



The Amsterdam Library of Object Images

JAN-MARK GEUSEBROEK, GERTJAN J. BURGHOUTS AND ARNOLD W.M. SMEULDERS
*ISLA, Informatics Institute, Faculty of Science, University of Amsterdam, Kruislaan 403,
1098 SJ Amsterdam, The Netherlands*
mark@science.uva.nl

Received June 23, 2004; Revised July 29, 2004; Accepted August 5, 2004

First online version published in September, 2004

Abstract. We present the ALOI collection of 1,000 objects recorded under various imaging circumstances. In order to capture the sensory variation in object recordings, we systematically varied viewing angle, illumination angle, and illumination color for each object, and additionally captured wide-baseline stereo images. We recorded over a hundred images of each object, yielding a total of 110,250 images for the collection. These images are made publicly available for scientific research purposes.

Keywords: scientific image collections, sensory information, photometric invariance, geometric invariance, color constancy, wide baseline stereo, image retrieval, appearance modelling, pose estimation, 3D object reconstruction, super resolution

1. Introduction

Kant in his Critique of Pure Reason (Kant, 1781) described the basic principles of human knowledge, made up from a priori and empirical knowledge, the latter originating from sensory information. He argued that for knowing the a priori component in knowledge it is important to isolate the sensory part of the information and its subsequent, immediate denominations. The ability of perception of space must be truly a priori.

Analogously, the cognition of objects is based on the true, a priori characteristics. We need a map of the conditions by which the perception of the object has been transformed by the scene it is in. It should be noted that these a priori characteristics cannot be visible as we never see an object without a scene.

We take Kant's distinction literally by aiming for a proper and eventually rich description of the object in order to attribute the part of the scene. We do not believe in one universal law, nor that completeness is within reach. Rather, we aim for representations of object views independent of the accidental conditions

of the recording. When these factors have been isolated, pure sensory information about the object is what remains.

The accidental conditions of recording can be largely characterized by direction of view, the incident light, the color of the light and the other accidents as the presence of a foreground or background. In order to study the general laws of object appearance in a scene, we need an encyclopedia of views under a large variety of accidental conditions. The most important are viewing direction, illumination direction, and illumination color.

We have recorded a large variety of object shapes, transparencies, albedos, surface covers, and material interfaces to study the general mechanisms of object cognition. We were inspired by the successful Columbia University Object Image Library COIL-20 and COIL-100 database (Nene et al., 1996a, b). The COIL database consists of 100 objects recorded under 72 different viewing angles. Although the database is a successful testing ground for object recognition, it is small, limited in its repertoire of albedo and surface

types, and it is restricted as illumination is kept fixed during recording. Small variation in illumination was added by the Surrey Object Image Library (SOIL) (Koubaroulis et al., 2002), where illumination intensity was changed to generate two recordings of the same object. A collection of 20 objects under various illumination sources is presented by Barnard et al. (2002), representing a small but valuable data set to test color constancy algorithms. Here we present the ALOI (pronounce as alloy) collection. The collection covers a wide variety of illumination and viewing directions, roughly characterizing the Bidirectional Reflectance Distribution Function (BRDF) (Nicodemus et al., 1977; Horn, 1986, p. 209ff) for one-thousand objects. Additionally, we recorded the objects under different illumination colors, and we recorded wide-baseline stereo pairs.

2. Experimental Setup

The setup used to acquire the image collection consists of a rotating stage, a bow with five light sources attached, and three cameras, see Fig. 1. The light sources are mounted at the light bow at -60° , -30° , 0° (top of the bow), 30° , and 60° , enumerated $l1 \dots l5$, respectively. The light beams are directed towards the rotating stage, which is at the heart line of the light bow, but 30 cm behind the bow. The cameras are positioned on the opposite side of the light bow, at a working distance of 125 cm from the heart of the rotating stage. The cameras are at 0° ($c1$), -15° ($c2$) and -30° ($c3$)

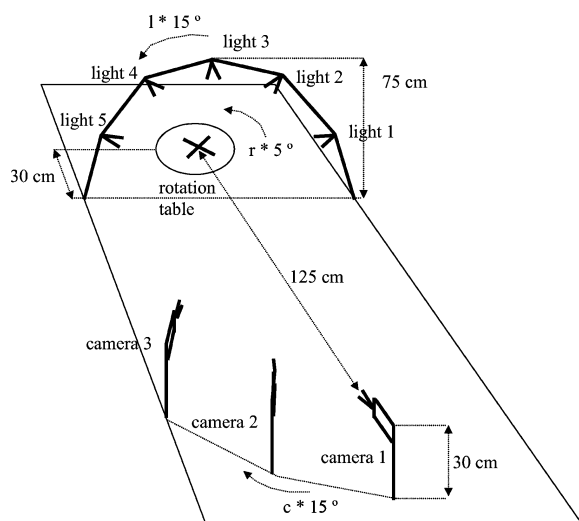


Figure 1. Experimental setup for capturing the collection.

azimuthal viewing angle, 30 cm above the plane of the turntable. Objects are positioned manually at the center of the rotation stage.

