Journal of Korea TAPPI
Vol.40. No.5, 2008
Printed in Korea

A New Characterizing Method for Recycled Paper and the Application of Image Segmentation on the Measurement Sub-visible Dirt

Shijin Dong[1], Haitao Zhang[1], Xuejun Cui[2], Junfeng Li[3], and Hongyan Wang[3†]

(Received December 1, 2008: Accepted December 15, 2008)

ABSTRACT

The paper established a new method for fast measurement and characterizing of sub-visible dirt of recycled paper which is too small to be seen with naked eye. This method provided a new way for the evaluation of recycled paper that is hard to be characterized by the conventional method. Two effective thresholding algorithms HA and SDA were compared and their applicable conditions were discussed. Results showed that the HA could be used for un-printed paper while SDA is suited for recycled papers. The gloss of paper samples was measured and the relation between gloss and sub-visible dirt was investigated. The significant effect of this method for characterizing and comparing paper was exhibited. The experiment results indicated that sub-visible dirt measuring method could be a complementariness of the conventional methods.

Keywords : *recycled paper, sub-visible dirt, image analysis, threshold, gloss*

1. Introduction

Dirt speck is defined as any object embedded in paper which shows a different color with the rest of paper sheet (background). Dirt speck is one of the most valuable optical characteristics to assess the de-inking efficiency of recycled paper. An important prerequisite for using the recycled paper again is to reduce the amount of dirt specks to a satisfying low level (1). Dirt is also an important property for un-printed paper Therefore, accurate measurement of dirt count and dirt area ratio is absolutely necessary in paper and pulp industry (2,3,4).

Generally speaking, dirts are classified into 3 categories. The speck covering an area that is equal or larger than 0.04mm^2 is defined as visible dirts, the one whose equivalent physical diameter (EPD) is within the range of 8-160 micrometers is named as sub-visible dirts, and the rests are called transitional dirts (3). In the traditional method, the operators compare each dirt speck of sample with a group of dots in a standard chart then choose the dot which perfectly

• Qianjin Street 2699#, Department of Chemistry, Jilin University, Changchun 130012, P.R. China
1 Doctor student
2 lecturer
3 professor
† Corresponding Authour: E-mail; wang_hy@jlu.edu.cn

matches the dirt to determine its real size (4,5).

There are many shortcomings of the conventional standard method such as time-consuming, and highly dependant on subjective judgment by human so that the measuring result will vary from person to person and sometimes the result may have a bad reproducibility (5).

The shade on the paper is an important appearance property influencing the marketability of recycled paper. Sub-visible particles of residual ink which is too small to be seen with naked eye could still produce shade on the paper sheet so that the level of sub-visible dirt can impact its usefulness in application (6). The conventional method mainly takes the visible dirt into consideration, but it will fail to work when measuring sub-visible dirt. For the conventional method can not satisfy the requirements of characterizing high-quality paper, it become necessary to develop a more sensitive and effective measuring method for paper evaluation. Similar considerations may be appropriate for recycled paper and pulp (7).

The published literatures seldom discuss the specificity and the adaptability of thresholding algorithms and the relation between sub-visible dirt with other properties. In this paper, the specificity and the adaptability of thresholding algorithms are investigated and the relation between sub-visible dirt and gloss has been discussed. With the development of computer and the high resolution scanner, it becomes easier to measure the sub-visible dirt by image analysis technology (8).

2. Experiment

2.1 Apparatus

This sub-visible dirt measuring system is mainly composed of two parts. One is image acquisition device that consists of a CCD sensor matrix and illuminating device .The sensor has 256 gray levels sensitivity and a high physical pixel resolution. The light emitted from illuminating device is diffused and un-polarized. The spectrum of the light is concentrated between 650nm-710 nm. The other part is called image processing device that includes a computer and a self-programmed software.

2.2 Testing procedure

2.2.1 Sample process

Carefully brush away any loose surface dirt specks. Make sure measuring specimens do not contain smudge marks and wrinkles that could dramatically affect the results.

2.2.2 Image scanning

Image samples are carefully scanned by image acquisition device. For getting the most representative images, three areas should be measured to reach a total area of 12 square inches for each sample to be measured.

