0.207)	5.387 (0.314)
9 × 9 × 9	1.936 (0.119)	1.897 (0.114)	3.353 (0.160)	4.188 (0.181)	4.969 (0.288)
11 × 11 × 11	1.710 (0.097)	1.828 (0.109)	3.156 (0.150)	3.949 (0.165)	4.742 (0.261)
13 × 13 × 13	1.661 (0.099)	1.641 (0.098)	3.187 (0.150)	3.785 (0.158)	4.536 (0.252)

[a] Standard Error as defined in [31].

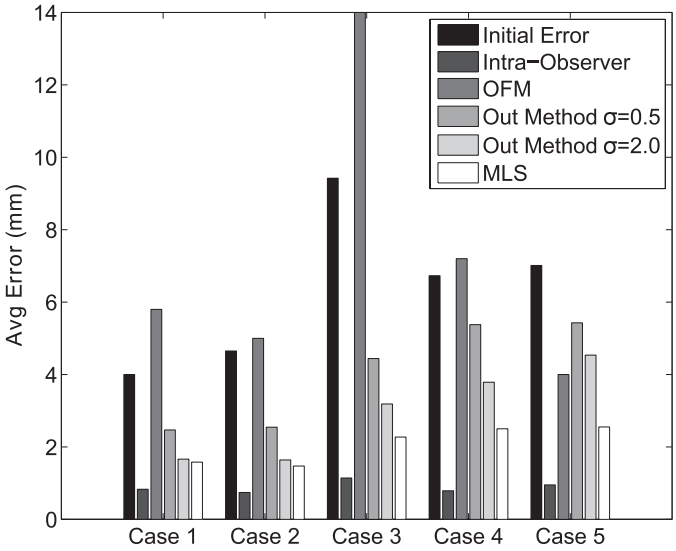

Fig. 4. Average registration errors of our method for each case. Corresponding average errors from [31] and [33] are also shown for comparison.

lab, Lyon, France" [32] has three images with 100 manually annotated landmarks in all frames.

The first five cases of [31] are used here to study the effect of the size of the Smoothing Window on estimation accuracy. We have varied the size from 3 × 3 × 3 to 13 × 13 × 13 voxels and the results are shown in Table II.

The above results, similar to the ones obtained using synthetic data, show that the increase in overall accuracy due to increase in the size of Smoothing Window is plateauing when window sizes are larger. To illustrate this further, average errors are plotted against Smoothing window size in Fig. 3. To bring all the results to the same scale, the average errors are divided by the average error of 13 × 13 × 13 window size for each case.

TABLE III

P-VALUES AND HYPOTHESIS TEST RESULTS (WITH CONFIDENCE INTERVAL OF 99%) OF PAIRED T-TESTS CONDUCTED TO TEST THE SIGNIFICANCE OF IMPROVEMENTS ACHIEVED BY INCREASING THE SIZE OF THE SMOOTHING WINDOW. $H = 1$ MEANS THAT THE DIFFERENCES ARE STATISTICALLY SIGNIFICANT

		3 × 3 × 3 ↓ 5 × 5 × 5	5 × 5 × 5 ↓ 7 × 7 × 7	7 × 7 × 7 ↓ 9 × 9 × 9	9 × 9 × 9 ↓ 11 × 11 × 11	11 × 11 × 11 ↓ 13 × 13 × 13
Case 1	*P*	9.56E-06	0.008	0.068	0.142	0.722
	H	1	1	0	0	0
Case 2	*P*	2.00E-04	0.012	0.021	0.660	0.201
	H	1	0	0	0	0
Case 3	*P*	2.35E-06	0.003	0.460	0.370	0.738
	H	1	1	0	0	0
Case 4	*P*	2.77E-09	1.00E-04	0.033	0.330	0.473
	H	1	1	0	0	0
Case 5	*P*	1.20E-03	0.226	0.328	0.560	0.570
	H	1	0	0	0	0