The light sources consist of five Osram tungsten halogen lamps (type 64637, 12 V, 100 W), with a color temperature of 3100 K at 12 V. Duration time for setting-up, calibrating, and recording the database (approx. 200 hours) was only a fraction of the average life time of the lamps (specified as 1500 hours), hence aging effects affecting color temperature may be considered minimal. Each of the light sources is connected to an electronic dimmer (Osram HT1-10DIM with Osram HT150/230/12L transformer), which is under computer control via an RS-232 controller (in-house made).

The rotating stage (Parker Hannifin, type 20505RT with power supply P25L, stepper driver L50, and motor SY563T) is set up for 800 steps per revolution, 72 revolutions per cycle of 360° , at a maximum speed of 5.22 rpm. Tolerance is specified to be better than 0.05° . The stage is under computer control via an RS-232 interface (in-house made).

Images are recorded by three Sony triple-CCD color cameras (DXC390P). The camera records 768×576 pixels on a $1/3$ inch CCD. All settings were switched off, yielding linear response between scene irradiance and pixel intensity within the dynamic range of the camera. To increase the sensitivity, needed for capturing the illumination color collection at low lamp voltages, gain was set to 6 dB. Cameras were white balanced once with all lights turned on at 11.76 V (default voltage in all recordings unless noted otherwise). All cameras were controlled by the computer through their RS-232 interface. The cameras were equipped with Computar 12.5–75 mm f1:1.2 zoom lenses. Aperture was closed to f5.6. Zoom was fixed at 48 mm for collecting the first 750 small objects, and at 15 mm for collecting the remaining 250 larger objects. Images are captured by a Matrox Corona-II frame grabber.

Intensity balancing was performed once to compensate for the amount of light viewed by each camera under the various recording conditions. Therefore, we adjusted the integration time (camera “shutter speed”) for each recording circumstance, and for each camera individually. We positioned a grey reference board at the stage, visible to all cameras. We controlled the lights to match the specific recording condition, and tuned the average grey value of the viewed image to be at a pre-defined intensity level ($\pm 2\%$). Hence, the overall image intensity for each recording condition per object is comparable.

3. ALOI Image Collections

We used the setup to acquire four image collections, each designed to a specific field of computer vision. The collection for *illumination direction* ALOI-ILL consists of 24,000 images recorded under varying illumination angle; the *illumination color* collection ALOI-COL consists of 12,000 images for which illumination color temperature was varied; 72,000 images of the objects under in-plane rotation aim to describe *object view* (ALOI-VIEW); and the *wide-baseline stereo* collection ALOI-STEREO consists of 2,250 images. For each collection, we recorded 750 small objects (zoom lens at 48 mm) and 250 larger objects (zoom lens at 15 mm). Every object is annotated by: a number; a textual description identifying the object (e.g. “yellow box toy”); the material(s) it is made from (e.g. “plastic”); whether its staining is uniform or pluriform; and some additional surface properties (e.g. “shiny”, “composite”, “coarse”).

3.1. Illumination Direction Collection

Each object was recorded with only one of the lights turned on, yielding five different illumination angles. By switching the camera, and turning the stage towards that camera, the illumination bow is virtually turned by 15° and 30° , respectively. Hence, the aspect of the objects viewed by each camera is identical, but light direction has shifted by 15° and 30° in azimuth. In total, this results in 15 different illumination angles.

Furthermore, combinations of lights were used to illuminate the object. Turning on two lights at the sides of the object yielded an oblique illumination from right (condition 16) and left (condition 17), respectively. Turning on all lights (condition 18) yields a sort of hemispherical illumination, although restricted to a more narrow illumination sector than true hemisphere. In this way, a total of 24 different illumination conditions were generated (see Fig. 2).

