Gehry versus Zumthor: Machine Recognition of Architectural Styles

VERONICA FINNEY
University of Idaho

FRANK JACOBUS
University of Idaho

JAY MCCORMACK
University of Idaho

JOSH HARTUNG
University of Idaho

INTRODUCTION

The Pre-Design and Programming process in architecture is often time consuming for the architect and therefore expensive for the client. Within this process an attempt is made to understand the needs of the client in an effort to focus the Schematic Design phase of the architectural project. A lack of quality information in Pre-Design often leads to a lengthened and costly Schematic Design phase during which multiple iterations are required to ensure client satisfaction. Numerous computational design programs and strategies have attempted to mitigate these problems by automating some of the research gathering process. Unfortunately, failures in the ability to capture and reuse expert knowledge have made it difficult to create computational tools that are effective for designers at a practical level. Emerging visual recognition software offers a unique opportunity to capture formal design characteristics that historically have fallen within the purview of the architect. This alludes to the potential use of these tools as instruments that can aid expert and knowledge based systems.[1]

In this experiment, we test the ability of a computational model known as Hierarchical Temporal Memory (HTM) in order to assess how well the tool recognizes the designs of two architects whose work could be said to fall on either end of an architectural style spectrum. HTMs are built on a model of human neurological function and have shown great promise in areas of ambiguous pattern recognition in vision and speech applications.[2] [3]Most significantly, they have the ability to store data on a series of training patterns and subsequently identify incomplete or out of sequence pattern data that they have not seen before.[4] This allows the system to learn the characteristics of a series of anecdotal data and then evaluate new data based on its similarity to the exemplars.[5] In a design context this is a first step toward a technique for computational evaluation of quality by analogy, where solutions are judged based on their similarities to known, preferred solutions. The ability to find meaningful connections between groups of objects, while allowing for differences and computing with ambiguity, means that new solutions can be dramatically different and even novel while retaining some core familiarity that is preferred by the designer or client.

This experiment provides support that HTMs can be effective tools for recognizing stylistic differences in building design, thereby enabling the tool to assess success of novel designs that are to meet certain stylistic parameters set forth during the Pre-Design effort. A computational model of stylistic quality that is consistent with that of a human is a powerful first step toward general automated design tools and toward the effective population of expert and knowledge-based systems.

BACKGROUND

Hierarchical Temporal Memory, (HTM) developed by Numenta, Inc., is a recent form of artificial intelligence that excels at ambiguous pattern recognition. Based on an emerging theory of human neurological function, HTMs can solve ambiguous and multi-modal problems which traditional computing systems find difficult or impossible. Promising results have been observed in machine vision, voice recognition, and objective character recognition.[6]

HTMs are not programmed in a traditional sense; rather they are trained on input data. They have two modes: training and inference. During training, learning nodes arranged in a hierarchy identify and store patterns in space, and then in time. Information vectors are passed first into a spatial pooler where it is evaluated for similarity to other vectors that have already been seen. If the vector is spatially the same or similar (within some defined range) to a vector that has already been seen, it is identified as a coincidence of the exemplar vector. If the vector is beyond the defined range, it is recorded as a separate coincidence. The coincidence information for each vector is then passed to a temporal pooler.[7]

The temporal pooler identifies patterns in time. As new vectors are received, coincidences come from the spatial pooler in a sequential manner. These sequences are stored and evaluated for coincidence in a similar manner to the spatial pooler. Finally, the temporal coincidences are output from the node. This information is then propagated up the hierarchy, each level feeding its parent the information from several nodes. The result at the highest level is an invariant understanding of the problem data, which may be used for inference.[8]

During inference the HTM is presented novel data from the same category as the training data. This data may be highly ambiguous. For instance, if an HTM was trained on photos of full cows, then novel data might be a picture of a cow partially obscured behind a barn. This is a completely different data set than was initially presented: large amounts of data are missing. Perhaps the head of the cow, the hind legs, or the midsection are obscured. Each of these circumstances presents a problem to traditional pattern recognition systems. The HTM however, still sees a significant portion of the patterns it identified during training. As this data propagates up the hierarchy, incomplete patterns are filled in by probabilistic analysis. The HTM returns the result: a cow. Just as a human would do, the system assumed that because it saw the front of the cow, then the rear of the cow must be attached and standing behind the barn.[9]