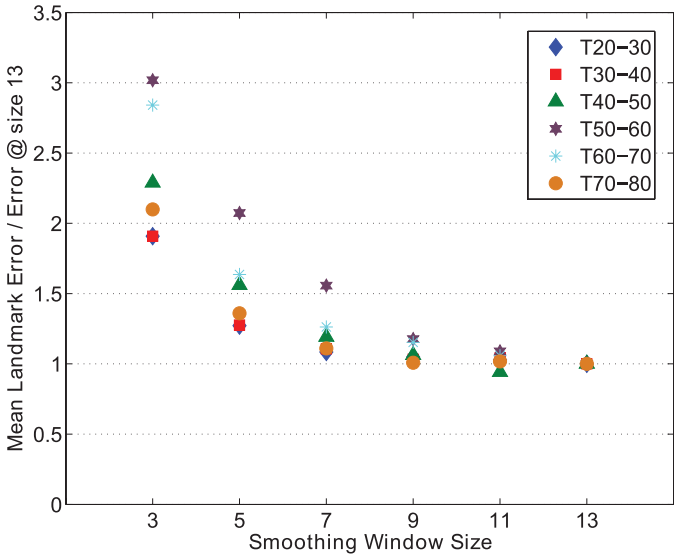

Fig. 5. Effect of varying the size of the smoothing window on average landmark errors in images of Popi Model data set [32].

To show that slight improvements in the mean accuracy of larger window sizes are not statistically significant, Paired t-test was conducted and the results are shown in Table III. For a given set of images, we calculated the p-value using the landmark registration errors of two adjacent smoothing window sizes. These results were then used to compare the mean errors against a null hypothesis (results are not significantly different) based on a confidence interval of 99%. A p-value $< 1\%$ means that the null hypothesis is rejected ($H = 1$) and the difference between the two results are statistically significant. As expected we can see that in all cases the results improve when the smoothing size increase from 3 × 3 × 3 to 5 × 5 × 5 but increases beyond 7 × 7 × 7 are not statistically significant.

The average errors of our method for all those cases are also compared with the results in [31], [33] [2] and are shown in Fig. 4. In particular, the OFM (Optical Flow Method) in this

[2]The mean errors for our method are calculated using the publicly available landmark set (containing 300 landmarks per image), which is a subset of the full landmark set used to calculate the results in [31] and [33].

