Designing a Topological Algorithm
for 3D Activity Recognition

Maria-Jose Jimenez[1], Belen Medrano[1], David Monaghan[2],
and Noel E. O'Connor[2]

[1] Applied Math Department, School of Computer Engineering, University of Seville,
Campus Reina Mercedes, 41012 Sevilla, Spain
{majiro,belenmg}@us.es
[2] INSIGHT Centre for Data Analytics, Dublin City University, Dublin, Ireland
{david.monaghan,noel.oconnor}@dcu.ie

Abstract. Voxel carving is a non-invasive and low-cost technique that
is used for the reconstruction of a 3D volume from images captured from
a set of cameras placed around the object of interest. In this paper we
propose a method to topologically analyze a video sequence of 3D recon-
structions representing a tennis player performing different forehand and
backhand strokes with the aim of providing an approach that could be
useful in other sport activities.

Keywords: 3D video sequence · Voxel carving · Volume reconstruction ·
Persistent homology · Activity recognition

1   Introduction

The combinatorial nature of a 3D digital image is a suitable material to homology
computation by taking as input the (algebraic) cubical complex associated to the
image (whose building blocks are vertices, edges, squares and cubes). Homology
is a topological invariant that characterizes "holes" in any dimension (in the case
of a 3D space, connected components, tunnels and cavities). *Persistent homology*
[5,20] studies homology classes and their life-times (persistence) in an increasing
nested sequence of subcomplexes (a filtration on the cubical complex).

Space or voxel carving [2,4,12,18] is a technique for creating a three-dim-
ensional reconstruction of an object from a series of two-dimensional images
captured from cameras placed around the object at different viewing angles.
The technique involves capturing a series of synchronised images of an object,

and, by analysis of these images and with prior knowledge of the exact three-dimensional location of the cameras, deriving an approximation of the shape of the object.

There are numerous research papers dealing with the problem of human activity recognition from 3D data (see [1] for a recent review). An important subgroup of these works provide algorithms for activity recognition from a set of silhouettes of the subject, such as [19] or [11]. In [19] Fourier Transform in cylindrical coordinates is performed to compare *motion history volumes* representing different actions. In [11], the so-called *action volume* is produced from a set of human body silhouettes from the same view angle. They combine multiview angles to obtain a set of representative action volumes that are used to classify the action.

There have been some papers (see [13–15]) dealing with the application of persistent homology to the problem of gait recognition. Different silhouettes are extracted from a whole gait cycle (from only one viewpoint) and stacked together to form some kind of action volume to be topologically analyzed by using persistent homology.

In this paper we focus on sequences of 3D reconstructions of volumes that are captured from a small set of cameras with different viewpoints in a tennis court. From that input, we construct another 3D object containing the motion history information and that we analyze it from a persistent homology perspective.

In the following section, we describe the specific context in which we develop our work. Section 3 describes the design of our method to apply persistent homology to such specific context with the aim of recognizing the activity in a video sequence of voxel carving reconstructions. Reports on the computations performed as well as some conclusions are collected in Sect. 4. We draw some ideas for future work in the last section.

2 Voxel Carving Video Sequences

Voxel carving techniques are very useful for 3D reconstruction since they are non-invasive and they can cover a very large environment. They can be implemented with an array of low-cost cameras to produce a synchronised set of images. In each image, the subject of interest is identified and then segmented from the background of the image (this is commonly known by silhouette extraction). The subject silhouette is segmented from the background and a 3D bounding box is then drawn around the subject's approximate position in 3D space. This bounding box defines a volume that has a corresponding real world three-dimensional coordinate system. The different silhouettes are used to "carve" the defined volume accordingly. A sequence of reconstructed volumes can be seen in Fig. 1.

