Walk This Way: Footwear Recognition Using Images & Neural Networks

Valentin Cedric Gazeau
Sam Houston State University
Department of Computer Science
Huntsville, Texas 77340, USA
vcg006@shsu.edu

William Bradley Glisson
Louisiana Tech University
Department of Computer Science
Ruston, LA 71272, USA
glisson@latech.edu

Cihan Varol
Sam Houston State University
Department of Computer Science
Huntsville, Texas 77340, USA
cxv007@shsu.edu

Qingzhong Liu
Sam Houston State University
Department of Computer Science
Huntsville, Texas 77340, USA
qxl005@shsu.edu

Abstract

Footwear prints are one of the most commonly recovered in criminal investigations. They can be used to discover a criminal's identity and to connect various crimes. Nowadays, footwear recognition techniques take time to be processed due to the use of current methods to extract the shoe print layout such as platter castings, gel lifting, and 3D-imaging techniques. Traditional techniques are prone to human error and waste valuable investigative time, which can be a problem for timely investigations. In terms of 3D-imaging techniques, one of the issues is that footwear prints can be blurred or missing, which renders their recognition and comparison inaccurate by completely automated approaches. Hence, this research investigates a footwear recognition model based on camera RGB images of the shoe print taken directly from the investigation site to reduce the time and cost required for the investigative process. First, the model extracts the layout information of the evidence shoe print using known image processing techniques. The layout information is then sent to a hierarchical network of neural networks. Each layer of this network is examined in an attempt to process and recognize footwear features to eliminate and narrow down the possible matches until returning the final result to the investigator.

1. Introduction

The continued escalation of digital evidence into court proceedings is prevalent in today's society. Researchers regularly investigate the role, impact, and challenges that residual digital data encounters in legal environments [1–4]. The integration of digital evidence is continuously expanding from born-digital to digitally converted artifacts. The continued amalgamation of traditional evidence into the digital ecosystem creates opportunities to improve and broaden investigation capabilities.

According to the FBI's Uniform Crime Reporting (UCR), forensic footwear evidence is regularly introduced into legal proceedings to establish the presence of a shoe at a crime scene [1]. They go on to indicate that it is often the most abundant form of evidence at a crime scene. According to UCR, it is estimated that 1,197,704 time-sensitive violent crimes were committed around the nation in 2015 with a little over thirty per cent (30%) of them analyzing an average of five (5) shoe prints [5].

It has also been reported that at the scene of a crime, about thirty per cent (30%) of shoe prints can be retrieved and a lifted shoe-print could be used in two separate tasks: either match the print with a database of known shoe prints, or match it against other shoe prints found at the scene [5]. Unfortunately, matching it with a database is not a trivial task, in fact, the limitations of such systems are obvious in large databases due to the need to match the sample collected with all database samples (one by one) [6] . Moreover, it is more difficult to decide on the classification between many users, and often in the case of deteriorated images of the shoe label.

Girod et Al. point out that due to the variety of surfaces on which the impressions are produced, there is variation in the consistency of footwear prints [6]. The authors go on to note that the information preserved in a shoe print may be inadequate to distinguish an individual shoe in a specific way, but are still very valuable [6]. This implies that a very small fraction of the general population will own any particular shoe model. This is due to the wide variety of shoes available on the market, with most having distinctive outsole patterns [6]. Furthermore, the same outsole motif can be used on many different brands and styles of footwear. If a shoe's model or brand pattern may be identified from its print then the hunt for a specific suspect can be greatly narrowed [6].

There is also variability in the quality of footwear impressions because of the variety of surfaces on which the impressions are made. Detail retained in a shoe print may be insufficient to uniquely identify an individual

URI: https://hdl.handle.net/10125/80264
978-0-9981331-5-7

shoe but is still very valuable [6]. Due to the wide variety of shoes available on the market, with most having distinctive outsole patterns, this implies that any specific model of shoe will be owned by a very small fraction of the general population [6]. Furthermore the same outsole pattern can be found on several different footwear brands and models. If the outsole pattern of a shoe can be determined from its mark, then this can significantly narrow the search for a particular suspect [7].

Forensic footwear evidence is frequently used in legal proceedings to help prove that a shoe was at a crime scene and is often the most abundant form of evidence at a crime scene [6]. There currently are techniques with remarkable accuracy in terms of recognizing the evidence shoe. The issue is that the majority of them take more than a few days to process, making them not very reliable in time-sensitive investigations [7]. Furthermore, these techniques require training and specific resources in order to extract the shoe print. These reasons make the whole shoe print recognition process very costly and challenging depending on the case [7]. The techniques used nowadays extract the layout of the shoe using lifting, casing and/or 3D-imaging to minimize the clutter and partial occlusion problem [8].

Currently, authorities use techniques such as lifting, casting and 3D-imaging in order to extract the shoe print from the investigation area. Lifting and casting involve extracting the shape of the shoeprint with tools such as adhesive lifters, gelatin lifters, electrostatic lifting devices or a plaster cast. The object is then sent to the lab in order for processing to produce an image of the sole patterns that is archived in a database. These techniques suffer from long processing times as they can take multiple days and are also prone to human error when extracting the shoeprint on site. Another popular technique uses a 3D-imaging capture device which is faster and more accurate than traditional techniques, but it is a costly and bulky device and thus can not be used for every investigation. The FID-300 is a public data set of shoeprint images that was created based on these techniques. This data set serves as a foundation for most current research.

The shoe prints differ in shape, color and appearance [9]. They are surrounded and partly obscured by other objects and can be placed on a large range of different surfaces [9]. Because of all of these issues, it is an inherently difficult job to identify a shoe print based on an image. These challenges have piqued the curiosity of many researchers and practitioners that span from the use of many image processing techniques to machine learning techniques. These include whole image fractal decomposition, gradient location, orientation histograms, scale invariant feature transforms, combinations of computational models and deep networks, using hybrid features and neighboring images, hierarchical classification models and the application of machine learning models towards the field of image pre-processing and shoe print identification [10].

This information prompts the hypothesis that the digitization of the extraction process using existing camera technology will improve processing capabilities for identifying forensic footwear. To address this hypothesis, the following research questions need to be addressed: Can current camera hardware produce images that can be used to extract the layout of the shoe print for recognition? Can current image processing techniques provide a good base to extract the layout of a shoeprint from an image taken at an investigation site? Can neural networks help in the recognition of a specific footwear with images of the shoe print produced by a smartphone or camera? Can the techniques be used to obtain results quicker than the current investigation process?

This research investigates these questions by creating a data set of raw digital images of shoeprints taken on a sandy surface and processing them to extract the sole pattern which are used to train a hierarchical machine learning algorithm tasked in the recognition of the shoe based on the shoeprint extraction.

The remainder of this paper is structured as follows. Section two covers the literature review. Section three describes the research methodology. Section four presents the experimental results and an analysis of the data. Section five draws conclusions and presents future work.

