**Proceedings of ASME 2009 International Design Engineering Technical Conferences &
Computers and Information in Engineering Conference
IDETC/CIE 2009
August 30-September 2, 2009, San Deigo, USA**

DETC2009-87398

A TESTING METHOD AND COGNITIVE MODEL OF HUMAN DIAGRAM UNDERSTANDING FOR AUTOMATING DESIGN SKETCH RECOGNITION

Mark D. Fuge
Department of Mechanical Engineering
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213
Email: mfuge@andrew.cmu.edu

Levent Burak Kara*
Department of Mechanical Engineering
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213
Email: lkara@andrew.cmu.edu

ABSTRACT

Sketches, whether hand-drawn or computer generated, are a natural and integral part of the design process. Despite this fact, modern day computational design tools are ill-equipped to take full advantage of sketching input. The computational challenges of recognizing sketches are easily overcome by human visual recognition and much insight stands to be gained by emulating human cognitive processes. Creating robust, automated tools that overcome the ambiguity of sketching input would allow for advances not only in the practice of engineering design, but in the education of design itself. One first step toward the development of a robust sketching tool is to determine how humans interpret mechanical engineering diagrams. This paper presents two contributions toward the goal of an automated diagram understanding system. First, a method is presented to gain insight into human diagram recognition using techniques analogous to peripheral vision and human attention. Following this, a cognitive model of human diagram understanding is presented from which to further develop computational design tools. With this work, researchers should be able to (1) improve understanding of human diagram recognition and (2) use our model to emulate human diagram recognition in future computational design tools.

INTRODUCTION

Sketches and diagrams are some of the oldest and most widely used tools by engineers. They are the fastest way for us to record our ideas in visual form and they act as an essential element in group communication. Engineers and architects [1] are famous for "back of the envelope" sketches, and some even use drawing as a thinking tool [2]. Decades of study into the role of sketching in the design process have demonstrated its importance. Ullman *et al.* [3], in a seminal study on the importance of drawing in mechanical design, concluded that "CAD systems must allow for sketching input". Recent work has echoed Ullman's findings, further necessitating the development of intelligent sketch understanding systems. Shah *et al.* [4] show that collaborative sketching methods produce designs of higher quality when compared with non-sketching methods. Work by McKoy *et al.* [5] furthers Shah's research, and concludes that "Sketching is best for representing ideas generated during conceptual design, compared to textual representations." Schütze *et al.* [6] showed a strong positive correlation between sketching and resultant design quality, concluding that "digital sketching tools... can create potentially large time and cost savings for computer-aided design in mechanical engineering." Work done by Tversky [7, 8], Yang *et al.* [9, 10], Song and Agogino [11], and others [12] have reinforced the importance of sketching tools in the engineering design process.

Despite the importance of sketching in the engineering design process, most modern day Computer Aided Design (CAD) systems have difficulty tapping into this ubiquitous form of communication. Sketching is inherently ambiguous and the same symbol can have different meanings depending on both the context and the domain (the symbol for a spring and a resistor are

*Address all correspondence to this author.

identical, yet they mean different things). In addition, unlike sketches generated via a tablet PC, scanned sketches lack temporal information about stroke order which might help identify or segment symbols. Despite these challenges, humans are capable of performing sketch identification with relative ease. By understanding the ways that humans perform sketch recognition, it may be possible to develop computational methods that overcome sketch ambiguity and perform more akin to human recognition.

Recognizing diagrams, which are cleaner and less ambiguous than sketches, is the first step in enabling more efficient sketch-based computational tools. For the purpose of this work, the type of diagrams studied are mechanical engineering textbook diagrams, examples of which can be seen in Fig. 11. In this work, a new method is presented that dissects the human diagram recognition process in order to understand how humans overcome diagram ambiguity. Results from a small five person user study are presented and the significance of the results is discussed. We also present a cognitive model of human diagram understanding that lays the groundwork for the future development of computational tools that emulate human recognition. This new method, along with the cognitive model, represents a new avenue for solutions to the challenges of sketch recognition noted above. Our main contributions are two-fold: (1) the new testing method enables a new direction of research into human diagram understanding, using techniques based on peripheral vision and human attention, and (2) the cognitive model provides a foundation of knowledge from which new computational tools can arise that emulate human cognitive processes.

RELATED WORK

The problem of sketch and diagram understanding has been explored for a number of years in a wide variety of fields, including computer science, engineering, and cognitive psychology. Relevant research toward solving this problem can be broken down into three main areas: Sketch Identification, Sketch/Diagram Understanding, and Human Visual Understanding. The following section will present related research and publications in each of these areas.

The goal of Sketch Identification research is to use visual information to identify elements within a sketch or diagram. This not only includes research in identifying symbols themselves, but also in how to group or segment relevant symbols together. Ramani *et al.* have demonstrated the power and relevance that sketch identification techniques can have on the mechanical design process. Through the use of probability based classifiers, Ramani *et al.* were able to use three sketches of a part to accurately identify and retrieve a corresponding solid model [13]. Igarashi *et al.* showed how identifying certain geometric relations between lines, such as parallelism, could allow for the beautification of sketches [14]. Igarashi's methods provide insight, from both a human and computational perspective, on how geometric relations play a part in sketch recognition. Saund *et al.* demonstrate how the use of gestalt principles, such as smooth continuation and spatial proximity can group and segment sketches in a way that better emulates human performance [15, 16]. Kara *et al.* showed how stroke clustering algorithms based on minimum spanning trees can be used effectively to segment and identify elements in an online sketch environment [17, 18].

In contrast with sketch identification, the field of sketch and diagram understanding undertakes the challenge of drawing qualitative meaning from symbols. Rather than purely identifying symbols, sketch and diagram understanding attempts to assess the qualitative behavior of elements within a sketch, for the purposes of identification and simulation. For the past decade, Robert Futrelle has been developing the *Diagram Understanding System*, utilizing context-based constraint grammars and spatial indexing to identify 2D graph-based scientific data [19, 20]. By formulating the task of symbol identification and simulation as a constraint satisfaction problem, Kurtoglu and Stahovich have produced a symbol recognition system capable of correctly determining the meaning of a limited set of symbols used across multiple domains [21]. Similar goals were achieved by Alvarado and Davis through dynamically constructed bayes nets, wherein both user stroke data and contextual information informed the recognition process [22, 23], and through "categorical and situational rules" [24]. Causal reasoning techniques, specifically Qualitative Configuration Spaces, have been explored by Stahovich and Kara, and have been used to not only computationally simulate the behavior of mechanical diagrams but also to synthesize new designs [25–27].

Lastly, cognitive psychologists have been studying attention and visual perception for a number of years in an attempt to understand human cognitive functions. Specifically for diagram recognition, Tversky *et al.* have produced a number of publications exploring the roles of demarcations such as arrows in mechanical diagrams [7], the cognitive processes of spatial cognition [28], and the attentional focus that designers give to their own sketches [8].

