Sensor systems for high speed intelligent sorting of waste paper in recycling

M. K. Ramasubramanian¹, R. A. Venditti², P. K. Gillella³

¹Associate Professor Mechanical and Aerospace Engineering
²Associate Professor, Paper Science and Engineering
³Graduate Student, Mechanical and Aerospace Engineering
North Carolina State University
Raleigh, NC 27695-7910, USA
rammk@ncsu.edu
(919) 515-5262

Abstract

Sorting of paper into compatible grades is a necessary step prior to recycling. Current method of manually sorting is tedious, slow, and expensive. Hence, significant part of the waste paper stream is sent to the landfill. High speed automation of the sorting process will improve the cost efficiency and increase the amount of paper recycled significantly. There have been recent developments in automation of this process. Mechanisms for the distribution of the papers from a bale onto a moving conveyor and the pneumatic actuation to deflect papers into different streams are well established. Correct identification of the sample in real-time before the sample reaches the actuation station still remains a challenge since different types of paper and board samples are mixed in a waste stream, and has a lot of variation in terms of color, chemical composition, coating, and prints in black and white and color to different degrees. This is primarily due to the lack of satisfactory sensors and sensor fusion algorithms for sample identification in real-time. In this
paper, we identify key parameters that must be measured, sensor design, and integration of the output from sensors to interpret the type of sample using a fuzzy inference system. Results show that the sensor system proposed is capable of identifying the samples at 90% accuracy. The sensor system can be integrated onto a conveyor and actuation system for automated sorting.

*Keywords: Sorting, waste paper, lignin, stiffness, gloss, color, fuzzy logic*
1.0 INTRODUCTION

Recycling of paper and paper related products are very important for sustainable economic growth. It helps save landfill space and costs and reduces the energy requirements for paper manufacturing since the energy that is needed to produce recycled fibers is much lower than energy required to produce virgin fibers from wood. Paper recycling reduces the rate of consumption of precious natural resources (wood, water, minerals, and fossil fuel). It is estimated that one ton of paper fibers from recycled stock saves approximately 17 to 31 trees, 7000 gallons (26500 liters) of water, 4,000 KWh of electricity, and 60 pounds (27.2 Kg) of air pollutants that would otherwise be produced [1]. In the US 100 million tons of paper a year is consumed—for everything from daily newspapers to books and cardboard boxes. It has also been established that 50% of all the paper being produced is discarded immediately after the first use and that almost all of it can be recycled [2]. Current recycling rate is about 50%. Paper products are the number one contributor to landfill volume. This highlights the importance of having an efficient recycling system for paper. Before waste paper can be recycled, it has to be sorted into the various grades. Paper is broadly classified into the following grades for the purposes of recycling groups: Newsprint, Coated Sheet, Colored Free Sheet, White Free Sheet, Board, and Mixed Waste.

Currently waste paper is either sorted by the consumer prior to disposal or not sorted at all. At the municipal waste sorting plants, it is mostly carried out manually. Here the paper waste is broadly classified as brown paperboard, newspaper, and mixed office waste. But the inability of the human eye to distinguish between coated inferior grades and good grades result in sorted office waste containing high-grade paper contaminated by similar looking low-grade paper, which reduces the overall fiber quality when they are recycled together. The throughput from these manual sorting facilities is also very low. A typical manual conveyor operates at 75 feet per minute (22.86 m/min). A recently conducted research has also concluded that laborers working at the manual sorting facilities are exposed to microorganisms, organic dust, fungi which cause severe infections [3]. Hence there is an important need to automate the sorting process for efficient recycling.

To develop an efficient automated sorting system, robust sensors that can measure various physical and optical properties of paper are necessary. The physical characteristic of paperboard is its high stiffness compared to other grades of paper; the characteristic of newsprint is its high lignin content; coated papers have high gloss; and non-office papers are generally not white and have high color content. In order to sort the papers into various grades, one needs to measure the stiffness of specimens, the chemical content, the color, the gloss, and make an inference all in real-time. This poses many challenges that have been systematically addressed in our work. Further challenges arise when a newsprint paper is printed
with colors that mask the lignin underneath, lignin free sheets that are printed with colors that falsely trip the lignin sensor or a stack of office papers on the conveyor that produce high stiffness and mimic a paperboard. Added together with high throughput requirements, intelligent sorting of mixed waste is a challenging problem.

