9

Robustness Design of Industrial Strength Recognition Systems

Hiromichi Fujisawa

9.1 Characterization of Robustness

There are a few research works on robustness of OCRs for machine-printed documents [28]. An interesting work to be noted is Baird's document image defect model [2]. The types of defects in machine-printed document images were studied and a quantitative defect model was devised. To estimate the parameters of the model, "calibration" methods were studied. More importantly, it can be used to generate synthetic degraded document images with control parameters [22]. By using such synthetic images, it is hoped that we are able to evaluate the robustness of an existing OCR, and to train the OCR for such degraded images more systematically.

Another work to be mentioned is by Nartker and Taghva, who showed how OCRs from different vendors performed on documents with different qualities [1,35]. They classified 2200 pages of machine-printed English document images into five groups (G1–G5) depending on their image quality measured in terms of the page-wise recognition accuracy (correct recognition rate). Figure 9.1 shows the result of their experiments in terms of recognition accuracy of eight OCRs for five groups of document images [1]. Then, they found that the lower the accuracy for G1 is, the steeper the performance degradation curves are. It meant that the difference between good and bad OCRs is amplification of the error rate for low quality documents.

These works give some insights into how to evaluate the robustness. So, we try to apply the same approach to a more complex recognition system like a postal address recognition system. Figure 9.2 shows how the profiles of read rates[1] of a postal address recognition system look like.

[1] Here, "read rate" refers to the accept rate of the recognition system. It is often used for postal address recognition systems, where the error rates are kept around 1%.

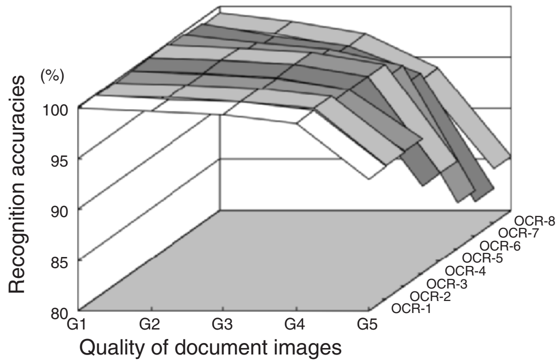

Fig. 9.1. Recognition accuracies of eight OCRs for five datasets of different image qualities. (The graph reproduced from the data shown in [35].)

The profiles shown there are for three different kinds of postal mails, i.e. bulk mail (BK), machine-printed mail (MP) and handwritten mail (HW). The x-axis represents the datasets of postal images numbered so that the corresponding read rates come into decreasing order. Actually, there are three separate sets of datasets corresponding to the kind of mails.

The former discussion on the robustness, saying that the steeper the performance degradation curve, the less robust the recognition system, suggests to measure the robustness in terms of the gradient of the curves. For simplicity, we may calculate the standard deviation (SD) of the plots of read rates instead of the gradient. So we have calculated the average and SD of the read rates for BK, MP and HW mails shown in Figure 9.2. The pairs of average and SD of these read rates are (0.86, 0.10), (0.82, 0.05) and (0.64, 0.06), respectively. When we compare the SDs so calculated, we may conclude that the system is most robust for MP mails, because the MP's standard deviation 0.05 is the smallest. However, it is counterintuitive, because BK has the highest read rate 0.86. Here we need some more discussion.

In the above discussion, the concept of "degradation" requires information about the origin, specifying "from where", to be complete in the meaning. This means that the primary performance measure of pattern recognition systems should be any representative performance such as the performance in the normal operations (or average performance measure), and the robustness is a *secondary* performance measure representing the dispersion of the performance fluctuation influenced by PIFs. Industrial strength recognition systems require not only higher average performance, but also smaller performance fluctuations.