3D Rotation-Invariant Description from Tensor Operation on Spherical HOG Field

Kun Liu[1,3]
liu@informatik.uni-freiburg.de

Henrik Skibbe[1,3]
skibbe@informatik.uni-freiburg.de

Thorsten Schmidt[1,3]
tschmidt@informatik.uni-freiburg.de

Thomas Blein[2,3]
thomas.blein@biologie.uni-freiburg.de

Klaus Palme[2,3]
klaus.palme@biologie.uni-freiburg.de

Olaf Ronneberger[1,3]
ronneber@informatik.uni-freiburg.de

[1] Computer Vision and Image Analysis Group
University of Freiburg
Germany
http://lmb.informatik.uni-freiburg.de

[2] Institute of Biology II - Botany
University of Freiburg
Germany

[3] BIOSS Centre for Biological Signalling Studies
University of Freiburg
Germany

Abstract

Rotation-invariant descriptions are required in many 3D volumetric image analysis tasks. The histogram-of-oriented-gradient (HOG) is widely used in 2D images and proves to be a very robust local description. This paper concentrates on how to use the HOG feature in 3D volumetric images when rotation-invariance is concerned. This is challenging because of the complexity of 3D rotations. We present a decent solution based the spherical harmonics theory which is an effective tool for analysing 3D rotations, together with the spherical tensor operations which explore high order tensor information in spherical coordinates. The design is quite general and could be used for different applications. It achieves high scores on Princeton Shape Benchmark and SHREC 2009 Generic Shape Benchmark, and also produces promising results when applying on biological microscopy images.

1 Introduction

In 3D volumetric images, it is very common that similar objects presented in different orientations need to be recognized as the same class [10, 19, 22]. When objects have unknown poses, starting from rotation-invariant descriptions can make the following analysis easier since they are pose-independent. Besides, a good description should be able to capture substantial image patterns, and be robust to small object deformation or image distortion. Gradient-histogram based features, like HOG [6] and SIFT [15], are widely used for 2D image description and prove to be very robust [9, 26]. Concerning the rotation-invariance, HOG always needs to be constructed on a certain coordinate, and SIFT achieves the invariance on

| (a) | (b) | (c) | (d) |
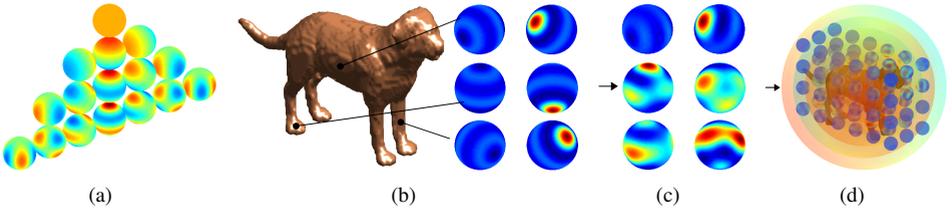
Figure 1: The basic ideas: (a) *Spherical harmonics* for representing HOG (real part of the first 4 bands are shown). (b) Individual gradient as function on a sphere (reconstructed from the *SH* coefficients). (c) The spatially aggregated *spherical HOG*. (d) *Spherical tensor* operations create rotation-invariant descriptions from the *spherical HOG field*.

detected interest points by normalizing (aligning) the local coordinate to the dominant gradient directions. The basic HOG feature employs a histogram binning on gradient direction and a spatial aggregation, with soft binning. The spatial aggregation works as a "blurring" effect, which discounts the importance of exact alignment between features. Because the important gradient patterns in different directions have already been separated into histogram bins, they can be well preserved under this "blurring".

Some work is recently reported on extending SIFT or HOG feature to 3D data [2, 10, 13, 21]. The approach proposed in [21] computes HOG features on pose-normalized 3D shapes. Allaire *et al.* [2] proposed a 3D SIFT. Their method normalizes the local coordinate directions by checking dominant gradient directions in two steps. A similar method is employed in [10]. However, these normalization methods only work well for objects or positions with certain conditions. Like in the 2D case, this rotation-invariant SIFT is based on an interest point detection, because the decision for a local coordinate system is not reliable for arbitrary positions. In fact, the interest point detection is usually insufficient for complex tasks. For example, many state-of-the-art 2D object recognition algorithms skip this step and use the non-invariant dense HOG features with a sliding window [9, 26]. On another aspect, the complexity of 3D rotations makes the normalization methods even less reliable in 3D images. While only one angle needs to be determined for an aligned coordinate in 2D, three angles are required for the 3D case (like Euler angles).

In contrast, there exist analytic methods which guarantee the rotation invariance directly, without any dependence on pose normalization. A fundamental method to compute such invariant features is group integration [18]. Closely related methods come from Fourier analysis in spherical coordinates [12, 25], where spherical harmonics (*SH*) are used as the angular basis, and rotation-invariant features can be created by taking the magnitude of *SH* coefficients in different bands. Recently, high order information like vector and tensors are also utilized in the *SH* framework [7, 24]. They bring better representation for local structures. However, these *SH* based methods have no special ability to deal with object deformation and image distortion, which proves to be important for robust description. So we are motivated to take advantage of some HOG-like feature to enhance the performance of *SH* based descriptions.

