Analysis of quantitative indicators of Kanji text

Masaaki Sato
Fuji Xerox Co., Ltd.
Ebina-shi, Kanagawa-ken, Japan

Abstract

After printer resolution becomes higher than 600 dpi, and black-and-white printer with image enhancement technology has appeared in market, text quality has improved very much day by day. But in the case of Kanji in Asian markets including Japanese market, reproduction of Kanji text is very difficult even using image enhancement, because curved lines, slightly slanted lines, and periodic lines are included much more in structure of Kanji. This paper shows about quantitative indicators of Kanji text quality which are estimated based on visual perception. The indicators describe output resolution, fidelity of line width and edge reproduction.

Finally, this paper demonstrates different trend of important parameters observed among font types, especially noticeable for difference between Ming and Gothic fonts.

Quantitative indicators affecting text

Before we develop evaluation method for Black text quality, we need to find out attributes and factors affecting the text quality. As we generally know by the experiment of lithography technology until now, text quality is divided into 2 important indicators, which are darkness and structural reproduction.

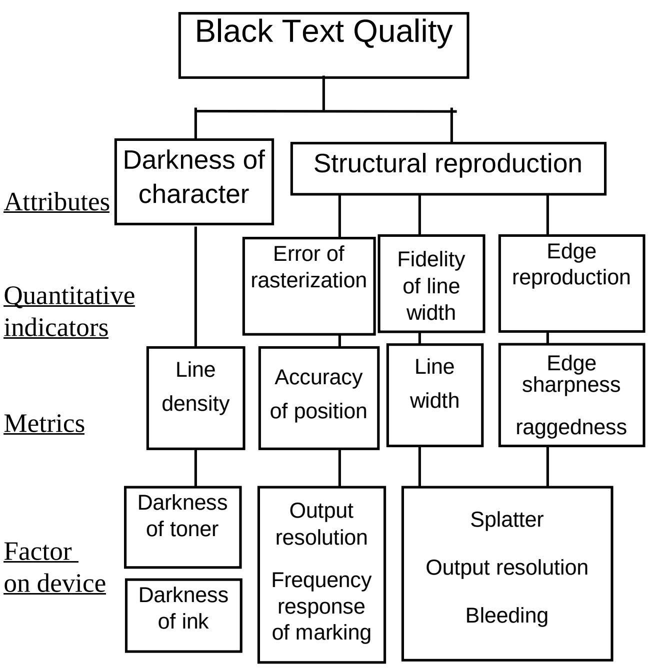

FIG.1 Image factors affecting text quality

FIG.1 is describes attributes affecting to text quality and corresponding factors. It is assumed here that darkness of text correlates with density of lines. Japan color of lithography standardizes as density at coverage 100 % solid area gives density around 1.4[1)]. In my study, density of line width around 100 micron becomes to 1.3 by measurement of micro densito-meter. In the case of xerography technology, black-and-white printer reproduces the line of around 1.0 density, and color printer reproduces the line of around 0.7 density. In the case of ink jet technology, line density drops down to 0.4 by using plain papers, but it is rare that line density drops down by using the papers recommended by ink jet manufacturers. This paper describes about quantitative analysis of structural reproduction except for density, and also indicates high priority parameters affecting to text quality.

Decision of a set of characters for analysis

Regarding the following points, I decided a set of Japanese characters for analysis.

1) Characters which are used for checking by lithography traders.
2) Characters which are used for indicating pronunciation of Kanji

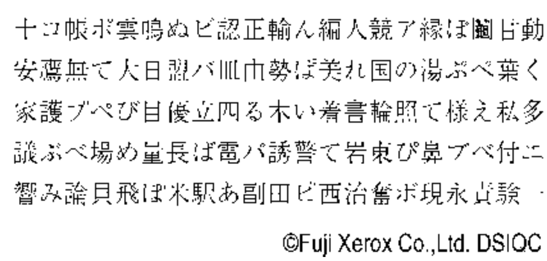

FIG.2 A set of characters for analysis of Kanji quality

In the past, Japanese fonts were classified roughly into 2 types; so-called Ming type and Gothic type.

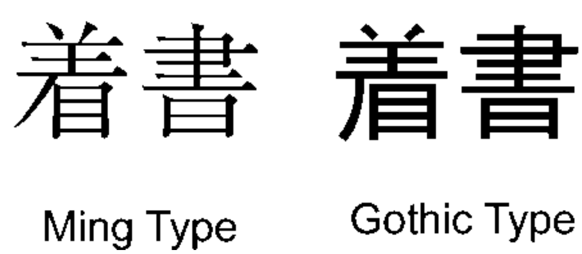

FIG.3 Ming type and Gothic type

Ming type has the characteristics of line width that the horizontal is lighter than the vertical and it has so-called "Kaeshi" or "Tome" like serifs of English fonts at the edge of lines. On the other hand, Gothic type has the characteristics of line width that the horizontal is the same as the vertical, and it has much simpler structures than Ming type(See FIG.3). Since Japanese fonts are almost classified into 2 types, this paper explains the analysis of correlation between observed score and predicted score from quantitative parameters against 6 point and 8 point which are often used in Japanese documents, both for Ming and Gothic type.