In the real world coordinate system the approximate subject volume is populated with voxels, that are set at a particular distance apart or spatial resolution, i.e. if the distance between voxels decreases then the spatial resolution increases. From experimental observation, authors in [16] found that a three dimensional
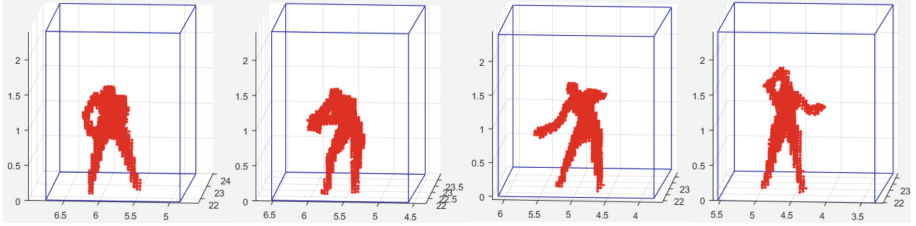
Fig. 1. A sequence of 3D reconstructions by voxel carving. Each frame is a 3D point cloud.

spatial frequency of 4 cm, i.e. 15,625 samples per cubic metre, was sufficiently adequate for their purposes and in [7] they concluded that higher resolutions did not contribute to a better topological model in the reconstruction process. That means that the spacing considered between each voxel is 4 cm in the OX, OY and OZ directions. This way, it is satisfied that the final reconstructions are qualitatively detailed enough to be used as a 3D visualisation tool, and, at the same time, based on the computational performance of a single PC, this resolution allows to run the algorithm at near to real-time. Regarding the quality of space carving results, persistent homology was proposed first in [8] as a tool for a topological analysis of the carving process along the sequence of 3D reconstructions with increasing number of cameras.

The general voxel technique proposed in [12] was modified and adapted to a specific task, as fully detailed in [16,17]. And it is, in fact, that specific voxel carving technique that we are using in this paper, fixing the number of cameras to five, since this is the usual constraint we can find in practise.

Once we get the sequence of voxel carving results, the first step is to segment the frames involving each action accomplished by the subject. This can be done by a visual inspection of the video, but an attempt to automatically recognize the beginning and end of each movement (either forehand or backhand strokes) led us to compute the variation of the mass center of each 3D frame with respect to previous and next ones with a kind of second derivative. That is, for each frame F_i, consider the mass center $(c_{i,1}, c_{i,2}, c_{i,3})$ and compute the list of values $|2c_{i,1} - c_{i-1,1} - c_{i+1,1}| + |2c_{i,2} - c_{i-1,2} - c_{i+1,2}| + |2c_{i,3} - c_{i-1,3} - c_{i+1,3}|$ whose graphic representation can be seen in Fig. 2. One can observe that peaks are mainly grouped around five points corresponding to five movements of the player (three forehand and two backhand).

3 Persistent Homology for 3D Activity Recognition

Persistent homology has been proved to be a useful tool in the study of 3D shape comparison. For example, in the paper [3] the authors provide an algorithm to approximate the matching distance (which is computationally costly) when comparing 3D shapes represented by triangle meshes.
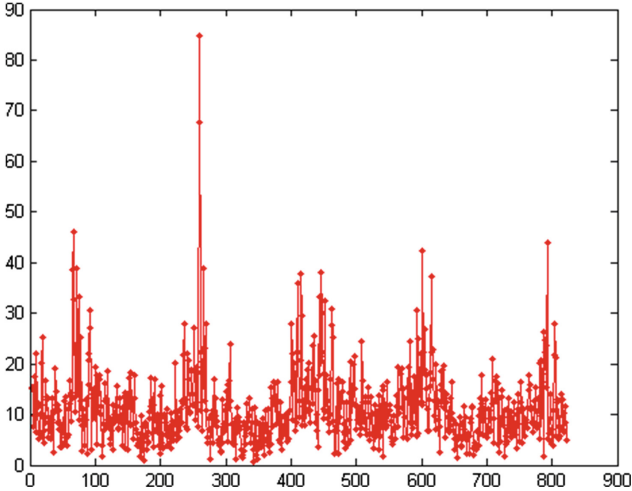
Fig. 2. Graphic representation of variation of mass center of each frame in the sequence with respect to previous and next ones.

However, as far as we know there is no work on activity recognition using persistent homology, except for the related topic of gait recognition which has already been explored from the persistent homology viewpoint in [13–15].

We are concerned with the application of persistent homology computation to provide topological analysis of a time sequence of 3D reconstructions by the voxel carving technique. We consider a sequence of voxel carving results under a fixed number of cameras, so it is convenient to have in mind that each frame is referring to a 3D reconstruction, that is, a set of 3D points in space.

