

Алгоритм Брэдли

Подключаем нужные библиотеки

In [1]:

```
%matplotlib inline

import pathlib
import imageio
import cv2
import numpy as np
from tqdm import tqdm
import matplotlib.pyplot as plt
import matplotlib.image as mpimg
from skimage.color import rgb2gray
```

Загружаем ДатаСет

In [2]:

```
training_paths = pathlib.Path('1 - Image/Dataset').glob('**')
training_sorted = sorted([x for x in training_paths])
```

Реализация алгоритма бинаризации изображения

d - размер окна, по которому идет оценка

d_h - половина размера окна, т.к нам удобнее работать именно с ней

sz - подобранный параметр, для измерения размера окна

coef - подобранный параметр для нахождению трешхолда

S - матрица интегрального изображения

Реализация

1. В цикле заполняем нашу матрицу интегрального изображения, добавляем отсутп в виде нулей, для недопущения уменьшения размера картинки
2. Трешхолд высчитывается как среднее по окну + coef * среднее
3. За полняем изображение поэлементно

In [15]:

```
%%time

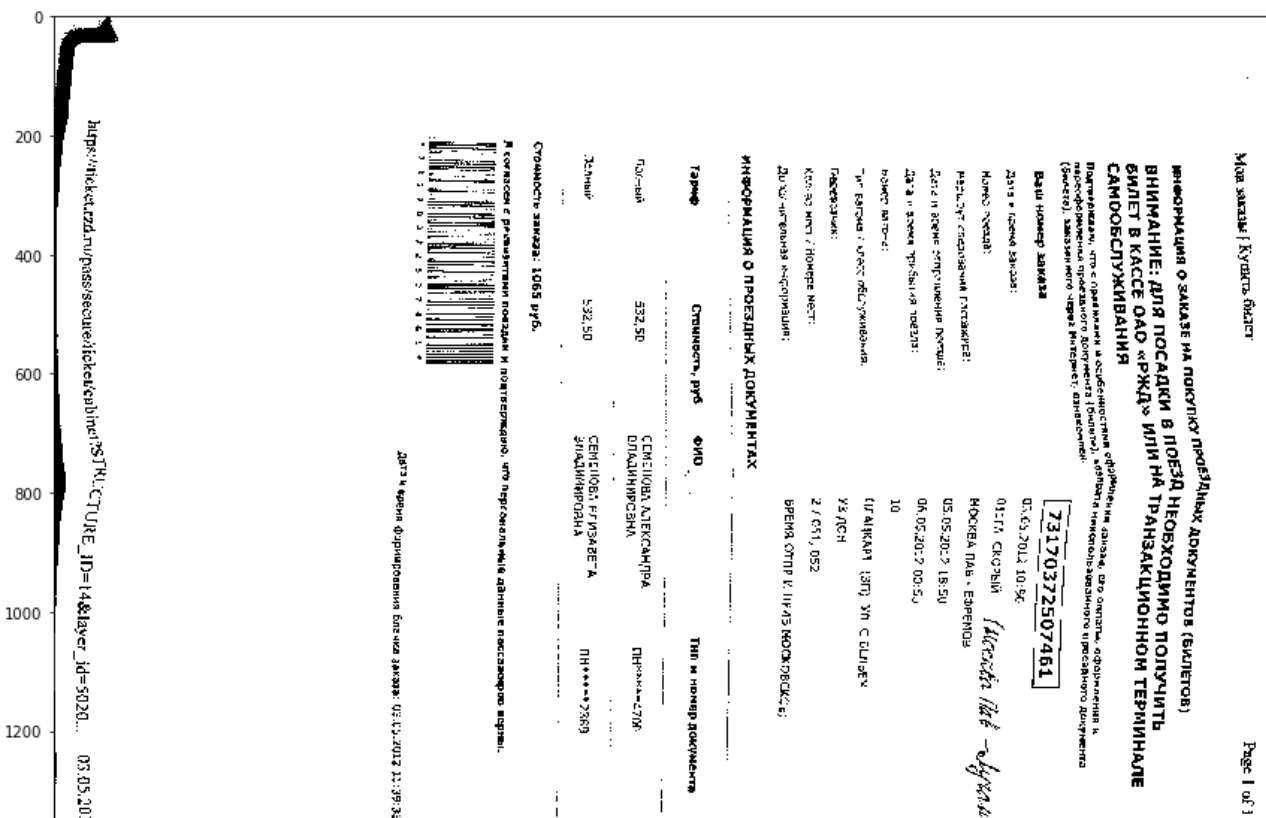
def binarizeImage(image, sz, coef):
    d = image.shape[0] // (int)(sz)

    d_h = d // 2

    s = np.pad(image, d_h, mode = 'constant')

    for i in range(d_h, image.shape[0]):
        for j in range(d_h, image.shape[1]):
            s[i, j] = s[i, j] + (s[i - 1, j] + s[i, j - 1] - s[i - 1, j - 1])

    def thresholding(ar):
        i = (int)(ar[0])
        j = (int)(ar[1])
```



1400

0 250 500 750 1000 1250 1500 1750 2000

5. Все остальные галки должны быть выключены.

