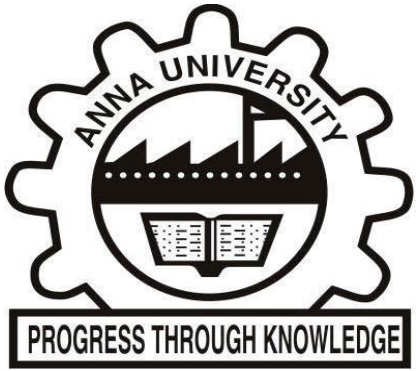
LIBRARY MANAGEMENT SYSTEM USING IMAGE PROCESSING



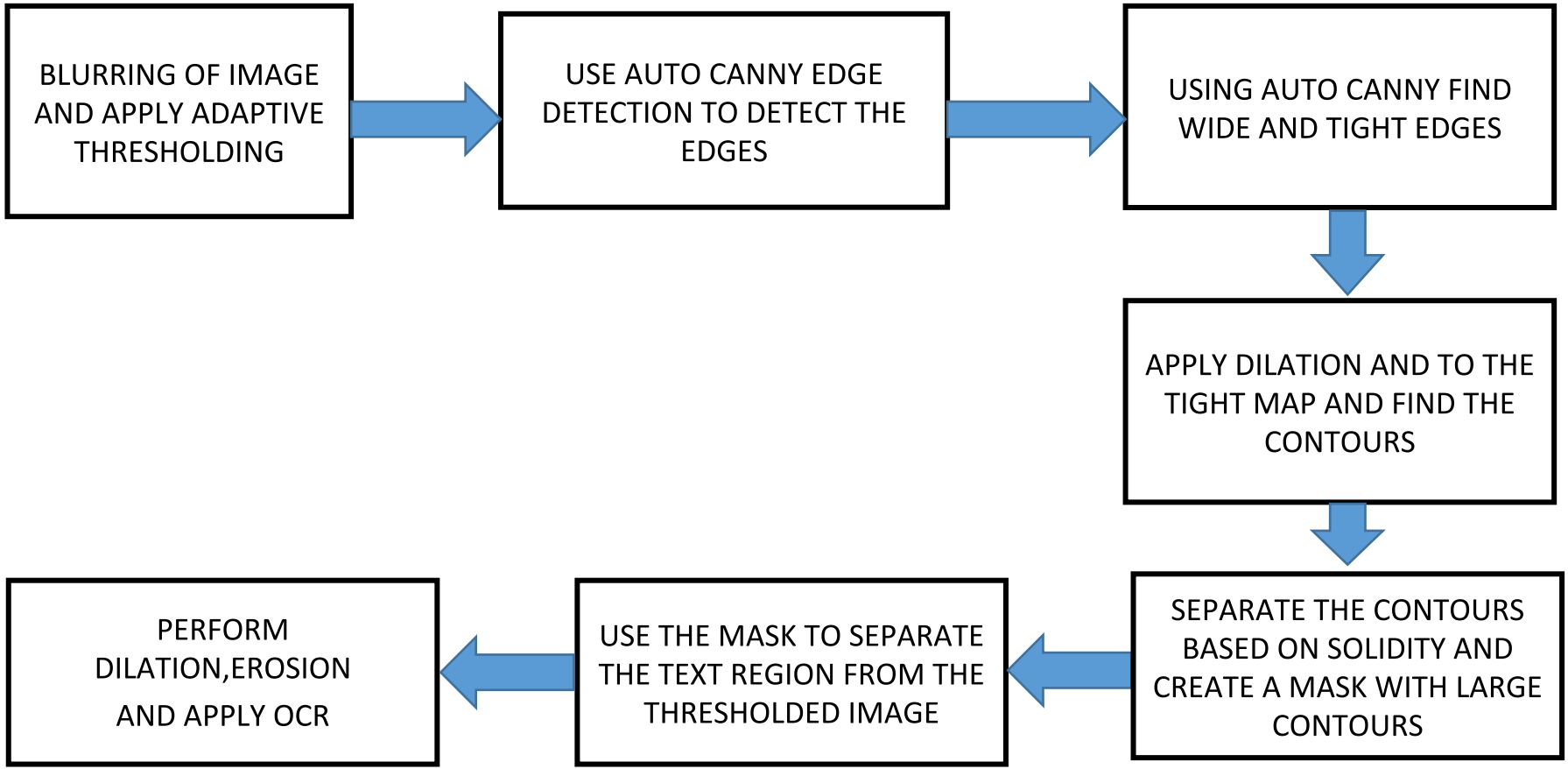
Submitted by

Sai Mukkundan R - 2014504036
Melvin A - 2014504030
Shoban Chander E -2014504040

Under the guidance of

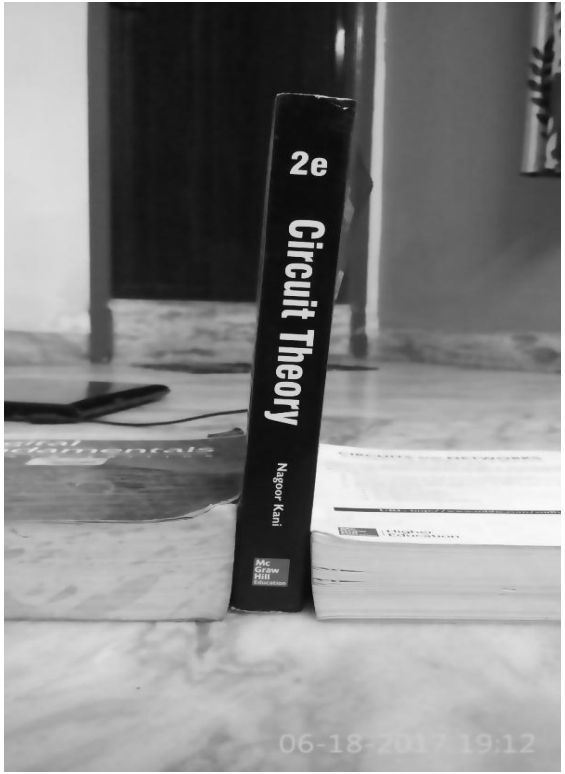
Dr.V.Sathiesh Kumar
Assistant Professor, Department of Electronics Engineering

BLOCK DIAGRAM-METHOD A

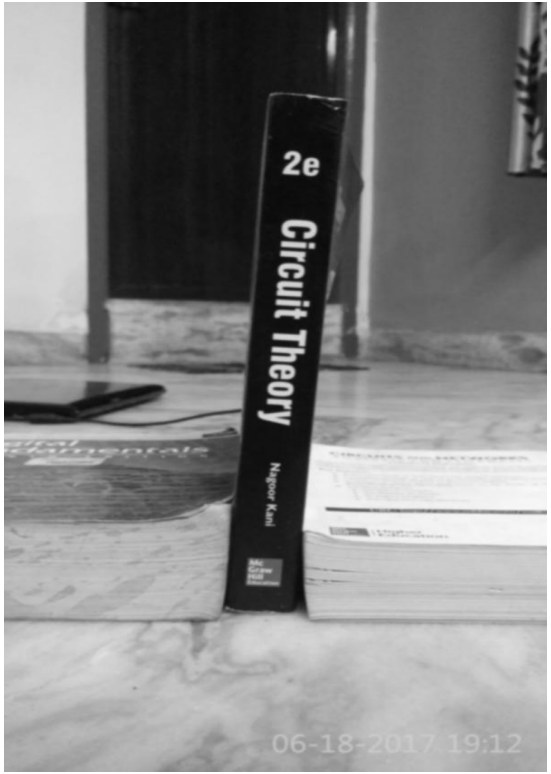


ORIGINAL IMAGE

BILATERAL BLURRED IMAGE



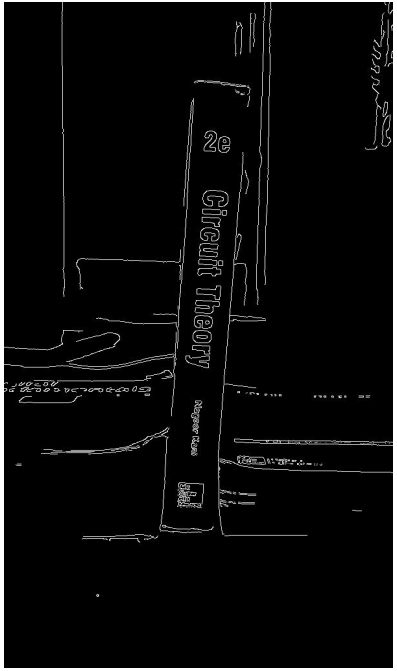
GAUSSIAN BLURRED IMAGE



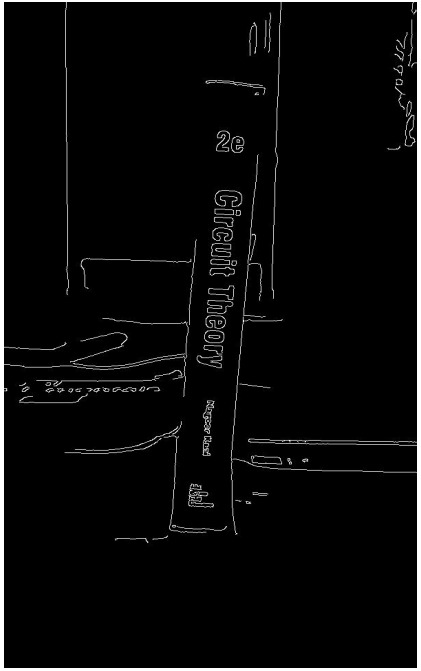
ADAPTIVE THRESHOLD OF BILATERAL BLURRED IMAGE

ADAPTIVE THRESHOLD OF GAUSSIAN BLURRED IMAGE

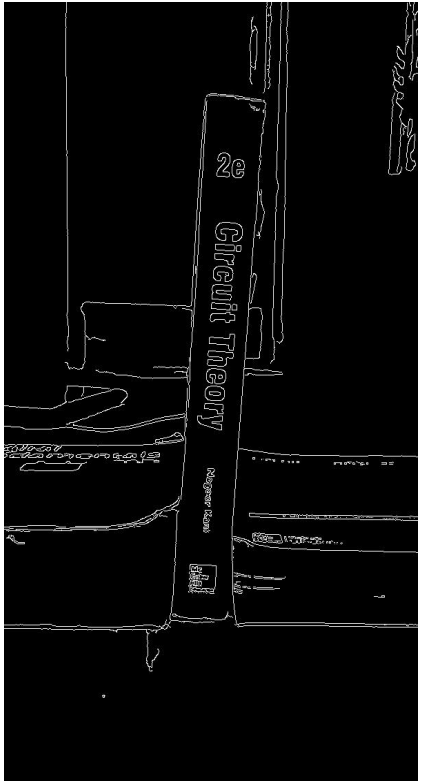
AUTO CANNY EDGE DETECTED IMAGE (BILATERAL)