In this example, the HTM stored the underlying patterns that represent all cows in all situations from a limited set of data containing a few cows in a few situations. Rather than storing specific information about the form of the animal, the HTM stores the hierarchy of coincidences over space and time that are common to the observed cows. In this way, the HTM stores underlying patterns and uses them to interpret vague or incomplete data in a similar manner to a human.[10]

EXPERIMENT

Recent research into HTM systems by the authors of this paper has included an experiment to recognize chair back styles made by three different designers. Within this experiment the HTM tool was tested to determine how well it recognized stylistic qualities in relation to human recognition tendencies. To verify this, human subjects were asked to develop a qualitative representation of style for the chair backs using image data that was identical to that which the HTM was trained on. By correlating qualitative statements made by the human subjects with the results obtained from the HTM, the study found the system's recognition of quality to be analogous to that of the human.[11] The experiment cited in this paper tests the HTM software to determine whether it can make a distinction between the work of Frank Gehry and Peter Zumthor. Testing the ability of human beings to make a distinction between the work of Gehry and Zumthor did not seem necessary due to the stark and obvious difference in the two styles and based on the discovery in the previous paper.

Defining an Architectural Style

The designation of an architectural style is an attempt to classify architecture in terms of form, techniques, time period, region, materials and other influences.[12] For the purpose of this paper we will focus on form and technique as providing enough of a stylistic determinant to be useful in exploring human readings of style versus that of the HTM tool.

The majority of the architectural work of Frank Gehry has been formally classified as expressionist, deconstructivist and postmodern.[13] This fragmented, collaged style helps to distinguish his work from that of other architects. Visually speaking, the overall form of his most seminal works, such as the Guggenheim Museum in Bilbao and the Walt Disney Concert Hall in Los Angeles, is composed predominantly of complex and fragmented shapes and lines. Gehry's work can be described as consisting of an array of diagonal and curvilinear lines that form individual masses superimposed atop or adjacent to one another. These individual masses form an asymmetrical yet balanced relationship and are often disconnected from the ground plane at various entry locations. Because the lines of his structures often curve in the x, y, and z-axis, and are typically composed of a smooth surface, a complex network of shadows exist on the building envelope. In many of Gehry's seminal works the asymmetrical masses rise in space toward the center of the building.[14]

Radically contrasting the style of Frank Gehry is the work of Swiss-born architect Peter Zumthor. Zumthor's work indicates an appreciation for the climate, landscape, and vernacular architecture that exists in his homeland, Switzerland. The visual language of Zumthor's buildings is typically simple, and their lines are predominantly rectilinear and orthogonal. As opposed to the tendency to create buildings that look as though they are a composition of irregular masses, as described above for Frank Gehry, many of Zumthor's buildings look as though they are a single solid mass from which material has been removed to form voids.[15] Simplicity of form and simplicity of line define the majority of Zumthor's oeuvre as opposed to the complexity of overall form and line in Gehry's work.

Experiment Preparation

After defining the visual language with which humans classify these radically different styles, we began to compile the images for the experiment. Enough images had to be gathered to adequately populate the training data. Though there is no specific number of images required by the software, through past experiments and early testing of this experiment, we found that we needed at least seventy images to populate the training data. The first images we gathered were photos of existing buildings, predominantly from a front elevation viewpoint to the extent possible. With only these images as training and testing sets we had some difficulty achieving levels of recognition that indicated that the HTM tool was achieving success. Rather than preparing sophisticated drawn images of the architects' projects, we began by simplifying their work into a very blunt and straight forward visual language. If we were to casually classify "Gehry-style" and "Zumthor-style" into one form on a page, the Gehry-style would be a curvilinear line and the Zumthor-style would be a straight line. Eventually we learned that a mixture of black and white photographic images and hand drawn graphics, acting as simple line work that represented the building elevations from each architect, was the most effective representation of their style characteristics. We didn't consider it to be a problem that the tool had trouble recognizing the style characteristics from a set consisting only of photographs because the goal was style recognition generally, not through any specific means. Ultimately, by creating graphic representations of "Gehry" and "Zumthor" as styles we aided the HTM tools training catalog. The proof as to whether our own graphic representation of these styles was effective would eventually come when we tested the system for its recognition of "Gehry" or "Zumthor" as a style. If the test itself was on photographic images of the building and the tool recognized the style then our graphic representations would be justified.

