IMPROVEMENT FOR MONO PRINTER’S INVOICE COPYING

HP INC
Improvement for Mono Printer’s Invoice Copying

A method for improving the invoice copying’s image quality is disclosed. This method involves the image binarization algorithm optimization.

Usually the paper for invoice is thin, and people use needle printer or thermal printer to print the invoice, the blue text in the duplicate invoice is not very clear. Likewise, if it is handwritten invoice, the handwritten text in the duplicate invoice is very light. Sometimes people need to copy the printed invoice’s duplicate or the handwritten invoice’s duplicate, but the original light blue text is not legible in the copy, that’s because the light blue text is subject to being ignored when binarization with a fixed threshold. Some printers allow people to adjust copy settings like the light and contrast, but it is not easy for customers to find out a good contrast value and the distortion is obvious if too extreme contrast. Now this disclosure aims to raise an optimized algorithm in printer FW to improve the invoice’s copy quality automatically.

After scanning, FW can get a gray scale image firstly, then FW starts converting the gray scale image to binary image, during this phase a self-adaptive threshold instead of the fixed threshold is employed. For each pixel, the threshold to determine whether this pixel should be mapped to zero or mapped to maximum (255) will depend on its neighborhood pixels. Firstly, the mean of neighborhood pixels’ gray scale value can be calculated (using Gaussian distribution weighted mean here), then if this pixel’s gray scale value is smaller than C (C=mean-delta), this pixel will be mapped to zero in the binary image, and if this pixel’s gray scale value is bigger than C, then this pixel will be mapped to 255 in the binary image.

The whole flow for this adaptive threshold method like below.

\[ 	ext{Mean-delta} \]

\[ 0 \rightarrow \text{Mean} \rightarrow \delta \rightarrow 255 \]

*Figure-1 How to map to binary image*
Figure-2 Adaptive threshold algorithm flow

This new algorithm advantage: It can choose the most appropriate threshold dynamically to do binarization instead of using a fixed threshold for all the pixels. So, the pixels which comprise the light blue text will have much higher chance to be treated as valid content.

Experiment:

Original document’s gray scale image, please see figure-3
Figure-3 original gray scale

Current FW: fixed threshold for binarization. Please see figure-4

![Image of invoice]

Figure-4 current FW, copy output

New adaptive threshold algorithm in experiment: please see figure-5

![Image of invoice]

Figure-5 experiment, copy output

Disclosed by Jian-Qun Zhang, HP Inc.