2.2.3 Image analysis

Image detectors digitize the signal of each picture element (pixel) into discrete values which are called gray levels. According to the theory of image analysis, dirt specks are measured by self-programmed software.

2.2.4 Data report

When the total measuring process is finished, the system could report a detailed result report including total dirt speck amount, dirt count per square centimeter, dirt area ratio, the specific position of dirt specks, histogram of size distribution and histogram of gray-value distribution.

3. Theory of Image Analysis

3.1 Image segmentation

The aim of the segmentation process is the detection and classification of the dirt specks presented in the original image (Fig. 1(a)). Let G= {0, 1, 2⋯..255} be the set of 256 gray levels. The image function of gray value at pixel (x, y) is denoted by $f(x, y) \in G$. After comparing gray value of each pixel with a given threshold T, the function $f(x, y)$ will be transformed into a binary function $f'(x, y)$. The explicit expression is as follow:

$$f'(x,y)=\begin{cases} d, f(x,y) \le T \\ b, f(x,y) > T \end{cases} \qquad [1]$$

According to formula [1], the original image is converted into a binary image (Fig. 1(b)) automatically. This binary image presents white background when gray value of pixel is over threshold T and a group of dark regions which are candidates for dirt specks if their gray values are lower (darker) than T. Thus the whole binary image has been divided into two parts.

3.2 Dirts recognition

Fig. 2 is a schematic image of amplified binary image for dirt recognition. In Fig. 2, the letters a, b, c, d and e stand for dirt specks which are consisted of dark pixels, and the unit grids such as C3 and H4 stand for the pixels whose side length is set as 20μm. By judging the relation among adjacent dark pixels (connecting horizontally or vertically), the local spacial relations would be clearly investigated and the composition of each dirt specks is determined. There are 5 specks in Fig. 2. Speck a and speck b are not recognized as one speck because the adjacent pixels F5 and G6 are not connected horizontally and vertically but diagonally.

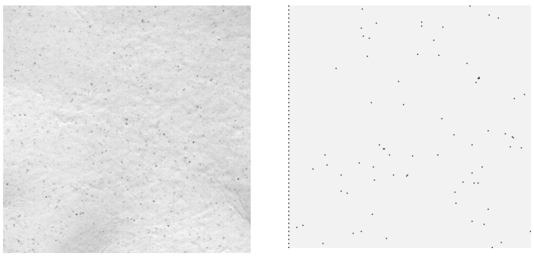

Fig. 1. (a) Original image (b) Converted binary image.

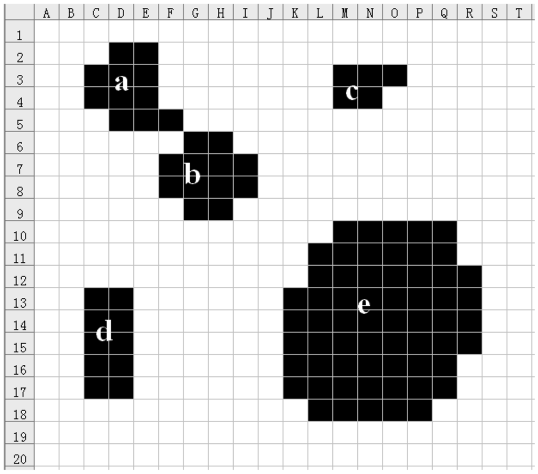

Fig. 2. Schematic image for dirts recognition

After the spacial relations of each pair of adjacent pixels in binary image are investigated, all the dimensional properties of dirt specks such as size, shape, position are determined. For the shape of ink speck is almost irregular, equivalent physical diameter (EPD) is used to characterize particle size. The EPD of a speck is calculated according to:

$$EPD = 2 \times \left(speck\ area \div \pi\right)^{0.5} \qquad [2]$$

When EPD of a particle fall into the range of 8-160μm, the particle will be marked as sub-visible particle. The EPD of dirt speck E in Fig. 2 is 180.5 (E= $2\times(61\times202/3.14)^{1/2}$=180.5) by formula [2]. For its EPD is over 160μm, speck E can not be classified as sub-visible dirt and is eliminated for the result calculation.

