Automatic Document Scanning

Open Software Project

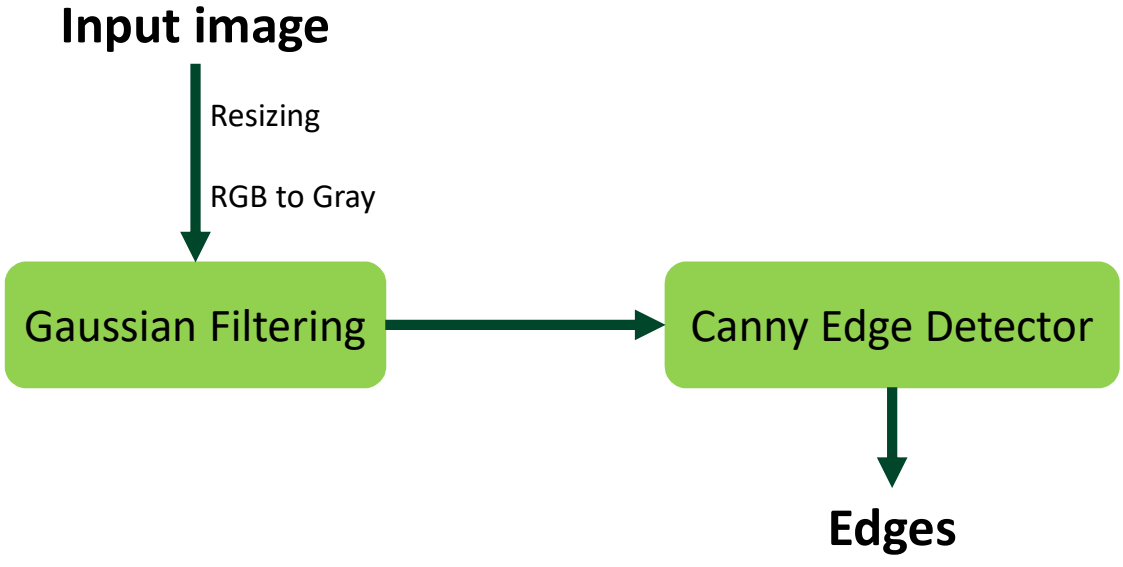
Yujin Hong, Dasom Jang

Outline

- **Step 1** Find Edges with Canny Edge Detector
- **Step 2** Sort regions and select a valid one
- **Step 3** Adjust the valid region within the rectangular output shape
- **Step 4** Make it clear using adaptive thresholding and denoise

