

Project information
CS 395T: Object Recognition
Spring 2007

Proposals

Below are some ideas you could pursue for the final project. However, you may choose to define your own project instead. In either case, your project proposal will need to describe the following points:

- **Summarize** the problem and main idea of the project.
- **Overview relevant related work.** Depending on your project choice, the relevant work is not necessarily limited to papers on our syllabus, so you should do a literature search. Following references in papers we are covering is a good place to start.
- **Technical plan:** What representation(s) and algorithm(s) will you explore? Describe how you will incorporate existing techniques. Or, if you are proposing a new approach, what is the basic idea?
- **Experimental plan:** What experiments will you run to evaluate the idea? What language, libraries, software do you intend to use? Will the experiments show off certain properties of the algorithm? Will they involve a direct comparison with an alternate technique?
- **Sources of data** you will use: Is there an existing image database that is relevant? Will you need to collect new images? If so, how will you do it?
- **Partner plan:** As partners, how will you share the work this project involves? You do not need to write a list of divided tasks. But, you should give evidence that you have discussed how you will work together effectively.
- **Speculate** on what will come out of this project. Do you have a hypothesis about the results? What is most unclear about the project plan at this point?

Proposals are due on March 8. You and your partner should submit one proposal together, and you should both understand and agree with everything described in it.

Project ideas

Learning categories from a keyword-based image search

Current image search engines rely heavily on text to retrieve images: a user provides keywords, and images having that keyword in the filename or in nearby html are candidates for retrieval. Develop a method to learn category models from the images returned from such a keyword-base search (e.g., using Google Image Search). Apply the learned models to recognize novel instances of that category. Your method will have to cope with the fact that retrievals using keywords alone are imperfect and noisy. How will the method determine which examples are “inliers”? Could the text *and* images be complementary? What are the implications for unsupervised category learning?

Some relevant previous work:

- R. Fergus, L. Fei-Fei, P. Perona, and A. Zisserman. Learning Object Categories from Google's Image Search, ICCV 2005.
- R. Fergus, P. Perona, and A. Zisserman. A Visual Category Filter for Google Images, ECCV 2004.

Semantic Robot Vision Challenge

Address the vision problem [above](#), *and* integrate your solution into a mobile robot that can classify objects autonomously in an indoor environment based on what it has learned. This is an ongoing challenge problem sponsored by NSF called the Semantic Robot Vision Challenge.

For this challenge, the robot should be able to: 1) autonomously connect to the Internet and build an object classification database sufficient to identify a number of objects found on a textual list and 2) use this classification database to autonomously search an indoor environment for the objects in its list.

Read more about the competition details and data [here](#) (final rules on the contest will be posted March 1.)

Location recognition

Develop a system that would allow a user to snap a picture of a particular building or scene to find out *what* they are looking at and/or *where* they are. For example, by submitting a picture of the UT Tower, a user could be told that this is the particular building they are looking at. Or, by submitting a view of Chicago's skyline from any of a variety of viewpoints, the city would be identified.

Below are some potentially useful sources of data (or you might choose to collect your own images from the web, or around Austin or campus...)

- The [Where am I? contest](#) held as part of ICCV 2005 posed a location recognition challenge where contestants were given images from a calibrated camera together with GPS locations. The goal is to predict the GPS location of unlabeled images.
- The [Zurich Building Image Database](#) from ETH-Zurich contains 1005 images of buildings in Zurich.

Scene categorization

Design a technique to identify the *category* of a scene. The method might distinguish between different types of rooms, for example (bedroom, family room, kitchen), or it might distinguish highways from forests from mountains.

Some relevant previous work, and data set references therein:

- A. Oliva and A. Torralba. Modeling the shape of the scene: a holistic representation of the spatial envelope, IJCV 2001.
- J. Vogel and B. Schiele. A semantic typicality measure for natural scene categorization. In DAGM'04 Annual Pattern Recognition Symposium, Tuebingen, Germany, 2004.
- M. Szummer and R. Picard. Indoor-outdoor image classification. In Workshop on Content-based Access of Image and Video Databases, Bombay, India, 1998.
- L. Fei-Fei and P. Perona, A Bayesian Hierarchical Model for Learning Natural Scene Categories, CVPR 2005.

Detecting copyright infringement in videos

Video sharing has exploded on the internet recently. Could a vision system automatically find cases where copyrights are being violated? Develop a system to detect duplicate videos. Videos that have the same underlying content can still vary across instantiations, and a robust duplication detection technique will need to ignore irrelevant changes to the actual video frames to determine what's similar.

Celebrity look-alikes

Design a system that can take a photo of a face and determine the celebrity who looks most similar. (Check out the system from myheritage.com.) Use face detection techniques to automatically locate a face in a cluttered image, then apply the search.