4. Results and Discussion

4.1 Thresholding algorithm

As far as sub-visible dirt measuring is concerned, our purpose is to segment the dirt from the rest of paper. Thresholding is an important factor for measuring dirt specks accurately. The goal of using thresholding is to divide an image into two or more regions (10). There are lots of available thresholding algorithms in image analysis, but most of thresholding algorithms do not suit for sub-visible dirt measuring (10, 11). Standard deviation algorithm(SDA) and hierarchical algorithm (HA) are selected to be applied in sub-visible dirt measuring(12).

Standard deviation algorithm is an effective threshold method. In this method, we only take 1-dimentional gray value distribution into consideration. According to the definition of dirt in image analysis, pixels corresponding to ink specks are detected if their gray values are lower than the threshold. The threshold is calculated as the following equation:

$$Treshold = \overline{X} - 2\sqrt{\frac{1}{N}\cdot\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2} \qquad [3]$$

where $\overline{X}$ is the mean gray value of the paper image. $\sqrt{\frac{1}{N}\cdot\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$ stands for the standard deviation of 1-dimentional gray value distribution of paper image.

Fig. 3 is the histogram of gray value distribution of Fig. 1(a). In Fig. 3, we can find that the gray values of pixels are mainly concentrated in the range from 140 to 200. Base on the data of Fig. 3, all the information for computing standard deviation could be found according to formula [3],

$$Threshold = 180 - 2\sqrt{\frac{1}{N}\cdot\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$$

$$=180-2*15.14=149.72$$

After rounding, the value 150 is selected as the threshold for the image sample shown in Fig. 1 by SDA algorithm.

Hierarchical algorithm is a complicated dynamic threshold method. The purpose of this algorithm is to find a threshold which will demonstrate the largest contrast of dirt specks and the background.

When giving a threshold T, the segmentation is made automatically and a group of dark regions PT is found as the candidates of dirt specks.

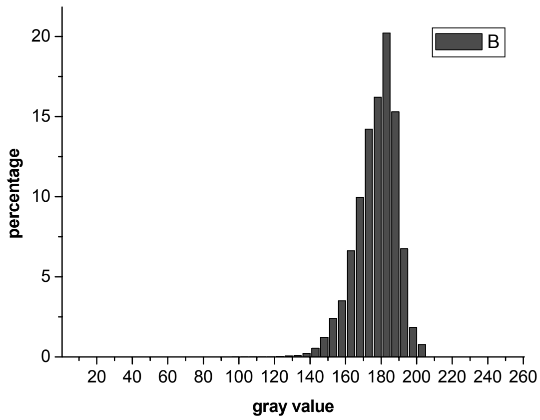

Fig. 3. Histogram of 1-dimentional gray value distribution

$$PT = \{ S_1^t\ S_2^t\ S_n^t \}, \qquad [4]$$

Considering the T changing from 0 to 255, a large set Φ including 256 groups of dark regions is obtained.

$$\Phi = \{P0,P1,P2,\cdots.P255\} = (S_1^0 , S_2^0\ \ S_{n1}^t\ , S_1^1\ S_2^1\ , S_1^{255}\ S_2^{255} \ldots S_{n255}^{255}) \qquad [5]$$

PT stands for a group of dirt candidates which are segmented by threshold T(seeFig. 4), S_i^t stands for one dirt speck candidates, and Φ stands for the universal set of all sub dirt candidates groups. In all

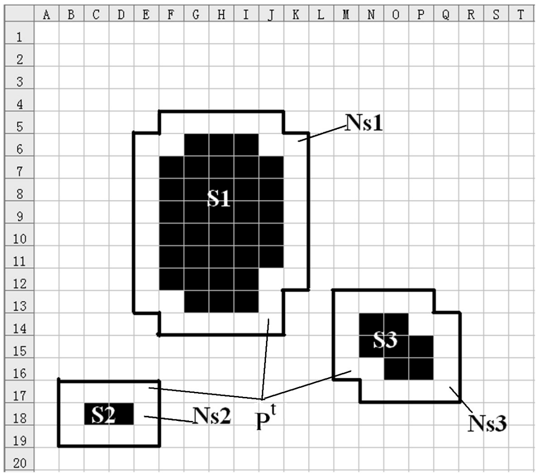

Fig. 4. Schematic image of set P^T and Ns.

