An Active Patch Model for Real World Texture and Appearance Classification – Supplementary Material

Junhua Mao, Jun Zhu, and Alan L. Yuille

University of California, Los Angeles

1 Introduction

The supplementary material is organized as follows: Firstly, we show the experimental results about the threshold value for matching score $M(\mathbb{A}, I, pos)$ in Section 2. Secondly, we discuss the role of color information in the three datasets we tested (i.e. AniTex, VehApp and KTH-TIPS2 [1]) in Section 3. Thirdly, in Section 4, we describe the psychophysics experiments in detail. Finally, we show more example images from the main image set of AniTex and VehApp (3120AniTex and $\mathrm{VehApp_{crop}}$) in Section 5 and the visualization of the active patch dictionaries in Section 6.

2 The threshold for matching score

We adopt a threshold for the matching score $M(\mathbb{A}, I, pos)$ to determine whether an active patch $\mathbb{A}$ is fired on a target position pos in image I. $M(\mathbb{A}, I, pos)$ can be calculated by Equ. (1) in line 230 of the main paper. We use 0.8 as the value of the threshold in the main paper to ensure the perceptual similarity between the fired active patch and the corresponding image region, as well as the high performance. The performance comparison of different threshold values on the 3120AniTex dataset is shown is Fig. 1. On one hand, if we use a threshold that is too large, the accuracy of our method will drop. The reason is that although larger threshold will enforce a more precise firing between the active patch and the target image region, it will also make the firing more possible to be affected by the noise in the image. In addition, using larger threshold can lead to more ”blank” image regions (image regions that are not fired by any active patches), which will decrease the representative ability of the active patch dictionary.

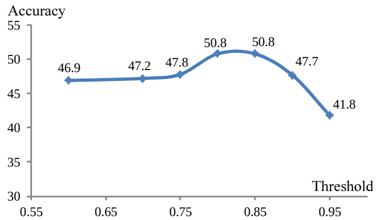

Fig. 1. Performance comparison on the 3120AniTex dataset using different threshold values.

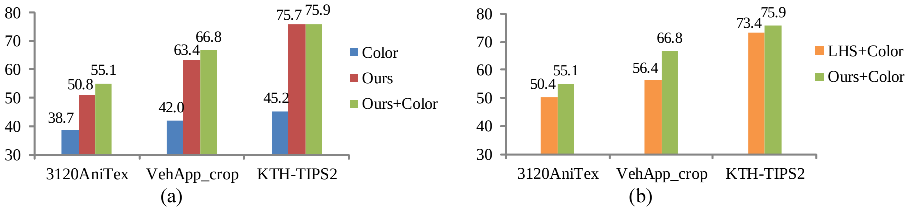

Fig. 2. (a). Performance comparison of the color descriptor, our method and the combination of our method and the color descriptor on the three datasets. (b). Performance comparison of LHS [2] combined with the color descriptor and our method combined with the color descriptor. (Best viewed in color)

On the other hand, if we use a threshold that is too small, the active patches and some of their fired image regions might not be similar enough to each other. It will lead to an inaccurate description of the image region by the active patches, and thus decrease the performance.

3 Color

For a fair comparison with the state-of-the-art texture analysis methods [2–5], we focus on the analysis of gray scale texture information for the dataset images in the main paper. We also investigate the role of color information for the three datasets (i.e. 3120AniTex, $\mathrm{VehApp}_{\mathrm{crop}}$ and KTH-TIPS2 [1]). The results are shown in Fig. 2.

We adopt a simple color histogram descriptor. The images are converted to HSV color space. We quantize the hue channel into 16 bins, saturation channel into 8 bins and value channel into 2 bins. This leads to a descriptor with dimension 256. We apply L1 norm on the descriptor. The experimental results show that color provides additional information for the classification of different types of texture and appearance. When concatenating the simple color descriptor with our image descriptor, the performance will increase as shown in Fig. 2(a). The improvements of accuracy are 4.3%, 3.4% and 0.2% for 3120AniTex, $\mathrm{VehApp}_{\mathrm{crop}}$, and KTH-TIPS2 respectively. It implies that the color descriptor are complementary to our method based on active patches. We also combine the color descriptor with LHS [2] by concatenating these two descriptors. The results in Fig. 2(b) show that the color descriptor can also increase the accuracy of LHS by 2.3%, 3.1%, and 0.4% for 3120AniTex, $\mathrm{VehApp}_{\mathrm{crop}}$, and KTH-TIPS2 respectively. The performance gain is generally lower than or equal to that of our method combined with the color descriptor.

