

Scantron95945 class Documentation

Bubble Scan AI module currently uses `Scantron95945` class to process scanned images of Scantron 95945 sheets. The class takes a PDF file path as input and performs the following operations:

1. Extracts all images from the PDF file and saves them in a subfolder named after the PDF.
2. Aligns the extracted images using template matching with a provided template image.
3. Extracts regions of interest (ROIs) from the aligned images, including the first and second answer columns and the student ID section.
4. Processes the ROIs to detect filled bubbles and extract the student's answers and ID.
5. Returns the processed data as a JSON object containing the student ID and their answers for each question.

The code utilizes various image processing techniques such as thresholding, contour detection, and template matching using tools like OpenCV library, PyMuPDF library for extracting images from PDF files.

Member functions

`__init__(self, pdf_path)`

- Initializes the `Scantron95945` object with the provided PDF file path.
- Calls the following methods in sequence:
 - `extractImagesFromPdf()`: Extracts images from the PDF file.
 - `template_matching()`: Aligns the extracted images using template matching.
 - `extractROIs()`: Extracts ROIs from the aligned images.

`extractImagesFromPdf(self)`

- Extracts all images from the PDF file and saves them in a subfolder named after the PDF.
- Scales the images to match the size of the scantron95945 template (1689x2186 pixels).

`align_image(self, image, template)`

- Aligns the input image with the provided template image using feature-based matching (ORB) and homography estimation.
- Returns the aligned image.

`template_matching(self)`

- Aligns all extracted images using the `align_image()` method and saves the aligned images in a separate folder.

`crop_roi(self, image_path)`

- Crops the ROIs (first answer column, second answer column, and student ID section) from the aligned image.
- Saves the cropped ROIs in separate folders for each image.

`extractROIs(self)`

- Calls the `crop_roi()` method for each aligned image to extract ROIs.

`get_responses_bubble_row(self, image, num_choices=5)`

- Processes a single row of bubbles to detect the filled bubble(s).
- Returns the corresponding letter(s) (A-E) for the filled bubble(s) or None if no bubble is filled.

`find_rows(self, image)`

- Detects the boundaries of each answer row in the image using vertical projection.
- Returns a list of tuples representing the start and end coordinates of each row.

`roi(self, image, start_question_num, num_choices=5)`

- Processes an answer column ROI to extract the responses for each question.
- Returns a dictionary mapping question numbers to their corresponding answers.

bubble_column(self, column, num_bubbles=10)

- Processes a single column of bubbles to detect the filled bubble.
- Returns the index of the filled bubble (0-9) or None if no bubble is filled.

student_id(self, roi, num_columns=10, num_bubbles=10)

- Processes the student ID ROI to extract the student ID.
- Returns the student ID as a string, with 'X' for any undetected digits.

extract_responses(self)

- Processes all the extracted ROIs to extract the student ID and answers for each student.
- Returns the processed data as a JSON object containing the student ID and their answers for each question.

Custom class Documentation

This documentation provides a step-by-step guide on how to create a custom class for processing scanned images of custom sheets, leveraging the functionality of the `Scantron95945` class. By following these steps, you can adapt the code to work with your specific custom sheet layout and extract the desired information from the scanned images.

Step 1: Analyze Your Custom Sheet Layout

Before creating a custom class, you need to analyze the layout of your custom sheet to identify the key markers and regions of interest (ROIs) that will be used for cropping and extracting data. Take note of the following:

- The positions of the top markers that indicate the start and end of each answer column and student ID column.
- The positions of the left markers that indicate the start and end of the answer rows within each column.
- Any additional markers or regions specific to your custom sheet which have essential information.

Step 2: Create a Custom Class

Create a new Python class that inherits from the `Scantron95945` class. This custom class will contain the modified `crop_roi()` function to find the coordinates of the following markers. Here's an example:

```
class CustomSheet(Scantron95945):  
    pass  
    # replace pass with crop_roi function specific to custom sheet
```

Step 3: Override the `crop_roi()` Function

Override the `crop_roi()` function in your custom class to match the specific layout and marker positions of your custom sheet. This function is responsible for cropping the answer column and student Id ROIs from the aligned image based on the positions of the top and left markers.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

FEED IN THIS DIRECTION

Left Markers

1 2 3 4 Top Markers

KEY ID
A B C D

SCORING & PRINTING OPTIONS:
☐ RESCORE ☐ MULTIPLE ANSWER SCORING
☐ CORRECT ANSWER ☐ MARK X ☐ TOTAL ONLY
MARK ONLY ONE

Scantron Test Sheet
50W/W
Form No. 95945

ANSWER KEY INFO.
OF KEYS
ITEM COUNT

PERFORMANCE ASSESSMENT
% OF TOTAL SCORE
POINTS EARNED

STUDENT ID NUMBER

SCANTRON.