Step 1

Edge Detection



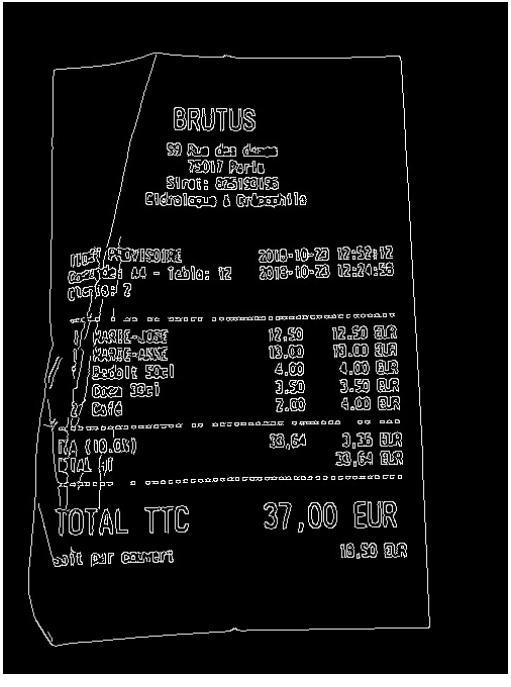
Step 1

Edge Detection

```
#Read the image
img=cv2.imread('input_1.jpg')

#Image resizing if needed
# -> When an image was too big or too small,
    #when an image resized too much, it makes a strange output.
if(img.shape[1]>1000 or img.shape[0]>1000):
    r=1000.0 / img.shape[1]
    dim=(1000, int(img.shape[0] * r))
    img=cv2.resize(img, dim, interpolation = cv2.INTER_AREA)
if(img.shape[1]<500 or img.shape[0]<500):
    r=500.0 / img.shape[1]
    dim=(500, int(img.shape[0] * r))
    img=cv2.resize(img, dim, interpolation = cv2.INTER_AREA)

#It shows the resized original image.
cv2.imshow('INPUT',img)

#Find edges
gray=cv2.cvtColor(img,cv2.COLOR_BGR2GRAY)
gray=cv2.GaussianBlur(gray,(5,5),0)
edge=cv2.Canny(gray,50,150)
```

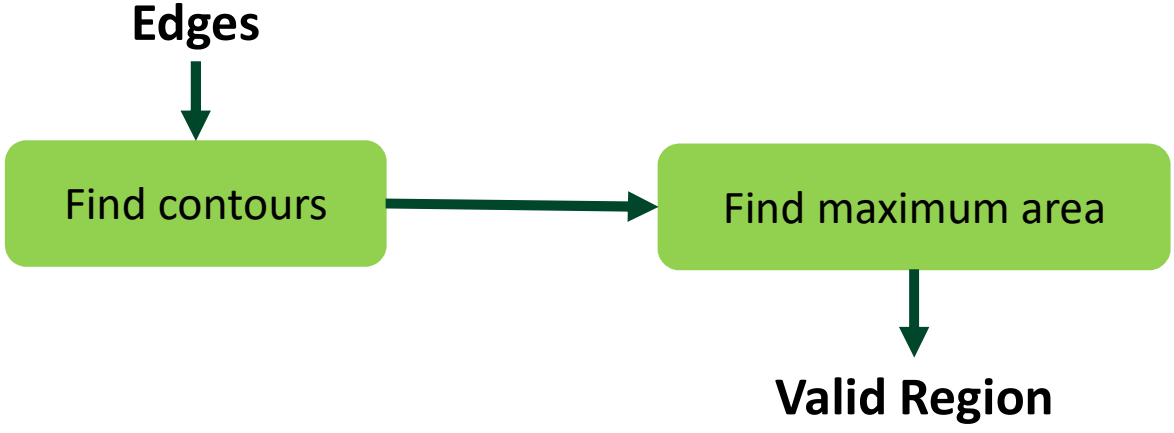
Step 1

Edge Detection



Step 2

Selecting a region



Step 2

Selecting a region

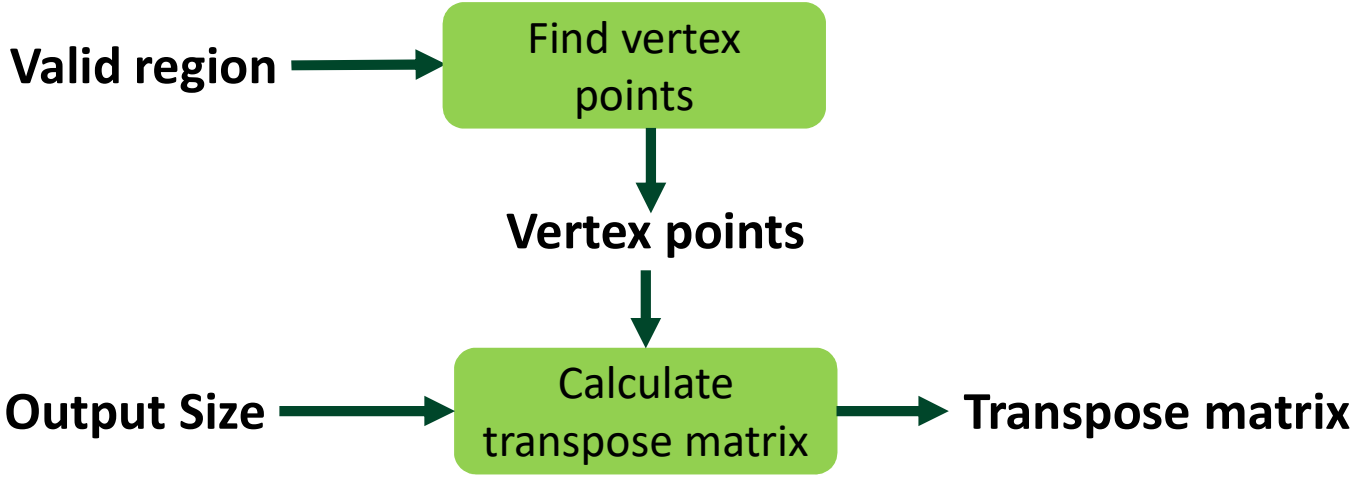
```
#Finding and drawing contours
contours,_=cv2.findContours(edge.copy(), cv2.RETR_EXTERNAL, cv2.CHAIN_APPROX_SIMPLE)
#                                        outermost among contours
#                                        return the points that can draw contour lines only
cv2.drawContours(img,contours,-1,[0,255,0],2)

#It shows the resized grayscale image with contours found above
#cv2.imshow('Contours',img)

#Find the part of the document in the image by contours
n=len(contours)
max_area=0
pos=0

for i in contours:
        area=cv2.contourArea(i)
        if area>max_area:
                max_area=area
                pos=i
```

