3D Non-separable Moment Invariants

Jan Flusser[1], Tomáš Suk[1(✉)], Leonid Bedratyuk[2], and Tomáš Karella[1]

[1] Czech Academy of Sciences, Institute of Information Theory and Automation, Pod Vodárenskou věží 4, 182 08 Prague 8, Czech Republic
{flusser,suk,karella}@utia.cas.cz

[2] Khmelnytsky National University, Instytuts'ka, 11, Khmelnytsky 29016, Ukraine
LeonidBedratyuk@khmnu.edu.ua

Abstract. In this paper, we introduce new 3D rotation moment invariants, which are composed of non-separable Appell moments. The Appell moments can be substituted directly into the 3D rotation invariants instead of the geometric moments without violating their invariance. We show that non-separable moments may outperform the separable ones in terms of recognition power and robustness thanks to a better distribution of their zero surfaces over the image space. We test the numerical properties and discrimination power of the proposed invariants on three real datasets – MRI images of human brain, 3D scans of statues, and confocal microscope images of worms.

Keywords: 3D recognition · 3D rotation invariants · non-separable moments · Appell polynomials

1 Introduction

Recognition of 3D objects is particularly important in bio-medical imaging, where modalities such as CT, MRI, and confocal microscopes yield full 3D volumetric data. Two main approaches to this problem are via "handcrafted" and "learned" features. While in 2D the convolutional networks and deep learned features have almost completely replaced traditional handcrafted features, the situation in 3D recognition is not so clear-cut.

For volumetric data, there are several 2D-inspired architectures operating on voxels such as convolution networks [15], residual networks [17], U-Net [10], generative models [4] and transformers [14]. However, one faces many practical problems when applying neural networks to 3D data. The data size and dimension imply the demand of large-scale annotated training sets. Such public

This work was supported by the Czech Science Foundation under the grant No. GA21-03921S and by the *Praemium Academiae*. Thanks also to the Ministry of Education and Science of Ukraine for funding it under the grant No. 0119U100662. Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA CZ LM2018140) supported by the Ministry of Education, Youth and Sports of the Czech Republic.

N. Tsapatsoulis et al. (Eds.): CAIP 2023, LNCS 14184, pp. 295–305, 2023.
https://doi.org/10.1007/978-3-031-44237-7_28

datasets do not exist, unlike for instance ImageNet, that serves as a universal training set in 2D applications. We can find only few specialized benchmarks for narrow areas like Kitty (dataset for autonomous driving) [9] and fastMRI [24] containing knee and brain MRI snaps. These training data can be used in specific areas, but do not have a potential of pre-training general backbones suitable for transfer learning. The problem of geometric invariance of the network, widely investigated in 2D [16], has been studied in a few very recent papers [20,23]. So, there is still a clear demand to develop efficient handcrafted invariant features.

Among many possible choices, *moment invariants* were proven to be very powerful descriptors of 3D bodies, because they provide invariance to the object pose and scale [8]. 3D moment invariants have been studied much less than their 2D counterparts, which means there are still many open questions concerning namely numerical stability and ability to represent objects by low-dimensional vectors. Both these issues are connected with the orthogonality of the moments (more precisely, with the orthogonality of the corresponding polynomial bases). Orthogonal moments provide generally better representation, stability and discrimination power than non-orthogonal ones. On the other hand, rotation invariants from OG moments are generally more difficult to construct than those from standard non-orthogonal moments [18,19]. Two families of popular 3D rotation moment invariants composed of OG moments are those based on Zernike moments [5] and Gaussian-Hermite moments [22].

Both these systems (and actually all other ones that have been used in object recognition so far) are *separable*, which means their basis functions can be factorized as $\pi_{pqr}(x, y, z) = P_p(x)P_q(y)P_r(z)$. Zernike moments are separable in polar domain, Gaussian-Hermite moments are separable in Cartesian domain. Separability is convenient from computational point of view but results in certain limitations of the representation ability. The distribution of zeros of separable functions is constrained such that the zero surfaces fill a rectangular or polar grid (see Fig. 1). Hence, separable basis functions provide good representation in the grid directions while the representation in "diagonal" directions may be worse. It may lead to the drop of discriminability, if characteristic object structures exhibit a diagonal-like orientation and/or if we employ only a few low-order basis functions. This has led recently to introducing *non-separable* bases, however so far in 2D only.