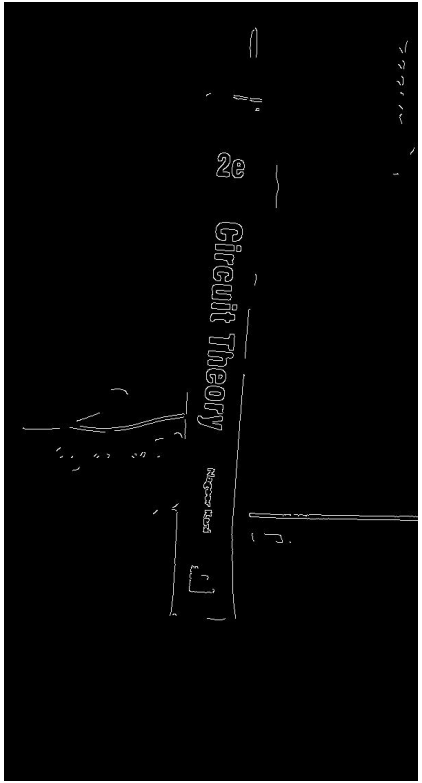
AUTO CANNY EDGE DETECTED IMAGE (GAUSSIAN)



WIDE CANNY

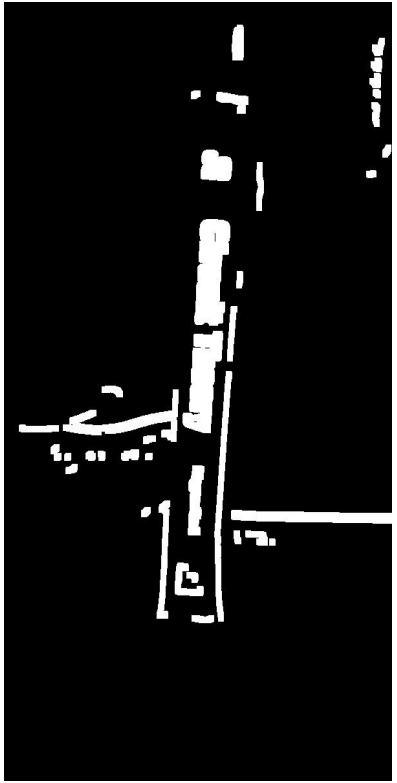


TIGHT CANNY

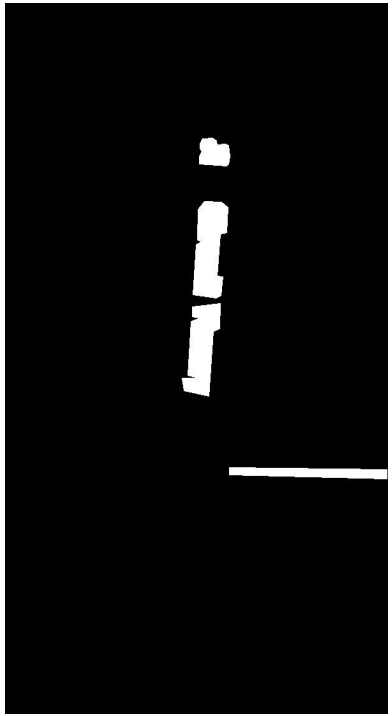


CONTOURS

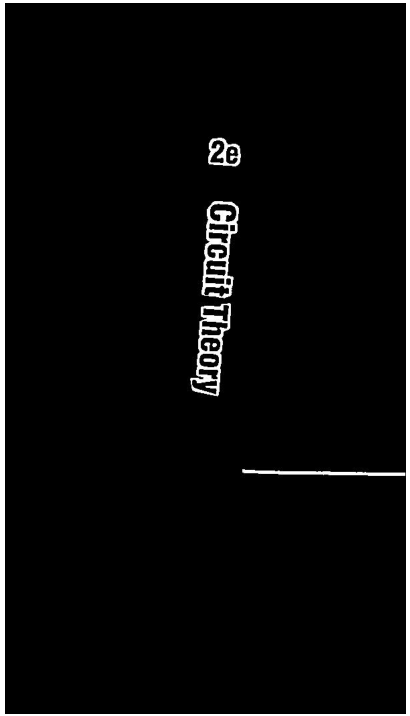
DILATED IMAGE



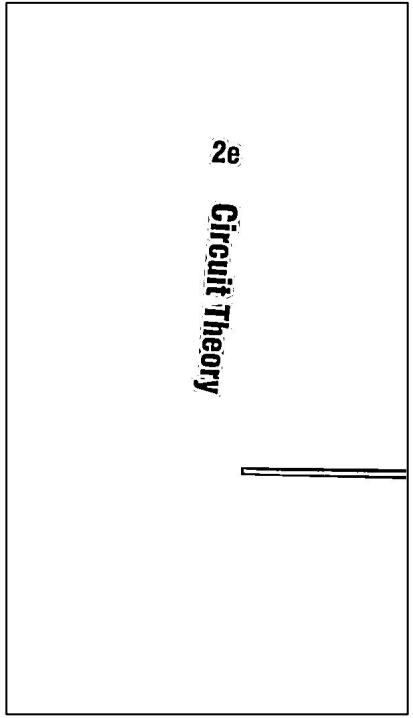
CONTOUR SEPARATED AND SOLIDITY APPLIED



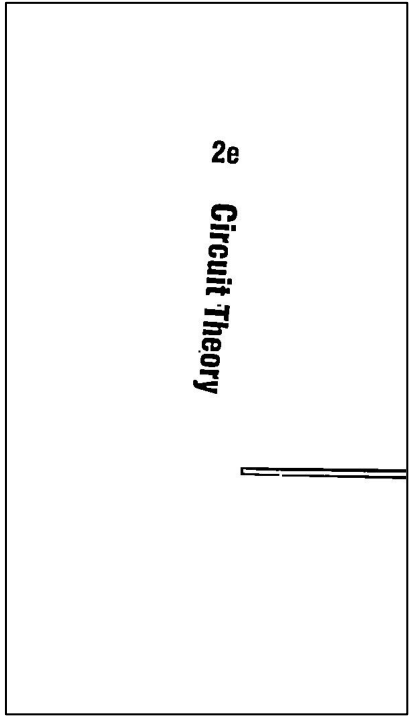
BITWISE AND WITH PREVIOUS IMAGE AS MASK



INVERTED IMAGE



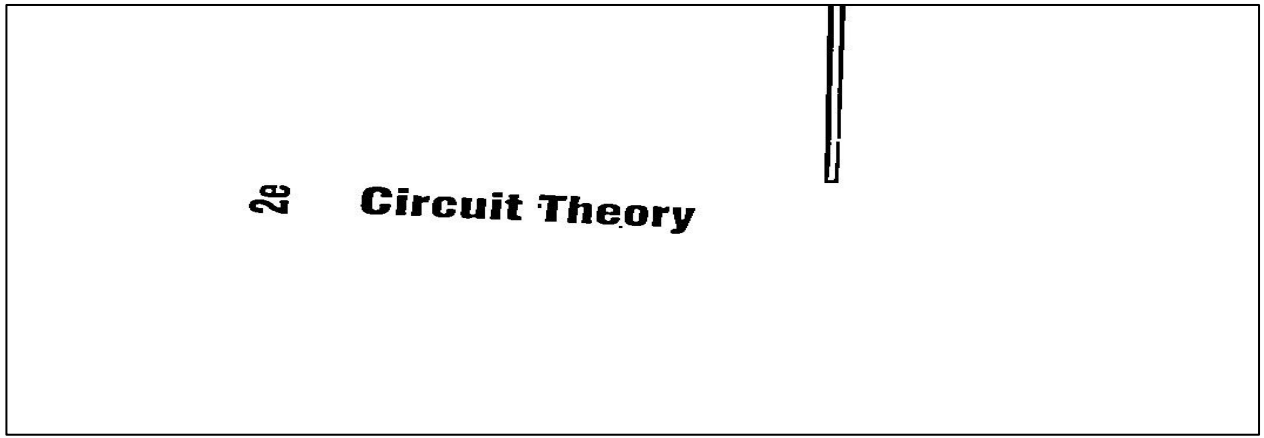
DILATED AND ERODED IMAGE



DRAWBACKS

- OCR doesn’t support different kind of fonts.
- As the books contain varieties of fonts, every font can’t be loaded in the OCR dataset.
- Also keeping many books aside leads to detection of names from many books at the same time. Hence this method fails.

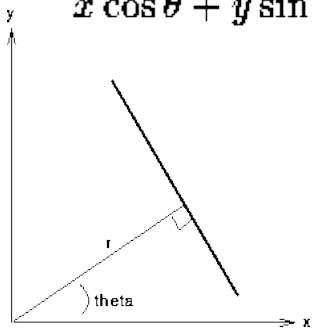
FINAL IMAGE ROTATED AND SENT TO OCR



HOUGH TRANSFORM BASED EDGE SEPARATION - METHOD 2

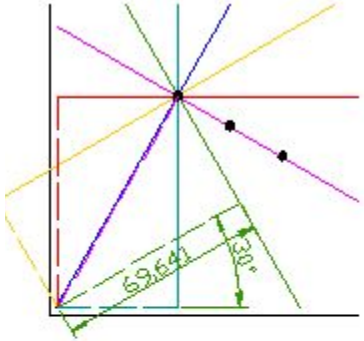
- Hough Transform is used to detect the continuity of points in the edge regions.
- It converts edge points into lines.If they obey the same parametric line equation model.

$$x\cos\theta + y\sin\theta = r$$

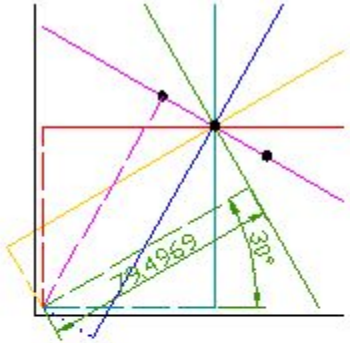


- It uses a voting system in parametric space (r,theta),to find the line which consist of maximum points lying on them.