In addition to the above we ensured that all images were the same size and file type. We also made sure that the datum of the ground plane was relatively consistent from image to image. These measures were taken to ensure consistency and avoid as much confusion as possible for the HTM tool itself. These strategies were important for us because of a limitation of training data. A more robust system would analyze thousands, if not millions of images, as a training mechanism, whereas we trained our system on less than one-hundred images.

For Gehry, the drawings consisted of a series of sketches based off of the visual language of his work described above, where the sketches all shared a common scaled ground line, and resembled some of his seminal architectural work. Also there were some included that didn't directly reference a form of one of his projects but to the human eye could be classified as such. The Gehry-style graphic is shown in Figure 1 below.

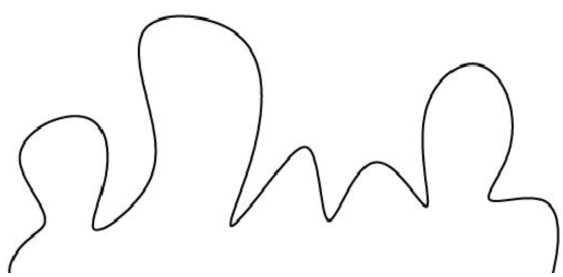

Figure 1. Training Sketch for Gehry Network

For Zumthor, the drawings consisted of a series of simplified sketches based off the visual language describing his work above. The sketches all shared a common ground line that matched the ground line in the Gehry sketches to ensure a similar graphic structure. Just as with Gehry's images, the sketches were drawn to epitomize the overall gestures of his seminal projects.

Populating the Experiment

The majority of the preparation process prior to conducting the experiment entailed populating both the training and testing images as described above. The use of hand drawn images as training and testing data proved to be ineffective in our initial experiments due to their inherent inconsistencies. Based on these inconsistencies in early trials a decision was made that all future training and testing data would be produced digitally. Using Adobe Illustrator, we were able to size all of our image files according to the end image size that we required, thus maintaining a consistent size in resolution and file type such that neither of these factors would influence the results of the experiment. Every attempt possible was made to maintain the drawing's size and shape characteristics. In addition, we used the same line types and weights so as not to over-complicate the image with background data that might confuse the software. With this new criteria we produced thirty plus images representing the iconic buildings of each architect using very simplified and streamlined graphics; this is demonstrated in Figure 2 below.

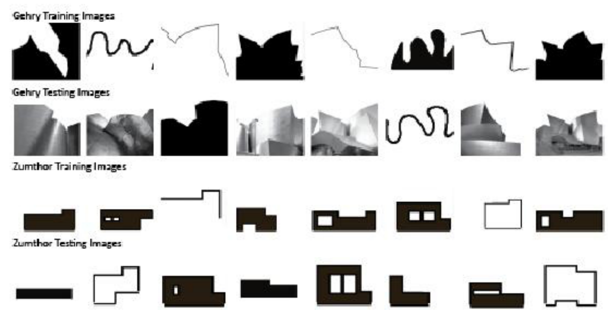

Figure 2. Portion of Gehry Training Set

Next, we further populated the experiment by digitally tracing over actual photographic images of the iconic works of each architect. We still attempted to use views that were as close as possible to a straight elevation so as to simplify the input data and we cropped the background of each image such that line work behind each building would not obscure the HTMs ability to recognize the work. We produced these by tracing images and technical drawings of specific seminal works from the architects.

Preliminary Training

Using Numenta's Vision Toolkit we uploaded the prepared images described above, labeling the curvilinear dominant images as "Gehry" and the orthogonally dominant images as "Zumthor", creating a visual representation that the computer would associate with each architect respectively.

After uploading the training images into the Numenta Vision Toolkit we ran the systems training exercises. The system first inverted the images and then placed them on a black background, eliminating the white background present in the originals. This was done in-situ by the Vision Toolkit's onboard image conversion features. We then tuned network parameters individually to ensure spatial learning was storing unique coincidences and there was temporal coherence between the spatial patterns.[16] This process of storing the similarities and unique coincidences prepared the system for the actual testing and successful classification of new graphics to take place.

HTM Trial

Upon completion of training, the HTM had a stored classification for each set of training images. The software was trained on a total of one-hundred fifty original images, approximately seventy-five for each category, and tested on fifty images, twenty-five in each category. The majority of the fifty testing images were original and were compiled from actual photographs of the architect's work. The inclusion of non-original images allowed us to test how well the HTM recognized these images in relation to the novel images. As in the testing images, the backgrounds of the training images were eliminated so that extraneous information would not influence the test. The images were also reduced or expanded such that they were the same size and they were set to grey scale to create consistencies among the testing set. A number of the testing images were cropped and reused as detail images to help populate the experiment and to allow us to test the HTMs ability to recognize micro scale line work and detail. Once the images were uploaded into individual testing sets, the Numenta Vision Toolkit examined the training images for similarities and consistencies so as to begin forming a classification for each image.

