Almanac To iCal

Team Name: Not Devoid

Members

- Rishabh Agrawal - 202012038 - ECE
- Atharva Gogate - 2020112001 - ECD
- Akshit Gureja - 2020112004 - ECD
- Puru Ojha - 2022701007 - MS-CS

Mentor TA: B.V.K

Repo Link:

Final Presentation

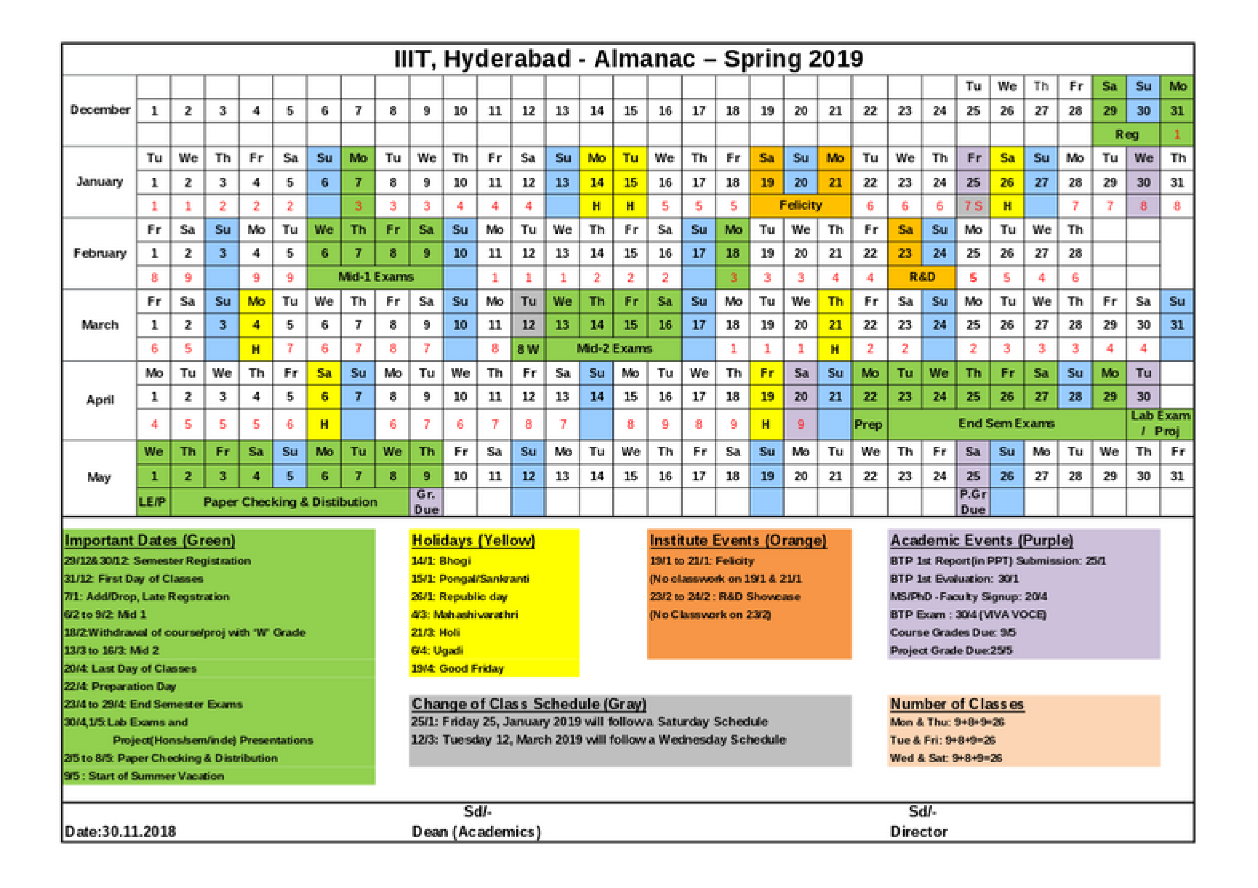
Problem Statement

We are given an almanac which we have to export to an online calendar. We have assumed that the standard almanac template to be followed is that provided by IIIT-H and the project is done to fit the requirements of that particular format. This project has a significant practical advantage: it allows the user to automatically set their calendar without having to go through the arduous process of manually setting dates.