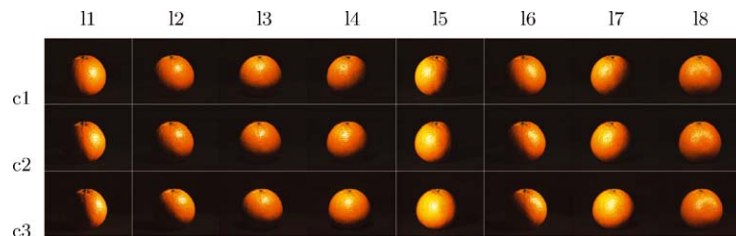


Figure 2. Example object from ALOI-ILL viewed under 24 different illumination directions. Each row shows the recorded view by one of the three cameras. The columns represents the different lighting conditions used to illuminate the object.

3.2. Illumination Color Collection

Each object was recorded in frontal view, with all five lamps turned on. Voltage of the lamps was controlled to be $i/255 * 12$ V, $i \in \{110, 120, 130, 140, 150, 160, 170, 180, 190, 210, 230, 250\}$. Note that changing the voltage of the lamp will affect its emitted spectrum, as halogen acts like a blackbody radiator, and approximately follows

$$T_{\text{eff}} \approx T_{\text{nom}} \left(\frac{U_{\text{run}}}{U_{\text{nom}}} \right)^{0.42} \quad (1)$$

where T_{eff} is the effective color temperature of the lamp running at U_{run} V. Furthermore, U_{nom} , T_{nom} is the nominal voltage and color temperature of the lamp, being 12 V and 3100 K, respectively. Hence, by controlling the running voltage, the illumination color temperature is changed from 2175 to 3075 K. Cameras were white balanced at 3075 K, resulting in objects illuminated under a reddish to white illumination color (see Fig. 3).

3.3. Object View Collection

The frontal camera was used to record 72 aspects of the objects by rotating the object in the plane at 5° resolution (see Fig. 4). This collection is similar to the COIL-20 and COIL-100 collection recorded by Nene et al. (1996a, b).

3.4. Wide-Baseline Stereo Collection

For 750 objects (objects 251–1000), we also captured wide baseline stereo images. By turning the object 15° to view the second camera, a center image, a right image and a left image could be made. The combination of left-center and center-right images yields two pairs of 15° -baseline stereo, whereas the combination of the

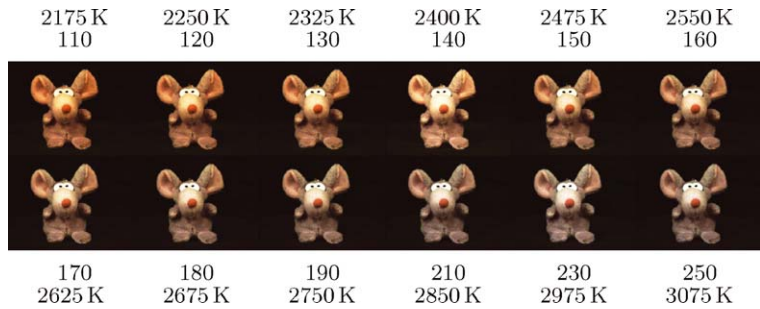


Figure 3. Example object from ALOI-COL viewed under 12 different illumination color temperatures.

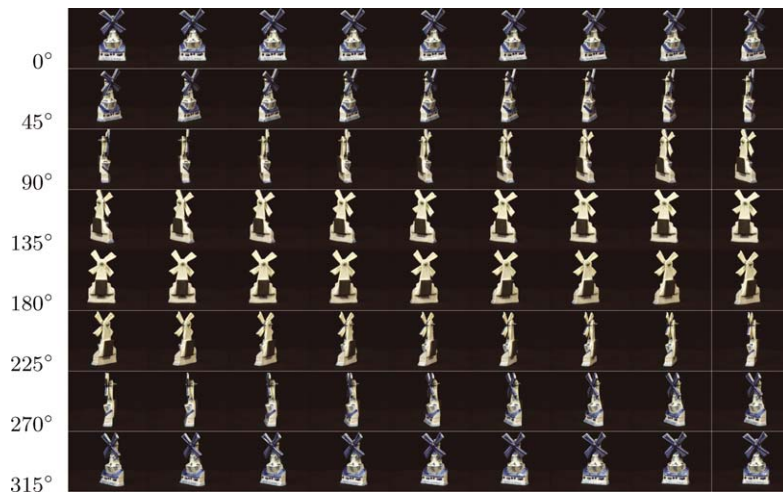


Figure 4. Example object from ALOI-VIEW viewed from 72 different viewing directions.



