

# Formation Pattern Classification of Paper Based on Imaging Using Multiple Light Sources

Ka Young Lee\*, Seongho Lim, Youngsoo Lee and Young Bin Pyo

Digital Analysis Division, National Forensic Service, 10 Ipchun-ro, Wonju, Gangwondo, 26460, Republic of Korea

## Abstract

The analysis and identification of paper is a major research topic in the field of forensic document examination. Damage to documents serving as test specimens must be reduced to a minimum; thus, a non-destructive analysis method must be used for analyzing the paper that comprises the documents. Various non-destructive analysis and identification methods are currently being developed for this purpose. Herein, multiple light sources were used to perform non-destructive optical inspection of office paper specimens of major brands used in South Korea to verify the possibility of analysis and identification. Additionally, the images obtained through the non-destructive analysis were applied in a deep learning algorithm to test whether paper specimens from the same brand could be automatically classified.

**Keywords:** Forensic science • Document examination • Paper identification • Non-destructive optical inspection • Formation pattern • Deep learning

## Introduction

### Paper analysis and classification in forensic science

Paper is one of the most commonly used materials for documentation, and its importance is indisputable [1]. There are many ways in which paper can be classified, one of which is according to its use, such as office (printing) paper, writing and painting paper, packaging paper, and sanitary paper [2]. As types of paper differ according to their use, basic raw material, and fabrication process, the physical and optical properties of paper vary [3-5]. The analysis of paper involves identifying the main raw materials and pulp and additives as well as investigating the composition and physical properties. After pulp, a filler is the material that accounts for the largest proportion in the composition of paper; a filler fills the inside of paper for improving the writing and printing properties. The type and composition ratio of a filler in paper can be determined through Scanning Electron Microscopy with Energy Dispersive X-ray (SEM-EDX) spectroscopy, and the results can be utilized to approximately identify the brand of the office paper [6,7].

In the field of forensic science, the same documents may be used multiple times for investigations and in trials between interested parties after investigation, as well as commonly reanalyzed for other issues; therefore, it is common to use non-destructive methods [8-11]. However, the most commonly used analysis methods in other fields destroy the specimens; thus, such methods cannot be effectively applied in forensic science. Formation is a physical property indicating the uniformity of distribution and texture of pulp fibers in paper, and can be observed and analyzed from an image of the pulp when viewed through a projected light source. Formation is predominantly determined by the manufacturing process, which is particularly dependent on

the production line of the manufacturing factory of the paper; other factors such as the ratio of composition also affect the formation. As the manufacturing lines of paper companies are large structured systems that are specific to each company and difficult to modify once established, the properties of paper produced are unique to each company and are expected to not undergo changes for a long period of time. In this study, we conducted preliminary research on the formation of different paper specimens. We tested whether office paper could be classified according to the manufacturing company based on their formation patterns observed via optical analysis. Furthermore, we investigated whether automated classification is possible using deep learning and the results derived therefrom.

### Convolutional neural network

Pattern recognition technology automatically extracts pattern information from various types of data. It is classified broadly as a branch of computer science and narrowly as artificial intelligence technology. Pattern recognition technology aims for computers to be able to mimic the audio-visual cognitive abilities of humans, for whom pattern recognition refers to the identification and understanding of external phenomena via the processing, analysis, and integration of information transmitted through sensory organs. It is frequently used for applications such as economic forecasting using economic indicators, understanding the meaning of sentences by recognizing.

### The shape of the letters and identifying people by recognizing biometric data

Convolutional Neural Network (CNN) is a deep learning algorithm used for extracting salient features or attributes from pattern recognition data and classifying the input data into identifiable classes [12-14]. CNN has recently become a commonly used method in the field of pattern recognition as it can find unique features in a large number of training images and classify them, and subsequently automatically classify new images into predefined categories by extracting the pre-trained unique features [15].

