Image Database for 3-D Object Recognition

M. Reinhold, Ch. Drexler, and H. Niemann
Computer Science Department, Chair for Pattern Recognition,
University of Erlangen-Nürnberg,
Martensstrasse 3, D-91058 Erlangen, Germany
Phone: +49 9131 8527872, Fax: +49 9131 303811
e-mail: {reinhold,drexler,niemann}@informatik.uni-erlangen.de

Institut für Informatik, Universität Erlangen-Nürnberg

Lehrstuhl für Mustererkennung Technical Report Series

LME-TR-2001-02

May, 2001

Lehrstuhl für Mustererkennung
Institut für Informatik
Universität Erlangen-Nürnberg

TECHNISCHE FAKULTÄT

Abstract

This technical report describes the DIROKOL image data set that consists of 13 real world objects taken from the office and health care domain. The objects were placed on a turntable and the images were taken by a camera that was mounted on a robot arm which could be lifted and lowered. So, the camera could reach every position on a hemisphere above the objects. The background was arranged to be dark from all viewpoints. For each object 3,720 images were taken from viewpoints that were uniformly distributed on the hemisphere; thereby the angle between two adjacent viewpoints was 3°. Three different lighting conditions were applied in such a way that the lighting between adjacent viewpoints was different. For 7 of the 13 objects a second image set exists: the sampling points for these images lie between those of set 1, i. e. 1.5° shifted to the positions of set 1, resulting in 3,600 images at disjunct positions to set 1. The data set is available online via ftp.

1 Introduction

This database was generated for comparing the various object recognition and localization systems that are developed at the Chair of Pattern Recognition of the University Erlangen–Nürnberg and are used in several projects (e. g. Graduate College “3-D Image Analysis and Synthesis”, *DIROKOL*, SFB 603). It consists of objects taken from the office domain, e. g. hole punches and staplers, and the health care domain, e. g. NaCl-bottle and pillbox, in Figure 1 all the objects are shown. We decided to take these objects, because on the one hand in the chair we deal with office scenes and on the other hand in the project *DIROKOL* a hospital environment was investigated. For each of the 13 objects 3,720 color images were taken from viewpoints uniform distributed about a hemisphere with a pose interval of 3°, resulting totally in 48,360 images. For 7 of the 13 objects a second set of images exists taken from other viewpoints as set 1. This second set comprehends 3,600 color images for each object, resulting totally in 25,200 images.

In section 2 we describe the image acquisition. Especially in subsection 2.1 we explain the used devices, in subsection 2.2 the viewpoints and the illumination conditions of the images and in subsection 2.3 the image operations we did to get undistorted quadratic images. Finally in section 3 we present the standard training and test set.