2.0 BACKGROUND

2.1. Lignin Sensor

An optical sensor that can measure the amount of lignin present in paper based on fluorescence has been developed at North Carolina State University [4] [5] [6] [13] [14]. This sensor is highly efficient in measuring the amount of lignin in samples. Lignin is the non-cellulosic polymer present in wood. When present in paper, lignin weakens the paper and causes discoloration when exposed to light. When wood is pulped primarily by mechanical action (groundwood pulp) or a combination of heat and mechanical action (thermomechanical pulp, TMP), the yield from the process is very high but a substantial part of the lignin stays with the fibers. Newsprint is typically made with mechanical pulp. On the other hand, fibers used in office ledger paper are produced using a chemical process, namely, the kraft pulping process. In this case, the fibers are released from lignin in wood through a chemical process that dissolves the lignin. In these grades, the lignin content is low. Further, these fibers are bleached to increase the brightness. Hence the lignin sensor can distinguish between newsprint and other paper grades. The lignin sensor measures the intensity of fluorescence to predict the actual amount of lignin present in the paper. Fig. 1 shows a schematic of the lignin sensor.

Details of lignin sensor design and its performance are presented in detail elsewhere [4]. The lignin sensor produces a proportional output to the amount of lignin present in the sample. The sensor was also tested on moving paper samples on a conveyor at 2.5 m/min and the sensor was able to successfully identify between, office paper, TMP, groundwood (newsprint), and bleached sulfite pulp [4]. An example plot from a dynamic test with a high lignin and a low lignin sample strips alternating on a moving belt (a belt sander) is shown in Fig. 2. Variations within sample are due to color and printed text on the sample surface.

It was found that the sensor results are affected by color present on the surface of the paper and the distance of the sensor from the sample at the time of measurement [4]. A distance versus intensity calibration curve was generated and used as a correction for the measured intensity. The presence of color poses a more serious problem. For example, a red patch on the top of an office ledger paper will produce a normalized voltage that matches that of newsprint, meanwhile, a
blue or green patch printed on newsprint surface will produce signals that match with a ledger paper and cause misidentification. Thus, the lignin sensor alone cannot detect the presence of newsprint. Additional considerations from other sensors are necessary to make a determination of the sample type.

2.2 Stiffness Sensor

Regardless of content, a board sample should be separated into a separate stream. There are several kinds of boards, namely, brown carton used in packaging, manila folder covers used in offices, thick magazine covers, six-pack can holders, etc. All of them have a common physical property, namely, high stiffness. Stiffness sensing on discrete samples, randomly oriented, on a moving conveyor is challenging. It is not possible to hold the sample and apply a load to measure the elongation and calculate the stiffness. Hence, we developed a quick, non-contact method for measuring the relative stiffness of paper samples moving on a conveyor at high speeds. The method consists of impinging an air jet on a moving paper sample while it is crossing a gap during transfer from one conveyor belt to another. The deflection caused by the impinging air jet can be measured and used to categorize different grades of paper. We carried out finite element simulations of the proposed method to understand the effects of speed, sample orientation, nozzle pressure, and other parameters to make certain that the stiffness estimation has sufficient resolution to distinguish between board and paper grades at high conveyor speeds. The sensing concept schematic is shown in Fig. 3. A finite element model of a discrete orthotropic paper sample crossing a conveyor gap at high speed, subjected to an air jet impingement was developed using ABAQUS finite element code. Details of the model and the results are described elsewhere [7].

Simulations show correctly the cases where low stiffness samples do not clear the gap due to an increasing deflection when subjected to air pressure, and high stiffness samples are predicted to clear the gap with minimal deflection. The sensing technique was tested in the static mode in the laboratory and in a dynamic mode on a sorting conveyor moving at 100 m/min [8]. The deflection decreased with increasing sample thickness and increasing basis weight, as both contribute to bending stiffness. For the case of dynamic measurement, it was necessary to use a continuous air flow at lower air pressure (34.5 kPa) compared to static tests [8]. The sensor was able to identify paperboard samples clearly. When the difference in the thickness of the tested samples is not significant, the difference in deflection was due to differences in elastic modulus [8]. Thus the stiffness sensor is successful in identifying board from paper on a moving conveyor. However, it cannot distinguish between a stack of newsprint versus a single paperboard. Additional sensor input must be used to resolve such conflicts. Additional sensors to measure other unique properties must be developed for a
robust sorting system, followed by a sensor fusion algorithm to realize the full potential of sensor-based sorting. Further details regarding the performance of the sensor has been presented elsewhere [15] [16].