256 kinds of segmentation, the optimum result will be determined, which is a subset of the large set Φ.

Fig. 4 is a schematic image for showing relations of Ns and S_n^1 in HA. Ns is defined as the smallest white region involving the dark region S_n^1. The gray-value contrast C_s^T is defined by

$$C_s^T = G_n^T - G_s^T \quad [6]$$

G_s^T is the mean gray-value of the dark region, and G_n^T is the mean gray value of all Ns. Threshold T is selected when the C_s^T reach its highest value and this segmentation will represent the most significant contrast between dirts and white background.

The same paper sample (Fig. 1 (a)) is selected to exemplify this algorithm. Threshold T is generally thought to locate in set G= {100,101,102···, 200}. Firstly, when setting 100 as the threshold, a segmentation result is automatically made sure that two dirt candidates covering 0.002% of the total surface area of sample are found. After that, the smallest involving regions Ns1 and Ns2 are easily determined. When the values of G_n^{100} and G_s^{100} are calculated, we can easily compute the value of C_s^{100} according to Formula [6] ($C_s^{100} = G_n^{100} - G_s^{100}$ = 117.1-94 = 23.1). As the threshold T changing from 101 to 200, the rest values could be calculated in the same way. All G_s^T Nsand G_n^T valuescouldbeseenin Fig 5(a).

Fig. 5(a) shows the change trends of Gs and Gn when increasing threshold T, and the data of Fig. 5 (b) shows the relation between C_s^T which is calculated by the data of Fig. 5(a) according to formula [6] and threshold T.

As shown in Fig. 5(b), function Cs reaches its highest value 26.8 when threshold T is 156, which means this segmentation can represent the most significant contrast for this paper sample. Therefore, the gray value 156 is selected to be the threshold in this hierarchical algorithm. The calculated thresholds by HA (156) and SDA (150) do not reveal significant difference for this sample, but there may be large differences for other samples.

4.2 Optimum selection for two thresholding algorithms

The standard deviation algorithm and the hierarchical algorithm are both successful threshold methods to solve the problems of image segmentation for the inspection of sub-visible dirt. They both take the original information of the image such as gray value into consideration. The former mainly considers the total 1-dimentional gray value distribution of the original image of paper sample and threshold is selected by statistics calculation. The latter investigates all the possible conditions of the threshold, and the threshold which produces the largest contrast between

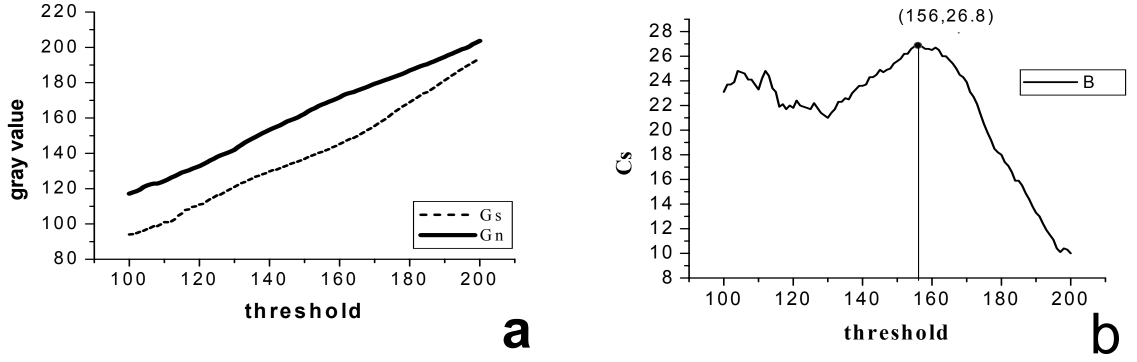

Fig. 5. (a) The relation of G_s^T and G_n^T with threshold T(b) relation between C_s^T and threshold T.

