Detection of Unique Point Landmarks in HARDI Images of the Human Brain

Henrik Skibbe and Marco Reisert

Department of Radiology, University Medical Center Freiburg, Germany
{henrik.skibbe, marco.reisert}@uniklinik-freiburg.de

Abstract. Modern clinical image acquisition techniques like diffusion magnetic resonance imaging (dMRI) or functional MRI (fMRI) allow us to study tissue and organisms in their natural volumetric configuration. Group studies often require the co-registration of images or partial image structures of different individuals. In such applications the detection of characteristic landmarks is often an indispensable prerequisite. However, the reliable detection of unique anatomical landmarks in medical images is a challenging problem which is often solved in a semi-automated or even manually manner. Landmarks are often sought to be located and analyzed at every position, and in any orientation. In particular the latter makes landmark detection challenging in 3D. In this paper we propose a new landmark detection system that reliably detects arbitrarily placed landmarks in High Angular Resolution Diffusion Images (HARDI) of the human brain.

1 Introduction

The study of Magnetic Resonance (MR) imaging modalities is of great interest in fundamental neuroscience and medicine. MR imaging offers a wide range of different contrasts, providing scientists insights into anatomical and functional properties of the human brain. In this paper we focus on High Angular Resolution Diffusion Images (HARDI [8]) of the human brain. The HARDI-technique combines different measurement parameters to infer underlying tissue properties and allows for studying the neuronal fiber architecture in the human brain without harming the patient.

In this paper we introduce a new approach for detecting unique point landmarks in HARDI images of the human brain. The reliable detection of landmarks [2] plays an indispensable role in registering brain structures from different images and thus is an important prerequisite for many registration and segmentation algorithms [3]. Our approach is based on the computation of a dense feature map of the HARDI signal. Similar to [9], where features are used to find correspondences in scalar valued MR contrasts, we propose features offering a unique signature of a voxel's surrounding in tensor-valued HARDI signals. Thanks to these features a large number of corresponding points can be reliably found in images of different individuals using a linear classifier. The parameters for the linear classifier are learned from a training set of landmarked images.

The challenges in our scenario are manifold: the HARDI technique leads to images with poor quality in terms of resolution and signal to noise ratio. Furthermore, the tensor valued HARDI signal can be considered as function $\mathbf{f} : \mathbb{R}^3 \times S_2 \to \mathbb{R}$, where S_2 denotes the unit sphere in $\mathbb{R}^3$. This means at each voxel position $\mathbf{x} \in \mathbb{R}^3$ we have an angular dependent measurement $\mathbf{f}(\mathbf{x}, \mathbf{n})$ represented as a function on the unit sphere. This function represents the diffusion weighted MR signal with respect to different diffusion directions $\mathbf{n}$. Due to this fact we cannot use intensity based standard techniques to describe a landmark; suitable representations of the signal are required in order to uniquely detect landmarks within the images.

The contributions of the present paper are the following: (1) we derive new rotation invariant features for HARDI images of the human brain. These features are yielding a unique description of each voxel in the images which allows for reliably detecting landmarks within images of different individuals. Positioning landmarks is not restricted to specific areas or points; landmarks can be positioned anywhere in the brain. (2) We propose a linear coarse to fine classification scheme to detect a large number (more than several thousands) of different, unique landmarks in reasonable time. (3) We demonstrate the effectiveness of our method in an experiment based on a dataset of brain images of 21 healthy volunteers. Furthermore, preliminary results on datasets of patients with pathologies are promising. (4) The source code will be made publicly available upon acceptance of this paper.