6. Нажать ОК.

12. Выделить все изображения из архива и выбрать команду Tools->File->Recognize. Распознавание для создания баз нужно делать только с помощью этой команды.

13. После того, как распознавание будет закончено, файлы с базами символов появятся в папке Imagebase, откуда была запущена FlexCapture.exe, файлы имеют расширение *.bas. В большинстве случаев это единственный файл - "boxed.bas".

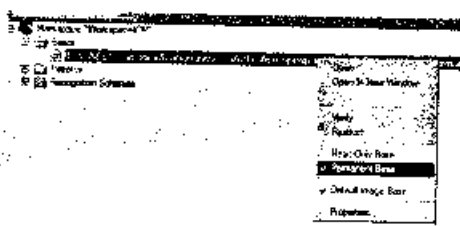
14. Готовый файл базы нужно удалить из папки WinRelease, предварительно скопировав в папку WinRelease\BAD Deployment\Images\Bases\Dirty и присвоив ему название по следующему шаблону: СокращенноеНазваниеЯзыка_ICR_ТипРазметки_Настр_Настр_ALL. Например: PT_ICR_Boxed_001_100_ALL.bas (№ страниц - такой же, как в названии проекта).

Разбиение файла базы на мелкие части для облегчения верификации.

Вся работа с базами ведется через программу UnifiedTeachNP.exe: находится в WinRelease технологический бинар 1.1 версии (последний). Далее базу необходимо предварительно быстро проверить:

15. Запустить UnifiedTeachNP.exe, создать/открыть существующий Workspace и открыть базу: СокращенноеНазваниеЯзыка_ICR_ТипРазметки_Настр_Настр_ALL.


16. Выставить следующие атрибуты у базы:



1. Снять галку Read-Only Base.

2. Поставить галку Permanent Base.

17. Пометить как dirty все или мусорные символы. Пометить как unknown символы, которые хоть и ясно читаются глазами, но к которым прилепился мусор от черного фона. Например:



Допустимо наличие одной-двух точек, если они не растягивают рамку графемы.

18. Затем разбить эту базу на две базы. Первая база с чистыми символами будет называться так: СокращенноеНазваниеЯзыка_ICR_ТипРазметки_Настр_Настр. Например: PT_ICR_Boxed_001_100.bas. Вторая база: UNKNOWN будет называться так: СокращенноеНазваниеЯзыка_ICR_ТипРазметки_Мусор_Мусор_UNKNOWN.

6

500

1000

0 500 1000 1500 2000

18. Затем разбить эту базу на две базы. Первая база с чистыми символами будет называться так: СокращенноеНазваниеЯзыка_ICR_ТипРазметки_Настр_Настр. Например: PT_ICR_Boxed_001_100.bas. Вторая база: UNKNOWN будет называться так: СокращенноеНазваниеЯзыка_ICR_ТипРазметки_Мусор_Мусор_UNKNOWN.