2. Literature Review

Current techniques employed to analyze shoe print evidence include lifting, casting, and 3D-imaging [1]. Lifting involves extracting the shoe print with tools such as adhesive lifters, gelatin lifters or electrostatic lifting devices to later analyze at a lab, this technique suffers from the fact that it is extremely complicated to extract fragile evidence such as a shoe print without tampering it making it harder to analyze [2]. The casting technique involves extracting the footprint layer by pouring a plaster mix onto the evidence shoe print which also suffers from a long process and is prone to human error [3]. The third option involves 3D-imaging technique, which is the best technique in terms of accuracy but is very costly according to the Uniform Crime Reporting Program (UCR), a branch of the

FBI [1]. The advantages and disadvantages of the extraction methods are listed in Table 1: Shoeprint Extraction Methods. Using these techniques, the FID (Footwear Impression Database) data set was created with 1175 images and is publicly available and serves as a foundation for most current research.

Table 1. Shoeprint Extraction Methods

Extraction	Advantages	Disadvantages
Lifting	-Accurate	-Prone to human error -Takes days to process -Needs to be restocked
Casting	-Accurate	-Prone to human error -Takes days to process -Needs to be restocked
3D-imaging	-Accurate -Quick	-Costly -Bulky -Quality depends on capture angle
Digital Photo	-Accurate -Quick -Mobile -Easy to use	-Partial occlusion

Rida et al. [6] present a literature survey that reviews several shoe print identification techniques. Their work identifies two main methods for the automatic classification of extracted evidence shoe prints but all of them go through the following steps: image pre-processing, feature extraction, and classification matching. The holistic or global methods seek to process the entire image as a whole using techniques such as Fourier, Gabor or Radon spatial transformations and/or fractal decomposition combined with a classification technique such as k-nearest neighbors based on Euclidean distance. The local techniques, as opposed to the global methods, extract discriminative features from sub-regions of the image with algorithms such as Maximally Stable Extremal Region (MSER) to detect points of interest combined with Gradient Location and Orientation Histogram. The local techniques have proven to be much more efficient with the prowess of machine learning, especially Convolutional Neural Networks (CNNs). Although, Rida argues that the main issues with shoeprint recognition, as a whole, include the lack of public datasets, the lack of pre-defined and standardized evaluation protocols, and that the most published techniques are evaluated on nonrealistic and/or synthetically generated images.

Wang et al. [7] proposed a multi-layer feature extractor model to compute the similarity values of two shoe prints on each layer. This method showed the success of the multi-layered Convolutional Neural Network (CNN), especially in the area of an occluded or partial image recognition. The main issue with this research is that it compares two shoe print images, which requires the investigator to have a picture of the suspected shoe print for evaluation.

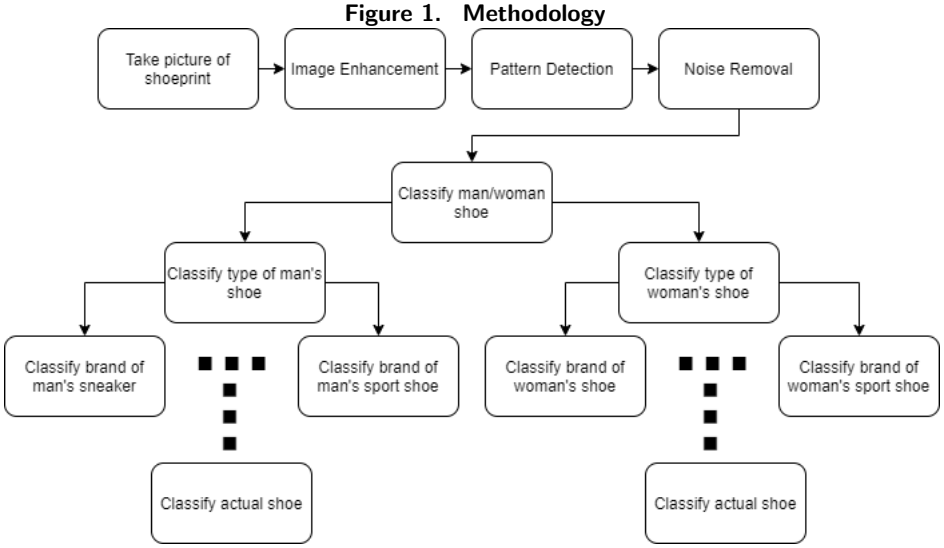
Kortylewski [5] used a holistic model-based image analysis framework to represent shoe prints as a hierarchical composition of a Laplacian-of-Gaussian basis elements that are used to detect the geometry and appearance of a shoe print in an image. A coherent occlusion model is then used to reduce the global clutter. The final model takes into account the shape, appearance variation, partial occlusion, and background clutter of the shoe print image. These image processing techniques are applied to shoe print images after it has been extracted by the traditional lifting, and casting techniques. Kortylewsky reached an accuracy of seventy-one (71) per cent with a data set comprising of 1,175 images of shoe prints that have already been extracted using the traditional methods.

Vor der Bruck et al. [8] used a hierarchical tree representation of shoe prints consisting of feature types and attribute nodes, which allows for establishing matches between entries of different features and attribute pairs to remove levels of abstraction. Recognizing the shoe based on the features and attributes of the shoe print such as tip position, size of the shoe, heel position. This technique greatly improves the accuracy and speed of the recognition process by ruling out impossible matches, especially because there are so many different kinds of shoes. However, the authors' research suffers for the same reason that it uses images of shoe print layouts that have already been extracted by the timely and costly methods discussed previously.

Qing liu et al. [9] combines compositional model techniques with Deep Convolutional Neural Networks (DCNNs) because of their powerful models that yield very impressive results at image processing and object classification. The main problem with DCNNs is that they do not generalize well to partially occluded objects, which are dominant in extracted shoe print layouts. In contrast to DCNNs, compositional models are very robust to partial occlusion, but aren't as discriminative as DCNNs. Here they retained the best of both approaches to create a discriminative model that is robust to partial occlusion which yields great results in terms of recognition accuracy. However, the application of these techniques to extracted shoe print images does not speed up the extraction process.

Ma, Zhanyu et al. [10] use a deep Convolutional Neural Network (CNN) for its praises in image processing and classification but argue that most other techniques for shoe print recognition pay too

Figure 1. Methodology



much attention to feature extraction while ignoring its distinctive characteristics. They created a model that divides the input image vertically into two parts and extract sub-features for each part with a shared feature extraction network. The model then calculates the importance of each sub-feature based on the informative pixel weight matrix, which are then concatenated as the final feature. The model finally uses the triplet loss function to measure the similarity between the query image and the gallery images. The model greatly improves accuracy as they analyze the importance of each feature in the CNN. This model also uses images previously extracted by traditional techniques.

