Eurographics Workshop on 3D Object Retrieval (2016)
A. Ferreira, A. Giachetti, and D. Giorgi (Editors)

Position paper: Shape Retrieval and 3D Gestural Interaction

A.Giachetti[1], F.M. Caputo[1], A. Carcangiu[2], R. Scateni[3], L.D. Spano[3]

[1] Dipartimento di Informatica, Università di Verona
[2] Dipartimento di Ingegneria Elettrica ed Elettronica, Università di Cagliari
[3] Dipartimento di Matematica e Informatica, Università di Cagliari

Abstract

Despite the emerging importance of Virtual Reality and immersive interaction research, no papers on application of 3D shape retrieval to this topic have been presented in recent 3D Object Retrieval workshops.
In this paper we discuss how geometric processing and geometric shape retrieval methods could be extremely useful to implement effective natural interaction systems for 3D immersive virtual environments. In particular, we will discuss how the reduction of complex gesture recognition tasks to simple geometric retrieval ones could be useful to solve open issue in gestural interaction. Algorithms for robust point description in trajectories data with learning of inter-subject invariant features could, for example, solve relevant issues of direct manipulation algorithms, and 3D object retrieval methods could be used as well to build dictionaries and implement guidance system to maximize usability of natural gestural interfaces.

Categories and Subject Descriptors (according to ACM CCS): I.3.6 [Computer Graphics]: Interaction Techniques—

1. Introduction

One of the emerging fields where geometry-based 3D shape rertieval methods will be fundamental for the development of real-world applications is certainly Human-Computer Interaction (HCI). The availability of a variety of sensors capturing dynamic 3D shape and tracking software able to record body part trajectories (Microsoft Kinect, LeapMotion, Intel Realsense) enables the realization of several types of gesture based interaction that are particular interesting for all the emerging Virtual Reality applications exploiting cheap Head Mounted Displays (HMD).

Indeed, existing software finding and tracking body parts in depth sequences are based on shape analysis and retrieval of shapes from training sets matching labelled examples, or other kinds of classification or regression schemes. Static hand gestures are often recognized through retrieval of models given acquired color, silouhette or depth data. Hand and full body pose is successully obtained even from single depth views with approaches involving matching of acquired data with simulated depth patterns with known pose [SSK*13, KKKA13]. Software for 3D interaction based on these methods obtained a relevant success for gaming and is tested in heterogeneous applications in different fields (medicine, education, cultural heritage, etc.). Methods based on time sequence processing algorithms (e.g. Markov models,...) and multimodal input have been applied in several experiments and compared in specific contests [CWS*15, EGB*13]. Gestures are recognized usually from of time evolution of keypoint positions, orientation of shapes, etc. The effort of geometry processing community for improving the quality of gesture tracking is relevant, as shown by examples like the use of 3D descriptors like shape context to analyze 2D+t tracking patterns [GME08] or the recent advances in finger tracking methods based on Iterative Closest Points, see for example [TST*15]. Many low cost tracking systems can now provide through specialized APIs tracking sequences for body or hand keypoint in 3D and allow the development of gesture detectors. The use of pre-processed sensor makes the development of higher level interaction paradigms easier and probably more effective. In fact, as underlined in [Bow13], control problems in 3D interfaces are often hardware specific and require adaptation to the input devices and to the application environment.

Despite the good performances of Computer Vision tools for body and hand pose recognition and tracking, the quality of the user experience in many 3D gestural interaction methods is still poor and there is room for relevant improvements. Take as an example the research on gesture-based manipulation: even if a large number of methods have been proposed in the literature, most demo systems do not provide easy to use interfaces and the possibility of reaching a high level of precision. This is due to the inaccuracy in tracking, occlusions, difficulty in segmentation of different gestural primitives, and other annoying problems. However, possible improvements creating a better user experience are not necessarily related to a better tracking of body landmarks, but can be obtained through a smart global processing of 3D keypoints trajectories. Our claim is that a relevant role in the solution of open issues

DOI: 10.2312/3dor.20161079

in 3D gestural interaction can be played by the geometry processing community, as many interaction problems could be solved by mapping them on simple geometrical problems and applying strategies typically used by researchers in basic 3D shape retrieval algorithms. Several methods have been recently presented for characterizing salient points and for global shape description, with invariance properties and robustness against various kinds of perturbation. Methods with similar characteristics, but adapted to the different kind of shape data provided by tracking devices could be, in principle, applied for the solution of open issues in gestural interaction.

