

Semantic Robot Vision Challenge (SRVC) Rules 2009

Version 1.7, last updated September 2, 2009

Leagues:

Because not every computer vision practitioner has immediate access to a robot, we offer different leagues in which teams can participate.

- **Software League :**

In this league, the organizers will capture a video log of the environment using a camera mounted on a platform that simulates the view from a mobile robot. The teams are provided with a set of images sampled from this video log. The teams provide their own laptops on which to execute the Internet search phase before receiving the log. During the environment search phase, the laptop will be disconnected from the Internet and their code will be executed on the images in the log file.

While spared from the complexities of integrated robotic systems, software teams also lose some of the advantages. In particular they have no control over the position and orientation of the robot's camera nor do they have access to the intrinsic/extrinsic camera parameters.

Robot League:

Teams provide their own robot hardware and software. The robots will be in the room one at a time. The robots must be able to navigate safely in the indoor setting without damaging the room or the objects within it in the event of an inadvertent collision. They must be able to fit through a doorway and should not damage or mark the floor (or carpet) with their wheels (e.g. skid-steer robots are discouraged). Please contact the organizers to make sure that your robot will qualify.

Teams that participate in this league may use a camera and/or range scanning system to locate and identify the objects. Stereo or monocular vision systems are both allowed so long as the output of the vision system (see below) is of the standard form for scoring purposes.

Objects to find:

A textual list of 20 objects will be given to the agents. This list will consist of both generally-named objects, such as "shoes" or "umbrella", and specifically-named objects, such as a book that has a specific title. To encourage the use of 3D information from images, the generally-named objects are divided further into two groups, according to when they are announced to the teams. Thus, the following three groups of objects are considered:

1. ***Generally-named objects announced in advance:*** Before the competition, the teams will receive a set of 10 names corresponding to general types of objects. At the competition, 5 of these 10 objects will be selected randomly and will appear on the textual list. Teams will not know which 5 of the 10 objects in this category will be selected until the start of the competition. For this category, the organizers will select objects for which proper CAD models are available on the web. Google 3D Warehouse (sketchup.google.com/3dwarehouse/) which is a free, online repository of 3D models will be used as reference for the competition. The teams can use these available models, or they can also create their own 3D models of the selected objects, for example, by scanning such objects and preloading them on their robot hard drive if they wish.

The selected generally-named objects will not have large intra-class variability. However, the models on the web will not be identical with the 3D models of the objects at the competition. For example, "chair" is not a likely candidate object, as chairs can have very different appearance. Possible examples are "banana" or "coffee mug".

The remaining 15 object names will not be known in advance. These objects will be both generally-named and specific but the exact percentages of each type will not be known until the start of the competition:

2. ***Generally-named objects announced at the competition.***
3. ***Specifically-named objects announced at the competition.***

The objects in the environment will be on the furniture and on the floor. Objects on the floor will have a footprint that is larger than 10 cm on a side. Objects found on the furniture might be smaller than this.

The list of objects given to the robot will have 'n' objects on it but not all of those objects will be found in the environment. However, at least half of the objects on the list will be

found in the environment. There will be 'm' objects in the environment where 'm' > 'n'. There will not be duplicate objects.

A few conventions will be applied to the names of objects:

1. Proper nouns will have their first letter capitalized while general nouns will be lower case.
2. Titles of books or names of people or places that refer to specific objects will be placed in quotes. The general name of that object will precede the title. For instance the book "Mary had a Little Lamb" or a poster of "Star Trek" would be listed like:

book "Mary had a Little Lamb"

poster "Star Trek"

3. General objects can have multiple words that describe them. For instance, a blue dry erase marker placed in the environment might be described as:

marker

dry erase marker

blue dry erase marker

where the only caveat is that the description of the object will be grammatically correct (e.g., nothing like *marker blue dry erase*).

Details of the environment:

The competition will be run in an open arena that will be arranged *roughly* like a living room with various pieces of furniture, such as coffee tables, chairs, arm chairs, a sofa and book cases. No guarantees are made as to the exact type of furniture that will be present nor are any guarantees made about the specific positions of that furniture. The arena will be viewable by an external audience. The exact size and space of the arena will depend greatly on the available space at the venue where the competition is held. It will have the following characteristics:

1. No spectators will be allowed in the arena with the robot when it is running. Only the judges will be allowed in the arena. The judge can allow a single team member into the arena to stop the robot if there is a problem.

2. The arena will be delimited by a low bounding wall. However, teams should expect to be able to see spectators outside of the wall. There will be additional clutter and unexpected objects that can be seen beyond the arena. The teams should be ready to filter this out.
3. The floor of the environment will vary from venue to venue and may be simple or complex (a colored carpet).
4. A few pieces of simple furniture (e.g. table, chair, bookshelf, etc...) will be placed in the environment, making it effectively impossible for the robot to stay in one place and view the entire environment.
5. Objects will be placed on the floor as well as on the pieces of furniture.
6. When selecting objects and their placements in the environment, we will attempt to take "context" into account and place objects in expected locations. For instance, books are not typically found on the floor but instead on tables and bookshelves. However, we make no guarantees that all object placements will always make perfect contextual sense.

Running the competition:

The competition consists of two phases: an Internet search phase where the competitor's robots will autonomously search the Internet for information about the objects in question and an environment search phase where the robots will actually search the environment for those objects. Teams must be prepared to surrender their robots and laptops first thing in the morning of the competition to the referees. Intervention with the computers or hardware by the team members is only allowed under the direct supervision of a referee. Teams must use the same computer(s) for the Internet search phase that they use for the environment search phase.