RESULTS OF EXPERIMENT

Of the fifty images tested the HTM software successfully recognized forty-nine, giving it a ninety-eight percent success rate. The only image that the software did not recognize was a Zumthor drawing that we had constructed as a random simplified image of line work that consisted of straight vertical and horizontal lines. It is difficult to tell exactly

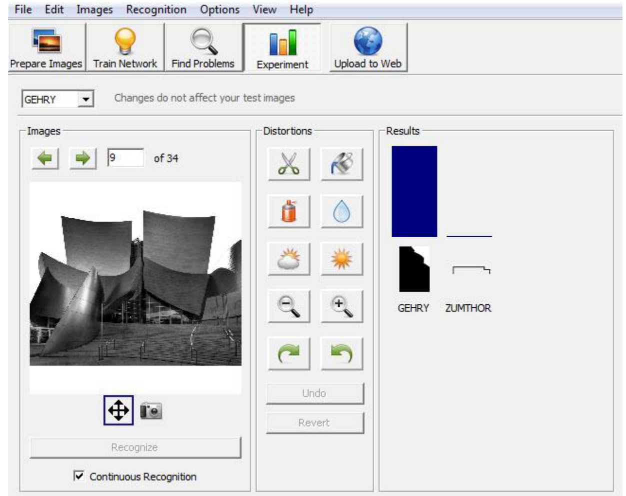

Figure 3. Numenta Vision Toolkit Interface

why the HTM assumed that this was Gehry's work and not Zumthor's, though the Vision Toolkit does provide ranked images that the tool felt were similar to the exemplar. (Figure 4) A ninety-eight percent rate of success, especially with such limited training data, indicates that the HTM tool has the ability to recognize and categorize architectural styles.

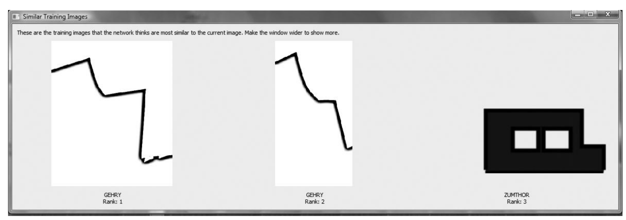

Figure 4. Comparing Ranked Images

CONCLUSION

Hierarchical Temporal Memory is widely used for speech recognition, optical character recognition (OCR) and even statistical trend analysis.[17] Some form of visual recognition using computational tools is used for emerging applications such as Google Goggles which allows a person to take a picture with their smart phone and retrieve immediate information about the image or about their geographic location based on the image.[18] The emergence and now pervasiveness of visual recognition tools begs the question about their potential effects in the architectural realm. Testing whether HTM can recognize the difference between Frank Gehry and Peter Zumthor's work is the start of a fundamental questioning about the ability for our computational tools to recognize architectural styles and ultimately to recognize what makes those styles unique. In conducting these experiments we were interested in both the short term and long term questions involved in the study. Short term questions involve the HTMs capability of making a distinction between the work of two architects given a relatively small database of precedents. The tools ninety-eight percent rate of success indicates that it is able to make this distinction. One can readily see where this machine learning ability can be applied in smart phone applications for travel and education as is moving forward in the Android market through the use of Google Goggles currently. Other applications such as the automated population of specific databases based on a machine search of broad internet databases perhaps are on the horizon or even being implemented currently. But what are some long term applications of this type of software? What does the arrival and evolution of this software imply for the architectural realm? For instance, if the software can recognize macro scale form and associate it with wiki style data compilation, can it also begin to recognize minute detail within that form? Will the software eventually recognize material detail and associate that detail with the objects physical properties? Finally, if the machine learning software can recognize form, material, and patterns in space, will it eventually be able to reconstitute these properties into a new physical reality? In other words, how soon will the learning machine become a creating machine, capable of rapid iterations on the way to a new product?