3.0 ADDITIONAL SENSORS

3.1 Gloss Sensor

The light reflected from any surface has two components. Diffuse reflection is the component whose angle of reflectance is totally random with respect to the angle of incidence. Fig. 5(a) schematically demonstrates the concept of diffuse reflection from a rough surface. Specular reflection is the component whose angle of reflectance is equal to the angle of incidence. Fig. 5(b) shows specular reflection from a smooth surface. The degree to which each component is present in the reflected light gives us an indication of the surface roughness of the sample. Surface gloss is a measure of the intensity of the specular reflectance of a sample [9]. It is measured using a specular gloss meter. The gloss meter works by focusing unpolarized white light onto the sample at certain standard angles of incidence (20°, 60° and 75°). The specular reflection is picked up by a receptor lens and focused onto a photo-detector. The intensity of the specular reflection serves as a measure of the surface gloss.

While industrial grade gloss sensors are available from manufacturers like Inprox (www.inproxsensors.com, model MTG5F, gloss/finish sensor), and image processing based gloss measurement techniques such as image processing [10], are available; we did not have ready access to an online sensor. Hence, for the purposes of establishing the use of gloss measurement in the sorting algorithm development, we used a Technidyne Gloss meter (model T480A) in the laboratory.

The purpose of gloss measurement is to identify coated cover stock on magazines and annual reports. The coated paper should be sorted into a separate group as it will pose problems in recycling due to its high mineral content. Further, some of the coated board could have a polyethylene coating on it and these highly glossy surfaces must be removed from recyclable high quality fibers.

3.2 Color Sensor

Most of the modern digital cameras work on the same principle as that of the Charge Coupled Device (CCD) based color sensors. Some sophisticated digital Single Lens Reflex (SLR) cameras also provide the option of capturing the
RAW data from the sensor. Digitized values of the analog signals generated from each of the pixel in the sensor are logged into a file. This essentially converts the digital camera into a highly flexible color sensor. These cameras provide the ability to alter the aperture size, sensitivity and exposure time of the sensor. In order to obtain color information in real-time, we used a commercially available digital SLR camera (Nikon D50). The RAW data from the camera’s sensor is logged into a memory card (Secure Digital) card. The logged data is then processed using proprietary software from the camera manufacturer to convert the images into standard RGB format. This RGB data is further processed using MATLAB to extract various color parameters for each of the samples.

In order to acquire images of samples, an aperture setting of 8, shutter speed 250 (.004 s), ISO sensitivity 200, Custom white balance (Pure white sample was used to calibrate the sensor), and focal plane 60 cm (distance from camera lens to the sample) were used in setup. For every sample, RAW data was collected with every picture. A sample sheet of color magazine paper and the RGB values across the scan line are shown in Fig. 6. The scan was done at low resolution, with 50 dpi averaging over a square of 5x5 pixels. It is seen that the RGB versus position varies depending on the color at the particular pixel as expected. However, no meaningful summary could be made to represent the color of the overall sample. In order to use the information for sorting, it is important to reduce the data to a parameter that can be computed in real time. The method of using RGB data for identification has been described by Su et al [11]. In their work, they used a color line scan camera and generated RGB values. They further used a threshold value for the sum of RGB values to identify the presence of a contaminant in wool. The method worked satisfactorily as they were trying to identify small contaminants on large wool fabric surface, generally different in color from background. In our work, it is not sufficient to identify the presence or absence of color, but need to estimate the overall color of the sample. In addition, it is necessary to determine if the sample overall is a color paper or a white paper with color prints on it as they would go to different recycling streams.

3.2.1 Color data processing

Since the RGB color data by itself did not yield any obvious trend, we transformed the data into a more practical, L* a* b* space, through the standard coordinate transformations. While interpreting RGB primary colors is obvious, the L*a*b* space needs some explanation. A detailed version can be found in reference [12]. The L*a*b* space is designed to approximate human vision and perception of color. The L component closely matches the human perception of brightness. The a* component represents the position of the color in green-magenta axis, and the b* component represents the position
of the color on the yellow-blue axis. The reason for choosing to work with this color space is because it closely resembles the way we perceive color i.e., Light/Dark, Warm/Cool descriptors.

