EXPERIMENTAL COMPARISON OF IMAGE THRESHOLDING METHODS FOR DEFECT DETECTION IN THE PAPERMAKING PROCESS

Luca Ceccarelli[(a)], Francesco Bianconi[(b)], Stefano A. Saetta[(c)], Antonio Fernández [(d)] and Valentina Caldarelli [(e)]

[(a,b,c,e)] Department of Industrial Engineering
Università degli Studi di Perugia
Via G. Duranti, 67 – 06125 Perugia (ITALY)

[(d)] Department of Engineering Design
Universidade de Vigo
Campus Universitario – 36310 Vigo (SPAIN)

[(a)] lucap3600@gmail.com, [(b)] bianco@ieee.org, [(c)] stefano.saetta@unipg.it, [(d)] antfdez@uvigo.es,
[(e)] vale.caldarelli85@alice.it;

ABSTRACT
Automatic detection and assessment of dirt particles in pulp and paper plays a pivotal role in the papermaking industry. Traditional visual inspection by human operators is giving the way to machine vision, which provides many potential advantages in terms of speed, accuracy and repeatability. Such systems make use of image processing algorithms which aim at separating paper and pulp impurities from the background. The most common approach is based on image thresholding, which consists of determining a set of intensity values that split an image into one or more classes, each representing either the background (i.e.: an area with no defects) or an area with some types of contraries. In this paper we present a quantitative experimental evaluation of four image thresholding methods (i.e.: Otsu's, Kapur's, Kittler's and Yen's) for dirt analysis in paper. The results show that Kittler's method is the most stable and reliable for this task.

Keywords: machine vision, image thresholding, paper, quality assessment

1. INTRODUCTION
Product and process control through machine vision has been receiving increasing attention during the last years. Applications in the industry now cover many produts, such as textile (Carfagni *et al*. 2005), wood (Bianconi *et al*. 2013), ceramics (Kukkonen *et al*. 2001), natural stone (Bianconi *et al*. 2012), food (Furferi *et al*. 2010) and vehicles (Furferi *et al*. 2013) – to cite some.

In the papermaking industry, machine vision proved effective in a number of problems, including printability analysis (Kalviainen et al. 2003); control of stripes and holes (Navarrete et al. 2003); assessment of the coating layer (Prykary et al. 2010); curl estimation (Synnergren *et al*. 2001), analysis of microstructural changes (Sjödahl and Larsson 2004) and automatic segregation of waste paper for recycling (Rahman *et al*. 2011). Among them, dirt inspection has always played a central role, due to the strong effect that such defects have on the quality of the final product. An excessive presence of contraries and impurities may cause the pulp or paper to be off-specification, with negative consequences for the producer. The detection and characterization of contraries is also a crucial step to track down and remove (or at least reduce) the source of impurities in the production process. The potential advantages are: a more efficient use of materials and energy, and a reduction of chemicals in the bleaching phase, with beneficial effects on the environment.

Various prototypes and systems for automatic dirt analysis and counting have been described in the literature – for an overview of methods see the works of Torniainen *et al*. (1999); Corscadden and Trepanier (2006) and Ricard *et al*. 2012. From a technical standpoint, the detection of whatever type of particles in pulp and paper can be viewed as an image segmentation process aiming at separating the contraries (foreground) from the rest of the product (background). Most commonly, defects are dark spots on a bright area; but in some types of paper they may well be both brighter and darker than the background. In the paper recycling process, for instance, we expect to find not only traces of toner and wood particles – which tend to be darker than the background – but also stickies – which are likely to be brighter than the background. In either case the segmentation process requires determining one or more intensity values (thresholds) for separating whatever type of defects from the background in the correct way. In this paper we present a quantitative experimental evaluation of four image thresholding methods that can be used for this task. Of each method we consider both the standard single-threshold version, which can be used when defects are all darker or brighter than the background, and the more challenging double-threshold version, which is required when defects are either darker or brighter than the background. To assess the accuracy of the methods in a quantitative way, we compare the results of automatic segmentation against a 'ground truth' of contraries manually generated and cross-validated by two human experts.