In 2022, Bedratyuk et al. [3] introduced 2D non-separable Appell moment invariants. In this paper, we generalize their idea into 3D.

2 Basic Idea Behind 3D Invariants

To design 3D rotation invariants form non-separable moments, we basically need to find polynomial basis functions that are *quasi-monomials*, are not separable, and there exists a stable and fast algorithm for their evaluation. Quasi-monomials are polynomials, that are transformed under coordinate rotation exactly in the same way as monomials $x^p y^q z^r$ [1]. This property is crucial for invariant design. We can simply substitute the quasi-monomial moments into well-known invariants of geometric moments (i.e. moments w.r.t. the monomial basis) [8]. The

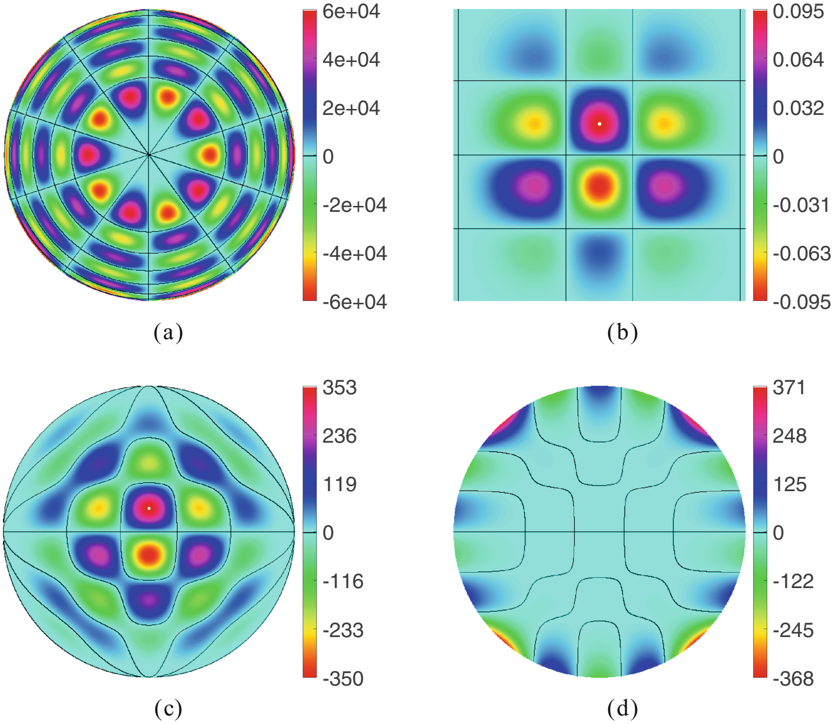


Fig. 1. Slices of 3D polynomials showing the zero distribution: (a) separable Zernike $\mathcal{R}e\left(Z_{15,9}^{5}\right)$, xy plane, (b) separable Gaussian-Hermite G_{456}, xy plane, (c) non-separable Appell U_{456}, xy plane, (d) non-separable Appell V_{456}, xy plane. The black curves are the zero sets.

problem is that quasi-monomials are rare. Among all separable polynomials, Hermite polynomials were proved to be the only quasi-monomials [21]. Among non-separable polynomials, there is no such necessary and sufficient condition. Fortunately, Bedratyuk et al. [3] proved, that Appell polynomials [7] are quasi-monomials in 2D. This key property is preserved in 3D as well. In the next section, we present 3D Appell polynomials, Appell moments and original recurrent relations for their efficient computation.

3 3D Appell Polynomials and Moments

The term *A*ppell polynomials (APs, named after P.E. Appell, a French mathematician) denotes two families of multivariate non-separable polynomials U and V. Appell polynomials are *b*i-orthogonal, which means any two polynomials, one being from U and the other one from V, are orthogonal (with a weight) on a unit sphere. The definition of Appell polynomials in 3D is the following (for more details on the APs see [7]).