The techniques predominantly used in industry involve extracting the layout of the shoeprint and then cross-reference with a database to identify the shoe, which is a costly and timely process [2]. Predominately, research focuses on a robust and accurate way to find the model of the shoe once the layout has already been extracted by the traditional methods such as lifting, casing and/or 3D-imaging, meaning that it would require a few days before the actual recognition process can begin. The literature review also suggests that while machine learning and image processing yield positive results in terms of accuracy, most models have weaknesses; for example, convolutional neural networks are weak to partially occluded objects. This suggests that there are research opportunities in investigating combination models to compensate for individual algorithm flaws.

Most of the current research uses the FID-300 or the CSFID-170 (which is a subset of the FID-300) data set which rely on the traditional extraction techniques that take days to process and are prone to error or are costly. Other researchers use digital images that are generated by making participants stepping over a powder then walking over a pad or chemical paper in order to achieve a near-perfect shoeprint extraction which resembles that of the FID-300 data set. The issue with this approach is that these images do not reflect real-world shoeprints found at investigation sites unless extracted from traditional methods. While research is being conducted that applies machine learning to shoe print data, minimal research examines methods to improve shoe print layout extractions methods on a digital photography of a shoeprint taken directly from the site of the investigation.

3. Methodology

This research uses a series of Artificial Neural Networks, particularly Multilayer Perceptrons, in a hierarchical model of classification to investigate improvements in accuracy and processing times. The idea is to capture a shoe print image directly from a picture taken by a smartphone or similar device, quickly process the image on-site with a high enough accuracy to make quick decisions in reference to an investigation. The entire recognition process is depicted on Figure 1: Methodology.

Before being used in the Neural Networks, the images first need to be pre-processed in order to extract the layout of the sole of the shoe. This pre-processing involves enhancing the image, sole pattern detection, noise detection and removal and finally pattern extraction.

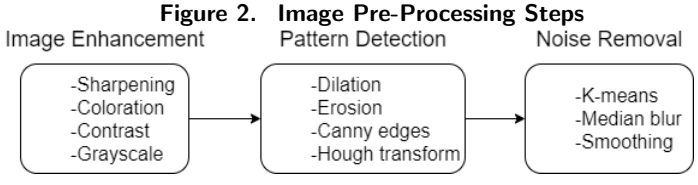
The data set created in this research consists of

raw shoeprint images acquired from the shoe collection of three different participants. The participants had to walk over a 5-meter surface of sand and the best four prints were photographed using three different devices. The first device is a Iphone X smartphone device with version 13.5.1. This mobile phone has dual 12-megapixel cameras on the rear with the main wide-angle camera being optically stabilized with an aperture widened from f/2.8 to f/2.4. The second device is a EOS T5 Rebel Cannon camera which is a digital, single-lens reflex, AF/AE camera with built-in flash that can take pictures with approximately 18.0 megapixels and was marketed in March 2014. The third device used in this experiment is Microsoft's Surface Pro 4 which features an 8-megapixel rear-facing camera with autofocus. Table 1 displays all the shoes that were used to create the data set.

3.1. Image Pre-Processing

In a picture of a shoeprint, the print is most characterized by the darkest areas of the picture which is hard to extract because of all the inconsistencies of the surface area around the actual print. For that reason, the images need to be pre-processed in order to extract the features that will best help the neural networks the recognize the shoes. The images go through a number of image processing techniques in order to achieve that goal. These steps can be grouped into three (3) categories: the image enhancement step, which improves the features of the image, the pattern detection, which helps locate where the shoe print is in the image, that way, the unnecessary information around the shoe print can be removed. The final step is the noise removal step, which extracts the most accurate layout of the shoe without all the inconsistencies within the image such as rocks, sand and others. These steps are depicted in Figure 2.

Figure 2. Image Pre-Processing Steps



3.1.1. Image Enhancement All enhancements were done using OpenCV's automated image processing methods. The first step to extract the dark areas of the print, which represent the main features of the shoe print, is to enhance the picture to differentiate the dark areas within the shoe print from those that reside outside. The first enhancement is to increase the sharpness of the picture, which refers to an image's overall clarity in terms of both focus and contrast. When the subject of an image is sharp, the image appears clear and lifelike, with detail, contrast, and texture rendered in high detail [11]. The second enhancement is to improve the color of the picture, this helps in pictures that look monotone and helps to create a gap between light and dark pixels [11]. The final enhancement done to the picture is to improve the contrast, which is a process that makes the image features stand out more clearly by making optimal use of the colors available [11] which helps to make the dark areas within the shoe print darker to ease the extraction process without getting the dark areas outside of the print. Once the image has been

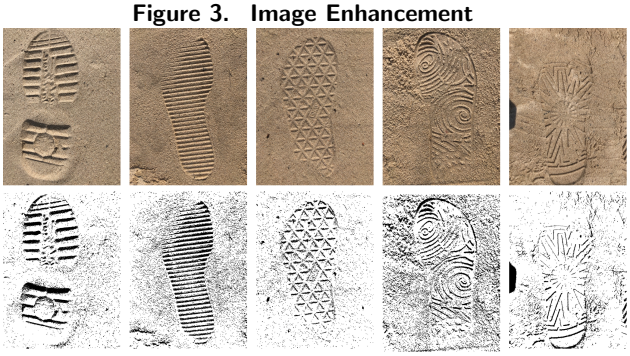
Table 2. Shoes

Brand	Model	Type	Size	Gender
UGG	Bailey Button	boots	7	woman
CASTT City	NA	boots	8	woman
1.4.3 Girl	Replay	rain boots	8	woman
Doc Martens	Flora	boots	9	woman
Addidas	Cloud Foam	sport	8	woman
Bobs	Desert Kiss	casual	8	woman
Roxy	Bayshore	sport	8	woman
Sugar	Microchip	Casual	8	woman
Lip Lover	NA	Sandals	8	woman
L'artiste	NA	sandals	8	woman
Crocs	Crocband	sandals	8	woman
Steve Madden	Annika	sandals	8	woman
Nike	Free run	sports	8	woman
Rock&Candy	Castine slide	sandals	8	woman
Ryka Aunora	Aurora	sandals	8	woman
Unionbay	NA	sandals	8	woman
Havainas	Slim	flip flops	8	woman
Billabong	Kai Sandal	flip flips	7	woman
Steve Madden	Donddi	sandals	7	woman
Carlos	Brie	boots	7	woman
Havainas	Slim	flip flips	7	woman
Teva	Original	sandals	7	woman
Athletic works	Soft Running	sport	7	woman
Shein	Nubuck	block heel	7	woman
Sugar	Noelle	block heel	7	woman
Report	NA	block heel	7	woman
Nike	Free run	sport	7	woman
Adiddas	Swift run	sport	8	woman
Converse	Shoreline slip	sneaker	8	woman
Birkenstock	Arizona	sandals	8	woman
Asics	GEL-Venture 7	sport	11	man
Parker&Sky	Oxford	dress	9	man
Tom's	Jean	dress	9	man
Levi's	Brawley Wax	dress	9	man
Timberland	Millworks 6"	boots	8	man
Sperry	Striper CVO	sneakers	9	man
Sanuk	Sknuer	sneakers	9	man
Justin	SuperIntendent	boots	9	man
Nike	Air	sport	9	man
Sperry	Gold cup	sneaker	9	man
Sperry	Penny	loafer	9	man
State street	Flex groove	loafer	9	man

enhanced, a copy is made which is used in the noise removal step later on.

