

Semisynthetic Ground Truth for Dirt Particle Counting and Classification Methods

Nataliya Strokina¹ Aki Mankki² Tuomas Eerola¹ Lasse Lensu¹
 Jari Käyhkö² Heikki Kälviäinen¹

¹ Machine Vision and Pattern Recognition Laboratory and ² FiberLaboratory
 Lappeenranta University of Technology
 P.O.Box 20, FI-53851 Lappeenranta, Finland
 firstname.lastname@lut.fi

Abstract

In the evaluation of dirt inclusions in paper, the attention is paid not only to the quantity of dirt but also to the type of dirt particles. Automatic classification methods can be designed for the task, but there should also exist proper evaluation data to truthfully compare the methods. For such comprehensive evaluations, reliable ground truth is essential. To get suitable samples of paper with expert-annotated dirt particles in each can be considered as a too complicated, laborious, and time-consuming task. In the present work, laboratory personnel produced pulp and dirt particles for the samples as pure as possible. Consequently, the initial data was provided as a set of paper sheets with a single dirt type in each. In order to combine dirt particles of different types in one image of a paper sheet and to know the exact location and the type of the particles, a semisynthetic method for generating the ground truth was developed. This paper introduces the algorithm for the purpose. It is shown how the background for dirt particle images is generated, what is the principle of dirt particle placing over the background, and how the problem of color normalization is solved. Finally, the generated background is statistically evaluated, and segmentation result of the dirt particles is used to estimate the quality of the semisynthetic images.

1 Introduction

In the pulp and paper industry automated dirt counting systems have been used for many years. Typically, these systems only segment and calculate the dirt particles. For example, in [3] Pulp Automated Visual Inspection System is introduced. The system performs the segmentation of dirt particles, as black areas in an image, and then standard dirt counts are calculated. Another example is presented in [1], where the InsPulp, an on-line visual inspection system, is described. The image acquisition and processing techniques allow the system to segment and detect the impurities in pulp with low error rate.

Classification of the particles is not done or it is done by dividing the particles in few groups based on their shape [11]. Accurate classification of particles would be a great benefit. Savings in chemical consumption and energy consumption could be attained since bleaching and screening processes could be adjusted according to the type of particles present in process. Also, accurate

classification of particles would have greatest benefit in the problem solving situations. Fast and precise information about the type of particles present in process in a problem situation could reveal the source of the problem and the process could be adjusted to overcome the problem.

For the classification and segmentation of dirt the ground truth is essential. Ground truth generation is a rather complicated task, assuming a tedious work of experts and creation of special conditions. A large amount of required data can become an obstacle as well. In some systems the difficulties with the evaluation of the performance are mentioned, e.g., in [3] there was no opportunity to compare the results with manually segmented particles. In [1], it is shown that the human inspection can be also subjective: the number of dirt particles detected by different inspectors was different.

The present paper focuses on the ground truth data generation for dirt particle segmentation and classification methods. Here, the expertise of laboratory personnel is not used to the spatial annotation of dirt particles in some images but to their most important task: produce dirt and pulp as clean as possible. The data generation for such systems is conditioned by the necessity of getting the images of paper with dirt particles of known type to guarantee the reliability of the training and test sets. The research is carried out in the “PulpVision” project [10]. The initial data and the approach for ground truth generation are introduced in Section 2. The experiments are described in Section 3, and the conclusions are drawn in Section 4.

2 The ground truth data generation

The section introduces the ground truth generation workflow and presents the description of each stage.

2.1 Method overview

A general workflow of the method, from samples preparation to color normalization of the generated images, is introduced in Algorithm 1. As pure as possible samples with a single type of dirt are prepared in the laboratory. In this case there is no need for the experts to specify type of each dirt particle. The goal is to generate an image with a uniform background and dirt particles of different types, so that the type and the location of each particle are known. Except from the semisynthetic image, a binary labeled mask is provided as an output. In the mask the dirt particles correspond

to white regions. All the stages are described in the following subsections.

Algorithm 1 Ground truth data generation

1. Prepare a set of paper samples.
 2. Acquire digitized images of the samples.
 3. If needed, normalize the color of the images.
 4. Segment dirt particles from the images and create a database of the dirt particles.
 5. Choose an image as a basis for the background.
 6. Remove the dirt particles from the basis image.
 7. Fill the holes in the basis image with a generated texture.
 8. Place a desired number of dirt particles from the database on the background.
-

2.2 Sample preparation

Samples for the ground truth consist of three different pulp types and four different types of dirt particles. Pulp types were bleached hardwood, bleached softwood and softwood pulp after the second chlorine dioxide bleaching stage, the color of which is not completely white. This way more variation for the background was gained.