$$
\begin{aligned}
&U_{m,n,o}(x,y,z) = \\
&= (m+n+o)! \sum_{i=0}^{[m/2]} \sum_{j=0}^{[n/2]} \sum_{k=0}^{[o/2]} \frac{(-1)^{i+j+k}(2i-m-1)!(2j-n-1)!(2k-o-1)!}{4^{i+j+k} i!j!k!(i+j+k)!(2i-1)!(2j-1)!(2k-1)!} \cdot \\
&\qquad \cdot x^{m-2i} y^{n-2j} z^{o-2k} (1-x^2-y^2-z^2)^{i+j+k} \\
&V_{m,n,o}(x,y,z) = \\
&= 2^{m+n+o} \sum_{i=0}^{[m/2]} \sum_{j=0}^{[n/2]} \sum_{k=0}^{[o/2]} \binom{m}{i}\binom{n}{j}\binom{o}{k} \frac{\Gamma\left(\frac{3}{2}+m+n+o-i-j-k\right)}{\Gamma\left(\frac{3}{2}\right) 4^{i+j+k} (i-1)!(j-1)!(k-1)!} \cdot \\
&\qquad \cdot (i-m-1)!(j-n-1)!(k-o-1)! x^{m-2i} y^{n-2j} z^{o-2k} .
\end{aligned}
\tag{1}
$$

The above formulas are, however, not convenient for numerical evaluation due to possible overflows. In Appendix, we present recurrent formulas for stable and fast computation.

The Appell moments M of a 3D image $f(x,y,z)$ are its projections onto the set of Appell polynomials

$$
M_{pqr}^{(P)} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P_{pqr}(x,y,z) f(x,y,z) \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z \,, \tag{2}
$$

where P stands either for U or for V. To obtain Appell invariants, these moments are substituted into geometric moment invariants [2,6,8] (this is possible because APs are quasi-monomials), so we end up with formulas such as

$$
\Phi_1 = M_{200} + M_{020} + M_{002},
$$

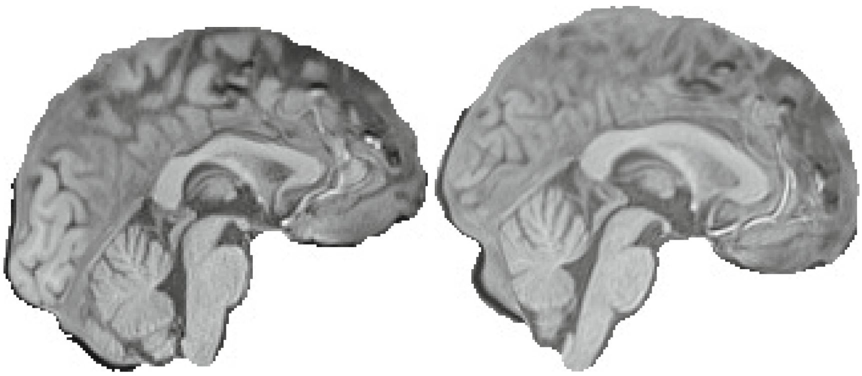
$$
\Phi_2 = M_{200}^2 + 2M_{110}^2 + 2M_{101}^2 + M_{020}^2 + 2M_{011}^2 + M_{002}^2.
$$

Using the list from [6], we obtain a complete and independent set of 213 invariants up to the 9th moment order.

4 Experiments

4.1 Human Brain MRI

The aim of the first experiment is to numerically verify the rotation invariance. We used two MRI measurements of the brain of the same patient (Fig. 2) downloaded from [11]. Their original sizes are $192\times224\times224$ and $193\times229\times193$ voxels. We generated 8 random 3D rotations of each snap with bilinear interpolation and then computed 77 rotation invariants up to the sixth order. We computed the Appell moment invariants both of U and V families by recurrence formulas (4)–(9) and compared them with the invariants from complex moments [19], geometric moments [18], Gaussian-Hermite moments [22] and Zernike moments [5].



(a) (b)

Fig. 2. Brain MRI images used in the experiment: (a) slice 96 (out of 192) of the first snap, (b) slice 97 (out of 193) of the second snap.

Table 1. ERAs of the rotation invariants in %. The averages over all invariants are used.

invariants	Appell U	Appell V	Complex	Geometric	G-H	Zernike
brain 1	1.2067	0.9720	2.6408	2.6392	3.4373	1.4609
brain 2	1.4592	1.1898	3.5169	3.5168	3.8445	1.8552
average	1.3329	1.0809	3.0788	3.0780	3.6409	1.6580

As a measure of quality we used the error relative to average (ERA)

$$ERA = \frac{100\%}{n_i} \sum_{j=1}^{n_i} \frac{\frac{1}{n_r} \sum\limits_{i=1}^{n_r} \left| I_j^i - \frac{1}{n_r} \sum\limits_{i=1}^{n_r} I_j^i \right|}{\frac{1}{n_i n_r} \sum\limits_{i=1}^{n_r} \sum\limits_{j=1}^{n_i} \left| I_j^i \right|}, \tag{3}$$

where n_i is the number of invariants ($n_i = 77$ for sixth order), $n_r = 8$ is the number of rotations, and I_j^i is jth invariant of ith rotation. ERA is similar to more common mean relative error (MRE), which is, however, unstable for invariants being close to zero. The average ERAs of all invariants are shown in Table 1. It is apparent that both Appell U and V invariants actually exhibit the rotation invariance, even with smaller error than traditional separable invariants.