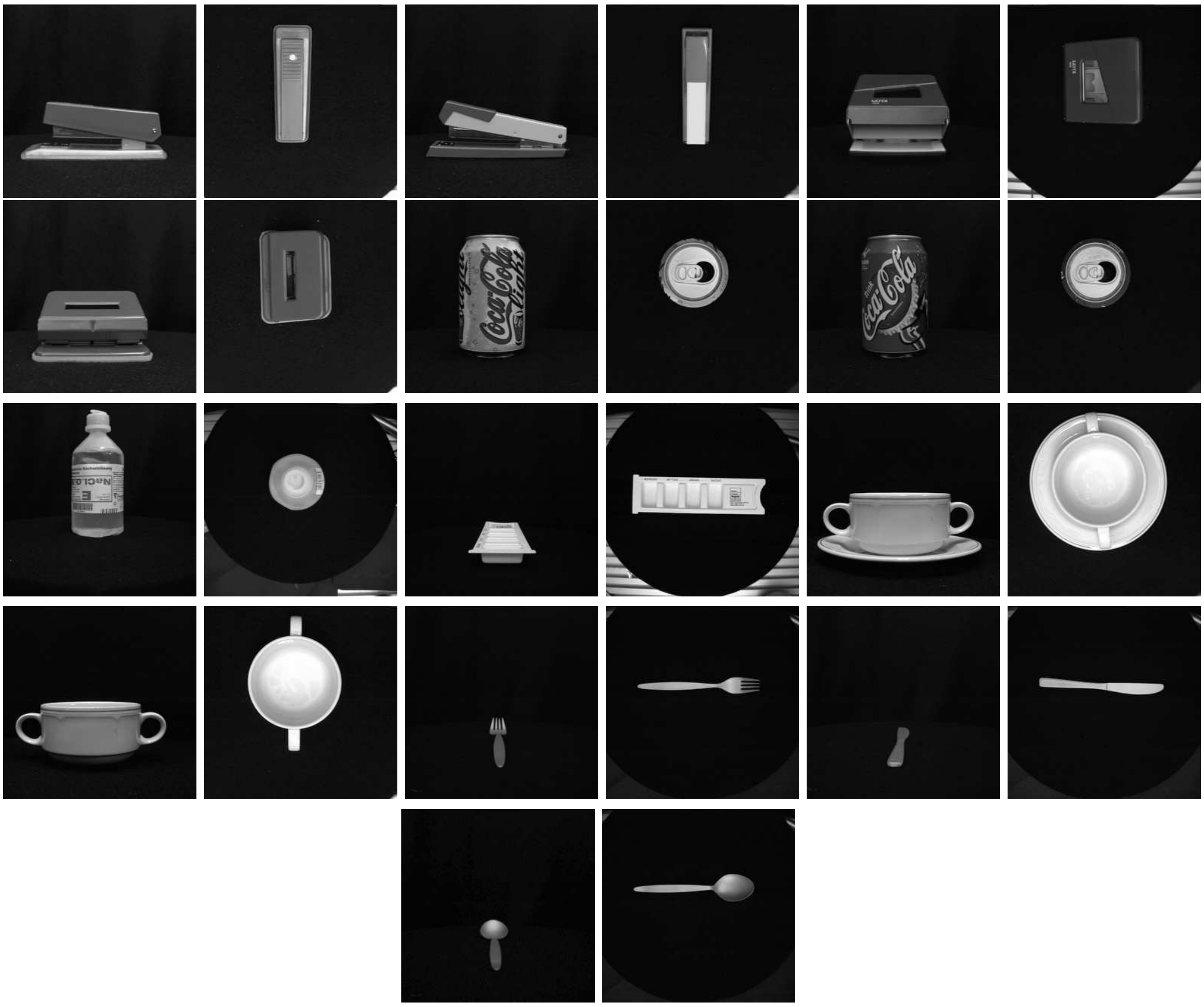

Figure 1: The objects for the experiments, seen from front ($\phi_t = 0°$, $\phi_a = 90°$) and top ($\phi_t = 90°$, $\phi_a = 0°$): on the one hand office tools like the green and the white-green stapler, the red and the green hole punch, the gray and the red can, on the other hand hospital objects like NaCl-bottle, pillbox, cup with and without saucer and cutlery (fork, knife, spoon)

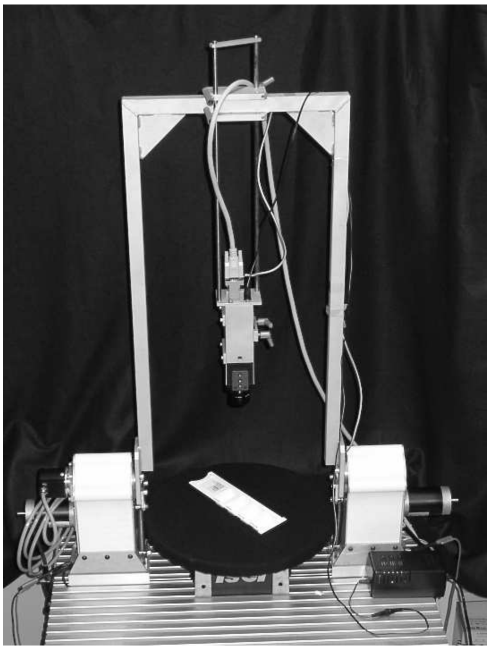

Figure 2: The experimental setup: the turntable and the robot arm with the camera