The next step is to convert the image into grayscale in order to go from a three-dimensional color image (RGB) to a one-dimensional image, making it easier for processing. The grayscale image is loaded into a numpy array, and the dimensions are extracted as well as the highest and lowest pixel values. The picture is then converted to a binary image by flattening the numpy array into a one-dimensional list in order to process each pixel individually. The value of each pixel is read and converted to a 1 if the pixel value is close to the maximum pixel value based on a threshold, which is calculated with the average brightness of the image; while The rest of the pixels are converted to 0. This process eliminates a lot of the unwanted pixels and carves out the shape of the shoe print but still leaves a lot of noise around the shoeprint, which represents the darkest areas around the shoeprint created by shadows and/or other inconsistencies left by the surface on which the shoeprint is imprinted (sand, mud, rocks). The result of the enhancement techniques are shown in Figure 3.

Figure 3. Image Enhancement



3.1.2. Pattern Detection To best extract the pattern of the shoe print, the morphology functions from the OpenCV library were used. The first mathematical morphology process applied to the image is the dilation operation. The dilation's fundamental effect is to progressively expand the border of the pixels in the foreground, usually the white pixels, making those areas increase in size, while holes within that area become smaller, and for smaller holes, they simply disappear [11]. This process works with two inputs: the binary image and the kernel window. For each background pixel, the kernel is overlayed with the binary image such that the location of the kernel matches the location of the input pixel. If, in the kernel, at least one pixel matches the foreground pixel in the binary image, then the pixel is set to a foreground value which is a black pixel.

The second mathematical morphology process is called erosion. The erosion morphology is the inverse of the dilation process. This time, the borders of the foreground pixel regions are eroded, making them shrink in size and gaps grow bigger. This process works very similarly to the dilation operation, where for each foreground pixel, the kernel is overlayed on top of the binary image such that the location of the kernel matches the location of the input pixel [11]. In this case, for every pixel in the kernel window, if the corresponding pixel in the binary image is a foreground pixel; then the input pixel is left as is. However, if any of the corresponding pixels in the image are background, then the input pixel is also set to a background value which is a white pixel.

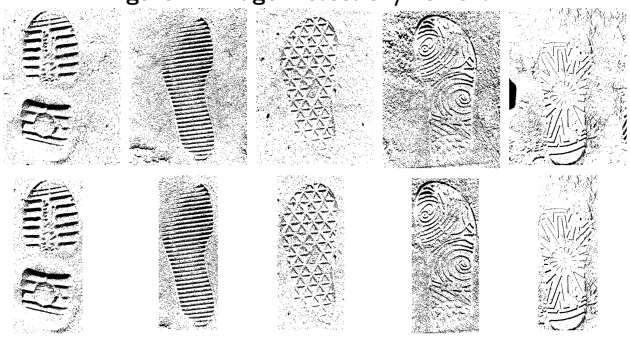
Dilation and erosion are often used in pairs; the combination of the two is called closing an image [12]. One of the dilation applications is to fill in images containing a lot of pepper noise [11]. However, one of the issues with this process is that the dilation can indiscriminately deform all pixel boundaries as well. This effect is minimized by performing an erosion operation on the picture following the dilation. Three (3) closing operations are made on the images to remove most small noise around the shoe print while obtaining the smoothest lines possible which can be detected.

The results of the removal is a picture with smooth lines that outline the shoe print. At this point, the Hough lines transform algorithm helps us detect lines in order to only process what is of interest in the image [12]. Before using the Hough transform algorithms, an image of the edges of the objects obtained from the closing operations is created. To detect the edges, OpenCV's Canny algorithm is used, which simply determines areas where the brightness or intensity of the pixels drastically change [12].

The Hough transform algorithm works by comparing points on the image of edges with it's Hough space, which is a 2D plane with a horizontal axis that represents the slope and the vertical axis on the edge image for the intercept of a line [12]. Any line on the edge image can be expressed in the form of $y = ax + b$. That means that a line from the image will create a point on the Hough space defined by its slope a and intercept b. The main issue with the original Hough transform is the fact that it can't detect vertical lines as they would have a slope of infinity. This can cause issues in processing of footprint images, as they often contain vertical lines. The enhanced Hough transform instead represents lines in the form of $\rho = xcos(\theta) + ysin(\theta)$. A line is then represented by its length ρ and its angle θ. With the enhanced Hough transform, the edge point instead generates a cosine curve on the Hough space, which eliminates the problem of unbounded

values when dealing with vertical lines and also helps in detecting curved lines. Using this technique the shape and location of the sole patterns are calculated by finding the size and center of the shoe print image which can be used to eliminate a lot of the unnecessary noise around the shoe print. The results of these techniques are shown in Figure 4.

Figure 4. Edge Detection/Removal



3.1.3. Noise Removal Even though the process up to this point helps to remove noise around the shoe print, there is still a lot of noise within the shoe print. The next step implements K-means to remove the remaining noise. K-means clustering is a vector quantization technique aiming to separate a number of observations (x) into a set of clusters (k) [13]. In those clusters, each observation corresponds to the closest cluster center (mean). The first step is to find the cluster center μ_i for each cluster by calculating the mean of all the points within cluster S_i. This is done by minimizing the squared deviations of pairs belonging to the same cluster:

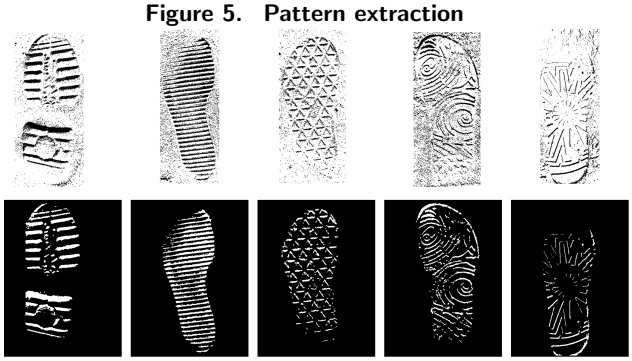
$$\arg min \sum_{i=1}^{k} \frac{1}{2\,|S_i|} \sum_{x,y \epsilon S_i} \|x - y\|^2 \qquad (1)$$

