THE KTH-TIPS database

Mario Fritz, Eric Hayman, Barbara Caputo and Jan-Olof Eklundh

Computational Vision and Active Perception Laboratory (CVAP)
Dept. of Numerical Analysis and Computer Science
KTH, SE-100 44 Stockholm, Sweden
{mjfritz,hayman,caputo,joe}@nada.kth.se

May 7, 2004

1 Background

This document provides a brief Users' Guide to the KTH-TIPS image database (KTH is the abbreviation of our university, and TIPS stands for Textures under varying Illumination, Pose and Scale). The guide describes which materials are contained in the database (Section 2), how images were acquired (Section 3) and subsequently cropped to remove the background (Section 4), and we also discuss some non-ideal artifacts, like poor focus, in some pictures (Section 5). This document concludes by outlining how we intend to extend the database in the future (Section 6).

The objectives with this database were to supplement the CUReT image database [1] in two directions, both of which concern extending material classification algorithms to function in real-world conditions. While the CUReT database contains images of 61 materials (over varying pose and illumination, but at constant viewing distance), the aims with KTH-TIPS were:

1. to provide variations in *scale* as well as variations in pose and illumination. This allows a systematic study of how important unknown viewing distance is to material classification, and provides data for evaluating algorithms intended to be robust to such variations;
2. to provide images of *other samples* of a subset of the CUReT materials, taken under different settings. We wanted to see whether it would be possible to actually classify materials in the real-world, as opposed to recognising exemplars of materials within a single database.

The cropped database is freely available on the internet [2]. Those interested in the full-size images should contact Eric Hayman (hayman@nada.kth.se).

The database was first presented and used in [3].

2 Imaged materials

While the CUReT database [1] images 61 materials, the KTH-TIPS database currently contains images of 10 of those materials as outlined in Table 1 and depicted in Figure 1. Each of the samples is planar. The orange peel was flattened by placing it inside a CD case.

Material	**Corresponding CUReT sample number**
Sandpaper	06
Crumpled aluminium foil	15
Styrofoam	20
Sponge	21
Corduroy	42
Linen	44
Cotton	46
Brown bread	48
Orange peel	55
Cracker B	60

Table 1: The materials present in the KTH-TIPS database.

Figure 1: Images of the materials present in the KTH-TIPS database.

3 Image acquisition

The images were taken with an Olympus C-3030ZOOM digital camera at a resolution of 1280×960 pixels. Many of the full-size images contain not only the sample, but also some background.

A single light source (standard desk lamp with a 60W tungsten light bulb) was used.

Images were taken at 9 different scales spanning two octaves. At the central scale the distance between the camera and the target was 28cm, this was selected to correspond roughly to the default scale in the CUReT database. The scales used are described in full in Table 2, and full-resolution images from one material (Cracker B) are shown in Figure 2.

Scale number	Relative scale	Distance to camera (cm)
1	$2^{-1.00} = 0.500$	14.00
2	$2^{-0.75} = 0.595$	16.65
3	$2^{-0.50} = 0.707$	19.80
4	$2^{-0.25} = 0.841$	23.55
5	$2^{0.00} = 1.000$	28.00
6	$2^{+0.25} = 1.189$	33.30
7	$2^{+0.50} = 1.414$	39.60
8	$2^{+0.75} = 1.682$	47.09
9	$2^{+1.00} = 2.000$	56.00

Table 2: The scales present in the KTH-TIPS database.