The input data is a sequence $\{F_t\}_t$ of 3D binary digital images or subsets of points F_t of $\mathbb{Z}^3$ considered under the $(26, 6)$–adjacency relation for the foreground (F_t) and background $(\mathbb{Z}^3 \setminus F_t)$, respectively. Due to the nature of our input data, we focus on a special type of cell complex: *cubical complex*. A cubical complex Q in $\mathbb{R}^3$, is given by a finite collection of p-cubes such that a 0-cube is a vertex, a 1-cube is an edge, a 2-cube is a square and a 3-cube is a solid cube (or simply a cube); together with all their faces and such that the intersection between two of them is either empty or a face of each of them. The cubical complex $Q(F_t)$ associated to F_t is given by identification of each 3D point of F_t with the unit cube centered at that point and then considering all those 3-cubes together with all their faces (square faces, edges and vertices), such that shared faces are considered only once. Sometimes we will refer to p-cubes with the more general term of *cells* (corresponding to the more general concept of cell complex, see [10]).

Given a cell complex, homology groups can be computed using a variety of methods. Incremental Algorithm for computing AT-model (Algebraic Topological Model) [9], computes homology information of the cell complex by an incremental technique, considering the addition of a cell each time following a

full order on the set of cells of the complex. In [6], the authors revisited this algorithm with the aim of setting its equivalence with persistent homology computation algorithms [5, 20] working over $\mathbb{Z}/2\mathbb{Z}$ as ground ring. We make use of algorithm in [6] for the persistent homology computation, though any other algorithm for computing persistent homology, adapted to cubical complexes, could have been applied. We will use the generated *persistence barcode* as a source to create a feature vector characterising the movement. Recall that a persistence barcode encodes "times" (indexes in the ordering) of birth and death of each homology class (see [5, 20]).

The method described in this paper consists in the following steps starting from a segmented sequence of 3D frames reconstructed by voxel carving: (1) from each reconstructed volume, take the projection on a plane parallel to the net in the tennis court; (2) produce a stack with the 2D images from the previous step; (3) topologically analyze the volume by considering different directions; (4) create several topological feature vectors associated to the volume; (5) compare vectors by using a similarity measure.

Step 1. In this specific context, a particular viewpoint that can be useful for recognizing the action is a front view from the net in the tennis court. Having a 3D reconstruction obtained from different viewpoints allows to reproduce the result from a viewpoint of interest even though there is no camera in that viewing angle. For each 3D reconstructed volume, hence, we project the points onto a plane parallel to the net (see Fig. 3). If necessary, this projection could be done onto other planes of interest depending on the action to be recognized. Even more, one could combine the information obtained from different projections, that is the advantage of having a 3D reconstruction of the subject.

Step 2. Form a stack with all the 2D images from the previous step, by aligning the mass centre of every 2D projection. This way, a volume is constructed that can be considered a motion history volume since contains information of the whole movement. In this volume, we will convene that OX is the axis that is perpendicular to the net (in the tennis court), OY is parallel to the net and OZ means the hight of the points in the volume (see Fig. 4).

Step 3. Consider the cubical complex Q associated to the 3D digital image from previous step. We must consider a full ordering of its cubes $\{c^1, \ldots, c^n\}$ such that if c^i is a face of c^j, then $i < j$. Such ordering will be determined by different filter functions given by the distance to certain planes in the 3D space. Then we will have a nested sequence of subcomplexes $\emptyset = Q^0 \subseteq Q^1 \cdots \subseteq Q^m$ (a filtration over Q determined by the value of the filter function induced on the cells of the complex) for which persistent homology can be computed.

Set the minimum and maximum coordinates of the points in the considered volume, $\{x_{min}, x_{max}, y_{min}, y_{max}, z_{min}, z_{max}\}$, and consider the following "directions" to provide the filters:

