Video Degradation Model and Its Application to Character Recognition in e-Learning Videos

Jun Sun[1], Yutaka Katsuyama[2], and Satoshi Naoi[2]

[1] Fujitsu R&D Center Co., LTD.
Eagle Run Plaza B1003, Xiaoyun Road No.26, Chaoyang District
Beijing, 100084, P.R. China
sunjun@frdc.fujitsu.com

[2] Fujitsu Laboratories LTD.
4-1-1, Kamikodanaka Nakahara-ku, Kawasaki, 211-8588, Japan
{katsuyama,naoi.satoshi}@jp.fujitsu.com

Abstract. As the rapid popularization of digital imaging equipment, video character recognition becomes more and more important. Compared with traditional scanned document, characters in video document usually suffer from great degradation and meet trouble in recognition. Thus, a systematically study of video degradation will be very useful for video OCR. In this paper, a video degradation model is proposed to imitate the process of video character image generation. The generated character images are used to make synthetic dictionaries to improve the recognition performance of real degraded characters in e-Learning videos. Experiments on 24317 e-Learning video characters prove the effectiveness of our method.

1 Introduction

Recently, digital imaging equipments like digital camera, digital video, and video mobile phone become more and more popular. The demand of OCR for video characters also increases dramatically. For good quality images, traditional OCR can always get good performance. But many video images suffer from great degradation caused by low image resolution, small contrast, large visual angle, compression, etc. These degradations will cause problems for OCR. A systematically study of character degradation is in great need. Many degradation models have been proposed [1][2][3] recently for scanned document. In this paper, a video degradation model is presented to imitate the process of video image generation. The synthetic video character images can be used for performance evaluation and dictionary data augmentation. In our experiment, the generated images are used to make degraded dictionaries to recognize characters in e-Learning videos. Evaluation results show that these synthetic character images are very helpful for real degraded character recognition.

2 Video Degradation Model

Our video degradation model is based on perspective transformation and super sampling. As shown in Fig. 1, a clear character image is rendered into scene plane. The

scene plane is rotated around 3 axes and the character image is projected into image plane. Using the knowledge of perspective transformation, a pixel in the image plane is mapped into a quadrangle region in the scene image. So the grayscale value for the pixel is calculated as the mean grayscale value of the region in the scene plane.
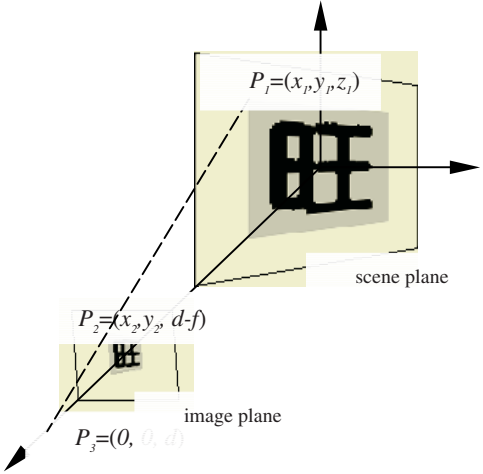

Fig. 1. Video degradation model

3   Recognition for e-Learning Video Characters

The generated characters can be used for performance evaluation and training data augmentation. Currently, we focus our application on character recognition in e-Learning videos because many characters in e-Learning video have very small size and different background color. Standard OCR can't get satisfactory recognition performance on these data since the training data is obtained from scanned clear character images. Our idea is using degraded image generated from our model to make degraded dictionary to improve the recognition rate.

1) Model simplification

For e-Learning videos, the rotation angle of scene plane is usually not very large. Also, the distance between camera center with scene plane is far larger than focal length. Thus, the influence of scene plane rotation for character image is mainly on the changes of aspect ratio. Since size normalization is one preprocessing of OCR that can counteract this influence, we set the rotation matrix to unitary matrix.

2) Contrast estimation

Histogram of every text line image is obtained and the grayscale values for the two peaks in the histogram which corresponding to background and character strokes is located. The two values are used to adjust the grayscale of the generated character image so that the contrast value is the same as the real cases.

3) Deal with blurring

In our model, we use super sampling to imitate the blurring effect caused by various reasons. The degree of blurring is controlled by the shrinking rate of the generated character size to the original character size. We don't estimate the exactly level of blurring, but instead generate several patterns for every category with different shrinking size to make several dictionaries with different degradation level.

4) Experiment and discussion

The testing data contains 24,317 binary multi-font characters extract from 22 real lecture videos, including Japanese class 1 and partly class 2 Kanji character, alphabet, digital, symbol, katakana and hiragana. The original training data come from scanned character images with 4299 categories. The binary character images are normalized to size of 64*64. Feature extraction method is contour chain code based method [4]. Simple NN classifier is used in classification.