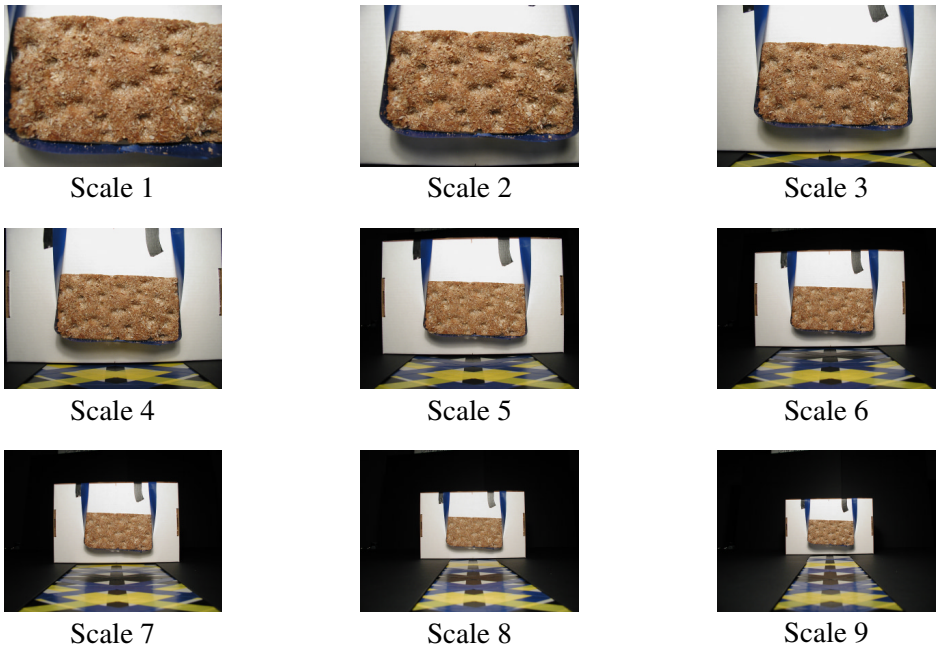

Figure 2: Full-size images depicting the variation of scale present in the KTH-TIPS database.

At each scale 9 images were taken in a combination of three poses (frontal, rotated $22.5\,^\circ$ left and rotated $22.5\,^\circ$ right) and three illumination conditions (from the front, from the side at roughly $45\,^\circ$ and from the top at roughly $45\,^\circ$). The adopted labeling scheme is shown in Table 3 and sample images in Figure 3.

Image number	**Object pose**			**Illumination direction**		
	Frontal	**$22.5\,^\circ$ right**	**$22.5\,^\circ$ left**	**Frontal**	**$\approx 45\,^\circ$ from top**	**$\approx 45\,^\circ$ from side**
1	x			x		
2	x				x	
3	x					x
4		x		x		
5		x			x	
6		x				x
7			x	x		
8			x		x	
9			x			x

Table 3: The labeling of images within each scale in the KTH-TIPS database.

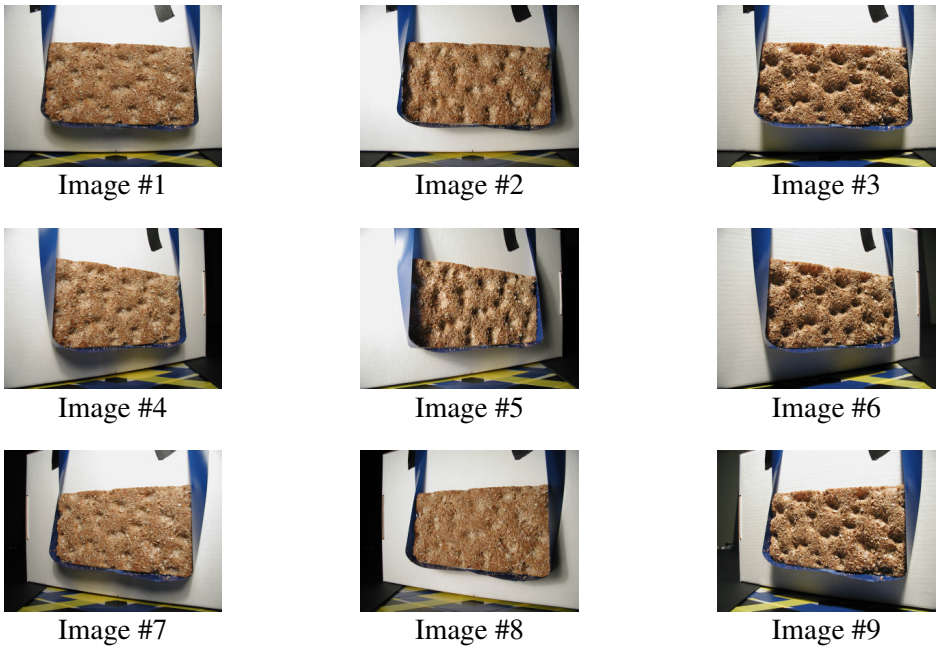

Figure 3: The variation of pose and illumination present in the KTH-TIPS database. In each row the pose is constant, whereas in each column the illumination is the same (frontal, side, or top illumination).