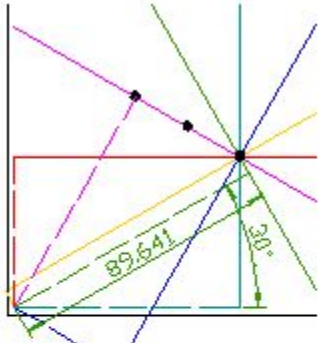
Theta \ Rho	2-2.5	2.5-3	3-3.5	3.5-4	4-4.5	4.5-5	5-5.5	5.5-6	6-6.5	6.5-7	7-7.5	7.5-8
0-5												
5-10												
10-15												
15-20		1	2		2	1	1					
20-25												
25-30					8							
30-35												
35-40												
40-45		1		1	1	1		1	1		1	



Angle	Dist.
0	40
30	69.6
60	81.2
90	70
120	40.6
150	0.4

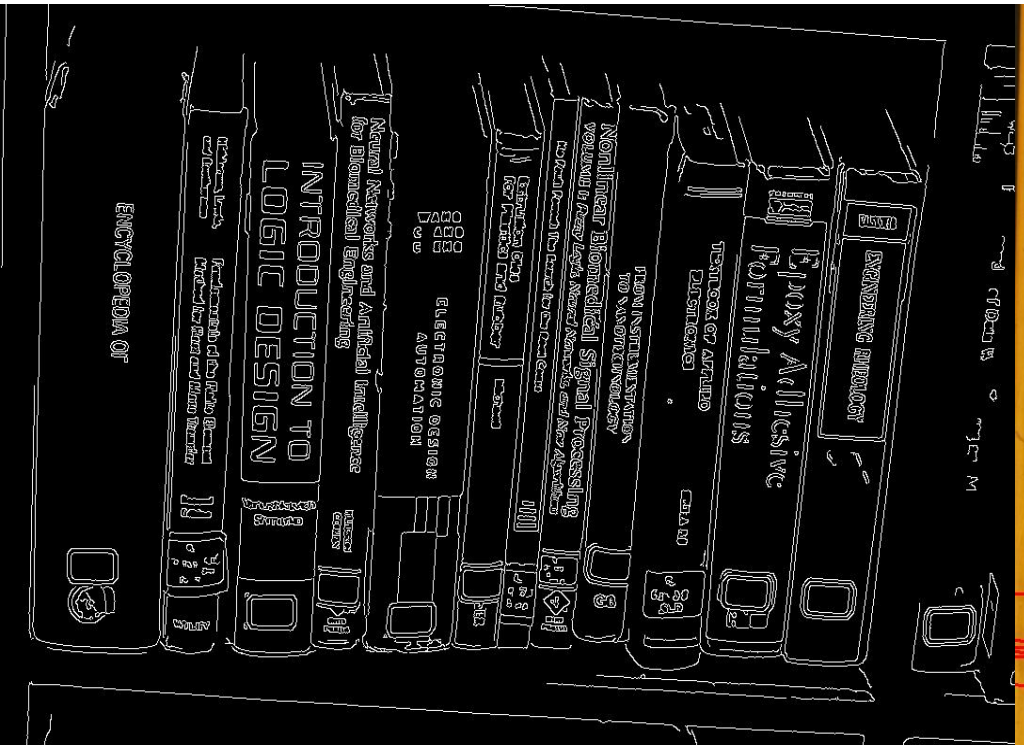


Angle	Dist.
0	57.1
30	79.5
60	80.5
90	60
120	23.4
150	-19.5



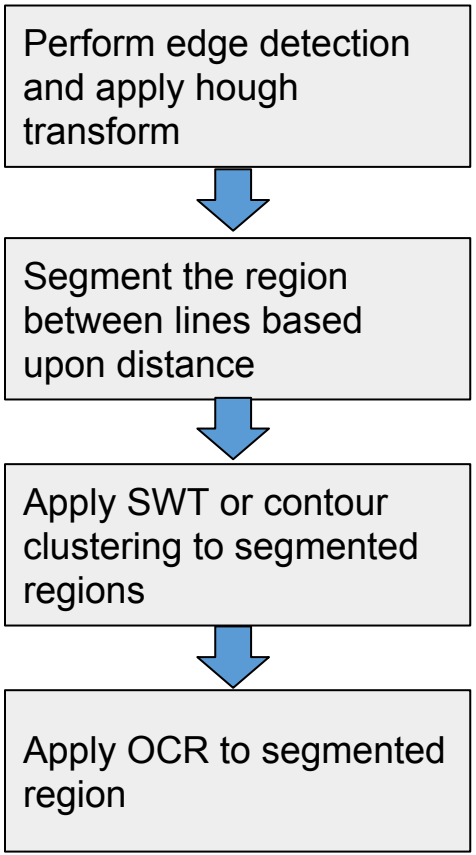
Angle	Dist.
0	74.6
30	89.6
60	80.6
90	50
120	6.0
150	-39.6

Edge detection using canny edge detector



After applying hough transform and drawing the lines between the books

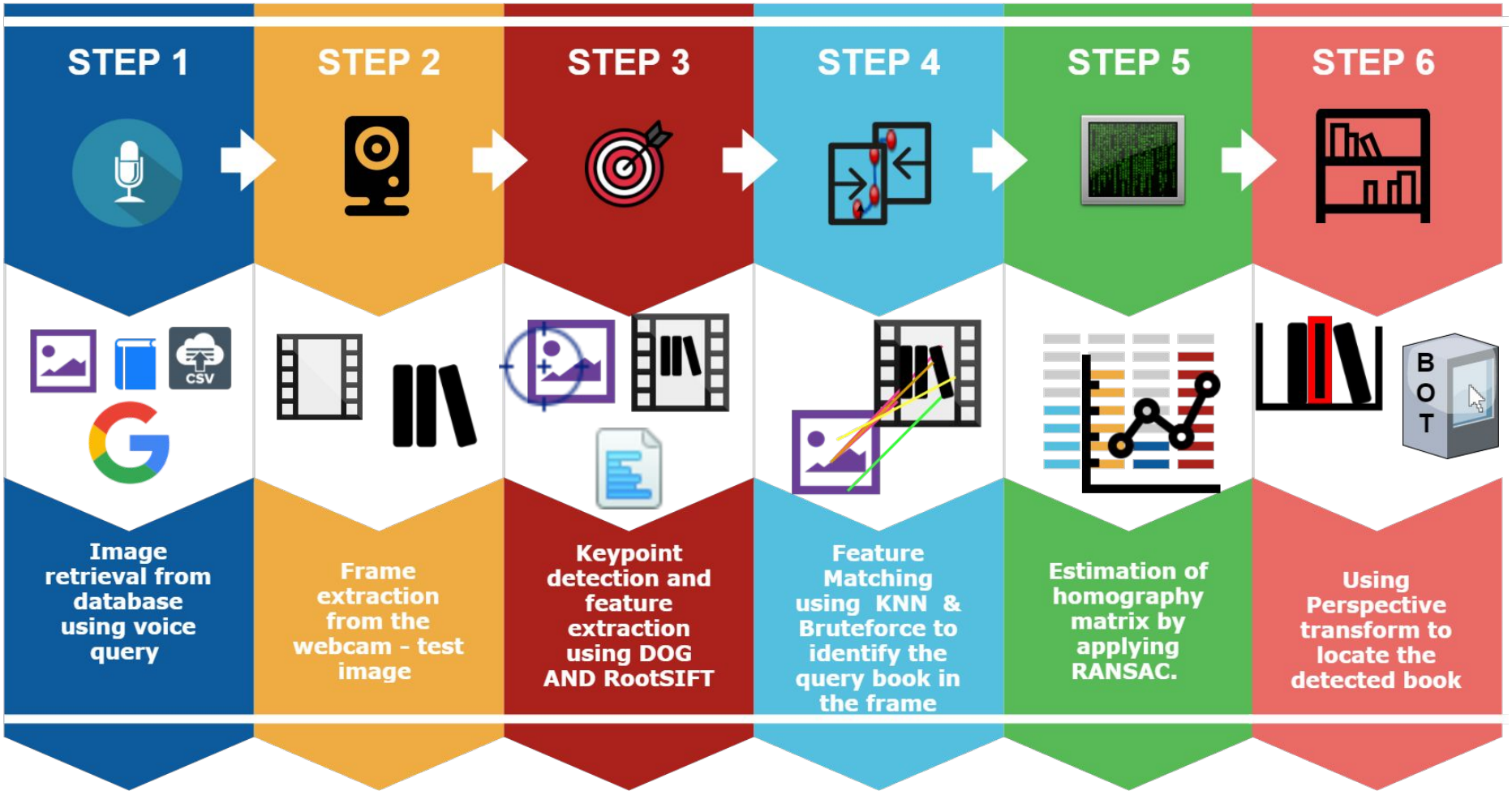
ALGORITHM



DRAWBACKS

- It is unreliable as the lines detected are subject to change due to uneven illumination
- You can’t decide an appropriate threshold for distance between adjacent lines since the book width varies widely
- Stroke Width Transform fails if applied on the segmented window whose viewpoint is not straight.
- Even applying contour based segmentation between the adjacent lines won’t yield proper output
- You can’t rely on ocr since font size is and shape are different for various books

Workflow


STEP 1
STEP 2
STEP 3
STEP 4
STEP 5
STEP 6
CSV
B
O
T
Image retrieval from database using voice query
Frame extraction from the webcam - test image
Keypoint detection and feature extraction using DOG AND RootSIFT
Feature Matching using KNN & Bruteforce to identify the query book in the frame
Estimation of homography matrix by applying RANSAC.
Using Perspective transform to locate the detected book

Voice Recognition using Google Api

- Collecting details of book name ,author name and along with its edition.
- Voice recognition with ambient noise control,speech to text conversion
- Speech is converted to text using the Google Speech Recognition API.
- The speech input is sent as a http post request to the server.POST Url: “https://www.google.com/speech-api/fullduplex/v1/up”.
- The text is sent back through get request to the user. GET Url:

  “https://www.google.com/speech-api/fullduplex/v1/down”

.

Advantages:

- No pre-trained voice data is required.
- Pronunciation ,autocorrect features with better recognition performance than other voice recogniser (CMU Sphinx)
- Provides security -POST request

Disadvantages:

- Internet is required for proper operation.