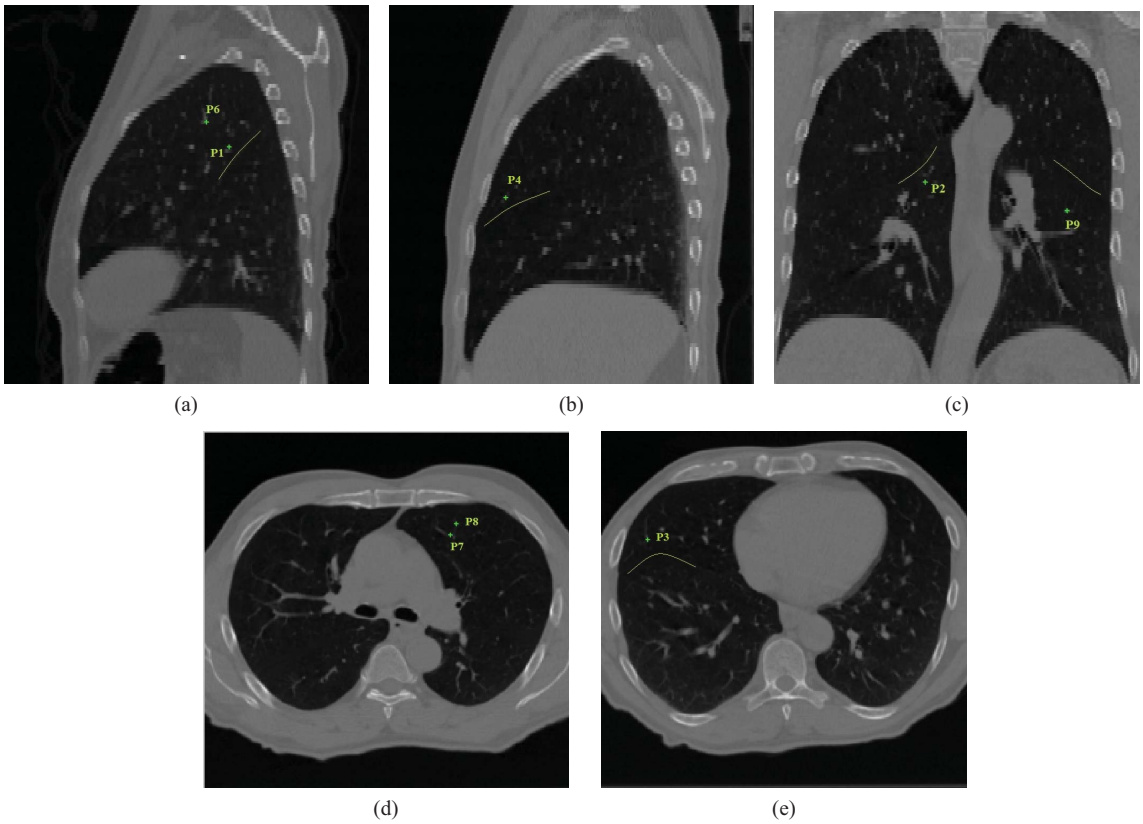

Fig. 6. Position of different landmark points used in our calculations (P1–P9). Where possible, the fissure is also highlighted in these images.

Fig. (from [31]) represents the results of the Horn & Schunck based flow calculation method which is a global method. The average error in landmark registration across all five cases for OFM is 7.3 mm whereas for our method with $\sigma = 0.5$ it is 4.051 mm and with $\sigma = 2.0$ the average error is 2.962 mm.

The MLS (moving least-square) in Fig. 4 (from [31]) is a landmark based registration algorithm which uses external information (expert knowledge) for identifying the landmarks. MLS has achieved an average error of 2.074 mm across all five cases. The difference between the average errors of our method ($\sigma = 2.0$) and the MLS is less than the intra-observer error (this is the error in identifying the Landmark points by an expert in repeated experiments) for the first 3 cases. This shows that for small displacements, local optic flow based method can achieve similar results to that of hand tuned registration methods.

It is important to note that a hierarchical approach would have yielded more accurate results for the last two cases which have higher displacements. We have not implemented this because improving accuracy by itself is not the aim here. Also, the landmark dataset used in evaluating our method consists of 300 publicly available landmarks.

In the next set of experiments, we used the Popi model dataset provided by [32] to demonstrate the effect of smoothing window size on the accuracy of optic flow calculations. The flow field was calculated using two adjacent frames in each 4DCT dataset. All landmarks were used to calculate the mean landmark error for each time step and results are shown in Fig. 5. In this figure, each time frame of 4DCT image is identified by a frame number e.g. T00, T10, T20, etc. and the mean landmark errors of all landmarks at each time step are averaged over the three cases. These results, similar to ones achieved for DIR data, are in line with our theoretical predictions.

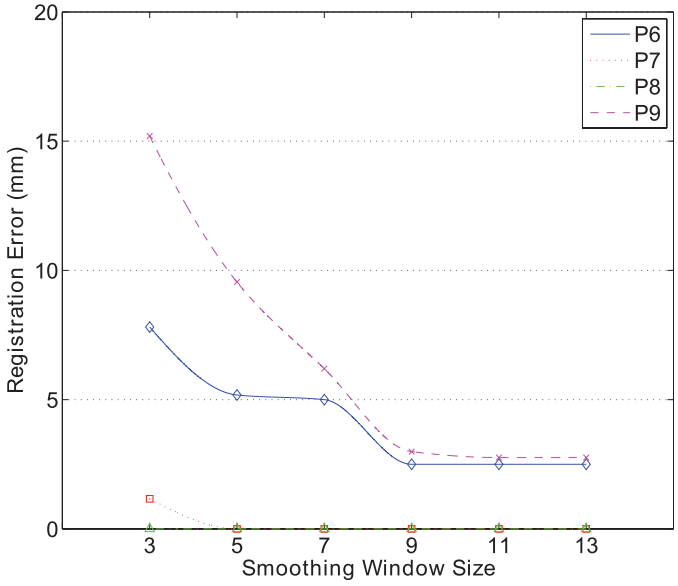

Fig. 7. Variation in estimation error with smoothing window size for landmark points near a fissure.