забўе большую базу на малым кіры

- Cont. Name** **Page**

☐ List all items, including all images by description

☐ List only by description

Form for 1

Form for 2

Form for 3

Form for 4

Form for 5

Form for 6

Form for 7

Form for 8

Form for 9

Form for 10

Form for 11

Form for 12

Form for 13

Form for 14

Form for 15

Form for 16

Form for 17

Form for 18

Form for 19

Form for 20

Form for 21

Form for 22

Form for 23

Form for 24

Form for 25

Form for 26

Form for 27

Form for 28

Form for 29

Form for 30

Form for 31

Form for 32

Form for 33

Form for 34

Form for 35

Form for 36

Form for 37

Form for 38

Form for 39

Form for 40

Form for 41

Form for 42

Form for 43

Form for 44

Form for 45

Form for 46

Form for 47

Form for 48

Form for 49

Form for 50

Form for 51

Form for 52

Form for 53

Form for 54

Form for 55

Form for 56

Form for 57

Form for 58

Form for 59

Form for 60

Form for 61

Form for 62

Form for 63

Form for 64

Form for 65

Form for 66

Form for 67

Form for 68

Form for 69

Form for 70

Form for 71

Form for 72

Form for 73

Form for 74

Form for 75

Form for 76

Form for 77

Form for 78

Form for 79

Form for 80

Form for 81

Form for 82

Form for 83

Form for 84

Form for 85

Form for 86

Form for 87

Form for 88

Form for 89

Form for 90

Form for 91

Form for 92

Form for 93

Form for 94

Form for 95

Form for 96

Form for 97

Form for 98

Form for 99

Form for 100

Form for 101

Form for 102

Form for 103

Form for 104

Form for 105

Form for 106

Form for 107

Form for 108

Form for 109

Form for 110

Form for 111

Form for 112

Form for 113

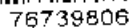
Form for 114

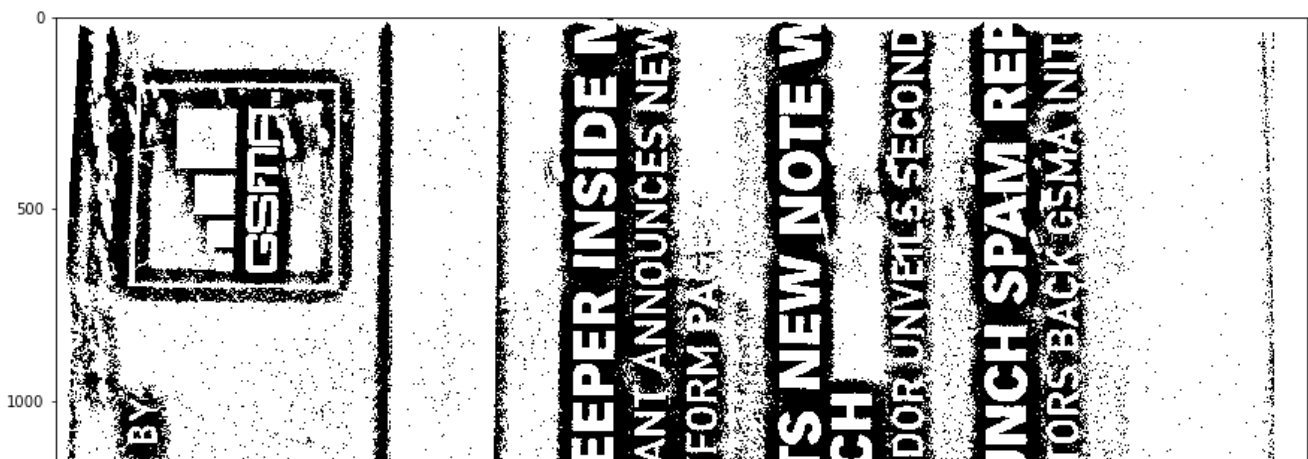
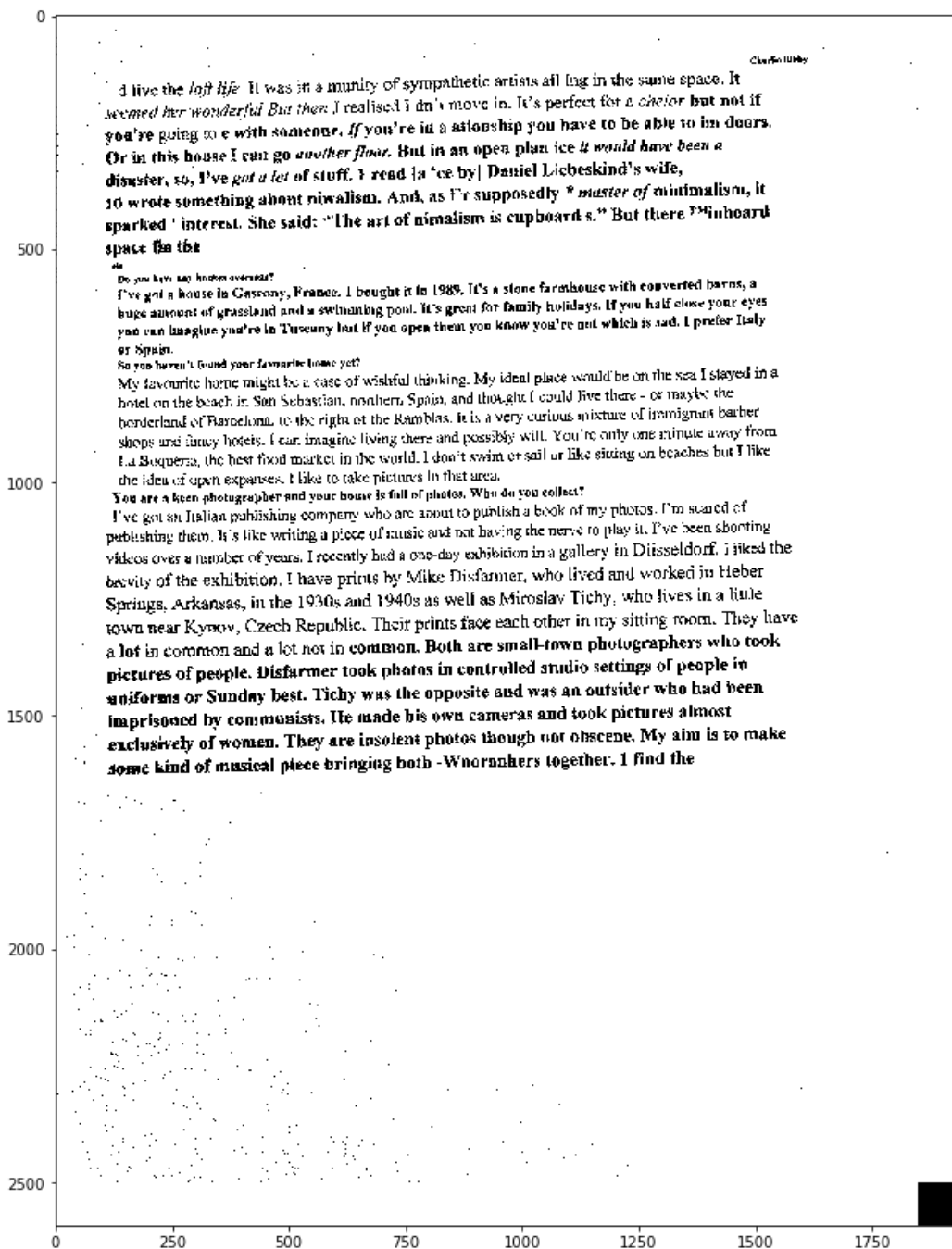
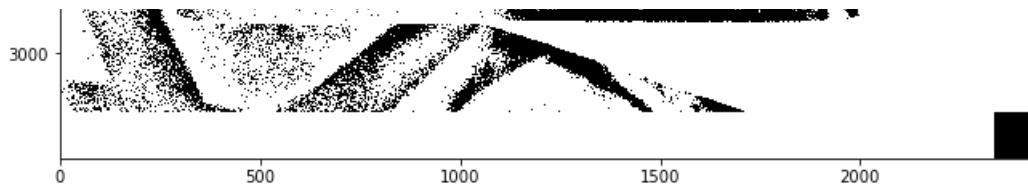
Form for