Figure 5. Example object viewed by 15° and 30° wide-baseline stereo pairs. Collections are denoted by ALOI-STEREO-15L (left-center pair), ALOI-STEREO-15R (center-right pair), and ALOI-STEREO-30 (left-right pair).

left-right pair yields a 30°-baseline stereo image (see Fig. 5).

3.5. Image Downscaling and Grey Conversion

It is unlikely that images will be needed at full resolution for all applications in computer vision. Therefore, we have created two new collections also available by resampling all images by a factor $r = 2$ (yield-

ing 384×288 pixels ALOI-RED2-...) and a factor $r = 4$ (yielding 192×144 pixels, ALOI-RED4-...). To reduce the effects of aliasing, images were smoothed by a Gaussian filter at scale $\sigma = 0.75$ ($r = 2$) and $\sigma = 1.5$ ($r = 4$) before resampling. Furthermore, we have converted all images (including the downsampled versions) to greyscale (ALOI-GREY, ALOI-GREY2, ALOI-GREY4) by calculating intensity $I = 0.21R + 0.72G + 0.07B$ (ITU-R Recommendation BT.709, 1990).

4. Conclusion

We have presented the ALOI collection of one-thousand objects, which we recorded under 72 in-plane viewing angles, 24 different illumination angles, and under 12 illumination colors. Additionally, we have recorded wide-baseline stereo images for 750 of the objects. The ALOI recordings are made publicly available for research purposes.¹ In total, the

collection contains 110,250 images, occupying 140 GB of disk space (uncompressed tiff, 60 GB lossless compressed png). For an overview, see the appendix Fig. 6.

The data set offers a testing and evaluation ground for a variety of computer vision algorithms, amongst others: object recognition, pose estimation, color

constancy, invariant feature extraction, stereo algorithms, super resolution from multiple recordings, and image retrieval systems. Furthermore, the annotated collection enables learning of true object characteristics from images of every day objects, hence, allowing to disambiguate a-priori object information from the empirical sensory information.