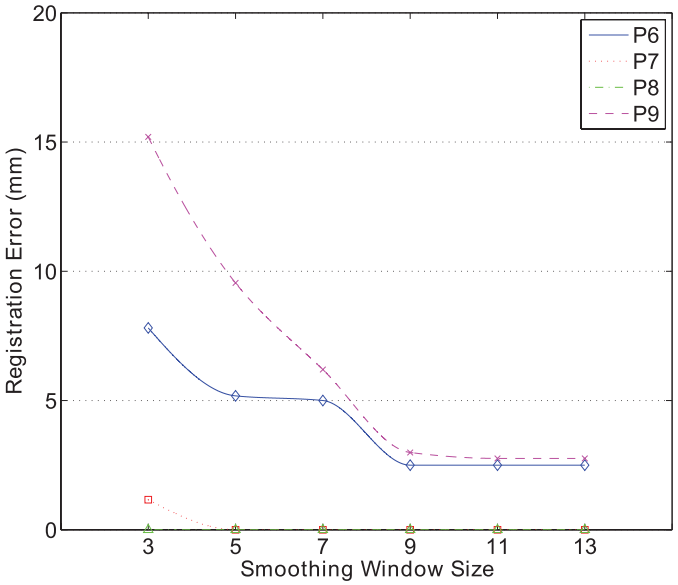

Fig. 8. Variation in estimation error with smoothing window size for landmark points not near a fissure.

C. Side Effects of Smoothing Increases

A lung contains separate lobes with different deformation patterns and the motion at the lobes boundaries (fissure) is discontinuous. The juxtaposition of lobes in a typical lung is shown in Fig. 6.

To see the effect of the size of Smoothing window on the accuracy of estimation near the fissure, we compared the estimated error for different smoothing window sizes for a small selection of points near the fissure and points in the same image, but further away from the fissure. The results of estimation error for different smoothing window sizes are shown in Fig. 7. As was predicted, the results show that the increase in smoothing window can indeed deteriorate the overall accuracy.

Four points (P6, P7, P8 & P9) are selected from cases 1 and 2 where P6 is in the same sagittal slice as P1 and P7 & P8 are in the same axial slice as P2 but are further apart from the fissure than P1 and P2. The errors are plotted against the Smoothing window size in Fig. 8.

D. Detection of Local Variations in Flow Fields

To highlight the importance of controlling the size of smoothing window, we concentrate on the detection of anatomically induced variation in flow field. We first introduce a novel way of visualizing the variations of 3D flow fields in 3D space and use that method to show the detectability of anatomical variation of flow fields on both synthetic and real 4D CT images. To visualize the variation, we rewrite the 3D optic flow constraints (equation 1) as

$$\frac{I_x}{I_t}u + \frac{I_y}{I_t}v + \frac{I_z}{I_t}w = -1. \tag{3}$$

The optic flow constraint is commonly viewed as a plane in the velocity space of (u, v, w). However, using the duality of points and planes in perspective geometry, we can view the above equation as a point in the derivative space $(I_x/I_t, I_y/I_t, I_z/I_t)$ and therefore all the points having the same (constant) velocity will form a 3D plane parameterized by (u, v, w). The importance of this transformation stems from the fact that while it's very difficult to visualize a number of 3D planes crossing each other around the vicinity of a single point, we can easily visualize a set of points scattering around a plane in a 3D Cartesian space.