1	1 10
2	2 13
3	3 15

Репертуар:
 «Патриот» - Тони - Джек М. Кутсон - «Секретный материал»
 - Рэй Брайбери - «Век из одуванчиков» - Жюль Сарамбо
 «Письма от Миссури» - Джонни Харрис - «Шоколад» - Морис Дини
 «Желтый король» - Харуки Мураками - «Оста на оленя» - Арно
 Ренес-Риверс - «Колла для барабана» - Кен Киз - «Над пропастью во ржи»
 - Джозеф - Дороти Д. Салливан - «Над пропастью во ржи»

OZON.RU





would like to congratulate the Global Mobile

MWOMEN BEST MOBILE PRODUCT OR SERVICE FOR WOMEN IN EMERGING MARKETS

Etisalat, Qualcomm, D-Free International and Great Connection Inc. - Etisalat Mobile Baby

BEST USE OF MOBILE IN EMERGENCY OR HUMANITARIAN SITUATIONS

UBL Omni - transparent and efficient Cash Disbursement service after the 2009 IDPs Crisis and the 2010 Flood Crisis

THE GREEN MOBILE AWARD

Flexenclosure - E-site

BEST MOBILE HEALTH INNOVATION

Etisalat, Qualcomm, D-Free International and Great Connection Inc. - Etisalat Mobile Baby

BEST MOBILE INNOVATION FOR EDUCATION OR LEARNING

OnPoint Digital - CellCast Solution

BEST MOBILE INNOVATION FOR AUTOMOTIVE, TRANSPORT OR UTILITIES

Ford Motor Company Ltd. - Ford SYNC with Emergency Assistance

BEST MOBILE MONEY INNOVATION

Etisalat, MasterCard and Oberthur Technologies - Etisalat Commerce

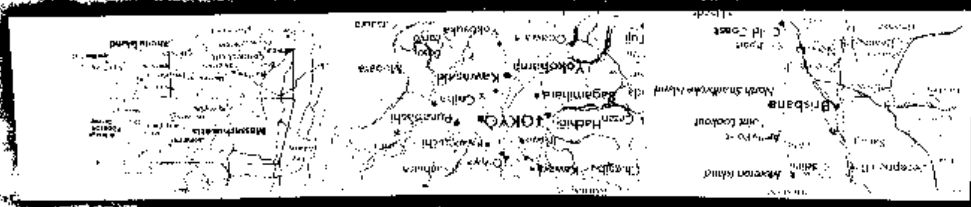
BEST MOBILE INNOVATION FOR PUBLISHING

Financial Times and Assanka - The Financial Times Web

Collins mobile Coverage is a web based roaming coverage map service made available through CollinsBartholomew's partnership with the GSMA. Using the latest mapping technology, Collins mobile Coverage combines up-to-date world base maps with unique network coverage data provided by operators from around the world. These search pages are delivered straight to network operators' mobile phones to help them find where they can use their phones when abroad. Collins mobile Coverage is a web based roaming coverage map service made available through CollinsBartholomew's partnership with the GSMA. Using the latest mapping technology, Collins mobile Coverage combines up-to-date world base maps with unique network coverage data provided by operators from around the world. These search pages are delivered straight to network operators' mobile phones to help them find where they can use their phones when abroad.

BROUGHT TO YOU BY
INTEL GETS D
SEMICONDUCTOR GI
THE MEDFIELD PLAT
SAMSUNG HI
TABLET LAUN
SOUTH KOREAN VEN
CARRIERS LAU
BIG FOUR US OPERA

COME AND VISIT US AT THE GSM A PAVILION IN HALL 8



NETWORK ROAMING COVERAGE MAP SERVICE FOR

mobile COVERAGE

Experience a world
where everything
intelligently connects
The Connected Day.