An example of smart application of simple geometric processing to realize effective interaction comes from 2D touchscreen interaction, where many gesture recognition applications are not based on complex time series analysis, but on the reduction of the problem to a simple template matching of 2D shapes. The popular 1-dollar recognizer [WWL07] and similar derived methods (also proposed for 3D interaction, e.g. [KR10]), decoupling and normalizing geometrical information from dynamic data and solving Nearest-Neighbor matching problems are a clear demonstration of the usefulness of this simplification. This may seem to indicate that in this case the dynamic information may be neglected without losing the meaningful part of the signal.

3D gestural interaction is clearly more challenging, but we think that, at least for some practical interaction tasks briefly described in the following sections, a smart use of geometric tools could in our opinion be used to solve open issues. Participants to the EG Workshop on 3D Object Retrieval could be extremely interested to be involved in the related research.

2. Open issues in mid-air virtual manipulation

In order to clarify our ideas, we first present three open issues in a particular 3D gestural interaction type we are currently investigating, e.g. 3D manipulation. Our goal is to propose and test novel gestural paradigms based on hands and fingers tracking, possibly robust against inaccuracy of the tracker performances. Many problems related to an effective interaction of this kind, involving selection of objects, grabbing, translation, rotation and possibly scaling, could be, in our opinion, solved with the tools developed in the geometry processing domain and we plan to investigate this approach in the future. In the following we briefly discuss some of these problems, like detection of gesture start and end, accuracy in manipulation, smart dimensionality reduction allowing a more effective interaction.

2.1. Detection of gesture start and end

Gestural interaction is particular successful on 2D touchscreens where it is easy to determine the beginning and the end of the gesture by using the contact between fingers and display surface. In 3D deviceless interaction, the beginning of a gesture must be automatically detected from the gesture itself. The use of pattern recognition tools may allow a robust recognition of a gesture learned after the gesture realization, that may be ok for a sign language interpreter, but not for a manipulation tool that should give visual feedback within a reasonable time. A possible trick to solve this issue typically applied in interface is to use a second hand or a vocal interface, or the recognition of a coded gesture to tell the system that the gesture is starting or it is finished. Furthermore, in a manipulation interaction gesture start involves also the "grabbing" of the object of interest and gesture end involves also its release. This means that the algorithm should localize with a sufficient accuracy the position of the grab and, more difficult, the desired location of the object release. Especially the last task requires, in our opinion, both a smart geometrical representation of the hand/finger trajectories, possibly invariant to users' gestures realization and a smart learning procedure in order to characterize the key actions and the corresponding desired object position.

It is a really challenging task, but the ideas that could be applied to find a reasonable solution may be the same applied in classical robust point matching and landmark location. Keypoint trajectories could be simplified and normalized, mappings between trajectories could be encoded as functions, evolution of connected point could be treated as a surface. And, in the same way learning approaches are used to find discriminative keypoints for specialized recognition tasks on 3D meshes [CPA13], it is possible to think that on geometric encoding of hand trajectories, keypoints able to have a user-independent recognition of gesture limits could be learned through the collection of example data. The collection of users' tracking data obtained in specifically designed tests to learn gestural features is often applied in HCI research, for example in gesture elicitation experiments [NDL*09, AWB*12], and, with the same approach we could learn how users behave when doing "naturally" simple gestures like grabbing, translating and rotating a virtual object. Registering example gestures and decoupling 3D trajectories and velocity patterns it would be possible to find specific and invariant keypoints to identify beginning and end of gestures.

2.2. Localization of gestures

Another big problem is related to the accuracy of gestures, that in manipulation tasks is particularly important. Also in this case geometry processing and learning can surely be used to increase the precision of the mapping between the real gesture and the virtual world.