In the remainder of the paper we first give an account of the materials and image acquisition devices used in our study, followed by a description of the thresholding methods included in the comparison. Then we outline the experimental set-up, summarize the main results of the study and conclude the paper with some final considerations.

2. MATERIALS

We considered two different classes of recycled paper. According to their appearance, we conventionally refer to the two classes as 'White' and 'Brown' (see Fig. 1).

(a) Sample of class 'White' (b) Ground truth

(c) Sample of class 'Brown' (d) Ground truth – defects darker than background (e) Ground truth – defects brighter than background

Figure 1 (a) Sample of class 'White' and (b) the corresponding ground truth; (c) sample of class 'Brown', and the corresponding ground truth for defects (d) darker and (e) brighter than the background.

Each class includes three sub-classes of different density. The characteristics of each class are reported in Tab. 1.

Table 1 Summary table of the materials used in the experiments.

Class	Sub-class	Density (g/m^2)	No. of samples	Image resolution
White	W1	137	20	400 × 400
	W2	154		
	W3	174		
Brown	B1	154	20	400 × 400
	B2	137		
	B3	137		

For each class we obtained a set of 20 specimens and acquired them through the imaging system described in Sec 2.1. Samples of class White present only defects that are darker than the background. We therefore used this set of samples to test the single-threshold version of the algorithms. By contrast, samples of class Brown show defects that are either brighter or darker than the background (see Fig. 1). Their analysis therefore requires the two-threshold version. The 'true' location and extension of the defects of each sample ('ground truth') have been manually determined and cross-validated by two skilled operators.

2.1. Image acquisition

The imaging system used in the experiments (Fig. 2) is composed of the following parts: one dome illuminator (Monster Dome Light 18.25"), one industrial CMOS camera, one support for the camera, one base and one slot to accommodate the paper specimen. The imaging apparatus can operate either by transmitted or reflected light. The lens can be selected to suit the specific application needs. In this activity we used a 12 mm fixed focal length objective (Pentax H1214-M). The whole imaging system provides a spatial resolution of approximately 370 dpi. The acquisition was carried out in reflected light mode.

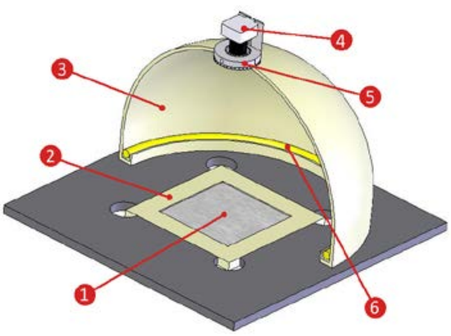

Figure 2 The image acquisition system: 1) paper sample; 2) slot; 3) hemispherical Lambertian surface; 4) camera, 5) rotatable support and 6) illumination ring.

3. METHODS

The problem of segmenting the image of a paper specimen through thresholding consists of determining a set of intensity values $G = \{G_0,\ldots,G_C\}$ that splits the image into a set of C classes, each corresponding to intensity values $i \in [G_{c-1},G_c[$. One of these classes will represent the background of the product; the others different classes of impurities. The case $C = 2$ is the most common, and occurs whenever we need to detect dark/blackish particles on a bright background (but the reverse may also occur). The cases $C > 2$ represent more complex scenarios, in which we have to look for more than one class of impurities. As we mentioned in the preceding sections, we limit our investigation to the cases *C = 2* and *C = 3*. The determination of a proper set of thresholds for a given image (thresholding) has been studied extensively in literature, and several methods exist – for a comprehensive review on the subject see the work of Sezgin and Sankur (2004). Nonetheless, no quantitative data are available, in the literature, as for the effectiveness of the methods for dirt analysis in pulp and paper. In this study we considered four parameter-free, computationally light and easy to implement methods. They are: Kapur's, Kittler-Illingworth's, Otsu's and Yen's. Here below we summarize the basics of each method. References are provided for the benefit of readers interested in the technicalities. All methods take as input the first-order probability distribution (histogram) of gray-levels; we therefore assume that the original images are converted to grayscale before processing. In Equations 1-4 we preliminarily define the weight, mean, standard deviation, entropy and correlation of each *c*-th class:

