Report

In this research paper, to figure out a helpful cue in the direction of rapid scene identification, the author has proposed a simple texture analysis of the image. Across various scene categories, this model learns texture features and then identify new scenes using the knowledge from different scene categories. As subjects with limited processing time, the texture analysis leads to similar identifications and confusion. The texture model performs best with smaller exposure time but as the exposure time increases, the human vision outperformed the texture model for scene identification.

The paper describes an experiment in which 48 undergraduates with normal or corrected-to-normal vision have participated for 1-h experiment. The database contains 1000 images of scenes with 10 basic-level categories like forest, kitchen, bedroom, city, farm, etc. These scenes further categories in man-made/indoor, man-made/outdoor and natural/outdoor. Out of these images, 250 images are selected randomly as the training set for the model and the remaining 750 images were used as the test set. To reduce visual distractions, the experiment was executed in a dim lit room. The images are never repeated during experiment. The procedure of this experiment includes fixation for 2000 ms, displayed grayscale image for 37, 50, 62 and 69 ms then displayed jumbled image (Mask) for 20 ms, displayed blank screen for 500 ms and finally asked two word choices and participants needed to select the word which best described the target.

For computational texture based model, the training set consists of 250 images which were not used in the testing phase. From the training set, the model learned universal textons. In the generic natural images, texton are fundamental micro-structures. The texton helps for image segmentation and recognition as a part of image modeling for reduced dimensions and less dependence between coefficients of a decomposes image representation [1].

In the model, to estimate texture features at each pixel in the image, uses a filter bank of first and second derivatives of Gaussian. The measure of standard deviation 2/standard deviation 1 is the elongation of the filter. The filters include two phases which are even and odd, three scales which are spatial frequency, selectivity and orientation selectivity, and six orientations which is equally spaced from 0 to pi.

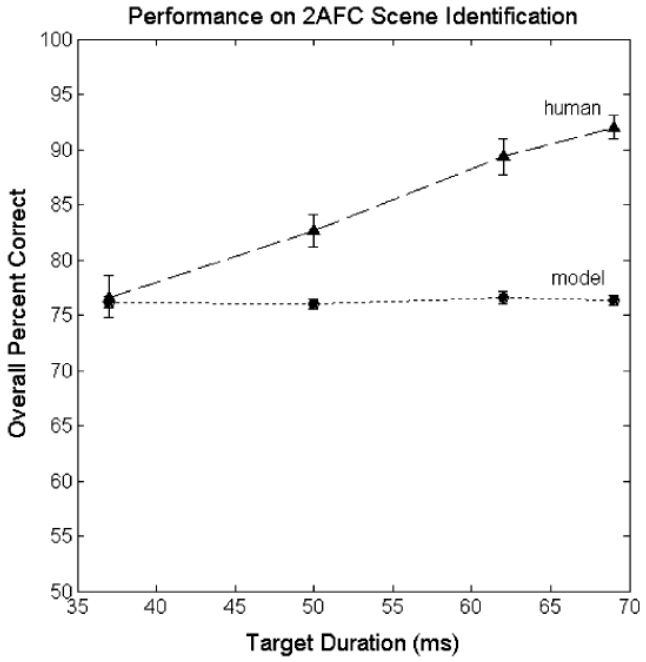
After this, the image is convolved with the created filter bank which yield a vector of filter responses. Texture is created when it has repeating properties of vector. The texture feature reoccurred in the image when it has similar vector of responses. Using k-means clustering algorithm, clustered the resulting response vectors to find texture feature in the training image. Based on the vector responses, every single pixel of an image is assigned a texton channel. The number of pixels in the texton channel n, represents the value of nth histogram bin for an image. Thus, the representation of histogram is nothing but frequencies of texton in the image.

In the experiment, for identifying new scenes, 750 images were taken into consideration to test the texture model for identifying scene. The histogram was created as a descriptor of texture for every single new image. To find the best match, the new image (test image) histogram is

Ishika Prasad | ip1262@rit.edu

compared with the training image histogram using the similarity measure. If the x2 similarity measure is small then based on texture feature, two images are similar.

The texture model is tested on 2AFC for scene identification. From the figure, it is evident that at basic-level categories, texture model identified correct scenes. In comparison with human vs texture model, with the longer exposure time, human identified the scene correctly with higher confidence than with less exposure time but for model, it worked best for less exposure time and with the increase in exposure time, giving the less confidence to identify the scene.



In summary, the human visual system is quite rapid for scene identification and useful for object identification during real-world tasks. The paper shows the comparison of the model and human performance in which it is shown that for scene identification texture provides a strong cue for subordinate and basic category levels during early scene processing. The author describes that only texture was able to account for correct categorization and error patterns on 8/10 scene categories. In conclusion, for early scene identification, simple texture recognition model suits well.

To identify the scene with small exposure time, the texture model works good for basic level category as like the humans. The texture model will not work correctly if the background texture matches with the scene image as determining the texture will be hard in these cases [2]. If the scenes which have similar visual textures, then both the human and model will be confused and might not determine the scene category. For the images which have vertical orientation like forest and street scenes, the model can be confused between outdoor/nature and outdoor/man-made scenes while the human rarely gets confuse to identify categories of these scenes.

References:

[1] http://www.stat.ucla.edu/~ywu/research/papers/eccv2002_texton.pdf

[2] https://ieeexplore-ieee-org.ezproxy.rit.edu/stamp/stamp.jsp?tp=&arnumber=7435688