In this paper, we will present an approach to embed HOG into the *SH* framework, producing a high-performance 3D rotation-invariant description. First the preliminaries about the *SH* and spherical tensors are given in the next section. Sec. 3 introduces a *spherical HOG* representation which connects HOG with the *SH* framework. Then Sec. 4 presents how to create rotation-invariant descriptions from the obtained *spherical HOG* field. The experi-

mental part demonstrates the performance of our approach on 3D shape retrieval benchmarks and a practical application on 3D confocal biological images.

2 Preliminaries for Working on the 2-sphere

In 2D, an angular signal on a certain radius (a circle) could be well described by a one-dimensional Fourier transform [25]. In 3D, to describe a signal on a sphere with 3D rotations, the tools we need are the spherical harmonics (*SH*) and the rotation group *SO*(3). The gentle introduction from Green [11] might be a good start for readers who are unfamiliar with *SH*. Here we give an introduction about some important conceptions for working on the 2-sphere(S^2), in the scope of this paper. For more details, we refer to [3, 17, 20].

2.1 Spherical Harmonics

Spherical harmonics $Y_m^\ell : S^2 \to \mathbb{C}$ form an orthonormal basis for the 2-sphere. Any square-integrable scalar function on a sphere, can be expanded into a linear combination as:

$$f(\theta,\phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \overline{\hat{f}_m^\ell} Y_m^\ell(\theta,\phi) = \sum_{\ell=0}^{\infty} \hat{\mathbf{f}}^{\ell\top} \mathbf{Y}^\ell(\theta,\phi) \tag{1}$$

where $\mathbf{Y}^\ell(\theta,\phi)$ and $\hat{\mathbf{f}}^\ell \in \mathbb{C}^{2\ell+1}$, ℓ denotes the band of expansion, $m \in [-\ell,\ell]$ is the order in the ℓ-th band, $\mathbf{x}^\top = \overline{\mathbf{x}}^T$ is the conjugate transpose. The angular part of spherical coordinate is defined as $\theta \in [0,\pi]$ and $\phi \in [0,2\pi)$ (colatitude and longitude). To describe an integrable function $f(\mathbf{r}) : \mathbb{R}^3 \to \mathbb{C}$, we can write

$$f(\mathbf{r}) = \sum_{\ell=0}^{\infty} \hat{\mathbf{f}}^\ell(r)^\top \mathbf{Y}^\ell(\mathbf{r}) \tag{2}$$

where $r = |\mathbf{r}|$, and $\mathbf{Y}^\ell(\mathbf{r})$ only depend on the angle of $\mathbf{r}$. The Schmidt semi-normalized *SH* are defined as $Y_m^\ell(\theta,\phi) = \sqrt{\frac{(\ell-m)!}{(\ell+m)!}} P_m^\ell(\cos\theta)e^{im\phi}$. P_m^ℓ are the associated Legendre polynomials. They have the normalization relationship $< Y_m^\ell, Y_{m'}^{\ell'} > = \int_{S^2} Y_m^\ell \overline{Y_{m'}^{\ell'}} d\Omega = \frac{4\pi}{2\ell+1}\delta_{\ell,\ell'}\delta_{m,m'}$. Considering the normalization and Equation 1, the expansion coefficients are computed as $\hat{f}_m^\ell = \frac{2\ell+1}{4\pi}<f,Y_m^\ell> = \frac{2\ell+1}{4\pi}<Y_m^\ell,f>$.

Without loss of generality, here we can always assume the rotation is around the origin. To analyse the rotation behaviour of expansion coefficients, we need the so called Wigner D-matrices [21], which are the irreducible representation of a rotation $\mathfrak{g} \in SO(3)$. They are determined by the 3D rotation angles, and denoted by unitary matrices $\mathbf{D}^\ell \in \mathbb{U}^{2\ell+1}$, for the rotation in the ℓth band. Let $\mathbf{U}_\mathfrak{g}$ be the rotation matrix in Cartesian coordinate, $\mathbf{D}_\mathfrak{g}^\ell$ will keep the *SH* basis fixed under the rotation as

$$[\mathfrak{g}\mathbf{Y}^\ell](\mathbf{r}) = \mathbf{D}_\mathfrak{g}^\ell \mathbf{Y}^\ell(\mathbf{U}_\mathfrak{g}^T \mathbf{r}) = \mathbf{Y}^\ell(\mathbf{r}) \tag{3}$$

If we apply a rotation on the 3D function $f(\mathbf{r})$ in Equation 2, we get $[\mathfrak{g}f](\mathbf{r}) = f(\mathbf{U}_\mathfrak{g}^T\mathbf{r}) = \sum_\ell \hat{\mathbf{f}}^\ell(r)^\top \mathbf{Y}^\ell(\mathbf{U}_\mathfrak{g}^T\mathbf{r}) = \sum_\ell (\mathbf{D}_\mathfrak{g}^\ell \hat{\mathbf{f}}^\ell(r))^\top \mathbf{Y}^\ell(\mathbf{r})$. So the *SH* expansion coefficients for a rotated function just look like $\mathbf{D}_\mathfrak{g}^\ell \hat{\mathbf{f}}^\ell$.

