WORD: Wiry Object Recognition Database

About

The Wiry Object Recognition Database (WORD) is a collection of labeled images of “wiry” objects such as chairs, ladders, and carts, along with some software utilities for use in. Wiry objects are distinguished by a prevalence of very thin, elongated, stick-like components; examples include tables, chairs, bicycles, and desk lamps. They are difficult to recognize because their shapes are complex and they tend to lack distinctive color or texture characteristics. Recognizing them in images is important in a number of problem areas because they are relatively common.

This database provides benchmark image sets with ground truth for evaluating shape-based object recognition and tracking algorithms. In most image sets, the ground truth consists of binary detected edges, each of which has been hand-labeled as corresponding to an object of interest, or to clutter. In other image sets, the ground truth consists of a set of polygonal regions that map onto the object.

Overall Comments On Images

The stool images were acquired by taking videos on a handheld Sony camcorder; all other images were acquired by taking photographs on an Olympus digital camera. All images were lit by fluorescent light, except the bike set, which was lit by a combination of fluorescent and natural light. The target objects, clutter objects, and camera rotate and translate between images; however, we did not carefully measure exact 3D poses for any object in any image since the goal of these data sets was to quickly acquire large sets of images which overall span a wide range of object and scene variation. Scale of the target object does not vary significantly in any images except those of the stool. We detected binary edges in the scene using the Vista line finder; edge detection parameters, where appropriate, are listed.

File Formats and Directory Structure

Each image set is in a directory with two subdirectories: `images/` contains the raw images and `corrs/` contains files with the label information for the images. Additionally, image sets captured in front of a bluescreen (chair, redchair, and cart) have a third subdirectory, `fgimages/`, which has the raw images before the background was segmented out. For all image sets except the stool, the label files are ASCII lists of the line segments detected by vista. The format for a line segment running from pixel `(<x_begin>,<y_begin>)` to pixel `(<x_end>,<y_end>)` in the image is:

`<class> <x_begin> <y_begin> <class> <x_end> <y_end>`

`<class>=1` if the line segment projects onto the object of interest, and `<class>=-1` if it projects to any other part of the image: clutter objects, background elements, etc. Line segments are separated by newlines.

For the stool images, the `corrs/` directory contains a mask JPEG for each image in the `images/` directory. Pixels at, or near, 255 project onto the object of interest.

Chair

Example image

Example image with labeled edges

Number Of Images: 116
Image Size: 1024 x 768
Comments: This image set contains images of an office chair in front of a bluescreen. The images span the full revolution of the chair in the plane parallel to the floor. Across all images, the elevation of the camera varies by approximately 25 degrees with respect to the object, and the extent of scale variation across images is about 10%. The height of the camera varies between 1.5 m and 2 m and the distance from the camera to the object is around 2-3 m. The background was segmented using the fill tool in Adobe Photoshop. The chair is in the lower left in all images to facilitate compositing with the clutter A, clutter B, and clutter C images.
Vista Parameters: `(sigma, lowThreshold, highThreshold,minimumEdgeLength, accuracy, granularity, magnitude) = (3,10,35,50,2.0,4.0,20.0)`

Red Chair

Example image

Example image with labeled edges

Number Of Images: 40
Image Size: 1024 x 768
Comments: The red chair and the other chair are made by the same company and are very similar. The main differences are in the color of the cushions and the lack of bars along the ground connecting the front legs to the back legs. The two chair image sets could be useful in investigating how recognition algorithms handle minor variations in object shape (*i.e.* narrow "classes" of target objects). Imaging conditions are the same as those for the other chair image set.
Vista Parameters: `(sigma, lowThreshold, highThreshold,minimumEdgeLength, accuracy, granularity, magnitude) = (3,10,35,50,2.0,4.0,20.0)`

Cart

Example image

Example image with labeled edges

Number Of Images: 174
Image Size: 1024 x 768
Comments: This image set shows a 4-wheeled push cart at arbitrary orientations with respect to the camera. The images span the full revolution of the cart in the plane parallel to the floor. Across all images, the elevation of the camera varies by approximately 25 degrees with respect to the object, and the extent of scale variation across images is about 10%.

The height of the camera varies between 1.5 m and 2 m and the distance from the camera to the object is around 2-3 m. The background was segmented using the fill tool in Adobe Photoshop. The cart is in the lower left in all images to facilitate compositing with the clutter A, clutter B, and clutter C images.
Most edges detected by vista are from occluding contours, however some of the edges are generated by patterns of rust on the object (see above example image). In some images, the cart wheels are in deep shadow and are poorly segmented.