Then, the algorithm uses a set of observations $(x_1, x_2, ..., x_n)$ and divides them into groups $S = \{S_1, S_2, ..., S_k\}$ with the purpose to minimize the sum of squares within each clusters:

$$\arg min \sum_{i=1}^{k} \sum_{x \epsilon S_i} \|x - \mu_i\|^2 = \arg min \sum_{i=1}^{k} |S_i|\, Var S_i \qquad (2)$$

The K-means technique is used on the copy of the enhanced image which occured during the enhancement process. The technique clusters pixels based on their values and positions into four different groups

Figure 5. Pattern extraction



along with the mask image produced with the pattern detection. Since the darkest areas of the picture represent the shadows within the shoeprint, then the pixels within those shadows are close in value and position and they are all clustered in the same groups. After testing, it was decided that four groups yield the best results because of the different objects and the different shades of lighting around the shoe print. The resulting image is then converted to binary by keeping only the group of pixels within the shoeprint and then compared with the original binarized image to create a new image. This step removes the majority of the noise around the shoe print making it easier to further process for analysis.

The next step employs OpenCV's median blur algorithm, which processes each pixel by replacing it by the median of all the pixels in the kernel area, which is a moving window of all neighboring pixels. This operation processes the edges while removing the noise. The result image is then passed to skimage's morphology process, which removes objects and holes that are smaller to a set threshold. This helps in removing the little noise that is left around the shoeprint layout. Finally, a smoothing algorithm was created to perform a majority count of pixel values on a kernel window of 5x5 pixels in order to replace the pixel in the center of the kernel, which again helps to remove remaining noise and smooth the layout of the shoeprint itself. The results can be seen in Figure 5.

3.2. Shoe recognition

To test the quality of the extracted shoe print layouts, Python's TensorFlow and Keras libraries were used to train a series of MLPs (Multilayer Perception) each tasked to classify a specific feature of the shoe in order to recognize the actual shoe from the image of the shoe print. Before the image pre-processing, all images of extracted shoe prints were sampled to a resolution of 3024x4032 using bicubic interpolation as

Table 3. Feature Summary

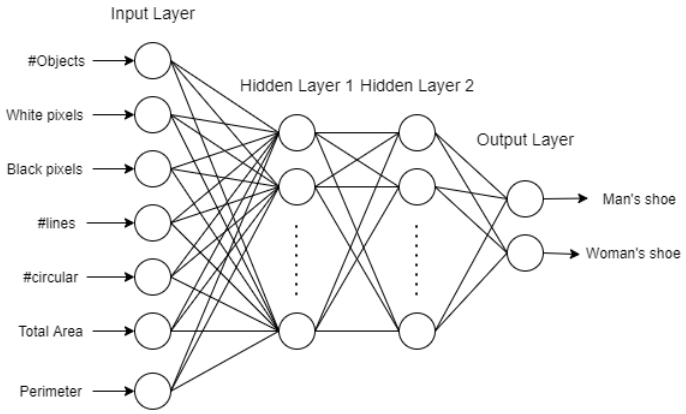
Feature	Summary
Objects	The number of objects in the shoe print. Sole patterns can be simple or complex.
White pixels	The number of white pixels in the image. Surface area of the shoe sole.
Black pixels	The number of black pixels in the image. Surface area of the shoe sole.
Lines	The number of lines in the image. Some sole patterns have a lot of lines.
Circles	Number of circles. Some sole patterns have a lot of circles.
Total area	Total surface area of the print.
Perimeter	The perimeter of the outer sole layout.

this technique creates sharp images by using the closest 4x4 neighborhood of known pixels.

The MLPs use seven (7) features that were calculated to train and classify the images. The first feature calculated is the number of objects in the shoe print, for this, the OpenCV's findContours() function was used. The second and third features are simply the number of black and white pixels in the extracted shoeprint, this is explained by the fact that some shoe soles cover greater surfaces than others and hence can help in the classification. Then the number of linear and circular shapes were calculated using OpenCV's HoughLines() and HoughCircles() functions respectively. These help in classification because each shoe sole patterns either exhibit a lot circular shapes, a lot of linear shapes or a combination of the two. The total area of each objects is also calculated to get insight on how much surface the shoe print covers. This was calculated with OpenCV's contourArea() method. The last feature that was used is the perimeter of the print, this was calculated by tweaking OpenCV's findContours() arguments and then using the arcLength() function. The feature set is shown in Table 3.

Consequently, each neural network has an input layer of 7 neurons, 2 hidden layers of variable neuron sizes depending on the number of output classes and an output layer. All networks were trained using all the data available for their specific class. The training algorithm used for all MLPs is Adam which is similar to the stochastic gradient descent which updates parameters using the gradient loss function but Adam adjusts the update parameters based on adaptive estimates. The number of epochs required varied depending on the number of training data available for each neural network. The level 1 neural network was trained using 150 epochs while the level 2 for men used 80 epochs, the level 2 for women used 100 epochs and all level 3 neural networks used 50 epochs for training. The architecture of the artificial neural networks is depicted in Figure 6.

Figure 6. Neural Network Architecture

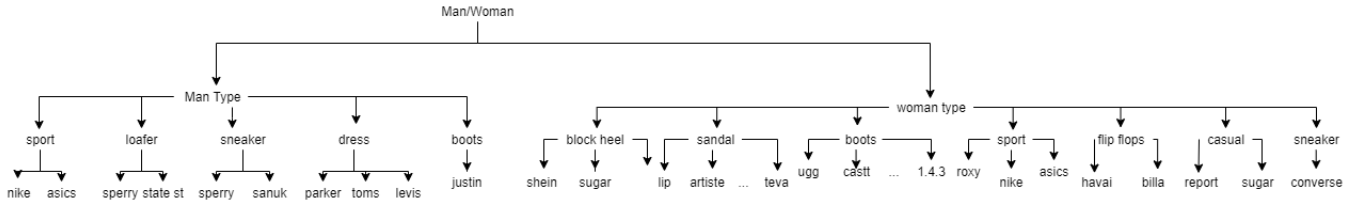


With neural networks, the more output classes there are, the less accurate it will be because it will have more room for error. This is potentially a problem for shoeprint recognition as there are hundreds of thousands of different kinds of shoes, if a neural network has that many output classes, its accuracy would suffer proportionally. To mitigate this problem, a number of neural networks are used, that way each shoe is classified hierarchically in different categories. This hierarchical strategy is shown in Figure 7. For example, the first level of neural network analyzes the layout and simply outputs whether the shoe belongs to a man or woman, which removes around half of the possibilities and reduces the output possibilities of this neural network to just two, tremendously reducing the error rate.