## Methods

**Specimen preparation and optical image acquisition:** For this study, 200 sheets of commercially available office paper from eight major paper companies in South Korea were used. Transmitted light images of the office paper specimens were obtained using a video spectral comparator with multiple light sources under the conditions summarized in Table 1 [16]. Prior to

\*Address for Correspondence: Ka Young Lee, Digital Analysis Division, National Forensic Service, 10 Ipchun-ro, Wonju, Gangwondo, 26460, Republic of Korea, Tel: +821046085628; E-mail: leeky39@korea.kr, 39leeky@gmail.com

Copyright: © 2023 Lee KY, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Received: 28 August, 2023; Manuscript No. jfr-23-111402; Editor Assigned: 30 August, 2023; PreQC No. P-111402; Reviewed: 15 September, 2023; QC No. Q-111402; Revised: 21 September, 2023, Manuscript No. R-111402; Published: 28 September, 2023, DOI: 10.37421/2157-7145.2023.14.566

image acquisition, a white calibration standard (Serial No. 8804577) was used to adjust the white balance of the video spectral comparator [17]. The images were obtained from random locations on the sheets of paper (Figure 1).

**Machine learning method and training data acquisition using CNN:** For training data, we prepared 1,600 transmitted light microscopy images of each of the eight types of office paper selected for the study. To extract the unique properties of the formation pattern and color observed using transmitted light, 25 random areas with dimensions of 128 × 128 pixels each were extracted from each input image. A total of 40,000 images were used as training data.

## Results and Discussion

### Transmission image analysis

The paper formation images of the specimens were obtained using an infrared light source with a 530–630 nm wavelength range and a filter. Then, combining with the transmitted light and changing the infrared light filter to obtain a wavelength range of 645–1000 nm, the formation patterns of the specimens were observed. When the 830-nm wavelength was combined with the transmitted light, the specimen identification results were superior to those obtained using the other wavelengths. In particular, it was difficult to identify and distinguish the formation patterns of specimens when using low wavelengths [18].

### Ultraviolet transmission image analysis

Different wavelengths of ultraviolet light—365, 312 and 254 nm—were used to screen the samples and obtain paper formation images, and were then combined with the transmitted light to observe the formation patterns. The results of combining the 365-nm wavelength and the transmitted light are summarized in Figure 2. However, it was difficult to distinguish the specimens from one another [19].

During the fabrication process of office paper, fluorescent substances are added to the surface to turn the yellow color of the pulp into white. Each manufacturer varies the fluorescent substance and treatment processes used to achieve their desired whiteness of the paper. The fluorescent substances

on the paper surface can be differentiated by ultraviolet light detection. The pattern and extent of fluorescence that reacts with the ultraviolet light varies according to the content of the fluorescent substances. Therefore, it was hypothesized that different paper specimens can be more easily discerned by observing images with combined ultraviolet and transmitted light. However, the image analysis results from the combined ultraviolet and transmitted light showed that distinguishing the paper was more difficult than when using visible and infrared light, as the formation patterns were harder to observe in the images obtained from ultraviolet light.

### Visible light transmission image analysis

The specimens were identified from the images using a combination of visible and transmitted light. The visible light allowed for the intrinsic colors of each paper to be observed clearly, and the uniformity and texture of the pulp fibers distributed throughout the paper were easy to observe as well. The experimental image results are presented in Figure 3. To investigate whether paper specimens could be distinguished and identified based on the above results in Figure 4, a representative image in Figure 5 was compared with randomly selected specimens. The best rate of identification achieved as a result of this comparison was 92% (737 out of 800 specimens, as identified by a document examiner).