Free for all delegates

There is incredible excitement around Near Field Communications (NFC) today and the news has been awash with service launches, NFC-enabled devices and new alliances. While NFC is perhaps most closely associated with mobile payments, it goes far beyond; NFC will facilitate a wide range of new applications for consumers, such as mobile ticketing to board public transportation, the exchange of information and content, control access to cars, homes, hotels, offices and car parks and more. This is just the tip of the iceberg - the possibility for innovation is endless.

The market potential for NFC is significant - nearly 1.5 billion SIM-based handsets will have been sold worldwide between 2010 and 2016, reporting transactions of more than \$50 billion globally over the same period according to Strategy Analytics, and momentum is growing.

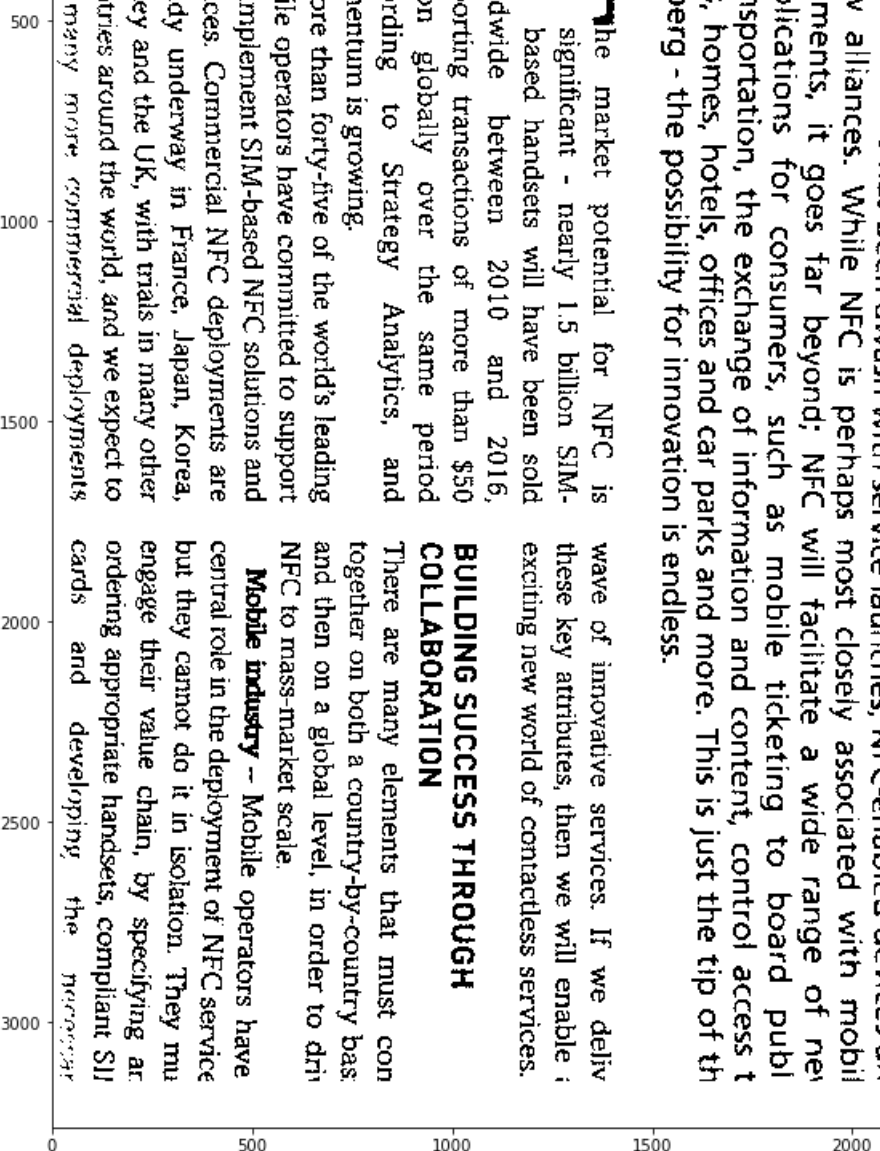
More than forty-five of the world's leading mobile operators have committed to support and implement SIM-based NFC solutions and services. Commercial NFC deployments are already underway in France, Japan, Korea, Turkey and the UK, with trials in many other countries around the world, and we expect to see many more commercial deployments

in wave of innovative services. If we deliver these key attributes, then we will enable an exciting new world of contactless services.

BUILDING SUCCESS THROUGH COLLABORATION

There are many elements that must come together on both a country-by-country basis, and then on a global level, in order to drive NFC to mass-market scale.

Mobile industry – Mobile operators have a central role in the deployment of NFC service but they cannot do it in isolation. They must engage their value chain, by specifying and ordering appropriate handsets, compliant SIM cards and developing the necessary



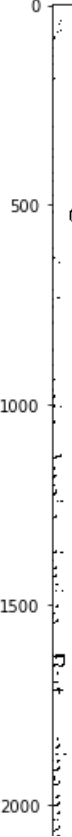
China mobile connections, Q4 2011

China Mobile		China Unicom*
Connections (m)	648.7	199.7
Connections, 3G* (m)	51.8	40
% 3G*	8%	20%
Market Share	67%	21%
Market Share, 3G*	24%	18%
Net Additions (m)	15.2	10.6
Net Additions, 3G* (m)	8.7	9.8
Growth, YoY	11%	19%

Source: Wireless Intelligence • *Includes CDMA2

China Mobile (CHT) have been able to migrate more subscribers to their respective 3G networks is partly due to them being able to tap into a broader range of 3G smartphones compared to what is currently available for China Mobile's homegrown TD-SCDMA network.