The transformation from the RGB color space to the L*a*b* color space is essentially a coordinate transformation and is carried out in two stages. The first stage involves the conversion from the RGB space into a device independent XYZ space. The transformation is as follows [12].

\[
\begin{pmatrix}
X \\
Y \\
Z
\end{pmatrix}
= 
\begin{pmatrix}
0.4124 & 0.3576 & 0.1805 \\
0.2126 & 0.7152 & 0.0722 \\
0.0193 & 0.1192 & 0.9595
\end{pmatrix}
\begin{pmatrix}
g(R_{normalized}) \\
g(G_{normalized}) \\
g(B_{normalized})
\end{pmatrix}
\]

(1)

where

\[
g(K) = \begin{cases} 
K^{\gamma}, & \text{for } K > 0.04045 \\
\left(\frac{K+a}{1+a}\right)^{\frac{1}{\gamma}}, & \text{for } K \leq 0.04045 \\
K^{\frac{1}{12.92}}, & \text{otherwise}
\end{cases}
\]

and \( \gamma = 2.2, \ a = 0.055 \) \hspace{1cm} (2)

The normalized values for RGB values are obtained by dividing the raw value by 255 (8 bit resolution) so that the intensities will range from 0 to 1. The L*, a*, and b* values are computed by

\[
L^* = 116 f\left(\frac{Y}{Y_n}\right) - 16
\]

\[
a^* = 500 f\left(\frac{X}{X_n} - f\left(\frac{Y}{Y_n}\right)\right)
\]

\[
b^* = 200 f\left(\frac{Y}{Y_n} - f\left(\frac{Z}{Z_n}\right)\right)
\]

(3)

where, \( X_n, Y_n, \) and \( Z_n \) are the tristimulus values for the reference white point. For incandescent illumination, these values are \( X_n = 0.94809, Y_n = 1, \) and \( Z_n = 1.0730 \). The functions \( f(X/X_n), f(Y/Y_n), \) and \( f(Z/Z_n) \) are computed from:

\[
f(t) = \begin{cases} 
\frac{1}{t^3}, & \text{for } t > 0.008856 \\
7.787t + 16/116, & \text{otherwise}
\end{cases}
\]

(4)

Fig. 7 shows the distribution of the L*a*b* parameters across the surface of a newsprint sample, along the scan line. The L* component, indicating brightness, is found to be high for shades of white and drops to a low value for darker colors. Also, between white and black, there is very little change in the a* and b* components. There is a significant change in those components only when a colored portion is encountered.
For colored sheets, the presence of printed text not only affected the L* component, but also affected either the a* or b* components. An algorithm that could eliminate the effect of printed text on surfaces and calculate the L* a*b* color components was developed in MATLAB. This algorithm makes a significant but a reasonable assumption that no sample will be coated entirely in black. It needs data from both RGB and L*a*b* spaces for a given point. It begins scanning from one end of the sample. Printed text is assumed to be either black in color or any dark shade of gray. In numerical terms, this is evaluated using the following conditions: If the mean of the R, G and B components is less than 50 and the sum of the deviations of each of the components from the mean value is less than 50, then the color found is black or a dark shade of gray. Once a printed portion is detected, the algorithm digitally erases the printed text by replacing the color for those points with the data from the most recent points which were non-black. The rare scenario where the sensor starts scanning at a black region, the algorithm continues scanning till it finds a non-black region and then replaces all the data corresponding to the printed regions with the new color that is detected retroactively. The same procedure is repeated till the end of the sample is reached. The result would contain all the data corresponding to the printed regions replaced with data from neighboring regions.

To illustrate the working of the digital erase algorithm, a sample that was studied. The sample is white with printed text and images on it. The red line shows the scan line for which L*a*b* data was obtained after image processing. The result is shown in Fig. 8. It is seen that the original data without correction (solid dots) shows a large drop in brightness value L* as the scan crosses over the black print. Hollow circles represent corresponding ink removed points digitally or filtered data points.

The digitally erased points change the brightness value back to that of the previous point thereby restoring the brightness to that of the underlying sheet. Results for a color flyer paper printed with black ink were analyzed. The results are shown in Fig. 9 and the case for a magazine paper is shown in Fig. 10. Although the digital erasure was carried out for the magazine paper, there is no printed text meeting the erasure criteria that the data is significantly unaltered. Since there are several colors printed on the surface, the L*a*b* data shows high variability along the length of the scan.

In order to reduce the data, the following parameters were calculated for the scan.

1. LAB Deviation: The standard deviation of L*, a* and B* components

2. LAB Average: The mean of the L*, a* and b* components
3. LAB Deviation filtered: The standard deviation of L*, a* and b* components of the filtered data

4. LAB Average filtered: The mean of the L*, a* and b* components of the filtered data.

Results are shown in Tables 1, 2, and 3 corresponding to samples shown in Figures 8, 9, and 10, respectively. Some observations can be made from the data in the tables.