The work presented in this paper, while related to the previous research described above, differentiates itself by integrating elements from both the cognitive psychology and computer vision communities. This paper proposes a cognitive model for diagram understanding, but aligns itself with the goal of developing a computational infrastructure upon which future research can be based. Unlike much prior cognitive psychology research, our work focuses specifically on the recognition and mental simulation of mechanical diagrams. Yet our work is developed in a broader context than much of the sketch identification or diagram understanding research to date. By providing a common platform from which these fields can work together, our work aims to make the computational emulation of human diagram-

Figure 1. Humans use top-down processing to identify the dalmatian

matic processing achievable.

BACKGROUND: CURRENT HUMAN VISUAL RECOGNITION THEORIES

For decades, biologists, psychologists, and many others have studied the way our brains perceive images. Current understanding groups our abilities into two camps, often called Top-Down processing and Bottom-Up processing. These two processes work together simultaneously to help us recognize images.

Top-Down processing, which is more knowledge-driven, involves using information about context or experience to help identify patterns. An example of top-down processing can be seen in Fig. 1. The individual blots in the picture do not mean much to us, but when viewed in the context of the entire picture we can use them to identify the image of the dalmatian. The efficacy of top-down processing is largely dependent on user experience, and interprets new evidence in the context of what is already known. In contrast, Bottom-Up processing, or *data-driven* processing, uses the aggregate of individual features in certain symbols in order to identify them. This is the same way many modern computer symbol recognizers work, by identifying specific features and associating them with a distinct classification. Within the psychology community there are four common theories to explain bottom-up processing: Template matching, Prototype theory, Feature Analysis, and Recognition by Components [29, 30]. Full coverage of these four areas is beyond the scope of this work, but our user experiences during our studies suggest that a combination of prototype theory and feature analysis is used during the recognition process.

TECHNICAL APPROACH

In order to better understand human diagram recognition, which occurs in a matter of seconds, we first need to slow down the recognition process. By slowing down the diagram recognition process, key insights can be gained about how humans gather and use visual information to identify images. Our approach to slowing down human diagram recognition is based on separating the human ability to concurrently perform top-down and bottom-up processing. This is accomplished through a tablet PC interface that restricts viewing to areas consciously selected by the user. This section will provide a brief overview of how our experiment takes advantage of current theories to slow down human diagram recognition. Lastly, the findings of our experiment will be discussed.

Slowing Recognition Through Human Disabling

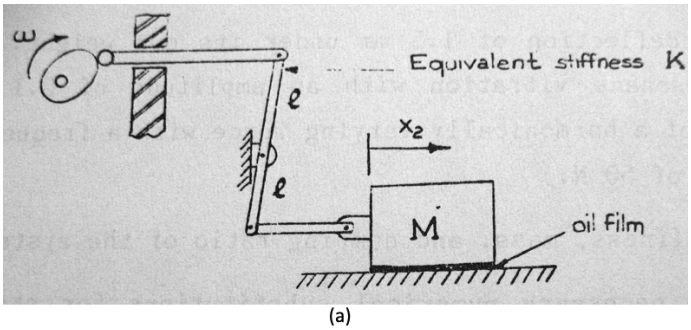
We have devised a combination of two techniques that allow us to control when users are able to conduct bottom-up versus top-down processing. Our approach is analogous to the way that peripheral vision works [29]. In peripheral vision, the eyes can only focus clearly on one specific portion of an image at any one time, and everything outside of a certain radius appears blurred. Our method mimics this phenomenon, forcing the user to consciously select the areas of the image they wish to focus on while limiting their field of view.

To eliminate bottom-up processing, we first blur the entire image. In this way the user can only see rough clusters of points, without the feature detail necessary to perform bottom-up processing. With bottom-up processing temporarily suspended, the user is forced to use top-down processing. In this way, the user attempts to recognize the image using only contextual information and their past experience. An example of such a blurred image can be seen in Fig. 2.

Once the user has extracted as much information as possible from the blurred image, the user can interact with the image using a tablet PC interface by circling or otherwise demarcating a portion of the image that they wish to uncover. An example of a user interacting with our interface can be seen in Fig. 3. This will de-blur the image in the selected portion, allowing access to a small amount of feature information from which the user can conduct bottom-up processing. In order to prevent the user from circling and uncovering the entire image at once, a blur is applied to the selected area in proportion to the area of the selection. The smaller the circled area, the crisper the underlying image will become. This relationship is described in Eqn. 1, and an example of this can be seen in Fig. 2.

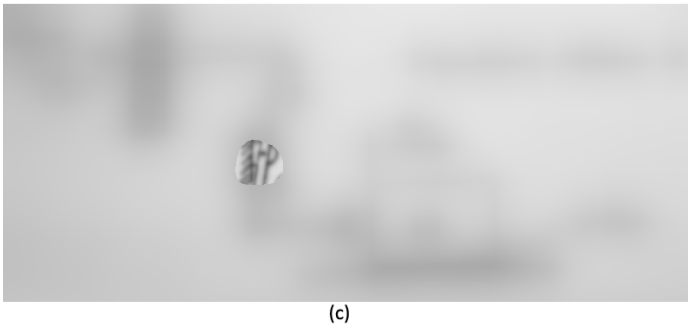
$$Selection\ Blur \propto C * \sqrt{\frac{Area_{selection}}{Area_{image}}} \tag{1}$$

Where $A_{selection}$ is the area of the polygon selected by the user,



(a)

(b)



(c)

Figure 2. (a) The original textbook image (b) The blurred image initially seen by the user (c) Selected area de-blurred by the user

A_{image} is the area of the original image, and C is a blurring constant that can be adjusted by whoever is conducting the experiment. The user can then repeat this process of selection multiple times until the problem is identified. It should be noted, however, that the user is only able to de-blur one section of the image at one time. This means that previously clear areas will be blurred once again when the user selects a new focal area. This choice better mimics the process of shifting attention around an image, and discourages the user from simply uncovering the entire image over time.

Throughout the experiment, as the user begins to uncover more and more information, we ask the user to verbalize their thought processes and decisions, recording them using a built-in microphone. By recording their verbal descriptions along with their pen strokes, we are able to play back the entire session without the use of video recording equipment. The user is also provided with a yellow canvas, as seen in Fig. 10, upon which

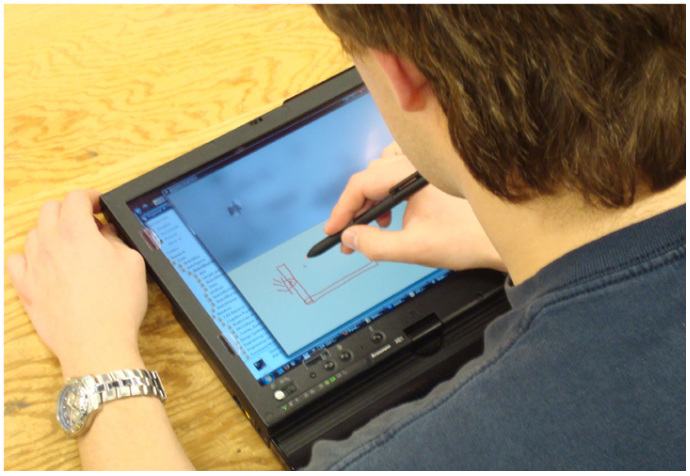


Figure 3. Users interact with the program using tablet interface

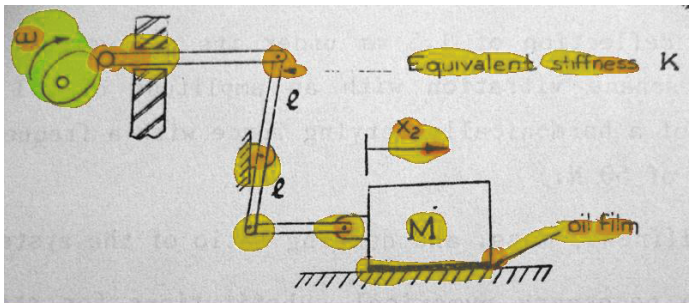


Figure 4. An example heat map showing the areas most important to the user. The user was able to correctly identify the image using only the selected information.

