Synthetic data generation for optical flow evaluation in the neurosurgical domain

Abstract: Towards computer-assisted neurosurgery, scene understanding algorithms for microscope video data are required. Previous work utilizes optical flow to extract spatio-temporal context from neurosurgical video sequences. However, to select an appropriate optical flow method, we need to analyze which algorithm yields the highest accuracy for the neurosurgical domain. Currently, there are no benchmark datasets available for neurosurgery. In our work, we present an approach to generate synthetic data for optical flow evaluation on the neurosurgical domain. We simulate image sequences and thereby take into account domain-specific visual conditions such as surgical instrument motion. Then, we evaluate two optical flow algorithms, Farneback and PWC-Net, on our synthetic data. Qualitative and quantitative assessments confirm that our data can be used to evaluate optical flow for the neurosurgical domain. Future work will concentrate on extending the method by modeling additional effects in neurosurgery such as elastic background motion.

Keywords: Neurosurgery, surgical microscope, optical flow, evaluation

1 Introduction

Worldwide, more than 13 million neurosurgical interventions are conducted annually [1]. Thereby, neurosurgeons need a
well given challenging visual conditions: low texture tissue, blur, specularly reflective surfaces, and large amplitude motion. While in other medical domains methods exist to generate synthetic data for optical flow evaluation [7], there are presently no such methods for neurosurgery.

**Contributions.** Here, we present a method to generate synthetic neurosurgical microscope data for optical flow evaluation. This data takes into account the variety of surgical instruments and challenging visual conditions present in neurosurgery. In our experiments, we compare the accuracy of two state-of-the-art optical flow algorithms on the generated data.

## 2 Method

We present the approach for synthetic data generation through simulation of the microscope field of view (FOV).

### 2.1 Synthetic data generation approach

Evaluation of optical flow requires a sequence of at least two images and the corresponding optical flow ground truth. To evaluate optical flow algorithms, we generate short sequences of abstract surgical actions in the microscope FOV. From clinical recordings, we concluded that the majority of motion is caused by surgical instruments. Therefore, our approach focuses on simulating instrument motion.

Our simulation is inspired by the geometric constraints during neurosurgical interventions. Typically, the surgeon operates under a microscope with an instrument in each hand (see Fig. 2 (a)). We model this situation in our simulation (see Fig. 2 (b)). According to discussions with surgical experts, the surgeon’s instruments point towards the location of surgical action. We refer to this location as activity center, \( p_{ac} \) (see Fig. 2 (c)). We model this by simulating instrument tip motion around a (non-visible) activity center. To achieve plausible instrument body motion, we constrain each instrument body \( i \) by a prismatic-spherical joint at the random location \( p_{hi} \) (possible surgeon's hand locations).

The data generation process consists of three steps.

**1) Motion modelling.** Since optical flow evaluation requires only short temporal motion sequences, we abstain from simulating the complete surgery. Instead, we describe instrument motion by a simple randomized model (see Fig. 2 (d)). Additionally, a randomly oriented force (green) is applied to the tips.
For each new sequence, the activity center $p_{ac}$ is set randomly. Its acceleration $\frac{d}{dt}v_{ac}$ is modeled by a Gaussian distribution in eq. 1,

$$\frac{d}{dt}v_{ac} = \mathcal{N}(0, \Sigma_{ac}^2),$$

where $\Sigma_{ac}$ is tuned manually to match the clinical data. Possible motions are limited inside the FOV.

The instrument tip motion $v_{ti}$ is modeled by an attraction force towards the activity center. Additionally, we introduce a random acceleration term. The velocity of the instrument tip $v_{ti}$ is described by eq. 2,

$$\frac{d}{dt}v_{ti} = k_i d_{i-ac} + \mathcal{N}(0, \Sigma_i^2),$$

whereas $d_{i-ac}$ describes the vector to $p_{ac}$. Attraction of activity center $k_i$ and random motion component $\Sigma_i$ are also adjusted manually. We enforce the instruments to stay within the FOV by increasing $k_i$ upon leaving the FOV. Collisions of instrument bodies are avoided by collision detection routines.

(2) **Instrument shape variation.** Neurosurgical interventions are characterized by a high variability of instrument types and instrument shapes, [8] lists more than 100 different instruments. Thus, accurate modeling of all existing neurosurgical instruments is impossible. Instead, we simulate the shape variety by generating arbitrary instrument shapes using structured domain randomization.

Our approach consists of three main operations. Each operation is inspired by properties of existing instruments (see Fig. 3 (a)). First, we create the instrument axis as straight line in cylinder coordinates $(z, \phi, r)$. Then, we manipulate the instrument axis by inserting a random kink (see Fig. 3 (b)) and/or bending the axis by inserting Beziére curves (see Fig. 3 (c)). Third, we extrude a cross-section along the instrument axis. All instruments generated in our approach have rotational symmetric cross-sections, which depend on the radius $r(z)$ along the z-axis. The radius function $r(z)$ is determined by random combination of elementary functions, including low-order polynomials, exponential and trigonometric functions.