Face recognition: inspiration from the human visual system

Explore the computational parallels for the properties of the human visual system for face recognition as summarized in the paper we read by Sinha and colleagues. Choose one or two of those points and design experiments that will provide a computational interpretation of the result(s).

Object recognition challenges

Benchmark data sets can be a useful way to compare different algorithms. Implement an existing object recognition technique(s) and run experiments with a benchmark database(s). Are there extensions that would lead to stronger performance? What role do feature choices play? If you work with multiple kinds of data sets (e.g., scenes vs. objects, specific recognition vs. categories), do you observe trade-offs between different types of approaches? Analyze the results.

- [Caltech-256 web page](#)
- [Caltech-101 web page](#)
- [Pascal Visual Object Classes Challenge](#)
- [GRAZ-02 database](#)

Efficient distance computation via embeddings

Specialized distance or similarity measures can provide a meaningful way to compare image and shape representations, particularly when the representation has a particular structure (i.e., is non-vector). BoostMap is a boosting-based technique by Athitsos and colleagues that we are reading about in class. BoostMap learns an embedding to map a computationally complex distance to a vector space where distances can be computed efficiently. Implement this technique, and apply it to distance functions of interest for recognition, for example, the Earth Mover's Distance, Hausdorff distance, shape context matching, Bhattacharya's affinity, dynamic time warping, or others. Evaluate the results on appropriate dataset(s).

References:

- V. Athitsos, J. Alon, and S. Sclaroff. Efficient Nearest Neighbor Classification Using a Cascade of Approximate Similarity Measures, CVPR 2005.
- V. Athitsos, J. Alon, S. Sclaroff, and G. Kollios. BoostMap: A Method for Efficient Approximate Similarity Rankings, CVPR 2004.

Low-maintenance photo organization

Develop a system to organize a user's personal photo collection with little or no interaction required from the user. In this setting, what cues could be useful beyond the image content (e.g., time stamps)? Think about which dimensions a user might wish to sort the photos—this could lead you to consider distinct recognition problems. For example, can we pull all images of Aunt Linda? Or any group shots that have most of the family in them? All images taken at the beach? Images that together could form a panorama?

Note that for this project a graphical user-interface is unnecessary; if you do choose build one, wait until the vision components are entirely completed.

Names and faces in the news

Implement the approach of Berg et al. for finding faces in images and associating them with the appropriate names from caption text. Evaluate your results, and consider modifications or extensions to the method.

References:

- T. Berg, A. Berg, J. Edwards, M. Maire, R. White, Y. Teh, E. Learned-Miller, D. Forsyth. Names and Faces in the News. CVPR 2004.
- T. Berg, A. Berg, J. Edwards, D. Forsyth. Who's in the Picture. NIPS 2004

Unsupervised category learning (I)

Implement the unsupervised category learning technique developed by Russell et al, and consider modifications or extensions to the method.

Reference:

- Russell, B. C. , Efros, A. A. , Sivic, J. , Freeman, W. T. and Zisserman, A. Using Multiple Segmentations to Discover Objects and their Extent in Image Collections. CVPR 2006

Unsupervised category learning (II)

Implement the unsupervised category learning technique developed by Grauman and Darrell, and consider modifications or extensions.

Reference:

- K. Grauman and T. Darrell. Unsupervised Learning of Categories from Sets of Partially Matching Image Features. CVPR 2006

Automatic cast listing for movies or TV shows

Develop a system that takes a movie and automatically determines the key actors and actresses and labels them each time they appear in the show. In addition to the visual cues, consider also exploiting information from the subtitle text, as is done in Everingham et al.

Some relevant references:

- W. Fitzgibbon and A. Zisserman. On affine invariant clustering and automatic cast listing in movies. ECCV 2002.
- Sivic, J. and Zisserman, A. Video Data Mining Using Configurations of Viewpoint Invariant Regions. CVPR 2004.
- T. Berg, A. Berg, J. Edwards, D. Forsyth. Who's in the Picture. NIPS 2004
- Everingham, M. , Sivic, J. and Zisserman, A. Hello! My name is... Buffy -- Automatic Naming of Characters in TV Video. British Machine Vision Conference, 2006.

Detecting irregularities or unusual objects

A method for detecting unusual objects or regions in images (or similarly, activities in video) could be of great use as a cue for attention, or within surveillance and monitoring applications. The references below address the problem of identifying what is unusual, and their solutions do so without requiring explicit instructions about what *is* normal. Implement one of these techniques, or design your own. Be creative about the dataset with which you evaluate it. Consider modifications or extensions to the existing techniques.

References:

- H. Zhong, J. Shi, and M. Visontai. Detecting unusual activity in video. CVPR04.
- O. Boiman and M. Irani, Detecting Irregularities in Images and in Video. ICCV 2005.
- O. Boiman and M. Irani, Similarity by Composition, NIPS 2006.