Appendix

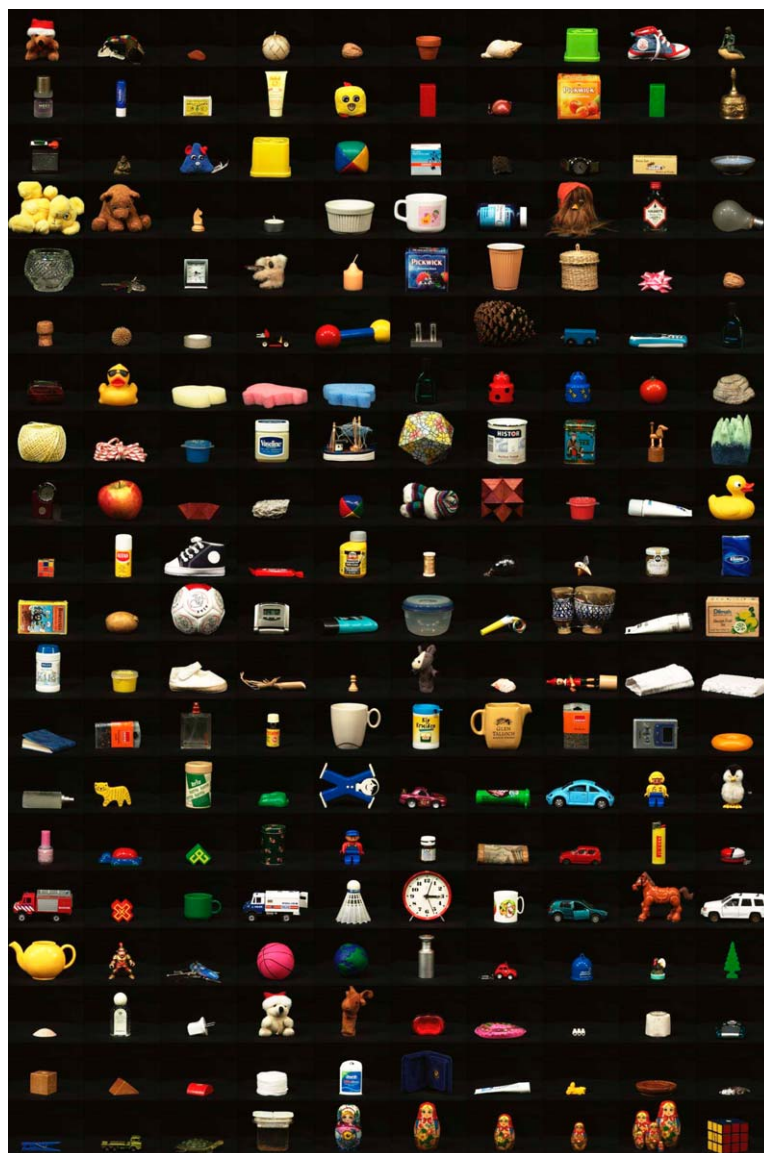


Figure 6. Overview of all objects recorded in the ALOI collection.

(Continued on next page.)

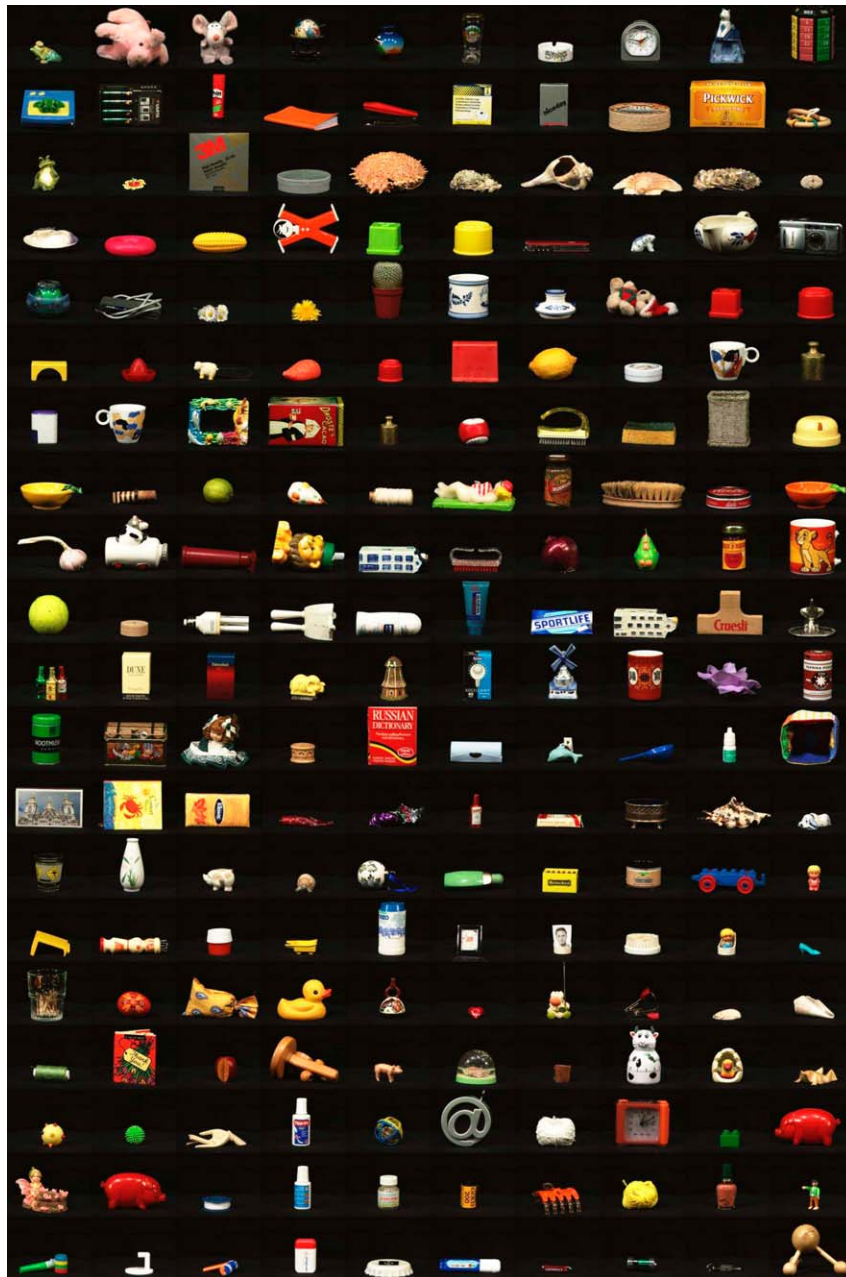


Figure 6. (Continued on next page).