Weight $$\omega_c = \sum_{i=G_c}^{G_{c+1}-1} p_i \tag{1}$$

Mean $$\mu_c = \sum_{i=G_c}^{G_{c+1}-1} \frac{ip_i}{\omega_c} \tag{2}$$

Variance $$\sigma_c = \sum_{i=G_c}^{G_{c+1}-1} (1-\mu_c)^2 \frac{p_i}{\omega_c} \tag{3}$$

Entropy $$E_n = \sum_{i=G_c}^{G_{c+1}-1} p_i \log_2 \frac{1}{p_i} \tag{4}$$

Correlation $$CR_n = \log_2 \sum_{i=G_c}^{G_{c+1}-1} \left(\frac{\omega_c}{p_i}\right)^2 \tag{5}$$

where p_i is the probability of the *i*-th grey-value.

3.1. Kapur

In Kapur's method (Kapur *et al*. 1985) the set of optimal thresholds, indicated as $\overline{\mathbf{G}}$ in the following equations, are the intensity levels that maximize the sum of the entropy of each class (Eq. 6). For this reason the procedure is also referred to as *maximum entropy criterion*.

$$\overline{\mathbf{G}}_{\text{Kapur}} = \underset{\mathbf{G}}{\arg\max} \left(\sum_{c=1}^{C} E_c \right) \tag{6}$$

3.2. Kittler-Illingworth

This approach assumes that the gray-scale histogram of the whole image can be approximated through a mixture of *N* Gaussian distributions, one for each class. Optimal thresholds are the values that minimize the error between the original histogram and the mixture of the approximating distributions (Kittler and Illingworth 1986). In formulas we have:

$$\overline{\mathbf{G}}_{\text{Kittler}} = \underset{\mathbf{G}}{\arg\max} \left[\sum_{c=1}^{C} \omega_c \ln\left(\frac{\omega_c}{\sigma_n}\right) \right] \tag{7}$$

3.3. Yen

Yen's method (Yen *et al*. 1995) is formally very similar to Kapur's, but instead of maximizing the sum of the entropy of each class, it sets the optimal thresholds at the values that maximize the sum of the correlation of each class (Eq. 8). Therefore the method is also known as *maximum correlation criterion*.

$$\overline{\mathbf{G}}_{\text{Yen}} = \underset{\mathbf{G}}{\arg\max} \left(\sum_{c=1}^{C} CR_c \right) \tag{8}$$

3.4. Otsu

Otsu's method determines the set of thresholds that maximizes the between-class variance. Originally designed for two level thresholding (Otsu 1979), it has been later extended to the multi-class domain (Liao *et al*. 2001). Mathematically, the method can be formalized as follows:

$$\overline{\mathbf{G}}_{\text{Otsu}} = \underset{\mathbf{G}}{\arg\max} \left[\sum_{c=1}^{C} \omega_c (\mu_c - M)^2 \right] \tag{9}$$

where *M* is the average intensity of the whole image.

4. EXPERIMENTS AND RESULTS

We carried out a set of experiments to quantitatively evaluate the goodness of the thresholding methods at separating paper impurities from the background. To assess the effectiveness of each method we considered the following parameters: overall accuracy, normalized number of false positives and normalized number of false negatives.

