Computer Vision - Project

VideoBookScanner: From Flipping Pages to Seamless PDFs

Group Members:

Yash Maniya (B20CS033)

Tatvam (B20CS077)

Ruthvik K (B20AI037),

Dhruv Viradiya (B20CS079)

Github

Problem Statement:

This project aims to develop a computer vision solution to automate the conversion of a video showing page flips into a structured PDF document. The primary challenge involves accurately detecting page flips amidst motion and deformations in the video frames. To address this, the system utilizes advanced computer vision techniques such as ShiTomasi corner detection and for identifying key features and tracking motion between frames. Additionally, GrabCut is employed for segmenting and isolating page regions, while perspective transformation techniques are utilized to correct distortions caused by camera angles. The system must robustly handle variations in lighting, camera angles, and page sizes to ensure accurate extraction of content. The goal is to streamline the document digitization process by automating the conversion of video recordings into easily readable and organized PDF documents, suitable for archival purposes and efficient access to information.

Literature Review:

- **ShiTomasi Corner Detection:**
  - ShiTomasi corner detection is a popular method for identifying key features in an image.
  - It operates by detecting corners with high "cornerness" scores, indicating distinctive points in the image.
- **Lucas-Kanade Optical Flow:**
  - Lucas-Kanade optical flow is utilized for tracking motion between consecutive frames in a video.
  - It estimates the motion of key points by minimizing the error between observed and predicted image intensities.
- **GrabCut Algorithm**:
  - The GrabCut algorithm is employed for segmenting and isolating regions of interest within an image or video frame.
  - It utilizes iterative graph cuts to partition the image into foreground and background regions, facilitating object extraction.
- **Perspective Transformation:**
  - Perspective transformation techniques are applied to correct distortions caused by the camera's viewpoint.
  - These techniques involve estimating transformation matrices between corresponding points in source and destination images and applying them to warp the image accordingly.
- **Importance of Robust Techniques:**
  - The literature underscores the significance of robust feature detection, motion tracking, segmentation, and perspective correction in document digitization tasks.
  - By leveraging these advanced computer vision techniques, researchers aim to automate and streamline the process of converting video recordings of flipping pages into structured and easily readable PDF documents.

Methodology:

The methodology involves integrating optical flow analysis, page flip detection, and image processing techniques to convert a video of page flips into a PDF seamlessly.

1. Optical Flow Analysis *(Lucas-Kanade method)*:

Optical flow analysis, specifically the Lucas-Kanade method, is a technique used to track the movement of feature points within a sequence of frames in a video. This method is particularly useful for tasks like motion analysis and detecting events such as page flips in documents. Let's break down the process into more detail:

Goal: The goal of using the Lucas-Kanade method for optical flow analysis is to track specific feature points within a document across consecutive frames of a video. By doing so, we can analyze the motion of these points and potentially detect significant events like page flips.

Method:

- ➢ **Initialization**: The first step is to read the initial frame from the video. This frame is then converted to grayscale, resulting in an image referred to as old_gray. Parameters for the Lucas-Kanade optical flow algorithm are defined at this stage, including those for corner detection.
- ➢ **Feature Detection**: The Shi-Tomasi corner detection algorithm, implemented as ***cv.goodFeaturesToTrack***, is used on the old_gray image to detect feature points. These feature points are denoted as p0.
- ➢ **Optical Flow Calculation:**

    For each subsequent frame: The frame is converted to grayscale (referred to as frame_gray). The Lucas-Kanade method, implemented as ***cv.calcOpticalFlowPyrLK***, is employed to calculate optical flow between the old_gray and frame_gray images. This process tracks the feature points (p1) from the initial frame to the current frame, estimating their movement.

- ➢ **Trajectory Visualization:** Finally, the trajectories of the tracked feature points are plotted on the document. This visualization helps in understanding the movement of these points over the course of the video frames.

2. Page Flip Detection:

Goal: The goal of page flip detection is to differentiate between stable frames (where the document remains stationary) and frames where a page flip occurs, indicating a change in the document's state.

Method:

- ➢ **Motion Strength Calculation:** The motion strength between consecutive frames is calculated by comparing the positions of tracked points, likely obtained through optical flow analysis such as the Lucas-Kanade method. This calculation assesses how much movement or displacement has occurred between frames, which can indicate the presence of a page flip.
- ➢ **Thresholding**: A threshold value is defined based on the calculated motion strength. This threshold serves as a cutoff point to distinguish between frames

with minimal motion (stable frames) and frames with significant motion (potential page flips). If the motion strength between consecutive frames exceeds this threshold, it's considered indicative of a page flip event.