Unicom has been the exclusive provider of the iPhone since 2009, while China Telecom is thought to be close to launching a CDMA



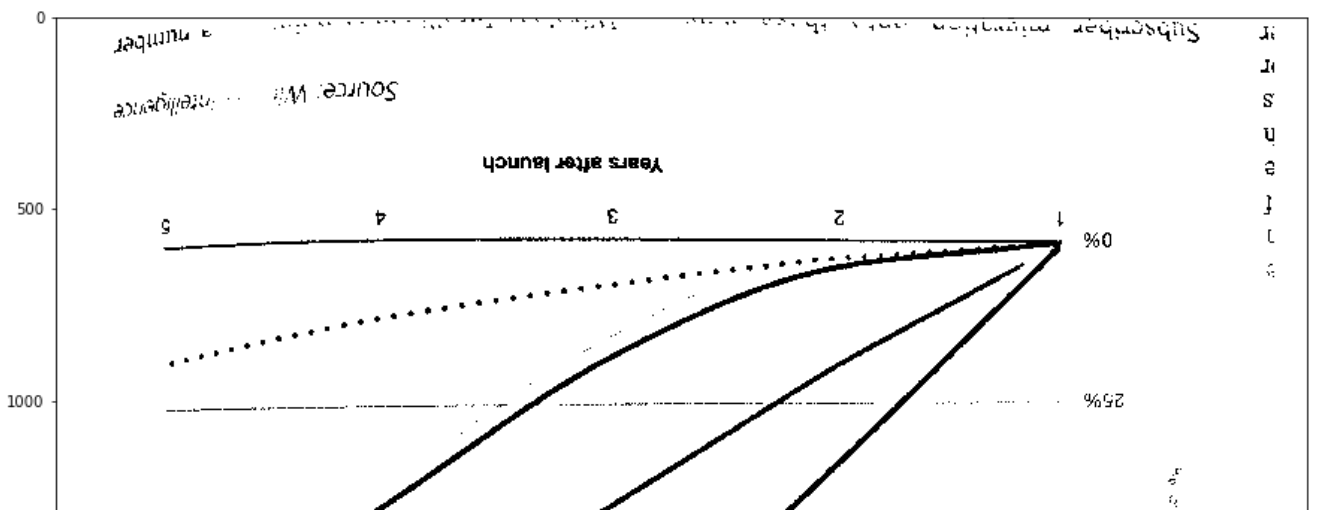
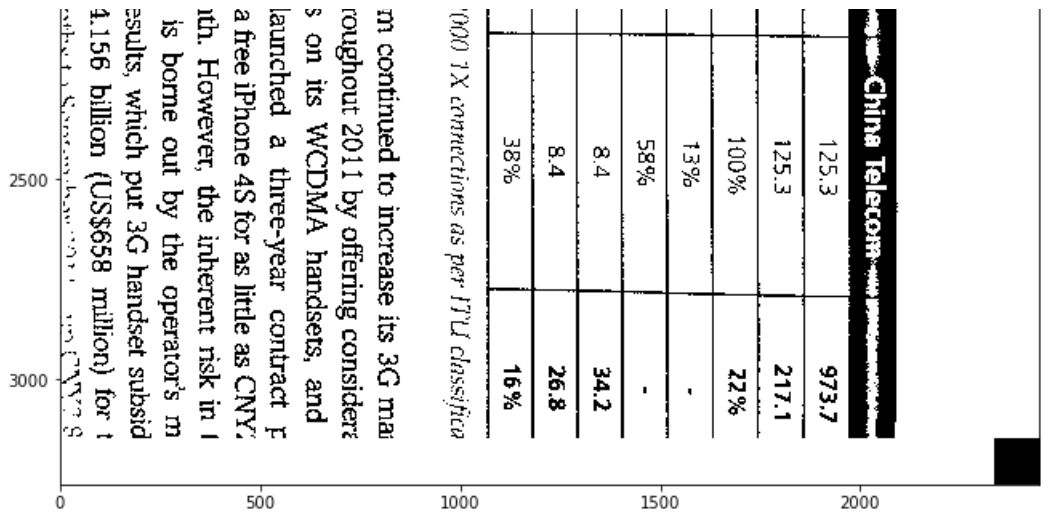
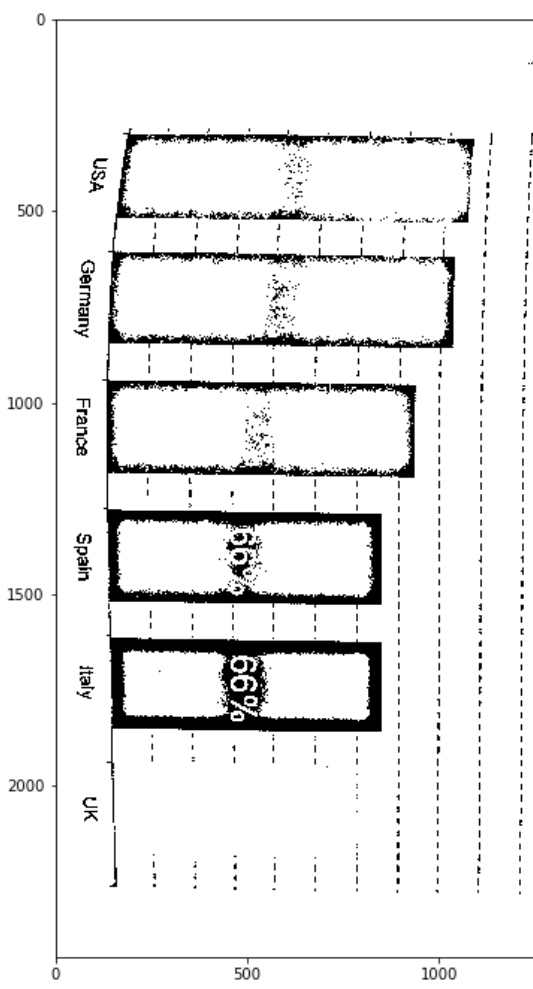
campaigns running on mobile. The term 'rich media' is quickly entering the mobile advertising lexicon and is certainly creating a lot of hype. Although potentially quite broad in definition, it generally refers to ads that are based on the benefits of HTML5 technology and offer consumers the opportunity to interact with the branded ad content itself. Essentially, a rich media ad can replicate the power of a TV ad on the device screen, delivering the TV experience combined with touchscreen interactivity that is unique to mobile.