2 Image Acquisition

2.1 Devices and Cameras

The experimental setup employed for the image acquisition is shown in Figure 2. We used a turntable and a robot arm that are moved by Nanotec step motors and controlled by an isel-CNC-Controller C142-1. Thereby the angles was defined so: ϕ_t is the angle of the turntable, whereby for $\phi_t = 0$ you see the front of the object, and ϕ_a is the angle of the robot arm, whereby for $\phi_a = 0$ the robot arm is at the azimuth. The minimal angle for the step motors is 1.8°, but because of a gearing the turntable can be moved in $\Delta\phi_t = 0.15°$ steps, and the robot arm in $\Delta\phi_a = 0.03°$ steps. The camera that was mounted on the robot arm was a Pulnix TMC-9700 with 2/3" progressive scan interline transfer CCD and a digital RGB output. Thereby the resolution of the camera was $768(h) * 484(v)$ pixels, and a imaging Technology AM-STD-RGB-CSC frame-grabber in a Linux PC was used. Since the size of the objects varies very much and we want that all the objects has approximately the same size (the size of the image), we used two objectives, one with a focus of 8 mm and one with a focus of 12 mm, respectively the object size and varied the distance between the camera and the objects between 25 and 50 cm. Thereby for the same object the distance was fix for all viewpoints.

For the lighting we applied four 500 W halogen headlights that were directed to reflectors so that we had indirect illumination and no strong shadows. For illumination 1 all halogen headlights were switched on, for the other two illumination (2 and 3) respectively one of the four headlights was switched off, for illumination 2 the right one and for illumination 3 the left one so that we had three different lighting conditions. For having a dark background the environment was covered by black cloth. In spite of this some contours of the background like the edge of the turntable could be seen, and at the border of some images the robot arm is visible.

2.2 Database Structure

2.2.1 Initial Setup (Set 1)

For the image acquisition we put each of the 13 objects at the center of the turntable. At the beginning of the image acquisition the robot arm with the camera was vertical above the object. We moved the turntable in steps to $\Delta\phi_t = 3^\circ$ and took at each position an image of the object, resulting in 120 images for one rotation $0^\circ \leq \phi_t \leq 357^\circ$. After every full rotation $\phi_t = 360^\circ$ we lower the robot arm for $\Delta\phi_a = 3^\circ$ and rotate the turntable again. We repeat this, until the robot arm is horizontal, i. e. $\phi_a = 90^\circ$. So the robot arm was lowered in 31 steps ($0^\circ \leq \phi_a \leq 90^\circ$), together with the 120 positions for a full turntable rotation, this leads to 3,720 images per object. Thereby the viewpoints are uniform distributed on a hemisphere and the angle between two adjacent viewpoints is 3°. The lighting condition was chosen so that the illumination of adjacent viewpoints is different, as one can see in Figure 3. For example for $\phi_a = 0$ the illumination sequence is for a rotated turntable: illumination 1, illumination 2, illumination 3, illumination 1, illumination 2, illumination 3 etc. . It is the same for a lowering robot arm.

2.2.2 Extension (Set 2)

For 7 of the 13 objects, i. e. the red hole punch, the red can, the pillbox, the cup, the knife, the fork and the spoon, we took 3,600 images additionally. Also here, we move the turntable and the robot arm in steps $\Delta\phi_t = \Delta\phi_a = 3^\circ$, but the positions of this set 2 were shifted for 1.5° compared to set 1: $1.5^\circ \leq \phi_t \leq 358.5^\circ$ and $1.5^\circ \leq \phi_a \leq 88.5^\circ$. Also the lighting conditions are shifted for $\Delta\phi_t = \Delta\phi_a = 1.5^\circ$ as shown in Figure 3.

2.3 Image Processing

Since the CCD cells of the camera were not quadratic, but $11.6\,\mu m * 14.6\,\mu m$ (data sheet) the images were distorted and had to be scaled. If we would use the information of the data sheet, the images had to be scaled from the original size $768 * 484$ to $655 * 484$ pixels. We tested this