4 Image cropping

To remove the background, and to be consistent with the experiments in [4, 5], we manually cropped images to 200×200 pixels when possible. However, for some samples (Brown bread and Cracker B) this was not possible at large camera-target distances since the subject did not fill a sufficiently large part of the image. In these cases, images were instead cropped as follows:

1. If possible, an "equivalently sized" rectangular region was selected, with an aspect ratio as close as possible to 1. "Equivalently sized" implies that the same number of pixels should be made available to the computer vision algorithm (e.g. material classification algorithm) once early processing (e.g. filtering) has been performed. In our work [3] the early processing involved applying a filter bank, and removing pixels which were not entirely covered by the support region of the filter kernel. These pixels are located at the edges of the image patch. In particular we used 41×41 filter kernels, so with a 200×200 patch, after filtering we were left with $(200-40) \times (200-40) = 160^2 = 25600$ pixels which were input to the classification algorithm. Therefore, we selected $x \times y$ patches such that $(x-40) \times (y-40) \approx 25600$.
2. If the largest possible x and y did not satisfy the "equivalently sized" criterion above, we simply took the largest possible rectangular region corresponding to the foreground texture.

We must emphasize that the "equivalent size" condition is dependent on the employed image processing strategy and might very well be poorly suited to *your* application.

Table 4 lists where these cropping strategies were necessary. With Brown bread the texture round the edges of the slice is somewhat different (denser) to that in the middle, so these edges were also removed.

Material	Scale	Images	Cropping strategy
Brown bread	6	All	Equivalent size
	7	8,9	Equivalent size
	7	1,2,3,4,5,6,7	Largest possible
	8 and 9	All	Largest possible
Cracker B	7	All	Equivalent size
	8	1,2,3	Equivalent size
	8	4,5,6,7,8,9	Largest possible
	9	All	Largest possible

Table 4: Images where it was not possible to extract 200×200 pixels foreground patches.

Additionally we would like to point out that with Orange peel it was not always possible to extract 200×200 pixel foreground patches either. However, with this material the CUReT database exhibits similar problems; in the CUReT images some background is present. Since one of our main objectives was to attempt to recognise our samples using models trained on the CUReT database, we decided against cropping the Orange Peel to a smaller size. It is, however, undoubtedly a problem that the amount of background varies from scale to scale, and our background was not quite as uniform as the CUReT background.

5 Some poor quality images in the database

In the previous section we highlighted two problems present in the KTH-TIPS database, namely the *cropping* and the *orange peel – background* problems.

A further source of information degradation is problems with the automatic focus of the camera, and this severely limits the value of any image processing/classification result for those images. Scale 1 (subject closest to the camera) was probably chosen to have an insufficient camera-target distance. Furthermore, in an attempt to get patches which were in focus, non-central patches were often selected, implying that the estimate of camera-target distance is inaccurate. Indeed, one should not be too surprised if poor results are obtained with Scale 1.

Table 5 summarises the images with poor focus.

Material	**Scale**	**Images**
Sandpaper	1	All
Crumpled aluminium foil	1	1,2,3, 7,8,9
Styrofoam	1	1,2,3, 7,8,9
Sponge	1	1,2,3, 7,8,9
Corduroy	1	1,2,3, 7,8,9
Linen	1	1,2,3, 7,8,9
Cotton	1 8 9	1,2,3,4, 7,8,9 8 1,4,7,8,9
Brown bread	1	1,2,3, 7,8,9
Orange peel	1	All
Cracker B	1	1,2,3, 7,8,9

Table 5: Images poorly focused in the KTH-TIPS database.

Further issues are

- perspective effects at some of the closer distances (this is fairly minor);
- minor creases in cloth (linen and cotton) induce spurious edges. On the other hand, real-world samples of cotton and linen can also have creases;
- specularities with orange peel caused by placing the peel in a CD case. This is prominent at Scale 9.

6 Future extensions to the database

In our future work we hope to extend the database with multiple samples of each material, and with more than the 10 materials currently present. Future versions of the database might also eliminate some of the problems detailed in Section 5 in this document.

References

[1] K.J. Dana, B. van Ginneken, S.K. Nayar, and J.J. Koenderink. Reflectance and texture of real-world surfaces. *ACM Transactions on Graphics*, 18(1):1–34, January 1999.

[2] M. Fritz, E. Hayman, B. Caputo, and J.-O. Eklundh. The KTH-TIPS database. Available at `www.nada.kth.se/cvap/databases/kth-tips`.

[3] E. Hayman, B. Caputo, M. Fritz, and J.-O. Eklundh. On the significance of real-world conditions for material classification. In *Proc. 8th European Conf. on Computer Vision, Prague*, 2004.

[4] M. Varma and A. Zisserman. Classifying images of materials: Achieving viewpoint and illumination independence. In *Proc. 7th European Conf. on Computer Vision, Copenhagen*, page III: 255 ff., 2002.

[5] M. Varma and A. Zisserman. Texture classification: are filter banks necessary? In *Proc. Computer Vision and Pattern Recognition*, pages II: 691–698, 2003.