4.2 The Statues

This experiment demonstrates the ability of the Appell invariants in a simple object recognition task. We scanned five visually similar small sculptures by a 3D scanner. The scanner uses 8 scanning directions to create a 3D model (see Fig. 3 (a)–(e) for the models). No texture is covering the models or inserted inside.

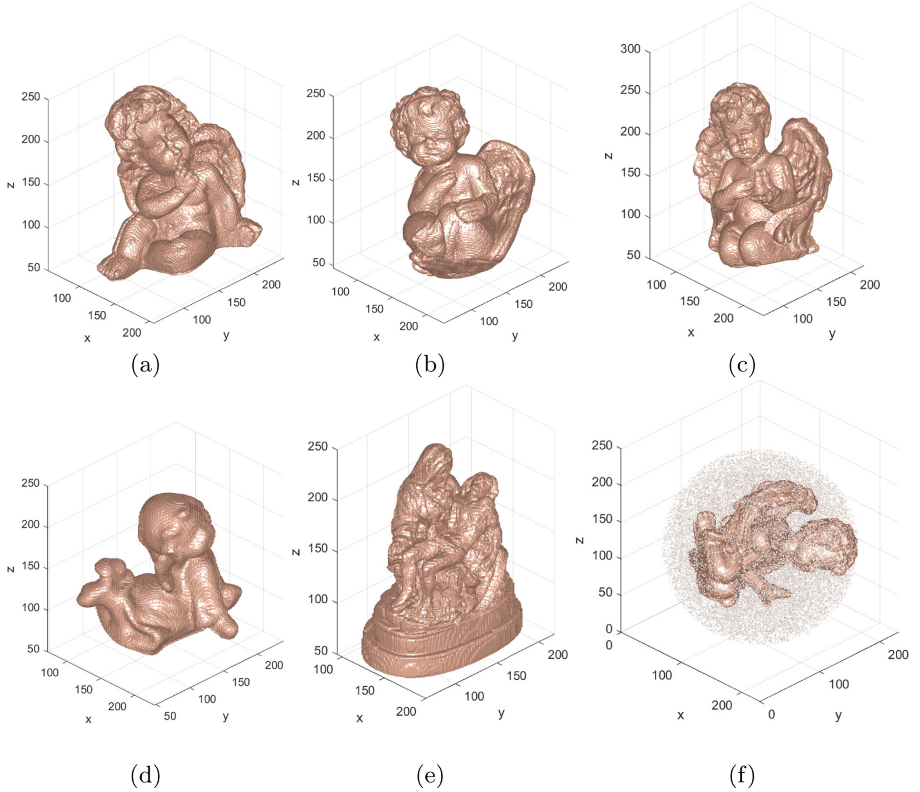


Fig. 3. From (a) to (e) the models of five statues used in the experiment, (f) rotated and noisy sample to be recognized.

The original models were used as the training samples. Eight random rotations of each statue were classified by the same invariants that were used in the MRI experiment. We applied a simple nearest-neighbor classifier in the space of invariants. If there is no noise, all methods classified all statues correctly. To make the problem more challenging, we added random noise inside the circumscribed sphere around each test sample (see Fig. 3(f) for an example), that simulates scanner errors in recovering 3D surface. Noisy objects are more difficult to recognize and performance differences of individual methods become apparent, as is documented in Table 2.

We can see that the Appell U moments are the best performing ones, the only unsatisfactory result is for low order of the moments. Looking at the other results, it is interesting that good recognition rate does not necessarily correspond with low ERA value (compare Complex and Geometric invariants).

