The Information is in the Maps: Representations & Algorithms for Mapping among Geometric Data

Leonidas Guibas
LELAND STANFORD JUNIOR UNIV CA

09/30/2015
Final Report

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to the Department of Defense, Executive Service Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
30-11-2015	Final Performance Report	09/01/2012-08/31/2015

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
The Information is in the Maps: Representations and Algorithms for Mappings among Geometric Data	NA
	5b. GRANT NUMBER
	FA9550-12-1-0372
	5c. PROGRAM ELEMENT NUMBER
	NA

6. AUTHOR(S)	5d. PROJECT NUMBER
Guibas, Leonidas, J.	NA
	5e. TASK NUMBER
	NA
	5f. WORK UNIT NUMBER
	NA

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
LELAND STANFORD JUNIOR UNIVERSITY STANFORD UNIVERSITY 450 SERRA MALL STANFORD, CA 94305-2004	NA

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
AF OFFICE OF SCIENTIFIC RESEARCH 875 NORTH RANDOLPH STREET, RM 3112 ARLINGTON VA 22203	AFOSR
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)
	NA

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION A

13. SUPPLEMENTARY NOTES
NA

14. ABSTRACT
The project has developed tools for understanding informative, structure-preserving and composition-revealing maps between multiple geometric data sets in 2D, 3D, and higher dimensions. The discovery of such informative maps is deeply interesting, because our comprehension of a geometric data set is often based on understanding the relationships among its parts, or its connections to other data sets that may share the same or similar structure. Towards this end, we have studied a variety of neighborhood descriptors in geometric data sets and their properties. Such descriptors help establish useful correspondences and seed the process of finding good maps. We have also investigated networks of maps and the consistency of information transport in such networks. The space of all maps is a huge space and an important part of the project has addressed the problem of finding compact representations and encodings for the much smaller spaces of interesting maps within a specific application. The machinery developed here has proven of use across a broad spectrum of disciplines that use geometric data, has helped generate novel mathematics and algorithms, and has made connections with cognitive science concepts and ideas.

15. SUBJECT TERMS

Final Report.

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
					19b. TELEPHONE NUMBER (Include area code)

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18
Adobe Professional 7.0

Reset

INSTRUCTIONS FOR COMPLETING SF 298

1. REPORT DATE. Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

2. REPORT TYPE. State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

3. DATES COVERED. Indicate the time during which the work was performed and the report was written, e.g., Jun 1997 - Jun 1998; 1-10 Jun 1996; May - Nov 1998; Nov 1998.

4. TITLE. Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

5a. CONTRACT NUMBER. Enter all contract numbers as they appear in the report, e.g. F33615-86-C-5169.

5b. GRANT NUMBER. Enter all grant numbers as they appear in the report, e.g. AFOSR-82-1234.

5c. PROGRAM ELEMENT NUMBER. Enter all program element numbers as they appear in the report, e.g. 61101A.

5d. PROJECT NUMBER. Enter all project numbers as they appear in the report, e.g. 1F665702D1257; ILIR.

5e. TASK NUMBER. Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

5f. WORK UNIT NUMBER. Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

6. AUTHOR(S). Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES). Self-explanatory.

8. PERFORMING ORGANIZATION REPORT NUMBER. Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES). Enter the name and address of the organization(s) financially responsible for and monitoring the work.

10. SPONSOR/MONITOR'S ACRONYM(S). Enter, if available, e.g. BRL, ARDEC, NADC.

11. SPONSOR/MONITOR'S REPORT NUMBER(S). Enter report number as assigned by the sponsoring/ monitoring agency, if available, e.g. BRL-TR-829; -215.