Voice Query with Database

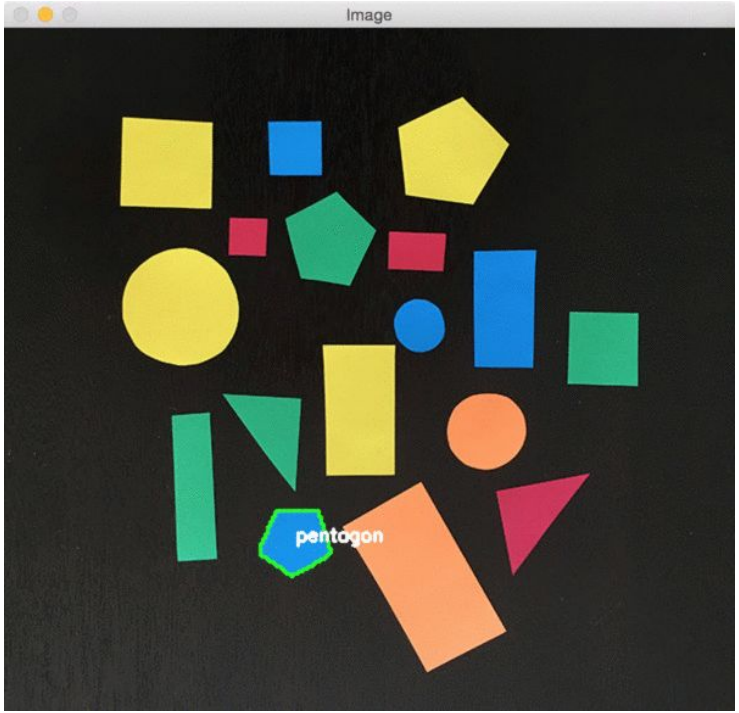
- Reading data from a csv file ,converting into python dictionary with image path as key.
- Retrieving the image path of the query book based on the voice data.
- Books are cropped such that their spine fits the entire frame

```
C#,XAVIER,,1.jpg
ENGINEERING PHYSICS,MUKHERJI,,2.jpg
ENGINEERING ELECTROMAGNETICS,HAYT,FIFTH,3.jpg
A TEXTBOOK ON C#,MURUGESAN,,4.jpg
MATHEMATICAL ANALYSIS,APOSTO,,5.jpg
ENGINEERING MATHEMATICS-II,VEERARAJAN,,6.jpg
OPERATING SYSTEMS,NIIT,,7.jpg
AIRCRAFT MATERIALS & PROCESSES,,FIFTH,8.jpg
ELECTRICAL ENGINEERING,HAMBLEY,SECOND,9.jpg
AN EMBEDDED SOFTWARE PRIMER,SIMON,,10.jpg
INTERNAL COMBUSTION ENGINE FUNDAMENTALS,HEYWOOD,,11.jpg
THE AUTOMOTIVE CHAS,GENTA,,12.jpg
DEFORMATION AND FRACTURE MECHANICS,HERTZBERG,FOURTH,13.jpg
CERAMIC PRECURSOR TECHNOLOGY,NARULA,,14.jpg
WAVELETS IN ELECTROMAGNETICS AND DEVICE,PAN,,15.jpg
HOW TO MAKE INJECTION MOLDS,MOHREN,,16.jpg
SUPERCONDUCTIVITY,BUCKEL,,17.jpg
PROBABILITY MODELS IN OPERATIONS RESEARCH,CASSADY
NACHLAS,,18.jpg
FUEL CELL FUNDAMENTALS,COLELLA,,19.jpg
TYRE AND VEHICLE DYNAMICS,PACEJKA,,20.jpg
COMPLEX ANALYSIS WITH MATHEMATICA,SHAW,,21.jpg
POLYMER MATRIX COMPOSITES AND TECHNOLOGY,WANG,,22.jpg
BRITANNICA,,,23jpg
VOICE OVER PACKET NETWORKS,WRIGHT,,24.jpg
ANALYTICAL INSTRUMENTATION,CURRELL,,25.jpg
INTELLECTUAL PROPERTY RIGHTS,GANGULI,,26.jpg
IMAGE PROCESSING FOR COMPUTER GRAPHICS,VELHO,,27.jpg
MECHANICAL PROPERTIES OF POLYMERS,MICHLER,,28.jpg
LINUX COMPANION FOR SYSTEM ADMINISTRATORS,HEIN,,29.jpg
FUEL CELL TECHNOLOGY,HOOGERS,,30.jpg
```

CSV file with fields for book name,author name and edition

Keypoint Identification and Feature Extraction

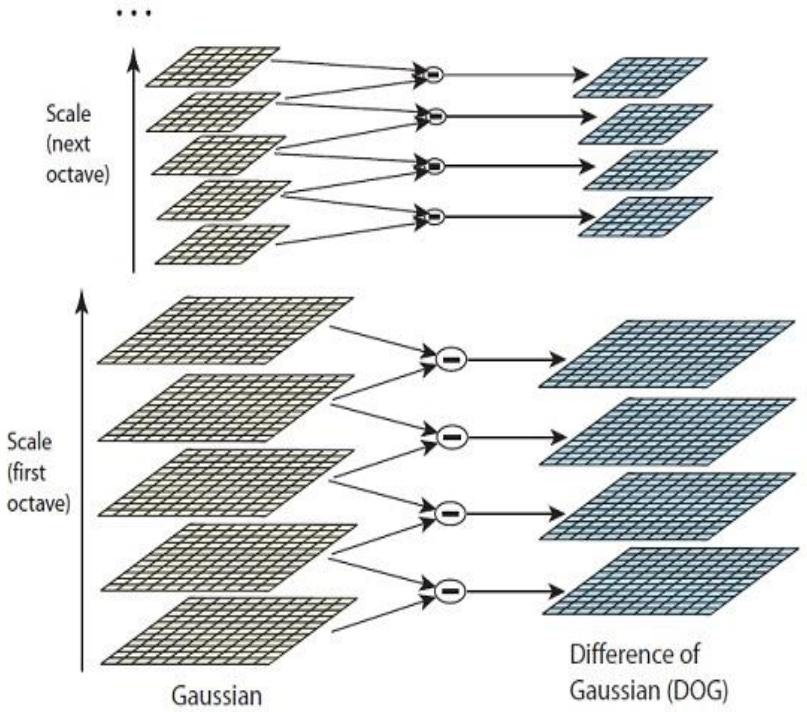
- Feature extraction is nothing but the process of quantifying an image and characterizing it based on texture,colour,shape or combination of all three.The output of feature extraction process is a feature vector representing an entire image (image descriptor)or local region in an image(feature descriptor).
- Keypoint detection involves locating interest points around which feature descriptors can be constructed.This interest point may be sharp gradient changes around corners and edges.
- Good keypoint detector should produce repeatable and localised keypoints for precision matching.



DOG Keypoint Detection

- DoG keypoint detector which works perfectly in detecting book spines in different orientation and scales.

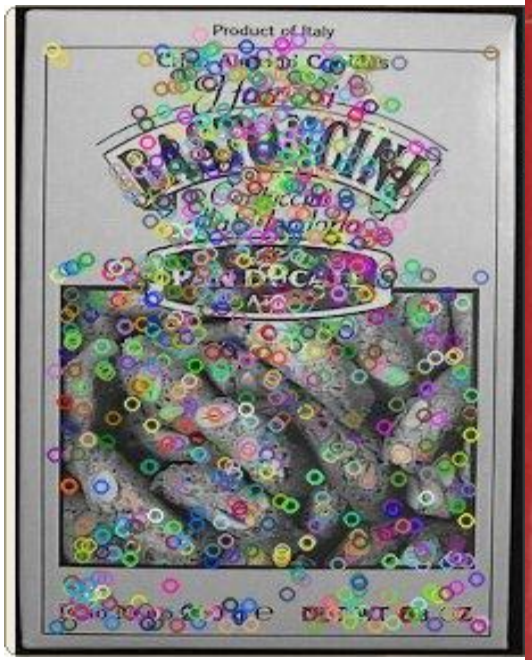
- DoG makes use of scale space where an octave(set of column images with same size) consisting of different versions of same image but gaussian blurred to different extent in each stage of that octave



Keypoint Detection and Feature extraction in book spine

- In our case we use feature extraction to identify interesting regions of specific book spine and extract features from them and later use those features to compare the books present in the library stack.

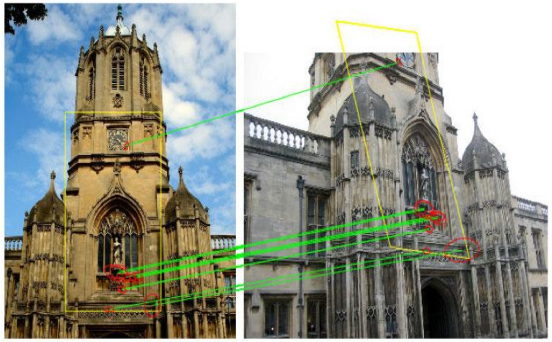
- There are several keypoint detectors like MSER,FAST,HESSIAN,HARRIS,STAR,DENSE but DOG (SIFT) keypoint detector found to perform well for this problem with minimum false positives.



Keypoint detection in front cover and spine

SIFT and RootSIFT Feature Description

- SIFT divides a region around a keypoint into 16X16 cells which are regrouped into 16 (4X4 cells).Now in this subregion,HOG is computed for orientation of zero degree to 180 degree divided into 8 bins.Each bins contain gaussian weighted gradient magnitude.Now each 4X4 cells contain 8 dimension vector which are clustered together into 128 dimension vector which forms our feature descriptor.
- ROOTSIFT which is a simple extension to SIFT feature descriptor.We modify the SIFT in such a way that when the feature vectors are measured with same old euclidean distance they appear to be measured with hellinger kernel.
- ROOTSIFT outperforms SIFT in terms of number of accurate keypoint matches.