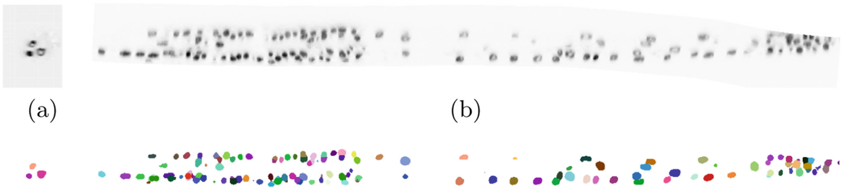
5 The Worms

In this experiment, we tested recognition via template matching. We used 3D data from confocal microscope that are publicly available [13]. The dataset was

Table 2. Success rates and relative errors of various rotation invariants in % for noisy objects. The first column shows the maximum order of the moments used.

max. order	Appell U	Appell V	Complex	Geometric	G-H	Zernike
2	60	62.2	100	60	93.3	95.6
3	100	91.1	97.8	100	100	100
4	100	100	97.8	100	100	95.6
5	100	88.9	97.8	97.8	80	95.6
6	100	93.3	97.8	100	86.7	95.6
ERA	0.246	0.303	2.675	0.324	2.744	2.506

captured by Leica microscope with 63× oil objective [12] and consists of 28 volumes of worms *Caenorhabditis elegans* at the larval stage[1] and corresponding stacks of 555 ground-truth annotated cell nuclei, see Fig. 4.



(a) (b) (c) (d)

Fig. 4. The worm used in the experiment: (a) cross-section, (b) longitudinal section, (c) ground-truth nucleus masks in the cross-section, (d) ground-truth nucleus masks in the longitudinal section.

Now we tried to detect the nuclei via template matching. Ten nuclei were chosen for training, i.e. we computed their invariants of all kinds up to the sixth order. Then we passed through the scan of the worm, computed invariants in the neighborhood of each voxel and compared them with the invariants of the training set. There is a hypothesis that the nuclei of different cells are very similar in their shape and appearance but differ from one another by orientation in 3D space, so rotation invariance of the features is required. We optimized the radius of the spherical neighborhood for each type of moments individually to get the best performance (the optimal radius depends on the shape of the basis functions, so it cannot be the same in all cases).

The voxel is considered to be the center of the nucleus if the two following conditions are satisfied:

[1] The dimension of the chosen volume is $1244 \times 140 \times 140$, the pixel size is $0.122 \times 0.116 \times 0.116$ µm.

- The feature distance must be below a user-defined threshold and must form the local minimum in the $3 \times 3 \times 3$ neighborhood of the voxel in question.
- The detected nucleus cannot overlap the nuclei detected before.

The quality of the detection was evaluated by means of the ground-truth masks. If the spatial distance between the detected nucleus and the nearest mask is less than 10 voxels, the detection is considered correct.

The results are summarized in Table 3. Again, Appell U invariants detected almost all nuclei and won the contest, followed by Complex, Geometric, and Zernike invariants.

Due to the high computation demand of a pattern matching problem, the source code was implemented in PyTorch framework allowing us to run the algorithm in parallel on Nvidia A100 GPU. Thanks to this, the task run by several orders faster than in case of traditional implementation, but still it took about two hours due to a large number of template positions to be tested. A speed up via pyramidal search and / or sparse space sampling would definitely be possible but the runtime was not the issue we were primarily interested in. Therefore, the invariant calculation in each voxel took about two hours using Nvidia A100 GPU. The source codes are available at `https://github.com/karellat/nuclei`.

Table 3. The numbers of correctly detected nuclei out of 545 instances.

Invariants	Appell U	Appell V	Complex	Geometric	G-H	Zernike
# detected nuclei	528	359	473	437	338	414
Radius [voxels]	13	11	11	13	15	17

6 Conclusion

We introduced new 3D rotation moment invariants, which are composed of non-separable Appell moments. To the best of our knowledge, this is the first application of 3D non-separable polynomials in object recognition. The design of the invariants was possible because the Appell polynomials are quasi-monomials. At this moment, we are not aware of any other non-separable quasi-monomials. Furthermore, we proposed recursive formulae for fast and stable computation.

To show the performance of the new Appell invariants in practice, we presented three experiments of different kind – invariance verification on MRI scans, object recognition of real 3D objects, and template matching in a volumetric microscopic images. In all of them, Appell invariants outperformed the competitors. This is mainly due to more even distribution of zeros of the Appell polynomials over the image space, which leads to a better representation ability of the Appell moments, especially if only low-order features are used.