they can sketch or record their hypotheses as they unfold. In addition to recording the user during the session, we also record a final "heat map" which overlays the areas the user selected on the original image. An example of this can be seen in Fig. 4. This allows us to focus on which areas the user thought were the most important for the understanding of the image. In our tests we define "understanding" to be the ability to describe, either orally or visually, the qualitative behavior of the mechanical system. This includes identifying each piece of the image and being able to describe how those pieces interact and change in relation to each other.

Our method of using selective de-blurring of the image allows constant information access to top-down processing, but only select information access to bottom-up processing. This allows us to slow down the process of human diagram recognition enough so that users are able to verbalize their thought processes and decisions more clearly. In the following section, user studies are presented which show some of the results obtained using our method.

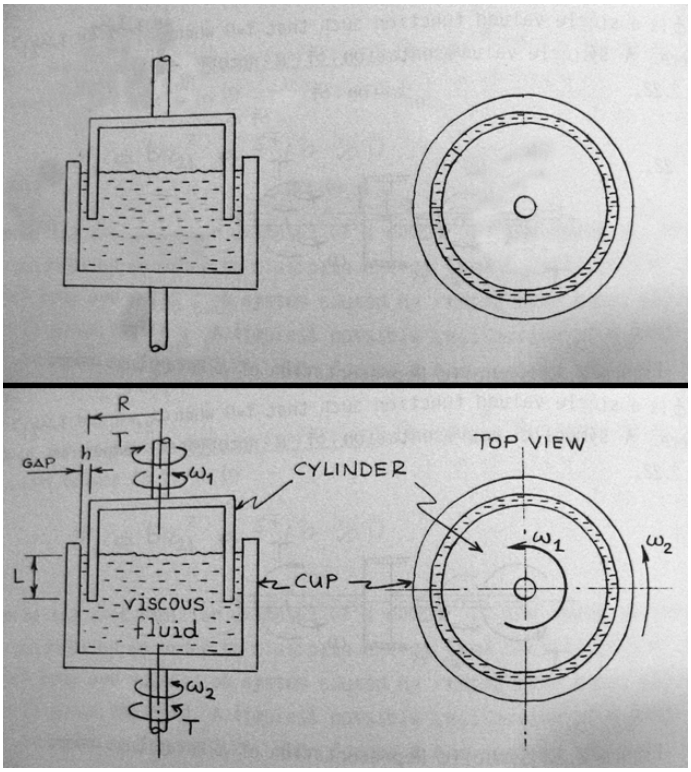


Figure 5. Each image existed with and without labels and demarcations

User Testing

In order to test the efficacy of our method at helping understand human diagram recognition, we tested it on five mechanical engineering students at Carnegie Mellon University. This section will present the implementation of our method, along with the results gained through our five subjects.

Experimental Implementation Our method was implemented using a Java Applet run on a Tablet PC, and used a set of 14 images selected from a variety of Mechanical Engineering textbooks across a number of disciplines [31–35]. The set of images used in this study is included at the end of the paper in Fig. 11. We duplicated the set of 14 images and removed all demarcations, such as labels and arrows, such that only the constituents of the diagram were left. This resulted in a total set of 28 images that were used in the experiment. An example comparison between the sets can be seen in Fig. 5. This set of 28 images was divided into 2 corresponding sets of mutually exclusive images, each set containing 7 images with labels and 7 without labels. The users were assigned one of these image sets and were shown a randomly selected image within that set. We ensured that the users never saw the same image twice, whether labeled or not.

In order to get the users to select only the portions of the image that were of greatest importance, the recognition task was proposed as a game. The object of the game was to get the lowest possible "score," which increased based on the number and size of the circles drawn. In this manner, we discouraged the user from excessively circling parts of the image that were unnecessary for understanding. The users were also given a yellow canvas upon which they could sketch what they thought the system looked like. In addition to helping the user, it also allowed us to roughly capture their mental model throughout the test.

In order to test the new experimental method, five seniors in mechanical engineering were independently evaluated in a controlled environment. Each subject was placed in a room with one of the researchers conducting the study, and only had access to the Tablet PC interface used during the study. Each student was told that they would be shown a set of textbook diagrams that may or may not have labels on them. The student was instructed that the goal was to identify the original image, as well as to describe the qualitative behavior of the objects shown in the image. Students were instructed to verbally describe their thought process after each selection, and that they could elect to use the canvas to draw out their ideas if desired. The was no time-limit imposed on the students, as this might have stifled each student's ability to express his or her thought process verbally. The test was conducted for 1 hour, or until the student had correctly identified 10 images. Each student's pen stroke information was recorded, along with a voice and video recording of the session so that it could be played back for further analysis.

Our user studies, in addition to testing our method's efficacy, were aimed at validating the following hypotheses:

1. Visual labels, such as arrows, letters, or phrases, that are not part of the individual constituents of the image, help us understand images at a much faster rate, with less information.
2. Visual labels help us delineate between cases where the model geometry is ambiguous, such as in cases where bars could be rigid as opposed to flexible.
3. Humans use clusters of points to help us define spatial relationships between subsections of an image and then focus on specific features such as intersections of lines, symmetry, or common symbols to reinforce our hypotheses regarding those spatial relationships.
4. Certain parts of images, such as long continuous lines, are not as essential to image understanding and many humans do not need this information to correctly identify the function of the diagram.

Results After performing the tests on five senior Mechanical Engineering students, several key observations were recorded. These observations later went on to form the basis for our cognitive model of human diagram understanding. Our results indicated that the presence of labels, such as arrows and equations decreased both the time required to identify the dia-

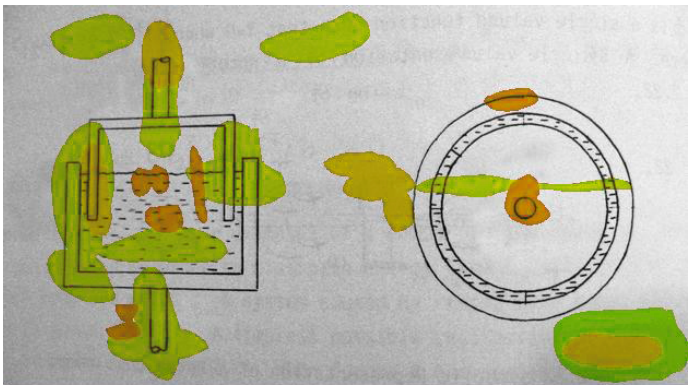


Figure 6. Users tried searching for words to resolve ambiguity

gram and the difficulty of overcoming the ambiguity within the diagram. Other results indicated that the attention of the users was drawn to dense clusters of points that defined "information rich" areas, particularly at interfaces between objects. The phenomenon of confirmation bias was also noted as users developed their final hypothesis.

