

Automatic determination of text readability over textured backgrounds for augmented reality systems

Alex Leykin

Department of Computer Science
Indiana University, Bloomington
oleykin@indiana.edu

Mihran Tuceryan

Department of Computer and Information Science
Indiana University - Purdue University Indianapolis (IUPUI)
tuceryan@cs.iupui.edu

Abstract

This paper describes a pattern recognition approach to determine readability of text labels in augmented reality systems. In many augmented reality applications, one of the ways in which information is presented to the user is to place a text label over the area of interest. However, if this information is placed over very busy and textured backgrounds, this can affect the readability of the text. The goal of this work was to identify methods of quantitatively describing conditions under which such text would be readable or unreadable. We used texture properties and other visual features to determine if a text placed on a particular background would be readable or not. Based on these features, a supervised classifier was built that was trained using data collected from human subjects' judgement of text readability. Using a rather small training set of about 400 human evaluations over 50 heterogeneous textures the system is able to achieve a correct classification rate of over 85%.

1. Introduction

There are many modes in which location aware information can be added to the scene in order to aid the user. One mode is to give the information in the form of 3D graphics properly aligned and overlaid on the images of the real world scene and the many technical problems that arise from this have been addressed extensively in the literature. Another mode of presenting information is to label objects or associate information about objects, by placing text labels on or near the 3D objects in the world. At first glance this appears to be a much more manageable task for many reasons. We do not have to worry about precise and realistic 3D alignments with 3D objects. Many technical issues that have to do with 3D information disappear or become easier. However, the overlay of text labels on the scene result in a

new set of problems, some of which have been addressed before.

For example, an important issue that arises in this respect is the layout of the text labels in a crowded scene such that the labels are readable and do not interfere with each other. Some aspects of text label placement in AR such as the layout and overlapping of the text have been studied and evaluation criteria and algorithms have been developed to generate the most readable layout of text labels [1, 10].

There are, however, considerations other than the text layout that affect the readability of the text labels. One such factor on text readability is the interference from the background texture. If the placement of text on such backgrounds is not carefully done, the visual properties of the background may interfere with the readability of the text so placed. Figure 1 shows an example of such a labeling which has become hard to read due to the background texture.

In this paper, we present a method to automatically determine if a given text will be readable or unreadable. In order to accomplish this, we use statistical pattern recognition techniques to design a classifier.

In Section 2, we present a brief summary of related work. In Section 3, we present our approach to classifying readability of text over textured background. In Section 3.1 we describe the experimental setup for obtaining the training data for the supervised classifier. In Section 3.2, we explain the visual features we use for the classification. Section 4 outlines the experimental results. Finally, in 5, we present our conclusions and outline potential research on this problem.

2. Prior Work

Prior research has studied the visual properties of the background that affect the readability of text [4, 11, 12]. Scharff et al. have identified the contrast and spatial frequency content of the background texture as factors affecting the readability of text. These studies have been con-



Figure 1. An example scene with text label placed over the buildings being labeled that are unreadable.

ducted in the context of augmented reality and web design applications, but the problem was never treated as one of automated readability classification. The goal of their experiment was to study the factors affecting the readability of text in the human perceptual system and from this derive certain rules, according to which the appropriate background texture and text properties could be selected. This goal, of course, is practical in web design applications, but less so in augmented reality applications. One cannot always control the texture characteristics of the background images in the real world, other than possibly change the position of the text and place it on a more suitable background. In order to even decide that a particular background is going to adversely affect readability, one needs to have an automatic way of analyzing the background image and how it will interact with a specific text to be overlaid on it.

Petkov and Westenberg have studied the effect of band-limited noise on perception of contours and, in particular, the effect of such noise (texture in our context) on the perception of text [9]. They have conducted psychophysical experiments to reach their conclusion that the frequency range of most efficient suppression is related to the letter stroke (or weight) and not to the letter size. This con-

clusion supports our use of Gabor based features in our work. Solomon and Pelli showed that the luminance contrast threshold above which the letters become readable depends on the intermediate frequencies [13].