1. The standard deviation of a* and b* components of the filtered data are generally lower for the printed white printed paper and the colored flyer paper when compared to that of color magazine paper. This is due to the presence of a lot of color images in the magazine sample. The sum of standard deviations of the LAB filtered data can be used as one of the color identifier parameters, being able to distinguish magazine type color versus plain color or white papers.

2. The sum of the absolute values of the a* and b* components of the LAB Average filtered in general is lower for white copy paper than that of colored paper. This is because pure white has a* = 0 and b* = 0 and most of the brightly colored sheets have either a* or b* very high. Thus, the sum of absolute values of LAB average obtained from filtered data can be used as the second color identification parameter.

These parameters were used as the two color sensor parameters in the design of sensing algorithm.

4.0 SENSOR FUSION

A sample moving on a conveyor should be assessed for all necessary attributes and the information from the sensors should be collectively interpreted to make a decision on the grade of paper that is. In this work, we have used the lignin sensor, the stiffness sensor, gloss sensor, and the color sensor parameters developed in the previous section and explored the use of artificial intelligence algorithms for sensor fusion. Due to their recent success in the fields of pattern recognition and classification, an algorithm based on artificial neural networks was first evaluated. The possibility of integrating human-like reasoning and expert knowledge into algorithms based on Fuzzy Logic was a primary reason for choosing Fuzzy Inference Systems (FIS) as another alternative.

4.1 Samples and Tests
We selected samples of different types from a typical unsorted waste paper and identified different regions in them to represent different variations of color, prints, and other variables. A total of 224 regions on 63 different samples were selected. Six parameters were determined for each of the 224 samples, namely, average lignin value across the sample, gloss meter reading from the centre of the sample, deflection in the downward direction, deflection in the upward direction, color variance of a* and b*, color average of a* and b*.

4.1.1 Lignin Sensor Data

The lignin sensor is calibrated by adjusting the control voltage and varying the gain of the photon counting module. The lignin sensor has currently been interfaced with LabVIEW via a USB-6001 data acquisition card. Since newsprint samples have the highest amounts of lignin in a waste stream, a newsprint sample is placed under the lignin sensor and the control voltage is adjusted till the voltage from the photon counting module is close to 4.5V. The reason for restricting the output to this value is to prevent any damage to the DAQ card due to voltages greater than 5V. These high voltages could be generated occasionally by samples with high lignin content if the spot being scanned is printed with red color. The calibration is also verified by testing a lignin-free sheet (copy paper) and checking if the sensor produces an insignificantly low value.

A LabVIEW program is started which sets the control voltage for the PCM and records the analog signal from the lignin sensor. The sample is moved under the sensor head to simulate the condition of the sample moving on a conveyor. Care is taken in sliding the sample at a constant speed. The sliding is done in such a way that the scan line passes through the center of one of the zones of interest. The values are logged into a data file when the program is stopped after the entire length of the sample has been scanned. The process is repeated many times for each sample to measure the data for each of the zones. An example raw data from the lignin sensor for color flyer paper with black prints is shown in Figure 11. The shaded region of the data corresponds to the entire width of the paper sample. The zone data will be a subset of the data in the shaded region. The data is further filtered through averaging over 10 data points to remove noise and the zone data is obtained by taking the corresponding data from position on the sample. The data for zone 3 shown on the sample is represented by the shaded region in Figure 12.