Our first hypothesis addresses the role of labels, arrows, and other descriptive marks. While statistically significant conclusions cannot be drawn from our sample, we did notice that the lack of labels caused significant problems for students, specifically when identifying the fluid damper seen in Fig. 5. In this case, the student was unable to resolve the ambiguity in the image without labels. As seen in Fig. 6, the user eventually started searching around empty parts of the image, trying to find labels to solidify his understanding of the image. This same example also shows that visual labels help us delineate between cases where the model geometry is ambiguous. In Fig. 6, the lack of labels caused a great deal of frustration for users as they tried to resolve the ambiguity inherent in diagram recognition.

The practice of using dense clusters of points to define areas of focus was used extensively by all participants. Many users based future selections off the amount of data that might be uncovered in a specific region. Areas with interactions or dense populations of points are more "information rich" and can lead to solution convergence faster. This behavior was noticed across all users.

In tandem with the notion that humans uses point clusters as visual anchor points, humans will also avoid paying attention to less "information rich" areas of an image, such as long continuous lines. This is due to the human ability to use gestalt principles to connect or associate things that cannot be seen directly. Users rationalized the entire image without having to see it directly. An example can be seen in Fig. 7, where the user was able to use symmetry and continuation to correctly recognize most of the image without the need to inspect the features. This result implies that not all the feature information contained within an

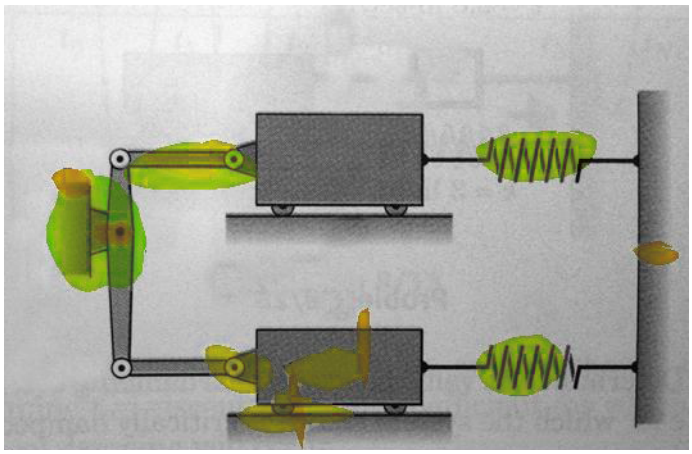


Figure 7. Humans use Gestalt principles, such as symmetry, to facilitate the recognition process

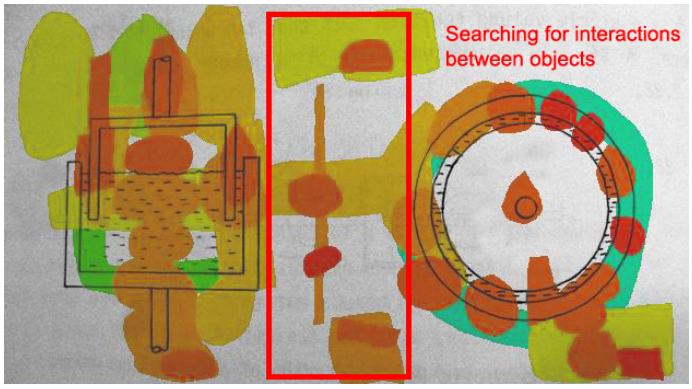


Figure 8. Users searched for interactions between objects to resolve ambiguity

image is needed in order to make a correct identification of the image.

One additional observation noted beyond our listed hypotheses was the importance of interactions between objects in identifying the purpose of the diagram. Users spent the most time observing how objects interacted with one another, and not with identifying the actual objects themselves. The interfaces between objects are information rich, since they dictate the constraints or relations that govern each object. For example, a rectangular block of mass contains only information about the object, while identifying a pin joint between an arm and a mass contains information about the objects themselves, as well as the kinematic constraints that govern them. Interfaces are so important that several users explicitly searched for interactions in order to reduce the ambiguity in the diagram, such as in Fig. 8.

The last major conclusion drawn from our results is the human tendency to fixate on one hypothesis while collecting evi-

dence. Initially, the user has many hypotheses about what the problem might be. However, as the user uncovers parts of the image, they select a "dominant hypothesis," which represents the best potential representation of the problem given what they have seen thus far. This dominant hypothesis is used by the user to simulate the system in their head as they uncover the problem. The user is unable to store multiple mental models in active attention, and as a result the user fixates on one hypothesis in an attempt to prove or disprove it. As the user gathers information from the diagram, the current hypothesis acts as a filter which considers the new evidence in the context of the dominant hypothesis. This behavior is expressed in more detail in the next section where the cognitive model is presented.

In summation, the results of our user testing indicated the following key issues:

1. Labels, such as arrows, words, or equations, substantially improve the recognition process by resolving ambiguity and providing additional, though often redundant, information.
2. Human attention is drawn towards "information rich" areas of a diagram, such as dense clusters of points, or the boundaries between objects. Humans want to maximize the information gained with every shift in attention.
3. Humans use geometric relations, such as parallelism, and gestalt principles, such as smooth continuation, in order to predict areas that cannot be seen directly. Localized assumptions about visual information are used to minimize the need for additional information.
4. Humans can suffer from confirmation bias, often filtering information to support their beliefs. When significant enough cognitive dissonance develops, they reconsider their beliefs.

COGNITIVE MODEL

Following the development and testing of the experimental approach described above, a cognitive model was developed to describe the process that humans go through when recognizing diagrams. The cognitive model presented here can act as the basis for the development of computational tools that emulate the human diagram recognition process. The initial development of the cognitive model was drawn from prior informal observations made when noticing how users interacted with sketches and diagrams, as well as from current visual recognition theories in cognitive psychology [29]. From this initial model, hypotheses were generated that could be tested using the experimental method described above. After performing user testing to validate these hypotheses, the initial model was refined into the current model presented below.

Overview

The cognitive model is broken down into four steps: Gather, Recall, Identify, and Reconcile. The first step, Gather, describes how we focus attention on parts of an image, as well as predict new areas upon which to focus. The Recall step pulls information from our past experience, such as domain knowledge or symbols, to aid in interpreting new information. The Identify step uses feature information gained from steps 1 and 2 to identify symbols or interactions between symbols. Finally, the Reconcile step uses new information to generate, update, and refine mental models, or "hypotheses," about the problem. These steps are repeated cyclically in order to refine a "dominant hypothesis" which is a mental model of the diagram seen by the user. The model itself is presented in Fig. 9, and each step will be discussed in greater detail throughout the coming sections.

Step 1 - Gather

When viewing images, humans are unable to focus on every area of the visual field at once. Instead, humans attend to select areas of an image at one time, shifting attention to different parts of the image as needed. This process has been explored by the cognitive psychology community, and studies have shown that humans actively focus on some parts of images over others, depending on the information they hope to gain from the image [36]. The proposed cognitive model emulates these behaviors in the Gather step, where the human selects an area to focus attention on.