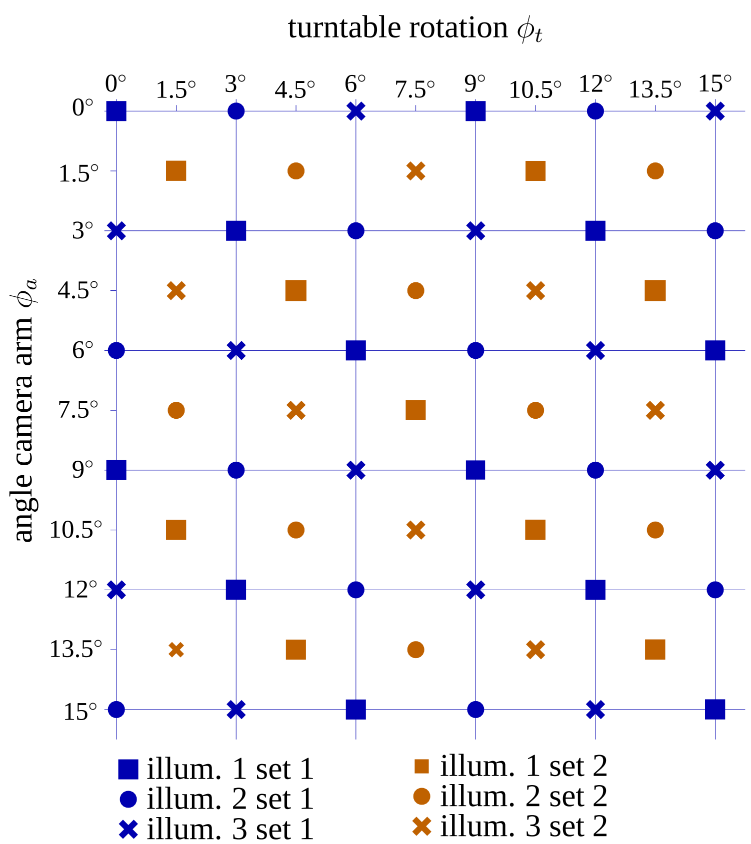

Figure 3: Illumination conditions for the different viewpoints and different image sets

size with a calibration pattern and found out that the right scaling was 623*484 pixels. Since we wanted to reduce the data size, i. e. quadratic images with a size of 256*256 pixels, we had to scale the images further to $330 * 256$ pixels and to cut the images at the right and left border to $256 * 256$ pixels.

We realized that in the following way: we summarized the two scaling to one scaling from $768*484$ to $330*256$ pixels. Thereby we transformed the image data with a Fourier transformation in the frequency domain and cut the respective high frequencies and then we transformed the data back. Afterwards we cut at the left and the right border respectively 37 rows so that the image size was $256 * 256$ pixels. Furthermore we transformed the color images to gray value images.

3 Standard Training and Test Sets

The half of the set 1 is used for the training, the other half for the tests. These sets are partitioned in that way:

- for the training set for $\phi_a = 0°, 6°, 12°, ...$ the images $\phi_t = 0°, 6°, 12°, ...$ and for $\phi_a = 3°, 9°, 15°, ...$ the images $\phi_t = 3°, 9°, 15°, ...$, see file trainDat.
- For the test the other half of set 1, i e. for $\phi_a = 0°, 6°, 12°, ...$ the images $\phi_t = 3°, 9°, 15°, ...$ and for $\phi_a = 3°, 9°, 15°, ...$ the images $\phi_t = 0°, 6°, 12°, ...$, see file testDat.

  Also, there are smaller test sets, whereby the larger sets contain the smaller sets and also in the smaller sets the viewpoints are distributed over the hole range of the external rotations:

  - testDatHalfA, it contains 960 images of testDat.
  - testDatQuarterA, it contains 480 images of testDatHalfA.
  - testDatEightA, it contains 240 images of testDatQuarterA.
  - testDatSixteenA, it contains 120 images of testDatEightA.

Experiments are performed for classification and pose estimation experiments.

Acknowledgement

The work on this data base was partially funded by the Graduate College "3-D Image Analysis and Synthesis" and *DIROKOL*.

4 Appendix A

4.1 Get the dataset

The dataset of the can be downloaded or requested ... The size of all gray value images is 2.08 GB.

4.2 Naming Conventions

The names of the images consist of three parts:
for example NaCl1__0100_1200.gvi

- the name NaCl1, i. e. the name of the object, whereby the 1 means that it belongs to set 1;
- the first number is the number of motor steps of the robot arm, multiplied with the 0.03° it is the angle ϕ_a of the robot arm, here $0100 \mathrel{\widehat{=}} 3°$;
- the second number is number of motor steps of the turn table, multiplied with the 0.15° it is the angle ϕ_t of the turn table, here $1200 \mathrel{\widehat{=}} 180°$;