Engineering sketch recognition: findings from recent bio-inspired and cognitive studies at VDEL

Luoting Fu
Visual Design and Engineering Lab
Carnegie Mellon University
5000 Forbes Ave.
Pittsburgh, PA 15213 USA
luoting.fu@cmu.edu

Mark D. Fuge
Visual Design and Engineering Lab
Carnegie Mellon University
5000 Forbes Ave.
Pittsburgh, PA 15213 USA
mfuge@andrew.cmu.edu

Levent Burak Kara
Visual Design and Engineering Lab
Carnegie Mellon University
5000 Forbes Ave.
Pittsburgh, PA 15213 USA
lkara@cmu.edu

CHI 2010, April 10–15, 2010. Atlanta, Georia, USA
ACM 978-1-60558-930-5/10/04.

Abstract

In this position paper, we review the latest findings in bio-inspired and cognitive studies in diagrammatic engineering sketch understanding in the Visual Design and Engineering Lab (VDEL) at Carnegie Mellon University. We discuss issues including (i) image-based recognition of off-line sketches with a bio-inspired visual pipeline, (ii) the stroke-image duality of online sketches and (iii) a cognitive model of human sketch recognition. We present our position of their implications for future sketch understanding research.

Keywords

sketch recognition; visual recognition

ACM Classification Keywords

I.5.4 Pattern Recognition: Applications

Background

Many engineering models are graphically depicted as sketches on two-dimensional media and are often acquired, preserved and circulated as such in the digital universe. For example, engineering diagrams in digitized legacy patents, textbooks and other technical archives are scanned into the computers as sketches; electric circuits posted on the Internet and indexed by search engines are often sketches in image format; pedagogical examples of control systems drawn on the white board or preliminary designs sketched on the back of an envelope or on a Tablet PC can be captured as digital sketches. The recognition of such diagrammatical sketches entails the automatic conversion from the sketch representation to another representation of interest, such as simulation-enabled, "live" engineering models.

Over the past few years, researchers in the Visual Design and Engineering Lab (VDEL) at Carnegie Mellon University have developed a number of engineering sketch recognition systems and performed several user studies.

A major theme of our sketch recognition research is to formulate diagrammatical sketch recognition as the sequential executions of two sub-tasks, *parsing* which locates the sketched engineering symbols and *recognition* which categorizes the identified symbols [6]. Earlier efforts included heuristic [9, 6, 4, 7] and graph-based [12] approaches for parsing and recognition. Recently we explored image-based approaches, ranging from image-based recognition of pre-segmented image patches [10] to a bio-inspired pipeline for image-based sketch parsing and recognition [2].

Our users studies sought to benchmark software performances, evaluate user experience [8] and probe the cognitive mechanism of human sketch recognition [3].

The following sections briefly summarize our works in the bio-inspired computational models [2] and cognitive user studies [3] of engineering sketch recognition and present our positions on future advances. Interested readers are referred to the full version of the cited works.

Image-based Recognition of Off-line Sketches with a Bio-inspired Visual Pipeline

In [2] and its recent extension under review, we implemented a bio-inspired, computational recognition approach to convert network-like, image-based engineering diagrams into engineering models. Such conversion enables a broad spectrum of computations of engineering interests, such as CAD modeling, simulation, information retrieval and semantic-aware editing.

Our overall approach, shown in Figure 1, leverages a Convolutional Neural Network (CNN) [11] as a trainable engineering symbol recognizer. The CNN emulates the visual cortex of mammals and is capable of learning the visual features of the defined symbol categories from a user-supplied, labeled training examples. Once trained, the CNN is able to output a categorical label of the input image patch. Compared with alternative image-based recognition techniques, the CNN recognizer has several useful properties, such as built-in invariance to moderate distortions that are inherent in freehand sketches, high processing speed due to the feed-forward nature of the pipeline and domain-neutrality that ensures easy re-targeting to new sketching domains.

To apply a trained CNN to an entire input sketch which contains many image patches, two parsing modules are developed. First, a generic approach is to apply a sliding window on a multi-scale image pyramid produced from the input sketch.

The exhaustive nature of this approach increase the chance of locating the symbols within the sketch. Secondly, if the diagramming syntax of a domain guarantees symbols isolated from each other, Connected Component Analysis (CCA) [5] can be used as a fast and accurate alternative to the sliding window.

Once the sketch is parsed and the symbols are detected, the connectivity between symbols are then analyzed and an attributed graph representing the engineering model conveyed by the diagram can be produced for subsequent engineering computation.