Vista Parameters: `(sigma, lowThreshold, highThreshold,minimumEdgeLength, accuracy, granularity, magnitude) = (3,10,35,50,2.0,4.0,20.0)`

Ladder


Example image


Example image with labeled edges

Number Of Images: Total of 1159 images taken in 7 rooms: classroom (170), conference room (183), cubicle (174), apartment (174), kitchen (163), lab (123), warehouse (171)
Image Size: 1600 x 1200
Comments: This is a large set of real images of a ladder at arbitrary out-of-image-plane rotation in 7 different prototypical environments : classroom, conference room, office, lab, apartment, warehouse, kitchen (the example image shows a typical “classroom” image). For each image, the camera was approximately 3m away from the objects in the scene; the elevation of the camera varied between 1.6m and 1.75m; the camera position across the set of all images of a particular scene covered about 60 degrees of rotation with respect to the scene objects in the plane parallel to the floor. The camera was moved between each view, and once every five views the ladder was rotated to an arbitrary angle with respect to the ground and the poses and configurations of clutter objects were randomly modified. The depth of the ladder with respect to the camera varied by a total of approximately 20% across all views. This data set provides a testbed for evaluating how shape-based recognition algorithms fare across a variety of realistic environments.
Vista Parameters: `(sigma, lowThreshold, highThreshold,minimumEdgeLength, accuracy, granularity, magnitude) = (1.5,10,50,25,2.0,4.0,20.0)` (lab, kitchen, conf. room)
`(2,10,50,25,2.0,4.0,20.0)` (cubicle, apartment)
`(3,10,50,25,2.0,4.0,20.0)` (classroom, warehouse)

Bicycle

Example image

Example image with labeled edges

Number Of Images: 151
Image Size: 2048 x 1536
Comments: These images provide a benchmark data set for the challenging problem of recognizing articulated wiry objects. The target object, the bicycle, is viewed at 3 different poses: with the long axis of the bicycle making an angle of 90 degrees, 45 degrees, and 0 degrees (shown in the example image) with the image plane. Moreover, at each bicycle pose, the handlebars rotate to one of 3 different articulations with respect to the rest of the bicycle: the front wheel makes an angle of 0 degrees, 45 degrees, and 90 degrees with the long axis of the bicycle. The camera height for these images was about 1.9 m; the objects are roughly 4-5 m away from the camera; and the camera rotates a total of about 60 degrees w.r.t. the bike. The camera shifts from side to side between each image. Once every 3-5 images the pose and articulation of the bike is modified, and the clutter objects are shuffled. Besides articulation, other challenging aspects of this data set are: the unstable nature of edges extracted from the plants, the sometimes large number of edges detected among the bricks in the background, and the gradual change in lighting over the course of the data set (caused by sunset).
Vista Parameters: `(sigma, lowThreshold, highThreshold,minimumEdgeLength, accuracy, granularity, magnitude) = (3,10,50,50,2.0,4.0,10.0)`

Clutter A


Example image


Example image with labeled edges

Number Of Images: 116
Image Size: 1024 x 768
Comments: This image set shows a background environment consisting of a set of “office” objects– for example lamps, a table, and boxes. The purpose of the image set is to provide a set of clutter scenes to composite with the bluescreened images in the chair, redchair, and cart image sets. The lower left portion of the images is mainly empty so that the bluescreened images can be added in to make realistic-looking views with minimal adjustment. The set of views spans roughly 60 degrees of rotation in the plane parallel to the floor, and variation in scale and camera elevation is about the same as for the cart and chair images. Specifically, the objects are about 2-3 m away from the camera, the height of the camera is between 1.5 m and 2m, and the total variation in object scale across all views is about 20%. To induce appearance variation in the background between views, we modified the poses of each background object and shuffled their relative positions every 5 to 8 images. The camera translates from side to side between views and rotates slightly.
Vista Parameters: `(sigma, lowThreshold, highThreshold,minimumEdgeLength, accuracy, granularity, magnitude) =`

Clutter B


Example image


Example image with labeled edges

Number Of Images: 139

Image Size: 1024 x 768
Comments: Like Clutter A, this image set shows a background environment consisting of a set of “office” objects. The purpose of the image set is to provide a set of clutter scenes to composite with the bluescreened images in the chair, redchair, and cart image sets. The lower left portion of the images is mainly empty so that the bluescreened images can be added in to make realistic-looking views with minimal adjustment. The main difference between Clutter A and Clutter B is that some of the objects are different: in particular the table and monitor are present here. Clutter B could be useful to test how algorithms trained using Clutter A as training data react to changes in clutter objects. The set of views spans roughly 60 degrees of rotation in the plane parallel to the floor, and variation in scale and camera elevation is about the same as for the cart and chair images. Specifically, the objects are about 2-3 m away from the camera, the height of the camera is between 1.5 m and 2m, and the total variation in object scale across all views is about 20%. To induce appearance variation in the background between views, we modified the poses of each background object and shuffled their relative positions every 5 to 8 images. The camera translates from side to side between views and rotates slightly.
Vista Parameters: `(sigma, lowThreshold, highThreshold,minimumEdgeLength, accuracy, granularity, magnitude) =`