12. DISTRIBUTION/AVAILABILITY STATEMENT. Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

13. SUPPLEMENTARY NOTES. Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

14. ABSTRACT. A brief (approximately 200 words) factual summary of the most significant information.

15. SUBJECT TERMS. Key words or phrases identifying major concepts in the report.

16. SECURITY CLASSIFICATION. Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

17. LIMITATION OF ABSTRACT. This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.

Final Progress Report, 09/01/2014 to 08/31/2015 for AFOSR FA9550-12-1-0372, "The Information Is In the Maps: Representations & Algorithms for Mappings among Geometric Data"

Final Executive Summary

The goal of this effort has been to develop a set of mathematical and computational tools for describing, analyzing, computing and exploiting relationships and mappings between geometric data sets, both pairwise and in higher-order combinations, or in loose networks of interrelated sets. The objective is to analyze geometric data sets jointly, organizing data collections into (possibly overlapping) groups of related sets or parts thereof, separating what is common from what is variable within each group and across groups, and understanding the main axes of variability.

The basic thesis of the work has been that geometric data sets are best understood not in isolation but within a ``social network'' of related data sets and their associated maps and correspondences. These relational networks can interconnect data sets into societies where the "wisdom of the collection" can be exploited in performing operations on individual data sets better, or in further assessing relationships between them. By creating such societies of data sets and their associations in a globally consistent way, we enable a certain joint understanding that provides the powers of abstraction, analogy, compression, error correction, and summarization over the data.

For example, given a collection of images with shared content across a number of object categories (e.g., airplanes), our network analysis techniques are able to learn the shared categories and discover the object(s) in each category contained in each image. Furthermore, this is accomplished in a fully unsupervised manner and the results surpass some state of the art methods that use supervision. Of course supervision can be added to further improve the results.

Of particular interest this past year has been algorithms for relating and interconnecting diverse modalities that provide information about objects in the world, including images, sketches, 3D scans, 3D models, and language. Different modalities often capture distinct types of information about the nature and state of objects in the world, so that the information fusion made possible by this integration creates new integrated knowledge not available separately from any of the modalities.

Unlike traditional data fusion, in our setting fusion is possible not only at the level of object instances but also across object categories, through the abstraction mechanisms we have developed in our networks. In particular, we have aimed to provide additional information or knowledge about captured signals (e.g., images), in a real time setting -- information that is

NOT present in the raw signal but is inferred from contextual knowledge present in the network. For example, when we see a chair partially occluded by a table, we can usually make a pretty good guess about what the occluded portions looks like, because we may see other identical chairs in the same environment, or because we have memory of having seen other similar chairs in analogous settings in the past. We have aimed to endow computers with exactly this ability to "imagine the unseen" and have made substantial progress on this front.

Specific 3rd Year Accomplishments/Findings

I. Common Embedding Spaces for Multimodal Data, Combining Visual Representations and Language

We have developed a new method for structuring multi-modal representations of shapes according to semantic relations. We learn a metric that links semantically similar objects represented in different modalities. First, 3D-shapes are associated with textual labels by learning how textual attributes are related to the observed geometry. Correlations between similar labels are captured by simultaneously embedding labels and shape descriptors into a common latent space in which an inner product corresponds to similarity. The mapping is learned robustly, by optimizing a rank-based loss function under a sparseness prior for the spectrum of the matrix of all classifiers. Second, we extend this framework to relating multi-modal representations of the geometric objects.

The key idea is that weak cues from shared human labels are sufficient to obtain a consistent embedding of related objects even though their representations are not directly comparable. Technically, we accomplish the assignment of labels to 3D geometry by learning a low-rank classifier matrix that recognizes similarities of labels through correlations in shape. This permits information sharing across geometrically similar objects as well as semantically related labels. In experiments, we can clearly see an advantage in performance over baseline methods that ignore this side information. Moreover, we have generalized the idea of multi-label classification through a low-dimensional latent space to obtain a novel cross-modal embedding of objects. This can be used for object retrieval across different modalities and for interactive explorations of complex data spaces.

We have evaluated our method against common base-line approaches, investigated the influence of different geometric descriptors, and demonstrated a prototypical multi-modal browser that relates 3D-objects with text, photographs, and 2D line sketches.

This work has appeared at the 2015 Eurographics Symposium on Geometry Processing.