Step 2

Selecting a region

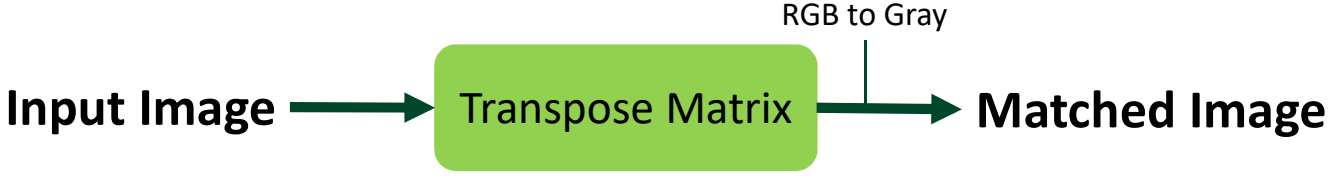
Step 3

Adjusting : using transpose matrix



Step 3

Adjusting : using transpose matrix



Step 3

Adjusting: using transpose matrix

```
# Find the corners of the object and the dimensions of the object
peri=cv2.arcLength(pos,True)
approx=cv2.approxPolyDP(pos,0.02*peri,True)
size=img.shape
w,h,arr=transform(approx)
# transform() :return the corners and the dimensions of the object

# Make a scanned document with perspective transformation
pts2=np.float32([[0,0],[w,0],[0,h],[w,h]])
pts1=np.float32(arr)
M=cv2.getPerspectiveTransform(pts1,pts2)

image=cv2.cvtColor(dst,cv2.COLOR_BGR2GRAY)
dst=cv2.warpPerspective(img,M,(w,h))
```

Step 3

Adjusting : using transpose matrix

BRUTUS
99 Rue des dames
75017 Paris
Siret: 825193196
Cidrologue & Crêpophile

TICKET PROVISOIRE 2018-10-23 12:52:12
Commande: A4 - Table: 12 2018-10-23 12:24:58
Clients: 2

1	MARIE-JOSE	12.50	12.50 EUR
1	MARIE-ANNE	13.00	13.00 EUR
1	Badoit 50cl	4.00	4.00 EUR
1	Coca 33cl	3.50	3.50 EUR
2	Café	2.00	4.00 EUR
TVA (10.0%)		33,64	3,36 EUR
TOTAL HT			33,64 EUR