Then, the second level has two versions: a men-shoes version and a women-shoes version. Depending on the output of the first layer, one of the adequate neural network is called to output the type of shoe that it is, again filtering out the remaining possibilities at each level to increase accuracy. There are plenty of types of shoes, for this reason this level of neural network needs more output classes, which in turn reduces the accuracy of the network, but each type of shoe has very specific characteristics which makes the recognition easier, for example, a boot footprint looks very different from a sport's footprint. The big differences between each type of shoe print helps to strengthen the accuracy of this neural network.

The third level recognizes the brand of the shoe depending on its output type from the previous level. Even though each type of shoe has very similar outlines, each brand typically has the same or very similar patterns of soles which helps a lot in the recognition of

Figure 7. Recognition Process



the brand. For example, in this case if the shoe print to recognize belongs to a man (from level 1) and is a sport's shoe (from level 2) then it will call the appropriate neural network to recognize the brand of the man's sport shoe.

4. Results

As this research investigates the usability of digital images in the process of shoe recognition the multilayer perception was chosen because of its simplicity and its flexibility as they can be used generally to learn a mapping from inputs to outputs. In this experiment, there were a total of forty-two (42) pairs of shoes. For each of these shoes, four images were taken with three of them being used for training and the last one was kept for testing. This is a total of 126 images for training and 42 for testing. The average pre-processing time for each picture takes ten (10) minutes on average while the entire recognition process can take up to thirty (30) minutes. These measurements were made on a machine running Windows 10 Pro with an Intel Core i7-4790k CPU and 8.00 GB of RAM.

A total of 16 neural networks were trained to complete this model based on the data available. The performance of each neural network was measured using the sklearn.metrics.classification.report libraries. Table 2 shows the accuracy and confidence of each neural network used in this hierarchical approach. The results indicate that the neural networks performed well as most of them have an F-score of 100% .This indicates that digital images of shoeprints found at an investigation site can indeed be used in the process of shoe recognition. This also shows that high accuracy can be achieved with the use of multiple neural networks each tasked to classify a few classes based on hierarchical features. The results suggest that a using digital images of a shoeprint to recognize a shoe is, first, quicker than traditional methods which rely on extraction techniques that can take a few days of processing (casting, lifting etc..), second, digital images aren't prone to human error like lifting and casting techniques are and finally they are cheaper and more mobile than 3D-imaging tools.

The level 1 neural network which was trained to classify men and women's shoes has an F-score of 97.56% which is explained by the huge gap in data between the two classes. Within the forty-two (42) pairs of shoes, thirty (30) of which belong to women and only twelve (12) belonging to men. The pair of shoe that resulted as a false positive for the men is the Nike Air sports shoe which can be the fact that all sports shoes have similar sole patterns and there were more sports shoes belonging to the women class. For the women class, the shoe that resulted as a false positive was the UGG boot, which is explained by the dimension of the sole. The neural network may have helped it's learning by using the size of the shoes since, in this data set, all women shoes are about the same size and the same can be said about the sizes for men's shoe. This may be a problem for some types of shoes (like the UGG boots), but since on average men have larger shoe sizes then women, this fact helps to strengthen the accuracy and confidence of the neural network.

The metrics for the level two and level three neural networks were calculated independently assuming that the previous level was well predicted. The level two (2) of neural networks were created to recognize the type of shoe based on the first neural network assumption that it was either a man's or a woman's shoe. The different output classes were different based on the available data and the output of the level one (1) neural network. The accuracy is very strong because of two main factors: first because the number of output classes of both neural networks was minimized by first figuring out if the shoe belonged to a man or woman with the level 1 of the neural network, limiting the output layer of the level two (2) to five (5) classes for men and seven (7) classes for women. The second reason that helped the neural networks to reach an F-score of 100% is that most soles of the same types of shoe are very similar. For example, all sport shoes have a very similar sole layout, the same can be said about the size of the sole of most boots etc. For the level two (2) for women's shoes, the data was more abundant, both neural networks performed well with the lowest having an F-score of 95.65%, namely the Casual type which is explained by the lack of data in those categories: the casual type had only two pairs of shoes for training since it only had two shoes.

Table 4. Neural Networks Performances

Level 1: Man's or Woman's shoe				
Accuracy(%)	Precision(%)	Recall(%)	F-score(%)	Avg. Confidence(%)
95.23	100.00	95.23	97.56	99.27

Level 2: Type of shoe					
Neural Network	Accuracy(%)	Precision(%)	Recall(%)	F-score(%)	Avg. Confidence(%)
Man Type	91.67	100.00	91.67	95.65	95.27
Woman Type	97.67	100.00	97.67	97.78	88.57

Level 3: Brand of shoe (Man)					
Neural Network	Accuracy(%)	Precision(%)	Recall(%)	F-score(%)	Avg. Confidence(%)
Men Boots	100.00	100.00	100.00	100.00	99.99
Men Dress	100.00	100.00	100.00	100.00	99.99
Men Sneaker	100.00	100.00	100.00	100.00	99.99
Men Sport	100.00	100.00	100.00	100.00	99.99
Men Loafer	50.00	50.00	100.00	66.67	76.35

Level 3: Brand of shoe (Woman)					
Neural Network	Accuracy(%)	Precision(%)	Recall(%)	F-score(%)	Avg. Confidence(%)
Women Block Heel	100.00	100.00	100.00	100.00	99.99
Women Sandal	85.71	85.71	100.00	92.31	94.76
Women Boots	100.00	100.00	100.00	100.00	99.99
Women Sports	100.00	100.00	100.00	100.00	99.99
Women Flip Flops	100.00	100.00	100.00	100.00	99.99
Women Casual	50.00	100.00	50.0	76.35	89.93

The level three (3) of neural networks were made to recognize a specific brand from a specific types of shoes. In this case, all the neural networks performed perfectly with an F-score of (100%) except for one. This accuracy is explained by multiple factors, firstly, because of the lack of data, most of these neural networks were binary classifiers (only two output classes) meaning that there was very little chance to classify the wrong class. Secondly, as mentioned earlier, most brands have very similar sole patterns on all their shoes, even for shoes that are of different types, like the Sperry brand made sneakers and loafers but they both had very similar sole patterns, making it easier to tell them apart from other brands. That said, one neural network did not perform as well, with an F-score of only 66.67%. The Men Loafer neural network had an accuracy of only fifty percent (50%), which can explained by the fact that training data was lacking for this neural network as there were only two loafer type shoes, one of which was a Sperry brand. The Sperry brand has very similar soles on all their shoes and two pairs of shoes of the same brand were used to train the Sneaker neural network, so there is no surprise that the Sperry loafer was classified as a Sneaker instead. Additionally, for women's shoes, the Sneaker type did not need a neural network to recognize the brand since the data set only had one (1) sneaker brand. These could have been strengthened with more abundant data.