2.2 Spherical Tensors

Spherical tensors were first proposed for image processing in [16]. A function $\mathbf{f}: \mathbb{R}^3 \to \mathbb{C}^{2\ell+1}$ is called a rank-ℓ spherical tensor field (denoted by $\mathscr{T}_\ell$), if it transforms with respect to rotations as

$$[\mathfrak{g}\mathbf{f}](\mathbf{r}) = \mathbf{D}_\mathfrak{g}^\ell \mathbf{f}(\mathbf{U}_\mathfrak{g}^T \mathbf{r}) \tag{4}$$

Clearly this rotation behaviour is simpler than Cartesian tensors. $\mathbf{Y}^\ell(\mathbf{r})$ is an example for a spherical tensor field of rank ℓ. Spherical tensors have the same component-wise addition rule as Cartesian tensors. But for multiplication, a different operation to couple two spherical tensors $\circ_\ell : \mathbb{C}^{2\ell_1+1} \times \mathbb{C}^{2\ell_2+1} \to \mathbb{C}^{2\ell+1}$ is defined as

$$(\mathbf{v} \circ_\ell \mathbf{w})_m = (\mathbf{e}_m^\ell)^T(\mathbf{v} \circ_\ell \mathbf{w}) := \sum_{m=m_1+m_2} <\ell m|\ell_1 m_1, \ell_2 m_2> v_{m_1} w_{m_2} \tag{5}$$

where $\mathbf{v} \in \mathscr{T}_{\ell_1}$, $\mathbf{w} \in \mathscr{T}_{\ell_2}$, $\mathbf{e}_m^\ell$ is the ℓ-dimensional natural basis. $<\ell m|\ell_1 m_1, \ell_2 m_2>$ are real co-efficients (called *Clebsch-Gordan* coefficients [20]), which only have non-zero values when $m = m_1 + m_2$ and $|\ell_1 - \ell_2| \leq \ell \leq \ell_1 + \ell_2$. The coupling result $\mathbf{v} \circ_\ell \mathbf{w}$ is a spherical tensor of rank ℓ. This operation provides a complete and orthogonal expansion for the angular part of arbitrary spherical tensor fields, that is, for $\mathbf{f} \in \mathscr{T}_\ell$:

$$\mathbf{f}(\theta, \phi) = \sum_{j=0}^{\infty} \sum_{k=-j}^{j} \hat{\mathbf{a}}_k^j \circ_\ell \mathbf{Y}^j(\theta, \phi) = \sum_{j=0}^{\infty} \sum_{k=-j}^{j} \sum_{m=-(j+k)}^{j+k} \hat{a}_{k,m}^j \mathbf{Z}_{k,m}^j(\theta, \phi) \tag{6}$$

where $\hat{\mathbf{a}}_k^j \in \mathscr{T}_{j+k}$ are tensorial expansion coefficients, $\mathbf{Z}_{k,m}^j := \mathbf{e}_m^{(j+k)} \circ_\ell \mathbf{Y}^j$ is the tensorial basis for rank-ℓ spherical tensor fields. Refer to [17] for more details and proofs. A variant of this coupling is defined as $\mathbf{v} \bullet_\ell \mathbf{w} := \frac{1}{<\ell 0|\ell_1 0, \ell_2 0>} \mathbf{v} \circ_\ell \mathbf{w}$, which will be used later.

3 Spherical HOG Representation

The most straightforward way to use gradient information in spherical coordinates is to transform the gradient into a spherical vector, namely the rank-1 spherical tensor. This kind of approaches are reported in [7, 24]. But those approaches do not have the advantage like HOG feature has. The local contrasting gradient information could be compromised when smoothing is applied to address small deformation and disturbance.

In contrast, we will start with the construction of a *spherical HOG* feature in 3D. A small trick which connects HOG to *SH* framework comes from a simple observation: although a histogram is often shown in a discrete manner, the original information it encodes is a continuous distribution. A typical image of 2D gradient histograms is shown in Figure 2(a). In comparison, the representation as circular signals shown in Figure 2(b) are equal and even more accurate, while only low-order Fourier series are enough to encode them. If we construct a 3D HOG as a signal on the 2-sphere, we can work in a fully continuous setting (some discrete quantization error can be avoided), and rotations could be easily addressed by using *SH* coefficients to represent the signal. Following is the detail. Refer to [6] for the construction of 2D HOG features in Cartesian coordinates.