II. The ShapeNet Repository

In order to develop and test our shape mapping algorithms at scale, we have initiated an effort to collect and annotate a large corpus of 3D CAD models that we call ShapeNet (http://shapenet.cs.stanford.edu). This is a joint effort with Professors Pat Hanrahan and Silvio

Savarese at Stanford, and Tom Funkhouser and Jianxiong Xiao at Princeton. The repository contains 3D models from a multitude of semantic categories and organizes them under the WordNet taxonomy. In addition to categories, ShapeNet currently provides consistent rigid alignments and bilateral symmetry planes for each 3D model. These annotations, as well as other planned semantic annotations, are made available through a public web-based interface to enable data visualization of object attributes, promote data-driven geometric analysis, and provide a large-scale quantitative benchmark for research in computer graphics and vision. Planned annotations include object parts and part names, local as well as global symmetries, physical properties such as size and weight, and affordances / functionality (how the shape is used). Maps and correspondences between shapes will be included, as well as between shapes and images.

ShapeNet aims to fill a large gap in the 3D repositories that currently exist, which are either large (e.g., the Trimble 3D Warehouse, 2.5M shapes) but poorly annotated, or annotated but small (e.g, the Princeton Shape Benchmark, 1.8K shapes). Recently large data sets of images, such as Imagenet (Deng et al. 2009, 14M images organized into 20K categories associated with "synsets" of WordNet) have played a major role in key data-driven advances in computer vision, in part by providing rich training data for machine learning algorithms. The same has occurred with natural language processing (NLP – e.g., in machine translation) and our belief is that similar breakthroughs can happen with 3D data.

This is a seed effort intend to lead to a separately supported project.

III. Shape Completion for Incomplete 3D Scans

Acquiring 3D geometry of an object is a tedious and time-consuming task, typically requiring scanning the surface from multiple viewpoints. In work this past year we focused on reconstructing complete geometry from a single scan acquired with a low-quality consumer-level scanning device, even in the presence of significant occlusions (and of course self-occlusions). Our method is class-based and uses a network of example 3D shapes to build structural part-based priors that are necessary to complete shapes in that class. In our representation, we associate a local coordinate system to each part and then learn the distribution of positions and orientations of all the other parts from the network, which implicitly also defines the positions of symmetry planes and symmetry axes. At the inference stage, this knowledge can be transported to the new scan and used to analyze incomplete point clouds with substantial occlusions, because observing only a few regions is still sufficient to infer the global structure. Once the parts and the symmetries are estimated, both data sources, symmetry and database, are fused to complete the point cloud, providing much better results than either of them alone could.

Our main technical contribution is a data-driven technique for estimating shape structure from incomplete point clouds. The key difference from previous approaches is that our method does not rely on a global coordinate system — instead every part defines local coordinates, and then all parts are jointly optimized to find their most plausible arrangement. This enables the

prediction of parts in occluded regions, and the estimation of symmetries even if the input partial scan is asymmetric due to occlusions.

We have evaluated our technique on a synthetic dataset containing 481 shapes, and on real scans acquired with a Kinect scanner. Our method demonstrates high accuracy for the estimated part structure and detected symmetries, enabling higher quality shape completions in comparison to all alternative techniques. Furthermore, we have publically released our benchmark data set so that others working in this area can evaluate their methods and compare them to ours.

This work, an instance of "imagining the unseen," will appear at Siggraph Asia 2015.

IV. 3D-Assisted Feature Synthesis for Novel Views

We are especially interested in being able to link images that represent views of the same object from very different viewpoints. Comparing different views has been a long-standing challenging problem in computer vision, as visual features are not stable under large view point changes. In work this past year, given a single input image of an object, we have developed tools able to synthesize its features for other views, leveraging an existing, modestly-sized 3D model collection of related but not identical objects. 3D models can provide strong prior information to help an algorithm "imagine" what the underlying 3D object should look like from novel views. To accomplish this feature transport to new views, we study the relationship of image patches between different views of the same 3D object, seeking what we call "surrogate patches" — patches in one view whose feature content predicts well the features of a patch in another view. These surrogate relationships are learned from the analysis of a co-aligned set of 3D models in a given class. Note that, indirectly, these surrogate patch relationships encode geometric global or local symmetries of the underlying 3D models, without having to first estimate the latter.