In the model, the dominant hypothesis provides a mental picture of the diagram which humans use to predict where future attention should be focused. For example, take the image and mental model shown in Fig. 10. The human can use his or her mental model to predict the location of new prominent features in the model. Instead of exploring all areas of the image, the human instead predicts the next area to focus on based on which areas will be most likely to either prove or disprove the mental model.

Step 2 - Recall

Once an area has been selected in the Gather step, humans need to retrieve information from their past experiences in order to help identify the new visual information. The cognitive model breaks this step into three areas: A pattern library, an interaction library, and a set of domain schemata. Each of these parts is utilized by the cognitive model to help identify symbols and hypotheses that are consistent with past experiences.

The pattern library is a storage area in memory for all of the past symbols seen by the human. This library stores visual prototypes of common objects such as instantiations of springs, dampers, masses, and pumps, among other symbols. These prototype images can be deformed to match similar images within a diagram. This library is akin to recognition approaches proposed by prototype theory within the cognitive psychology community [29].

In contrast with the pattern library, the interaction library

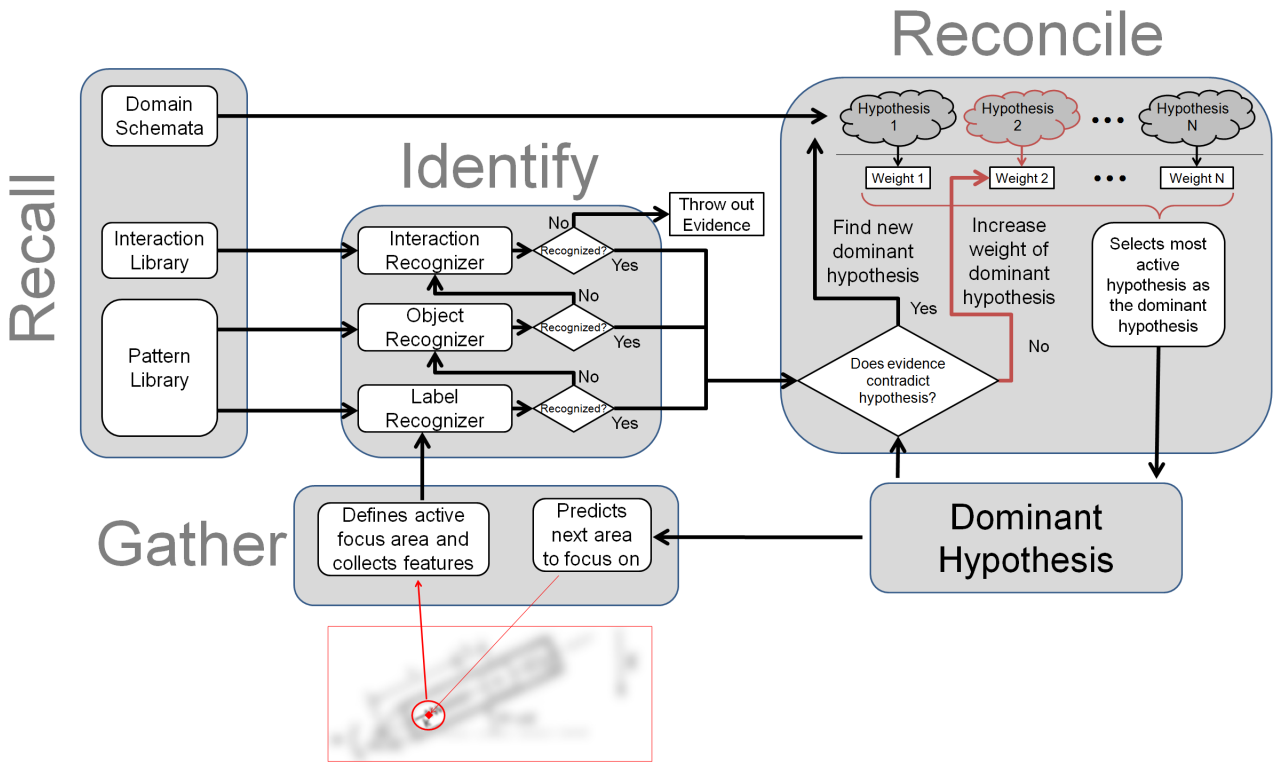


Figure 9. The cognitive model of human diagram understanding

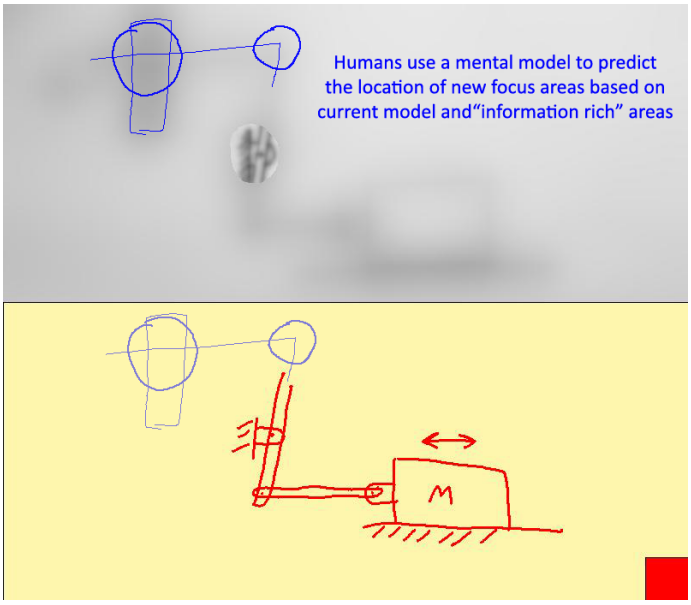


Figure 10. Mental model helps predict new focus areas

stores information about how objects are related or connected with one another. This purpose of the interaction library is to help identify how various objects in a scene are connected together. As an example, springs are known to connect with masses, dampers, springs, and other devices, but are not known to connect with electrical inductors, fluid valves, or resistors. In this way, the interaction library acts as a set of constraints on how symbols are interpreted. Its role in the recognition process is explained in more detail in Step 3.

Finally, the domain schemata is a previous set of rules and assumptions, learned through experience, that govern how elements should behave within a specific domain. For example, the domain schemata for an undergraduate dynamics problem may include:

1. A list of expected elements. Examples include springs, dampers, masses, cams, rope, etc.
2. Initial assumptions regarding those elements. Examples include that springs should be considered massless, each mass should be kinematically constrained to move in 1-2 degrees of freedom, bars should be considered rigid, etc.

The extent of the domain schemata is large and difficult to quantify, so the examples presented do not represent an exhaustive set. Future research will have to be conducted to explore the depth of this area in greater detail. The domain schemata is used to bring identified symbols and interactions together into a cohesive whole. It allows humans to tie in pieces of the diagram

into a mental model that can simulate the behavior of the system. The use of the domain schemata in generating a mental model is explained in greater detail in Step 4.

Step 3 - Identify

The identification step is responsible for taking in visual information and using past experience to correctly identify the meaning of a symbol. This process is two fold.