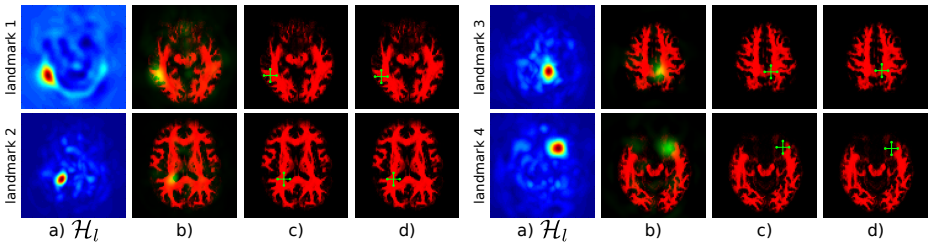

Fig. 1. a) Differently weighted linear combinations of the feature images lead to different detection results (Maximum intensity projection). b) The outcome of the linear classifier at the desired landmark position (green) together with the white matter mask (red). c) Position of the global maximum of $\mathcal{H}_l$. d) The landmark reference position.

2 Landmark Detection in HARDI Images

Our landmark detection method is based on local, rotation invariant feature images. The idea is that corresponding to each voxel position $\mathbf{x} \in \mathbb{R}^3$ in the HARDI signal there exists a feature vector $\mathbf{F}(\mathbf{x}) := (F_1(\mathbf{x}), F_2(\mathbf{x}), \cdots, F_N(\mathbf{x}))^T \in \mathbb{R}^N$ that uniquely describes the appearance of a voxel's surrounding. This includes

features describing the surrounding of a voxel in a coarser scale, thus the rough position of the voxel in the brain can be determined. On the other hand, the feature vector comprises features describing finer details of the close neighborhood of a voxel. Our experiments will show that the resulting features are highly discriminative. Such feature images of a HARDI signal of a human brain are shown in figure 2. Details are given later in this section.

Our assumption is that the features guarantee a unique representation of each landmark. Hence we can use the maximum response of a linear classifier to determine the landmark positions. This is probably the fastest way to detect the landmarks. In a supervised training step we learn weights for the linear classifier for all landmarks $l = \{1, \cdots, M\}$. We denote by $\mathcal{H}_l \in \mathbb{R}^3 \rightarrow \mathbb{R}$ the evidence image for the position of landmark l:

$$\mathcal{H}_l(\mathbf{x}) := \sum_i \alpha_i(l) F_i(\mathbf{x}) = \boldsymbol{\alpha}(l)^T \mathbf{F}(\mathbf{x}) \quad . \tag{1}$$

See figure 1 a) for some examples of $\mathcal{H}_l$ for different landmarks l. With $\boldsymbol{\alpha}(l) \in \mathbb{R}^N$ we denote the weights corresponding to landmark l. The global maximum of $\mathcal{H}_l$ is considered as the prediction for a landmark position. It is worth noting that due to the fact that eq. (1) is linear we can use a simple least square fit $\underset{\boldsymbol{\alpha}(l)}{\operatorname{argmin}} \| \sum_{\mathbf{x}} \mathcal{H}_l^{\text{train}}(\mathbf{x}) - \boldsymbol{\alpha}(l)^T \mathbf{F}^{\text{train}}(\mathbf{x}) \|^2$ to compute the weights in a training step. $\mathbf{F}^{\text{train}}$ are features of training images and $\mathcal{H}_l^{\text{train}}$ are binary valued label images where the desired landmark position has been set to 1. Note that due to the sparseness of $\mathcal{H}_l^{\text{train}}$ the system of equations can be solved for all landmarks simultaneously in a memory efficient way within seconds. Once the weights are determined we can find the landmarks in any new image by computing the features and attaching the weights according to eq. (1).

In our experiment we aim at detecting more than 10^4 landmarks. Successively computing the evidence images $\mathcal{H}_l$ followed by the determination of the global maximum for each single landmark is far too computational expensive. Since the landmarks are unique we can determine the predicted landmark position in two steps: 1) We down-sample the feature images by a factor of 4 (we just consider every fourth voxel). We then compute the evidence $\mathcal{H}_l(\mathbf{x})$ for each landmark voxel by voxel "on the fly" and store the position of the highest result for each landmark. 2) We then search for the precise location of the maxima in the surrounding of each previously stored position. Given the feature images we can find about 22000 landmarks in a $100 \times 100 \times 69$ brain image in about 5 minutes[1].