Four common types of dirt particles were selected based on the literature [7], [8] and expert knowledge: shives, bark, plastic and sand. Dirt particles were, either prepared or separated from the pulp, in a laboratory. Shives were separated from reject pulp from brow stock screening.

Sample sheets from three different pulp types were prepared according to the ISO 5269-1 standard. The amount of pulp equivalent to 1,63 g of dry pulp and adequate amount of one type of dirt particles were mixed before the sheet forming. This way sample sheets which contained the equivalent amount of fibers as in a standard 60 g/m² sample sheet, and adequate amount of dirt particles, were prepared. The amount of dirt particles was selected so that there would be over 20 particles per sample sheet, but not so many that particles would significantly overlap each other. All three pulp types were mixed with one of the four types of dirt particles one at a time. Five sample sheets per test point and also one sample containing sand separated from industrial pulp was prepared, so all together 61 sample sheets were prepared.

2.3 Imaging

Images of the sample sheets were taken with a Canon EOS 500D systems camera. The camera contains a 15,1 mega pixel CMOS sensor and it was equipped with a Canon EF-S 60 macro objective. The camera was attached to a Copylizer eVision executive HF stand. The stand contains a fixed light table. First, the camera was focused to the plain and an image from a graph paper was taken. This image was used to define the size of the imaged area. The imaged area was 127,5 mm x 85 mm.

Sample sheets were placed on the light table and a glass plate was put on them. The glass plate was used to ensure that the sample sheets would be smoothly on the light table. Transmitting light was used, but the room was not totally dark and the system was not

covered, so some reflecting light was also present. Examples of images can be found in Figure 1.

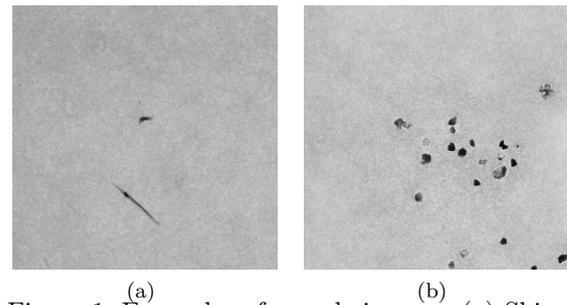
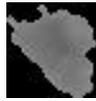
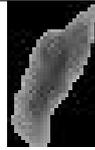


Figure 1. Examples of sample images: (a) Shive; (b) Sand.

2.4 Dirt segmentation

In order to construct a database of dirt particles dirt is segmented from all the images in the test set. The segmentation is performed on the grayscale images using the Kittler method [9]. Table 1 introduces the scheme of the database. An image of a bounding box for each of the segmented particle is stored in the database, as well as its area and type.

Table 1. A scheme of the dirt particle database.

Particle 1	Particles 2	Particle 3	...
			...
Area 1	Area 2	Area 3	...
Type 1	Type 2	Type 3	...

2.5 Background generation

A basis image for the background generation is chosen by a user. The dirt particles are segmented from the image as in Section 2.4. The segmented dirt is removed from the image so that the area occupied by dirt particles is substituted by white pixels. The holes in the image are filled in using Markov Random Fields (MRF) method [4]. In the concept of MRFs [6] a random field is composed pixel by pixel, each of them having the probability of having a certain intensity, depending on the intensity or color of its neighboring pixels. Before image restoration a user manually specifies a sample region in the image, containing no dirt particles with a uniform background. This sample image is used as an ideal background for the area filling. In Algorithm 2 the probability is not calculated explicitly to avoid the computational expenses. The number of iterations can be restricted, for example, by a threshold for the standard deviation of color, which should be close to the original background.

There are several issues to consider in this algorithm: one of them is how to determine the neighborhood, how many iteration should be carried out, and what is the best matching neighborhood. In the implemented version of the algorithm the neighborhood is determined by a user, as well as the number of iterations. The best matching is found by computing the distance between two neighborhoods and choosing the one with the least Euclidean distance.

Algorithm 2 Background generation

1. Construct the neighborhood of each pixel in the sample image.
 2. Repeat the steps 3-5 until a satisfactory result is obtained according to the threshold.
 3. Construct the neighborhood of the current pixel from the original image.
 4. Find the best matching neighborhood from the sample image.
 5. Substitute the current pixel by the pixel with the best matching neighborhood.
-

2.6 Dirt particle placing

During this stage a specified amount of dirt particles is spread over the uniform synthesized background. According to [5], the particles should be placed randomly, and there is also an option of random rotation of the particles. In order to prevent dirt particle overlapping there is a binary mask of the image where the area occupied by a particle is marked by white color. In Algorithm 3 the place for a new particle is selected from the non-occupied area.