The first step is to create raw histograms voxel-wisely. Let the intensity gradient computed in the image be a vector field $\mathbf{d}(\mathbf{r})$ and its spherical coordinate representation be $[\theta_d(\mathbf{r}), \phi_d(\mathbf{r}), d(\mathbf{r})]$. A raw *spherical HOG*, which only describes the gradient distribution

(a)   (b)   (c)
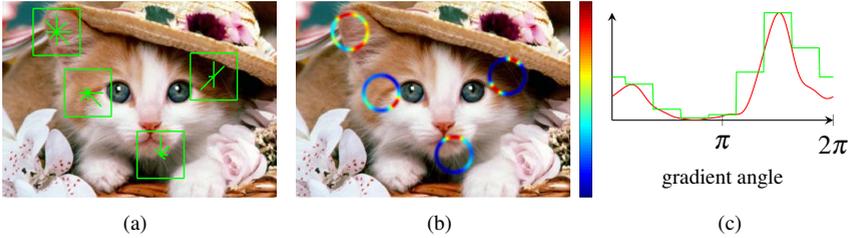
Figure 2: (a) Histogram of oriented gradient (b) Circles representing the histograms (c) 1D illustration of the histogram at bottom (All the continuous signals are reconstructed from first 8 Fourier coefficients after Gaussian filtering in the frequency domain.)

at one voxel, should be an oriented impulse signal on the 2-sphere. That is, a Dirac function $\delta(\theta_d(\mathbf{r}), \phi_d(\mathbf{r}))$ of height $d(\mathbf{r})$. Its *SH* coefficients compute to

$$\hat{\mathbf{d}}^\ell = \frac{2\ell+1}{4\pi} <\mathbf{Y}^\ell, d \cdot \delta(\theta_d, \phi_d)> = \frac{2\ell+1}{4\pi} d\mathbf{Y}^\ell(\theta_d, \phi_d) = \frac{2\ell+1}{4\pi} d\mathbf{Y}^\ell(\mathbf{d}) = \frac{2\ell+1}{4\pi}\mathbf{R}^\ell(\mathbf{d}) \quad (7)$$

where $\mathbf{R}^\ell := |\mathbf{r}|\mathbf{Y}^\ell(\mathbf{r})$. The projected *SH* coefficients look redundant as a representation for only one vector. However, what we start building are "histograms". Here *SH* is nothing more than a convenient basis chosen for the continuous histogram, which has the same role as the natural basis $\mathbf{e}_i$ for discrete histogram vectors.

"Soft binning" for both gradient direction and spatial aggregation are considered in the 2D HOG feature. In our setting, the angular smoothing for gradient direction surely could be done by a Gaussian smoothing on the sphere, which is called Spherical Diffusion[4]. Based on the convolution theory on the 2-sphere, the smoothing could be applied efficiently as a weighting in the frequency domain $\hat{\mathbf{d}}^{\ell'} = e^{-l(l+1)t}\hat{\mathbf{d}}^\ell$, where the diffusion time t controls the smoothing extent. The diffusion will restrict the bandwidth of the represented signal, so it also has the anti-aliasing effect for the band-limited *SH* transform. However, we found this weighting is not critical for performance, as we keep working in the frequency domain. Except aliasing, the band-limited *SH* transform fulfils the "soft binning" automatically.

For the spatial aggregation with "soft binning", we just need to convolve a Gaussian kernel g_σ with all *SH* coefficients component-wisely, as $\tilde{d}^\ell_m = \hat{d}^\ell_m * g_\sigma$. Finally, the local variance normalization can also be implemented with convolutions which have isotropic effect, instead of using any grid-block based operations.

Now we have the *SH* coefficients $\tilde{\mathbf{d}}^\ell(\mathbf{r})$ at every voxel, representing the local Gaussian-weighted spherical HOG. Because of the "soft binning", we only need low-band *SH* coefficients. The construction of spherical HOG is illustrated in Figure 1(b)(c). The signals shown on the spheres are reconstructed from *SH* expansion up the 5th band. The spherical HOG has all the advantages of the 2D HOG feature. The spatial smoothing blurs the localization of gradient, and only the neighbouring gradient signals in similar directions will aggregate together.

4 Tensor Operations for Regional Description

The spherical HOG constructed in the last section collects gradient information locally. Only with smoothing, we can not effectively expand the description range without losing detail.

In SIFT and sliding window techniques, histograms at neighbouring grids are concatenated into a region description. Now we come to the same step.

As the spherical HOG will rotate together with the image, it follows that the spherical HOG coefficients $\tilde{\mathbf{d}}^\ell(\mathbf{r})$ form spherical tensor fields, based on the preliminaries. Formally, we know that a gradient field $\mathbf{d}(\mathbf{r})$ transforms under a rotation $\mathfrak{g}$ as $[\mathfrak{g}\mathbf{d}](\mathbf{r}) = \mathbf{U}_\mathfrak{g}\mathbf{d}(\mathbf{U}_\mathfrak{g}^T \mathbf{r})$, so the spherical HOG coefficients computed on a rotated image $f(\mathbf{U}_\mathfrak{g}^T \mathbf{r})$ are

$$[\mathfrak{g}\hat{\mathbf{d}}^\ell](\mathbf{r}) = c\mathbf{R}^\ell([\mathfrak{g}\mathbf{d}](\mathbf{r})) = c\mathbf{R}^\ell(\mathbf{U}_\mathfrak{g}\mathbf{d}(\mathbf{U}_\mathfrak{g}^T \mathbf{r})) = c\mathbf{D}_\mathfrak{g}^\ell\mathbf{R}^\ell(\mathbf{d}(\mathbf{U}_\mathfrak{g}^T \mathbf{r})) = \mathbf{D}_\mathfrak{g}^\ell\hat{\mathbf{d}}^\ell(\mathbf{U}_\mathfrak{g}^T \mathbf{r}) \quad (8)$$