Computing the feature images F: We utilize spherical tensor algebra (STA) [4, 5], which allows to compute dense, multi-scale feature images from HARDI data in an efficient and rotation invariant way. The use of STA is quite reasonable, because it is common to represent HARDI images by Spherical Harmonics (SH). The orientation distribution in each voxel is encoded in the SH basis providing a memory efficient and smooth way to handle the data.

[1] using 4 cores of an Intel Xeon CPU X7560 with 2.27GHz

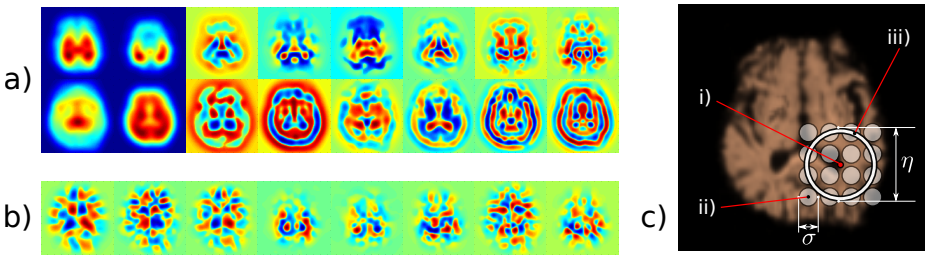

Fig. 2. a) Rotation invariant features according to eq. (4) of a human brain (center slice). They are invariant with respect to reflection symmetry, too. b) We additionally use features that can clearly distinguish between the left and right hemisphere (the sign differs, eq. (5)) c) Features for a landmark i) are computed in two steps: We first densely compute local nonlinear image features ii). Based on these features in a larger neighborhood iii) around i) we form the final rotation invariant features.

We first decompose the HARDI signal $\mathbf{f} : \mathbb{R}^3 \times S_2 \to \mathbb{R}$ into its basic angular frequency components by orthogonal projection onto the SH basis functions $\mathbf{Y}^\ell : S_2 \to \mathbb{C}^{2\ell+1}$ voxel by voxel [4]. With each index ℓ a certain angular frequency is represented. Since diffusion is symmetric, the HARDI signals are symmetric, too. Hence we only need to consider SH associated with an even index. Consequently we can represent $\mathbf{f}$ in terms of $\mathbf{Y}^\ell$ by $\mathbf{f}(\mathbf{x}, \mathbf{n}) = \sum_{\substack{\ell=0 \\ \ell \text{ even}}}^{\infty} \mathbf{a}^\ell(\mathbf{x})^T \mathbf{Y}^\ell(\mathbf{n})$. We denote by $\mathbf{a}^\ell(\mathbf{x}) \in \mathbb{C}^{2\ell+1}$ the vector valued expansion coefficients representing the HARDI signal of $\mathbf{f}$ at image position $\mathbf{x} \in \mathbb{R}^3$ in the SH-domain.

In our framework we only use the coefficient images $\mathbf{a}^0 : \mathbb{R}^3 \to \mathbb{C}$ and $\mathbf{a}^2 : \mathbb{R}^3 \to \mathbb{C}^5$. $\mathbf{a}^0$ represents the mean of the signal. $\mathbf{a}^2$ gives us information about the diffusion directions and heavily contributes in the white matter regions to the HARDI signal. Considering higher frequency components, i.e. $\mathbf{a}^4$ in our experiments did not lead to better results.

The raw coefficients $\mathbf{a}^0$ and $\mathbf{a}^2$ are only describing the very local properties of the tissue and thus are far not sufficient to yield enough information to uniquely represent a landmark in the HARDI signal. Due to this reason we designed new nonlinear features for representing the neighborhood around a respective voxel. The resulting features are rotation invariant thus no pre-alignment of the images is required. The features are computed in two steps: First, non-linear local image features are densely computed in a close neighborhood of a voxel (Fig. 2 ii)). A second step combines the non-linear features in a larger surrounding of a voxel and forms its unique feature signature (Fig. 2 iii)).