Internet Search Phase:

All groups at the same time will be given a file containing the names of 20 objects. Then the robots (or one of the robots' computers) will be connected to the internet and will autonomously access public-domain databases in order to search for images and other information (such as meta-data, 3D models, or videos) about the objects in question. Teams will be allowed to connect their computers to the Internet for *Two hours* (or potentially more depending on the available connection speed) after receiving the list of objects in order for them to generate their classification database. In this phase, the robots will be unable to see the environment or move

Environment Search Phase:

Once the classification database has been built, the robots will be disconnected from the Internet and will be placed at the entry point to the environment. They will one at a time navigate autonomously through the environment (up to a maximum amount of time) to search for the objects. At the end of the time limit, the robots will be removed from the arena and immediately surrender their answers to the referees. The robot will not be allowed to connect to the Internet during this phase and must instead rely on the results of the search performed in the previous phase.

For the software league, a log captured beforehand by an official competition robot (provided by the organizers) will be provided to the team's computers. This log will consist of a stream of image data as the robot moves through the environment. Because the images will be taken sequentially as the robot moves about, there is no guarantee that a particular image will actually contain any of the objects.

The robots will have **30 minutes** to navigate through the arena to collect visual data, parse through the data that they have obtained, and generate their answers. Once the 30 minutes is up, the robots must be removed from the arena and immediately surrender their answers to the referees. The software league is allowed a total of 30 minutes of time to operate on the log file that they are given.

Audience Visibility

Due to the nature of the event, most of the interesting processing performed by the competitors' computers is not externally visible to the audience. We are very interested in allowing people to see what is going on inside the algorithms as they do their processing. Thus, all competitors will be required to have some sort of graphical display that can be visible on their laptop screens which will also be projected on a wall. One member of the team will be required to narrate to the audience what their algorithm is doing so that everyone watching can understand what the robot or laptop is currently doing.

Ideally the display should be graphical and show the audience what objects have already been discovered by the robot, or else provide some sort of description of what the robot is currently attempting to do. For this effort, robot teams should have an external "monitoring" laptop that can connect to their robot to receive an uplink of data that can be displayed to the audience via an overhead projector or else a large display monitor.

Other mechanisms for providing feedback are encouraged including the use of text-to-speech. The exact specifics of the interface must be included in the team's qualification paper and be approved by the organizers of the event.

Scoring the Competition

At the end of the competition, the robots are required to return at **most** 'n' image files (jpeg or some other common format) on a memory stick, where 'n' is the number of objects specified in the initial list of objects to find.

These image files should have the following naming convention:

objectname.ext

Where *objectname* is the name of one of the objects in the list. This file will consist of one of the images from the log file which has been augmented with a bounding box that clearly identifies the position of the specific *objectname* in the image. Thus, if one of the objects to find is *umbrella*, one of the images would be labeled *umbrella.jpg* and would be a photo taken from the robot's camera where an umbrella in the image is clearly marked with a bounding box. If the name contains characters that cannot be properly reproduced as a filename on the competitor's operating system, a space or underscore character can be substituted for them.

There are several important restrictions that must be followed regarding the contents of the returned images:

1. Only a *single* bounding box can be drawn in an image. If there is more than one bounding box in the image, that image will not be scored. However, if more than one object is found in an image, that same image can be reused for each of the different objects there. Just be sure that each instance of that image has a single bounding box and the filename reflects the unique object that is found.
2. Only a *single* file is allowed for each object. If there are multiple images that depict the same object, then none of those images will be scored.
3. The bounding box must be drawn with lines that are no wider than a single pixel. If the box is wider than this, it risks occluding parts of the object and causing difficulties and inaccuracies with scoring. Images with bounding boxes drawn with lines that are wider than a single pixel will thus be discarded.

4. The bounding box must be drawn with colors that will be visible in all backgrounds. If the judges have difficulty finding the box (for instance if it is a white box on an area of the image that is overexposed), the image will be discarded. Teams may want to select colors that change along the bounding box's perimeter so that the box is clearly visible along all its borders.

Scoring of the images will be performed by a jury consisting of at least 3 members. The scoring criteria will be based on the quality of fit and the correctness of the bounding box over the object in the image. To determine the correctness of each image submitted by the team, the judges will follow the following procedure:

1. The judges will first determine whether the object being recognized is actually in the image.
2. If the object is not present, the image will receive no points.
3. If the object is present, the judges will manually compute a bounding box around the object as seen in the image.
4. The bounding box returned from the team will be manually measured.
5. The match between the true and returned bounding box will be computed as the ratio of (*intersection/union*) of the two bounding boxes.
6. Based on the computed value of this ratio, points will be assigned for each image of a specific object as follows:
 - o score $\geq 75\%$ = 3 points
 - o score $\geq 50\%$ = 2 points
 - o score $\geq 25\%$ = 1 points

Images of **generally-named objects** will be given **twice the score**. That is:

- o score $\geq 75\%$ = 6 points
- o score $\geq 50\%$ = 4 points
- o score $\geq 25\%$ = 2 points

Bonus Point Rule #1: To encourage active exploration and vision techniques, all teams will be given bonus points for actively finding and announcing objects that are discovered in real-time during the run. An image taken from the agent's overhead (public) display that contains the object must be visible on the projected display so that the audience and the referees can see the discovered object. Finally, announced objects will only be worth credit if the team also has a correctly labeled image of that object that will be scored at the end where the bounding box around that object must have a non-zero

ratio of (intersection/union). If all the conditions above are true then the team **receives 1 point** for that object.

Bonus Point Rule #2: We want to encourage the development of efficient implementations of computer vision algorithms as this is highly relevant for real-time robotics applications. After all the points have been tallied, the team that was able to finish the environmental search phase the fastest (*and which had a non-zero image recognition score*) will **receive 3 additional bonus points**. This award will be given to the fastest team in the robotics league as well as to the fastest team in the software league.