Because the data set is small, some types of shoes don't have multiple brands, which makes the recognition easier, for example, there was only one (1) pair of sneakers in the training, so if the level 2 recognized a sneaker then it has to be that specific brand. Having multiple shoes of the same type and of the same brand can be problematic as they could have different models. To solve this problem a level 4 neural network can be implemented to recognize the specific model of a particular brand. It is also important to note that this model is invariant to rotations of the picture for two reasons: first, since the investigator will take the picture on the investigation scene, they can make sure that the picture is taken correctly. They can even take multiple pictures and keep the best one for processing. Second, because the features used in the recognition are not affected by the rotation as the number of pixels, lines, circles, surface area etc... should be the same, unless part of the shoe is cut from the picture.

Using multiple levels of neural network classification strongly strengthens the accuracy at the cost of some processing speed, but it is still much faster than the traditional processes used nowadays by

lifting, plaster casts, and/or 3D-imaging techniques, which usually takes a few days. The main strength with the hierarchical approach, aside from its accuracy, is that any level of depth can be achieved and it is very modular. If a new shoe needs to be added to the training samples, instead of having to retrain every neural network; for example it can simply be retrained at the appropriate neural network level for it to gain that experience, and if an entirely new neural network needs to be trained, for example, for another brand or model, then it can simply be placed at the adequate level without having to change the whole structure of neural networks. On the other hand, the main issue with the hierarchical approach is that a lot of different neural networks were trained in order to reach maximum accuracy. This is fine for a proof of concept but since each neural network was trained differently based on their required classification, each additional neural network and/or level takes a lot of time to select the appropriate training data. This takes time to train which adds complexity to the overall architecture. This means that determining exactly what kind of classification is needed at each level is crucial.

To answer the questions asked in the introduction, current camera hardware produce images of very high resolution which can be used to extract the layout of the shoe print for recognition especially with the image processing techniques available nowadays. However, there are some factors to take into account such as the amount of lighting, and the position of the light source as it creates very different shadows across the shoe print which can produce very different results. In this case the camera flash helps create the best images. Neural networks are also helpful in the classification of images, especially if used in a hierarchical model to minimize potential errors.

5. Future Works

There are multiple aspects that can greatly enhance the results. The first is to have a more abundant data set. There are thousands of types of shoes making the manual creation of the data very complicated. That said, since the hierarchical design presented in this research is very modular, it makes it really easy to add new data since it doesn't require to re train every neural networks, the new data can simply be added to the relevant ones. For example if there is a new woman's shoe to be added to the training data, then only the relevant women neural networks will have to be retrained while the men's neural network can remain unchanged, taking advantage of the fact that each neural network is separate.

One of the main challenges is the achievement of a high accuracy level 1 neural network rate with abundant data. In this case, if the first neural network makes an error in classification, then the rest of the neural networks called will also be erroneous in their classification. Even though on average, men have larger shoe sizes then women, which is a feature that can help strengthen this accuracy, there are always some exceptions which the neural network can easily be fooled by. Research needs to investigate multiple architectures of neural networks to find the best machine learning algorithm for the level 1 in order to minimize the risk of erroneous classification.

Some brands of shoes have the same sole patterns on different types of shoes creating another challenge. For example, within the data set, the Sperry brand makes two different types of shoes (loafers and sneakers) but both types of shoes have the same sole pattern, which greatly lowered the accuracy of the neural network tasked to recognize the type. This is common among some brands of shoes and it will remain a problem, but even though the exact model of the shoe can't be found, this model still classifies the brand of the shoe with strong confidence, which can still be extremely useful for an investigation. Moreover, this data set did not exhibit different models of shoes that have the same type and the same brand. Nevertheless, knowing the brand of the shoe is still valuable for an investigator.

An additional improvement that can increase the performances of the neural networks are the selection of the neural networks' features. In machine learning, the quality of the features in the data set has a major impact on the quality of the insights you will get while using the data set for machine learning. Finding the best features that can help to solve the problem has always been a very important task in artificial intelligence. In this case only seven (7) features were used, the features were selected based on assumptions about what kind of information will best help the neural networks to recognize the specific characteristics of the shoe print. The more meaningful features that are available, the better the neural networks will perform. There is an array of features that can be used in this kind of classification. Research needs to investigate which features are the most relevant in shoe print recognition.

This hierarchical neural network architecture can also be applied to a variety of problems with the main advantage being that the more the problem is separated into sub-problems, the higher the performance of each neural network. Another advantage is that the model can be tweaked and new training data can be added without having to re-train every neural network. On the other hand, having multiple levels of neural networks adds a lot of complexity and can be time consuming to initially

design and train all the neural networks, however the execution is still quicker than traditional methods. An additional weakness is the fact that it takes more time to use as it calls multiples neural networks, making this type of model only usable on non-time sensitive recognition that require high accuracy.

References

[1] FBI. 2020. The Forensic Analysis Of Footwear Impression Evidence. [online] Available at: https://archives.fbi.gov/archives/about-us/lab/forensic-sci-ence-communications/fsc/july2009/review/2009_07_review02.htm.

[2] A. Girod, C. Champod, and O. Ribaux, Trace de Souliers. Lausanne,Switzerland: Presse polytechniques et universitaires romandes, 2008.

[3] V. S. S. Mikkonen and P.Heinonen, "Use of footwear impressions incrime scene investigations assisted by computerized footwear collectionsystem," Forensic Science International, vol. 82, no. 1, 1996, pp. 67–79.

[4] R. Davis, "An intelligence approach to footwear marks and toolmarks,"Journal of the Forensic Science Society, vol. 21, no. 3, 1981, pp. 183–193.

[5] A. Kortylewski, "Model-based image analysis for forensic shoe print recognition," Ph.D. dissertation, Basel University, 2017

[6] I. Rida, S, Bakshi, X. Chang, and H. Proenca, "Forensic Shoe-print Identification: A Brief Survey," Journal of Latex Class files, vol. 14, no. 8, august 2015

[7] Wang, Xinnian & Wu, Yanjun & Zhang, Tao. (2019). "Multi-Layer Feature Based Shoeprint Verification Algorithm for Camera Sensor Images". Sensors. 19. 2491. 10.3390/s19112491.

[8] T. Brück and T. Stadelmann, "Semi-Automated Footwear Print Retrieval Using Hierarchical Features," The Eleventh International Conference on Emerging Networks and Systems Intelligence, 2019.

[9] Qing Liu, Huiyu Wang et al. "Combining Compositional Models and Deep Networks For Robust Object Classification under Occlusion" Preprint/early-stage research.

[10] Ma, Zhanyu, et al. "Shoe-Print Image Retrieval With Multi-Part Weighted CNN." IEEE Access, vol. 7, 2019, pp. 59728–59736., doi:10.1109/access.2019.2914455.

[11] A. Alahi, R. Ortiz, and P. Vandergheynst. Freak: Fast retina keypoint. In Computer vision and pattern recognition (CVPR), 2012 IEEE conference on, pages 510–517. Ieee, 2012. 14

[12] Mukhopadhyay, Priyanka, and Bidyut B. Chaudhuri. "A survey of Hough Transform." Pattern Recognition 48.3 (2015): 993-1010.