DATASET

Datasets

Data Set is taken from IIIT's Intranet portal, and it contains Almanacs from Monsoon'17 until present. Among them we have excluded the Almanac's made for trimester format during online classes because their overall structure is different from the usual almanac.

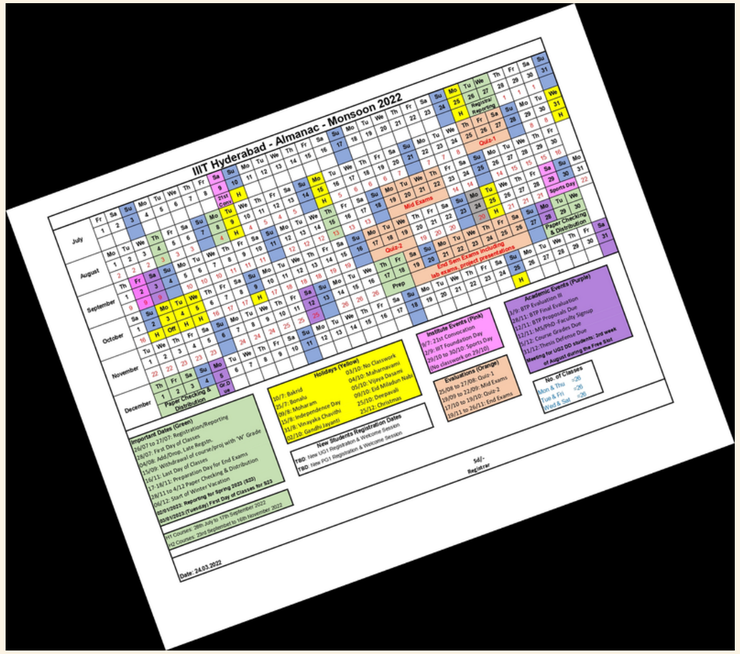


https://intranet.iiit.ac.in/offices

Pre-processing on datasets

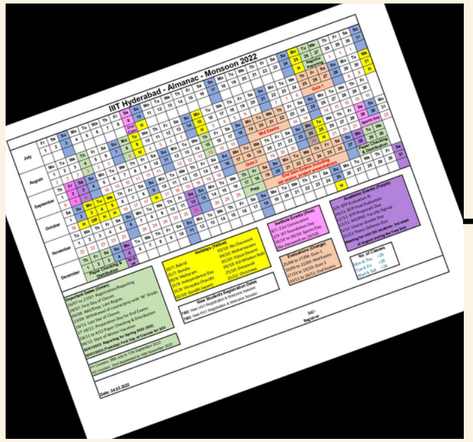
Online free tools are used to convert the almanacs pdf to PNG format.

To test the code for rotation invariance, the images are rotated using python functions.