Algorithm 3 Dirt particle placing

1. Initialize a binary mask to zeros.
 2. Repeat the steps 3-5 to reach the desired number of dirt particles.
 3. Randomly select a dirt particle from those that have not been placed yet.
 4. Find the place for the particle in the binary mask not occupied by any other particle.
 5. Place the dirt particle on the background.
-

One thing that should be taken into account is to control the amount of dirt in an image. It can be controlled by the percentage of the dirt covering the surface. To form the list of dirt particles to be positioned on the image, the user should specify what type of dirt is needed to be added to the image and its proportion.

As a result Algorithm 3 outputs the image of uniform background and dirt particles placed on it in a random way. Additionally the information about the dirt particles is stored in a labeled binary mask.

2.7 Color normalization

In the case the samples are produced from different types of pulp, where color may vary, color normalization of sample images is needed. The method, described in [2], is taken as a basis. In this work, color normalization assumes transferring color schemes of all the images from the set to the color scheme of the first image. The transform is done separately for dirt and background. Instead of working in the RGB color space, phosphor chromaticity of which might not be known, the images are transferred to Ruderman et al.'s perception-based color space $l\alpha\beta$ [2]. The transformation minimizes the correlation between color channels, making it less probable to get undesired cross-channel effect. Once all the images are in the $l\alpha\beta$ color space, the color characteristics of the primary image are applied to the other images by the normalization of the

images means to the mean of the target image. After that the images should be transferred to the RGB color space. As a result images with normalized color schemes are produced.

3 Experiments and Discussion

The experiments have been performed on the provided set of images. Figure 2 presents the results of filling the holes with synthesized background, and scattering the dirt particles over the obtained background. The experimental set consisted of three images of the same pulp type with different dirt classes: plastic, shive, and bark. One of the images was chosen as the background, and the certain amount of dirt particles from each image is placed on the background.

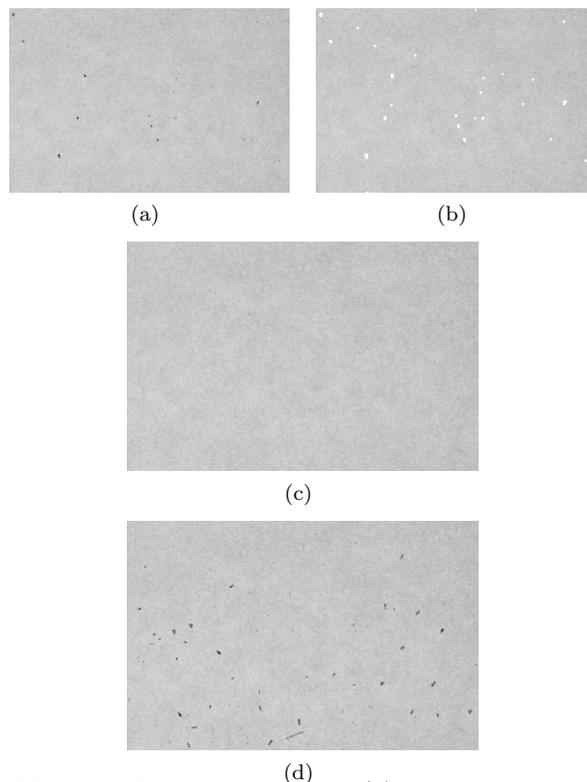


Figure 2. Image generation: (a) Original image; (b) Image with removed dirt; (c) Generated background; (d) Resulting image.

Three statistical moments are calculated (see Table 2) for the original and semisynthesized background: standard deviation, skewness and kurtosis. Skewness measures the lack of symmetry and kurtosis characterizes whether the data is peaky or not.

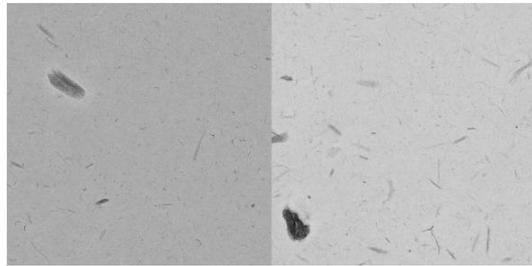
Table 2. Statistical moments for original and semi-synthetic backgrounds.