Clutter C


Example image


Example image with labeled edges

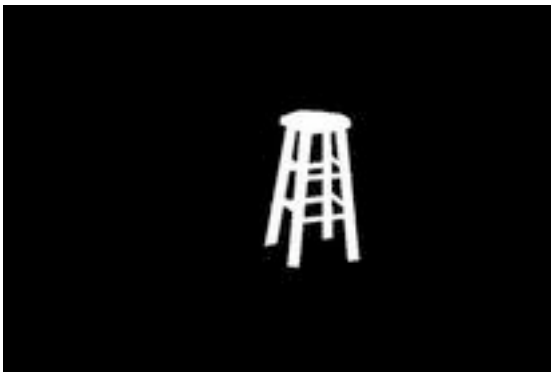
Number Of Images: 26
Image Size: 1024 x 768
Comments: This image set shows a background office environment on the opposite side of the room from where the Clutter A and Clutter B images were captured. The purpose of the image set is to provide a set of clutter scenes to composite with the bluescreened images in the chair, redchair, and cart image sets. Furthermore, this data set could be useful for testing how algorithms trained to account for the environment in Clutter A/B reacts to moderate changes in background. The lower left portion of the images is mainly empty so that the bluescreened images can be added in to make realistic-looking views with minimal adjustment. The set of views spans roughly 60 degrees of rotation in the plane parallel to the floor, and variation in scale and camera elevation is about the same as for the cart and chair images. Specifically, the objects are about 2-3 m away from the camera, the height of the camera is between 1.5 m and 2m, and the total variation in object scale across all views

is about 20%. To induce appearance variation in the background between views, we modified the poses of each background object and shuffled their relative positions every few images. The camera translates from side to side between views and rotates slightly.

Vista Parameters: `(sigma, lowThreshold, highThreshold,minimumEdgeLength, accuracy, granularity, magnitude) =`

Stool

Example image



Example image with labeled edges

Number Of Images: Total of 8201 images taken in 3 rooms: A401 (4423), A408 (2202), SH201 (1574)
Image Size: 360 x 240
Comments: These images are frames from a set of videos taken of a bar stool in a set of office rooms on a handheld Sony camcorder. The location of the stool in each frame is demarcated by a set of polygons. To label the location of the stool over an entire video, we initialized a set of polygons to cover the stool as it appeared in the first frame of the video—one polygon per leg, cross-brace, and seat of the stool. We then applied a 2-frame multiresolution Lucas-Kanade optical flow estimator (`cvCalcOpticalFlowPyrLK` in the OpenCV library) to estimate pixel flows between frames, and estimated polygon motion based on the flows. Polygon positions were manually adjusted at points of tracking failure. We followed this procedure for a total of 29 videos shot in 3 rooms: A401 (20 videos), A408 (3 videos), and SH201 (2 videos). In each video, the cameraman keeps the stool in view and translates and rotates around the stool. This large image set allows the investigation of challenging problems: some of the images are blurry, the stool changes scale significantly, and the change in background between rooms is significant. Besides providing a large data set for wiry object recognition, this image set provides benchmark data for a problem area that (to my knowledge) has yet to be addressed: tracking complex-shaped wiry non-human objects.

Software Tools

Coming soon: A simple program for hand-labeling points and edges in images, and another program for hand-labeling regions in movies. These programs use elements of Qt and the OpenCV library for GUI and image processing functionality.

To Contribute

If you have any images or software tools that would be appropriate to contribute to this database, please email Owen Carmichael (otc AT andrew.cmu.edu). Adding to this community resource would be an immense help in spurring progress in shape-based recognition by allowing researchers to quickly try out their algorithms on a wide range of images. Thanks in advance!

Related Publications

Owen Carmichael . *Discriminative Techniques For The Recognition Of Complex-Shaped Objects*. PhD Thesis, The Robotics Institute, Carnegie Mellon University. September 18, 2003. Technical Report CMU-RI-TR-03-34. [pdf, 7.4 MB]

To Reference

If you make use of any of these images, please include this reference in your publications:

Owen Carmichael and Martial Hebert. *WORD: Wiry Object Recognition Database*. Carnegie Mellon University; http://www.cs.cmu.edu/~owenc/word.htm

Bibtex:

```
@Misc{WordDatabase,
  title =        {WORD: Wiry Object Recognition Database},
  author =       {Owen Carmichael and Martial Hebert},
  howpublished = {www.cs.cmu.edu/\~{}owenc/word},
  month =        {January},
  year =         2004,
  note =         {Carnegie Mellon University}
}
```

Copyright Notice

Contact

Owen Carmichael, otc AT andrew.cmu.edu