where $c = \frac{2\ell+1}{4\pi}$, and we need Equation 3 to know $\mathbf{R}^\ell(\mathbf{U}_\mathfrak{g}\mathbf{d}) = \mathbf{D}_\mathfrak{g}^\ell\mathbf{R}^\ell(\mathbf{d})$. Hence, $\hat{\mathbf{d}}^\ell(\mathbf{r})$ and the smoothed $\tilde{\mathbf{d}}^\ell(\mathbf{r})$ are both spherical tensor fields of rank ℓ, and we will use tensor operations to create regional descriptions from the spherical HOG field.

The basic approach to describe a region around one point, is to compute features on multiple concentric shells centred at the selected point [1]. Here we can also sample $\tilde{\mathbf{d}}^\ell(\mathbf{r})$ at multiple shells, and project them onto the tensorial basis, as Equation 6. For efficiency, the *SH* expansion of each individual tensor component will be computed first, as proposed in [24]. According to Equation 2, for the Mth component of $\tilde{\mathbf{d}}^\ell(\mathbf{r})$, we have the scalar expansion $\tilde{d}_M^\ell(\mathbf{r}) = \sum_{j=0}^\infty \sum_{n=-j}^j \hat{b}_{M,n}^j(r)\mathbf{Y}^j(\mathbf{r})$. Then the tensorial expansion coefficients for $\tilde{\mathbf{d}}^\ell(\mathbf{r})$ can be computed from these component-wise expansions by a derived relation as

$$\hat{a}_{k,m}^j(r) = \frac{2(j+k)+1}{2\ell+1} \sum_{M,n} \hat{b}_{M,n}^j(r) <\ell M|(j+k)m, jn> \quad (9)$$

See [1] for proofs. Using this orthogonal expansion, we can get a large group of expansion coefficients $\hat{\mathbf{a}}_k^j(r,\ell) \in \mathscr{T}_{j+k}$ for each tensor field $\tilde{\mathbf{d}}^\ell$ on the shell of radius r. For any coefficients of the same rank $(j+k = j'+k')$, they will transform with the same Wigner-D matrix under rotations. Because the Wigner-D matrices are unitary, their effect will be compensated in the complex inner product. So a general formula of our rotation-invariant features is

$$||\sqrt{<\hat{\mathbf{a}}_k^j(r,\ell), \hat{\mathbf{a}}_{k'}^{j'}(r',\ell')>}|| = ||\sqrt{<\mathbf{D}_\mathfrak{g}^{j+k}\hat{\mathbf{a}}_k^j(r,\ell), \mathbf{D}_\mathfrak{g}^{j'+k'}\hat{\mathbf{a}}_{k'}^{j'}(r',\ell')>}|| \text{ when } j+k = j'+k' \quad (10)$$

Clearly, there is no restriction about ℓ and r. This formula is also consistent with the *L2*-norm of coefficient $||\hat{\mathbf{a}}_k^j||$. In fact, under certain conditions, the tensor coupling $\mathbf{v} \bullet_0 \mathbf{w}$, which creates rank-0 tensors, just coincides with the complex inner product. The features created in this way can be very effective, as the coupling between tensors of the same rank provides the possibility to create much more invariant features comparing to only taking magnitude of expansion coefficients.

Till now the approach is only defined for computing features for a selected centre point, so that the the shell-wise orthogonal expansion is affordable. However, in some applications, it is necessary to get voxel-wise descriptions in the whole volume. We then need a more efficient method for that. In general, to create voxel-wise output with clear rotation behavior, we just need a filtering operation that maps a spherical tensor field to a spherical tensor field. The Spherical Gaussian Derivative (*SGD*) filtering introduced in [17] provides such a tool. It is comparable to the common Gaussian derivatives[14], but working on spherical tensor fields. The filtering can be quite efficient when multiple derivatives are required, as it can be implemented as a Gaussian filtering followed by differentiations. The only special element here is the differentiation defined on spherical tensor fields. With the spherical gradient operator $\nabla = (\frac{1}{\sqrt{2}}(\partial x - i\partial y), \partial z, -\frac{1}{\sqrt{2}}(\partial x + i\partial y))$, the spherical up-derivative ∇^1 :

$\mathscr{T}_\ell \to \mathscr{T}_{\ell+1}$ and the spherical down-derivative $\nabla_1 : \mathscr{T}_\ell \to \mathscr{T}_{\ell-1}$ are defined as $\nabla^1 \mathbf{f} = \nabla \bullet_{\ell+1} \mathbf{f}$ and $\nabla_1 \mathbf{f} = \nabla \bullet_{\ell-1} \mathbf{f}$.