5.1.1 Stiffness Sensor Data
The stiffness sensor setup involves setting the gap between the two conveyors (or between two supports in case of static setup), nozzle air pressure, duration of air pulse for static tests, distance between the nozzle and the conveyor, speed of the conveyor, sensitivity of the ultrasonic distance sensor, and the sensor to sample distance. The sensor is described in detail elsewhere [7]. The parameters are chosen in such a way that three calibration samples with varying degrees of stiffness all produce considerably different amounts of deflection. Care is taken when using the static sensor setup to make sure that the sample returns to its original position after the pneumatic load is removed. If the sample fails to return to the original position, it is an indication that if the same parameters were used in the dynamic setup, the sample would fail to clear the gap between the conveyors, based on several observations. In general, under pneumatic load, the sample deflects down and upon release bounce back. More compliant samples show a pronounced bounce after load is removed while a board sample shows less of a bounce back. The ultrasonic sensor is positioned carefully in such a way that all these deflections are within the sensor’s working range. The upper bound and lower bound values of the sensor’s detection window are programmable. We use this function to maximize the sensor’s resolution by using the deflection data from the three characterization samples to set these values. This was done by placing the sensor at a distance of 25cm below the undeflected sample and programming its output to be 0V at 18cm and 5V at 28cm. This gave us a resolution of 0.5 V/cm which was excellent for capturing the deflection profiles. Each sample’s deflection profile is obtained from the data logged during the tests. The position of the paper sample before the air pulse is applied is considered to be its mean position. This is obtained by averaging out the distance sensor readings from the beginning of the test until the air pulse is applied. Once the pulse is applied, the maximum and minimum distances from the sample are recorded. By subtracting the value of the mean position from the maximum and minimum values, we obtain the maximum and minimum deflection parameters for each sample.

4.1 Neural Network

We implemented our neural network algorithm based on radial basis functions. We selected samples of different types from a typical unsorted waste paper and identified different regions in them to represent different variations of color, prints, and other variables. A set of 203 sample data were used to train the neural network and the remaining 21 sample data sets were used for model validation. Five sample categories were used, namely; 1. Newsprint, 2. White lignin free sheet, 3. Coated sheet, 4. Board, and 5. Color lignin free sheet. The predictions of the neural network were rounded off to
the nearest integer between 1 and 5 and associated with the sample identification. As shown in Table 4, the accuracy of prediction (correct identification of sample) was not satisfactory. Possible reasons include data inconsistencies that are inherent to waste paper systems. The lignin sensor data for various samples when tested repeatedly and randomly is shown in Figure 13. It is seen that the average values show the expected trend. However, individual test results can be confounding. For example, newsprint could register a low lignin sensor value due to the presence of blue or green shades on the surface which is known to quench the amount of fluorescence given off by the sample. Some newsprint samples made from recycled paper also register low values of lignin if during recycling they had been mixed with fibers from higher grade paper to increase the strength of the paper. These variations are not easily controllable in a waste paper stream. Another possible explanation is that the neural network builds the model that is purely based on the training set without the possibility of incorporating rule based decision making, in addition. For instance, it is known that if the stiffness of a sample is very high, it is almost always a board sample and if the lignin content is very high and the stiffness value is low, then the sample is newsprint. Newsprint can also be identified by another rule which states that, if the stiffness value is very low and the gloss value is also very low then the sample is newsprint. This can be easily implemented in a fuzzy logic scheme. When all but one sample was used as model training data and the one was used as test data, the accuracy improved somewhat, as shown in Table 5.

4.2 Fuzzy Inference System

Fuzzy inference is the process of formulating the mapping from a given input to an output using fuzzy logic. The membership functions for the sensor inputs are shown in Fig. 14. The experimental data were organized for each sensor and the sample was reviewed by a human judge to determine the various cases of memberships for each sensor data. Next, the fuzzy rules need to be developed. The following is a listing of conditions that have currently been implemented as the fuzzy rules for the FIS algorithm for this application.

- If lignin is high and stiffness is low then the sample type is newsprint
- If lignin is medium and stiffness is low and gloss is low then the sample type is newsprint. (This rule assists when a newsprint sample may have significant recycled fiber, causing low lignin reading.)
- If lignin is moderate and stiffness is moderate and gloss is moderate then the sample is a free sheet (depending on the color variance parameter, it can either be a colored or white free sheet)
• If stiffness is low and gloss is high then sample type is magazine (color variance is optional, if it is high, confidence increases, if low, it can be ignored)

• Any sample with very high stiffness is generally considered to be board.

It should be noted that some sensors cannot identify a particular type of paper due to the arbitrary variation of a parameter across different samples. For example, the color variance cannot predict if a sample is newsprint or not, since newsprint sample may or may not have colorful images printed on it. This issue is handled by simply ignoring certain sensor data when identifying certain samples.

The membership functions and the fuzzy rules that have been derived are input into the MATLAB’s FIS toolbox. This toolbox is a design time tool that helps us quickly visualize the inputs, outputs and the various fuzzy parameters. Upon successfully entering all the data into the FIS toolbox, it was tested using the rule viewer. The rule viewer lets us simulate the FIS algorithm by entering an input vector. It shows the various calculations being carried out and gives us an output vector. The output vector is an array of the degrees of confidence with which the algorithm states that a sample is of a particular type. The element in the output vector with the highest value indicates the type of paper. Consider the following case,

• Newsprint confidence - 0.45
• White free sheet confidence - 0.54
• Coated confidence - 0.89
• Board confidence - 0.14
• Colored free sheet confidence - 0.27

From the output vector, it can be interpreted that the algorithm classifies the particular sample as a coated sheet.