- **Visualization**: The detections of page flips are visualized in a plot that represents the motion strength over the sequence of frames. This visualization helps in understanding when and how frequently page flips occur throughout the video footage. Peaks or spikes in the plot above the threshold value correspond to instances where page flips are detected.

3. Image Processing for Region of Interest (ROI) Extraction:

Goal: The primary aim is to identify stable frames in a video where no page flips occur and extract the relevant content from these frames for further analysis.

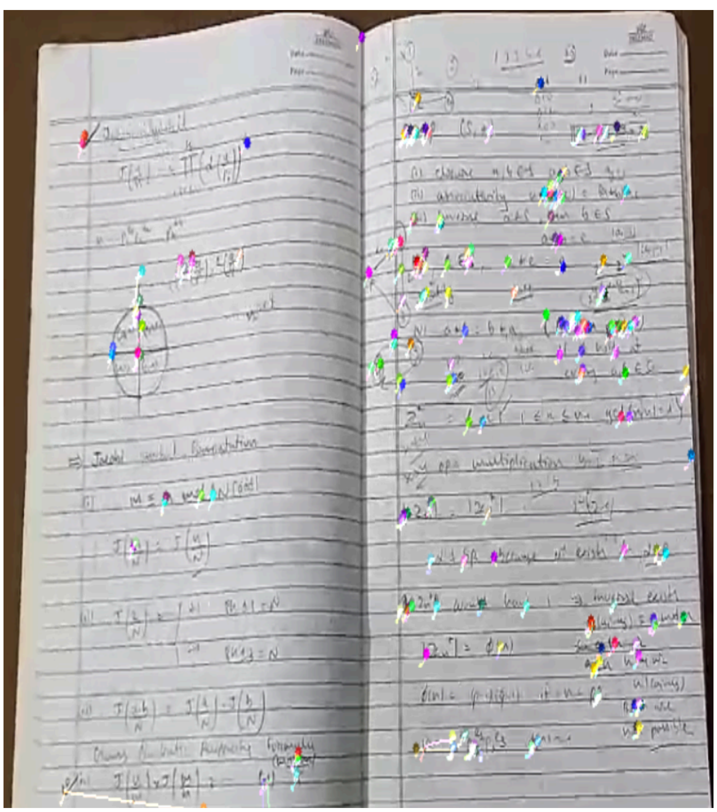
Method:

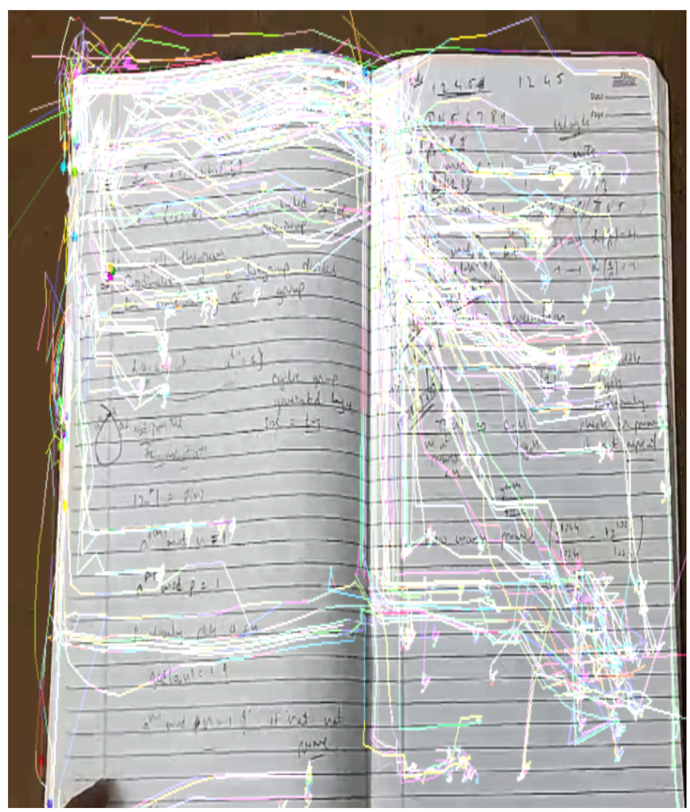
- **Stable Frame Capture:** In this step, frames without any page flips are singled out for processing. These frames are considered stable, indicating that the document content remains unchanged.
- **Preprocessing**: Once a stable frame is identified, preprocessing steps are applied to prepare it for content extraction. This includes:
  - **Resizing**: If the frame's size is too large, it may be resized to make it more manageable for subsequent operations.
  - **Morphological Operations:** Techniques such as closing are employed to remove any noise or unwanted text from the document, ensuring that only relevant content is retained.
- **Edge Detection:** Edge detection techniques, like the Canny edge detector, are applied to the preprocessed frame to identify the edges within the document. This step helps delineate the boundaries of the document content.
- **Contour Detection:** Contours are then identified within the edge-detected image. Contours represent the outlines of objects in the image—in this case, the boundaries of the document.
- **Perspective Transformation:** With the document's boundaries identified, the next step involves approximating its corners using contour approximation techniques. This provides a basis for transforming the document into a flat, scanned-like view. A perspective transformation matrix, also known as a homography, is calculated. This matrix enables the conversion of the document from its original perspective to a flat representation, akin to a scanned image.
- **ROI Extraction**: Finally, the perspective transformation matrix is applied to the stable frame to extract the region of interest (ROI). The ROI contains the document content, now presented in a flat, easily readable format.