When an image of an object is provide, we first estimate its pose and then develop local linear models for predicting its features using features from the same views of our 3D models. We finally use our surrogate relationships for transferring the same linear combination to estimate features for the new view. Based upon these surrogate relationships, we can in fact create feature sets for all views of the latent object on a per patch basis, providing us an augmented view-independent representation of the object. We note that our method can work with many common image feature sets, including HoG, CNN, etc.

The method allows us to compare images of objects from very different views. In addition to demonstrating that we do much better in such cross view comparisons than traditional image-based methods, we have explored a number of other applications of our techniques. These include (a) part-based image retrieval, where we query for similar images within a specified region, aiming again at view independent results, (b) fine-grained image retrieval and object categorization, and (c) instance retrieval (looking for exactly the same object in other images).

This work, a second instance of "imaging the unseen," will appear at ICCV 2015 (oral).

V. Analyzing and Using the Shape of White Matter Brain Structures

In recent years, a focus of the neuroscience community has been to understand the role of white matter in human cognition and function. We have developed a set of tools to study the 3D shape and shape-variability in major human white matter tracts. First, a mapping tool that extracts the skeleton of a tract and performs fine grained spatial clustering to identify correspondence between areas of tracts from different brains. This tool allows comparison of tissue properties on a fine-scale between individuals, mapping the 3D morphology of these structures across large human populations, thus exploring the available neuroimaging datasets in a principled manner. Second, a synthesis and simulation tool that applies a set of simple shape-deformations, such as bending and shearing, on a given tract. This enables parametrization of the underlying shape space and analysis of inter-individual normal and pathological shape variability. This work is currently ongoing. We have tested the fine-mapping tool on a cohort of subjects from the ADNI dataset and showed that we're able to accurately map corresponding areas on a variety of tracts across individuals.

Refereed Publications (past year only)

Robert Herzog, Daniel Mewes, Michael Wand, Leonidas Guibas, Hans-Peter Seidel. *LeSSS: Learned Shared Semantic Spaces for Relating Multi-Modal Representations of 3D Shapes*. In: Computer Graphics Forum (Proc. SGP 2015), 2015.

Minhyuk Sung, Vladimir Kim, Roland Angst, Leonidas Guibas. *Data-Driven Structural Priors for Shape Completion*. ACM Transaction on Graphics (Proc. Siggraph Asia), vol 34, 6, 2015 (to appear).

Hao Su, Fan Wang, Eric Yi, and Leonidas Guibas. *3D-Assisted Feature Synthesis for Novel Views of an Object*. International Conference on Computer Vision (ICCV), Santiago, Chile, 2015 (to appear).

Personnel Supported (past year only)

1. Leonidas Guibas, Faculty PI, 1.00 month
2. Peter Huang, Postdoctoral Fellow, 0.30 month [Shapenet repository]
3. Vladimir Kim, Postdoctoral Fellow, 7.25 months [shape completion]
4. Tany Glozman, Graduate Student, 1.50 months [white matter brain structures]

A Stanford undergraduate student (Ivan Robles) also worked with us his summer on topics related to this grant.

Education

We are planning a new course, *CS233: The Shape of Data – Geometric and Topological Data Analysis*, based to a significant part on material developed under this grant. This course will be taught at Stanford in the spring of 2016.

Interactions/Transitions

The work described this year's and in previous reports has been or will be presented in the very top venues in computer vision, computer graphics and machine learning. Earlier work was also presented at the AFOSR Computational Cognition annual meetings.

Several companies have expressed strong interest in our work. We are currently collaborating on topics related to this grant with Adobe, Apple, Autodesk, and Google.

PI Honors/Awards

- 2013 Eurographics Symposium on Geometry Processing, Best Paper Award
- 2013 International Conference on Computer Vision Helmholtz Award (recognizes ICCV papers from ten years ago with significant impact on computer vision research)

1.