The utility of the proposed approach includes different application scenarios, such as the construction and simulation of control system or mechanical vibratory system models from hand-sketched or camera-captured images, content-based image retrieval for resonant circuits and semantic-aware image editing for floor plans.

The CNN-based recognizer is a generalized approach applicable to multiple engineering domains regardless of whether the diagrams are manually drawn, captured by image acquisition devices, or rendered from digital ink on a tablet PC. In comparison, previous works in sketch understanding mostly rely on stroke-level, online features and hence are limited to diagrams drawn with a digitizer on a tablet PC. In contrast, the CNN-based approach, and other similar visual recognition approaches independently developed at MIT at the same time [14, 13], opens up new directions for recognition approaches and application niches. Examples of potential applications include digitizing archived engineering drawings, performing semantic-based sketch retrieval and sketch-based modeling on traditional media with the aid of a camera, as shown in Figure 2.

The Stroke-Image Duality of Online Sketches

The online representation of sketches captured from Tablet PCs features the stroke-image duality. On one hand, online sketches are 1D streams of sample points on the pen trajectory. On the other hand, they represent the 2D spatial distributions of pixels of the ink rendered on the computer display.

It is this dual access to the stroke-based pen trajectory and pixel-based representation of ink that distinguishes online sketch recognition from its purely 1D sibling fields such as speech recognition or purely 2D fields such as optical character recognition.

While certain tasks (*e.g.*, parsing, recognition) within certain engineering domains are better served with one representation, the whole sketch recognition pipeline can be enhanced by the introduction of the other, complementary representation. For example, image-based representation and convolutional neural nets provide invariance to local distortions of hand-sketched engineering symbols. Conversely, the speed and curvature profile of a stroke can facilitate the parsing of amorphous symbols with vastly different trajectories for certain parts, symbols of highly varying aspect ratio, or symbols with rotation invariance. This suggests that a combined recognition approach tapping into both representations could be very fruitful. Our previous work [9] and recent work by MIT [13], utilizing both stroke-based parsing (*i.e.*segmentation) and image-based symbol recognition, offer evidence in this regard.

The sub-tasks of parsing and recognition do not have to rely on one representation exclusively. It would be interesting to further investigate combined parsing cues or recognition features derived from both strokes and pixels. For example, a "rectangular-ish" cluster of pixels laying on the trajectory featuring a few drops in the speed profile would very likely be a rectangle.

A Cognitive Model of Human Sketch Recognition

Other than the neural models presented earlier, we also believe that another possible step toward the development of a robust sketch recognition is to determine and emulate how humans interpret sketches. To explore this, we developed a method in [3] to gain insights into human diagram recognition using techniques analogous to peripheral vision and human attention. Following this, a cognitive model of human diagram understanding was developed from which a computational structure can be constructed to mirror human diagram recognition.

As detailed in [3], our cognitive model breaks the human diagram recognition process down into four main steps, each responsible for adjusting a single "Dominant Hypothesis". This "Dominant Hypothesis" is an abstract internal model of the diagram that can be simulated by the mind, and is progressively updated as new information from the diagram is incorporated. As a result of this internal model, the human mind is able to converge towards a consistent sketch interpretation without requiring a full understanding of the entire image. Similar computational efforts based on Dynamic Bayesian Networks [1] have demonstrated the potential of this approach. We found that while humans are very good at interpreting sketches, they also suffer from critical issues that future computational tools may not want to inherit: rather than treating each hypothesis equally, humans select a Dominant Hypothesis as the working model and start to become blind to potential interpretations of symbols that are not supported in their mental model. While the use of a single hypothesis reduces the cognitive load on humans and speeds up recognition under uncertainty, it might increase the rate of false recognitions or decrease the speed at which problems are correctly recognized by computers. For this reason, emulating some specific aspects of human behavior may not be desired. Nonetheless, we still believe the general strategy of modeling computational tools on human behavior is fundamentally sound.

Conclusions

We sought biological and cognitive inspirations from believably the best performing sketch understanding system: the human mind. From the biological visual cortex, we implemented a neural network model for image-based sketch recognition. As an alternative to conventional stroke-based approaches, it features broadened application niches.

We note that the unique stroke-image duality of online sketches implies rich information that must be leveraged by combined stroke-image recognition approaches in future. We believe that bio-inspired learning models such as neural nets and Conditional Random Fields have great utility here.