First, the mind identifies all elements that have the same geometric representation as the current visual input. An example of this step is where a jagged line is interpreted as either a spring, an electrical resistor, or a thermal resistor. In order to accomplish this, information is pulled from the pattern library, discussed in the Recall step. Our research did not investigate the exact cognitive process by which recognition takes place, but it aligns itself well with either prototype theory or feature recognition theories proposed by the cognitive psychology community [29]. These recognition techniques have also been explored computationally, and form a good basis for future development within the model.

Our current model separates the identification step into 3 main components: The label recognizer, object recognizer, and interaction recognizer. The delineation between components is made based on the purpose of different parts of an image, as well as the scope and type of symbols being identified.

Label Recognizer The label recognizer is responsible for identifying elements such as letters, equations, and other labels which are demarcations in the diagram, rather than objects of the diagram themselves. Labels are differentiated from the objects within the diagrams due to both their size and their function within the diagram. Labels such as letters or arrows are often small when compared with the objects they describe, and are spatially segmented from the diagram. A comparison with and without these labels can be seen in Fig. 5.

The ability to locate and identify labels such as letters or arrows allows a gain in knowledge regarding both the domain and behavior of the overall diagram. Arrows often indicate key directions of motion or critical dimensions relative to the qualitative behavior of the diagram. Likewise, letters, equations, and phrases provide similar knowledge regarding how the diagram behaves. While the exact model cannot be determined from labels alone, they are able to quickly define the domain of the problem.

Object Recognizer The object recognizer is where the physical elements of the diagram itself are determined. Similar to the label recognizer, the object recognizer draws information from the pattern library in order to determine whether incoming visual information is part of a set of already known images. If the object is located within the pattern library, it is considered "recognized" and is sent on to be reconciled with the dominant hypothesis.

If only part of an object is seen, such as the corner of a box, the recognition process uses context to help identify the parts of an image that are unseen. In order to do this, geometric relations, such as parallelism and symmetry [14], as well as gestalt principles, such as smooth continuation [16], are used to predict what the entirety of the object might be. If enough evidence exists in support of the prediction, then information is passed to the Reconcile step. If not enough evidence exists, additional information must be obtained from the diagram.

Interaction Recognizer If visual evidence cannot be identified by either the label or object recognizer, it is passed off to the interaction recognizer. The interaction recognizer is different from the previous recognizers in that its focus is not on the objects themselves, but rather the boundaries between objects. This recognizer links objects together in ways that are compatible with the definitions defined in the interaction library of the Recall step.

An example of visual evidence that is processed by the interaction recognizer is the connection between the spring and the wall in Fig. 7. By using the knowledge that the end of the spring connects perpendicularly to the wall, the mind can assume, with higher probability, that the spring operates in a one dimensional fashion in a direction normal to the wall.

Step 4 - Reconcile

The final step of the recognition process occurs when incoming visual information is reconciled with the current mental model. Throughout the recognition process, the mind adjusts an internal model of the problem that includes information about not only the geometric configuration of the parts, but also about the behavioral characteristics of each element. The process by which this model is transformed over time involves the generation of candidate hypotheses, the weighting of these candidates in accordance to current evidence, and the selection of a "dominant hypothesis" which is then used to assist in other parts of the recognition process. Once the first piece of evidence is collected, multiple low-fidelity candidate hypotheses are created. As additional visual information is gained, the mind contrasts the evidence against different hypotheses and gives more support to some hypotheses over others. Eventually, the mind is unable to maintain all candidate hypotheses in the same level of detail, and selects the most probable idea as the dominant hypothesis.

With an influx of new information, the dominant hypothesis can undergo three types of changes: An addition to the model, an object-level change, or a domain-level change. An addition to the model occurs when a new piece of evidence is found which extends the current model, without affecting any pre-existing components of the model. An object-level change occurs when a new

piece of evidence causes a change in one of the existing pieces of the model. An example of this could be finding the word "flexible" next to a bar, and changing the object from a rigid bar, to a bar that can elastically deform. A domain-level change occurs when a piece of evidence is found which alters the fundamental domain of the problem, from a Dynamics to a Thermal Fluid domain for instance. All three changes occur as new evidence is collected, and the mind eventually converges on a mental model that matches the goal problem.

ANALYSIS AND IMPLICATIONS

While the work presented here describes our initial attempts to develop a cognitive model of human diagram recognition, a number of insights remain for further analysis. These issues include the relationship between the proposed cognitive model and the results seen through our user testing, further extensions of our testing method through which the model can be tested, future directions of research that this work allows, and an overview of some of the pitfalls of human recognition that should be avoided when developing computational tools.

Implications for studying human diagram recognition

Though the proposed testing method is designed to slow down the human diagram recognition process without substantially altering it, extensions to our method would allow for additional results not covered in this work. While the selective de-blurring technique does not slow down cognitive processes in all steps, it does slow down the speed at which information can be gathered. While this substantially affects recognition speed, it only minimally affects the overall cognitive process. Our method, being analogous to peripheral vision, tracks the field of view in a way similar to that of modern eye tracking [36]. The difference in our technique is that it establishes a verbal protocol through which the internal mental model can be recorded. Results from an eye tracking study using images similar to our work could further validate this claim.

This work provides a platform upon which many future research directions can be pursued. Two main areas of future research include the validation of the cognitive model through specific experiments designed to test subsections of the model, and the development of a computational model based on the cognitive model presented in this work. In order to increase the accuracy and utility of the current model, further research should be conducted to better understand how hypothesis solutions are generated and maintained over the course of the recognition process. Our current testing method can slow down the gathering of information within the model, but is not best suited for observing how components of the model are assembled into a hypothesis, leaving this area open to future research.

Implications for development of computational tools

The cognitive model was developed specifically to facilitate implementation in a software architecture, and as such represents a serial process. However, it is not clear that the human cognitive process completely follows this serial sequence and evidence has shown that human cognition is capable of both serial and parallel processing [29]. Since our experimental method is not yet able to isolate parallel processing steps, the cognitive model does not yet reflect these attributes of human cognition.

Since our cognitive model is based on our user testing, there exist a number of similarities that emulate common human cognitive processes. User testing revealed that labels, words, and demarcations such as arrows helped reduce ambiguity within the diagram and lead to faster, more confident diagram recognition. Our model takes this into account by designating a separate label recognizer specifically designed to detect these features. Research indicated that humans pay specific attention to how objects interact with one another in order to determine qualitative behavior. As a result, the model includes a recognizer designed to facilitate the recognition of interactions between objects. A curious phenomenon noted in user studies was the presence of confirmation bias, wherein the user would filter and evaluate new evidence only in the context of the dominant hypothesis, rather than taking previous hypotheses into account. For this reason, our model includes the fact that once a dominant hypothesis is selected, new evidence is compared against only the dominant hypothesis. Only when evidence is found to contradict the dominant hypothesis is the hypothesis significantly altered. Lastly, our model takes into account the ability of humans to predict the location of new data, based off the use of geometric relations and gestalt principles seen during user testing.