The output vectors for all the test samples are presented in Table 5. It contains details about the number of samples in each paper type that was tested, number of samples that were identified correctly and the prediction accuracies for each type of paper. Also, a comparison is done with the prediction accuracy of the neural network algorithm. To obtain the neural network prediction accuracies shown in Table 5, simulations were carried out for one sample at a time as described earlier. All other samples except the one being tested were used as the training data. Then the output obtained was compared to that
required for that particular sample. This process was repeated for all other samples and the predictions of sample type were recorded. Due to the increase in the amount of the training data and larger number of validation data, an increase in the prediction accuracy was observed with the modified neural network. Out of 224 test samples, 204 samples were identified correctly by the FIS algorithm. This gives us a sorting accuracy of 90% for the FIS algorithm. Some observations can be made as to why some of the samples were wrongly detected.

• If for some reason, a newsprint sample made from recycled paper (low lignin) somehow registers a moderate deflection value at the stiffness sensor, both the rules that are meant to identify the newsprint would fail and at the same time satisfy the requirements for a white lignin free sheet.

• When a white free sheet fails to deflect as expected at the stiffness sensor and registers a low value of surface gloss, it ends up satisfying all the conditions necessary for the identification of a board sample.

• When a coated sheet registers low gloss values due to the effect of color on the surface of the sample and produces a moderately high value of lignin, it is identified as a newsprint sample or white free sheet depending on the extent of the lignin content measured.

• Colored free sheets are sometimes identified as white free sheet when the coloring is very close to white i.e., when the a* and b* coordinates of the base color are very close to 0.

• Some colors register a very high value of gloss even when printed on a low gloss paper. These kinds of paper can be identified as a coated sheet if the stiffness sensor reading also falls within range of those expected for coated sheets.

It can be seen in all these cases that it requires two parameters to be simultaneously wrong for a particular sample to be misidentified, the likelihood of which is not very high, thus providing high predictability considering the variations typically found in a waste stream.

5.0 CONCLUSIONS

A Mechatronic system for the high speed identification of paper types for has been developed. In addition to making use of current sensors, such as for gloss, lignin content, and stiffness, a new sensing technique to sense and interpret color information has been incorporated to increase the efficiency and robustness with which the paper samples can be identified. A large set of paper samples were collected to be tested with the sorting system. The samples were
chosen such that they represented most of the variations possible in a typical waste stream. All the tests performed under static conditions i.e., the samples were all at rest with respect to the sensors for the purpose of algorithm development. The lignin sensor and stiffness sensors were tested online to establish their dynamic behaviors. Specialized algorithms have been developed to process the color sensor data to ensure the compatibility with the sorting algorithms. Two algorithms based on artificial Neural Networks and Fuzzy Inference Systems have been evaluated using MATLAB. After extensive simulations, the algorithm based on Fuzzy Inference Systems was found to be the most useful with a sorting efficiency of nearly 90%. This is far higher than any existing automated sorting system efficiency.

ACKNOWLEDGEMENTS

The authors would like to thank the U.S. Department of Energy for supporting this research under the Industries for the Future Program, Forest Products Industry Agenda 2020; project number DE-FC07-00ID13880.

REFERENCES


Figure 1. Noncontact lignin sensor

Figure 2. Lignin sensor dynamic test- Alternating high and low lignin samples on a moving belt at 2.5 m/min.
Figure 3. Relative bending stiffness sensor [7].

Figure 4. Deflection of samples (a) newsprint, (b) Color free sheet, and (c) board
Figure 5. Diffuse and Specular reflections

Figure 6. Color waste sample, a newspaper insert
Figure 7. Newsprint sample and L*a*b* data
Figure 8. Black and white prints on white copy paper and processed data
Figure 9. Color flyer paper with black printing across scan line.
Figure 10. Processed L* a* b* data for a magazine paper
Figure 11. Lignin sensor data across a sample crossing different measurement zones.