4 Psychophysics experiments

In this study, we are interested in determining whether texture can help to identify the category of the animals in human vision. The results of this study also reveal the difficulty of the classification tasks on animals' texture.

4.1 Methods

We use the 250AniTex dataset for this experiment. A total of 11 participants took part in the psychophysical experiments. They were all familiar with the animal categories but possessed no specific proficiency or animal expertise. The 250 images (50 images per category) were presented in random order on a 13-inch screen set at 800×600 pixels resolution. The images were converted to gray scale ones. Participants sat at a distance of 50 cm to the screen during the experiment and were told to maintain this distance in order to keep the visual angle of the stimuli as constant as possible.

Participants first learned about the five animal categories by seeing examples of animals and the cropped patch images (these images were not used during the actual experiment). The experimental procedure involved a five-alternative forced choice task with brief exposure of 300 ms (It is a standard practice [6]). In the task, the participants need to decide the most probable animal category for each texture patch image as quickly as possible. The patch image was followed by a masking image consisting of randomized pixels, which was used to ensure that visual information can only be processed for the 300 ms duration. After the masking stage, an instruction, i.e., "Please select the most probable animal", was shown on screen. The participants were requested to press one of five keyboard buttons labeled with the five animal categories.

4.2 Results and Discussion

The dependent variables measured in this experiment are accuracy and response time for each category. They are analyzed using a standard one-way repeated-measures ANOVA (analysis-of-variance) with follow-up T-tests. We want to test whether the participants gave their response according to the animal category information contained in the testing images. The mean accuracy for the five categories are: cat= 56.36%, dog= 24.36%, sheep= 54.00%, cow= 55.09%, horse= 46.91%. The ANOVA reveals a highly significant effect of animal category on humans' response: $F(4, 40) = 12.36$, p-value< 0.001. It means that humans make their decisions based on the animal category information provided by the texture images instead of random guess. The corrected post-hoc tests show that the accuracy for dog images is significantly worse than that of all other animal categories (all p-value< 0.03). Except for dogs, the performance for all other

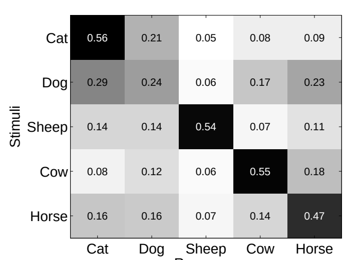

Fig. 3. The confusion matrix for humans (Fig. 6(b) in the main paper).

categories is better than chance (=20%). In order to gain more insight into the response patterns, Fig. 3 plots the confusion matrix for this experiment (as shown in Fig. 6(b) of the main paper). We can see that cats are often confused with dogs, whereas dogs are confused with cats and horses. One possible reason is that certain types of dog, cat and horse share very similar texture patterns (e.g., a smooth fur).

The average response time for these five categories ranges from 1.2 to 1.3 seconds. The ANOVA fails to find an significant effect of the animal category on the response time: $F(4, 40) = 0.87$, p-value$= 0.49$. The higher response time compared to [6] may be due to the larger number of choices the participants had to make (five choices in our study versus two choices in [6]).

To sum up, our results show that animal categorization based on briefly presented texture patches is feasible and the animal texture contains critical information about the animal categories.

5 More examples of the images from the two new datasets

Fig. 4 and Fig. 5 show more examples from the main image sets (3120AniTex and $\mathrm{VehApp}_{\mathrm{crop}}$) of the AniTex dataset and the VehApp dataset respectively.

6 Visualization of the learned dictionary

We visualize the learned active patch dictionaries for 3120AniTex, $\mathrm{VehApp}_{\mathrm{crop}}$ and KTH-TIPS2 in Fig. 6 respectively by showing the basic patches of the first 200 active patches in each dictionary.