4.1. Overall accuracy

The overall accuracy is the sum of the percentage of foreground pixels (i.e.: defects) correctly classified as foreground and that of background pixels (i.e.: non-defects) correctly classified as background. This parameter gives an overall estimate of the effectiveness of the segmentation process. In formulas we have:

$$A = \frac{|B \cap B_T|}{|I|} + \frac{|F \cap F_T|}{|I|} \tag{10}$$

where *A* is the overall accuracy; *I* the whole image; *B* and *F* the background and foreground produced by the thresholding method; B_T and F_T the 'true' background and foreground, which have been manually established beforehand. Symbol '| |' stands for 'the number pixels of'.

4.2. False positives

False positives represent 'type I errors': a false positive occurs each time a background pixel (i.e.: non-defect) is incorrectly classified as foreground (i.e.: defect). The normalized number of false positives can be expressed as follows:

$$FP = \frac{|F \cap B_T|}{|I|} \qquad (11)$$

4.3. Fale negatives

False negatives are also referred to as 'type II errors'. A false negative arises each time a foreground pixel (i.e.: defect) is incorrectly classified as background (i.e. non-defect). In formulas we have:

$$FN = \frac{|B \cap F_T|}{|I|} \qquad (12)$$

4.4. Results

Tables 2-4 summarize the performance of the image thresholding methods considered in the experiment.

Table 2 Overall results of the single-threshold (two-class) experiment.

	Kapur			**Kittler-Illingworth**			**Otsu**			**Yen**		
Dataset	FN	FP	A	FN	FP	A	FN	FP	A	FN	FP	A
W1	0,132	0,004	99,864	0,046	0,095	99,860	0,000	42,546	57,453	0,148	0,002	99,850
W2	0,225	0,003	99,773	0,097	0,128	99,774	0,000	40,619	59,381	0,249	0,001	99,750
W3	0,284	0,001	99,715	0,163	0,066	99,770	0,000	42,689	57,311	0,297	0,000	99,703
Avg	**0,214**	**0,003**	**99,784**	**0,102**	**0,096**	**99,801**	**0,000**	**41,951**	**58,048**	**0,231**	**0,001**	**99,768**

Table 3 Overall results of the double-threshold (three-class) experiment – defects brighter than the background.

	Kapur			**Kittler-Illingworth**			**Otsu**			**Yen**		
Data set	FN	FP	ACC	FN	FP	ACC	FN	FP	ACC	FN	FP	ACC
B1 (w)	0,053	0,027	99,920	0,052	0,023	99,925	0,150	0,000	99,850	0,073	0,016	99,911
B2 (w)	0,121	4,980	94,899	0,126	0,001	99,872	0,237	0,000	99,763	0,141	9,935	89,924
B3 (w)	0,129	0,008	99,863	0,166	0,000	99,833	0,292	0,000	99,708	0,164	0,003	99,833
Avg	**0,101**	**1,672**	**98,227**	**0,115**	**0,008**	**99,877**	**0,226**	**0,000**	**99,774**	**0,126**	**3,318**	**96,556**

Table 4 Overall results of the double-threshold (three-class) experiment – defects darker than the background.

	Kapur			Kittler-Illingworth			**Otsu**			**Yen**		
Data sets	FN	FP	ACC	FN	FP	ACC	FN	FP	ACC	FN	FP	ACC
B1 (b)	0,069	0,028	99,903	0,073	0,015	99,913	0,000	50,253	49,747	0,078	5,004	94,918
B2 (b)	0,260	9,947	89,793	0,276	0,016	99,708	0,000	47,171	52,829	0,324	9,937	89,739
B3 (b)	0,200	0,019	99,781	0,175	0,033	99,791	0,000	45,604	54,396	0,229	0,010	99,761
Avg	**0,176**	**3,332**	**96,492**	**0,175**	**0,021**	**99,804**	**0,00**	**47,676**	**52,324**	**0,210**	**4,984**	**94,806**

In the single-threshold experiment (Tab. 2) Kapur's, Kittler's and Yen's methods all showed good accuracy with comparable figures. By contrast, the performance of Otsu's algorithm was largely unsatisfactory. Among the first three approaches, Yen's and Kapur's produced less false positives, whereas Kittlers' produced less false negatives.