To apply this *SGD* filtering on the spherical HOG coefficients, we just need to adjust the scale of the Gaussian convolved with $\hat{\mathbf{d}}^\ell$, then compute the spherical tensor derivatives of different orders, and again collect features by coupling the output of the same rank with inner product, like in Equation 10

$$||(< \nabla_q^p(\hat{\mathbf{d}}^\ell * g_\sigma), \nabla_{q'}^{p'}(\hat{\mathbf{d}}^{\ell'} * g_{\sigma'}) >)^{\frac{1}{2}}||, \quad \text{when } \ell + p - q = \ell' + p' - q' \tag{11}$$

where g_σ is the Gaussian kernel $e^{-\frac{r^2}{2\sigma^2}}$ and $*$ denotes convolution with each component.

5 Experiments

5.1 Rotation-invariant description for 3D Shapes

First we evaluate the proposed description on 3D shapes, with Princeton Shape Benchmark (PSB) [23] test partition which has 907 objects in 92 classes (examples shown in Figure 3), and SHREC 2009 Generic Shape Benchmark (SHREC) [1] which run 80 queries on 720 objects in 40 classes. The models are presented in 150^3 voxels, after normalization for translation and scale. Besides the proposed "spherical HOG + spherical tensor"(HOG-ST) approach, we implemented the standard *SH* expansion and spherical tensorial expansion on structure tensor field (StrT-ST)[24]. Following the "SHD" method in [23], we first apply a distance transform on the binary shapes, then compute local features like HOG or structure tensor. Voxel-wise spherical HOG representations are only computed up to the 4th band, leading to 5 spherical tensor fields. Then the tensorial expansion is computed according to Equation9∼10. For our method, the tensorial expansion computes to $j_{max} = 9$. For scalar *SH* expansion and structure tensor field, we compute to $\ell_{max}(j_{max}) = 16$. All the features computed on shells are weighted by the square of radius, and concatenated into a vector. We only couple the tensorial coefficients from the same shell, except in one test (HOG-ST$_R$), where features from coupling tensors on different shells are also included. The evaluation is done by computing pair-wise distances among all 3D shapes, then five measures are evaluated using the tools provided with these benchmarks: Nearest Neighbour accuracy, First Tier, Second Tier, E-measure and Discounted Cumulative Gain. In several tested parameter settings, our method gets the best performance with large spatial smoothing $\sigma = 10_{voxel}$, 12 sampled shells and L1-norm distance measure. Results are listed in Table 1. Refer to [1, 23] for the definitions of these measures and much more methods evaluated on these benchmarks.

Figure 3: Examples of the 3D shapes in Princeton Shape Benchmark

Comparing to all the *SH* related methods, our method has a large improvement. It differs from the tensorial description from structure tensor or vector fields, not only by the rank of tensors, but also by the fact that the local gradient patterns can be better preserved under smoothing in our method, as they were represented by spherical histograms. As a frequency-domain approach, *SH* related descriptions could be suspected of being vulnerable to small

Table 1: The retrieval statistics for different methods

Data	Method	NN(%)	FT(%)	ST(%)	E(%)	DCG(%)
PSB	HOG-ST$_R$	**71.8**	**43.3**	**53.8**	**31.3**	**68.3**
	HOG-ST	67.4	37.4	47.6	28.0	63.8
	StrT-ST	61.7	30.7	39.6	23.2	58.2
	SH	56.0	28.4	37.6	22.3	56.0
	BoF$_{SHcorr}$[8]	62.4	/	/	/	/
	HOG$_{align}$[21]	58	27	35	21	55
	LFD[23]	65.7	38.0	48.7	28.0	64.3
SHREC	HOG-ST	**90.0**	**50.6**	**62.0**	**43.4**	**80.2**
	StrT-ST	81.2	39.0	49.3	34.1	71.2
	HOG$_{align}$[21]	75	41	52	35	71

Table 2: Classification accuracy for labelling root-layers/cell-wall using different numbers of training samples

$N_{training}$	SGD$_{HOG}$(%)	SGD(%)
400	92.7/95.1	77.1/94.6
800	95.6/96.1	82.9/95.5
1600	96.4/96.5	87.2/96.0
4000	96.5/97.8	90.6/96.5

disturbance. That might explain why using the HOG feature and large smoothing can bring a significant improvement. To some other compared methods, our method also benefits from the rotation-invariance which does not depend on any pose estimation. The approaches tested on SHREC in [1] employed many state-of-the-art techniques which are specially designed for 3D shape retrieval. Our approach can challenge them and performs favourably compared to some of them. Meanwhile, our invariant description is a more general design, which can be used on grey-value volumetric data too. A shortcoming of our method is that the dimension of description could go very high when computing on multiple shells. "HOG-ST" consist of more than 150 coefficients on each shell, and coupling between different shells in "HOG-ST$_R$" leads to several times more. A feasible solution is to introduce a radial basis (*e.g.* Bessel function) into the expansion [25], which should lead to more compact description and even better performance. With an unoptimized implementation, the running time for the feature extraction is about 10 seconds for each 3D model, on a 3.2GHz CPU. The spatial smoothing alone costs about 5 seconds, where we apply Gaussian convolutions on each component of the Spherical HoG.