TOTAL TTC 37,00 EUR
Soit par couvert 18,50 EUR

BRUTUS
99 Rue des dames
75017 Paris
Siret: 825193196
Cidrologue & Crêpophile

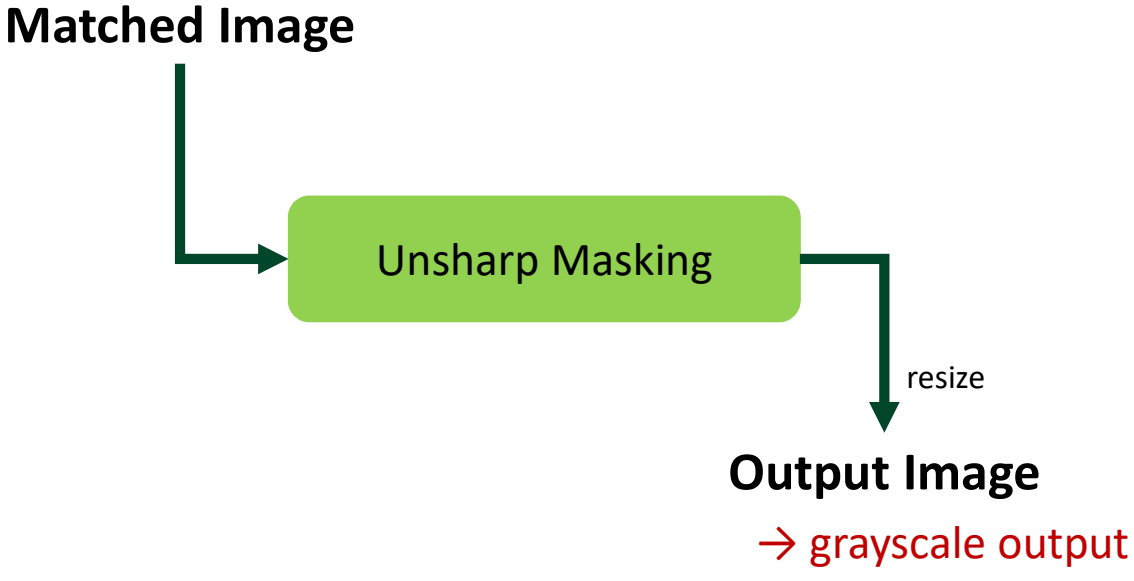
TICKET PROVISOIRE 2018-10-23 12:52:12
Commande: A4 - Table: 12 2018-10-23 12:24:58
Clients: 2

1	MARIE-JOSE	12.50	12.50 EUR
1	MARIE-ANNE	13.00	13.00 EUR
1	Badoit 50cl	4.00	4.00 EUR
1	Coca 33cl	3.50	3.50 EUR
2	Café	2.00	4.00 EUR
TVA (10.0%)		33,64	3,36 EUR
TOTAL HT			33,64 EUR