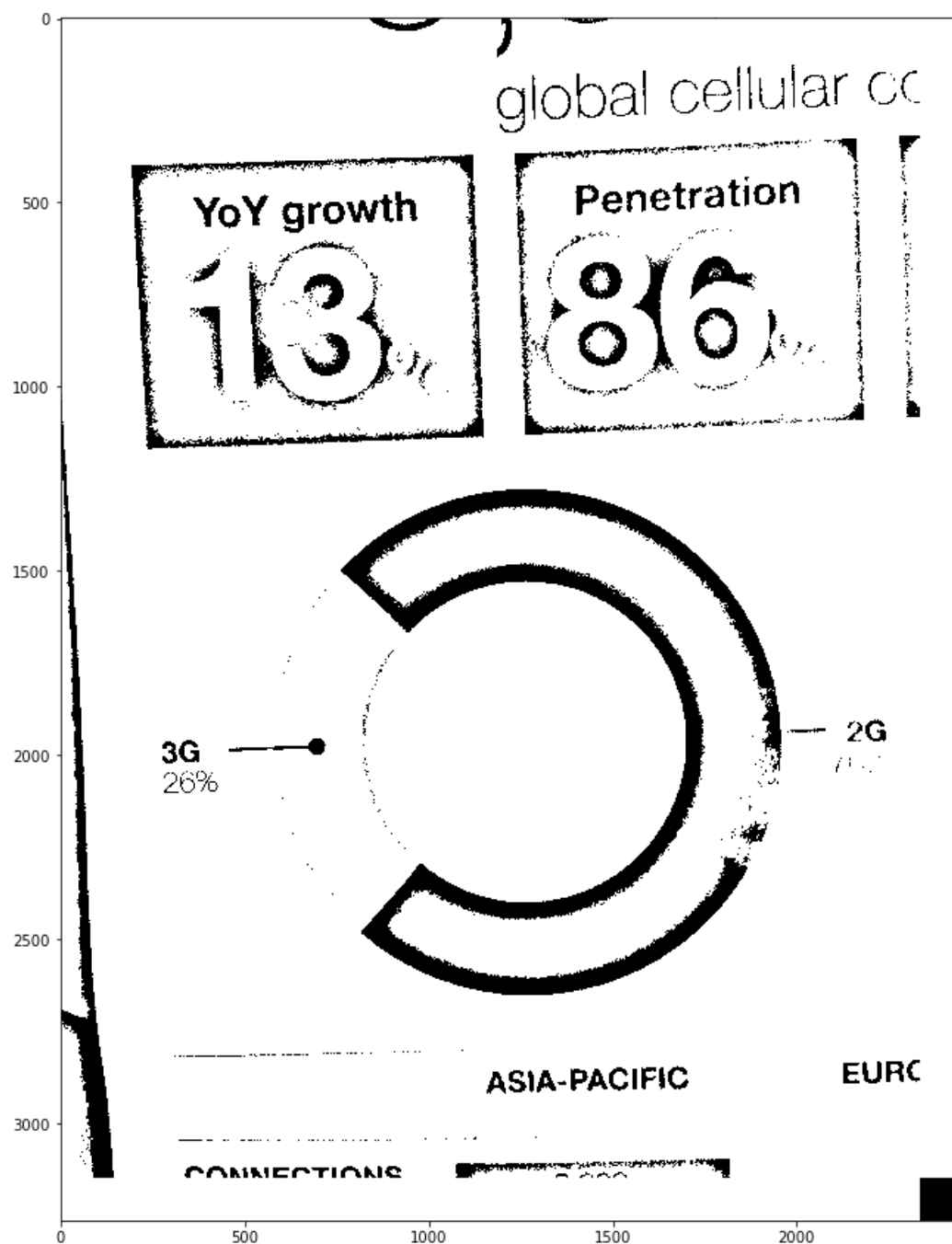
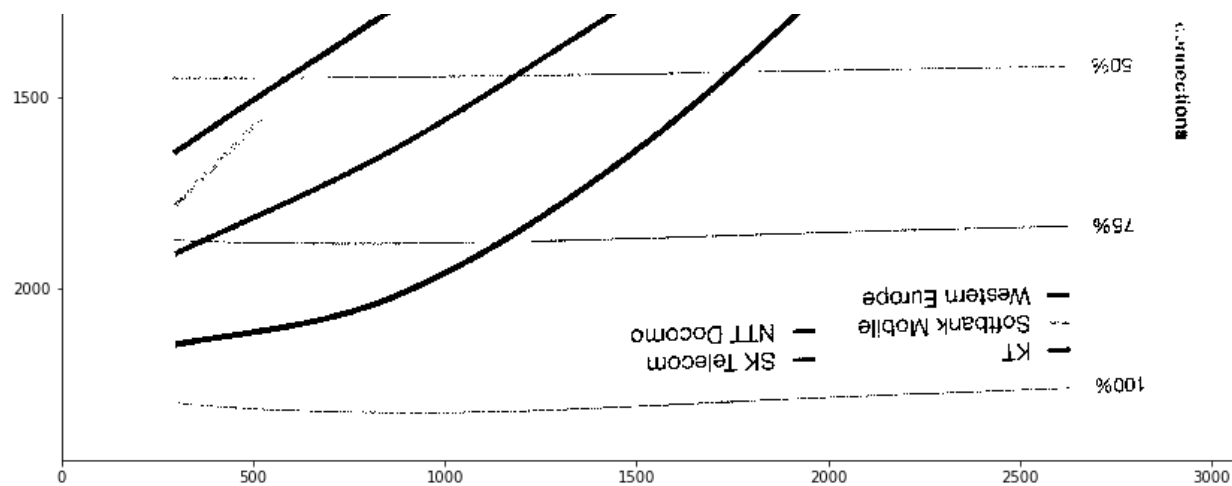
IT'S ALL ABOUT NUMBERS