Appendix

In this appendix, we present recurrent relations for fast and stable computation of 3D Appell polynomials. The polynomials $U_{m,n,o} = U_{m,n,o}(x, y, z)$ satisfy the recurrences

$$\begin{aligned}U_{m+1,n,o} &= x(2m+n+o+1)U_{m,n,o} + moxzU_{m,n,o-1} + mnxyU_{m,n-1,o}+\\ &+2mnoxyzU_{m,n-1,o-1} + m((y^2+z^2-1)m + (y^2+2z^2-1)o+\\ &+(2y^2+z^2-1)n)U_{m-1,n,o} + moz((y^2-1)(m+o-1)+\\ &+(3y^2-1)n)U_{m-1,n,o-1} + mny((3z^2-1)o+\\ &+(z^2-1)(m+n+1))U_{m-1,n-1,o} - 2mnoyz(m+n+o-2)U_{m-1,n-1,o-1}\end{aligned} \tag{4}$$

$$\begin{aligned}U_{m,n+1,o} &= y(m+2n+o+1)U_{m,n,o} + noyzU_{m,n,o-1} + mnxyU_{m-1,n,o}+\\ &+2mnoxyzU_{m-1,n,o-1} + n((x^2+z^2-1)n + (x^2+2z^2-1)o+\\ &+(2x^2+z^2-1)m)U_{m,n-1,o} + noz((x^2-1)(n+o-1)+\\ &+(3x^2-1)m)U_{m,n-1,o-1} + mnx((3z^2-1)o+\\ &+(z^2-1)(m+n-1))U_{m-1,n-1,o} - 2mnoxz(m+n+o-2)U_{m-1,n-1,o-1}\end{aligned} \tag{5}$$

$$\begin{aligned}U_{m,n,o+1} &= z(m+n+2o+1)U_{m,n,o} + moxzU_{m-1,n,o} + noyzU_{m,n-1,o}+\\ &+2mnoxyzU_{m-1,n-1,o} + o((x^2+y^2-1)o + (2x^2+y^2-1)m+\\ &+(x^2+2y^2-1)n)U_{m,n,o-1} + mox((y^2-1)(m+o-1)+\\ &+(3y^2-1)n)U_{m-1,n,o-1} + noy((x^2-1)(n+o-1)+\\ &+(3x^2-1)m)U_{m,n-1,o-1} - 2mnoxy(m+n+o-2)U_{m-1,n-1,o-1}\end{aligned} \tag{6}$$

and the polynomials $V_{m,n,o} = V_{m,n,o}(x, y, z)$ satisfy the recurrences

$$\begin{aligned}(2(m+n+o+1)+1)xV_{m,n,o} &= V_{m+1,n,o} - n(n-1)V_{m+1,n-2,o}-\\ &-o(o-1)V_{m+1,n,o-2} + m(m+2n+2o+2)V_{m-1,n,o}\end{aligned} \tag{7}$$

$$\begin{aligned}(2(m+n+o+1)+1)yV_{m,n,o} &= V_{m,n+1,o} - m(m-1)V_{m-2,n+1,o}-\\ &-o(o-1)V_{m,n+1,o-2} + n(2m+n+2o+2)V_{m,n-1,o}\end{aligned} \tag{8}$$

$$\begin{aligned}(2(m+n+o+1)+1)zV_{m,n,o} &= V_{m,n,o+1} - m(m-1)V_{m-2,n,o+1}-\\ &-n(n-1)V_{m,n-2,o+1} + o(2m+2n+o+2)V_{m,n,o-1}\end{aligned} \tag{9}$$

with the initial conditions $U_{0,0,0} = 1$, $U_{1,0,0} = x$, $U_{0,1,0} = y$, $U_{0,0,1} = z$, $U_{2,0,0} = 3x^2+y^2+z^2-1$, $U_{0,2,0} = x^2+3y^2+z^2-1$, $U_{0,0,2} = x^2+y^2+3z^2-1$, $U_{1,1,0} = 2xy$, $U_{1,0,1} = 2xz$, $U_{0,1,1} = 2yz$, $V_{0,0,0} = 1$, $V_{1,0,0} = 3x$, $V_{0,1,0} = 3y$, $V_{0,0,1} = 3z$, $V_{2,0,0} = 3(5x^2-1)$, $V_{0,2,0} = 3(5y^2-1)$, $V_{0,0,2} = 3(5z^2-1)$, $V_{1,1,0} = 15xy$, $V_{1,0,1} = 15xz$, $V_{0,1,1} = 15yz$.