The accuracy of object positioning in manipulation is limited by the lack in accuracy of tracking and the possible occlusions of keypoints. However, this problem could be mitigated by the redundancy of the data, and all the research on robust descriptors or partial retrieval can surely help in finding ad hoc solutions for the task. Approaches for partial shape retrieval, like Bag of Words [Lav12, WFB*14] could be, for example, applied to select only the partial information correctly describing the gesture and captured by the device/tracking library used. Furthermore, learning from example, using regression techniques, the relationships between keypoint positions and desired manipulation position could help in improving the accuracy of the gesture localization. Another problem in this particular case is the delay in visual feedback, as the detection of the release gesture requires a backwards analysis and if the grabbed object is moved together with the hand, there is a discrepancy between the expected manipulated object position and the visualized position in the virtual representation.

2.3. Smart dimensionality reduction

As pointed out in [Bow13], a weakness of mid-air 3D interaction is related to the lack of constraints when a lower dimensional gesture is actually performed. Think about a rotation: if want to rotate an object "naturally", we would actually performing a 1D gesture (or a 2D gesture on a plane), ideally assuming that the rotation axis is fixed. But this would require another gesture to be used to identify rotation axis or using bimanual gestures that are not very "natural". A smart solution in this case could be an automatic detection of the rotation axis from tracking data and a subsequent smart constraint of the gesture space. Geometric reasoning could help also in this case to find an optimal solution.

3. Interface design

Besides the challenges that gesture tracking devices pose in supporting manipulation interactions in a robust way, 3D shape retrieval may support designers in selecting gestures for triggering commands and actions in applications. In contrast to their 2D counterpart, 3D gestures do not have a well-established vocabulary for supporting similar interactions across different applications, such as e.g. the swipe or the pinch for zoom on touchscreens. Different taxonomies for interactive gestures have been defined in the literature (see for example [Kmcs05]), but most of them describe how gestures are performed, without a real consensus on the interaction semantics. Considering the technical difficulties in building robust gesture recognizers, designers often strive in defining gesture vocabularies. On the one hand, gestures should be selected in order to support the user's task through a "natural" movement, i.e. replicating the interaction with real world objects. On the other hand, gestures should be recognizable using the target apparatus, and they should be also clearly distinguishable from each other. This trade-off limits the usability of natural user interfaces (NUI) [Nor10].

3.1. Vocabulary definition

In many applications, designers define gestures that require the user to move one or more limbs on a given trajectory. In this case, representing gestures as 3D shapes provides advantages during the vocabulary definition process. Designers may visually understand, through the help of geometry processing tools, how similar two or more gestures are. In addition, shape retrieval may be used for supporting them in searching for a distinguishable shape, similar to one proposed by the designer, in order to minimize the trade-off between intuitiveness and separation. Finally, shape retrieval would be good also for promoting a shared set of gestures for similar functionalities in different applications.

How to effectively represent gestures as shapes, considering intra and inter user variability is still an open research question. If we limit the scope to 2D strokes, Leiva et al. [LMAP15] recently considered the Sigma Lognormal model for complex handwritten trajectories [PAYL93], which represent gestures as series of primitives (circular arcs) connecting a sequence of target points. The function describing the velocity of each primitive contains a set of parameters modelling the variability of their neuromuscular execution. The temporal overlap of the primitives produces the movement trajectory. Through such model, it was possible to accurately generate instances for training a gesture classifier, starting from a single example. An equivalent representation for 3D gesture shapes, which includes kinematic models for generating variations from a single or a few samples would lower the cost of producing robust classifiers. It would help also in retrieving gesture definitions by shape, useful while selecting the gesture vocabulary, and it would enable more robust gestural queries for retrieving 3D shapes.

3.2. Guidance

In NUIs, gestures should be self-revealing in order to be easily discovered by users while interacting with the application. Unfortunately, technical difficulties force designer in selecting less intuitive gestures and, on the contrary, most of the times the interaction is designed maximizing the recognition rate. Therefore, a guidance system for discovering which commands are available and how to trigger them is very helpful for users. Delamare et al. [DCN15] provide a categorization of the different aspects that may be covered by such guidance systems. Since gestures have a perceivable duration in time if compared to e.g. mouse clicks or key typing, the application should react to user's movement while performing the gesture, and not only at the end. In particular, it is important to represent the completed gesture portion, but also the expected completions of the movement, which may be more than one. In HCI jargon, the former representation is called the *feedback*, while the latter is called *feedforward* [VLvdHC13].