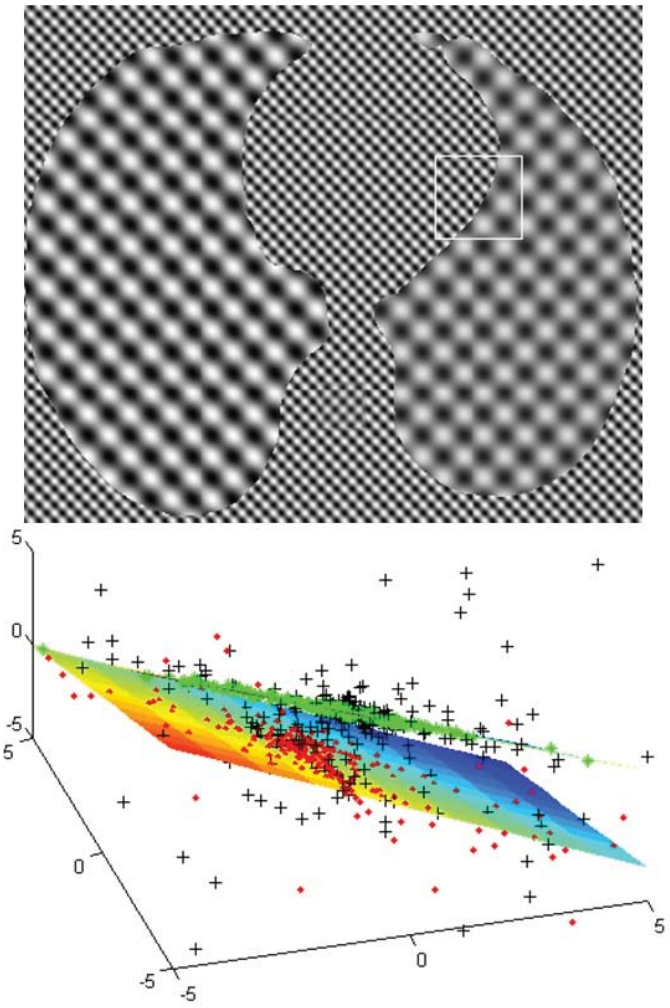

Fig. 9. The bottom image is an axial view of the sinusoidal image where the cubic region with two distinct motions is marked by a white rectangle. Bottom image plots the associated voxels in the derivative space (red dots and green asterisks are used for moving voxels, while black crosses represent the outliers). The points are segmented by fitting planes using MSSE. The bottom picture is rotated to show how the points are aligned with respect to the affine plane. Since a plane in the derivative space represents constant motion, the points with constant underlying motion (red dots) fit a plane better than the points with affine underlying motion (green asterisks).

To demonstrate the effectiveness of the above transformation to visualize the existence of different motions in a local area, we consider a cubic region in the synthetic lung image introduced earlier. The cubic region is shown in the top part of Fig. 9 by a white rectangle in the axial view and includes voxels with two different motions: constant and affine. A regular subset (one in nine) of voxels from this area is plotted in the derivative space and the points are segmented by applying MSSE sequentially (i.e. a fit and remove application of MSSE as explained in [27]).