Our cognitive studies suggested that humans tend to maintain only one Dominant Hypothesis in active memory in order to reduce cognitive load during recognition, adopting an accuracy-complexity trade-off different from most computational counterparts.

In conclusion, our explorations thus far have given us the confidence that emulating the biological and cognitive models of human sketch recognition is a promising venue for sketch understanding research in the future.

References

[1] C. Alvarado and R. Davis. Sketchread: a multi-domain sketch recognition engine. In *UIST '04: Proceedings of the 17th annual ACM symposium on User interface software and technology*, pages 23–32, New York, NY, USA, 2004. ACM.

[2] L. Fu and L. B. Kara. Recognizing network-like hand-drawn sketches: a Convolutional Neural Network approach. In *Proceedings of ASME 2009 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, San Diego, CA, 2009.

[3] M. D. Fuge and L. B. Kara. A testing method and cognitive model of human diagram understanding for automating design sketch recognition. In *Proceedings of ASME 2009 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, San Diego, CA, 2009.

[4] L. Gennari, L. B. Kara, and T. F. Stahovich. Combining geometry and domain knowledge to interpret hand-drawn diagrams. *Computers and Graphics*, 29(4):547–562, 2005.

[5] R. M. Haralick and L. G. Shapiro. *Computer and Robot Vision*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1992.

[6] L. B. Kara. *Automatic parsing and recognition of hand-drawn sketches for pen-based computer interfaces*. PhD thesis, Carnegie Mellon University, Pittsburgh, PA, USA, 2004.

[7] L. B. Kara, L. Gennari, and T. F. Stahovich. A sketch-based tool for analyzing vibratory mechanical systems. *J. Mech. Des.*, 130(10):101101, 2008.

[8] L. B. Kara, K. Shimada, and S. D. Marmalefsky. An evaluation of user experience with a sketch-based 3d modeling system. *Computers and Graphics*, 31(4):580–597, 2007.

[9] L. B. Kara and T. F. Stahovich. Hierarchical parsing and recognition of hand-sketched diagrams. In *User Interface Software Technology*, Santa Fe, New Mexico, 2004.

[10] L. B. Kara and T. F. Stahovich. An image-based, trainable symbol recognizer for hand-drawn sketches. *Computers and Graphics*, 29(4):501–517, 2005.

[11] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, November 1998.

[12] W. Lee, L. B. Kara, and T. F. Stahovich. An efficient graph-based symbol recognizer. *Computers and Graphics*, 31(4):554–567, 2007.

[13] T. Y. Ouyang and R. Davis. Learning from neighboring strokes: Combining appearance and context for multi-domain sketch recognition. In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems*, pages 1401–1409. MIT Press, Cambridge, MA, 2009.

[14] T. Y. Ouyang and R. Davis. A visual approach to sketched symbol recognition. In *Proceedings of the 21st International Joint Conferences on Artificial Intelligence*, pages 1463–1468, Pasadena, CA, 2009. Morgan Kaufmann.

About the Authors

Luoting Fu is a PhD student in the Department of Mechanical Engineering at Carnegie Mellon University. His research interests include sketch understanding and machine learning. He has a BS in aerospace engineering from Shanghai Jiao Tong University.

Levent Burak Kara is an assistant professor in the Department of Mechanical Engineering at Carnegie Mellon University. His research interests include intelligent design tools, pen computing, geometric modeling, human–computer interaction, and artificial intelligence. Kara has a BS in mechanical engineering from the Middle East Technical University, and an MS and PhD in mechanical engineering from Carnegie Mellon University.

Mark D. Fuge is a research associate in the Department of Mechanical Engineering at Carnegie Mellon University. His research interests include augmented reality for engineering design. He has a BS and MS in mechanical engineering from Carnegie Mellon University.