TOTAL TTC 37,00 EUR
Soit par couvert 18,50 EUR

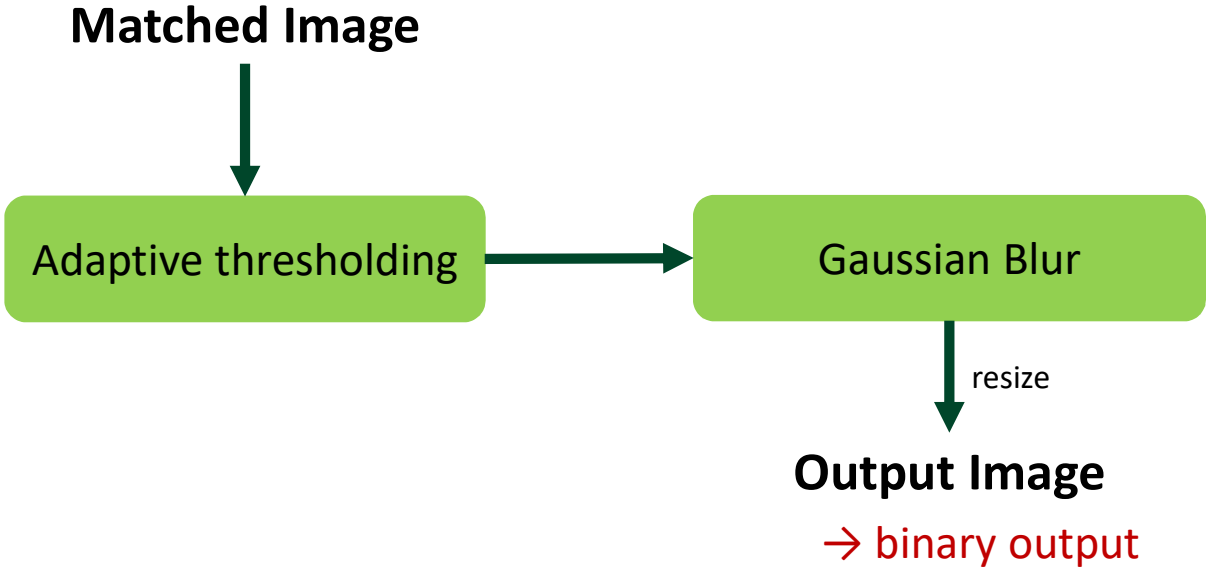
Step 4

Revision : sharpening



Step 4

Revision : thresholding and softening



Step 4

Revision : thresholding and softening

```
# Make the document clear with adaptive thresholding using moving averages
# and make letters smoother by Gaussian blurring
image=cv2.adaptiveThreshold(image, 255, cv2.ADAPTIVE_THRESH_MEAN_C, cv2.THRESH_BINARY, 7, 12)
image=cv2.GaussianBlur(image,(3,3),0)

image = cv2.resize(image,(w,h),interpolation = cv2.INTER_AREA)

#It shows the final output image (the scanned document)
cv2.imshow('OUTPUT',image)

#Save the final output image (the scanned document) and finish
cv2.imwrite('output.jpg',image)
```

Step 4

Revision : thresholding and softening

BRUTUS
99 Rue des dames
75017 Paris
Siret: 825193196
Cidrologue & Crêpophile

TICKET PROVISOIRE 2018-10-23 12:52:12
Commande: A4 - Table: 12 2018-10-23 12:24:58
Clients: 2

1	MARIE-JOSE	12.50	12.50 EUR
1	MARIE-ANNE	13.00	13.00 EUR
1	Badoit 50cl	4.00	4.00 EUR
1	Coca 33cl	3.50	3.50 EUR
2	Café	2.00	4.00 EUR

TVA (10.0%)	33,64	3,36 EUR
TOTAL HT		33,64 EUR

TOTAL TTC 37,00 EUR

Soit par couvert 18,50 EUR

BRUTUS
99 Rue des dames
75017 Paris
Siret: 825193196
Cidrologue & Crêpophile

TICKET PROVISOIRE 2018-10-23 12:52:12
Commande: A4 - Table: 12 2018-10-23 12:24:58
Clients: 2

1	MARIE-JOSE	12.50	12.50 EUR
1	MARIE-ANNE	13.00	13.00 EUR
1	Badoit 50cl	4.00	4.00 EUR
1	Coca 33cl	3.50	3.50 EUR
2	Café	2.00	4.00 EUR

TVA (10.0%)	33,64	3,36 EUR
TOTAL HT		33,64 EUR

TOTAL TTC 37,00 EUR

Soit par couvert 18,50 EUR

Results

Main Street Restaurant
6332 Business Drive
Suite 528
Palo Alto California 94301
575-1628095

Fri 04/07/2017 11:36 AM

Merchant ID: 9hqjxvufdr
Terminal ID: 11111

Transaction ID: #e6d598ef
Type: CREDIT

PURCHASE
Number: XXXXXXXXXXXX0041
Entry Mode: Swiped
Card Type: DISCOVER

Response: APPROVED
Approval Code: 819543

Sub Total USD$ 25.23
Tip: 3.78

Total USD$ 29.01

Thanks for supporting
local business!