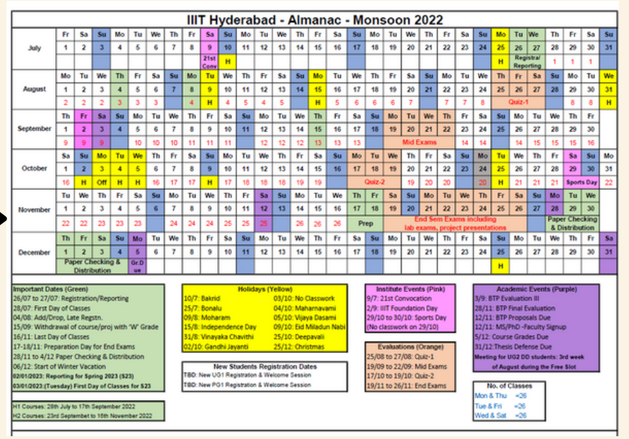
HOUGH TRANSFORM

For straining the image

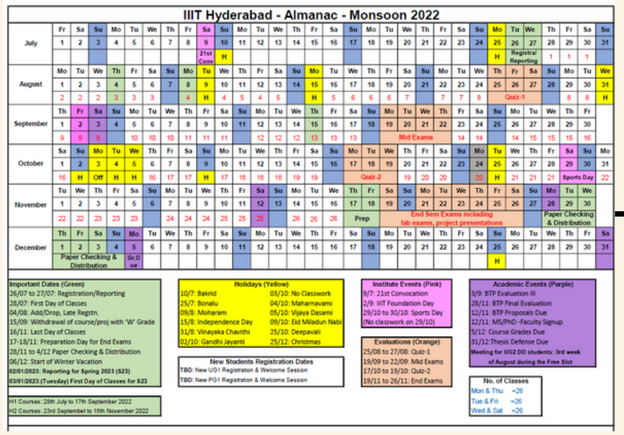


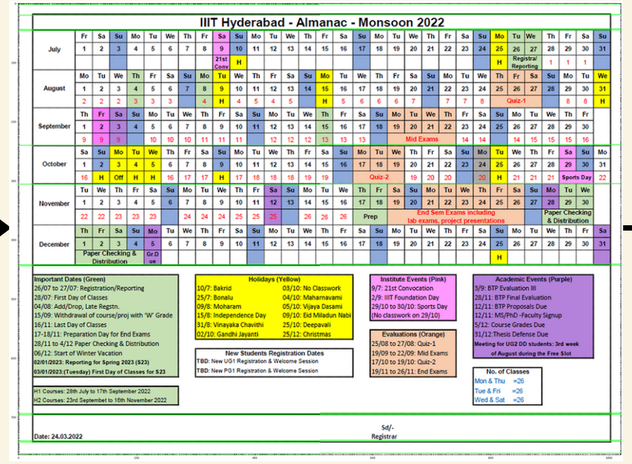
Hough Transform to get the angle of rotation