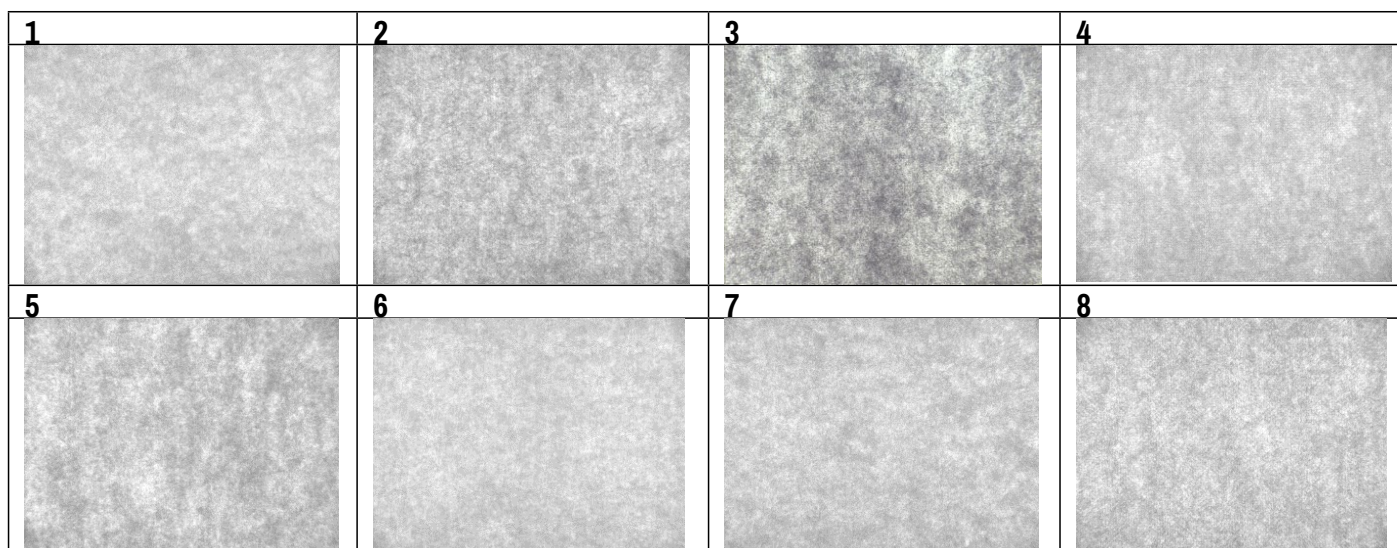
However, there were deviations in the results owing to the subjectivity of the document examiner. To resolve this issue, we implemented the CNN learning model to establish a more objective method for color and formation pattern analysis, and developed a system that automatically derives identification results [20].

### Automated identification using CNN

In this study, CNN learning was attempted by using the extracted random images as input data. The eight companies that the office paper specimens were sourced from were chosen as the categories. The unique properties of each paper that were observed from the transmitted light were automatically determined and stored in each layer of the CNN. Additionally, neural network weights, which minimized the error through error backpropagation, were calculated. The performance of the system was verified through the trained weights. For system validation, the images of paper specimens of known identities other than the images used in training were used as input, with 50

**Table 1.** Conditions for optical analysis of transmission images using a video spectral comparator with multiple light sources.

	Iris (%)	Intergration (ms)	Magnification (x)	Wavelength
Visible light-transmitted light	70	33 (30 ~ 35)	8	-
IR-transmitted light	90	140	8	830 nm
UV-transmitted light	90 ~ 91	15	8	365 nm



**Figure 1.** Formation images of each specimen obtained using combined infrared light (830 nm) and transmitted light.

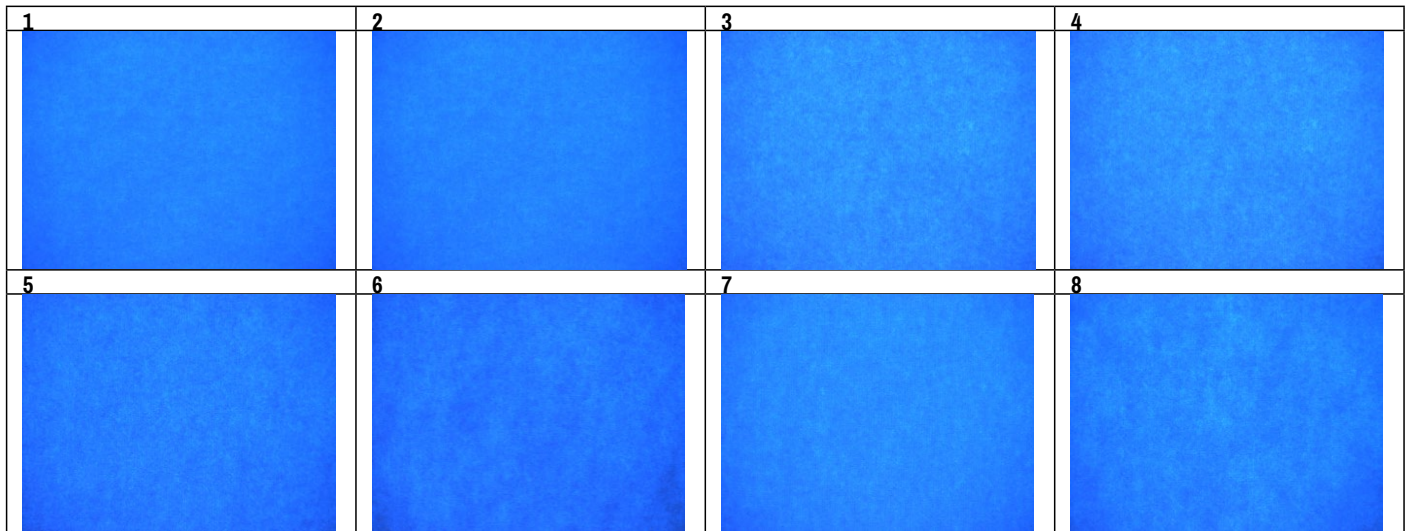


Figure 2. Formation images of each specimen obtained using combined ultraviolet light (365 nm) and transmitted light.