In the double-threshold experiment (Tab. 3-4), Kittler's method appreciably outperformed the others in terms of overall accuracy. This trend is even more evident when it comes to determining defects that are darker than the background (Tab. 4). Otsu's approach proved rather unreliable in this case too, with an overall accuracy far lower than the other methods. Kittler's method also produced fewer false positives in this experiment, whereas the number of false negatives is similar to that produced by the other methods.

5. CONCLUSIONS

Automatic dirt detection and analysis through machine vision plays a central role in the papermaking industry. A fundamental issue in this process is the problem of separating dirt particles from the background through suitable image processing methods. The typical strategy consists of determining a set of intensity values (thresholds) capable of separating the impurities from the background. In this context we have evaluated, experimentally, the performance of four thresholding methods on a dirt detection experiment. Among the four strategies considered here, the method proposed by Kittler and Illingworth (Kittler and Illingworth 1986) proved the most stable and reliable for dirt analysis.

ACKNOWLEDGMENTS

This work was supported by the European Union under project no. Life09-ENV/FI/000568: VOCless pulping waste waters.

REFERENCES

Carfagni, M., Furferi, R., and Governi, L. 2005, A real-time machine-vision system for monitoring the textile raising process. *Computers in Industry*, 56 (8-9), 831–842

Bianconi, F., Fernández, A., González, E. and Saetta, S.A., 2013, Performance analysis of colour descriptors for parquet sorting; *Expert Systems with Applications*, 40 (5), 1636-1644

Kukkonen, S., Kälviaïnen, K., and Parkkinen, J., 2001, Color features for quality control in ceramic tile industry. *Optical Engineering*, 40 (2), 170–177, 2001.

Bianconi, F., González, E., Fernández, A. and Saetta, S.A., 2012, Automatic classification of granite tiles through colour and texture features, *Expert Systems with Applications*, 39 (12), 11212–11218

Furferi, R., Governi, L., Volpe, Y., ANN-based method for olive ripening index automatic prediction, 2010, *Journal of Food Engineering*, 101 (3), 318-328

Furferi, R., Governi, L., Volpe, Y., Carfagni, M., 2013, Design and assessment of a machine vision system for automatic vehicle wheel alignment, *International Journal of Advanced Robotic Systems*, 10, art. no. 242

Kälviäinen, H., Saarinen, P., Salmela, P., Sadovkinov, A. and Drobchenko, A., Visual inspection on paper by machine vision, 2003, *Proc. of. SPIE Intelligent Robots and Computer Vision XXI: Algorithms, Techniques and Active Vision*. Vol. 5267, pp. 321-332

Navarrete, H., Cadevall, C., Bouydain, M., Antó, J., Pladellorens, J.M., Colom, J.F., and Tosas, A., System for off-line optical paper inspection and quality control, 2003, *In Optical Measurement Systems for Industrial Inspection III, volume 5144 of Proceedings of SPIE*, pp. 774–782

Prykäri, T., Czajkowski, J., Alarousu E. and Myllylä, R., 2010, Optical Coherence Tomography as an Accurate Inspection and Quality Evaluation Technique in Paper Industry, *Optical Review*, 17 (3), 218-222

Synnergren, P., Berglund, T. and Söderkvist, I., Estimation of curl in paper using a combination of shape measurement and least-squares modeling, 2001, *Optics and Lasers in Engineering*, 35 (2), 105-120

Sjödahl, P. and Larsson, L., 2004, Monitoring microstructural material changes in paper through microscopic speckle correlation rate measurement, Optics and Lasers in Engineering, 42 (2), 193-201

Rahman, M. O., Hussain, A., Scavino, E., Basri, H. and Hannan, M. A., 2011, Intelligent computer vision system for segregating recyclable waste papers, *Expert Systems with Applications*, 38 (8), 10398-10407