An interesting point we would like to discuss here is the similarity between this guidance process and the 3D shape retrieval: during the gesture performance the system interprets a partial input for selecting a subset of candidates for the final recognition. The partial input may be considered as a query, while the candidate gestures are the retrieved shapes. We see such parallelism also in the visual guidance representation. For instance, Octopocus [BM08] and the 3D pipe guidance [DCN15], both represent the gesture trajectories as lines that change their visual properties similarly to the autocompletion functionality in web search engines: the partial trajectory (the typed characters) is depicted with a solid line (a darker colour), while the list of possible completion traces (queries) is represented with dashed or thinner lines (a lighter text colour). Such guidance systems already focus on simple geometric properties of the trajectory for matching it with the available gestures in a naive way. More advanced geometry processing and shape retrieval techniques would provide a powerful tool for designing and implementing gesture interfaces. On the one hand, they would hopefully provide a robust technique for matching partial and complete gestures trajectories. On the other hand, their representation would provide a powerful tool for accessing gesture segments, allowing the definition of user interface reactions to partial recognition. The latter is a key feature for gesture representation, since many usability flaws in gestural interfaces are due to a lack of feedback or feedforward, with the consequent impossibility for the user to understand which actions are supported.

4. Discussion

We believe that the smart use of geometry processing, and, in particular, geometry-based shape retrieval techniques will be fundamental in order to develop 3D natural interaction system, so that

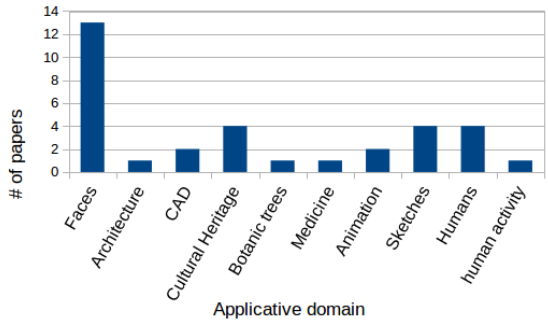

Figure 1: *Topics of applicative domain papers presented in 3DOR workshops.*

this is definitely a topic of interest for the 3D Object Retrieval Workshop community. However, if we consider the programs of the the 3DOR workshops (including Eurographics SHape REtrieval Contest (SHREC) tracks papers), even if many applicative papers have een presented on various topics (see figure 1, no papers on gesture-related 3D retrieval have been presented. One reason for this can be the limited interaction between the communities of HCI and geometry processing, another one the lack of test datasets that can be used by researchers not directly involved in interaction research to apply their methods on this specific application. We wrote this paper specifically to invite HCI and geometry processing community to increase their collaboration in order to solve the many challenging problems related to natural interaction in immersive virtual reality. We just discussed a few research directions of our interest to demonstrate our point, but, of course, there are many other relevant applications of geometry processing to gesture analysis, for example related to 4D characterization and joint analysis of gestures and shapes. For the proposed research directions, we plan to acquire and distribute labelled datasets to stimulate further work in the domain and, possibly, to organize future SHREC contests based on them.

References

[AWB*12] AIGNER R., WIGDOR D., BENKO H., HALLER M., LINDBAUER D., ION A., ZHAO S., KOH J.: Understanding mid-air hand gestures: A study of human preferences in usage of gesture types for hci. *Microsoft Research TechReport MSR-TR-2012-111* (2012). 2

[BM08] BAU O., MACKAY W. E.: Octopocus: A dynamic guide for learning gesture-based command sets. In *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, 2008), UIST '08, ACM, pp. 37–46. URL: http://doi.acm.org/10.1145/1449715.1449724, doi:10.1145/1449715.1449724. 3

[Bow13] BOWMAN D. A.: *3D User Interfaces*, 2nd ed. The Interaction Design Foundation, Aarhus, Denmark, 2013. URL: http://www.interaction-design.org/encyclopedia/3d_user_interfaces.html. 1, 3

[CPA13] CREUSOT C., PEARS N., AUSTIN J.: A machine-learning approach to keypoint detection and landmarking on 3d meshes. *International Journal of Computer Vision 102*, 1-3 (2013), 146–179. 2

[CWS*15] CHENG G., WAN Y., SAUDAGAR A. N., NAMUDURI K., BUCKLES B. P.: Advances in human action recognition: A survey. *arXiv preprint arXiv:1501.05964* (2015). 1