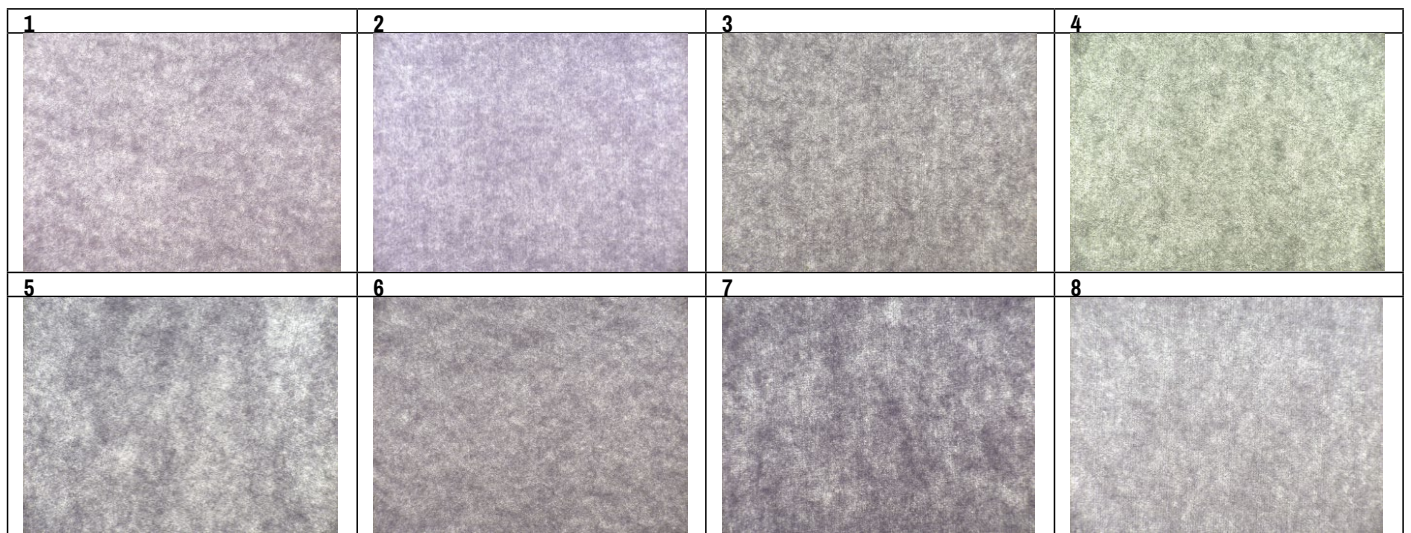


Figure 3. Representative images of the specimens observed obtained using combined visible and transmitted light.



Figure 4. Identification test: Eight types of paper formation images were presented, and the document examiner had to select the image that was most similar to the test specimen (green dotted line).

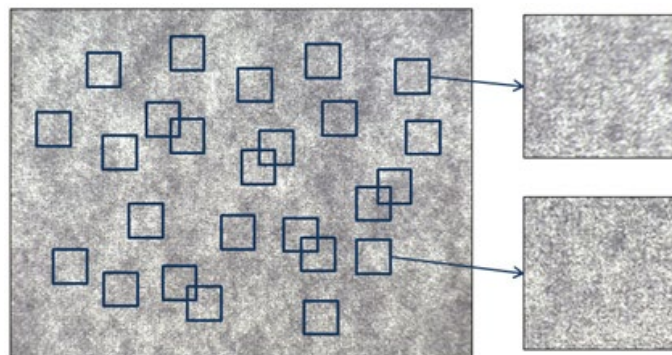


Figure 5. Extraction of random areas from specimen.