Making simulated samples for analysis

Three quantitative parameters were focused in making simulated samples for analysis text quality, which are resolution of rasterize image processor, line width and edge reproduction. The simulation process is ;

1) Resolution during rasterization was varies resolution of image setter when simulated samples are made by lithography

2) For varying line width and edge reproduction, conversion from gray level to binary image is performed after processing gaussian filtering.

The variation of line width may be caused by characterization of marking engine in addition to error of rasterization. And the variation of edge reproduction may also be caused by splatter of xerography engine or bleeding of ink jet engine. Since it is assumed that variation of edge reproduction distributes uniformly in statistic, gaussian filtering as shown FIG.4 is used.

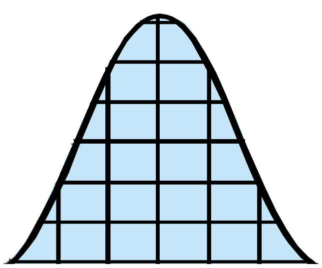

FIG.4 Distribution used to simulation of edge sharpness

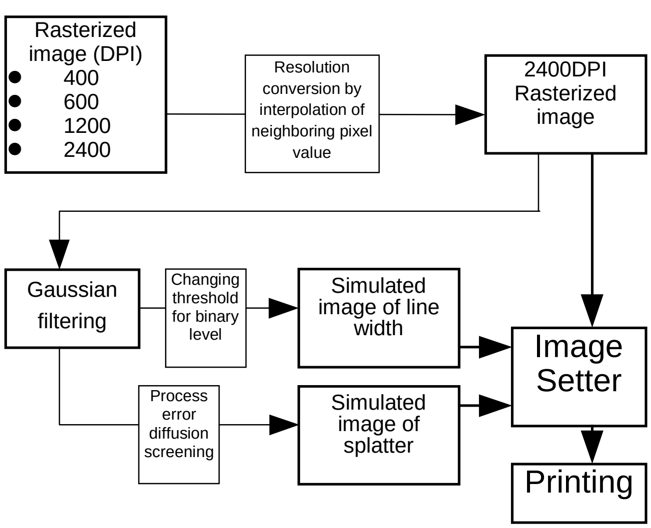

FIG.5 Image flow for making simulated image

Procedure for getting preference score

The experiment was performed for preference test based on the category as shown in FIG.6.

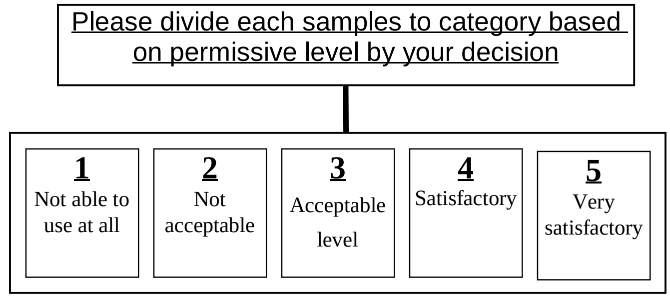

FIG.6 Definition of categories for preference test

When raters performed for preference test, we explained to raters as the follows.

- Acceptable level is purely decided by individual feeling of each raters.
- All levels are decided based on acceptable level.

As the result of preference test performed by 23 persons, each of 28 simulated samples had the score rated by 23 persons. Then average and standard deviation of 23 data were calculated for 28 samples. The average mean the score given to each samples as evaluated value, standard deviation mean variation of score rated by 23 persons. And the scores for analysis are calculated to the scale of 100 points by using equation(1).

In equation(1), ***MOS*** is the full-score points, and ***P*** is the score of 5 categories.

$$MOS = (P - 1) \times 25 \quad (1)$$

Measurement of lines for each samples

In order to find out quantitative indicators for expressing black Kanji text quality, line density and line width in line quality were measured for all samples. And I measured line widths for the horizontal and vertical as the follows by counting the number of pixel in 2400 dpi rasterized image.