Table 1. The comparison of two threshold algorithms

	1#			2#			3#		
	Threshold	Dirt count	RE	Threshold	Dirt count	RE	Threshold	Dirt count	RE
Verification Threshold	154	4.90		157	4.11		138	0.76	
SDA	156	5.24	7	158	4.26	3.6	169	1.98	160.5
HA	150	4.32	-11.7	155	3.91	-4.87	136	0.73	-4.3
	4#			5#			6#		
Verification Threshold	140	0.31		138	1.96		141	1.36	
SDA	170	1.45	367.7	168	3.32	69.3	173	2.78	104.4
HA	134	0.28	9.68	137	1.93	-1.53	140	1.34	-1.47

1# and 2# are recycled paper；3#, 4# are high quality un-printed paper；5#,6# are low quality un-printed paper. *The verification thresholds are obtained through human observation. After paper samples are amplified for more than 10 times by the high-resolution CCD sensor, the sub-visible dirt specks became visible to naked eye. The large sample images are carefully inspected by at least 5 operators. The dirt specks counted by naked eye could be converted into corresponding threshold by self- programmed software.

dirt specks and the background is selected. The results of two algorithms of six paper samples are compared as shown in Table1.

As we can see in Table 1, both SDA and HA algorithms could give out different segmentation thresholds when processing different paper images so that they both belong to dynamic threshold methods. However, there existed significantly differences between these two methods when measuring the same paper image. According to the verification threshold, both SDA and HA algorithms had got relative precise results of sample 1 and sample 2 which are recycled paper samples and the precision of SDA algorithm is a little higher than HA. The SDA algorithm failed to get the reasonable thresholds For 3#, 4#, 5# and 6 from two kinds of un-printed papers while the results of HA are quite acceptable. That is because when measuring un-printed paper that has little sub-visible dirt specks, there are not enough dark pixels of dirt specks for SDA algorithm to notice. Thus the influences of sub-visible dirt specks are seriously weakened by the large area ratio of white background when statistic calculation is carried out. However, the HA algorithm only paid attention to the smallest surrounding area, which would not be easy to be affected by the amount of dirt specks. On the contrary, when measuring the recycled paper sample which has large amount of sub-visible dirts, the information hiding in the widely distributed 1-dimentional gray value is more easily to be caught by SDA, but the large amount dirt and the complicated shape of the specks on the recycled paper may reduce the accuracy of HA algorithm.

In a word, HA algorithm is suggested to be used for un-printed paper and SDA algorithm could be used to measure recycled paper samples.

4.3 Functions of the measuring system

A recycled-paper sample is chosen as example to demonstrate the functions of sub-visible measuring system. The sample is made in accordance with Technical Association of the Pulp and Paper Industry (TAPPI) standard T400. The brightness and visible dirt counts of the samples had been measured before.

Measuring conditions is that brightness is 72.16, mean gray-value is 180, threshold by SDA is 154. The measuring system can give out three images, namely, the original image (Fig. 6(a)), the binary image (Fig. 6(b)) locating all the sub-visible dirts and the

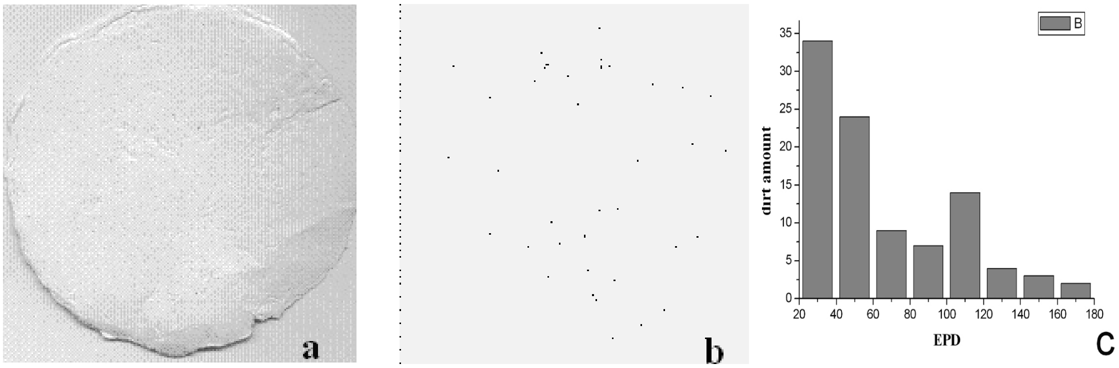

Fig. 6. (a) Original image, (b) binary image for sub-visible dirt, (c) histogram of EDP distribution.

histogram of EDP distribution (Fig. 6(c)). Besides, it can give out data report in detail including amount of dirt, dirt count, dirt area ratio and mean EPD of dirt specks. For the sample of Fig 6 (a) the amount of dirt, dirt count, dirt area ratio and mean EPD of dirt specks are 112, 1.86/cm2,0.0091percent,79.5μm respectively.