[13] Dhanachandra, Nameirakpam, Khumanthem Manglem, and Yambem Jina Chanu. "Image segmentation using K-means clustering algorithm and subtractive clustering algorithm." Procedia Computer Science 54 (2015): 764-771.

[14] Wu, Yanjun, et al. "Crime Scene Shoeprint Retrieval Using Hybrid Features and Neighboring Images." Information, vol. 10, no. 2, 2019, p. 45., doi:10.3390/info10020045.

[15] Egger, Bernhard, et al. "Occlusion-Aware 3D Morphable Models and an Illumination Prior for Face Image Analysis." International Journal of Computer Vision, vol. 126, no. 12, 2018, pp. 1269–1287., doi:10.1007/s11263-018-1064-8.

[16] Kong, Bailey, et al. "Cross-Domain Image Matching with Deep Feature Maps." International Journal of Computer Vision, vol. 127, no. 11-12, Apr. 2019, pp. 1738–1750., doi:10.1007/s11263-018-01143-3.

[17] Cui, Junjian, et al. "Robust Shoeprint Retrieval Method Based on Local to Global Feature Matching for Real Crime Scenes." Journal of Forensic Sciences, vol. 64, no. 2, 2018, pp. 422–430., doi:10.1111/1556-4029.13894.

[18] Ali, Redha, and Hussin K. Ragb. "Fused Deep Convolutional Neural Networks Based on Voting Approach for Efficient Object Classification." 2019 IEEE National Aerospace and Electronics Conference (NAECON), 2019, doi:10.1109/naecon46414.2019.9057795.

[19] Van der Walt, Stefan, et al. "scikit-image: image processing in Python." PeerJ 2 (2014): e453.

[20] Levin, Nadav. "The Forensic Examination of Marks." Interpols Forensic Science Review, Sept. 2017, pp. 51–69., doi:10.1201/ebk1439826584-3.

[21] Qi, Yonggang, et al. "Sketch-Based Image Retrieval via Siamese Convolutional Neural Network." 2016 IEEE International Conference on Image Processing (ICIP), 2016, doi:10.1109/icip.2016.7532801.

[22] Boonsuk, Ratcha, et al. "An Investigation on Facial Emotional Expression Recognition Based on Linear-Decision-Boundaries Classifiers Using Convolutional Neural Network for Feature Extraction." 2019 11th International Conference on Information Technology and Electrical Engineering (ICITEE), 2019, doi:10.1109/iciteed.2019.8929985.

[23] Zhou, Yuan, et al. "A Spatial Compositional Model for Linear Unmixing and Endmember Uncertainty Estimation." IEEE Transactions on Image Processing, vol. 25, no. 12, 2016, pp. 5987–6002., doi:10.1109/tip.2016.2618002.

[24] Lee, Sejeong, et al. "Occlusion Detector Using Convolutional Neural Network." 2017 International Conference on Control, Automation and Information Sciences (ICCAIS), 2017, doi:10.1109/iccais.2017.8217564.

[25] Baiker-Sørensen, Martin & Herlaar, Koen & Keereweer, et al. (2020). "The forensic examination of marks review: 2016 to 2018". Forensic Science International: Synergy. 10.1016/j.fsisyn.2020.01.016.

[26] Gazeau V, Varol C. Automatic spoken language recognition with neural networks. Int. J. Inf. Technol. Comput. Sci.(IJITCS). 2018;10(8):11-7.

[27] Graves, L., W.B. Glisson, and K.-K.R. Choo, LinkedLegal: Investigating social media as evidence in courtrooms. Computer Law & Security Review, 2020. 38: p. 105408.

[28] Brown, A.J., W.B. Glisson, T.R. Andel, and K.-K.R. Choo, Cloud forecasting: Legal visibility issues in saturated environments. Computer Law & Security Review, 2018.

[29] Kynigos, C., W.B. Glisson, T.R. Andel, and J.T. McDonald. Utilizing the Cloud to Store Camera-Hijacked Images in Hawaii International Conference on System Sciences (HICSS-49). 2016. Kauai, Hawaii.

[30] McMillan, J., W.B. Glisson, and M. Bromby. Investigating the Increase in Mobile Phone Evidence in Criminal Activities. in Hawaii International Conference on System Sciences (HICSS-46). 2013. Wailea, Hawaii: IEEE.

[31] Yang M, Jiang H, Tang Y. Shoe Pattern Recognition: A Benchmark. InChinese Conference on Biometric Recognition 2019 Oct 12 (pp. 405-414). Springer, Cham.

[32] Luostarinen T, Lehmussola A. Measuring the accuracy of automatic shoeprint recognition methods. Journal of forensic sciences. 2014 Nov;59(6):1627-34.

[33] Chengqing Tang and Xuejing Dai, "Automatic shoe sole pattern retrieval system based on image content of shoeprint," 2010 International Conference On Computer Design and Applications, 2010, pp. V4-602-V4-605, doi: 10.1109/ICCDA.2010.5540740.

[34] V. Ramakrishnan and S. Srihari, "Extraction of shoe-print patterns from impression evidence using Conditional Random Fields," 2008 19th International Conference on Pattern Recognition, 2008, pp. 1-4, doi: 10.1109/ICPR.2008.4761881.

[35] V. Ramakrishnan and S. Srihari, "Extraction of shoe-print patterns from impression evidence using Conditional Random Fields," 2008 19th International Conference on Pattern Recognition, 2008, pp. 1-4, doi: 10.1109/ICPR.2008.4761881.

[36] C. Wei and C. Gwo, "Alignment of core point for shoeprint analysis and retrieval," 2014 International Conference on Information Science, Electronics and Electrical Engineering, 2014, pp. 1069-1072, doi: 10.1109/InfoSEEE.2014.6947833.

[37] Petraco, Nicholas DK, et al. "Statistical discrimination of footwear: a method for the comparison of accidentals on shoe outsoles inspired by facial recognition techniques." Journal of Forensic Sciences 55.1 (2010): 34-41.

[38] Rida, I., Fei, L., Proença, H., Nait-Ali, A., & Hadid, A. (2019). Forensic shoe-print identification: A brief survey. arXiv preprint arXiv:1901.01431.

[39] M. Gueham, A. Bouridane, D. Crookes and O. Nibouche, "Automatic Recognition of Shoeprints using Fourier-Mellin Transform," 2008 NASA/ESA Conference on Adaptive Hardware and Systems, 2008, pp. 487-491, doi: 10.1109/AHS.2008.48.

[40] P. de Chazal, J. Flynn and R. B. Reilly, "Automated processing of shoeprint images based on the Fourier transform for use in forensic science," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 3, pp. 341-350, March 2005, doi: 10.1109/TPAMI.2005.48.