Besides improving the testing method, our current model can be used to start developing computational tools that emulate human diagram recognition. By taking each component of the model and substituting it with a computational process, it becomes possible to link processes together which emulate human recognition. In so doing, the cognitive model can not only be further validated, but also updated to include additional elements found while implementing a computational equivalent. Much research has already been conducted on certain aspects of the sub components, such as the recognizers [13, 14, 17, 22] and the physical reasoning required to generate qualitative hypotheses [21, 25, 37]. Much of this research could be used to assist in the creation of a computational tool that utilizes our current cognitive model.

Finally, the results of the user studies raised the question of whether the human mind is a good model to base computational recognition systems on. While humans are very good at identifying diagrams, they also suffer from critical issues that modern computational tools may not want to inherit. One of these issues, confirmation bias, increased the amount of time required for recognition in a number of the users tested. Since humans

select a dominant hypothesis as the working model, rather than treating each hypothesis equally, humans start to become blind to potential interpretations of symbols that are not supported in their mental model. In studies this often caused users to pursue a faulty hypothesis until enough evidence accrued to cause cognitive dissonance between what they saw and what they believed. This is caused by the fact that humans lack the working memory capacity that computers have, and are not capable of holding multiple competing hypotheses in memory equally. While the use of confirmation bias reduces the cognitive load on humans, it might increase the rate of false recognitions or decrease the speed at which problems are correctly recognized.

DISCUSSIONS AND CONCLUSIONS

Diagrams and sketches represent a key medium by which people exchange ideas throughout the design process. Developing a robust tool that recognizes diagrams and sketches would allow for substantial advances in both the speed and efficacy of the design process. This is particularly true when applied to communication within design teams that are not collocated. By allowing computational tools the ability not only to transfer visual information, but to also interpret and analyze it, these tools can be used as additive members within the visual design process. Even given the prevalence of computer aided design tools within the engineering design community, the need still exists to develop tools that are capable of interacting with visual information along with their human counterparts.

Despite the need of these tools, several challenges exist that have hindered the development of diagram recognition tools thus far. The most prevalent of these challenges is the inherent ambiguity in the symbol nomenclature used across multiple domains, along with isolating and identifying those symbols. However, humans have little to no trouble identifying diagrams quickly and easily. By understanding how humans are able to identify diagrams, much insight can be gained in how to develop better computational tools.

We have proposed a new technique to slow down the human diagram recognition process. The technique uses dynamic blurring of a target image to separate the top-down and bottom-up processing that humans normally perform simultaneously. By having humans select areas of interest, we can create a step wise record of how their internal hypothesis unfolds. This method can be used on any type of PC, on any set of images, and is applicable to a wide variety of domains. Since this method was only tested on a relatively small set of five users the true significance of these findings is not known. Despite the small test size, insights gained from these tests are still valuable for developing tools that aid in both single and multi-user sketch recognition.

Our user studies utilizing our testing method lead to the development of a cognitive model of human diagram understanding. Our cognitive model breaks the human diagram recognition process down into four main steps. The first step, Gather, predicts where to focus attention for new information and collects low-level visual information. The second step, Recall, pulls relevant information from past experience to assist in the recognition process. The third step, Identify, uses past experience and low-level visual information to recognize both objects as well as interfaces between objects. The last step, Reconcile, updates an internal mental model using newly identified information in order to produce the Dominant Hypothesis. The Dominant Hypothesis is an abstract internal model of the diagram that can be simulated by the mind in order to assess both the geometric and behavioral qualities of the diagram.

Since our test subject pool was limited to five seniors in Mechanical Engineering, our results are not readily generalizable to non-experts. The studies assumed that the subjects already had prior experience with common labels and graphical elements in Mechanical Engineering. The authors are currently pursuing a wider subject pool across different ages and academic fields in order to study how experience changes these results.

This work leads to many new areas of open study. In particular, our testing method allows for study of the human diagram recognition process. This method can be supplemented with additional experiments to further understand how humans generate internal models of diagrams. Our cognitive model provides a platform upon which a computational structure can be constructed to mirror human diagram recognition. While emulating some aspects of human behavior, such as confirmation bias, may not be desired, we believe the strategy of modeling computational tools on human behavior is still fundamentally sound.

REFERENCES

[1] Gross, M., 1996. "The electronic cocktail napkin: a computational environment for working with design diagrams". *Design Studies,* **17**(1), January, pp. 53–70.

[2] Henderson, K., 1999. *On Line and On Paper: Visual Representations, Visual Culture, and Computer Graphics in Design Engineering*. MIT Press, Cambridge, MA.

[3] Ullman, D. G., Wood, S., and Craig, D., 1990. "The importance of drawing in the mechanical design process". *Computer and Graphics,* **14**(2), pp. 263–274.

[4] Shah, J. J., Vargas-Hernandez, N., Summers, J. D., and Kulkarni, S., 2001. "Collaborative sketching (c-sketch)–an idea generation technique for engineering design.". *Journal of Creative Behavior,* **35**(3), pp. 168–198.

[5] Mckoy, F. L., Vargas-Hernández, N., Summers, J. D., and Shah, J. J., 2001. "Influence of design representation on effectiveness of idea generation". In ASME 2001 Design Engineering Technical Conferences and Computers and Information in Engineering Conference.

[6] Schütze, M., Sachse, P., and Römer, A., 2003. "Support

value of sketching in the design process". *Research in Engineering Design,* **14**(2), May, pp. 89–97.

[7] Tversky, B., 2005. *Visuospatial Reasoning*. Cambridge University Press, ch. 10, pp. 209–240.

[8] Suwa, M., and Tversky, B., 1997. "How do designers shift their focus of attention in their own sketches?". In Reasoning with Diagrammatic Representations: Papers from the 1997 AAAI Spring Symposium, M. Anderson, B. Meyer, and P. Olivier, eds., pp. 102–108.

[9] Yang, M. C., 2003. "Concept generation and sketching: Correlations with design outcome". In Proceedings of 2003 ASME Design Engineering Technical Conferences.

[10] Yang, M. C., and Cham, J. G., 2007. "An analysis of sketching skill and its role in early stage engineering design". *Journal of Mechanical Design,* **129**(5), pp. 476–482.

[11] Song, S., and Agogino, A. M., 2004. "Insights on designers sketching activities in new product design teams". In ASME Design Engineering Technical Conferences and Computers and Information in Engineering Conference.

[12] Chusilp, P., and Jin, Y., 2006. "Impact of mental iteration on concept generation". *Journal of Mechanical Design,* **128**(1), pp. 14–25.

[13] Hou, S., and Ramani, K., 2006. "Sketch-based 3d engineering part class browsing and retrieval". In EUROGRAPHICS Workshop on Sketch-Based Interfaces and Modeling.

[14] Igarashi, T., Matsuoka, S., Kawachiya, S., and Tanaka, H., 2007. "Interactive beautification: a technique for rapid geometric design". In SIGGRAPH '07: ACM SIGGRAPH 2007 courses, ACM.

[15] Saund, E., and Mahoney, J., 2004. "Perceptual support of diagram creation and editing". In Diagrammatic Representation and Inference, pp. 424–427.

[16] Saund, E., Mahoney, J., Fleet, D., Larner, D., and Lank, E., 2002. "Perceptual organization as a foundation for intelligent sketch editing". In AAAI Spring Symposium on Sketch Understanding, AAAI Technical Report SS-02-08, pp. 118–125.