References

1. Caputo, B., Hayman, E., Mallikarjuna, P.: Class-specific material categorisation. In: ICCV. (2005)

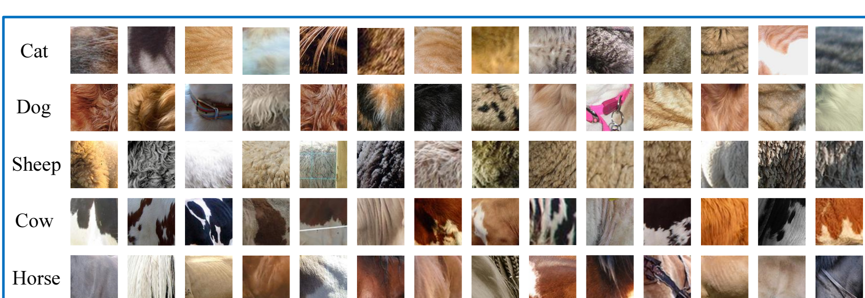

Fig. 4. More example images from the main image set (3120AniTex) of the AniTex dataset.

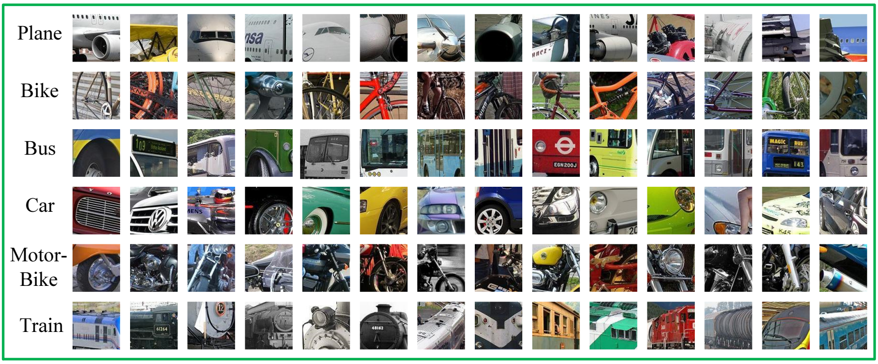

Fig. 5. More example images from the main image set ($\mathrm{VehApp}_{\mathrm{crop}}$) of the VehApp dataset.

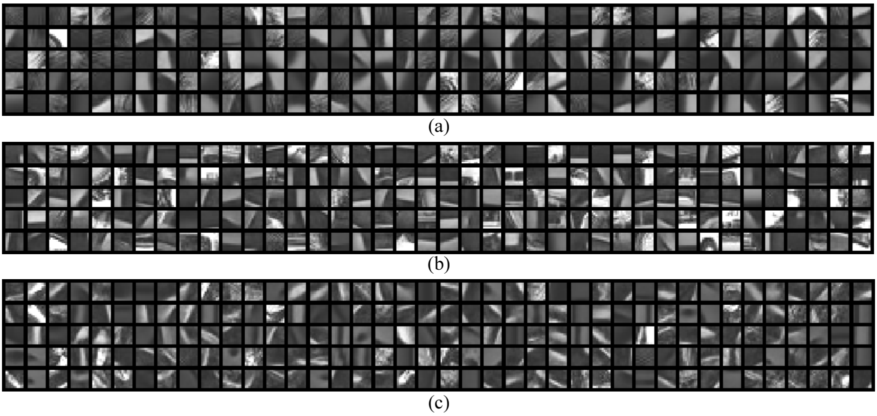

Fig. 6. The first 200 active patches in the learned active patch dictionaries for: (a) the 3120AniTex dataset, (b) the $\mathrm{VehApp}_{\mathrm{crop}}$ dataset, and (c) KTH-TIPS2 dataset [1]. The active patches are sorted by their selection order of the greedy algorithm described in Section 3.3 of the main paper. We visualize the active patches by showing their basic patches.

2. Sharma, G., ul Hussain, S., Jurie, F.: Local higher-order statistics (lhs) for texture categorization and facial analysis. In: ECCV. (2012)
3. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. TPAMI **24**(7) (2002) 971–987
4. Guo, Y., Zhao, G., Pietikäinen, M.: Texture classification using a linear configuration model based descriptor. In: BMVC. (2011)
5. Guo, Y., Zhao, G., Pietikäinen, M.: Discriminative features for texture description. PR **45**(10) (2012) 3834–3843
6. Thorpe, S., Fize, D., Marlot, C., et al.: Speed of processing in the human visual system. nature **381**(6582) (1996) 520–522