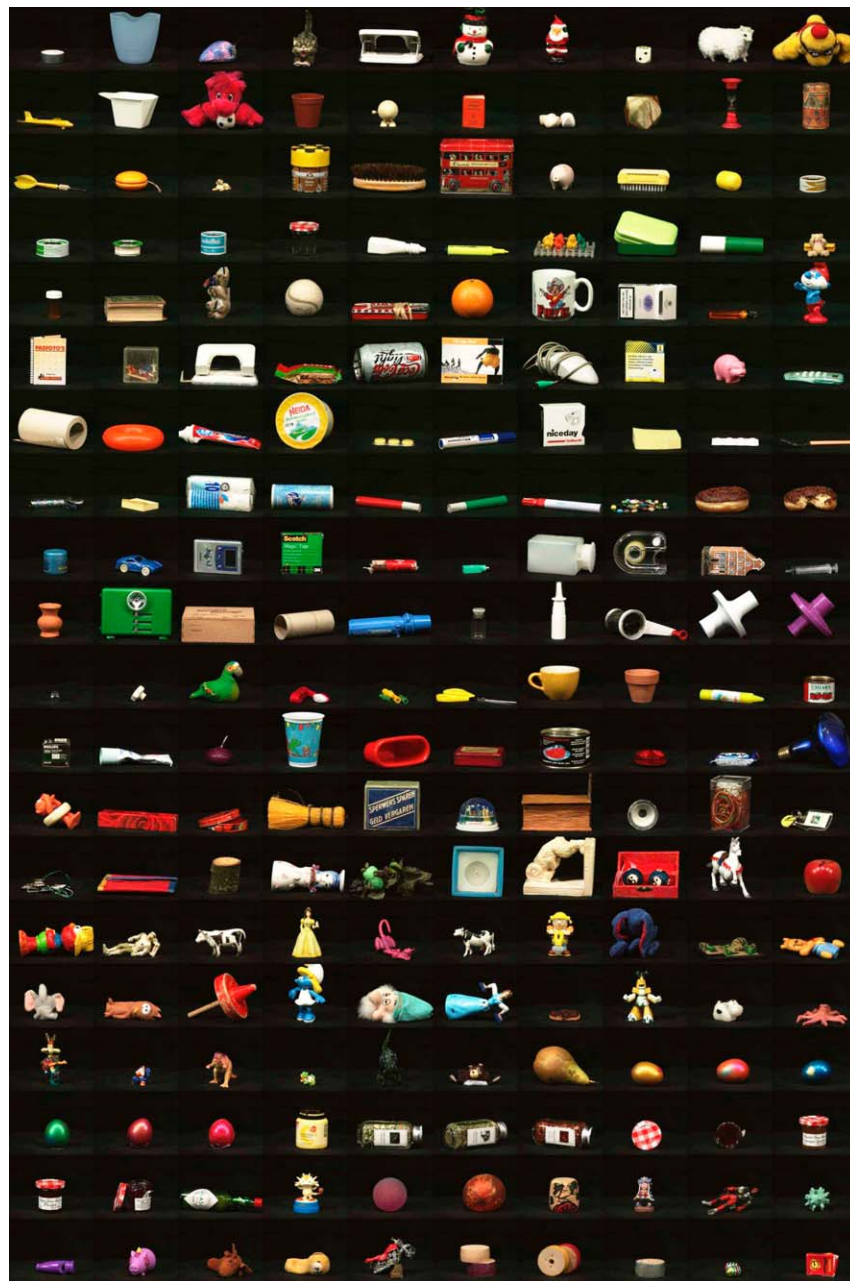


Figure 6. (Continued on next page.)