5.2 Voxel-wise rotation-invariant features for classification

We show a real application of our invariant description on a "segmentation" problem. The data comes from confocal microscopic imaging of Arabidopsis roots with stained cell walls (Figure 4(a)). The cell wall gives a representation of the cell outer shape, which could be used for analysing cell development or as reference structure for sub-cellular event description. Here we want a preliminary cell segmentation and an additional structural segmentation which separates cells into different layers, to support model fitting and further analysis. It's a difficult problem due to the uneven imaging quality in the data and the large variance of the cell shapes even in the same layer. We choose to do the segmentation by voxel-wise classification[18], which means embedding all neighbouring information into voxel-wise features and labelling each voxel by a trained classifier. Our rotation-invariant description is necessary for this job, as the cells in the same layer are oriented in different directions.

In this experiment, the pre-processing includes diffusion and Hessian-matrix based edge enhancement. The proposed approach starts from representing the local gradient with *SH* co-efficients up to the 4th band. To get dense descriptions, we apply *SGD* filtering on the spherical HoG field at 7 scales from $0.25\mu m$ to $16\mu m$ (the diameter of a root is around $100\mu m$), limiting $p + q \le 5$. The invariant descriptions used here only come from the magnitude (*L2-*

(a) 3D rendering of the training root

(b) Examples of computed Features

(c) Training labels

(d) Another test root(raw/enhanced)

(e) Test result

(f) Test result(cross section)
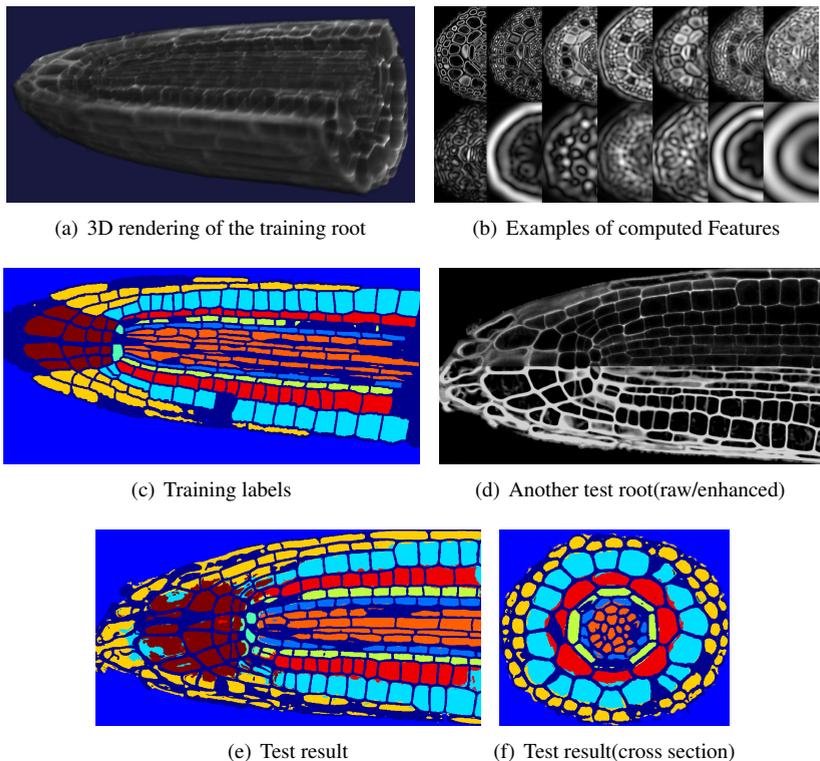
Figure 4: Segmentation on Arabidopsis root. (c) Training labels include 10 classes: background / cell wall / 6 layers and 2 regions. (e)∼(f) Classification result: some layers are not continuous because there are cell walls in the depth direction.

norm) of the tensor-valued filter output, without complex coupling. Finally, together with the local intensity value, we get 240-dimensional voxel-wise features. Some examples are shown in Figure 4(b). The features at different scales encode information from the local edge strength to the global geometrical feature. On these features, SVMs with RBF kernel [5] are trained from one root with manual labels, which is shown in Figure 4(c).[1] The classification are defined as two levels: a two-class problem (cell-wall/none-cell-wall) and a multi-class problem (different root layers in none-cell-wall regions). We test the performance of the proposed method as following: we sample 17000 points in the labelled root, and separate them into two parts by a mid-plane in the root. Then a cross-validation is carried out by using different numbers of training samples from one part and testing on another part. The performance is summarized in Table 2, with comparison to applying the *SGD* filtering on the intensity value. The clearly better performance of the HOG based approach with very few training samples demonstrates that it has better description ability, hence it should also have better generalization ability. Because of the high symmetry in one root, the ascending accuracy of *SGD* filtering on intensity possibly comes from over-fitting. When testing on other roots, the HOG-based feature still gives satisfying results (Figure 4(d∼f)). The ob-

---

[1]We created the labels by editing a watershed segmentation result manually. Some very badly segmented regions were discarded in labelling and were not used for training.

tained segmentation result will be a good start for model fitting and further analysis. Again, the most time-consuming part is the 3D convolution for Gaussian smoothing, while the other steps for feature extraction and classification is fast in comparison.