Reorder Form No. 95945
www.ScantronStore.com
800.722.6876

MARKING INSTRUCTIONS
Use a No. 2 Pencil
Fill oval completely
Erase cleanly

NUMBER CORRECT
PERCENT CORRECT
ROSTER NUMBER
SCORE
RESCORE

COMBINED POINTS EARNED
COMBINED PERCENT CORRECT
LETTER GRADE
SCORE
RESCORE

NAME
SUBJECT
PERIOD DATE

Scantron Mark Reflex® M-95945-8:654321 © 2008 Scantron Corporation ALL RIGHTS RESERVED.

Above is a reference image showing how region of interest for the scantron95945 sheet are cropped into 3 ROI, 2 answer columns and 1 student id column using top and left markers. The top – left corner of the sheet is not considered for this process which is marked in black.

KEY ID
☐ A ☐ B ☐ C ☐ D

SCORING & PRINTING OPTIONS:
☐ RESCORE ☐ MULTIPLE ANSWER SCORING
☐ CORRECT ANSWER ☐ MARK 1 ☐ TOTAL ONE
 MARK ONE ONE

Scantron Test Sheet
50W/W
 Form No. 95945

ANSWER KEY INFO
 # OF ALN
 ITEM
 COUNT

PERFORMANCE ASSESSMENT
 % OF
 TOTAL
 SCORE
 POINTS
 EARNED

STUDENT ID NUMBER

ROI 1
 ROI 2
 ROI 3

SCANTRON.
 Reorder Form No. 95945
 www.ScantronStore.com
 800.722.6876

MARKING INSTRUCTIONS
 Use a No. 2 Pencil
 Fill oval completely
 Erase cleanly

NUMBER CORRECT
PERCENT CORRECT
ROSTER NUMBER
SCORE
RESCORE

COMBINED POINTS EARNED
COMBINED PERCENT CORRECT
LETTER GRADE
SCORE
RESCORE

Modify the `crop_roi` function according to your custom sheet layout:

- Adjust the regions of interest for top and left markers based on the dimensions of your sheet.
- Customize the marker collection logic to identify the key markers specific to your sheet.
- Modify the ROI coordinate calculations based on the positions of the markers in your sheet.
- Crop the answer column ROIs from the image using the calculated coordinates.

Step 4: Customize Other Functions (Optional)

If your custom sheet requires additional processing or has a significantly different layout compared to the standard Scantron95945 format, you may need to customize other functions in your custom class. Some functions you might consider modifying include:

- `get_responses_bubble_row`: Modify this function if your custom sheet has a different number of answer choices per row or if the bubble positions differ from the standard format.
- `find_rows`: Adapt this function if your custom sheet has a different row layout or requires a different approach to identify the answer rows.
- `roi`: Customize this function if your custom sheet has a different number of answer choices or if the ROI extraction logic needs to be modified.
- `bubble_column`: Adjust this function if your custom sheet has a different number of bubbles per column or if the bubble positions differ from the standard format.
- `student_id`: Modify this function if your custom sheet has a different student ID format or layout.
- `extract_responses`: Adapt this function if your custom sheet requires additional processing steps or if the response extraction logic needs to be customized.

Make sure to thoroughly test your modified functions to ensure they work correctly with your custom sheet.

Step 5: Custom Class Usage

To use custom class for processing scanned images of custom sheet, create an instance of the class by providing the path to the PDF file containing the scanned images

```
custom_sheet = CustomSheet(pdf_path)
```

The `CustomSheet` class will inherit the functionality of the `Scantron95945` class, including image extraction, template matching, and ROI extraction. The modified `crop_roi` function will be used to crop the answer column ROIs based on the specific layout of your custom sheet.

You can then use the inherited methods from the `Scantron95945` class, such as `extract_responses`, to process the extracted ROIs and obtain the desired information from your custom sheet.