References

1. Bedratyuk, L.: 2D geometric moment invariants from the point of view of the classical invariant theory. J. Math. Imaging Vision **62**, 1062–1075 (2020)
2. Bedratyuk, L.P., Bedratyuk, A.I.: 3D geometric moment invariants from the point of view of the classical invariant theory. Matematychni Studii **58**(2), 115–132 (2023). https://doi.org/10.30970%2Fms.58.2.115-132

3. Bedratyuk, L., Flusser, J., Suk, T., Kostková, J., Kautský, J.: Non-separable rotation moment invariants. Pattern Recogn. **127**, 108607 (2022)
4. Brock, A., Lim, T., Ritchie, J.M., Weston, N.: Generative and discriminative voxel modeling with convolutional neural networks. arXiv preprint arXiv:1608.04236 (2016)
5. Canterakis, N.: 3D Zernike moments and Zernike affine invariants for 3D image analysis and recognition. In: Proceedings of the 11th Scandinavian Conference on Image Analysis SCIA 1999, pp. 85–93. DSAGM (1999)
6. DIP: 3D rotation moment invariants (2011). http://zoi.utia.cas.cz/3DRotationInvariants
7. Dunkl, C.F., Xu, Y.: Orthogonal Polynomials of Several Variables, Encyclopedia of Mathematics and Its Applications, vol. 155. Cambridge University Press (2014)
8. Flusser, J., Suk, T., Zitová, B.: 2D and 3D Image Analysis by Moments. Wiley (2016)
9. Geiger, A., Lenz, P., Stiller, C., Urtasun, R.: Vision meets robotics: the KITTI dataset. In: International Journal of Robotics Research (IJRR) (2013)
10. Hatamizadeh, A., et al.: UNETR: transformers for 3D medical image segmentation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 574–584 (2022)
11. Koschutnig, K.: Openneuro (2021). https://openneuro.org/datasets/ds003813/versions/1.0.0. Accessed 21 Sept 2021
12. Long, F., Peng, H., Liu, X., Kim, S.K., Myers, E.: A 3D digital atlas of C. elegans and its application to single-cell analyses. Nat. Methods **6**(9), 667–672 (2009)
13. Long, F., et al.: 3D Nuclei instance segmentation dataset of fluorescence microscopy volumes of C. elegans (2022). https://zenodo.org/record/5942575#.YoYxYVTP0uV. Accessed 1 Feb 2022
14. Mao, J., et al.: Voxel transformer for 3D object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 3164–3173 (2021)
15. Maturana, D., Scherer, S.: VoxNet: A 3D convolutional neural network for real-time object recognition. In: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 922–928 (2015)
16. Mumuni, A., Mumuni, F.: CNN architectures for geometric transformation-invariant feature representation in computer vision: a review. SN Comput. Sci. **2**(5), 1–23 (2021)
17. Sinha, A., Unmesh, A., Huang, Q., Ramani, K.: SurfNet: generating 3D shape surfaces using deep residual networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR, pp. 6040–6049 (2017)
18. Suk, T., Flusser, J.: Tensor method for constructing 3D moment invariants. In: Real, P., Diaz-Pernil, D., Molina-Abril, H., Berciano, A., Kropatsch, W. (eds.) CAIP 2011. LNCS, vol. 6855, pp. 212–219. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23678-5_24
19. Suk, T., Flusser, J., Boldyš, J.: 3D rotation invariants by complex moments. Pattern Recogn. **48**(11), 3516–3526 (2015)
20. Wu, H., Wen, C., Li, W., Li, X., Yang, R., Wang, C.: Transformation-equivariant 3D object detection for autonomous driving (2022). https://doi.org/10.48550/arXiv.2211.11962
21. Yang, B., Flusser, J., Kautsky, J.: Rotation of 2D orthogonal polynomials. Pattern Recogn. Lett. **102**(1), 44–49 (2018)
22. Yang, B., Flusser, J., Suk, T.: 3D rotation invariants of Gaussian-Hermite moments. Pattern Recogn. Lett. **54**(1), 18–26 (2015)

23. Yu, H.X., Wu, J., Yi, L.: Rotationally equivariant 3D object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1456–1464 (2022)
24. Zbontar, J., et al.: fastMRI: an open dataset and benchmarks for accelerated MRI (2019). https://doi.org/10.48550/arXiv.1811.08839