Torniainen, J.E., Soderhjelm, L.S.A. and Youd G., 1999, Results of automatic dirt counting using transmitted light, *TAPPI Journal*, 82 (1), 194-197

Corscadden, K.W., Trepanier, R.J., 2006, Online measurement of dirt specks in sheets, *Proc. of TAPPI Engineering, Pulping, and Environmental Conference*

Ricard, M., Dorris, G., Gendron, S., Pagé, N., Filion, D. and Castro, C., 2012, A new online image analyzer for macrocontaminants in recycled pulps, *TAPPI Journal*, 11 (2), 19-28

Sezgin, M., Sankur, B., 2004, Survey over image thresholding techniques and quantitative performance evaluation, *Journal of Electronic Imaging*, 13 (1), 146-165

Kapur, J., Sahoo, P. and Wong, A., 1985, A new method for gray-level picture thresholding using the entropy of the histogram, *Computer Vision, Graphics and Image Processing*, 29 (3), 273-285

Kittler, J. and Illingworth, J., 1986, Minimum error thresholding, *Pattern Recognition*, 19 (1), 41-47,

Yen, J.C., Chang, F. and Chang, S., 1995, A new criterion for automatic multilevel thresholding, *IEEE Transactions on Image Processing*, 4 (3), 370-378

Otsu, N., A Threshold selection method from gray-Level histograms, 1979, *IEEE Transactions on Systems, Man and Cybernetics*, 9 (1), 62-66

Liao, P., Chen, T. and Chung, P., A fast algorithm for multilevel thresholding, 2001, *Journal of Information Science and Engineering*, 17 (7), 13-727,

AUTHORS' BIOGRAPHIES

Luca Ceccarelli received a B.Sc. and M.Sc. in Mechanical Engineering from the University of Perugia, Italy. He is currently Research Assistant within the Department of Industrial Engineering of the same University. His research interests include computer vision and intelligent systems for industrial applications.

Francesco Bianconi received the M.Eng. degree in Mechanical Engineering in 1997 from the University of Perugia (Italy) and the Ph.D. in computer-aided design in 2001 from a consortium of Italian universities. He has been visiting research fellow at the University of

Vigo (Spain) and the University of East Anglia (UK). Currently, he is Lecturer within the Faculty of Engineering of the University of Perugia. His research interests include computer vision, image processing and pattern recognition, with a special focus on texture and color analysis. He is IEEE Senior Member.

Stefano A. Saetta is Associate Professor of Industrial Plants within the Department of Industrial Engineering of the University of Perugia, Italy. His research interests cover: modeling and simulation of logistics and production processes, life cycle assessment, discrete-event simulation, decision methods, lean production, networks of enterprises and environmentally friendly production systems. He has been visiting professor at Rutgers University (USA), the University of Arizona (USA) and the University of Göttingen (Germany). He has directed several national and international research projects supported by private and public companies. He authored/co-authored more than 80 scholarly papers in international journals and conferences.

Antonio Fernández received the M.Eng. degree in electrical engineering in 1993 and the Ph.D. degree (with honors) in applied physics in 1998, both from the University of Vigo, Vigo, Spain. He held a research fellowship in the Department of Applied Physics, University of Vigo, during the period 1994 through 1998. He was appointed to the Department of Engineering Design, University of Vigo, in 1999, where he is currently full-time Senior Lecturer in Engineering Drawing. He has worked as a visiting researcher at Centre for Research on Optics (Mexico), University of Perugia (Italy), Dublin City University (Ireland) and Computer Vision Centre (Spain). His research interests are in image processing, pattern recognition and computer vision, with a special focus on image texture analysis.

Valentina Caldarelli received a B.Sc. and M.Sc. in Mechanical Engineering from the University of Perugia, Italy. She is currently Research Assistant within the Department of Industrial Engineering of the same University. She works on the reduction of VOC emissions in papermaking process and on the simulation and analysis of the supply chain the simulation and analysis of the supply chain in the food industry.