There are a number of factors that have come together to drive the potential of rich media as an advertising format. The first is mass penetration of smartphone devices. There were approximately 252 million iOS devices and 243 million Android devices globally as of the end of 2011. That's a reach of almost half a billion people. In the UK, for example, there are approximately 15 million Android and iOS smartphone devices; a quarter of the UK population is already rich media enabled. The second is standards. OMA (Open Rich Media Mobile Advertising), an industry wide open initiative for advertisers, has been standardised by the AB Mobile Marketing Center of Excellence to create MPAU (Mobile Rich Media Ad Interface Definition) which defines a common API for mobile rich media ads that will help in mobile apps and web sites. This effectively provides the industry standard required for rich media to move mainstream.

Third, from a creative perspective, rich media gives advertisers complete design flexibility to deliver unique and memorable experiences. Expandable banners are the most common rich media format, whereby consumers tap on a banner and get presented with countless options, and experiences such as interactive mobile pages, videos or a combination of the two.

Table 1. Android and iOS penetration in USA and EU5 countries as a percentage of overall smartphone devices. Available rich media ad impressions are now at critical mass







Оценка времени работы

Проведем теоритическую оценку времени работы алгоритма:

Пусть размер изображения в пикселях N ($N = \text{width} * \text{height}$)

1. Перевод в черно-белое изображение, вернее в серое занимает порядка N операций
2. Вычисляем матрицу S за N операции
3. Проходим по изображению и делаем картинку черно-белой за N операции
4. Итого $3 * N$ операций

Код написанный на CPython выполняет примерно 70 000 000 операций за одну секунду,

значит время обработки изображения $\frac{3 * N}{70000000}$

Если изображение $3000 * 2000$ пикселей, то время работы составит 0.25 секунд.

Результаты

В некоторых местах текст получился белого, а не черного цвета, можно поправить, если проверить каков является фон изображения.

Некоторые фотографии зашумлены, можно попытаться изначально удалить шум, а потом бинаризовать, или еще больше поэкспериментировать с параметрами.

В некоторых изображениях текст бледный, на них можно уменьшить трешхолд. Вообще не мешало бы выбирать параметры трешхолда, т.е параметр размера окна относительно всего изображения и коэффициент, на который умножаем трешхолд, исходя из анализа каждого отдельного изображения

In []: