Image Processing using Color Space Models for Forensic Fiber Detection

Elisabeth Wetzer * Hans Lohninger **

* Vienna University of Technology, Austria; Uppsala University, Sweden (e-mail: elisabeth.wetzer@tuwien.ac.at, elisabeth.wetzer@it.uu.se).

** Vienna University of Technology, Vienna, Austria (e-mail: hans.lohninger@tuwien.ac.at)

Abstract: The purpose of this study is to investigate the feasibility of automating fiber analysis in forensic science applications. In order to make self-directed collection of spectral data possible the number of measuring locations needs to be restricted to fibers only. Full scans of the samples would result in very large amounts of data of which only a small part carries actual information about the objects of interest. Images obtained by optical microscopes are used for preprocessing to find suitable candidates for measuring locations that subsequently may be used to control the microscope stage for spectroscopic measurements. This paper presents a method based on a nonlinear transform known to enhance the contrast in a way that makes segmentation on grayscale images possible. It further introduces an approach using the differences between the color channels of RGB images combined with common morphological operators to segment color images. A third application is presented that enables the search for fibers matching a query object in color based attributes, for which it is necessary to consider multiple color models. This approach reduces time and efforts significantly when trying to match one specific fiber to other samples.

1. INTRODUCTION

Forensic science has become an essential aspect of criminal investigations, especially in murder or attempted murder cases. The examination of a crime scene involves evidence technicians and crime scene analysts to collect and identify physical evidence, but often also computer scientists to trace digital hints of the suspect’s motives or social connections to the victim. The registration of physical evidence includes the analysis of ear and finger prints, handwriting, hair, and paint to name only a few. Textile fiber analysis makes up an integral part in forensic criminal investigations. Fibers often provide an essential link between a victim and its murderer. The evaluation of such links requires profound knowledge of chemical technologies but also statistical science to account for the uncertainty that accompanies evidence collection, such as contamination of samples or chance for arbitrary reasons of evidence matching.

1.1 Current Methods in Fiber Analysis

The collection of fiber samples at crime scenes is performed by sticking and detaching adhesive tape in a way that fibers are captured on the glue and preserved by subsequently double folding the tape in order to prevent any further contaminations to the sample. These tape stripes are taken to the Criminal Investigation Service for further examination by forensic technical experts. The examination consists of an initial assessment of the present fibers using an optical microscope. The fibers are categorized by their optical features, such as color, thickness, length, or curvature. These features may reveal some information regarding their origin, however there is a large variability, especially in synthetic fibers. Therefore, as a follow up step in the examination chemical data is collected from fibers of special interest. A small cut is made into the tape and the fiber is dissected from the sample. In consequence the fiber has to be cleansed with ethanol to remove any residue of glue from the tape. The fiber is mounted onto a cover slip and thereby prepared for investigation by an infrared (IR) microscope. Diffraction limits the spatial resolution of IR microscopes, resulting in a resolution of about 1-3× the wavelength, depending on the instrument. The composition of the individual fiber can be found by manual comparison with reference spectra taken from extensive databases, e.g., Peets et al. (2017). This kind of analysis for fiber identification is very time consuming and highly dependent on the experience of the forensic analyst appointed to the task, as well as the amount of samples and reference data at hand. Often it is necessary to include other means of examination, such as fluorescence microscopy, polarized light microscopy or even microchemical tests to determine melting points, to gain enough information about the fiber to reduce the risk of incorrect characterization of the forensic samples. The prevailing steps taken for fiber identification bring a few unwanted aspects, such as
• High time consumption to examine all collected samples manually in an optical microscope.
• High personnel costs due to the advanced training requirements of the analyst.
• Tape dissection is required for IR microscope inspection, which is troublesome when dealing with forensic evidence since it increases the chance of contamination and eliminates the possibility to repeat the analysis by an independent laboratory for that specific fiber.
• In cases of high impurity of the sample fiber, such as clotted blood stains, the absorption behavior of the sample is altered in the IR examination, even after treatment with ethanol.
• Elevated risk of losing single fibers after dissection from the tape.

1.2 Problems Arising from Automation

Due to various arguments, some of which mentioned in the previous paragraph, as well as easy access to computational power, the demand for fiber analysis automation has grown bigger over the past years. Since optical features of fibers do not reveal enough information to determine the origin of a fiber, a microscopic technique is required that allows for automation in the first place. Such a tool is given by Raman spectroscopy. Using Raman microscopes simplifies sample preparation and allows for the detection of very low concentrations of present substances. Therefore it has become a valuable tool in forensic science (Cho (2007), Miller and Bartick (2001), Lepot et al. (2008), Adar (2013)).

Raman spectroscopy is a scattering technique based on the inelastic scattering of photons caused by atoms or molecules. The energy is transferred and results in a frequency shift of the emitted light compared to the incident light which is characteristic to the specific scattering atom or molecule.

If the size of a sample tape is assumed to be 100mm × 40mm, a full scan using a Raman microscope would lead to 100,000 × 40,000 pixels for a resolution of 1µm. At each pixel at least 1000 intensity values are sampled by Raman spectroscopy. Assuming only 4 bytes per intensity, this would result in 100,000 × 40,000 × 1000 × 4 = 1.6×10^13 bytes, which equals 16TB of data. In many criminal cases it is not unusual to collect up to a hundred such samples at a single crime scene. The sheer data size thus makes it unfeasible to scan the full sample by Raman spectroscopy, as well as the consideration of the time required to acquire the amount of measurements.

In this paper we will present a method for image segmentation which reduces the amount of measurement sites such that spectral data will only be collected along the fibers themselves and not from the background of the tapes which for the most part is made of glue. Furthermore we will investigate how different color models, which are abstract mathematical models describing multiple methods of how colors can be represented, can provide the possibility to look for a small candidate set of fibers similar to a query fiber of interest to speed up the search for matches in the optical microscope.

This paper will introduce two procedures with different applications: one is an image segmentation method to detect the full set of fibers present in an image to allow for an automated control of the microscopic stage; the second provides the possibility for a user to browse the collected data to find matches for a specific query fiber which was selected manually.

2. STATE OF THE ART TECHNIQUES

Definition 2.1. A digital image is a function \( f : \Omega \in \mathbb{Z}^2 \rightarrow \mathbb{Z}^n \) or \( \mathbb{R}^n \), where \( f(x, y, i) \) is the amplitude or intensity for the \( i \)-th layer, \( i \in \{ 1, 2, \ldots , m \} \) corresponding to the contributions of the individual bandwidths.

Hyperspectral images have many layers, each corresponding to wavelengths or bandwidths. In the case of RGB images taken from the optical microscope three layers are considered, corresponding to the red, green and blue channels. Image segmentation describes the process of partitioning an image to label every pixel in a way that labels can be associated with certain characteristics. The goal here is to segment optical microscope images in a way that labels are assigned to all pixels that contribute to fibers in order to operate the microscope stage such that Raman spectral measurements can be taken along the fibers exclusively.

One of the simplest approaches to image segmentation is thresholding. For grayscale images a threshold \( k \in [0, 255] \) is used such that every pixel \( f(i, j) < k \) is set to 0, corresponding to black and every pixel \( f(i, j) > k \) is set to 1, corresponding to white, resulting in a binary image. A common algorithm to calculate a suitable threshold for such binary masks is Otsu’s method, see Otsu (1979). It assumes two clusters of pixels and tries to maximize the inter-class variance of these two classes. However, in the images taken of the tape samples, the number of pixels that contribute to fibers, is small compared to the amount of background pixels. In Yang et al. (2009), an image enhancement model is introduced to cope with the low contrast between the objects and the background.

Figure 1 shows a typical image of a sample strip taken by an optical microscope to the left. It shows collected fibers but also some small pieces of dirt, air bubbles and shadows caused by the illumination. Otsu’s method assumes two clusters of pixels, among which it tries to find the best threshold for segmentation. The segmented image to the right shows that this approach is not successful. The reason becomes apparent in figure 2. The amount of pixels contributing to fibers is so small, that no two clusters representing the background and the fibers respectively.
The narrow object mask ensures that Raman measurements are being taken along the fiber when the microscope stage is moved along such a skeleton. The following section explains the details of the segmentation performed.

5. SEGMENTATION OF RGB IMAGES

The algorithm subtracts one color channel from another. In the second step it computes the resulting grayscale image and uses Otsu’s method to find a suitable threshold to convert the image to a binary image. Additionally a 2D median filter using a $3 \times 3$ neighborhood is applied to reduce noise on the images and all objects smaller than a certain amount of pixels is removed to clear the binary mask from artifacts.

In a next step skeleton models are formed for each object. A morphological skeleton models the ‘backbone’ of a binary object by removing pixels from the object’s boundary without breaking it apart or changing the Euler number of the image, which corresponds to the total number of objects minus the total number of holes in the image. Subsequently all endpoints and branchpoints are removed from the objects at hand. Endpoints denote the extreme points of a binary object, marked by circles in matrix $A_c$, while branchpoints mark the sites in an object at which the structure splits in two or more substructures, so called branches, marked by the circles in matrix $A_b$.

$$A_c = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, A_b = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The removal of all these points leads to splits in the objects corresponding to fibers we are interested in, but the structures remain large enough to be preserved after another clean up which removes all objects smaller than a second threshold that needs to be determined empirically, depending on the data at hand. Many of the objects which belong to one and the same fiber have been split into many smaller objects by the removal of end- and branchpoints. To reunite these pieces of fiber masks, a modified closing operator is applied. A closing operator is a morphological operator consisting of two steps: first a structuring element $A$, in this case in the form of $(a_{ij}) \in \mathbb{R}^{7 \times 7}$, $a_{ij} = 1, \forall i,j \in \{1,2,\ldots,7\}$ dilates the image, that means the for every binary object $B$ it results in $A \oplus B := \{a+b : a \in A, b \in B\}$. The results of the intermediate steps can be seen in figure 4. In the second step the resulting images is eroded by $(a_{ij}) \in \mathbb{R}^{3 \times 3}$, $a_{ij} = 1, \forall i,j \in \{1,2,3\}$, applied as the dual operator $A \ominus B := \{a-b : a \in A, b \in B\}$.

Although many small objects resulting from noise and artifacts in the samples could be removed in the steps taken thus far, there are still objects present in the binary mask that are non-concurrent to fibers. Hence the calculation of the skeleton is repeated. The fibers are characterized by their linear, long and narrow shape, which shows only a few branches. These features are used to distinguish between the objects of interest and artifacts. Figure 5 shows a typical artifact to the left which has
Fig. 4. Various steps of the image processing procedure to clear the mask from artifacts resulting from dirt, blood or air bubbles in the sample.

Fig. 5. Magnification of the morphological skeleton of an artifact in the binary mask to the left. Branchpoints marked in red, the center of mass in yellow. To the right the skeleton of a fiber in the sample showing the typical linear, long structure with only very few to no branches.

Fig. 6. Binary mask showing the center of masses of the white binary objects, as well as additionally their branchpoints on the right figure.

automated setup using other methods. The search is based on transforming the RGB color model of a fiber into other color spaces: the HSV, YCbCr and L*a*b* color spaces, perform segmentation in those spaces based on user input and combine the resulting binary masks into a single mask indicating the position of possible matches.

6. SEARCH FOR A SPECIFIC FIBER

In addition to the segmentation of the full microscopic images, it is favorable to provide the ability to search for a candidate set that matches a specific query fiber based on the color attributes in the optical microscope. This may be included as a partial automation to the prevailing method of fiber analysis using IR spectrometry, or used in a fully

6.1 Color Models

In the setup to look for fibers that have similar color parameters as a fiber the forensic analyst is interested in, the user can choose a suitable binary mask for the fiber of interest by setting lower and upper bounds to the parameter values of different color spaces with sliders in a graphical user interface (GUI). This is repeated for all parameters which determine the degrees of freedom in the HSV, YCbCr and L*a*b* color models.

6.2 RGB Color Model

RGB stands for red, green and blue and corresponds to the human color receptors as explained in Gonzalez and Woods (2007) and Gonzalez (2011). Although it is the most common representation of digital images, it is not an intuitive way to describe colors for humans. Fairchild (2013) gives a good overview of different color spaces.

6.3 HSV Color Model

HSV is short for hue, saturation and value and known for being a color system which corresponds very well to the human color perception. This is the reason why it is popular when a user is asked to choose colors. The transformation between the RGB and the HSV color model is described in Rogers (1985) and is based on regarding the RGB model as a cube, in which one corner corresponds to black with the coordinates (0,0,0) and another to white or (1,1,1) respectively. The diagonal connecting the two corners represents the gray axis and gives the value. Considering a plane perpendicular to the gray axis, hue is given by an angle around that hexagon, while saturation can be calculated by the distance on the plane from the value-axis. The transformation from RGB to HSV space is given by Hanbury and Serra (2003) and Bhatia (2008) as follows:

These many branches with branchpoints marks in red and the center of mass of the binary object denoted in yellow. It clearly shows a different structure than that a typical fiber identifies with, which can be seen on the right.

Figure 6 visualizes all branchpoints in red and centroids - or center of masses - of objects in yellow. Since most artifacts show many branchpoints in a dense area, this information is used to clear the mask further: all objects are being removed that have a high ratios of branchpoints to centroids.

This way the binary mask returns the skeletons of the fibers present in the grayscale image. Note that this procedure has to be repeated for every combination of color channels and these binary masks have to be united in the end.

6. SEARCH FOR A SPECIFIC FIBER

In addition to the segmentation of the full microscopic images, it is favorable to provide the ability to search for a candidate set that matches a specific query fiber based on the color attributes in the optical microscope. This may be included as a partial automation to the prevailing method of fiber analysis using IR spectrometry, or used in a fully
\[ V = \max(R, G, B) \]

\[
S = \begin{cases} 
\frac{\max(R, G, B) - \min(R, G, B)}{\max(R, G, B)} & \text{if } \max(R, G, B) \neq 0 \\
0 & \text{otherwise}
\end{cases}
\]

\[
H' = \begin{cases} 
\text{undefined} & \text{if } S = 0 \\
\frac{G - B}{\max(R, G, B) - \min(R, G, B)} & \text{if } R = \max(R, G, B) \\
\frac{G - B}{\max(R, G, B) - \min(R, G, B)} & \text{if } G = \max(R, G, B) \\
\frac{B - R}{\max(R, G, B) - \min(R, G, B)} & \text{if } B = \max(R, G, B)
\end{cases}
\]

\[ a = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112.000 \\ 112.000 & -93.786 & -18.214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \]

### 6.4 YCbCr Color Model

This model stores color information in two color difference components, Cb and Cr, as the difference between a reference value and the blue component as well as the red component respectively. The Y axis corresponds to the luminance in the given image. Conversion between the RGB and YCbCr model is done as follows:

### 6.5 L*a*b* Color Model

In this model L stands for lightness and the components \(a\) and \(b\) for the color opponents green-red and blue-yellow. More precisely the user has to determine suitable thresholds for segmentation in the CIEXYZ 1976 color space as given by ISO 11664-4:2008. It was introduced by the Commission Internationale de l’Eclairage to standardize color description independent of the device in use. As with the previously introduced color models, the user picks bounds which correspond to the fiber that shall be matched to fibers in other tape samples.

### 6.6 Example

A simple modification of a pre-implemented user interface in Matlab, the Color Thresholder App, can be used as a GUI in which the user easily sets the bounds in the color spaces mentioned in the previous section. All the color spaces covered here can be described by three degrees of freedom. Let \(C_i, i \in \{1, 2, 3\}\) denote the \(i\)-th degree of freedom for each color space model, then table 1 shows the chosen values for the lower bounds \(C_{1i}\) and upper bounds \(C_{3i}\) in the different color spaces to search the sample tape for a green fiber.

| Bounds | HSV | L*a*b* | YCbCr |
|--------|-----|--------|--------|
| \(C_{11}\) | 0.278 | 77.764 | 205.000 |
| \(C_{1u}\) | 0.533 | 100.000 | 238.000 |
| \(C_{21}\) | 0.102 | -19.000 | 127.000 |
| \(C_{2u}\) | 0.252 | -7.541 | 132.000 |
| \(C_{31}\) | 0.906 | -3.336 | 100.000 |
| \(C_{3u}\) | 0.992 | 2.089 | 121.000 |

Fig. 7. Overlay of the original image and the skeleton of a fiber resulting from the binary mask computed by the subtraction of two color channels from each other followed by various image processing steps in two different magnifications.

All images which shall be searched for fibers with similar color attributes as the one which was selected to determine the thresholds for segmentation are processed in the following way: The resulting three binary masks are subsequently added to each other. Morphological operators are applied to clean the binary mask from artifacts, including the following steps in the given order: dilation by a structural element \((a_{ij}) \in \mathbb{R}^{7 \times 7}\), \(a_{ij} = 1\), \(\forall i, j\), erosion by \((a_{ij}) \in \mathbb{R}^{3 \times 3}\), \(a_{ij} = 1\), \(\forall i, j\), the removal of all objects smaller than a chosen threshold, thinning the remaining objects to lines, followed by further removal of small objects and morphological closing to thicken the remaining fiber masks. In case the segmentation with the given parameters is successful, it is saved to a candidate set which can be displayed to the user, or used for further automation to take Raman measurements at the given locations for further verification of finding matches.

7. RESULTS AND CONCLUSIONS

The first task was to use RGB color images and segment them such that automatic scanning using Raman microscopes is allowed by reducing the data load from multiple terabytes per sample to only as many measurements as necessary to perform statistics on the spectral signal resulting from the fibers. Figure 7 shows the detection of an exemplary fiber in one of the images obtained by the optical microscope as an overlay of the binary mask computed by the method introduced in section 5. The benefit of examining the skeleton of the fiber where the fiber itself shows a larger diameter itself than its mask is that it assures that the microscope will certainly take measurements in the central part of the fiber, thereby insuring that the sampling locations will be in the same depth as well as possible. This is especially favorable when considering Raman spectroscopy, since the microscope operates on a confocal principle and hence is very sensitive to the depth of focal planes considered.

The results of the search for a specific fiber, demonstrated for a green exemplary fiber are shown in figure 8 and are encouraging. In a set of 77 tape images from an light microscope all optically matching fibers were detected that also a human analyst was able to differentiate. The fact that some of the fibers were not fully detected as visible in fig. 8 is not crucial, because every fiber was at least partly detected which is enough to include it to the candidate set or in case of automation, to take measurements along parts of the fiber. The algorithm was able to cope with the
Fig. 8. Results of the query for a green fiber based on the
thresholds that were chosen as best representatives
to a query fiber in HSV, YCbCr and L*a*b* space.
All returned fiber masks show candidates for possible
matches to the fiber in the top.

variances in the focus planes, the degree of contamination
by dirt and enclosed air bubbles.

8. OUTLOOK

There are multiple difficulties that result from automated
measurements that are not addressed in this paper. Es-
specially Raman spectroscopy is very sensitive to focal
depth because of its properties as a confocal microscope.
It enables the method to have a very high resolution in the
micrometer range, however it needs further evaluation in
how a suitable focal depth can be chosen automatically
without human supervision. Furthermore initial experi-
ments showed difficulties due to the coloring of the fibers.
In many cases the fibers absorbed the energy from the laser
excitation which in some cases resulted in deformation or
bleaching of the fibers.

REFERENCES

Adar, F. (2013). Raman microscopy analysis of molecular
orientation in organic fibers. Spectroscopy, 28(2).
Bhatia, P.K. (2008). Computer Graphics. I.K. Interna-
tional Publishing House pvt. Ltd., 1 edition.
Cho, L.L. (2007). Identification of textile fiber by raman
microspectroscopy. Forensic Science International, 6,
55–62.

Fairchild, M.D. (2013). ColorAppearance Models. Series
in Imaging Science and Technology. Wiley-IS & T, 3
dition.
Gonzalez, R. and Woods, R. (2007). Digital Image Pro-
cessing. Prentice Hall, 3 edition.
Gonzalez, R.C. (2011). Digital Image Processing Using
Matlab. MHE.
Hanbury, A. and Serra, J. (2003). A 3d-polar coordinate
colour representation suitable for image analysis. Tech-
nical report, Pattern Recognition and Image Processing
Group, Institution of Computer Aided Automation, Vi-
enna University of Technology.
Lepot, L., De Wael, K., Gason, F., and Gilbert, B. (2008).
Application of raman spectroscopy to forensic fibre
cases. Applied Spectroscopy, 48, 109–117.
Miller, J. and Bartick, E. (2001). Forensic analysis of single
fibers by raman spectroscopy. Applied Spectroscopy, 55,
1729–1732.
Otsu, N. (1979). A threshold selection method from gray-
level histograms. IEEE Transactions on Systems, Man,
and Cybernetics, 9(1), 62–66.
Peets, P., Leito, I., Pelt, J., and Vahur, S. (2017).
Identification and classification of textile fibres us-
ing atr-ft-ir spectroscopy with chemometric meth-
ods. Spectrochimica Acta Part A: Molecular and
Biomolecular Spectroscopy, 173, 175 – 181. doi:
https://doi.org/10.1016/j.saa.2016.09.007.
Rogers, D.F. (1985). Procedural Elements for Computer
Graphics. McGraw-Hill, Inc., New York, NY, USA.
Yang, W., Li, D., Zhu, L., Kang, Y., and Li, F. (2009).
A new approach for image processing in foreign fiber
detection. Computer and Electronics Agriculture, 68,
68–77.