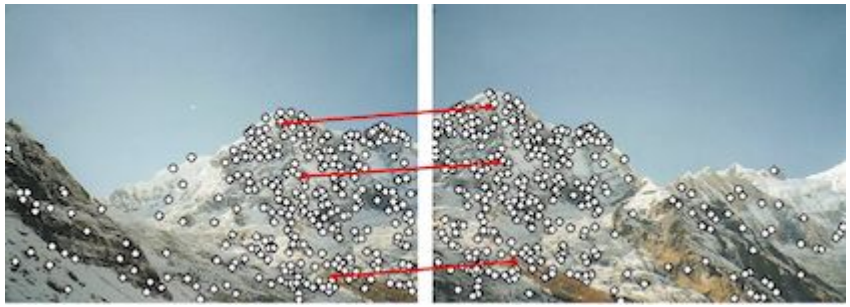


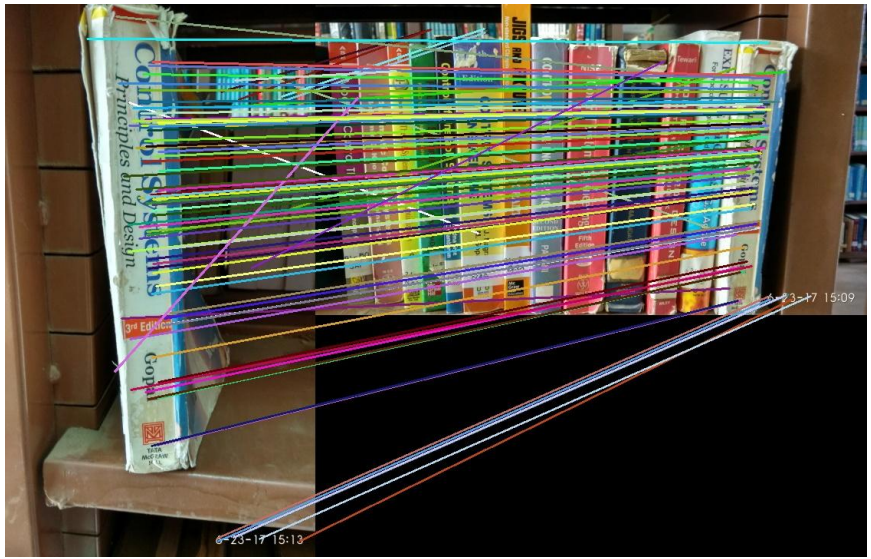
(a) SIFT (L2 distance): 10 matches

(b) RootSIFT: 26 matches

Feature Matching

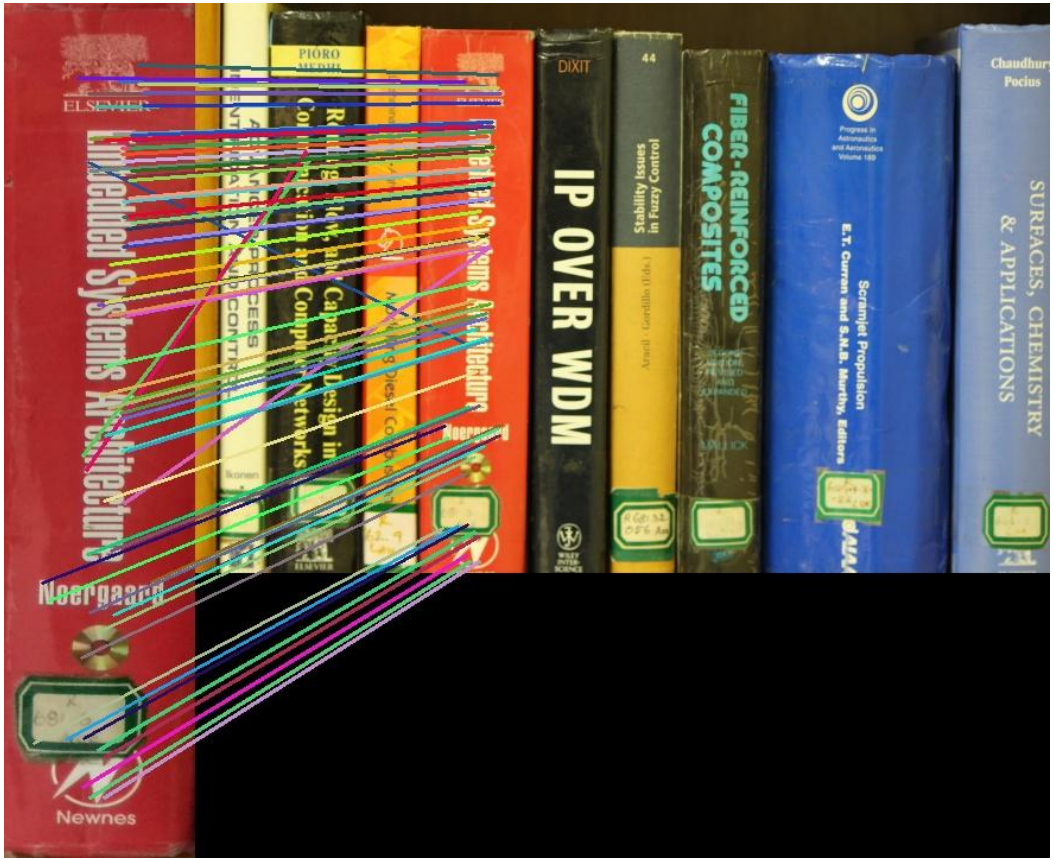
- Feature matching is done using any of the distance algorithm which hunts for similarity between points of two different images (maximum likelihood).
- The accuracy in finding matched keypoints between two images lies in the distance computation algorithm and the way it has been used.
- Here K-nearest neighbour algorithm along with Lowe's ratio test is used to find reliable and accurate matches to the keypoints in query image.





Result of applying Lowe’s ratio test and Knn algorithm

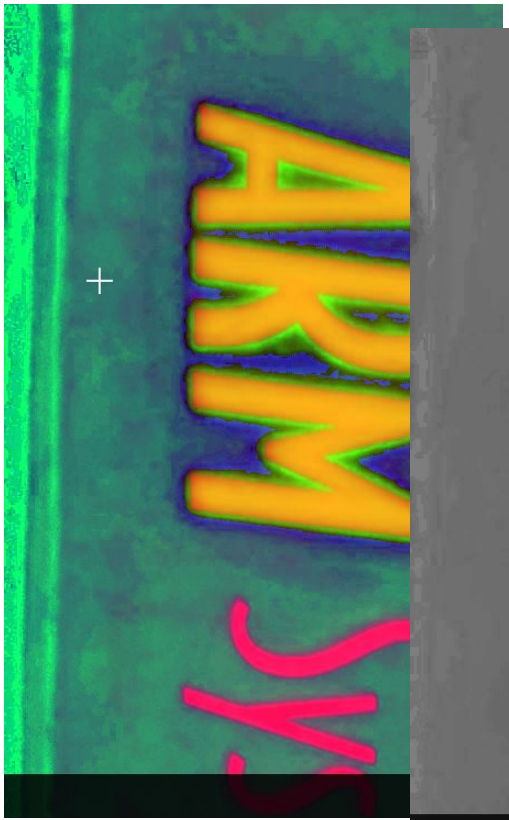
- The K nearest neighbour (Knn) method is used to identify at least two nearest neighbours(in this case)closer to the matched keypoint.
- Lowe’s ratio is based upon the fact that the neighbours of matched keypoints are always farther.By making lowe ratio to be 0.7,we could achieve accuracy by eliminating outliers.



Color feature descriptor

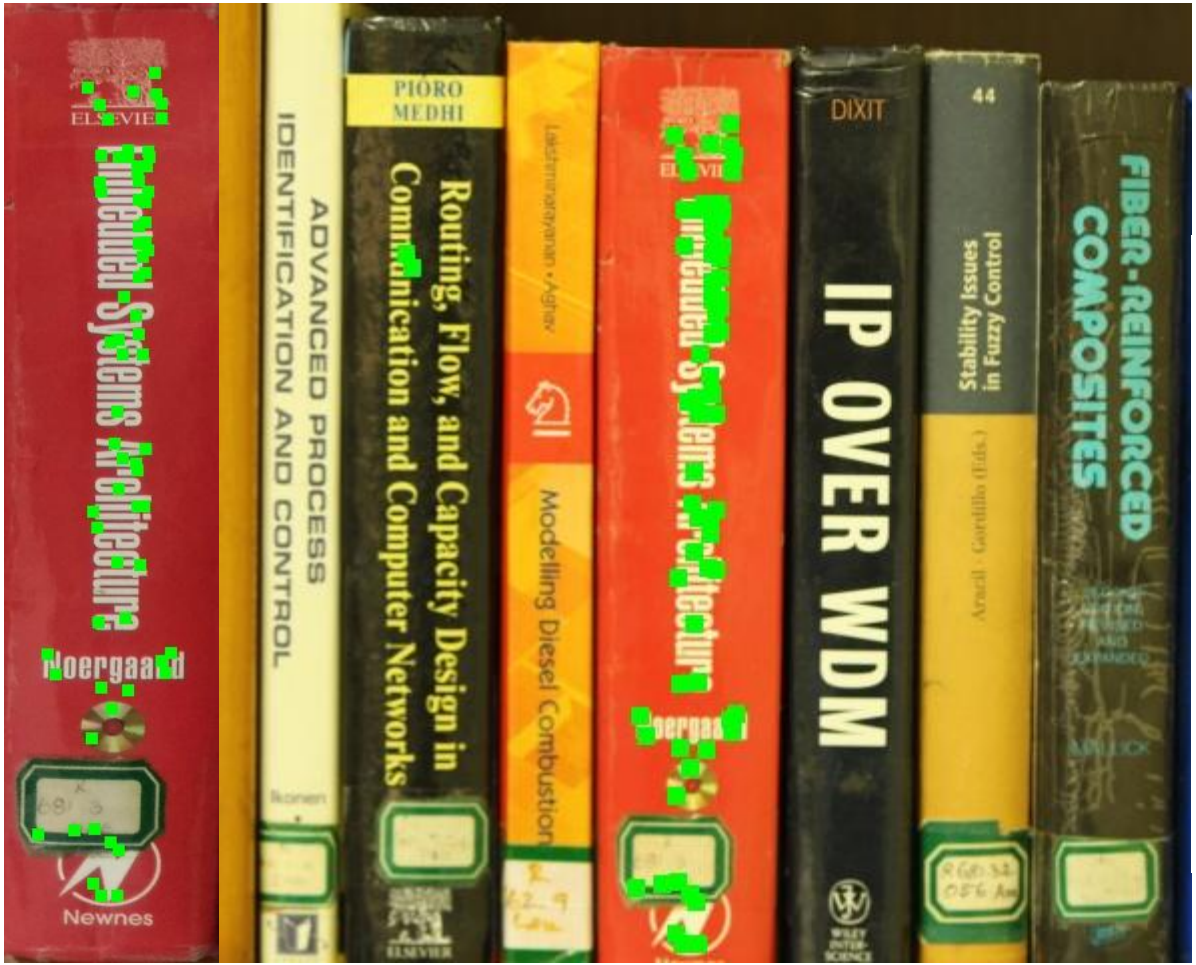
- To further prune the outliers from our matches,we reside to color description of neighbourhood pixels of our matched keypoints.
- Compute a 9X9 rectangular window around the keypoint and convert both images to hsv space.
- Split HSV plane ,calculate histogram around these widows in Hue plane for both the images.
- Append all the vectors and compute euclidean distance between them.
- Threshold the matches based on the color similarity and send those matches to homography