The degraded dictionaries are made by degraded patterns with shrinking rate of 20%, 40%, 60%, 80% and 4 levels of contrast: 64, 128, 192, 255. So the total number of degraded dictionaries is 16. The recognition rates by using different combination of dictionaries are recorded in Table. 1. Dict1 means original dictionary. Dict2 includes 4 degraded dictionaries from 4 shrinking rate with contrast estimation. Dict3 includes Dict2 plus original dictionary. Dict4 includes all 16 degraded dictionaries. Dict5 includes Dict4 plus original dictionary.

Table 1. Recognition rate using different dictionary

	Dict1	Dict2	Dict3	Dict4	Dict5
Recog. rate	84.54%	90.43%	90.55%	91.49%	91.57%

There are several interesting points we learned in the experiments:

1. The performance of Dict2 is inferior to Dict4. Because for every text line, we have only 4 levels of contrast values in the dictionary, so we have to choose the dictionary with the most similar contrast and there is some estimation error. What's more, using all 17 dictionaries gets the best result, but it also is the slowest one. If we don't consider computation power, much more patterns can be generated by using more contrast level and shrinking rate, and the recognition rate is expected to be even higher.

2. The degradation model shows different result for Kanji and alphabet. Fig. 2 shows the size distribution of all Kanji and alphabet characters in testing set, and that of mis-recognized Kanji and alphabet characters. For Kanji, the distribution changes a lot for error cases. The errors mainly happen in very small size (about 10 pixels) since the structure information is heavily degraded at such scale. But for alphabet, the size distribution doesn't change so much in error cases. The main reasons for the errors include stroke missing like 'i' degraded to 'l' and shape similarity like 's' misrecognized as '5'. Our video degradation model seems more effective on character set with abundant structural information like Chinese and Japanese. This is also proved in the recognition rate: the recognition rate improves from 81.41% to 96.55% for Kanji, but only from 75.70% to 84.64% for alphabet.

3. Degradation model doesn't work well for very small size characters. At very low image resolution, the binarization looses too much valuable information for recognition. Either an improvement in binarization or feature extraction directly from grayscale image is needed.
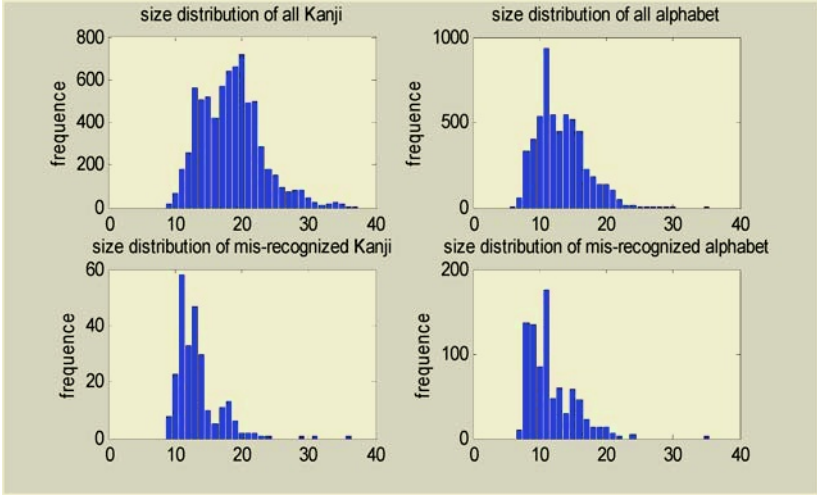

Fig. 2. Size distribution for Kanji and alphabet.

4   Conclusion

In this paper, a video degradation model is proposed to produce synthetic character patterns with different degradation level. The generated patterns are used to make degraded dictionaries to improve the recognition performance of characters in e-Learning videos. The advantages and shortcomings of the model are also discussed.

References

1. Barid, H. S.: State of the Art of Document Image Degradation Modeling. IAPR 2000 Workshop on Document Analysis Systems, (2000) 1-16
2. Sarkar, P., Nagy, P., Zhou, J., and Lopresti, D.: Spatial sampling of printed patterns. IEEE trans. PAMI, 20(3) (1998) 344-351
3. Barney Smith E. H., Qiu, X.H.: Relating statistical image differences and degradation features. Lecture Notes on Computer Science 2423 (2002) 1-12
4. Shridhar, M., Kimura, F.: Segmentation-Based Cursive Handwriting recognition. Handbook of Character Recognition and Document Image Analysis (1997) 123-156