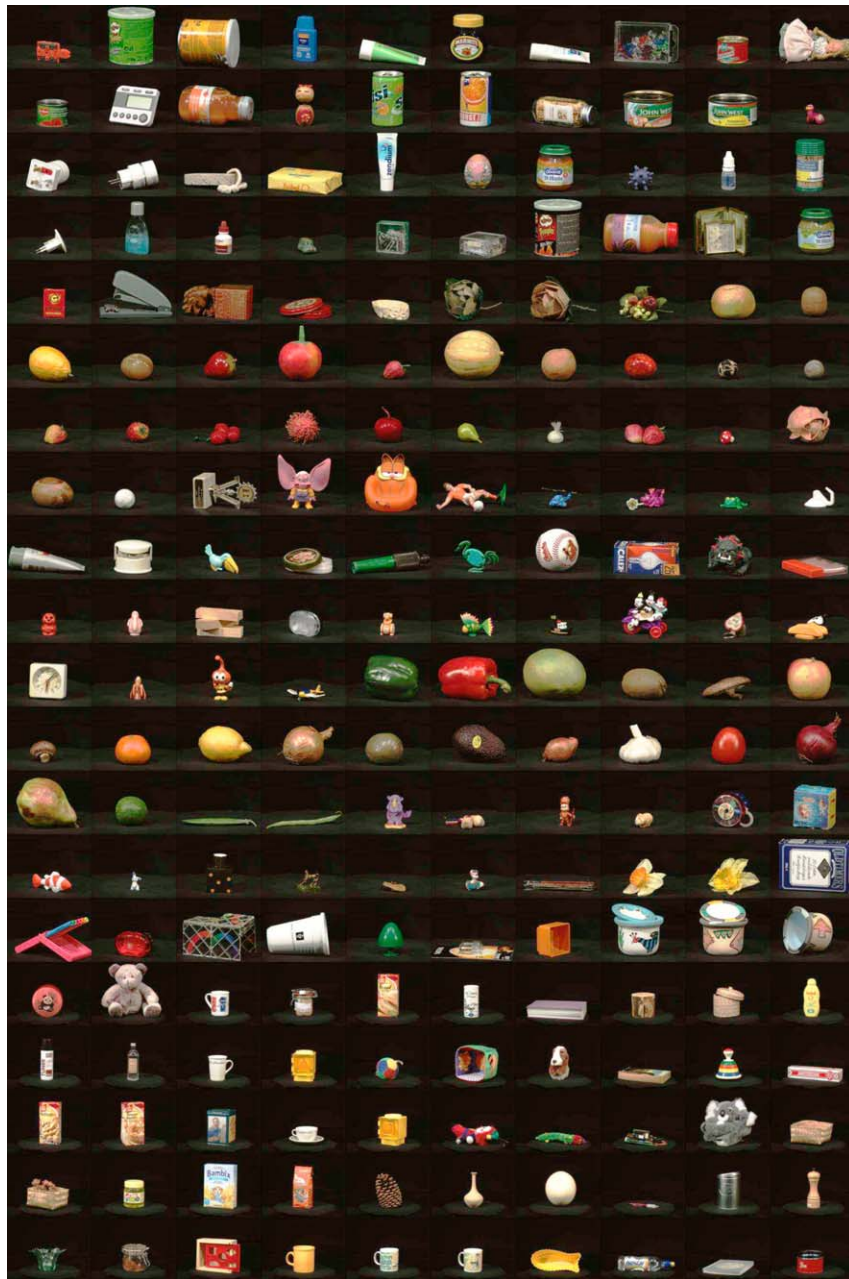


Figure 6. (Continued on next page).