Figure 12. Filtered lignin data for zone 3 in figure 11
Figure 13. Lignin sensor data variability due to sample variabilities due to print, color, and finer content.
Figure 14. Membership functions for sensor outputs based on experimental observations
Table 1. LAB statistics for printed white paper in Fig. 8

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>L*</th>
<th>a*</th>
<th>b*</th>
</tr>
</thead>
<tbody>
<tr>
<td>L.A.B.Deviation</td>
<td>27.4656</td>
<td>2.6173</td>
<td>2.8325</td>
</tr>
<tr>
<td>L.A.B.Average</td>
<td>56.7288</td>
<td>0.7371</td>
<td>-0.2394</td>
</tr>
<tr>
<td>L.A.B.Deviation_filtered</td>
<td>6.3750</td>
<td>1.6406</td>
<td>2.0179</td>
</tr>
<tr>
<td>L.A.B.Average_filtered</td>
<td>71.1617</td>
<td>0.0245</td>
<td>0.1700</td>
</tr>
</tbody>
</table>

Table 2. LAB statistics for color dyed flyer paper in Fig. 9

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>L*</th>
<th>a*</th>
<th>b*</th>
</tr>
</thead>
<tbody>
<tr>
<td>L.A.B.deviation</td>
<td>31.8480</td>
<td>3.9213</td>
<td>26.3924</td>
</tr>
<tr>
<td>L.A.B.Average</td>
<td>39.4070</td>
<td>-3.2899</td>
<td>31.9410</td>
</tr>
<tr>
<td>L.A.B.deviation_filtered</td>
<td>17.3055</td>
<td>3.9306</td>
<td>12.9763</td>
</tr>
<tr>
<td>L.A.B.Average_filtered</td>
<td>53.5225</td>
<td>-3.7342</td>
<td>44.7239</td>
</tr>
</tbody>
</table>

Table 3. LAB statistics for color magazine paper in Fig. 10

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>L*</th>
<th>a*</th>
<th>b*</th>
</tr>
</thead>
<tbody>
<tr>
<td>L.A.B.Average</td>
<td>43.8825</td>
<td>2.1940</td>
<td>24.8516</td>
</tr>
<tr>
<td>L.A.B.deviation_filtered</td>
<td>12.9507</td>
<td>39.6535</td>
<td>20.2449</td>
</tr>
<tr>
<td>L.A.B.Average_filtered</td>
<td>45.3308</td>
<td>2.5852</td>
<td>25.8290</td>
</tr>
</tbody>
</table>

Table 4. Prediction accuracy of neural network model

<table>
<thead>
<tr>
<th>Training data (sample range)</th>
<th>Test data (sample range)</th>
<th>Prediction accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - 199 ; 221 - 224</td>
<td>200 - 220</td>
<td>30%</td>
</tr>
<tr>
<td>1 - 199 ; 221 - 224</td>
<td>160 - 180</td>
<td>20%</td>
</tr>
<tr>
<td>1 - 139 ; 161 - 224</td>
<td>140 - 160</td>
<td>15%</td>
</tr>
<tr>
<td>1 - 199 ; 141 - 224</td>
<td>120 - 140</td>
<td>20%</td>
</tr>
<tr>
<td>1 - 99 ; 121 - 224</td>
<td>100 - 120</td>
<td>30%</td>
</tr>
<tr>
<td>1 - 39 ; 61 - 224</td>
<td>40 - 60</td>
<td>20%</td>
</tr>
</tbody>
</table>

Table 5. Prediction accuracy of FIS model and comparison with modified neural network

<table>
<thead>
<tr>
<th>Sample Type</th>
<th>Number of test samples</th>
<th>Number detected correctly</th>
<th>Prediction Accuracy</th>
<th>Neural Network Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newsprint</td>
<td>70</td>
<td>63</td>
<td>90.0 %</td>
<td>35.4 %</td>
</tr>
<tr>
<td>White Free Sheet</td>
<td>56</td>
<td>52</td>
<td>92.8 %</td>
<td>37.8 %</td>
</tr>
<tr>
<td>Coated Sheet</td>
<td>50</td>
<td>54</td>
<td>91.5 %</td>
<td>39.6 %</td>
</tr>
<tr>
<td>Board</td>
<td>21</td>
<td>21</td>
<td>100 %</td>
<td>31.4 %</td>
</tr>
<tr>
<td>Colored Free Sheet</td>
<td>18</td>
<td>14</td>
<td>77.8 %</td>
<td>38.6 %</td>
</tr>
<tr>
<td>Overall</td>
<td>224</td>
<td>204</td>
<td>90.4 %</td>
<td>36.6 %</td>
</tr>
</tbody>
</table>