[17] Kara, L. B., Gennari, L., and Stahovich, T. F., 2008. "A sketch-based tool for analyzing vibratory mechanical systems". *Journal of Mechanical Design,* **130**(10).

[18] Kara, L. B., 2005. "Automatic parsing and recognition of hand-drawn sketches for pen-based computer interfaces". Phd thesis, Carnegie Mellon University.

[19] Futrelle, R. P., 2007. The diagram understanding system - strategies and results. Tech. rep., Northeastern University, May.

[20] Futrelle, R. P., and Nikolakis, N., 1995. "Efficient analysis of complex diagrams using constraint-based parsing". In Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on, Vol. 2, pp. 782–790 vol.2.

[21] Kurtoglu, T., and Stahovich, T. F., 2002. "Interpreting schematic sketches using physical reasoning". In AAAI Spring Symposium on Sketch Understanding, AAAI Technical Report SS-02-08, pp. 78–85.

[22] Alvarado, C., 2003. Dynamically constructed bayesian networks for sketch understanding. Tech. rep., MIT Project Oxygen Student Workshop Abstracts.

[23] Alvarado, C., 2000. "A natural sketching environment: Bringing the computer into early stages of mechanical design". Phd thesis, Massachusetts Institute of Technology.

[24] Alvarado, C., 2000. "A natural sketching environment: Bringing the computer into early stages of mechanical design". Master's thesis, Massachusetts Institute of Technology, May.

[25] Stahovich, T. F., Davis, R., and Shrobe, H., 1998. "Generating multiple new designs from a sketch". *Artificial Intelligence,* **104**(1-2), September, pp. 211–264.

[26] Kara, L. B., and Stahovich, T. F., 2001. "Spatial reasoning about mechanical behaviors". In Proceedings of ASME Design Theory and Methodology Conference.

[27] Kara, L. B., and Stahovich, T. F., 2002. "Causal reasoning using geometric analysis". In Artificial Intelligence for Engineering Design, Analysis and Manufacturing, Vol. 16, pp. 363–384.

[28] Tversky, B., 2000. "Levels and structure of spatial knowledge". In *Cognitive Mapping Past, present and future*, R. Kitchin and S. Freundschuh, eds. Routledge, London, pp. 24–43.

[29] Smith, E. E., and Kosslyn, S. M., 2006. *Cognitive Psychology: Mind and Brain*, 1st ed. Prentice Hall.

[30] Biederman, I., 1987. "Recognition-by-components: A theory of human image understanding". *Psychological Review,* **94**, pp. 115–147.

[31] Holman, J. P., 2002. *Heat Transfer*, 9th ed. McGraw-Hill.

[32] Meriam, J. L., and Kraige, L. G., 2007. *Dynamics*, 6th ed. Engineering Mechanics. John Wiley and Sons.

[33] Platin, B. E., Caliskan, M., and Ozguven, H. N., 1991. *Dynamics of Engineering Systems*. METU.

[34] Gere, J. M., 2006. *Mechanics of Materials*, 6th ed. Thomson.

[35] Moran, M. J., and Shapiro, H. N., 2004. *Fundamentals of Engineering Thermodynamics*, 5th ed. John Wiley and Sons.

[36] Yarbus, A. L., 1967. *Eye Movements and Vision*. Plenum Press, New York.

[37] Stahovich, T. F., Davis, R., and Shrobe, H., 1993. An ontology of mechanical devices. working notes, reasoning about function, aaai-93.

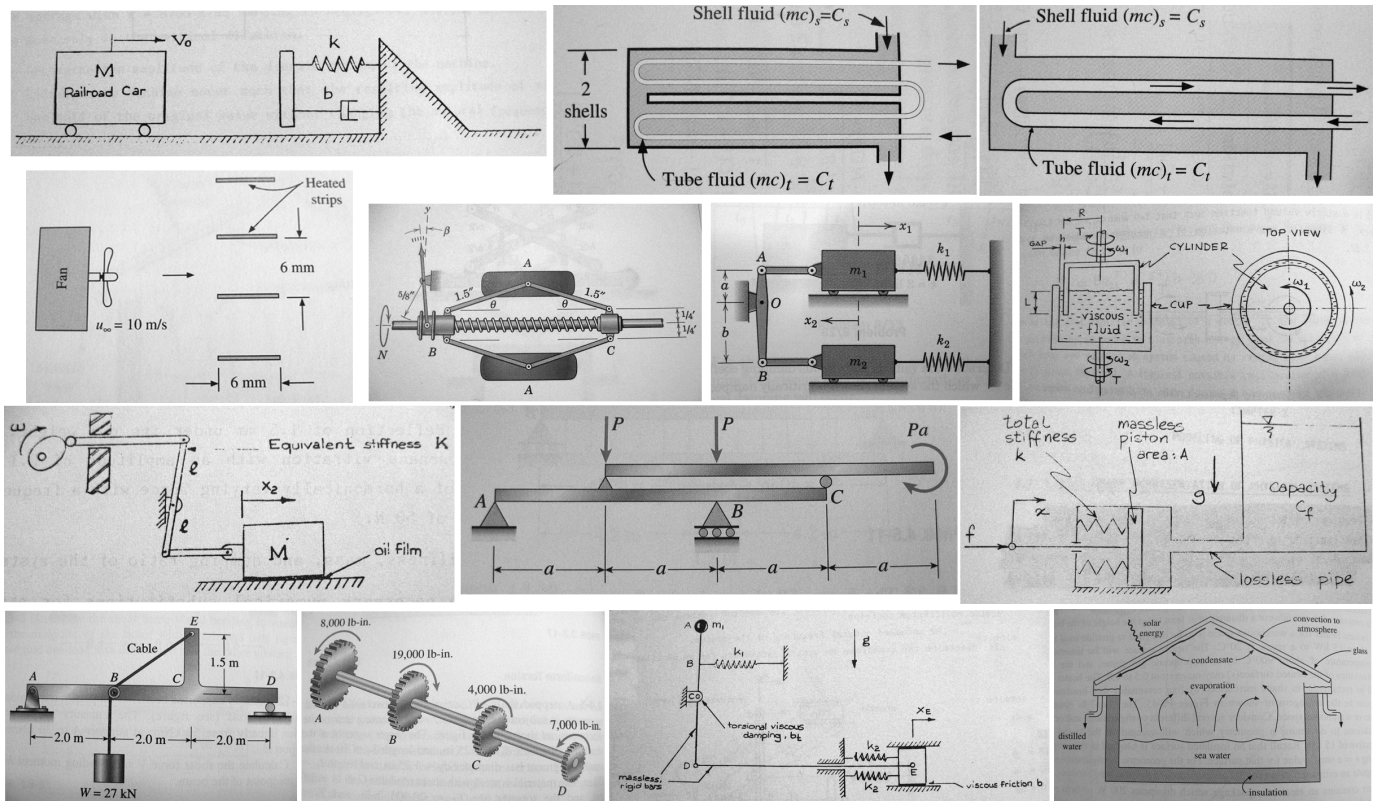


Figure 11. The 14 images used during our study. Images taken from [31–35].