Future Uses

This research tested HTMs capabilities of categorizing objects based on a pre-defined understanding of style. This test acts as a starting point for future discussions regarding the use of HTMs and systems like them in the architectural domain. Near future applications of this technology may include the following:

Object Classification and Categorization such that any uploaded image into a web database filters through the Vision Toolkit and is instantly classified by style, architect, era, size, use, typology, location, etc. by having its visual data filtered through pre-collected information in the database.

This software could also contribute to an inductive search engine that provides the possibility of entering textual characteristics of an object you are looking for and allowing the database to search for these objects based on its collected visual characteristics.

In addition, the HTM software could be used to help find architectural projects, or even products, that are similar stylistically to those that you are familiar with. For instance, one could input "Gehry" as a precedent and the visual recognition software could find projects that relate to Gehry stylistically and could provide feedback as to how closely they relate and why.

Further development in the environment of construction methods and specifications could include a user photographing pieces of an existing project and immediately extracting specification and detail information. The photograph would filter through an existing web database and connect to an existing material specification database giving immediate feedback to material and structural properties among other things.

The HTM tool could also be used to populate recommender systems, similar to those being used by companies such as Pandora, Amazon, and others. The user would fill in specific information and the software would help to link the recommender system to related needs; or simply to introduce the user to novel architectures. The software would literally find other architectures that one might like based on a single precedent.

HTM has the potential for future applications that would directly impact practice. First, the software could speed up the precedent search process and could extract valuable data from that process including spatial typologies linked with particular use strategies. The software could also link particular spatial types with code compliant uses; connecting space directly with building type precedents and code requirements. This could significantly reduce the amount of time dedicated to pre-design and schematic design efforts.

Highly reliant on technology already, practitioners will continue to move towards a technologically dominant future. Design methods and processes will evolve, continually changing the nature of the relationship between the architect and their machines. Architects will have to determine for themselves how they individual wish to engage with emerging digital tools which seem to be evolving toward the design companion paradigm.

Eventually we hope to utilize this system in a general design synthesis situation where it can leverage multi-modal data to create new designs that build off of existing ones. More testing and experimentation with this software is necessary to truly understand the future applications that may arise as a result of its application in practice.

ENDNOTES

1 Y.E. Kalay, *Architecture's New Media* (Cambridge: The MIT Press, 2004)
2 "Problems That Fit, Numenta, Inc.," cited February 27, 2009, http://www.numenta.com/for-developers/education/ProblemsThatFitHTMs.pdf.
3 "Hierarchical Temporal Memory – Comparison with Existing Models," cited February 27, 2009, http://www.numenta.com/for-developers/education/HTM_Comparison.pdf.
4 "Problems That Fit, Numenta, Inc."
5 "Hierarchical Temporal Memory – Concepts, Theory, and Terminology," Jeff Hawkins and Dileep George, cited February 27, 2009. http://www.numenta.com/Numenta_HTM_Concepts.pdf.
6 "Hierarchical Temporal Memory – Concepts, Theory, and Terminology"
7 "Hierarchical Temporal Memory – Concepts, Theory, and Terminology"
8 "Hierarchical Temporal Memory – Concepts, Theory, and Terminology"
9 Dr. Jay McCormack, Josh Hartung and Frank Jacobus, "Support for the Use of Hierarchical Temporal Memory Systems in Automated Design Evaluation: A First Experiment." (Proceedings of the American Society of Mechanical Engineers (ASME) IDECT/CIE. San Diego, California, 2009)
10 McCormack, Hartung and Jacobus, "Support for the Use of Hierarchical Temporal Memory Systems in Automated Design Evaluation: A First Experiment."
11 McCormack, Hartung and Jacobus, "Support for the Use of Hierarchical Temporal Memory Systems in Automated Design Evaluation: A First Experiment."
12 Dictionary.com, May 2011, http://dictionary.reference.com/
13 Mildred Friedman and Michael Sorkin. *Gehry Talks: Architecture + Process* (New York: Rizzoli, 1999)
14 Frank O. Gehry and Rosemarie H. Bletter, *The Architecture of Frank Gehry* (New York, NY: Rizzoli International Publications, 1986)
15 "Peter Zumthor", *Tokyo: A and U* (1998)
16 McCormack, Hartung and Jacobus, "Support for the Use of Hierarchical Temporal Memory Systems in Automated Design Evaluation: A First Experiment."
17 B.D. Ripley, *Pattern Recognition and Neural Networks* (Cambridge, 1996)
18 "Google Goggles," Cited September 2, 2011, http://en.wikipedia.org/wiki/Google_Goggles