Moreover, some neurosurgical instruments such as bi-polar coagulator or forceps consist of two body parts. Therefore, to model instruments with two body parts, we clone the extruded instrument bodies with a certain probability. We apply a random transformation $T$ to the two instrument body parts relative to the instrument tip $p_{ti}$. By translation and slight rotation, the intersection point of the bodies is moved along the instrument axis.

(3) **Background augmentation and visual property simulation.** To increase the degree of realism in our simulation, we include still images from neurosurgical video recordings as background. Including these images prior to rendering allows simulation of characteristic mirroring effects on the instruments. To simulate realistic instrument reflection, we incorporate specular textures. To achieve a realistic lighting set-up of the microscopes, we model a co-axial light source. Furthermore, to simulate the usually limited depth of field in neurosurgical video data we apply artificial blur.

**Implementation.** The simulation approach was implemented in the open-source framework Blender. First, we generate instruments according to the described model. Next, we create instrument motion sequences of arbitrary length. Optical flow output is directly obtained from Blender.
2.2 Experimental setup for optical flow evaluation

We evaluate our synthetic data w.r.t. accuracy of two state-of-the-art optical flow algorithms, Farneback\(^1\) [9] and PWC-Net\(^2\) [2]. For our analysis, we generate a dataset \textit{NeurOF} comprising 1087 sequences, each with 10 images at a resolution of 960 x 540. Background images were inserted from video recordings of 2 tumor, 2 vascular and 2 spine cases, which were conducted at Inselspital Bern, Switzerland with a ZEISS KINEVO 900 surgical microscope. For comparison, we evaluate two public datasets, namely FlyingChairs [5] and MPI-Sintel [6]. We evaluate the accuracy of estimated optical flow \((v_{i,j})_{\text{pred}}\) versus ground truth \((v_{i,j})_{\text{gt}}\) by means of endpoint error in eq. 3,

\[
AEPE = \frac{1}{n} \sum_{i,j} \left( (v_{i,j})_{\text{gt}} - (v_{i,j})_{\text{pred}} \right)^2.
\]  

3 Results

First, we present example images as generated by our method. We evaluate them qualitatively w.r.t. realism and image quality. Additionally, we present the generated optical flow ground truth. The synthetic scene shown in Fig. 4, 5 (a)-(d) prove that our method generates image data that resembles clinical neurosurgical data. Comparing clinical data (Fig. 1) and synthetic data (Fig. 4, 5) we observe similar visual appearances. The computed optical flow from the two algorithms in our evaluation, Farneback and PWC-Net (see Fig. 4, 5 (e)-(f)), indicate our data can be used to assess optical flow accuracy.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Farneback</th>
<th>PWC-Net (large)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FlyingChairs</td>
<td>8.18</td>
<td>1.44</td>
</tr>
<tr>
<td>MPI-Sintel (final)</td>
<td>11.35</td>
<td>3.7</td>
</tr>
<tr>
<td>NeurOF</td>
<td>2.48</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Next, we quantitatively assess the accuracy of these two optical flow algorithms on our synthesized data. According to the numerical results (see Table 1), PWC-Net performs at lower AEPE on \textit{NeurOF} than Farneback. Qualitative results in Fig. 4, 5 support our numerical findings. The lower AEPE for PWC-Net than Farneback are in line with numeric results on the public datasets FlyingChairs and MPI-Sintel (see Table 1).

---

\(^1\) OpenCV implementation. Parameters: pyramid scale = 0.5, levels = 3, window size = 15, iterations = 3, size pixel neighborhood = 5, STD for smoothing derivatives = 1.2

\(^2\) PWC-Net large, implementation and weights (FlyingChairs/FlyingThings3D cycle) by github.com/philferriere/tfoptflow
In direct comparison, the AEPE on NeurOF is lower than for FlyingChairs and MPI-Sintel. This observation is expected since NeurOF models only foreground instrument motion, while background motion is zero.

4 Conclusions

We demonstrate a synthetic data generation approach for the evaluation of optical flow on the neurosurgical domain. Our approach captures various effects, which are relevant for this domain such as large instrument motion amplitude and specular reflections. We verify the quality of the generated data through visual inspection w.r.t. optical flow calculation. Numerical results show that PWC-Net performs better than Farneback. However, our data models only instrument motions while the background is fixed. Future work will address simulation of a non-rigid background to further improve realism. Then, the benchmark can be extended by evaluating more state-of-the-art optical flow algorithms. Although our experiments focus on using the data generation approach for optical flow evaluation, our method potentially can also improve other computer-assisted surgery applications.

Author Statement

Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent has been obtained from all individuals included in this study. Ethical approval: The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration, and has been approved by the authors’ institutional review board or equivalent committee.

References