– direction given by OX axis, x^+: the filter function x^+ is provided then by the distance to the plane $x = x_{min}$;
– directions given by OY axis, y^+ and y^-: the filter function y^+ (resp. y^-) is provided then by the distance to the plane $y = y_{min}$ (resp. minus the distance);
– directions given by OZ axis, z^+ and z^-: the filter function z^+ (resp. z^-) is provided then by the distance to the plane $z = z_{min}$ (resp. minus the distance);
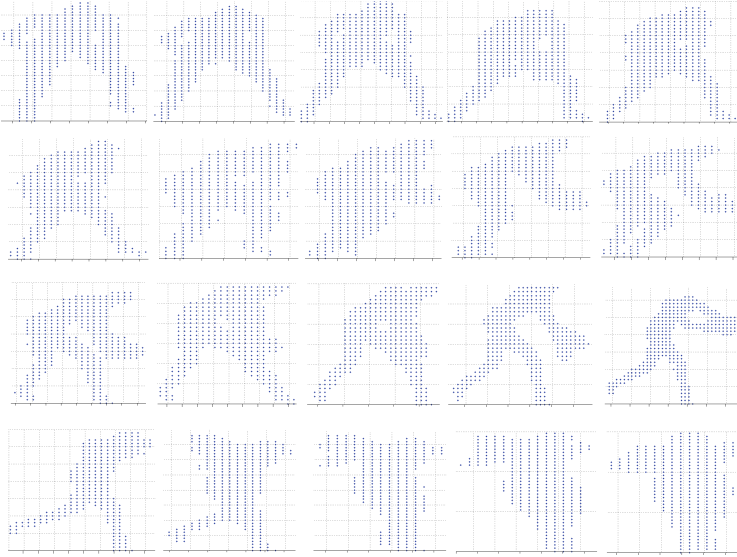

Fig. 3. Set of silhouettes obtained, from a sequence of 3D reconstructions, by projection on a plane parallel to the net in the tennis court
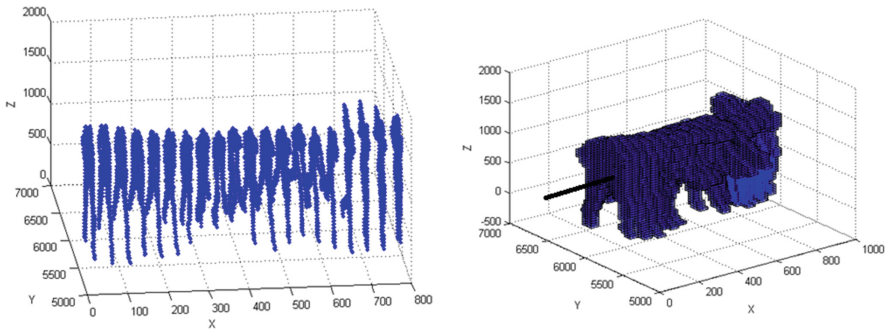

Fig. 4. Stack of silhouettes obtained, from a sequence of 3D reconstructions representing a backhand movement

- $45°$ direction on the OYZ plane, oyz^+ and oyz^-: the filter function oyz^+ (resp. oyz^-) is provided then by the distance to the plane $y+z = y_{min}+z_{min}$ (resp. minus the distance);
- $(-45)°$ direction on the OYZ plane, ozy^+ and ozy^-: the filter function ozy^+ (resp. ozy^-) is provided then by the distance to the plane $y-z = y_{max}-z_{min}$ (resp. minus the distance);

These directions are represented in Fig. 5. However, direction given by OX axis would provide poor information when applied to the whole complex, since in normal conditions, it will produce a unique connected component. That is why we propose a subdivision of the initial complex into 9 volumes (see Fig. 6) in order to compute persistent homology of each of these volumes separately along x^+ direction. This way, each silhouette is divided into a 3 by 3 array that may separate the evolution of movement of extremities from the central part of the body. More specifically, the volumes are given by $V_{ij} = \{(x,y,z), y_i \leq y \leq y_{i+1}, z_j \leq z \leq z_{j+1}\}$ for $i, j = 0, 1, 2$, with $y_0 = y_{min}$; $z_0 = z_{min}$; $y_i = y_{min} + \frac{i}{3}(y_{max}-y_{min})$ and $z_i = z_{min} + \frac{i}{3}(z_{max}-z_{min})$ for $i = 1, 2$; $y_3 = y_{max}$; $z_3 = z_{max}$.
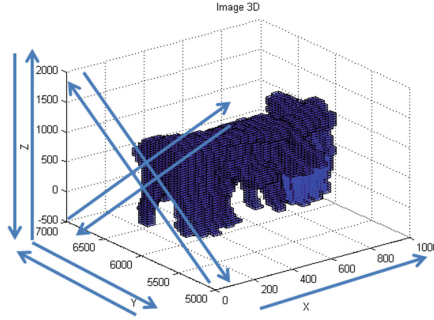

Fig. 5. Each of the 9 possible directions described to provide a filter function to order the cells in the complex.