[DCN15] DELAMARE W., COUTRIX C., NIGAY L.: Designing guiding systems for gesture-based interaction. In *Proceedings of the 7th ACM SIGCHI Symposium on Engineering Interactive Computing Systems* (New York, NY, USA, 2015), EICS '15, ACM, pp. 44–53. URL: http://doi.acm.org/10.1145/2774225.2774847, doi:10.1145/2774225.2774847. 3

[EGB*13] ESCALERA S., GONZÀLEZ J., BARÓ X., REYES M., LOPES O., GUYON I., ATHITSOS V., ESCALANTE H.: Multi-modal gesture recognition challenge 2013: Dataset and results. In *Proceedings of the 15th ACM on International conference on multimodal interaction* (2013), ACM, pp. 445–452. 1

[GME08] GRUNDMANN M., MEIER F., ESSA I.: 3d shape context and distance transform for action recognition. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on* (2008), IEEE, pp. 1–4. 1

[KKKA13] KESKIN C., KIRAÇ F., KARA Y. E., AKARUN L.: Real time hand pose estimation using depth sensors. In *Consumer Depth Cameras for Computer Vision*. Springer, 2013, pp. 119–137. 1

[Kmcs05] KARAM M., M. C. SCHRAEFEL: *A Taxonomy of Gestures in Human Computer Interactions*. Technical report, University of Southampton, 2005. 3

[KR10] KRATZ S., ROHS M.: A $3 gesture recognizer: simple gesture recognition for devices equipped with 3d acceleration sensors. In *Proceedings of the 15th international conference on Intelligent user interfaces* (2010), ACM, pp. 341–344. 2

[Lav12] LAVOUÉ G.: Combination of bag-of-words descriptors for robust partial shape retrieval. *The Visual Computer 28*, 9 (2012), 931–942. 2

[LMAP15] LEIVA L. A., MARTÍN-ALBO D., PLAMONDON R.: Gestures À go go: Authoring synthetic human-like stroke gestures using the kinematic theory of rapid movements. *ACM Trans. Intell. Syst. Technol. 7*, 2 (Nov. 2015), 15:1–15:29. URL: http://doi.acm.org/10.1145/2799648, doi:10.1145/2799648. 3

[NDL*09] NORTH C., DWYER T., LEE B., FISHER D., ISENBERG P., ROBERTSON G., INKPEN K.: Understanding multi-touch manipulation for surface computing. In *Human-Computer Interaction–INTERACT 2009*. Springer, 2009, pp. 236–249. 2

[Nor10] NORMAN D. A.: Natural user interfaces are not natural. *interactions 17*, 3 (May 2010), 6–10. URL: http://doi.acm.org/10.1145/1744161.1744163, doi:10.1145/1744161.1744163. 3

[PAYL93] PLAMONDON R., ALIMI A. M., YERGEAU P., LECLERC F.: Modelling velocity profiles of rapid movements: a comparative study. *Biological cybernetics 69*, 2 (1993), 119–128. 3

[SSK*13] SHOTTON J., SHARP T., KIPMAN A., FITZGIBBON A., FINOCCHIO M., BLAKE A., COOK M., MOORE R.: Real-time human pose recognition in parts from single depth images. *Communications of the ACM 56*, 1 (2013), 116–124. 1

[TST*15] TAGLIASACCHI A., SCHRÖDER M., TKACH A., BOUAZIZ S., BOTSCH M., PAULY M.: Robust articulated-icp for real-time hand tracking. In *Computer Graphics Forum* (2015), vol. 34, Wiley Online Library, pp. 101–114. 1

[VLvdHC13] VERMEULEN J., LUYTEN K., VAN DEN HOVEN E., CONINX K.: Crossing the bridge over norman's gulf of execution: Revealing feedforward's true identity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2013), CHI '13, ACM, pp. 1931–1940. doi:10.1145/2470654.2466255. 3

[WFB*14] WANG X., FENG B., BAI X., LIU W., LATECKI L. J.: Bag of contour fragments for robust shape classification. *Pattern Recognition 47*, 6 (2014), 2116–2125. 2

[WWL07] WOBBROCK J. O., WILSON A. D., LI Y.: Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes. In *Proceedings of the 20th annual ACM symposium on User interface software and technology* (2007), ACM, pp. 159–168. 2