STA provides two basic operations to deal with the coefficient images $\mathbf{a}^0$ and $\mathbf{a}^2$. These operations do not alter the rotation behavior of the SH representation, that is, they allow to compute rotation invariant features in a systematic way. The first class of operations are finite difference operators that connect SH representations of different degrees by differentiations, so-called *spherical tensor derivatives* $\boldsymbol{\nabla}^n$ [5]. We distinguish between spherical up-derivatives, where $n > 0$ and spherical down-derivatives where $n < 0$. The first operator increases

the tensor rank by n, the latter one decreases the tensor rank by n. The second class of operations are products that connect two different SH representations to form a new field with a different degree, called *spherical products* $\circ_\ell : \mathbb{C}^{2\ell_1+1} \times \mathbb{C}^{2\ell_2+1} \to \mathbb{C}^{2\ell+1}$ [5]. They couple *spherical tensors* associated with different *orders* ℓ_1, ℓ_2 to form new tensors of higher or lower order ℓ.

We obtain the local non-linear image descriptors (Fig. 2 ii)) in the following way: We first expand the local neighborhood of voxels in $\mathbf{a}^0$ and $\mathbf{a}^2$ in terms of spherical Gaussian derivatives by initially convolving the coefficient images with a Gaussian G_σ followed by successively computing the tensor derivatives [7]:

$$\mathbf{b}_0^\ell := \boldsymbol{\nabla}^\ell(\mathcal{G}_\sigma * \mathbf{a}^0), \qquad 0 \leq \ell < L,\ \mathbf{b}_0^\ell(\mathbf{x}) \in \mathbb{C}^{2\ell+1} \tag{2}$$

$$\mathbf{b}_2^\ell := \boldsymbol{\nabla}^{\ell-2}(\mathcal{G}_\sigma * \mathbf{a}^2), \qquad 0 \leq \ell < L,\ \mathbf{b}_2^\ell(\mathbf{x}) \in \mathbb{C}^{2(2+\ell)+1} \quad . \tag{3}$$

We denote by $\mathbf{b}_0^\ell(\mathbf{x}), \mathbf{b}_2^\ell(\mathbf{x})$ the expansion coefficients of the local neighborhood of $\mathbf{a}^0(\mathbf{x})$ and $\mathbf{a}^2(\mathbf{x})$, respectively. The upper bound $L \in \mathbb{N}$ restricts the number of expansion coefficients. The size of the local neighborhood representations is defines by σ, the size of the Gaussian. Then we use the spherical tensor products $\circ_\ell$ to form new, nonlinear representations $(\mathbf{b}_a^{\ell_1}(\mathbf{x}) \circ_\ell \mathbf{b}_a^{\ell_2}(\mathbf{x})) \in \mathbb{C}^{2\ell+1}$ voxel by voxel.

Finally, we follow ideas proposed by the Harmonic Filter framework (HF) [5] to form the final rotation invariant large neighborhood descriptors (Fig. 2 iii)). The HF is some kind of voting based approach for generic object detection where local image descriptors are voting for the presence of objects. Voting offers several advantages: The detection of objects is very robust with respect to occlusions, intra-class variations and deformations! In our framework we adopt the idea of voting and consider it as a collection of local descriptors in a voxel's surrounding. Mathematically this step coincides with the voting of the HF thus we gain rotation invariant features in the following way:

$$F_i(\mathbf{x}) := \mathcal{G}_\eta * (\boldsymbol{\nabla}^{(-\ell)}(\mathbf{b}_a^{\ell_1} \circ_\ell \mathbf{b}_a^{\ell_2})) \qquad \ell \leq L \quad . \tag{4}$$

A larger choice of η leads to image descriptors representing the rough position of the voxel in the brain. As small η leads to features representing local details of the HARDI signal. Figure 2 a) shows some examples of such features based on the HARDI signal of a human brain. Note that $L \in \mathbb{N}$ restricts the number of possible products i.e. the number of feature images.