HSV Image

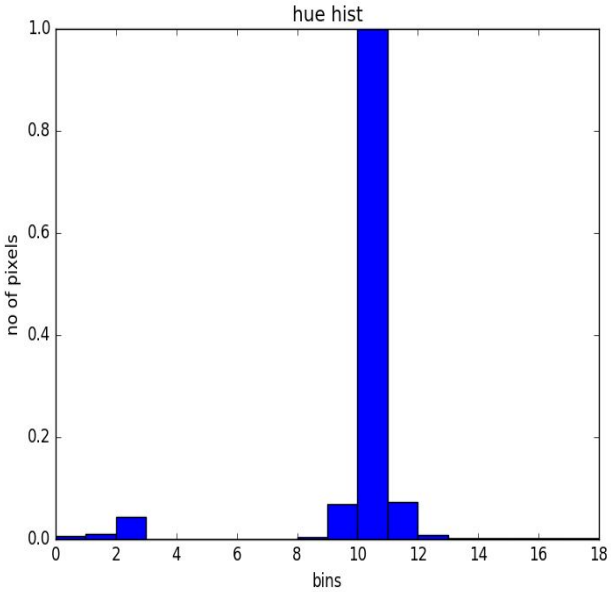


Hue Only Image

Window of 9X9 size around the matched keypoints in both query and test image



Hue Histogram around these keypoints

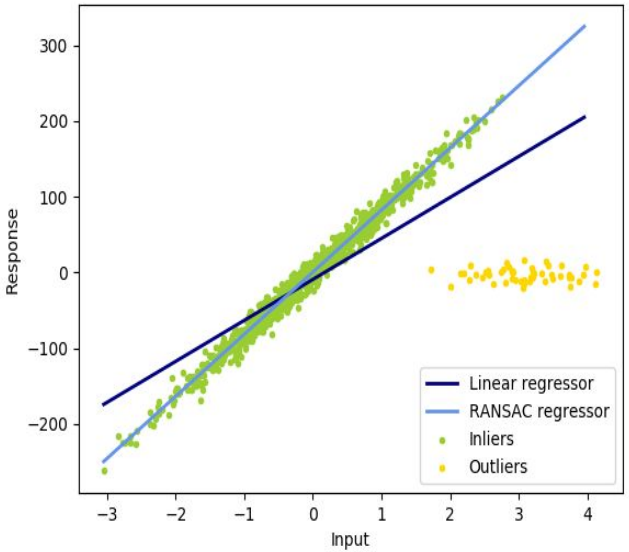


Homography Estimation

- A Homography is a transformation (a 3×3 matrix) that maps the points in one image to the corresponding points in the other image.

RANSAC

- Random Sample Consensus assumes that the training data consists of inliers that can be explained with the model and outliers that don't fit the model at all.
- In order to reject outliers during training, RANSAC uses a small set of samples to train a model rather than using all of the data and then enlarges set with other appropriate samples
- Then it determines the samples that are within an error tolerance of the generated model. These samples are considered as agreed with the generated model and called as consensus set.
- Data samples in the consensus behaves as inliers and the rest as outliers and trains the model with this new inliers.
- It randomly a selects a subset of points and repeats it until the estimation error becomes minimum,now it gives the final model parameters that gives the best fit for the points.



Ransac along with Homography

- Applying Ransac in homography matrix computation to eliminate the outliers.Ransac works with an estimation model and cost function just like linear regression.Here the estimation model is the homography matrix.
- To compute this accurately a threshold for allowing amount of outliers has to be specified in before.
- **Cost function:**

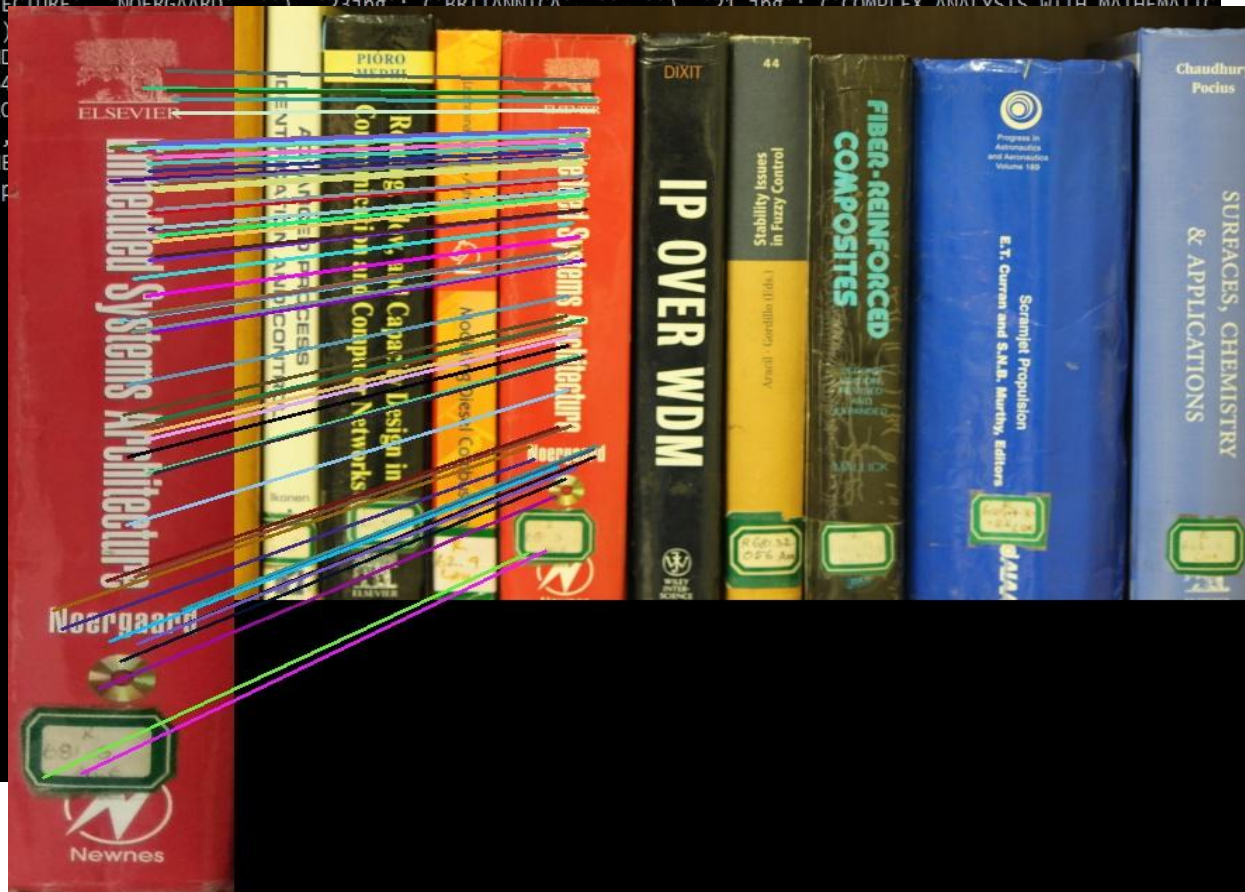
$$\left\| \mathtt{dstPoints}_i - \mathtt{convertPointsHomogeneous}\left(H * \mathtt{srcPoints}_i\right) \right\| > \mathtt{ransacReprojThreshold}$$

If this is true the point i is rejected as a outlier ,for homography computation maximum of four matching points is enough,for accurate matching,model with more inliers is taken

Result of Ransac based homography computation and matching

```
DMA', 'IPATOV', ''), '81.jpg': ('EMBEDDED SYSTEMS ARCHITECTURE', 'NOERGAARD', ''), '23jpg': ('BRITANNICA', '', ''), '21.jpg': ('COMPLEX ANALYSIS WITH MATHEMATIC
A', 'SHAW', ''), '57.jpg': ('GAS TURBINES', 'SOARES', ''
 'WILLIAMS', ''), '68.jpg': ('AIRCRAFT AEROELASTICITY AND
, ''), '48.jpg': ('ENGINE COMBUSTION', 'ROGERS', ''), '64
ECHANICS', 'HERTZBERG', 'FOURTH'), '72.jpg': ('SIGNAL PRO
'83.jpg': ('FIBER-REINFORCED COMPOSITES', 'MALLICK', ''),
CTURE DESIGN', 'MALEN', ''), '9.jpg': ('ELECTRICAL ENGINE
'), '17.jpg': ('SUPERCONDUCTIVITY', 'BUCKEL', ''), '69.jp
KA', '')}
descriptor shape:    (605, 128)
summed shape: (605, 1)
final shape of descp: (605, 128)
descriptor shape:    (2801, 128)
summed shape: (2801, 1)
final shape of descp: (2801, 128)
BOOK is not present in the frame - 7/11
descriptor shape:    (2881, 128)
summed shape: (2881, 1)
final shape of descp: (2881, 128)

# of keypoints from first image: 605
# of keypoints from second image: 2881
# of matched keypoints: 82
score:
0.841463414634
false positives 13
accurate matches 69
descriptor shape:    (4051, 128)
summed shape: (4051, 1)
final shape of descp: (4051, 128)
```



Ransac outlier elimination threshold 4.