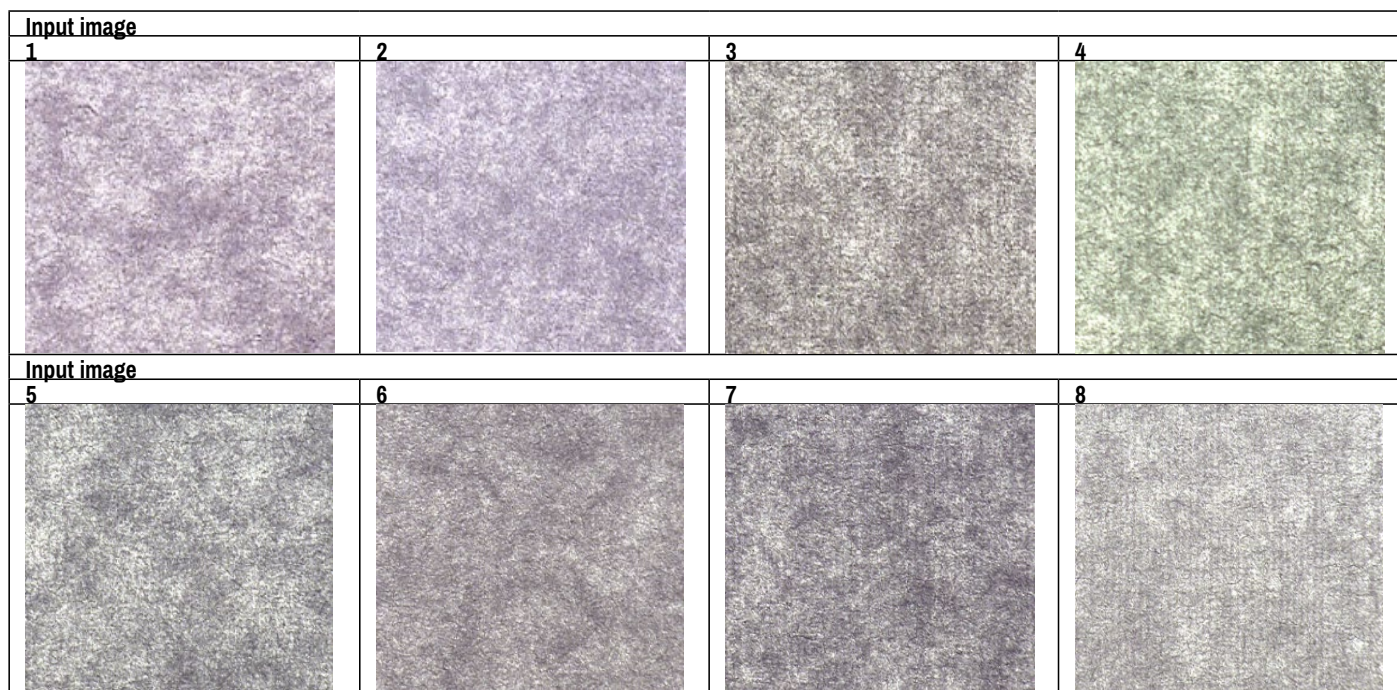
areas being extracted and entered into the trained CNN. After the results were derived, weights were given to the results of the 50 areas

A Gaussian kernel was applied to assign a higher weight to the extracted images closer to the centers of the original transmitted light images to consider optical properties such as the lens aberration of the imaging device. As clearer images were obtained from the areas closer to the centers of the images, the classification results were more reliable. The results for each area was assigned a non-negative real number between 0 and 1 such that the sum of all the results from all eight categories was 1. After applying the weight of the term to the probability value of each category, the sum of the derived results from all areas was calculated, and the category with the highest score was selected as the final result, and the sample was identified as belonging to that category.

The initial CNN model was initialized with random values, and was trained using gradient descent to produce the optimal results. Therefore, due to the initial random weight, different results were derived when the training was repeated, even for the same data. In this study, results were derived from three sets of training, which are shown in Table 2. The three sets of results exhibited an average identification performance of 97.28% (Figure 6).