Image	Std	Skewness	Kurtosis
Original	0.032	0.798	7.633
Semi-Synthetic	0.030	0.323	3.512

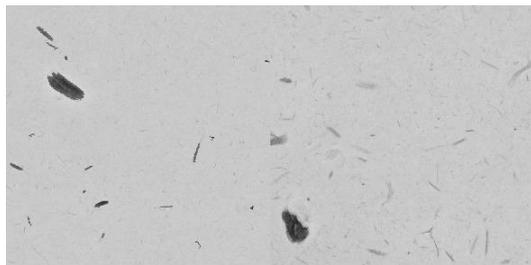
The results show that the variation of color for the generated image is similar to the variation of color in the original one. Although there some differences in the second and the third statistical moments.

In the next experiment images produced of two different pulps are considered. Figure 3 illustrates the

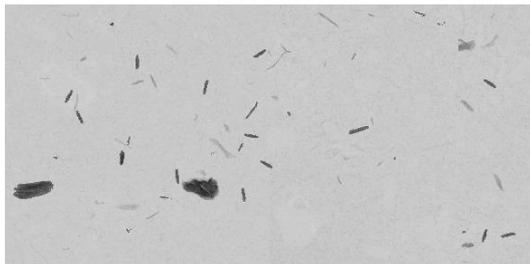
results of the color normalization performed on two images. In Figure 3(a), the original images are shown. The color scheme of the first image is transformed to the color scheme of the second image for dirt and background separately.



(a)



(b)



(c)

Figure 3. Color Normalization: (a) Original images; (b) Image with normalized color (c) Resulting image.

The resulting image is represented in Figure 3c. The color schemes of the images were normalized and the dirt particles from both images were scattered randomly over the obtained background. In the color normalization one should decide either only the background or the schemes of background and dirt separately need to be normalized.

In order to evaluate the performance of the system, for each original image a semisynthetic image was generated. The segmentation of dirt is performed on the original and the obtained image. The number of dirt particles detected in each image is presented in Table 3.

Table 3. Dirt counts for original and generated images.

Image	Bark	Plastic	Shive	Sand
Original	42	18	8	42
Semi-Synthetic	44	17	6	39

Comparing the results of segmentation, one can notice that the dirt counts in the generated images do not significantly differ from the original ones. In the future work, the performance of classifiers on the original and generated images should be evaluated.

4 Conclusions

The study presents a semi-synthetic approach to produce ground truth data for dirt particle segmentation and classification. The problems of background generation and dirt particles distribution over the background are addressed. Since color is a significant feature in image analysis, special attention is paid to color normalization of the images in the case of different pulp types.

Specifying the amount and type of dirt to be placed, the user obtains the semi-synthetic images, with known locations, types, and areas of dirt particles. There is also an option to use samples produced of different pulp classes. The further evaluation of the system is needed using the classifier of dirt particles. This will enable deeper understanding of how the method used for data generation can limit the quality of the produced images.

Acknowledgements

The research was carried out in the “PulpVision” project (TEKES project 70010/10) funded by the European Union and the participating companies. The authors wish to acknowledge the FiberLaboratory in LUT, TEKES, and the companies for their support and collaboration.

References

- [1] P. Campoy, J. Canaval, D. Pena: “InsPulp: An online visual inspection system for the pulp industry,” *Computers in Industry*, vol. 56, pp. 935-942, 2005.
- [2] E. Reinhard, M. Ashikhmin, B. Gooch, P. Shirley: “Color Transfer between Images,” *IEEE Computer Graphics and Applications: Special issue on Applied Perception*, vol. 21, pp. 34-41, 2001.
- [3] S. Parker, J. R. Chan: “Dirt Counting in Pulp: An Approach Using Image Analysis Methods,” *Proceedings of the IASTED International Conference on Signal and Image Processing (SIP)*, 2002.
- [4] L. Y. Wei, M. Levoy: “Fast texture synthesis using tree-structured vector quantization,” *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 479-488, 2000.
- [5] C. Zeyer, R. Venditti, K. Puangchinda: “The Distribution of Impurities in Pulp and Paper - The Effects of the Random Distribution of Impurities on Image Analysis,” *TAPPI journal*, vol. 78, pp. 168-175, 1995.
- [6] J. Zhang, P. Fieguth, D. Wang: “Random Field Models,” *Handbook of Image and Video Processing*, pp. 301-312, 2000.
- [7] J. Gullichsen, J.E. Levlin: “Chemical Pulping. Papermaking Science and technology,” 1999.
- [8] C. W. Dence, D. W. Reeve: “Pulp Bleaching, Principles and Practice,” 1996.
- [9] J. Kittler, J. Illingworth: “On Threshold Selection Using Clustering Criteria,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 12, pp. 652-655, 1985.
- [10] “PulpVision”: <http://www2.it.lut.fi/project/pulpvision/>
- [11] The Verity IA Color Image Analysis software: <http://www.verityia.com/dirt-counter.php>