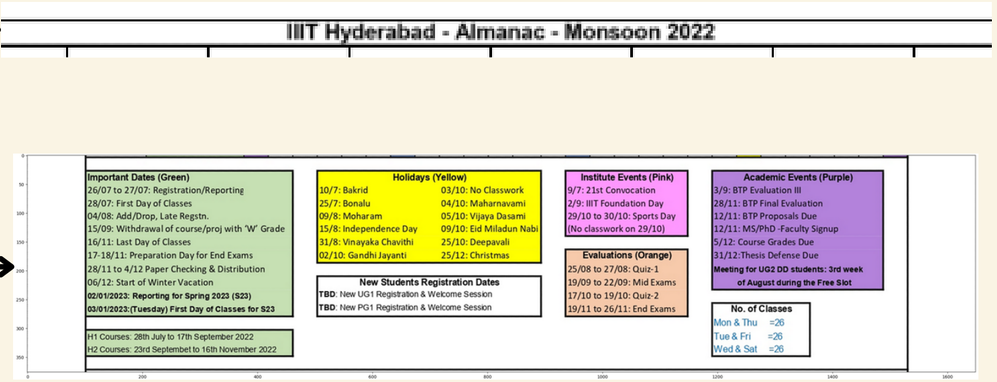
Rotation



Getting the Area of Interest





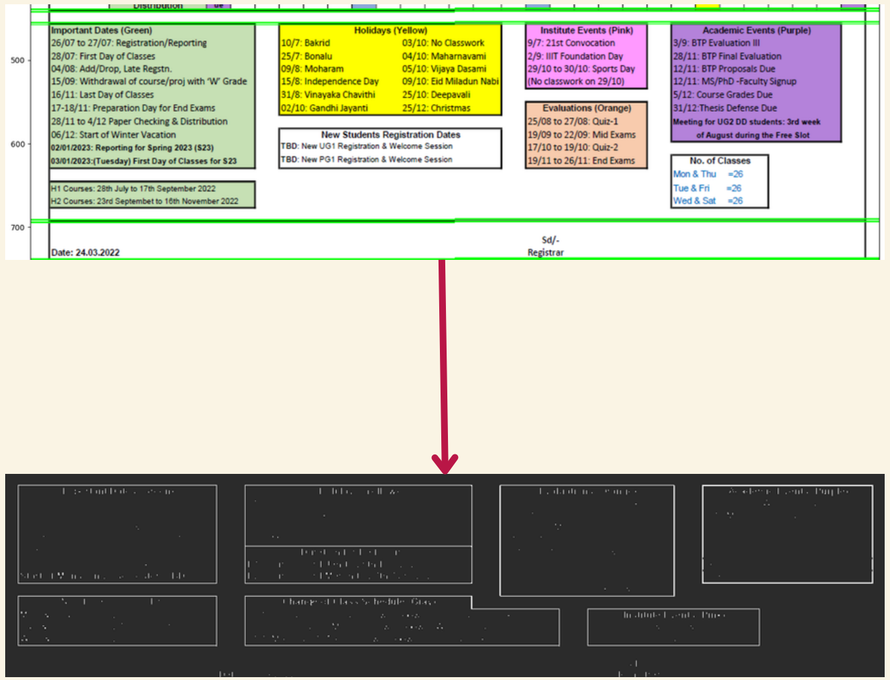


HARRIS CORNER DETECTOR

Harris Corner Detector

- After segregating the area of interest from the rest of the image. At first we threshold the image for better processing, after this erosion is used to get rid of texts from the box.

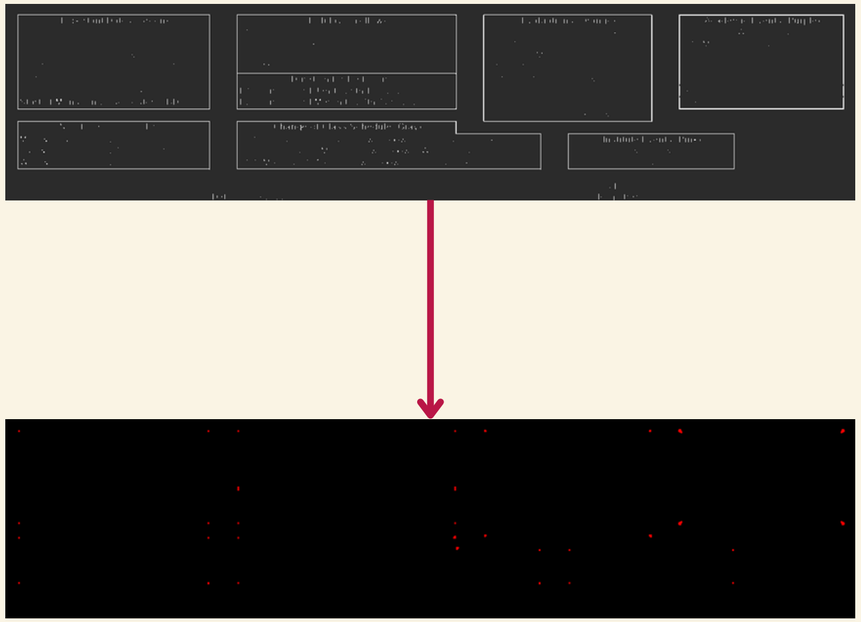
- Then Harris Corner Detector is used to detect the corners of the rectangles where the desired information is present.



Harris Corner Detector

- The Harris Corner detector gives Cluster of points for the corners so connected components is then used to get singular points.