1. Report Type

Final Report

Primary Contact E-mail
Contact email if there is a problem with the report.

guibas@cs.stanford.edu

Primary Contact Phone Number
Contact phone number if there is a problem with the report

650-723-0304

Organization / Institution name

Stanford University

Grant/Contract Title
The full title of the funded effort.

The Information Is In the Maps: Representations & Algorithms for Mappings among Geometric Data

Grant/Contract Number
AFOSR assigned control number. It must begin with "FA9550" or "F49620" or "FA2386".

FA9550-12-1-0372

Principal Investigator Name
The full name of the principal investigator on the grant or contract.

Leonidas J. Guibas

Program Manager
The AFOSR Program Manager currently assigned to the award

James H. Lawton

Reporting Period Start Date

09/01/2014

Reporting Period End Date

08/31/2015

Abstract

In this past year our focus has been algorithms for relating and interconnecting diverse modalities that provide information about objects in the world, including images, sketches, 3D scans, 3D models, and language. Different modalities often capture distinct types of information about the nature and state of objects in the world, so that the information fusion made possible by this integration creates new integrated knowledge not available separately from any of the modalities.

Unlike traditional data fusion, in our setting fusion is possible not only at the level of object instances but also within object categories, through the abstraction mechanisms we have developed in our networks. In particular, we have aimed to provide additional information or knowledge about captured signals (e.g., images), in a real time setting -- information that is NOT present in the raw signal but is inferred from contextual knowledge present in the knowledge network.

Specifically, we have shown how features for new views of an object can be computed from a single image, and how missing portions of a single 3D scan of an object can be completed -- in both cases exploiting network knowledge.

We have also demonstrated how to do compute joint cross-modal embeddings, including images, sketches, 3D models, and words -- and use them for retrieval across different modalities.

Distribution Statement
This is block 12 on the SF298 form.

Distribution A - Approved for Public Release

Explanation for Distribution Statement
If this is not approved for public release, please provide a short explanation. E.g., contains proprietary information.

SF298 Form
Please attach your SF298 form. A blank SF298 can be found here. Please do not password protect or secure the PDF The maximum file size for an SF298 is 50MB.

AFD-070820-035 Final Report SF 298 v2.pdf

Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF . The maximum file size for the Report Document is 50MB.

AFOSR Maps Final Year Progress Report_v2.pdf

Upload a Report Document, if any. The maximum file size for the Report Document is 50MB.

Archival Publications (published) during reporting period:

Robert Herzog, Daniel Mewes, Michael Wand, Leonidas Guibas, Hans-Peter Seidel. LeSSS: Learned Shared Semantic Spaces for Relating Multi-Modal Representations of 3D Shapes. In: Computer Graphics Forum (Proc. SGP 2015), 2015.

Minhyuk Sung, Vladimir Kim, Roland Angst, Leonidas Guibas. Data-Driven Structural Priors for Shape Completion. ACM Transaction on Graphics (Proc. Siggraph Asia), vol 34, 6, 2015 (to appear).

Hao Su, Fan Wang, Eric Yi, and Leonidas Guibas. 3D-Assisted Feature Synthesis for Novel Views of an Object. International Conference on Computer Vision (ICCV), Santiago, Chile, 2015 (to appear).

Changes in research objectives (if any):

Change in AFOSR Program Manager, if any:

Extensions granted or milestones slipped, if any:

AFOSR LRIR Number

LRIR Title

Reporting Period

Laboratory Task Manager

Program Officer

Research Objectives

Technical Summary

Funding Summary by Cost Category (by FY, $K)

	Starting FY	FY+1	FY+2
Salary			
Equipment/Facilities			
Supplies			
Total			

Report Document

Report Document - Text Analysis

Report Document - Text Analysis

Appendix Documents

2. Thank You

E-mail user

Sep 27, 2015 14:06:24 Success: Email Sent to: guibas@cs.stanford.edu