**Table 2.** The three sets of results exhibited an average identification performance of 97.28%.

	Experiment 1	Experiment 2	Experiment 3	Average
Training			8 types, 40,000 images	
Accuracy of classification results	97.34%	97.14%	97.35%	97.28%

**Figure 6.** Area images randomly extracted from specimens.

## Conclusion

In this study, we identified paper from formation images obtained from non-destructive optical analysis, which involved the combination of transmitted light with various wavelengths of light from different sources. The paper identification rate from formation images was highest when visible light and transmitted light were combined. Document examiners were able to identify each paper using formation images at a maximum identification rate of 92%. By training with a deep learning CNN model, the average rate of identification was increased to 97.28%. The results therefore demonstrate the advantage of using deep learning for obtaining objective and quantifiable analysis results in comparison to traditional paper identification, which depends on the observation and subjective determination of the examiner.

## Acknowledgement

This work was supported by National Forensic Service (NFS2023DTB02), Ministry of the Interior and Safety, Republic of Korea.

## Conflict of Interest

The authors state no conflict of interest.

## References

- Sjostrom, Eero. "Wood chemistry: Fundamentals and applications." *Elsevier* (2013).
- Nanko, H., A. Button and D. Hillman. "The world of market pulp. ce swann, ed. appleton, Wisconsin, USA: WOMP, LLC." ISBN (2005): 0-615-13013-5.
- CJ, Bier Mann. "Paper manufacture. Handbook of pulping and papermaking." (1996).
- Barnard, J. A. W., D. E. Polk and B. C. Giessen. "Forensic identification of papers by elemental analysis using scanning electron microscopy." *Scanning Electron Microscopy* (1975): 5-19.
- Spence, Lindsay D., Anthony T. Baker and John P. Byrne. "Characterization of document paper using elemental compositions determined by inductively coupled plasma mass spectrometry." *J Anal At Spectrom* 15 (2000): 813-819.
- Bown, Richard. "Particle size, shape and structure of paper fillers and their effect on paper properties." *Pap Technol* 39 (1998): 44-48.
- Zhao, Yulin. "Improvement of paper properties using starch-modified precipitated calcium carbonate filler." *Tappi J* 4 (2005): 3-7.
- Ellen, David, Stephen Day and Christopher Davies. "Scientific examination of documents: Methods and techniques." CRC Press (2018).
- Bisesi, Michael S. "Scientific examination of questioned documents." CRC press (2006).
- Brunelle, Richard L. and Robert W. Reed. "Forensic examination of ink and paper." Springfield, IL: CC Thomas (1984).
- Levinson, Jay. "Questioned documents: A lawyer's handbook." Academic Press (2000).
- Gatys, Leon, Alexander S. Ecker and Matthias Bethge. "Texture synthesis using convolutional neural networks." *Adv Neural Inf Process* 28 (2015).
- Cimpoi, Mircea, Subhansu Maji and Andrea Vedaldi. "Deep filter banks for texture recognition and segmentation." *Proc IEEE Int Conf Comput Vis Recognit* (2015): 3828-3836.
- Lin, Tsung-Yu and Subhansu Maji. "Visualizing and understanding deep texture representations." *Proc IEEE Int Conf Comput Vis Recognit* (2016): 2791-2799.
- Krizhevsky, Alex, Ilya Sutskever and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Adv Neural Inf Process* 25 (2012).
- Taigman, Yaniv, Ming Yang, Marc'Aurelio Ranzato and Lior Wolf. "Deepface: Closing the gap to human-level performance in face verification." *Proc IEEE Int Conf Comput Vis Recognit* (2014): 1701-1708.
- Wang, Jiang, Yi Yang, Junhua Mao and Zhiheng Huang, et al. "Cnn-rnn: A unified

- framework for multi-label image classification." *Proc IEEE Int Conf Comput Vis Recognit* (2016): 2285-2294.
18. Hou, Le, Dimitris Samaras, Tahsin M. Kurc and Yi Gao, et al. "Patch-based convolutional neural network for whole slide tissue image classification." *Proc IEEE Int Conf Comput Vis Recognit* (2016): 2424-2433.
19. Simonyan, Karen and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv* (2014): 1409-1556.
20. Girshick, Ross, Jeff Donahue, Trevor Darrell and Jitendra Malik. "Region-based convolutional networks for accurate object detection and segmentation." *IEEE Trans Pattern Anal Mach Intell* 38 (2015): 142-158.

**How to cite this article:** Lee, Ka Young, Seongho Lim, Youngsoo Lee and Young Bin Pyo. "Formation Pattern Classification of Paper Based on Imaging Using Multiple Light Sources." *J Forensic Res* 14 (2023): 566.