3. Architecture of the system and collection of training data

In this paper, we utilize pattern recognition techniques to determine if a text label overlaid on a textured background is readable. In order to automate this task, we designed a supervised classifier (see Figure 2) that used text features as well as texture features of the background image. We implemented a number of different such classifiers and used feature selection methods to identify the most important features in deciding readability. The classifiers were trained on data collected through a set of experiments with human subjects. Below we describe the details of the data collection, classifier design, and feature selection. As our features for designing the supervised classifier, we used contrast features for the text, font features such as size and weight, and Gabor filter based texture features at various frequencies and orientations to be used in our supervised classifier.

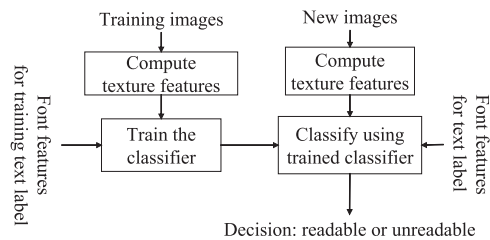


Figure 2. Brief overview of the architecture of our system.

3.1. Collection of training data

The readability of text is intimately tied to the perception and cognition of the humans who will be looking at such texts [7, 13]. Unlike many other computer vision or pattern recognition tasks, in which the class labels are well defined (e.g., what a handwritten letter is), the degradation of the readability of a text must be determined by the performance of the human who is looking at the text. In order to collect the training data needed by our supervised classifier, we set up a series of experiments in which human subjects assigned class labels “readable” vs. “unreadable” to the sample texts displayed to them.

Six independent participants were presented on a computer screen with a series of 100 images each, in which some sample text was overlaid on a textured background

(see Figure 3 for some examples). Three of the subjects were native speakers of English and three of them were non-native speakers. All participants had self-reported 20/20 or corrected to 20/20 vision. The subjects were approximately 50 cm away from the screen. As the overlaid text images were presented on the screen, the subjects had to assign a number between 1 to 8 for the level of readability of the text: 1 being least and 8 being most readable. The subjects were informed that numbers from 1 to 4 represent an unreadable text/texture combination, similarly numbers 5 to 8 represent a readable one. The textures were randomly chosen from a set of 50 texture images representing various materials, texture scales and objects. The textures were preselected to cover uniformly a range of frequencies and orientations. The texts presented to the subject were single words with approximately the same number of characters and the total set of characters was chosen so as to cover an entire Latin alphabet.

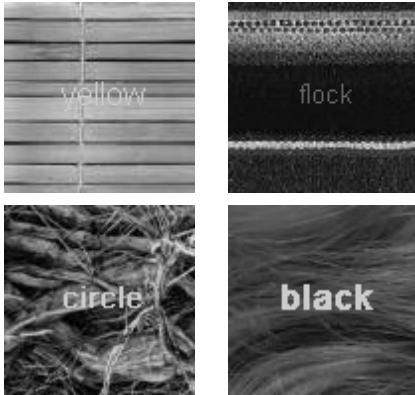


Figure 3. Samples presented to human subjects and the results of classification by our system. The first row shows samples classified as unreadable by the human subjects and the second row shows samples classified as readable by the human subject. The samples in the left column are classified by our supervised classifier as unreadable and the right column is classified as readable.

Several parameters were set as variable and further restrictions were imposed on them during the experiments. All the runs were performed on the same, calibrated monitor, under same lighting conditions (daylight). The viewing distance was set to be approximately an arm length (50cm) and viewing angle was fixed for all the subjects to position the eye-level at the same level as the center of the CRT screen. The dimensions of the display were 1024x768 with a pixel size of approximately 0.5mm.