Step 4. The filter function considered in the previous step set an ordering of all que cells in the cubical complex. Next step is to compute the persistence barcode for the cubical complex representing the motion volume. We make use of the concept of *simplified barcode* stated in [7] by which bars shorter than the distance between two consecutive subcomplexes in the considered filtration are discarded. In the case of the subdivision in the nine volumes, the computation is performed for each one of them. Hence, out of each computed barcode, a vector is produced in the style of Lamar et al. [13–15]. That is, consider the ordered set of cells in the whole volume $\{c_1, \ldots, c_N\}$ and a partition of this ordered set into n equal parts. Then, for each of the n intervals $(c_i^j, c_k^{j+1}]$, $j = 1, \ldots n$, compute

1. a_j = the number of homology classes living along the interval;
2. b_j = the number of homology classes that are born in the interval;

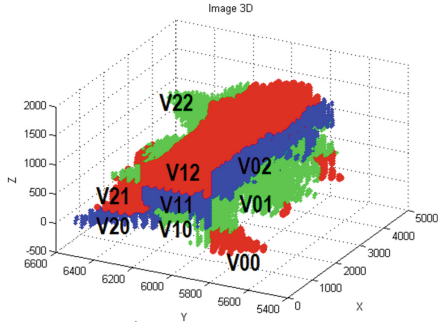and compose the vector $[a_1, b_1, a_2, b_2, \ldots, a_n, b_n]$.

Fig. 6. Color representation of the 9 volumes segmented from the motion history volume

Step 5. Finally, a similarity measure has to be considered for comparison of the feature vectors. We adopt the cosine of the angle between two vectors to measure how similar the corresponding barcodes are, that is, for two vectors V_1 and V_2 computed on the same direction, compute

$$S_{1,2} = \frac{V_1 \cdot V_2}{|V_1| \cdot |V_2|} \, .$$

Notice that each barcode produces a feature vector so the final similarity measure would be computed as the total sum of all the partial comparison measures between the corresponding vectors.

4   Experiments

We have considered 8 video sequences for forehand stroke and other 8 for backhand strokes. Such video sequences correspond to synthetic 3D reconstructions by voxel carving with coordinates on $0.4\mathbb{Z}^3$. Due to the fact that the result of voxel carving process may carry eventual numerical errors that produce some missing points, and after taking some experiments, we discarded the 1-homology classes and considered only dimension 0, that is, connected components.

By an initial evaluation on the computed barcodes (see Fig. 7, last column), we have confirmed the intuition that the direction z^- (that is, from top to bottom), is not very informative, so we have skipped it to compute the similarity measure. We have implemented the partition for $n = 5$ and $n = 10$ and realized that the latter provides much better results. This was also quite intuitive from observing Figs. 7 and 8 since $n = 5$ is too low to distinguish the numerous small bars from the few more significant bars that appear in the barcode.

We have also come up with the conclusion that the division into the 9 volumes to follow up the movement direction x^+ does not provide good results, what was also clear by watching the corresponding barcodes. The problem is that the connection of the whole object is lost and the division can be very different depending on the inclination of the subject yielding to different results. In the

first column of results of Fig. 9, the normalized similarity measure has been computed from the sum of similarity measures of each pair of vectors in directions y^+, y^-, z^+, oyz^+, oyz^-, ozy^+ and ozy^-, as well as those of volumes V_{00}, V_{01}, V_{02}, V_{10}, V_{12}, V_{20}, V_{21}, V_{22}, for $n = 5$ in direction x^+.
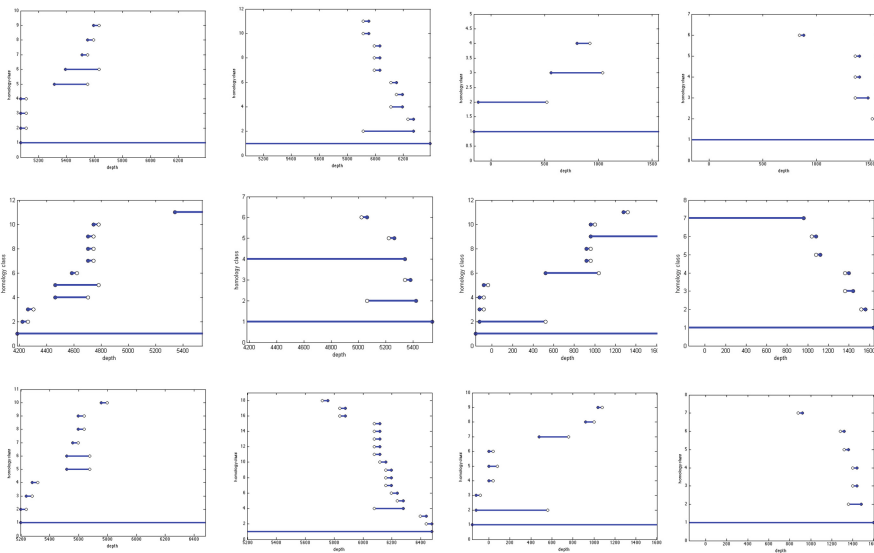