Considering our aims there exists one significant drawback of these features: they are invariant against reflection about an axis. Hence they can't distinguish the left and the right hemisphere. Figure 2 a) illustrates this problem. It is known that the spherical triple-correlation [1] yields complete rotation invariant features. Hence they must solve this issue. Based on this idea we designed new 3^{rd} order rotation invariant differential features fitting into our framework that are variant with respect to reflections about an axis:

$$F_j(\mathbf{x}) := \mathcal{G}_\eta * (\boldsymbol{\nabla}^{(-\ell_4)}((\mathbf{b}_a^{\ell_1} \circ_\ell \mathbf{b}_a^{\ell_2}) \circ_{\ell_4} \mathbf{b}_a^{\ell_3})), \quad \begin{array}{l} \ell_1+\ell_2+\ell_3+\ell_4 \text{ is odd} \\ \text{and } \ell_4, \ell \leq L \end{array} \quad . \tag{5}$$

The proof is given in the appendix. Figure 2 b) shows some examples.

Experiments For our experiments 21 in vivo diffusion acquisitions of human brains were acquired on a Siemens 3T TIM Trio using an SE EPI sequence with a TE of 95 ms and a TR of 8.5 s and an effective b-value of 1000. One voxel corresponds to $2mm^3$. We used 67 directions which we entirely fit to spherical harmonics. For evaluating the performance of our detection system we conducted the following experiment: The 21 HARDI images of healthy volunteers have been co-registered using the SPM[2] toolkit. Based on this co-registration we found the position of 20685 landmarks (per image) in the original image domain. The landmarks were densely distributed in the brain white and gray matter (one landmark at every second voxel position with respect to X,Y, and Z direction). We consider these co-registered landmarks as a reference which we use for training the classifier and for evaluating our approach. It is worth noting that it is difficult to co-register the noisy, tensor-valued HARDI images thus that the true "ground-truth" can hardly be provided. But if the positions of the landmarks detected by our algorithm are similar to the co-registered positions, than there is high evidence that most of the landmarks have been detected correctly.

The HARDI images have been transformed to the SH-domain and features have been computed for each image as described above. We first computed the expansion coefficients based on eq. (2) and eq. (3) using a Gaussian with $\sigma = 6mm$. We experienced that the signal corresponding to the brain white matter leads to the most reliable features thus we additionally expanded $\mathbf{a}^2$ with respect to a larger neighborhood $\sigma = 5mm$ (eq. (3)). We then computed features based on eq. (4) and eq. (5) using $L = 5$ and three different scales, namely with respect to $\eta = 4, 8$ and $12mm$. We found these parameters via a leave-one-out parameter grid-search on the training set. Using these parameters we gained 528 discriminative feature images per HARDI signal.

We used one third of the images (7 images) for training the linear classifier (eq. 1). The co-registered landmark positions where used for determining the filter parameters $\boldsymbol{\alpha}(l)$ for all landmarks. The remaining 14 images were used for evaluation.

Results & Discussion The displacements of the filter responses with respect to the co-registered reference positions can be found in table 1. The results clearly show that most of the detected landmark positions are very close to the ground

[2] SPM (Statistical Parametric Mapping version 5), http://www.fil.ion.ucl.ac.uk/spm/

Table 1. Correctly detected landmarks. Correctly means the detected landmark position is similar to its reference position. The column in red corresponds to the worst results. The table shows results for an increasing tolerated displacement.

tolerance	correctly detected landmarks for the 14 test datasets (totally 20685 unique landmarks)													
1 voxel	39.0%	26.6%	39.2%	36.5%	51.9%	40.4%	37.2%	51.0%	38.7%	49.4%	46.1%	38.7%	51.8%	44.7%
2 voxel	81.5%	65.7%	79.8%	77.0%	90.1%	81.1%	79.3%	89.3%	81.0%	88.6%	86.6%	79.1%	88.5%	85.8%
3 voxel	96.9%	90.7%	96.3%	95.4%	98.9%	96.7%	96.1%	98.5%	95.8%	97.5%	98.7%	95.7%	98.2%	97,8%
4 voxel	99.4%	96.1%	99.4%	98.7%	99.8%	99.4%	99.0%	99.6%	98.1%	99.1%	99.8%	98.3%	99.4%	99.6%
5 voxel	99.9%	97.9%	99.9%	99.6%	99.9%	99.9%	99.7%	99.9%	99.0%	99.4%	99.9%	99.3%	99.7%	**100**%