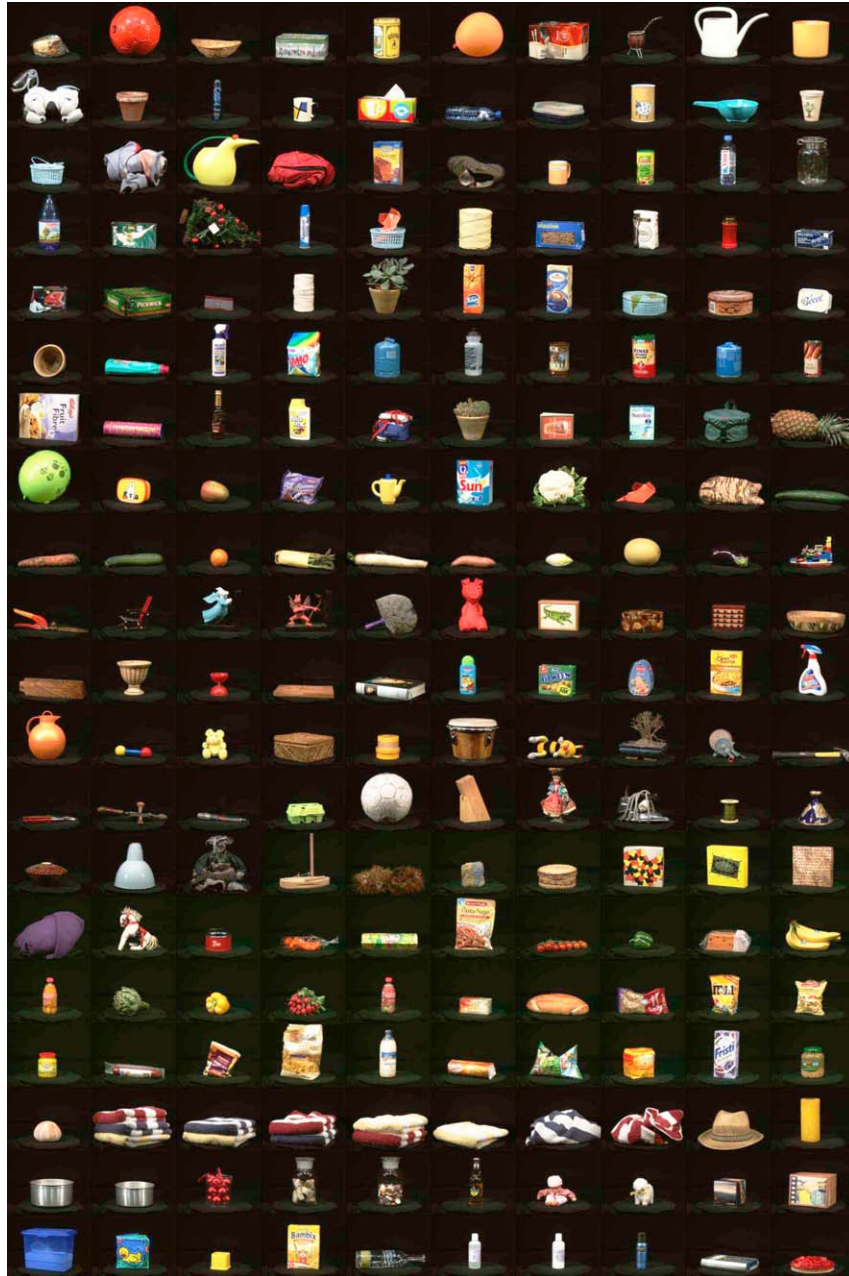


Figure 6. (Continued).

Acknowledgments

We are grateful to Edwin Steffens for constructing the acquisition setup and controller. We thank Ina Koning for recording a large portion of the objects. For recapitulating the data from the crashed server RAID, we are grateful to Jan Wortelboer and Jeroen Roodhart. For contributing all objects, we acknowledge Nora Geusebroek, Theo Gevers, Anna Grooten, Saskia Heijboer, Dennis Koelma, Virginie Mes, Giang Nguyen, Mireille Oud, Thang Pham, Diny Poelstra, Frank Seinstra, Cees Snoek, Joost van de Weijer, Marcel Worrying, and Elena Zudilova.

Note

1. <http://www.science.uva.nl/~aloi>

References

- Barnard, K., Martin, L., Funt, B., and Coath, A. 2002. A data set for colour research. *Color Res. Appl.*, 27(3): 147–151.
- Horn, B.K.P. 1986. *Robot Vision*. The MIT Press: Cambridge, MA.
- ITU-R Recommendation BT.709. 1990. Basic parameter values for the HDTV standard for the studio and for international programme exchange. Technical Report BT.709 [formerly CCIR Rec. 709], ITU, 1211 Geneva 20, Switzerland.
- Kant, I. 1781. *Critique of Pure Reason*. Hartknoch: Riga. translated by Guyer, P. and Wood, A.W. Cambridge University Press, 1998.
- Koubaroulis, D., Matas, J., and Kittler, J. 2002. Evaluating colour-based object recognition algorithms using the soil-47 database. In *Proceedings 5th Asian Conference on Computer Vision (ACCV 2002)*. D. Suter and A. Bab-Hadiashar (Eds.), Melbourne, Australia, pp. 840–845, Asian Federation of Computer Vision Societies.
- Nene, S.A., Nayar, S.K., and Murase, H. 1996a. Columbia object image library (COIL-100). Technical Report CUCS-006-96, Columbia University. <http://www.cs.columbia.edu/CAVE/research/softlib/coil-100.html>.
- Nene, S.A., Nayar, S.K., and Murase, H. 1996b. Columbia object image library (COIL-20). Technical Report CUCS-005-96, Columbia University. <http://www.cs.columbia.edu/CAVE/research/softlib/coil-20.html>.
- Nicodemus, F.E., Richmond, J.C., and Hsia, J.J. 1977. Geometrical considerations and nomenclature for reflectance. Technical Report Monogr. 160, Natl. Bur. Stand. (now NIST), 100 Bureau Drive, Stop 3460, Gaithersburg, MD 20899-3460.