To single out the luminance information all experiments

were conducted in gray-scale. Five font intensity levels from 0 to 1 with a step of 0.2 were used. We have fixed font sizes (10, 12 and 14 point Times New Roman) so that the height of a character is located around the same angle, which can be estimated as:

$$\alpha = \frac{360h}{2\pi r}, \quad (1)$$

where h is the height of the text and r is the distance from the screen. For the given font sizes (approx. 6 pixels high) and the viewing distance (approx. 500 mm) and measuring the size of the pixel as 0.5mm on the 1024 × 768 17 inch CRT display, we estimate the viewing angle height of the text to be approximately 0.35°

Note that the ranges used for intensity level, font size and weight were selected as a result of a preliminary experiment to eliminate outlying configurations and approximately equalize the volume of two classes. A priori probability for each class was set to 0.5.

Overall each subject was presented with 100 random configurations of the following: 50 textures, 13 input words, 6 font intensities, 4 font weights and 3 font sizes. This constitutes to a space of 46800 possible configurations, a space hardly covered by the total of 600 exemplars taken by us.

As one can see from Figure 3, the judgement about readability by the human subjects may be questionable in borderline cases (upper right and lower left images in the figure). For such borderline cases, the variability among human subjects was also high.

In order to eliminate inconsistent judgements across different subjects, we removed the outliers in the following way: all the responses for the same texture image and text intensity with rankings differing more than 2 points (e.g., 5 and 8) were removed. See section 4 for the results of the simulation.

We had two sets of experiments in order to take into account the attentive vs. pre-attentive nature of the processing of the visual inputs. The first set of experiments were conducted with no time limit on the visual input presented to the subject. The subject could take as long as he wished until a decision was made on the readability of the input. It is interesting to note that in the first set of experiments, response times were recorded for each subjects' choice. However they didn't demonstrate any significant correlation with the text readability, as defined by subjects' responses. This, of course, is a very deliberative process and we also wanted to see how much of the decision is affected by the pre-attentive processing by the visual system. Therefore, in the second set of experiments, all the original parameters were kept unchanged except that each image was shown for a brief period of time (150 msec or 300 msec). These intervals were chosen to capture pre-attentive and attentive visual system response correspondingly. According to Julesz preattentive

processing in the human visual system takes place under 200 msec [6]. The subject had this amount of time to pick the word he or she was presented with from a list of 8 alternatives. These two timings were randomly mixed with equal distribution throughout the experiment.

3.2. Texture and Contrast Descriptors

As our visual features to be used in the classifier, we used font size, font weight, intensity contrast, and responses to a bank of Gabor filters at four spatial frequencies and four orientations (for a total of sixteen filters) convolved with the background texture. As shown by [2] both luminance contrast and texture of the text and background determine a so-called “perceived contrast.” In other words, the clarity of the text and therefore the ease of reading is affected by these two factors.

We computed the contrast feature as the absolute difference between the mean intensities of the text and the surrounding region. Given the mean intensity of the text, I_t , and the mean intensity of the background, I_b , the contrast measure is given by $C = |I_t - I_b|$. I_b was computed using pixels in the immediate surrounding vicinity of the text and not the entire background image.

Gabor filters have been used as successful texture descriptors for classification and segmentation tasks in the past [3,5,14]. To extract both spatial and frequency descriptors of the background locally we used real-valued even-symmetric Gabor filters as given by Malik and Perona [8]:

$$h(x, y) = \exp \left\{ -\frac{1}{2} \left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) \right\} \cos(2\pi u_0 x), 3 \quad (2)$$

where σ_x and σ_y are the parameters controlling the width of the Gaussian envelope and u_0 is the frequency of a sine wave. For Gabor filters at different orientations, this basic function is rotated around the origin.

We have used filters at four orientations (0, $\pi/4$, $\pi/2$, and $3\pi/4$), and four spatial frequencies (1/2, 1/4, 1/8, and 1/16), producing a bank of 16 Gabor filters.

The resulting bank was convolved with the original background image before text was superimposed on it, generating 16 filtered background images. The mean value over the entire image for each feature was computed to obtain the texture features for the image. This assumes that the image contains a single texture.