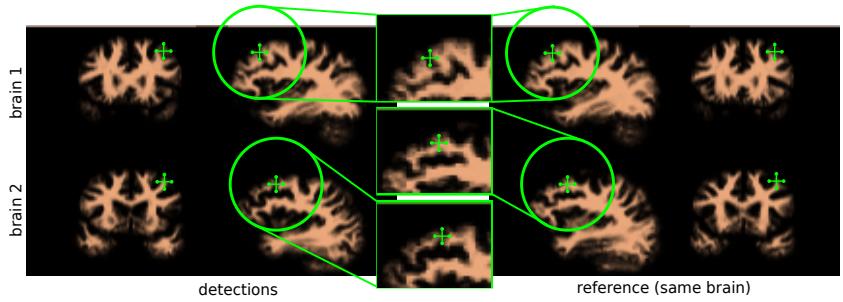

Fig. 3. The white matter mask of the HARDI signal [6] together with the landmark positions. We counted a detection as successful if the predicted landmark position was close to its reference position. Since the filter encodes the appearance of local structures in several granularities it can use coarser representations to find the global position in the brain while finer representations are used to adapt the position according to the local neighborhood configuration. This often shows more plausible results than the reference positions as exemplarily shown in the second row. the last row).

truth. Only a very small number (in the worst case $< 2.1\%$) of the landmark positions differ by more than 5 voxel, which corresponds to a displacement of $1cm$.

For a qualitative analysis of the results we computed a white matter mask directly from the HARDI signal [6]. This ensures consistency with the data. Figure 3 gives qualitative results. We observed that the detected landmarks were often more consistent with the local structure of the HARDI images than the co-registered reference location (Fig. 3). The detection of all 20685 landmarks takes about 16 minutes[1] per image.

We further conducted experiments based on five images of patients showing pathologies. Since no ground truth was available we visually compared the detection results of 50 landmarks with their position in images of healthy volunteers. Our approach was able to successfully detect all landmarks in the healthy areas of the brain. It is worth noting that we also detected the landmarks located very close to pathological areas (Fig. 4).

3 Conclusion

In this paper, we have presented a new framework that allows the detection of a large number of unique point landmarks within the tensor valued HARDI images of the human brain. Our experiment has shown that based on new image features in combination with a fast linear classifier the landmarks can be reliably detected in reasonable time. We make the source code publicly available after acceptance of this paper.

Appendix Eq. (5) is variant with respect to reflection about an axis: We consider w.o.l.g the reflection about the origin. It holds that $\mathbf{Y}^\ell(-\mathbf{n}) = (-1)^\ell \mathbf{Y}^\ell(\mathbf{n})$.

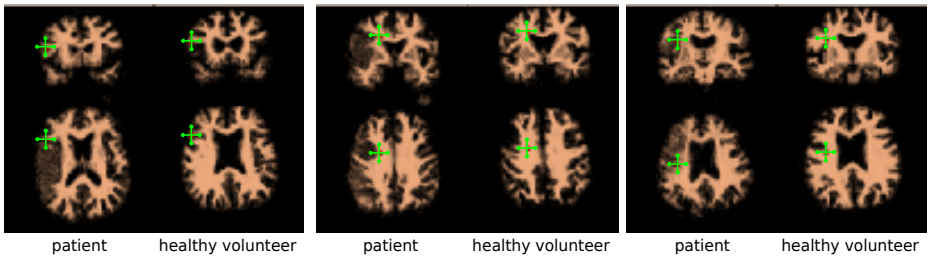

Fig. 4. Detected landmarks in patients with pathologies and healthy volunteers. Although the images strongly differ, the landmarks have been detected correctly in the healthy areas of the patients images.