So the methodology involves leveraging the Lucas-Kanade optical flow algorithm to track feature points within a document while visualizing their trajectories, enabling the analysis of motion patterns. Additionally, by analyzing the motion strength, the system can effectively detect page flips during the video playback phase. It also includes a process to capture stable frames devoid of page flips, ensuring that these frames can undergo further processing. During the processing stage, these stable frames undergo a series of operations such as preprocessing, edge detection, contour identification, and perspective transformation. These operations collectively facilitate the extraction of the document's region of interest (ROI), which is crucial for creating a seamless transition from flipping pages to generating a PDF.

Results:

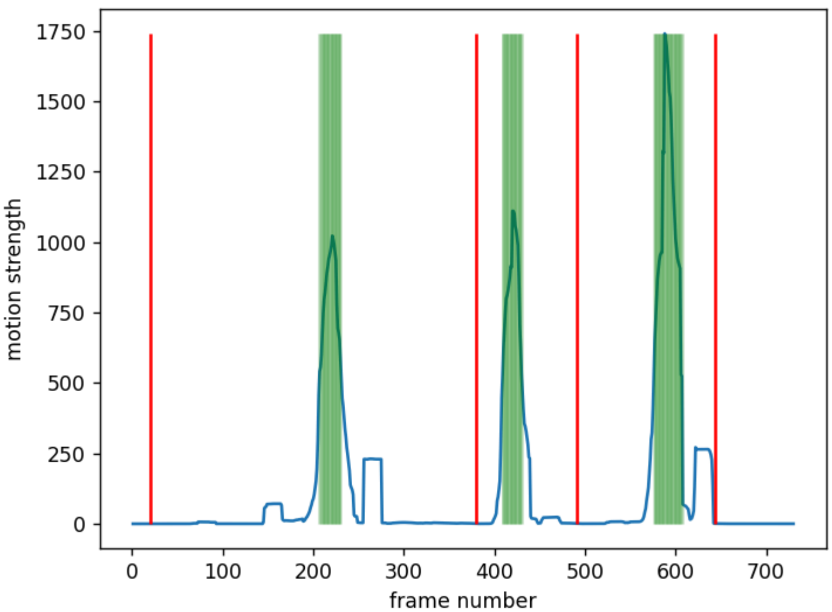
To generate a seamless PDF, users input a video depicting the page-flipping of a document. During the page-flipping process, significant features are identified, as depicted in the accompanying image, showcasing the capture of crucial features in the video. These features are subsequently utilized in the calculation of the Lucas-Kanade optical flow, as