6 Conclusion

In this paper, we have presented a new approach to create rotation-invariant descriptions for 3D images, which imports the HOG idea into the spherical harmonics framework. The trick is considering histogram as a continuous signal on a sphere and representing it with spherical harmonic coefficients. On the obtained spherical tensor fields, we could easily make rotation-invariance based on the rotation behaviour of spherical tensors. We have demonstrated the generality and effectivity of our description method on 3D shape retrieval benchmarks and a biological application - structural segmentation in Arabidopsis roots. They both show very good performance. With the spherical HOG representation as a bridge, the proposed approach inherits the robustness of the HOG feature and achieves rotation invariance on the theoretical base from spherical harmonics and tensor algebra.

Acknowledgement

References

[1] C. Akgül, A. Axenopoulos, B. Bustos, M. Chaouch, P. Daras, H. Dutagaci, T. Furuya, A. Godil, S. Kreft, Z. Lian, et al. SHREC 2009-Generic Shape Retrieval Contest. In *30th international conference on EUROGRAPHICS, workshop on 3D object retrieval*, 2009.

[2] S. Allaire, J.J. Kim, S.L. Breen, D.A. Jaffray, and V. Pekar. Full orientation invariance and improved feature selectivity of 3D SIFT with application to medical image analysis. *Computer Vision and Pattern Recognition Workshop*, 0:1–8, 2008.

[3] D.M. Brink and G.R. Satchler. *Angular momentum*. Clarendon Press, 1962.

[4] T. Bülow. Spherical Diffusion for 3D Surface Smoothing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1650–1654, 2004.

[5] C.C. Chang and C.J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[6] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *CVPR*, pages 886–893, 2005.

[7] J. Fehr. Local Rotation Invariant Patch Descriptors for 3D Vector Fields. In *International Conference on Pattern Recognition*, pages 1381–1384. IEEE, 2010.

[8] J. Fehr, A. Streicher, and H. Burkhardt. A bag of features approach for 3D shape retrieval. In *Proceedings of the 5th International Symposium on Visual Computing, Part I*, volume 5875, pages 34–43, 2009.

[9] P.F. Felzenszwalb, R.B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1627–1645, 2009.

[10] G. Flitton, T.P. Breckon, N. Megherbi, and UK Cranfield. Object Recognition using 3D SIFT in Complex CT Volumes. In *BMVC*, 2010.

[11] R. Green. Spherical harmonic lighting: The gritty details. In *Game Developers Conference*, 2003.

[12] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz. Rotation invariant spherical harmonic representation of 3D shape descriptors. In *Proceedings of the Eurographics/ACM SIGGRAPH symposium on Geometry processing*, pages 156–164, 2003.

[13] A. Kläser, M. Marszałek, and C. Schmid. A spatio-temporal descriptor based on 3D-gradients. In *BMVC*, 2008.

[14] J.J. Koenderink and A.J. van Doorn. Generic neighborhood operators. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(6):597–605, July 1992.

[15] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[16] M. Reisert and H. Burkhardt. Efficient tensor voting with 3d tensorial harmonics. In *Computer Vision and Pattern Recognition Workshops*, pages 1–7, 2008.

[17] M. Reisert and H. Burkhardt. Spherical tensor calculus for local adaptive filtering. In *Tensors in Image Processing and Computer Vision*, Advances in Pattern Recognition. Springer, 2009.

[18] O. Ronneberger, J. Fehr, and H. Burkhardt. Voxel-wise gray scale invariants for simultaneous segmentation and classification. In *Proceedings of the 27th DAGM conference on Pattern recognition*, 2005.

[19] O. Ronneberger, Q. Wang, and H. Burkhardt. 3D invariants with high robustness to local deformations for automated pollen recognition. In *Proceedings of the 29th DAGM conference on Pattern recognition*, pages 455–435, 2007.

[20] M.E. Rose. *Elementary theory of angular momentum*. Dover Publications, 1995.

[21] M. Scherer, M. Walter, and T. Schreck. Histograms of oriented gradients for 3D model retrieval. In *International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, pages 41–48, 2010.

[22] M. Schlachter, M. Reisert, C. Herz, F. Schlurmann, S. Lassmann, M. Werner, H. Burkhardt, and O. Ronneberger. Harmonic Filters for 3D Multichannel Data: Rotation Invariant Detection of Mitoses in Colorectal Cancer. *IEEE Transactions on Medical Imaging*, 29(8):1485–1495, 2010.

[23] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser. The princeton shape benchmark. *International Conference on Shape Modeling and Applications*, 0:167–178, 2004.

[24] H. Skibbe, M. Reisert, O. Ronneberger, and H. Burkhardt. Increasing the dimension of creativity in rotation invariant feature design using 3D tensorial harmonics. In *Proceedings of the 31st DAGM conference on Pattern recognition*, pages 141–150, 2009.

[25] Q. Wang, O. Ronneberger, and H. Burkhardt. Rotational invariance based on fourier analysis in polar and spherical coordinates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31:1715–1722, 2009.

[26] L. Zhu, Y. Chen, A. Yuille, and W. Freeman. Latent hierarchical structural learning for object detection. In *CVPR*, 2010.