THANK YOU

Main Street Restaurant
6332 Business Drive
Suite 528
Palo Alto California 94301
575-1628095

Fri 04/07/2017 11:36 AM

Merchant ID: 9hqjxvufdr
Terminal ID: 11111

Transaction ID: #e6d598ef
Type: CREDIT

PURCHASE
Number: XXXXXXXXXXXX0041
Entry Mode: Swiped
Card Type: DISCOVER

Response: APPROVED
Approval Code: 819543

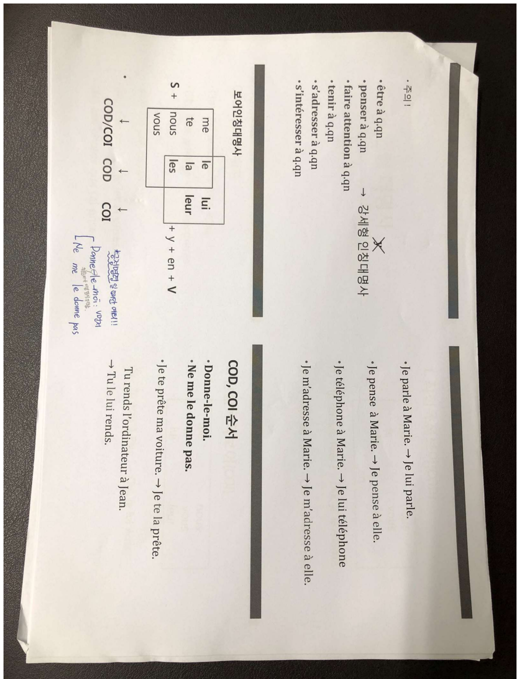
Sub Total USD$ 25.23
Tip: 3.78

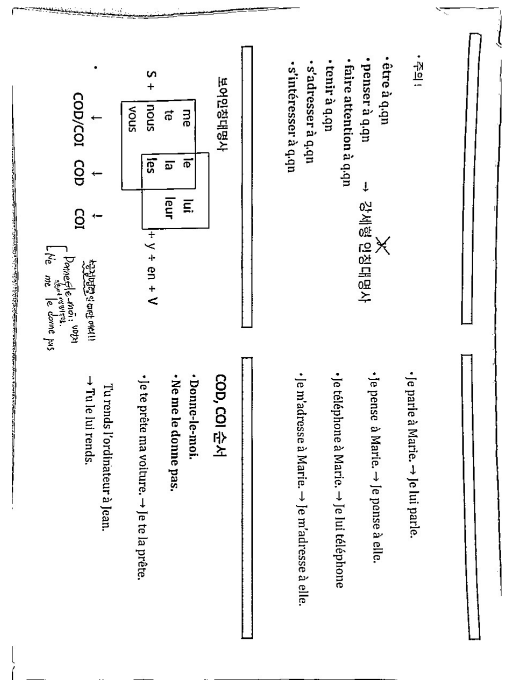
Total USD$ 29.01

Thanks for supporting
local business!

THANK YOU

Results





Results

We are GOPIC
Dear, Valued Homestay Families
We are GOPIC and we provide cheap and safe ride from your home to the airport services to students. We believe that you are the best partners for GOPIC, now and for the future.
As such, we would like to extend an offer to you and your household. Host families will receive $10 for every booking from your home to the airport.
Your opinions are important to us. If you have any questions or comments please call us at 416.855.3940 or 647.637.1831. You can also e-mail us at info@gopic.ca.
Thank you for your confidence and support.
Sincerely,
Team GOPIC
HOW GOPIC WORKS WITH HOMESTAY FAMILIES
-Introduce GOPIC's services to students when they need an AIRPORT DROP OFF service
To Pearson Airport: $60.00(Includes 2 Passengers)
•Extra Passenger: $10.00/person
Need a ride to the airport?
Call GOPIC
Get your commission
GOPIC
Call us at 416-855-3940 or 647-637-1831
With pick up request Date / Time & Address
www.gopic.ca tel: 416-855-3940 email: info@gopic.ca
office: #301-100 Sheppard Ave West Toronto ON M2N1M6

Thank You