- After detecting the corner points we use OCR between the points to gather the dates and the corresponding event from the Almanac.



Problem faced with Harris Corner Detector

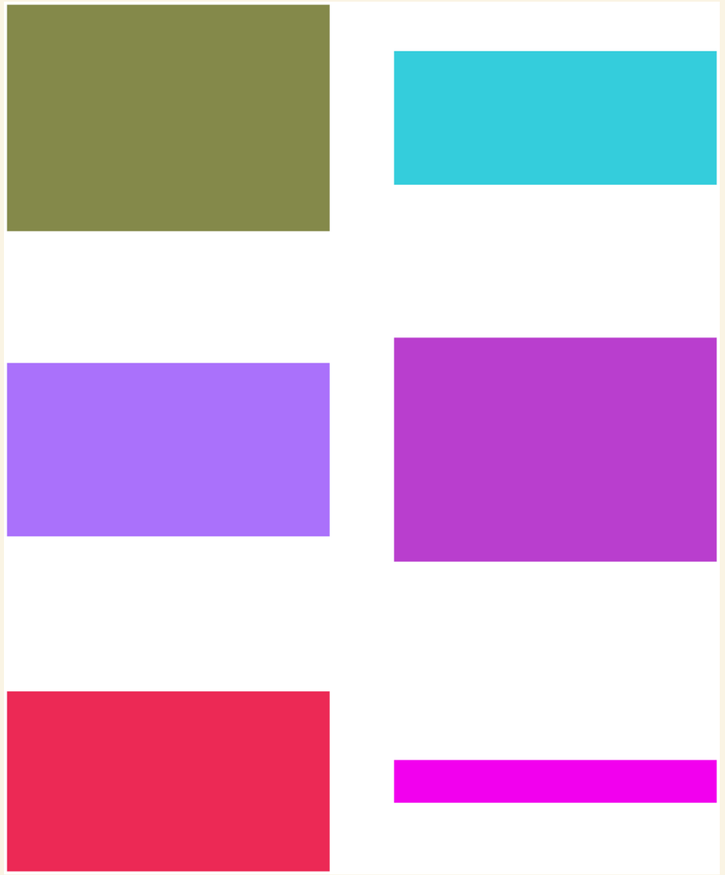
- Harris Corner detector gives a cluster of points for a single corner which in turn detects multiple boxes in place of one.

- Performing tasks such as min, max or CCL on the image is still not able to give a single aligned point per corner.

BOX DETECTOR

Getting different categories of events

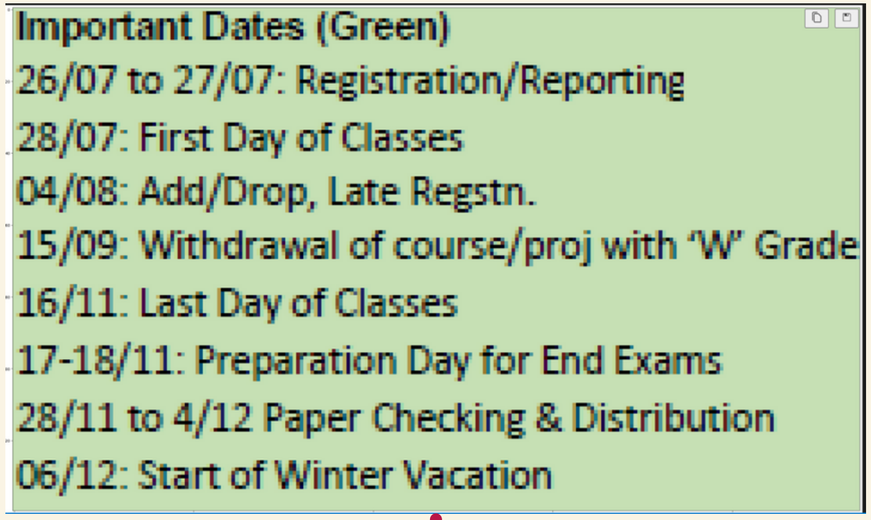
- We used thresholding and closing to get the shape boxes containing information about the dates and events.
- Connected components is then used to get the bounding boxes and the coordinates of the boxes are used for text recognition