Perspective Transform

- Perspective transform helps to find correspondence points between two images.
- We perform Perspective transform which calculates the amount of rotation,translation needed for identifying the corner coordinates in the test image.
- With the help of homography matrix,it computes the coordinates of the book spine.

$$H = \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{bmatrix} \qquad \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = H \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix}$$

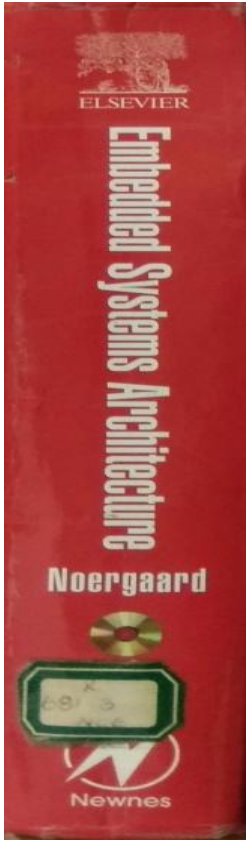
Homography Matrix obtained from matched keypoints

Unknown book corner coordinates

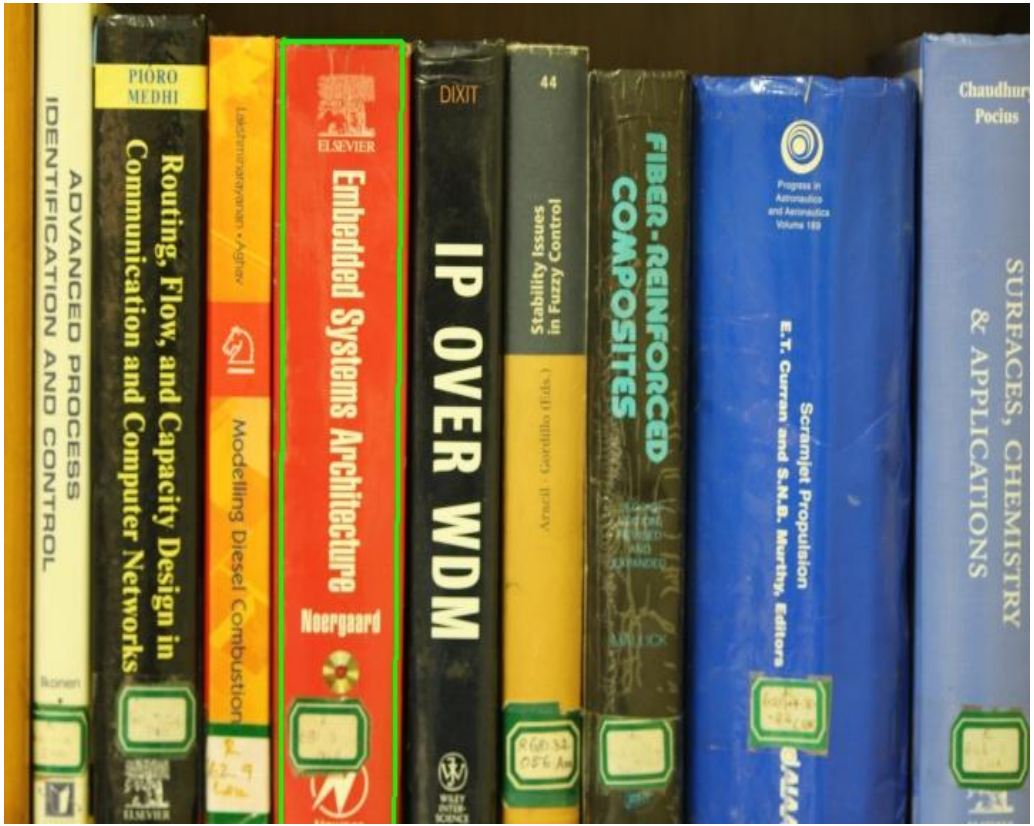
Cropped image corner coordinates

Bounding Box and Book Detection

After detecting the coordinates we draw the bounding box around the bounding box in the test frame



Query Image



Test Image with bounding box(green)

Merits and Demerits

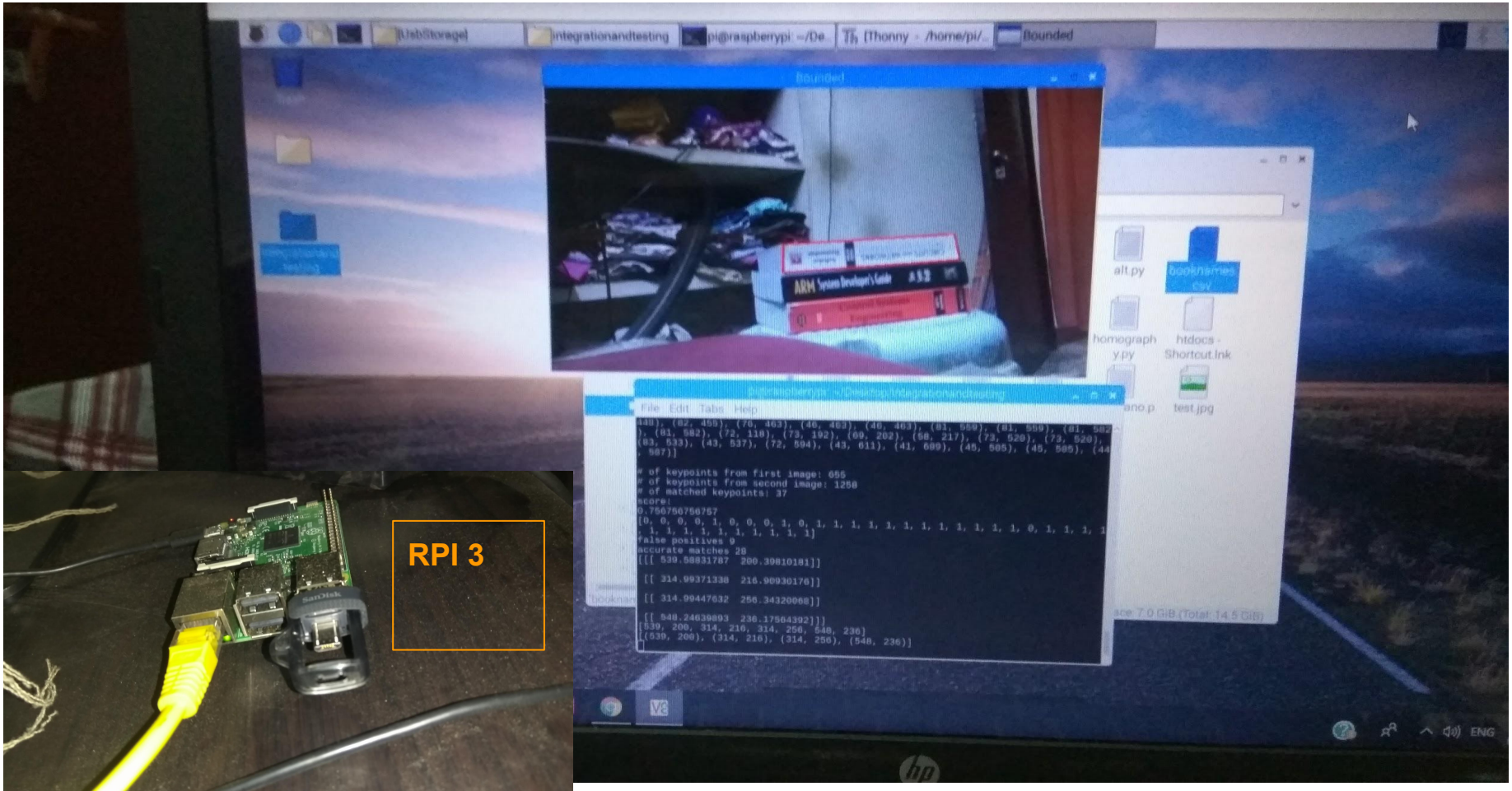
Merits:

- Only one image is required to identify any book.
- Computationally less expensive than machine learn based algorithm.
- Scale,affine,viewpoint,illumination,translation invariant.

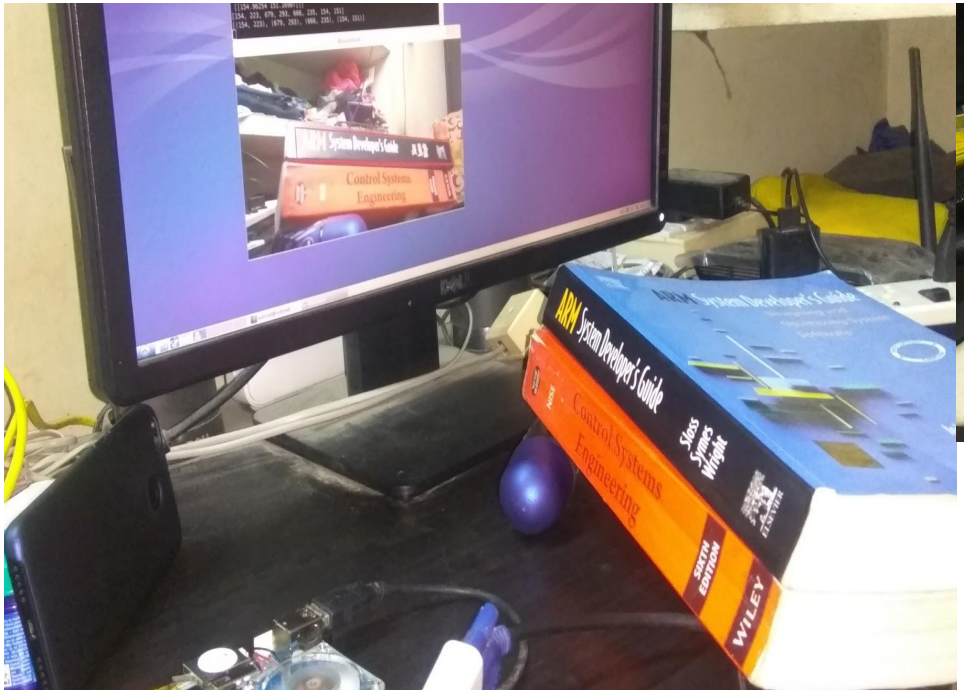
Demerits:

- Not hardware friendly algorithm.
- Less Ram based SBC suffers to run.
- Fails to identify same books of different edition since it is not based on ocr.