3.3. Supervised Classifier

We implemented a number of supervised classifiers using the training data obtained from our human experiments. The classifiers were trained using the features described

in Section 3.2 and the class labels “readable” or “unreadable” obtained from our data collection. We assigned the class label “unreadable” for samples ranked 1 to 4 and the class label “readable” for samples ranked 5 to 8. We implemented and tested the classifiers in Table 1, a total of seven classifiers: Normal densities based linear classifier (LIN); Normal densities based quadratic (multi-class) classifier (SQR); Nearest mean classifier (MEAN); Scaled nearest mean classifier (MEANSC); Uncorrelated normal densities based quadratic classifier (BAYES); Fisher’s discriminant (minimum least square linear classifier) (FISHER); and Support vector machine classifier (SVM).

4. Experimental Results

Before we describe the detailed quantitative experimental results, we give an example image of how this method can be used to determine readability. Figure 4 shows how the classifier we designed is used to determine which regions of the image given in Figure 1 result in readable text and which regions result in unreadable text. We divided the input image (of size 375×500 pixels) into 96×96 pixel regions. Then given the text label (font size 12, white color, and font weight normal), we classified each region of the image as readable (white regions in Figure 4(b)) or as unreadable (black regions in Figure 4(b)). Figure 4(c) then shows these regions using a semi-transparent red mask for the unreadable regions (only for purposes of visually presenting our classification result in this paper). In a typical AR system this information would be used to place the text label over the readable regions.

Now, we describe some of the more detailed quantitative results. We used 70% of the points as a training set to build a classifier and the remaining 30% to test its performance. Several classifiers mentioned above were used to classify the test data. The results are summarized in Table 1.

As we can see from this table, the best classification results are given by the Support Vector Machine classifier with $> 87\%$ classification accuracy. We looked at the reasons why we are unable to reach better classification rate, and our conclusion was the training data collected from the human experiments seems to be limiting the performance of the classifier. As discussed in Section 3.1, we did remove some outliers in the human classifications. The outliers were approximately 20% of the training samples. Column 1a in Table 1 gives the error rates that include the outliers in the training data and column 1b gives them after taking the outliers out.

As mentioned in Section 3.1, the selection of ground truth is a highly subjective process, because it relies on the judgment of a few subjects. We believe that further improvement can be achieved with increased size of human subject set. In addition, if the number of samples in the