OPTICAL CHARACTER RECOGNITION

Optical Character Recognition (OCR)

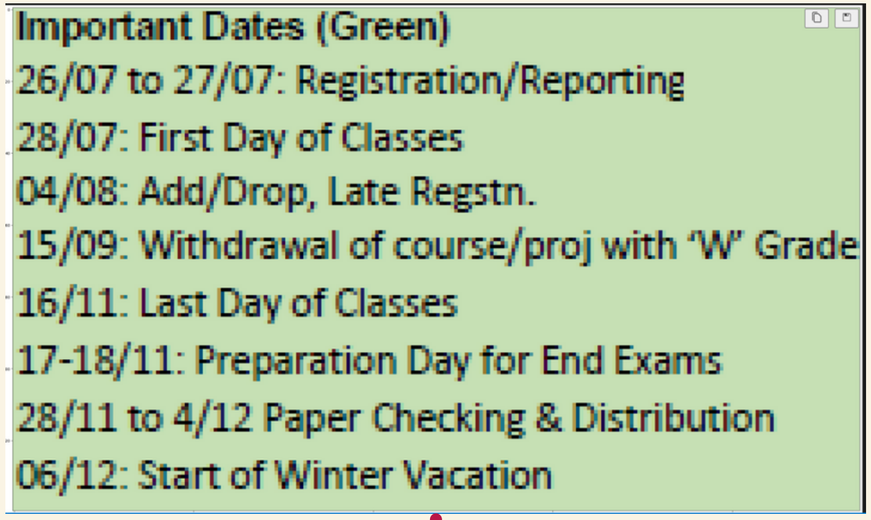
- This method is used to detect the text pertaining to various events in the almanac such as exam schedule, holidays, festivals, miscellaneous important dates, etc.
- With OCR, we can convert the image of the text to numbers (for the dates), the string against the date and accordingly assign in the online calendar.




```
Important Dates (Green)

26/07 to 27/07: Registration/Reporting

28/07: First Day of Classes

04/08: Add/Drop, Late Regstn.

15/03: Withdrawal of course/proj with "'W' Grade
16/1: Last Day of Classes

17-18/11: Preparation Day for End Exams

28/11 to 4/12 Paper Checking & Distribution
06/12: Start of Winter Vacation
```

OCR Contd.

- This method is used to detect the text pertaining to various events in the almanac such as exam schedule, holidays, festivals, miscellaneous important dates, etc.
- With OCR, we can convert the image of the text to numbers (for the dates), the string against the date and accordingly assign in the online calendar.




```
Important Dates (Green)

26/07 to 27/07: Registration/Reporting

28/07: First Day of Classes

04/08: Add/Drop, Late Regstn.

15/03: Withdrawal of course/proj with "'W' Grade
16/1: Last Day of Classes

17-18/11: Preparation Day for End Exams

28/11 to 4/12 Paper Checking & Distribution
06/12: Start of Winter Vacation
```

Limitations:

- If there are boxex of uneven shapes then is an additional noise in the box and the text recognition catches unwanted random letters
- If there is a different calender format the code will not be able to judge the new type and it will output a courrputed file.

Contributions

- Rishabh Agrawal: Implemented Hough Transform to align the image and get the interested region.
- Akshit Gureja: Converted the interpreted string to Ical format and debugged the string interpretation code
- Atharva Gogate: Used OCR to get the string from the calendar and interpret it to be converted into dates and events.
- Puru Ojha: Written the Harris Corner and the thresholding method to get the individual boxes of event categories from the area obtained after Hough Transform.

Thank you

end