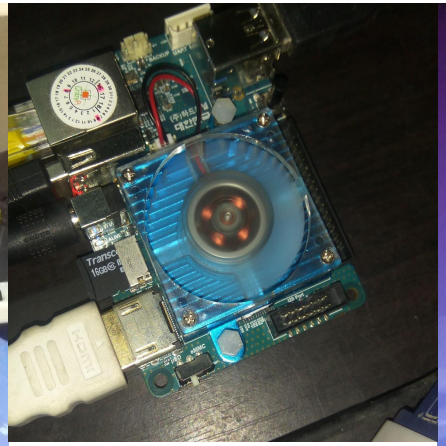
Hardware Implementation-Low end SBC



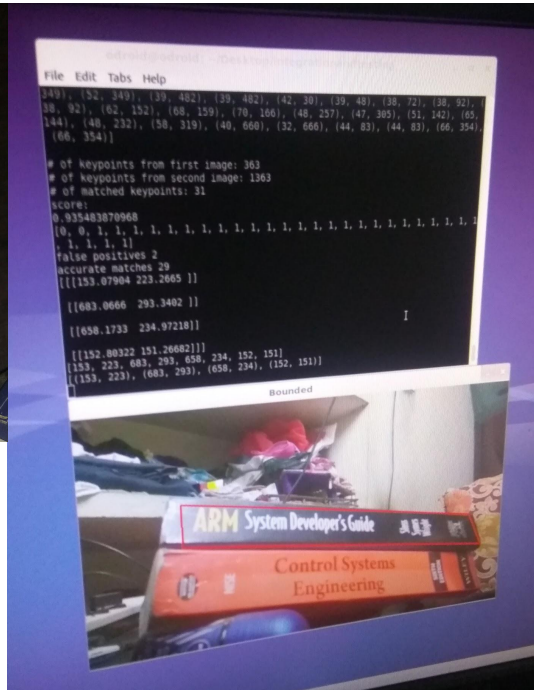
Hardware Implementation-High performance SBC



SOCKET BASED CAMERA CONNECTION

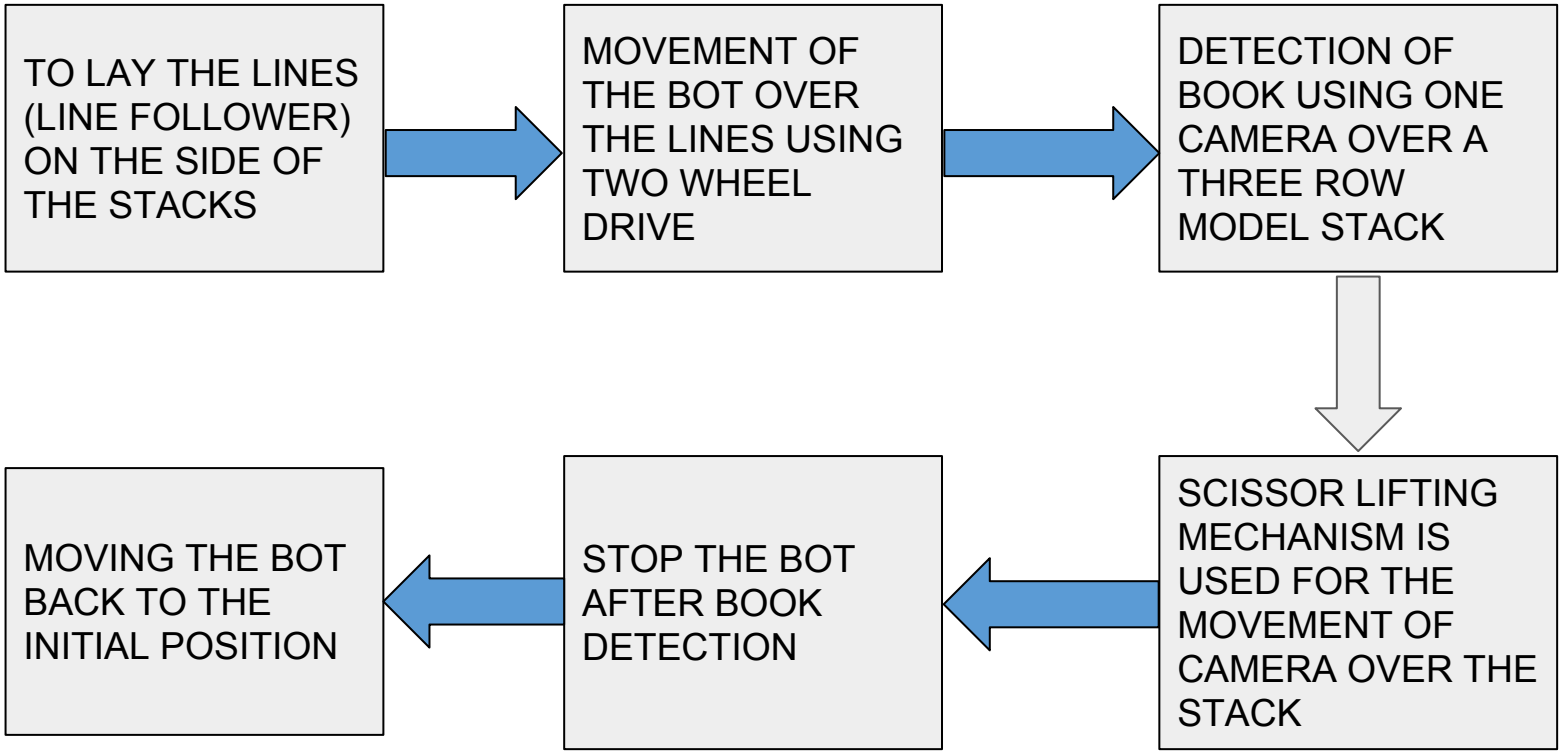


ODROID XU4

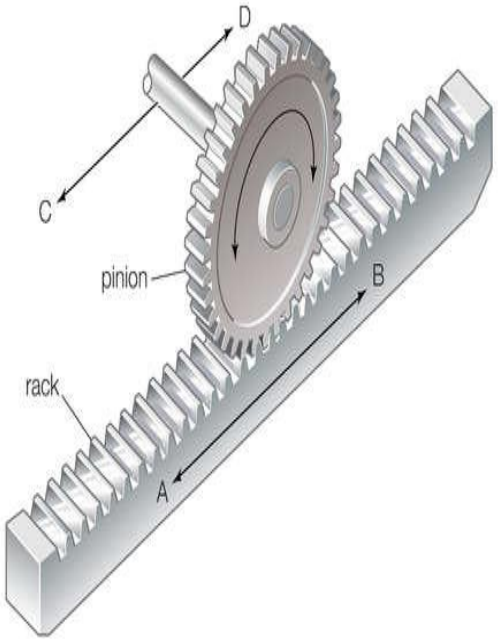


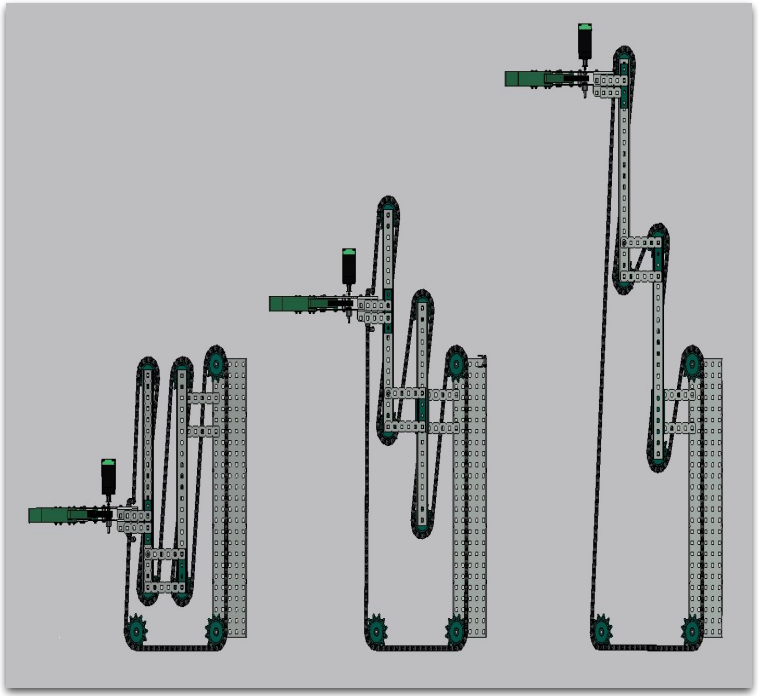
BOOK DETECTION (RED LINE)

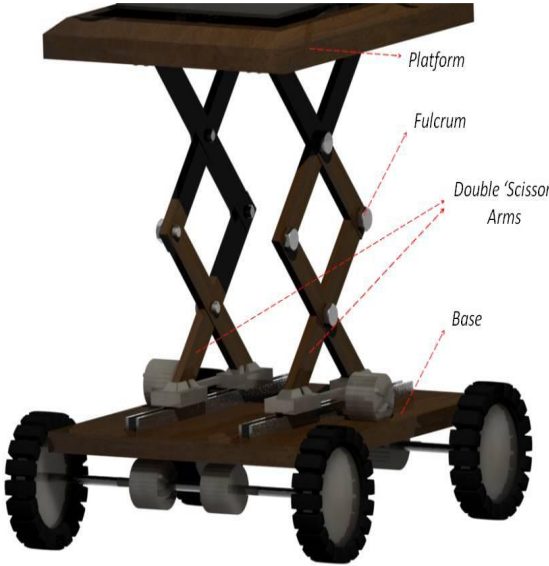
BOT DESIGN


TO LAY THE LINES (LINE FOLLOWER) ON THE SIDE OF THE STACKS
MOVEMENT OF THE BOT OVER THE LINES USING TWO WHEEL DRIVE
DETECTION OF BOOK USING ONE CAMERA OVER A THREE ROW MODEL STACK
SCISSOR LIFTING MECHANISM IS USED FOR THE MOVEMENT OF CAMERA OVER THE STACK
STOP THE BOT AFTER BOOK DETECTION
MOVING THE BOT BACK TO THE INITIAL POSITION

MECHANISM







RACK AND PINION MECHANISM

SCISSOR LIFT MECHANISM

References

[1]. *David G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints." International Journal of Computer Vision, 2004*

[2].*Mohammad Imrul Jubair; Prianka Banik, "A technique to detect books from library bookshelf image."2013 IEEE 9th International Conference on Computational Cybernetics (ICCC)*

[3].*Nuzhat Tabassum et al.**"**An approach to recognize book title from multi-cell bookshelf images".2017 IEEE International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*

[4].*Relja Arandjelović et al. "Three things everyone should know to improve object retrieval".2012 IEEE Conference on Computer Vision and Pattern Recognition*