TRACEFINDER – FORENSIC SCANNER IDENTIFICATION

Internship Project Documentation

Student Name: Vishnu Navghare

Project Duration: 2 Months

Domain: Artificial Intelligence / Image Forensics

INDEX

---

1. Introduction

In the digital era, scanned documents are widely used for official and legal purposes. However, document forgery and tampering have become common challenges.
The **TraceFinder – Forensic Scanner Identification** project focuses on identifying scanner-specific characteristics from scanned document images using image processing and machine learning techniques.

---

2. Project Objective

The main objectives of this project are:

- To collect scanned document images from multiple sources
- To analyze scanner-based image characteristics

- To preprocess images for forensic analysis
- To prepare data for machine learning model training

---

3. Dataset Description

The dataset consists of scanned document images collected from various sources.
The dataset is divided into the following categories:

- Original Documents
- Official Scanned Documents
- Wikipedia Document Scans
- Flatfield Images
- Tampered Document Images

Each category contains images in JPG / PNG format with different resolutions and quality levels.

---

4. Tools & Technologies

The following tools and libraries were used:

- Python
- OpenCV
- NumPy
- Matplotlib
- PIL (Pillow)
- Jupyter Notebook

---

5. Week 1: Dataset Collection & Analysis

5.1 Objective

To understand the dataset structure and analyze basic image properties.

5.2 Tasks Performed

- Loaded dataset folders using Python
- Counted number of images in each category
- Displayed sample images from every class
- Extracted image properties such as:
    - Resolution
    - Color mode
    - File size
    - Image format

5.3 Observations

- Images have varying resolutions depending on scanner devices
- Some images contain visible noise and brightness variations
- Flatfield images show uniform texture patterns
- Color and contrast differences are noticeable across sources

---

5.4 Conclusion (Week 1)

Week 1 successfully helped in understanding the dataset and identifying the need for image preprocessing to maintain consistency across images.

---

6. Week 2: Image Preprocessing

6.1 Objective

To standardize and enhance image quality for machine learning readiness.

6.2 Preprocessing Steps

6.2.1 Image Resizing

- All images were resized to a fixed dimension
- Ensures uniform input size for models

6.2.2 Grayscale Conversion

- RGB images converted to grayscale
- Reduces computational complexity

6.2.3 Noise Reduction

- Gaussian blur applied for denoising
- Removes scanner-induced noise

6.3 Results

- Improved image clarity
- Reduced noise and distortion
- Consistent image dimensions across dataset

6.4 Conclusion (Week 2)

Image preprocessing successfully prepared the dataset for feature extraction and model training.

---

7. Results & Observations

- Dataset is now clean and standardized
- Scanner-specific texture patterns are preserved
- Data quality improved significantly after preprocessing

---

8. Conclusion

The TraceFinder project has successfully completed dataset analysis and preprocessing stages.
These steps form a strong foundation for upcoming phases such as feature extraction, model training, and scanner classification.

---

9. Future Scope

- Feature extraction using texture-based methods
- Machine learning model training
- Deep learning-based scanner identification
- Real-time document authenticity verification