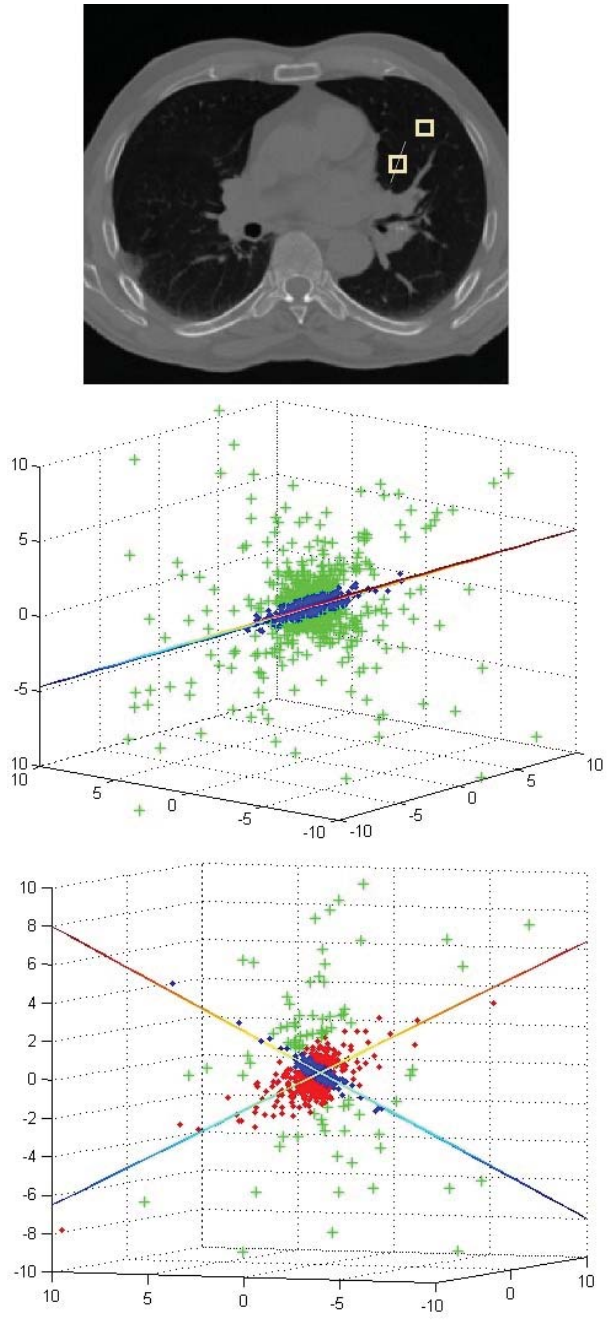

Fig. 10. Top image is an axial view of a 4-D CT image where the two cubic areas with either one (the top rectangle placed entirely on one lobe) or two (the lower rectangle straddles two lobes separated by the fissure - shown by a white line) distinct motions are marked by white rectangles. The middle and bottom images plot the associated voxels in the derivative space (red dots and green asterisks are used for moving voxels, while black crosses represent the outliers). The points in both areas are segmented by fitting planes using MSSE. The middle and bottom pictures are rotated to show how the points are aligned with respect to every plane. The bottom picture clearly shows that the difference between the motions of two lobes on different sides of a fissure is detectable using a local approach.

The resulting planes are shown in the bottom part of the Fig. 9 while the plot is rotated to show the spread of points around one of those planes. As expected, voxels with constant underlying motion fit well to a plane in the derivative space and therefore form a very thin scatter of points around a plane. The voxels with affine motions, on the other hand, are scattered wider while they still generally spread near a plane as long as they are spatially close to each other in the actual image. These results support our earlier hypothesis that good estimates of motion can be achieved using localized small smoothing area.

We can now apply the same transformation to the real 4D CT images and examine the possibility of detecting anatomical motion boundaries. We have examined the possibility of detecting the differences in motion between individual lung lobes by considering two small (local) cubic areas in a 4D CT image. As shown in the top part of Fig. 10, one of these cubes (shown by a white rectangle in the axial view) is entirely within a lobe and the other straddles two different lobes (where the fissure - highlighted by a white line - is located).

We have plotted all the points associated with the voxels in those areas and applied the MSSE [27] (with 30% minimum ratio and $T = 2$) to segment the motions. The segmentation results for both areas are shown in the middle and bottom parts of Fig. 10. While only one motion is detected in the upper area, the one straddling two lobes includes two distinct motions. This evidence again shows that imposition of "just enough" smoothing is important in preserving the local flow information.

V. Conclusion

A new approach to quantifying the minimum required smoothing based on the concept of the finite sample bias of a robust estimator is presented. The proposed approach is very general and makes predictions about required amount of smoothness to satisfy the sufficiency condition for a broad range of visual estimation tasks such as optic flow calculation. We particularly showed that smoothing over a cubic area as small as 5 to 7 voxels wide is sufficient to achieve the highest practical accuracy. This is a significant observation as it proves that very localized changes in motion in 3D data is directly observable. The predictions were tested for 3D optic flow estimation of 4D lung CT images and the results showed, inline with theoretical predictions, that only a very small amount of local smoothing is required to achieve high accuracy and observe local anatomically induced motion variations and that in some cases increasing the amount of smoothing deteriorates results.

Acknowledgment

The authors would like to thank Australian Research Council and the Danish Strategic Research Council (DSF) for supporting this paper.

References

[1] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.*, vol. 17, no. 1, pp. 185–203, 1981.

[2] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. Imaging Understand. Workshop*, pp. 121–130, 1981.

[3] A. Bruhn, J. Weickert, T. Kohlberger, and C. Schnörr, Discontinuity-preserving computation of variational optic flow in real-time in Lecture Notes in Computer Science. Berlin, Germany: Springer-Verlag, 2005, pp. 279–290.

