Alexei A. Efros

MSR-CMU Project Proposal

PI: Alexei (Alyosha) Efros
Project Theme:
Building the Memex: Organizing (Visual) Information without Categories

Project Context:

Categories (taxonomies, anthologies, etc) has been with us since the time of Plato and Aristotle, and are still the central paradigm for representing knowledge, allowing us to collapse huge amounts of information into human-manageable chunks. However, in the new digital age of infinite storage, could we be "categorizing away" too much, throwing away valuable information? In a physical library, a book can be placed only on one shelf, belong to a single category; in a digital library, it can be on infinitely many shelves at once – its subject matter being defined by what other books it's similar to. This is why on Amazon, nobody uses the build-in subject index. This is also why in the Yahoo! vs. Google battle of categorization vs. association, the latter emerged as a clear winner.

This high-risk, high-reward project aims at reconsidering the entire paradigm of concept representation in the new digital age. The goal is to build a new knowledge structure, based not on categorization, but on association, harking back to the Memex idea proposed by Vannevar Bush. While at first we will focus on *visual* information (because no good categorical representations exist for it), this approach, if successful, can have huge implications for representing many types of complex natural (non-man-made) information, such as sound and speech, tactile sensing, taste, and many other biological processes where categorical representations have been unsatisfactory. Furthermore, any progress made in this direction could potentially influence the understanding of human categorization in cognitive psychology and philosophy.

Project Description:

The use of categories (classes) to represent concepts (e.g. visual objects) is so prevalent in computer vision and machine learning (and much of modern science) that most researchers don't give it a second thought. Faced with a new task, one simply carves up the solution space into classes (e.g. cars, people, buildings), assigns class labels to training examples and applies one of the many popular classifiers to arrive at a solution. However, we believe that it is worthwhile to re-examine the basic assumption behind categorization.

Theories of categorization date back to the ancient Greeks. Aristotle defined categories as discrete entities characterized by a set of properties shared by all their members. His categories are mutually exclusive, and every member of a category is equal. This classical view is still the most widely accepted way of reasoning about categories and taxonomies in hard sciences. However, as pointed out by Wittgenstein, this is almost certainly not the way most of our everyday concepts work (e.g. what is the set of properties that define the concept "game" and nothing else?). Empirical evidence for typicality (e.g. a robin is a more commonly cited example of "bird" than a chicken) and multiple category memberships (e.g. chicken is both "bird" and "food") further complicate the Aristotelian view.

The ground-breaking work of cognitive psychologist Eleanor Rosch demonstrated that humans do not cut up the world into neat categories defined by shared properties, but instead use *similarity* as the basis of categorization. Her Prototype Theory postulates that an object's class is determined by its similarity to (a set of) prototypes which define each category, allowing for varying degree of membership. Going even further, Exemplar Theory rejects the need for explicit category representation, arguing instead that a concept can be implicitly formed via all its observed instances. This allows for a dynamic definition of categories based on data availability and task (e.g. an object can be a vehicle, a car, a Volvo, or Bob's Volvo).

But it might not be too productive to concentrate on the various categorization theories without considering the final aim – what do we need categories for? One argument is that categorization is a tool to facilitate *knowledge transfer*. E.g. having been attacked once by a tiger, it's critically important to determine if a newly observed object belongs to the tiger category so as to utilize the information from the previous encounter. Note that here recognizing the explicit category is unimportant, as long as the two tigers could be associated with each other. Guided by this intuition and evidence from cognitive neuroscience, Moshe Bar outlined the importance of analogies, associations, and prediction in the human brain. He argues that the goal of visual perception is not to recognize an object in the traditional sense of categorizing it (i.e. asking 'what is this?'), but instead linking the input with an analogous representation in memory (i.e. asking 'what is this *like*?'). Once a novel input is *linked* with analogous representations, associated representations are activated rapidly and predict the representations of what is most likely to occur next.

These ideas regarding analogies, associations, and prediction are surprisingly similar to Vannevar Bush's 1945 concept of the **Memex** – which was seen decades later as pioneering hypertext and the World Wide Web. Concerned with the transmission and accessibility of scientific ideas, Bush faulted the "artificiality of systems of indexing" and proposed the Memory Extender (Memex), a physical device which would help find information based on association instead of strict categorical indexing. The associative links were to be entered manually by the user and could be of several different types. Chains of links would form into longer "associative trails" creating new narratives in the concept space. For Bush "the process of tying two items together is the important thing."

This project proposes an entirely new framework for representing visual knowledge – we called it the *Visual Memex.* Our starting point is Bush's observation that strict categorical indexing of concepts has severe limitations. Abandoning rigid object categories, we embrace Bush's and Bar's belief in the primary role of associations, but unlike Bush, we aim to discover these associations *automatically*, from the data. At the core of our model is an instance-level representation of data, basically storing every object as is, instead of clumping it into a category. The Visual Memex can then be thought of as a vast graph, with nodes representing all the object instances in the dataset, and arcs representing the different types of associations between them, such as visual similarity, functional similarity, contextual dependency, shared attributes (material, geometry, function), even higher-level properties such a geographic and linguistic attributes, etc.

The beauty of this representation is that the act of "recognition" of a novel instance is simply the act of *associating* it with the rest of the memex graph in an optimal way. Instead of naming the new instance (an impossible task, since there are more things in the world than we have names for), we "explain" it by finding globally consistent connections to other, similar instances that have already been observed. As such, Visual Memex could be thought of as a process for taking an unordered heap of visual information, and organizing it in such a way that allows connections, similarities, dependences, and higher-order meaning to emerge naturally, "from the data".

While we have taken some initial, tentative steps in this direction [Malisiewicz09], most central challenges of this approach remain unsolved, making this a high-risk, high-reward project. These challenges span several areas, including computer vision (how to define visual similarity, how to model contextual relationships), theory (how to compute on graphs with multi-colored edges), and systems (how to efficiently process huge amounts of data to build the memex), etc.

Contacts at MSR:

The PI has had close interactions and past collaborations with a number of researchers at MSR, in particular Richard Szeliski's Interactive Visual Media Group in Redmond, and Andrew Blake's Machine Learning and Perception Group in Cambrdige, UK.

Budget Information:

Funding for one student per year plus some summer and travel support is requested for two years. For the first several months, this will support Tomasz Malisiewicz who should be graduating soon, and after that, this will allow PI to hire a new student on this project.