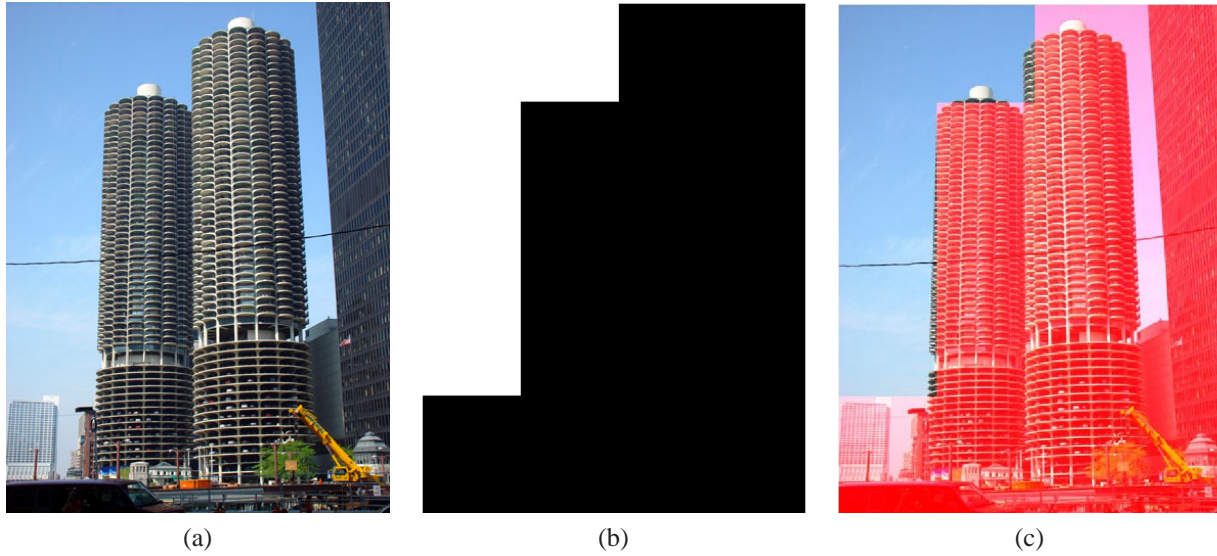


Figure 4. (a) An example scene image with textured and non-textured regions. (b) The regions of the image classified as readable (white) and unreadable (black) when a text label of 12 points normal weight and white color is overlaid. (c) The readable and unreadable classified regions superimposed on the original image (red mask is unreadable).

Classifier	1a	1b	2a	2b
LIN	0.160	0.131	0.250	0.197
SQR	0.207	0.165	0.283	0.255
MEAN	0.157	0.138	0.270	0.284
MEANSC	0.155	0.128	0.266	0.269
BAYES	0.170	0.141	0.303	0.281
FISHER	0.160	0.132	0.249	0.198
SVM	0.161	0.122	0.250	0.206

Table 1. Classification error rates for: 1a - results of first experiment used as training data; 1b - same as 1a with outliers removed; 2a - results of second experiment (with 150 msec delay) used as training data; 2b - results of second experiment (with 300 msec delay) used as training data.

training set is sufficiently high, the data can be subjected to more strict outlier removal process. For example removing identical samples with ranks differing at least by 1 point (as opposed to 3 point in the current setup) could lead to a more consistent definition of “readability.”

To make a preliminary assessment of the class separation we projected the data from the multidimensional feature space down to two dimensions while keeping the structure of the data in the feature space (Figure 5). Looking at

this figure, one can see that the two classes have a certain amount of overlap and we believe that this explains to some extent the classification error rates.

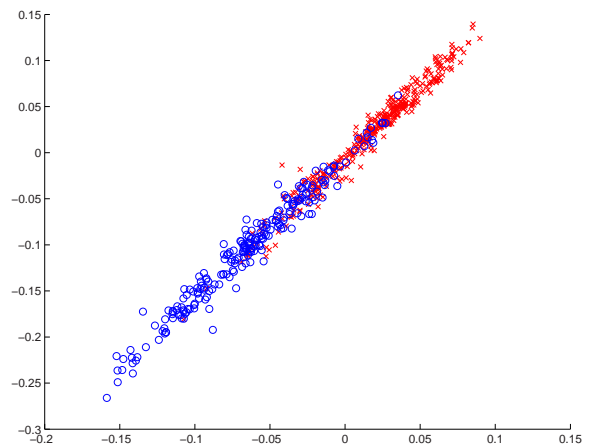


Figure 5. The features projected to two dimensions while maintaining the structure in feature space: Red (x) is one class, and blue (o) is another.

5. Conclusions and Discussion

The results of the human subject experiments confirm that background variations only affect readability when the text contrast is low [13]. The dominating frequency range that affects the readability: from 1.5 to 3.0 cycles per letter, also agrees with the results of [12, 13].

Figure 6 shows the frequency most correlated to the readability ranking to be 8 cycles per filter (32 pixels), which is equal to 4 pixels. Given letter size of approximately 6 pixels we get 1.5 cycles per character. Dominant orientation was $\pi/4$ and the least correlated orientation was $\pi/2$.