[4] N. Papenberg, B. B. Andrès, S. T. Didas, and J. Weickert, "Highly accurate optic flow computation with theoretically justified warping," *Int. J. Comput. Vis.*, vol. 67, no. 2, pp. 141–158, 2006.
[5] J. Modersitzki, "Flirt with rigidity-image registration with a local non-rigidity penalty," *Int. J. Comput. Vis.*, vol. 76, no. 2, pp. 153–163, Feb. 2008.
[6] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," *Int. J. Comput. Vis.*, vol. 92, no. 1, pp. 1–31, 2011.
[7] D. Sun, S. Roth, and M. Black, "Secrets of optical flow estimation and their principles," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Mar. 2010, pp. 2432–2439.
[8] A. Bruhn, J. Weickert, and C. Schnörr, "Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods," *Int. J. Comput. Vis.*, vol. 61, no. 3, pp. 211–231, 2005.
[9] L. Xu, J. Jia, and Y. Matsushita, "Motion detail preserving optical flow estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 9, pp. 1744–1757, Sep. 2012.
[10] T. Lindeberg, "A scale selection principle for estimating image deformations," *Image Vis. Comput.*, vol. 16, no. 14, pp. 961–977, 1998.
[11] P. Meer, G. Medioni, and S. Kang, "Robust techniques for computer vision," in *Emerging Topics Computer Vision*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2004.
[12] M. Otte and H. Nagel, *Optical flow estimation: Advances and comparisons*, in Lecture Notes in Computer Science. Berlin, Germany: Springer-Verlag, 1994, pp. 49–60.
[13] A. Bab-Hadiashar and D. Suter, "Robust optic flow computation," *Int. J. Comput. Vis.*, vol. 29, no. 1, pp. 59–77, 1998.
[14] V. Boldea, G. C. Sharp, S. B. Jiang, and D. Sarrut, "4D-CT lung motion estimation with deformable registration: Quantification of motion nonlinearity and hysteresis," *Med. Phys.*, vol. 35, no. 3, pp. 1008–1018, 2008.
[15] D. Yang, W. Lu, D. A. Low, J. O. Deasy, A. J. Hope, and I. E. Naqa, "4D-CT motion estimation using deformable image registration and 5-D respiratory motion modeling," *Med. Phys.*, vol. 35, no. 10, pp. 4577–4590, 2008.
[16] G. G. Zhang, T. C. Huang, T. Guerrero, K. P. Lin, C. Stevens, G. Starkschall, and K. Forster, "Use of three-dimensional (3D) optical flow method in mapping 3-D anatomic structure and tumor contours across four-dimensional computed tomography data," *Appl. Clinical Med. Phys. Amer. College Med. Phys.*, vol. 9, p. 2738, 2008.
[17] S. M. Song and R. M. Leahy, "Computation of 3-D velocity fields from 3-D cine CT images of a human heart," *IEEE Trans. Med. Imaging*, vol. 10, no. 3, pp. 295–306, Sep. 1991.
[18] A. Andreopoulos and J. K. Tsotsos, "Efficient and generalizable statistical models of shape and appearance for analysis of cardiac MRI," *Med. Image Anal.*, vol. 12, no. 3, pp. 335–357, 2008.
[19] A. Bruhn, J. Weickert, C. Feddern, T. Kohlberger, and C. Schnörr, *Real-Time Optic Flow Computation with Variational Methods*, in Lecture Notes in Computer Science. Berlin, Germany: Springer-Verlag, 2003, pp. 222–229.
[20] J. Weickert and C. Schnörr, "Variational optic flow computation with a spatio-temporal smoothness constraint," *J. Math. Imaging Vis.*, vol. 14, no. 3, pp. 245–255, 2001.
[21] B. Fischer and J. Modersitzki, *FLIRT: A Flexible Image Registration Toolbox*, in Lecture Notes in Computer Science. Berlin, Germany: Springer-Verlag, 2003, pp. 261–270.
[22] E. Haber and J. Modersitzki, "Image registration with guaranteed displacement regularity," *Int. J. Comput. Vis.*, vol. 71, no. 3, pp. 361–372, 2007.
[23] J. L. Barron and N. A. Thacker, *Tutorial: Computing 2-D and 3-D Optical Flow*. San Francisco, CA, USA: Academic Press, 2005.
[24] E. Castillo, R. Castillo, Y. Zhang, and T. Guerrero, "Compressible image registration for thoracic computed tomography images," *J. Med. Biol. Eng.*, vol. 29, no. 5, pp. 222–233, 2009.
[25] R. Hoseinnezhad, A. Bab-Hadiashar, and D. Suter, "Finite sample bias of robust estimators in segmentation of closely spaced structures: A comparative study," *J. Math. Imaging Vis.*, vol. 37, no. 1, pp. 66–84, 2010.
[26] R. Hoseinnezhad and A. Bab-Hadiashar, "Consistency of robust estimators in multi-structural visual data segmentation," *Pattern Recognit.*, vol. 40, no. 12, pp. 3677–3690, 2007.
[27] A. Bab-Hadiashar and D. Suter, "Robust segmentation of visual data using ranked unbiased scale estimate," *Robotica*, vol. 17, no. 6, pp. 649–660, 1999.
[28] E. P. Simoncelli, "Design of multi-dimensional derivative filters," in *Proc. IEEE Int. Conf. Image Process.*, vol. 1. Nov. 1994, pp. 790–794.
[29] J. L. Barron, D. J. Fleet, S. S. Beauchemin, and T. A. Burkitt, "Performance of optical flow techniques," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1992, pp. 236–242.
[30] D. J. Fleet and A. D. Jepson, "Computation of component image velocity from local phase information," *Int. J. Comput. Vis.*, vol. 5, no. 1, pp. 77–104, 1990.
[31] R. Castillo, E. Castillo, R. Guerra, V. E. Johnson, T. McPhail, A. K. Garg, and T. Guerrero, "A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets," *Phys. Med. Biol.*, vol. 54, no. 7, p. 1849, 2009.
[32] J. Vandemeulebroucke, S. Rit, J. Kybic, P. Clarysse, and D. Sarrut, "Spatiotemporal motion estimation for respiratory-correlated imaging of the lungs," *Med. Phys.*, vol. 38, no. 1, pp. 166–178, 2011.
[33] E. Castillo, R. Castillo, J. Martinez, M. Shenoy, and T. Guerrero, "4-D deformable image registration using trajectory modeling," *Phys. Med. Biol.*, vol. 55, no. 1, p. 305, 2010.
[34] X. Gu, H. Pan, Y. Liang, R. Castillo, D. Yang, D. Choi, E. Castillo, A. Majumdar, T. Guerrero, and S. B. Jiang, "Implementation and evaluation of various demons deformable image registration algorithms on a GPU," *Phys. Med. Biol.*, vol. 55, no. 1, p. 207, 2010.