Fig. 7. Persistence 0-barcodes of three forehand movements (three rows) on each of the directions y^+, y^-, z^+ and z^-
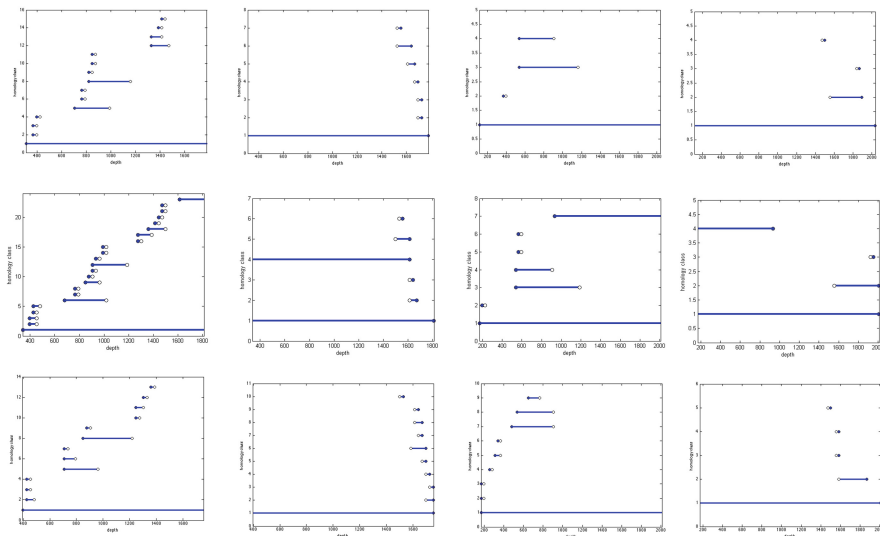

Fig. 8. Persistence 0-barcodes of three forehand movements (three rows) on each of the directions oyz^+, oyz^-, ozy^+ and ozy^-

Second and third columns of results in Fig. 9 have been computed without considering volumes V_{ij}, for partitions $n = 5$ and $n = 10$ respectively. It is clear that only for $n = 10$ does the method provide good results.

	All the filter functions for n=5	Filter functions OY+, OY-, OZ+, OYZ+, OYZ+, OZY+, OZY- for n=5	Filter functions OY+, OY-, OZ+, OYZ+, OYZ+, OZY+, OZY- for n=10
F1-F2	0,90	0,87	0,83
F2-F3	0,88	0,83	0,79
F3-B1	0,83	0,75	0,54
B1-B2	0,86	0,81	0,78
B2-B3	0,76	0,70	0,67
B1-B3	0,41	0,40	0,84
F2-B1	0,38	0,33	0,59
F3-B2	0,38	0,37	0,57
B1-B3	0,39	0,42	0,84
F1-B1	0,84	0,77	0,59
F2-B2	0,85	0,79	0,58
F3-B3	0,74	0,64	0,55
F1-B2	0,86	0,81	0,60
F2-B3	0,78	0,72	0,58
F1-B3	0,77	0,73	0,58

Fig. 9. Results of normalized similarity measures between three forehand and three backhand strokes with different partitions and filter functions.

5 Conclusions and Future Work

Fixing a certain number of cameras and considering a video sequence of 3D reconstructions (by voxel carving), we propose a method for activity recognition of a tennis player stroke based on persistent homology. This work could set the ground for extension to other activities recognition. Depending on the context, different projections could be used to form the stack of silhouettes to be analyzed and different directions of interest could be selected.