Table.1 Line width in structure of Japanese character

Font types	Horizontal line width (micron)	Vertical line width (micron)
Ming 6 point	53	116
Ming 8 point	63	159
Gothic 6 point	148	148
Gothic 8 point	212	212

Since reproduction of thinner lines is most important, I selected 0.1 point(=35 micron) line as a narrow line and 0.3 point(=105 micron) line as a middle line width. Then I measured line density, line width and edge raggedness by using micro-densitometer of aperture size 10 by 500 micron. From the profile of reflectance, line density is calculated the average of 3 positions at the peak of profile, line width is calculated by the distance between 50 % reflectance, and raggedness is calculated by measuring root mean square of edge profile. These methods are generally used for measurement from the past.

Analysis of correlation using statistics

As the result of statistical analysis using principal component, FIG.7 was obtained. This figure represents regression values which are calculated for each indicator; resolution, square of resolution, line width of 0.1 point, line width of 0.3 point and raggedness.

Resolution of output : ***DPMM*** (dot per millimeter)
Square of resolution : ***DPMM**2***
Line width of 0.1 point : ***L1*** (micron)
Line width of 0.3 point : ***L3*** (micron)
Raggedness : ***R*** (micron)

When the function is defined as the above, full-score points ***MOS*** is described as the follows.

MOS =a** x **DPMM** + **b** x **DPMM**2** + **c** x **L1** + **d** x **L3** + **e** x **R
- *Equation(2)*

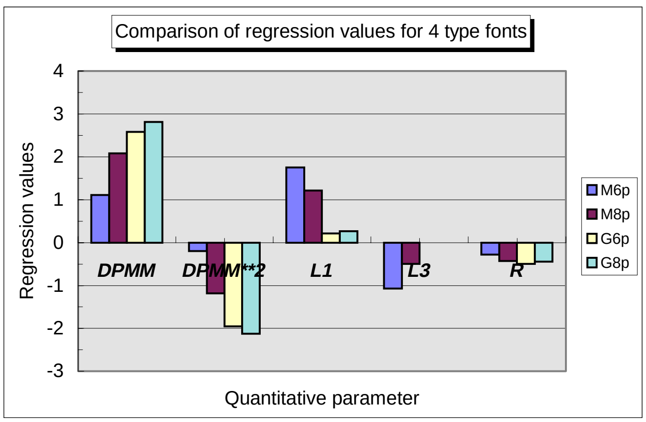

FIG.7 Regression values of 5 indicators

As seen in FIG.7, there are 5 indicators for 4 font types. Ming type has the characteristics which are linear by correlated to resolution and line width. On the other hand, Gothic type has no correlation with the line width, but it has the second-order correlation with the resolution.

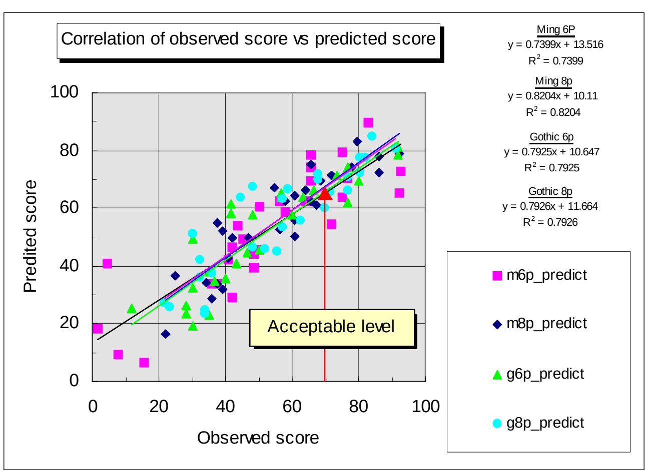

FIG.8 Correlation between observed scores and predicted scores

FIG.8 represents the correlation between predicted scores and observed scores given by preference test. This figure indicates that black text quality is explained 80 % for 4 type fonts using 5 parameters in the case of Ming, 4 parameters in the case of Gothic. This correlation may be improved by including raggedness of slant lines.

Conclusion

For Kanji text, I analyzed correlation between preference score and predicted score. The quantitative parameters are as the follows.

1) In the case of Ming type characters, resolution and the fidelity of line width are important parameters for reproducing good text printing. This trend is clear as character size is small. For example, it is clearer for 6 point size than 8 point size.

2) In the case of Gothic type characters, resolution is really important parameter comparing with other parameters. Gothic type has the second-order correlation with the resolution.

3) There is a trend that text quality is improved when a narrow line like 0.1 point width becomes thicker, and a middle line like 0.3 point width becomes narrower.

Maybe according to include the parameter for measurement of slant lines, contribution values will become to higher. It is important to know in measuring the lines the structure, angle, width etc. in order to find out important parameters for quantifying text quality.

References

1. ANSI/CGAT 1993 "Specifications for graphic arts printing - Type 1".