4.4 Relation between sub-visible dirt speck and gloss

In this paper, smoothness of paper samples is characterized by the 85o specular surface gloss. Gloss is another important optical character of paper, which could affect the appearance of paper and the printing properties. There are a lot of factors that affect gloss, such as coating component, solid content of coating, calendering and so on. In this paper, we only investigated the change of the gloss caused by sub-visible dirts using 24 pieces. The 24 pieces of paper include recycled papers and two kinds of un-print paper and each group of paper have the same coating environments respectively in order to eliminate the influences caused by other factors.

As is shown in Fig. 7, the gloss of the paper samples significantly declined when the dirt count of paper samples increased and the relationships between dirt area ratio and the gloss is quite similar. The reasons could be explained that on one hand, the sub-visible dirts could destruct the continuity and the smoothness of the coating on the paper surface, which will weaken the specular reflection effects; on the other hand, the dirt specks could aggravate the diffuse reflection effect which will also decrease the gloss of paper.

4.5 Complementarity for the conventional methods

As is shown in Table 2, sample 1 and sample 2 almost have the same brightness. The difference of visible-dirt count between sample 1 and sample 2 is quite negligible. Considering the conventional judging standards only, we can easily draw the conclusion from the testing data that the quality of two paper samples is quite similar. However, we can find more shades in sample 2 when observing by naked eye only. Using the sub-visible dirt measuring system, significant differences could be distinguished between the two samples. The sub-visible particles of sample 2 are nearly one and a half as much as that of sample 1 and the dirt area ratio of sample 2 is approximate 70% more than sample 1. Therefore the conclusion could be drawn that sample 1 is significantly better than sample 2 according to this new characterizing method. We can conclude that the sub-visible dirt measuring method could be complementa when the conventional methods including brightness and visible dirt failed to evaluate paper quality.