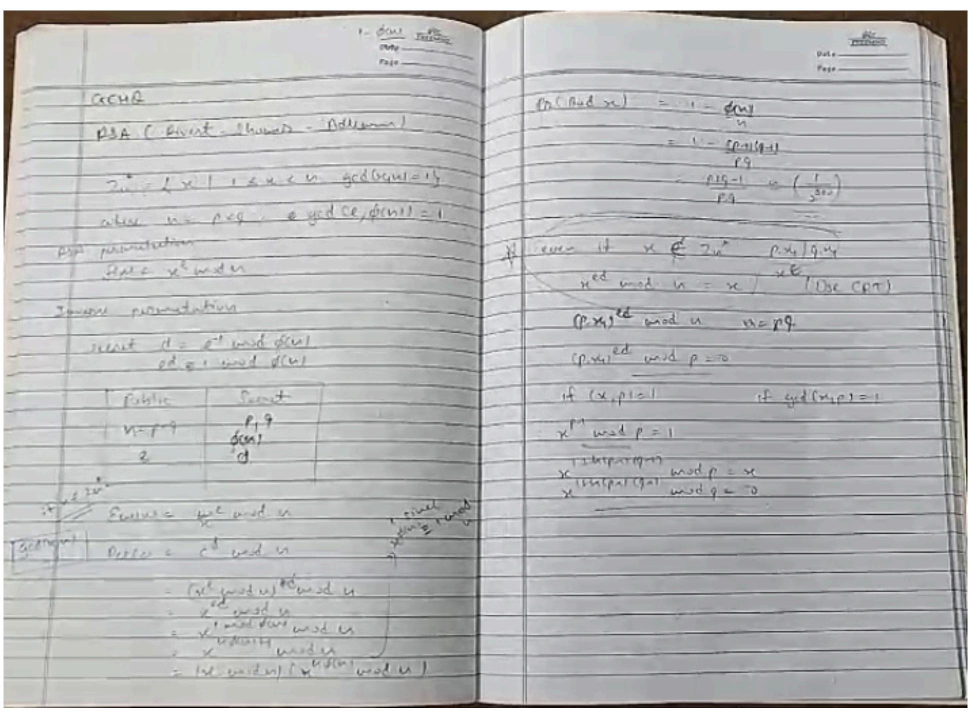


illustrated in the second image. The optical flow generated demonstrates the movement observed as pages are flipped within the document. This process aids in the analysis of motion patterns and the tracking of feature points across frames, contributing to the effective detection of page flips and the extraction of document content.

The above image illustrates the time-dependent motion of the optical flow. The blue line reflects the magnitude of the optical flow in this context. A frame is taken when the optical flow hits its minimum, signifying a consistent frame rate. When the pages are turned, a distinct green peak becomes visible in the optical flow, indicating the occurrence of a page flip event. After detecting this, the system gets ready to take the next frame. This iterative procedure of capturing frames takes place when there is minimal movement in the visual scene, guaranteeing that all frames in the movie are acquired in a sequential manner. The method effectively utilizes optical flow analysis to accurately detect page flips and adaptively modifies frame capture, enabling efficient processing of document footage.



This way we can create a seamless pdf file like the following:



Observations and Conclusion:

Observations:

- **Optical Flow Analysis**: The optical flow analysis reveals intricate motion patterns within the document video. By tracking feature points across frames, we observe the dynamic movement of document elements, such as text and images. This analysis provides valuable insights into the flow of information and changes in document structure over time.
- **Page Flip Detection:** The page flip detection mechanism effectively identifies instances where pages within the document are flipped. Through the analysis of optical flow changes, we detect distinct peaks corresponding to page flip events. This reliable detection method ensures that significant transitions in the document's state are accurately captured and recorded.
- **Frame Capture Synchronization:** The synchronization of frame capture with minimal optical flow intervals ensures the systematic extraction of stable frames containing document content. By leveraging periods of low motion activity, the algorithm optimizes the timing of frame capture, resulting in a comprehensive and coherent representation of the document's contents.
- **Robustness and Accuracy**: The methodology demonstrates robust performance in various document video scenarios, capturing subtle motion changes and accurately detecting page flips. The combination of optical flow analysis and page flip detection results in a robust framework for document analysis, capable of handling diverse document types and video conditions.

Conclusion:

The project successfully demonstrates the application of optical flow analysis and page flip detection in processing video footage of documents. By tracking feature points and analyzing optical flow patterns, the system accurately identifies page flips, enabling the extraction of document content from stable frames. This methodology provides a robust framework for automating the analysis of document videos, facilitating tasks such as motion analysis and page flip detection. Moving forward, further refinements to the algorithm and optimization of parameters could enhance its efficiency and applicability across a wider range of document types and video conditions. Overall, the project highlights the potential of optical flow-based techniques in document analysis and underscores their significance in streamlining document processing workflows.

References:

- https://learnopencv.com/automatic-document-scanner-using-opencv/
- https://medium.com/pixel-wise/detect-those-corners-aba0f034078b
- https://medium.com/@VK_Venkatkumar/optical-flow-shi-tomasi-corner-detection-sparse-lucas-kanade-horn-schunck-dense-gunnar-e1dae9600df
- https://docs.reportlab.com/