Alireza Bab-Hadiashar (SM'04) received the B.Sc. degree in mechanical engineering from Tehran University, Tehran, Iran, the M.E.S. degree in mechanical and mechatronics engineering from the University of Sydney, Sydney, Australia, and the Ph.D. degree in computer vision from Monash University, Melbourne, Australia, in 1988, 1993, and 1997, respectively.

He is currently a Professor with the School of Aerospace, Mechanical and Manufacturing Engineering, RMIT University, Melbourne, Australia. He has made significant contributions to the establishment of mechatronics engineering degree programs in both Australia and Malaysia. He has authored or co-authored numerous highly cited papers in journals and conferences on optical flow, robust segmentation, and model selection techniques. His current research interests include the development of robust estimation and segmentation techniques for computer vision applications and intelligent and robust sensory systems for new generation vehicle technologies.

Ruwan B. Tennakoon (M'09) received the B.Eng. degrees in electrical and electronics engineering from the University of Peradeniya, Peradeniya, Sri Lanka, in 2007. He is currently pursuing the Ph.D. degree with the Faculty of Engineering and Industrial Sciences, Swinburne University of Technology, Melbourne, Australia.

His current research interests include medical image registration, motion segmentation, and statistical signal processing.

Marleen de Bruijne received the M.Sc. degree in physics and the Ph.D. degree in medical imaging from Utrecht University, Utrecht, The Netherlands, in 1997 and 2003, respectively.

She is currently an Associate Professor of medical image analysis with Erasmus MC Rotterdam, Rotterdam, The Netherlands, and with the University of Copenhagen, Copenhagen, Denmark. She was an Assistant Professor and then an Associate Professor with the IT University of Copenhagen, from 2003 to 2006. She has authored or co-authored over 100 papers in peer-reviewed international journals and conferences. Her current research interests include model-based and quantitative analysis of medical images with applications a.o. in pulmonary imaging, neuro imaging, and cardiovascular imaging.

Dr. de Bruijne has been a member of the program committee of many international conferences in medical imaging and computer vision. She was an Associate Editor of *Image and Vision Computing and Medical Physics* and a member of the Editorial Board of *Medical Image Analysis*.