References

1. Aggarwal, J.K., Xia, L.: Human activity recognition from 3D data: a review. Pattern Recogn. Lett. **48**, 70–80 (2014)
2. Broadhurst, A., Drummond, T., Cipolla, R.: A probabilistic framework for space carving. In: Conference on Computer Vision, vol. 1, p. 388 (2001)
3. Cerri, A., Di Fabio, B., Jablonski, J., Medri, F.: Comparing shapes through multi-scale approximations of the matching distance. Comput. Vis. Image Underst. **121**, 43–56 (2014)
4. Culbertson, W.B., Malzbender, T., Slabaugh, G.G.: Generalized voxel coloring. In: Triggs, B., Zisserman, A., Szeliski, R. (eds.) ICCV-WS 1999. LNCS, vol. 1883, pp. 100–115. Springer, Heidelberg (2000)
5. Edelsbrunner, H., Letscher, D., Zomorodian, A.: Topological persistence and simplification. In: FOCS 2000, pp. 454–463. IEEE Computer Society (2000)

6. Gonzalez-Diaz, R., Ion, A., Jimenez, M.J., Poyatos, R.: Incremental-decremental algorithm for computing at-models and persistent homology. In: Real, P., Diaz-Pernil, D., Molina-Abril, H., Berciano, A., Kropatsch, W. (eds.) CAIP 2011, Part I. LNCS, vol. 6854, pp. 286–293. Springer, Heidelberg (2011)

7. Gutierrez, A., Jimenez, M.J., Monaghan, D., O'Connor, N.E.: Topological evaluation of volume reconstructions by voxel carving. Comput. Vis. Image Underst. **121**, 27–35 (2014)

8. Gutierrez, A., Monaghan, D., Jiménez, M.J., O'Connor, N.E.: Persistent homology for 3D reconstruction evaluation. In: Ferri, M., Frosini, P., Landi, C., Cerri, A., Di Fabio, B. (eds.) CTIC 2012. LNCS, vol. 7309, pp. 30–38. Springer, Heidelberg (2012)

9. Gonzalez-Diaz, R., Real, P.: On the cohomology of 3D digital images. Discrete Appl. Math. **147**(2–3), 245–263 (2005)

10. Hatcher, A.: Algebraic Topology. Cambridge University Press, Cambridge (2002)

11. Iosifidis, A., Tefas, A., Pitas, I.: Multi-view action recognition based on action volumes, fuzzy distances and cluster discriminant analysis. Sig. Process. **93**, 1445–1457 (2013)

12. Kutulakos, K.N., Seitz, S.M.: A theory of shape by space carving. Intern. J. Comput. Vision. **38**, 199–218 (2000)

13. Lamar-León, J., García-Reyes, E.B., Gonzalez-Diaz, R.: Human gait identification using persistent homology. In: Alvarez, L., Mejail, M., Gomez, L., Jacobo, J. (eds.) CIARP 2012. LNCS, vol. 7441, pp. 244–251. Springer, Heidelberg (2012)

14. Leon, J.L., Cerri, A., Reyes, E.G., Diaz, R.G.: Gait-based gender classification using persistent homology. In: Ruiz-Shulcloper, J., Sanniti di Baja, G. (eds.) CIARP 2013, Part II. LNCS, vol. 8259, pp. 366–373. Springer, Heidelberg (2013)

15. Lamar-Leon, J., Baryolo, R.A., Garcia-Reyes, E., Gonzalez-Diaz, R.: Gait-based carried object detection using persistent homology. In: Bayro-Corrochano, E., Hancock, E. (eds.) CIARP 2014. LNCS, vol. 8827, pp. 836–843. Springer, Heidelberg (2014)

16. Monaghan, D., Kelly, P., O'Connor, N.E.: Quantifying human reconstruction accuracy for voxel carving in a sporting environment. In: ACM MM, Scottsdale, AZ, 28 November–1 December 2011

17. Monaghan, D., Kelly, P., O'Connor, N.E.: Dynamic voxel carving in tennis based on player localisation using a low cost camera network. In: 2011 IEEE International Conference on Image Processing (ICIP 2011), Brussels, Belgium, 11–14 September 2011

18. Seitz, S.M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 519–528 (2006)

19. Weinland, D., Ronfard, R., Boyer, E.: Free viewpoint action recognition using motion history volumes. Comput. Vis. Image Underst. **104**, 249–257 (2006)

20. Zomorodian, A., Carlsson, G.: Computing persistent homology. Discrete Comput. Geom. **33**(2), 249–274 (2005)