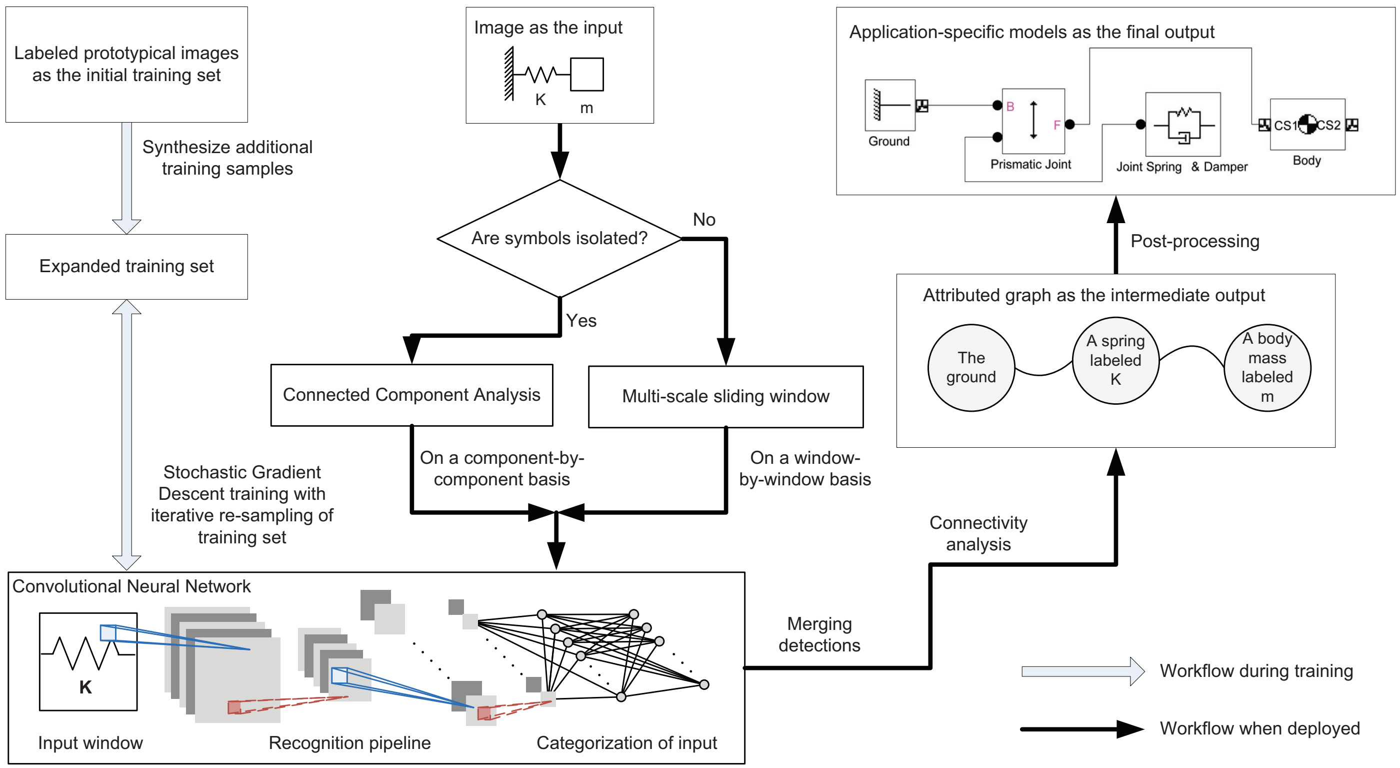

Figure 1: Overview of the proposed approach of [2] and its recent extension. At the heart of the approach is the Convolutional Neural Network as the symbol recognizer.

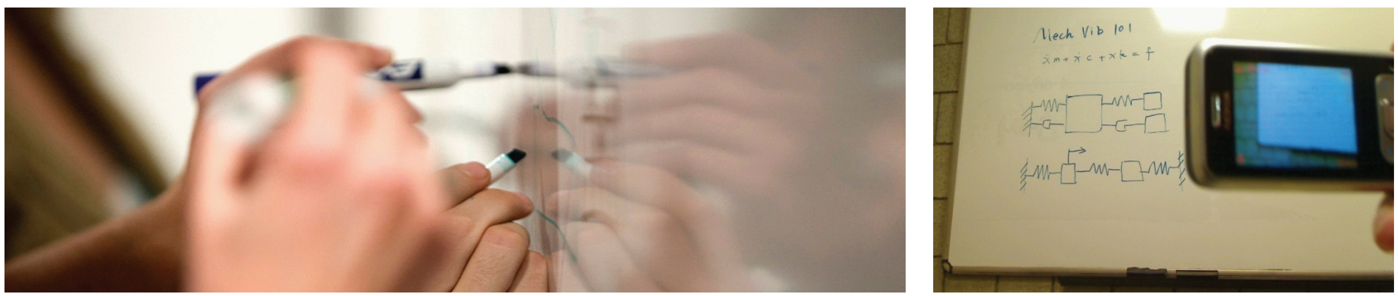

Figure 2: Sketch-based modeling on traditional media with the aid of a camera.