Let $f(\mathbf{r},\mathbf{n}) : \mathbf{R}^3 \times S_2 \to \mathbb{R}$. When $f(\mathbf{r},\mathbf{n}) = \sum_\ell \left(\mathbf{a}^\ell(\mathbf{r})\right)^T \mathbf{Y}^\ell(\mathbf{n})$ we have $f'(\mathbf{r},\mathbf{n}) = f(-\mathbf{r},-\mathbf{n}) = \sum_\ell \left(\mathbf{b}^\ell(-\mathbf{r})\right)^T \mathbf{Y}^\ell(\mathbf{n})$, with $\mathbf{b}^\ell(-\mathbf{r}) = (-1)^\ell \mathbf{a}^\ell(\mathbf{r})$. If $\ell_1 + \ell_2 + \ell_3$ is odd, then $((\mathbf{b}^{\ell_1}(-\mathbf{r}) \circ_\ell \mathbf{b}^{\ell_2}(-\mathbf{r})) \circ_{\ell_4} \mathbf{b}^{\ell_3}(-\mathbf{r})) = (-1)^{\ell_1+\ell_2+\ell_3}((\mathbf{a}^{\ell_1}(\mathbf{r}) \circ_\ell \mathbf{a}^{\ell_2}(\mathbf{r})) \circ_{\ell_4} \mathbf{a}^{\ell_3}(\mathbf{r}))$. With $(\boldsymbol{\nabla}^{(-\ell)}\mathbf{b}^\ell)(-\mathbf{r}) = (-1)^\ell (\boldsymbol{\nabla}^{(-\ell)}\mathbf{a}^\ell)(\mathbf{r})$ we can conclude that if $\ell_1 + \ell_2 + \ell_3 + \ell_4$ is odd, then $(\boldsymbol{\nabla}^{(-\ell_4)}((\mathbf{b}^{\ell_1} \circ_\ell \mathbf{b}^{\ell_2}) \circ_{\ell_4} \mathbf{b}^{\ell_3}))(-\mathbf{r}) = -(\boldsymbol{\nabla}^{(-\ell_4)}((\mathbf{a}^{\ell_1} \circ_\ell \mathbf{a}^{\ell_2}) \circ_{\ell_4} \mathbf{a}^{\ell_3}))(\mathbf{r})$.

References

1. R. Kondor. *Group theoretical methods in machine learning.* PhD thesis, Columbia University, 2008.
2. Jimin Liu, Wenpeng Gao, Su Huang, and W. L. Nowinski. A Model-Based, Semi-Global Segmentation Approach for Automatic 3-D Point Landmark Localization in Neuroimages. *IEEE Trans. Med. Imaging*, 27(8):1034–1044, 2008.
3. K. Rohr, H. S. Stiehl, R. Sprengel, T. M. Buzug, J. Weese, and M. H. Kuhn. Landmark-based elastic registration using approximating thin-plate splines. *IEEE Trans. Med. Imaging*, 20(6):526–534, jun 2001.
4. M. Rose. *Elementary Theory of Angular Momentum.* Dover Publications, 1995.
5. M Schlachter, M Reisert, C Herz, F Schluermann, S Lassmann, M Werner, H Burkhardt, and O Ronneberger. Harmonic Filters for 3D Multi-Channel Data: Rotation Invariant Detection of Mitoses in Colorectal Cancer. *IEEE Trans. Med. Imaging*, 29(8):1485–1495, 2010.
6. H. Skibbe, M. Reisert, and H. Burkhardt. Gaussian neighborhood descriptors for brain segmentation. In *Proc. of the MVA*, Nara, Japan, 2011.
7. H. Skibbe, M. Reisert, T. Schmidt, T. Brox, O. Ronneberger, and H. Burkhardt. Fast Rotation Invariant 3D Feature Computation utilizing Efficient Local Neighborhood Operators. *IEEE Trans. on PAMI*, 2012.
8. DS. Tuch, RM. Weisskoff, JW. Belliveau, and VJ. Wedeen. High angular resolution diffusion imaging of the human brain. In *Proc. of the ISMRM*, Philadelphia,USA, 1999.
9. Z. Xue, D. Shen, and C. Davatzikos. Determining correspondence in 3-D MR brain images using attribute vectors as morphological signatures of voxels. *IEEE Trans. Med. Imaging*, 23(10):1276–1291, October 2004.