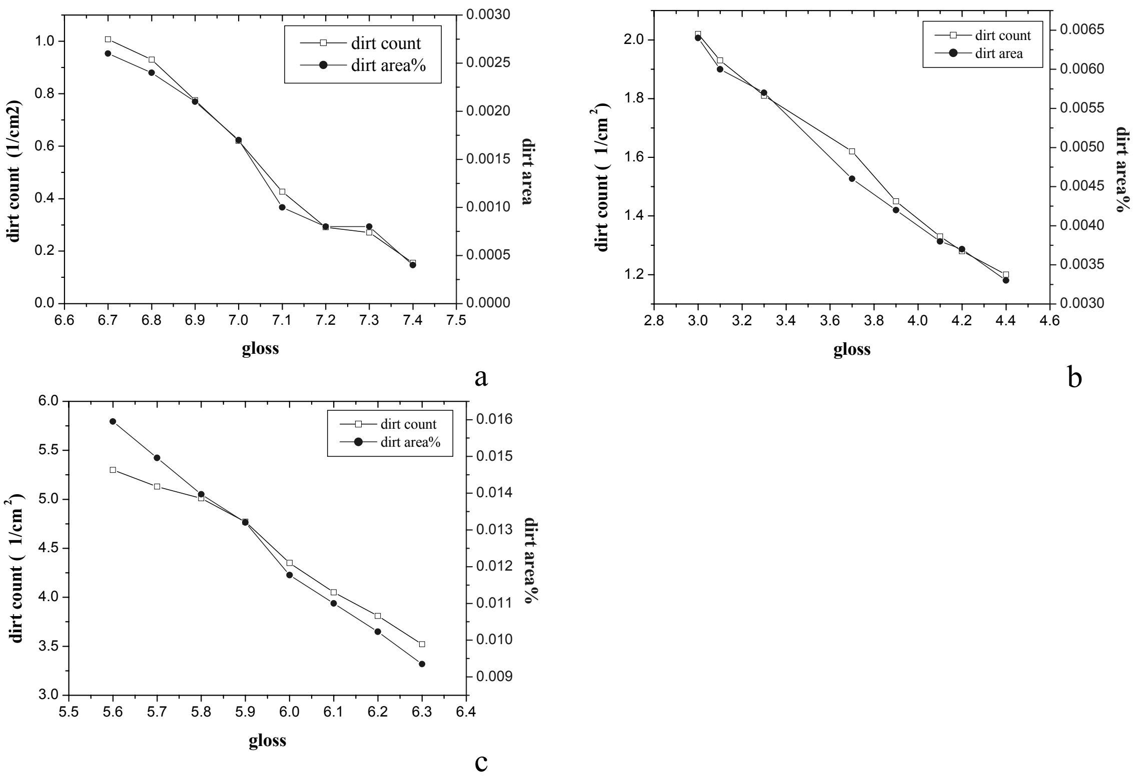

Fig. 7. Relation between gloss and sub-visible dirt of (a) high quality un-printed paper (b) low quality un-printed paper (c) recycled paper.

Table 2. The optical properties of two recycled paper samples

	Brightness	Visible-dirt count*	Sub-visible dirt count	Sub-visible Dirt area ratio	Gloss
Sample 1	72.16	1.38/cm^2	1.86/cm^2	0.0091	4.1
Sample 2	72.39	1.34/cm^2	2.70/cm^2	0.0145	3.1

5. Conclusions

It has been proved by experiment that both standard deviation algorithm and hierarchical algorithm are available to determine proper thresholds for measuring sub-visible dirt. HA algorithm is suggested to be used for un-printed paper and SDA algorithm could be used to measure recycled paper samples. This system could fast and automatically measure the sub-visible dirts of paper samples and it can give out a detailed result report including three images and the data of total dirt amount, dirt counts, dirt area ratio and mean EPD, which is quite valuable and effective for characterizing recycled paper and un-printed paper. The sub-visible dirt measuring method can be used as a complementarity of conventional method.

Acknowledgement

The author greatly appreciated the support of Hewlett-Packard (HP) Company.

Literature Cited

1. Voith, S.and Stoffaifbereitung, R., State of the art and new technologies in flotation deinking International Journal of Mineral Processing, 563: 17-333(1999).
2. U. Viestur, M. Leite M. and Eisimonte, Biological deinking technology for the recycling of waste papers, Bioresour.Technol.67; 255-267 (1999).
3. Jordan, D. B. and Nguyen, N.G., Dirt Counting with Microcomputers, Journal of pulp and Paper Science May; 73-77 (1985).
4. Faltas, R., Papadopoulos, P., Vanderhoek, N. and Reich, M., Measurement of dirt count using a scanner-based image analysis system, Apptia Journal. 53; 476-482(2000).
5. F. Duarte,H.Araujo, and A. Dourado, An automatic system for dirt in pulp inspection using hierarchical image segmentation, Computer & industrial engineering, 37: 343-346 (1999).
6. Kawabata, Y., Introduction of DCW: Fast and accurate measurement of on-line dirt count system, Japan TAPPI Journal, 56; 34-39 (2002).
7. Glowacki, J. J., Dirt analysis system: creating a new standard, Pulp& Paper, 70; 143-144, (1996).
8. Torniainen, J. E., Derhjelm.S.A.S, and Youd. G. Result of automatic dirt counting using transmiteed light, TAPPI Journal, 82; 194-197 (1999).
9. M. Sezgin and B. Sankur,Survey over image thresholding techniques and quantitative performance evaluation, J Electron Imag, 13; 146 – 168 (2004).
10. Keivan, T., Saed, S., Stephen, T. and Balke,A., image thresholding for real-time particle monitoring Imaging, Syst Technol, 16; 9-14 (2006).
11. Wang, Q. and Chi, Z.R.,Image thresholding by Maximizing the Index of Nonfuzziness of the 2-D grayscale histogram, Computer vision and image understanding 85; 100-116(2002).
12. W. T., Wen, C. H., Yang, C. W., A fast two-dimensional entropic thresholding algorithm, Pattern Recognition. 27; 885 – 893(1994).