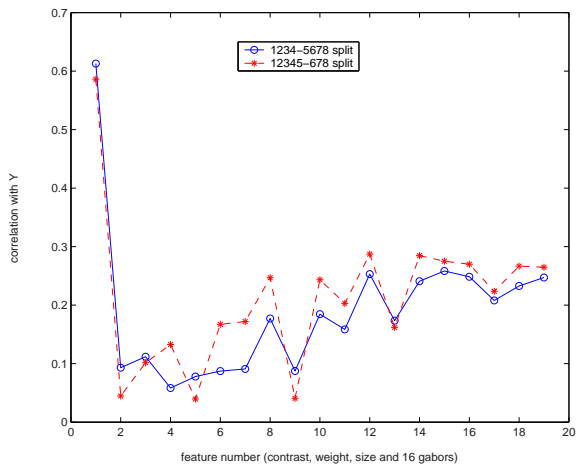


Figure 6. Correlation of features with classification

Another interesting observation is concerned with the alternative class separation 12345-678 which demonstrated improved classification performance. This indicates that these two classes are better separated than originally intended 1234-5678 split and suggests that subjects optimistically attribute barely readable text to the class of readable samples. In other words, what people deem almost readable in reality turns out to be unreadable by the very definition the same people have given for “readability”.

Our work described here does not use color to determine readability. All the texture features, contrast features, and text features used to determine readability are done on gray images and gray text. However, we expect the color information to influence the readability of text over textured backgrounds and we plan to incorporate this information in our future work.

Acknowledgments

We thank Information in Place, Inc. of Bloomington, Indiana for supporting this work.

References

- [1] R. Azuma and C. Furmanski. Evaluating label placement for augmented reality view management. In *Proceedings of the 2nd IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2003)*, pages 66–75, 2003.
- [2] C. Chubb, G. Sperling, and J. A. Solomon. Texture interactions determine perceived contrast. *Proc Natl Acad Sci U S A.*, 86(23):9631–9635, Dec 1989.
- [3] M. Clark and A. Bovik. Texture segmentation using gabor modulation/demodulation. *Pattern Recognition Letters*, 6:261–267, 1987.
- [4] A. Hill and L. Scharff. Readability of computer displays as a function of colour, saturation, and background texture. In D. Harris, editor, *Engineering psychology and cognitive ergonomics*, volume 4. Ashgate, Aldershot, UK, 1999.
- [5] A. K. Jain and F. Farrokhnia. Unsupervised texture segmentation using gabor filters. *Pattern Recognition*, 23(12):1167–1186, December 1991.
- [6] B. Julesz. Textons, the elements of texture perception, and their interactions. *Nature*, (290):91–97, 1981.
- [7] G. Legge, D. Pelli, G. Rubin, and M. Schleske. Psychophysics of reading i. normal vision. *Vision Research*, 25:239–252, 1985.
- [8] J. Malik and P. Perona. Preattentive texture discrimination with early vision mechanisms. *J. Opt. Soc. Am. A*, 7:923–932, May 1990.
- [9] N. Petkov and M. A. Westenberg. Suppression of contour perception by band-limited noise and its relation to non-classical receptive field inhibition. *Biological Cybernetics*, 88:236–246, 2003.
- [10] E. Rose, D. Breen, K. Ahlers, C. Crampton, M. Tuceryan, R. T. Whitaker, , and D. Greer. Annotating real-world objects using augmented vision. In *Proceedings of Computer Graphics International '95*, pages 357–370, Leeds, UK, 1995.
- [11] L. Scharff, A. Ahumada, and A. Hill. Discriminability measures for predicting readability. In *Human Vision and Electronic Imaging, SPIE Proc. 3644*, pages 270–277, 1999.
- [12] L. Scharff, A. Hill, and A. Ahumada. Discriminability measures for predicting readability of text on textured backgrounds. *Optics Express*, 6(4):81–91, 2000.
- [13] J. A. Solomon and D. G. Pelli. The visual filter mediating letter identification. *Nature*, 369:395–397, 1994.
- [14] M. Tuceryan and A. K. Jain. Texture analysis. In C. H. Chen, L. F. Pau, and P. S. P. Wang, editors, *The Handbook of Pattern Recognition and Computer Vision (2nd Edition)*, pages 207–248. World Scientific Publishing Co., 1998.