

```
In [1]: import numpy as np
import scipy as sp
from sklearn.model_selection import train_test_split
import gc
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from PIL import Image
from tqdm import tqdm
from scipy.signal import medfilt
sns.set(font_scale=0)
%matplotlib inline
```

```
In [2]: def ShowImage(img, isGrey = False):
plt.figure(figsize=(20, 15))
if not isGrey:
    plt.imshow(img, interpolation='nearest')
else:
    plt.imshow(img, cmap='gray')
plt.axis('off')
gc.collect()
```

Алгоритм Бредли

ссылка на алгоритм - <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.420.7883&rep=rep1&type=pdf>
(<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.420.7883&rep=rep1&type=pdf>)

Преобразование в изображение в серых тонах - <https://en.wikipedia.org/wiki/YUV>

<https://en.wikipedia.org/wiki/YUV>

Данный алгоритм был выбран мной, так как он дает хорошую биноризацию изображения, при этом работает очень быстро (на C++).

Все константы изначально были взяты из оригинальной статьи, но поэкспериментировав с `t_proc` (в статье просто `t`), я изменил константу с оригинальной 0.85 на 0.9 (в данном ноутбуке изображения с константой 0.85, в итоговых изображениях порог взят 0.9 (были получены из кода на c++)). Разница в улучшении больше всего заметна на фотографии с задней стороны книги (4 пример). Результаты времени работы, а также общие соображения по результатам будут описаны в конце.

```
In [4]: def create_grey_image(image, red_coeff = 0.299, green_coeff = 0.587):
        xr = np.copy(image[:, :, 0])
        xg = np.copy(image[:, :, 1])
        xb = np.copy(image[:, :, 2])
        I = red_coeff * xr + green_coeff * xg + (1 - green_coeff - red_coeff) * xb
        #I = medfilt(I, 3)
        return I
```

```
In [5]: def integrate_image(f):
        I = f.copy()
        N = I.shape[0]
        M = I.shape[1]
        for i in range(1, N):
            for j in range(1, M):
                I[i][j] = I[i][j] + I[i - 1][j] + I[i][j - 1] - I[i - 1][j - 1]
        return I
```

```
In [6]: def make_answer(I, F, t_proc = 0.85, s = None):
        N = I.shape[0]
        M = I.shape[1]
        if s is None:
            dh = 1 / 8 * N
            dw = 1 / 8 * M
            s = int(min(dh, dw) // 2)
        answer = np.zeros((N, M))
        for i in range(0, N, 1):
            for j in range(0, M, 1):
                imax = int(min(i + s, N - 1))
                jmax = int(min(j + s, M - 1))
                imin = int(max(i - s, 0))
                jmin = int(max(j - s, 0))
                size = (imax - imin) * (jmax - jmin)
                summ = I[imax, jmax]
                if imin >= 1:
                    summ -= I[imin - 1, jmax]
                if jmin >= 1:
                    summ -= I[imax, jmin - 1]
                if jmin >= 1 and imin >= 1:
                    summ += I[imin - 1, jmin - 1]
                if F[i][j] >= (summ / size) * t_proc:
                    answer[i][j] = 255
        return answer
```

```
In [33]: def Get_Bradley_Binorization(number, image, red_coeff = 0.299, green_coeff = 0.587, t_pr
        F = create_grey_image(image, red_coeff, green_coeff)
        print(F[100, 100])
        I = intagrate_image(F)
        print(I[100, 100])
        answer = make_answer(I, F, t_proc, s)
        plt.imsave("Result/bradley_" + i + ".png", answer, cmap='gray')
        return answer
```

```
In [19]: for i in tqdm(["01", "02", "03", "04", "05", "06", "07", "08", "09", "10", "11", "12", "  
    path = 'Dataset/' + i + '.JPG'  
    image = plt.imread(path)  
    image = image / 255  
    answer = Get_Bradley_Binorization(i, image)  
    ShowImage(answer, True)
```

100%|██████████| 15/15 [24:03<00:00, 102.71s/it]

Платителем, что с правомки и особыми условиями, возврата неиспользованного проездного документа, трансформации проездного документа (бумажный, электронный, билет), зачисление через интернет, единовремен-

73170372507461

03.05.2012 10:50

ОЛГА СКОПИН
МОСКВА ПЛБ - ЕДРЕМОВ
(Merveta Plab - edremov).

МОСКВА ПАВ - ЕДРЕМОБ

05.05.2012 18:50

06.05.2012 00:55

10

МІЛІТАРНА (3М) УО С БІЛІСЕН

NOTES

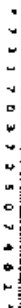
21051, 052

ВРЕМЯ ОТП И ПРИБ МОСКОВСКОЕ;

Тема	Степень, руб	ФНО	Тип и номер документа
------	--------------	-----	-----------------------

Стоймость заказа: 1065 руб.

и согласен с рекламными подзвонками и подтешиваю, что персональные данные пассажира вбросил



Дата и время формирования бланка заказа: 03.05.2012 11:39:35

https://ticket.rzd.ru/pass/secure/ticket/cabin?%7B%22STRUCTURE_ID%3D14&layer_id=5020.. 03.05.2012

5. Все остальные файлы должны быть выполнены.
6. Нажать OK.
12. Выделить все изображения из папки и выбрать команду Tools->Release. Распакованные для создания баз нужно делать только с помощью этой команды.
13. После того, как распаковывание будет закончено, файлы с базами символов появятся в папке WinRelease, откуда была запущена FlexCart6.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 \\Microsoft\\BDD-Development\\FontReader\\Images\\Bases\\Only и присвоив ему название по следующему шаблону: СоколициноНазваниеБаза_JCR_ТипРазметки_Нестр_Нестр_ALL. Например: RT_JCR_Boxed_001_100_ALL.bas (We-страниц – такой же, как в названии проекта).

Разбавление файла базы на мелкие части для облегчения верификации.

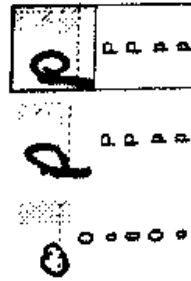
Вся работа с базами ведется через программу UnifiedTeachNIPR.exe: находится в WinRelease технологический билд 11 версии (последний). Далее базу необходимо предварительно быстро поместить.

15. Запустить UnifiedTeachNIPR.exe, создать/открыть существующий Workbase и открыть базу СоколициноНазваниеЯзыка_JCR_ТипРазметки_Нестр_Нестр_ALL.
16. Выставить следующие атрибуты у базы:



1. Снять галку Read-Only Base;
2. Поставить галку Permanent Base.

17. Пометить как илклоуп все явно мусорные символы;
Пометить как илклоуп символы, которые хоть и ясно читаются глазами, но к которым приклеился мусор от серого 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_Nestr_Nestr`
Например: `PT_ICR Boxed_001_100_bas`

Будет называться так:

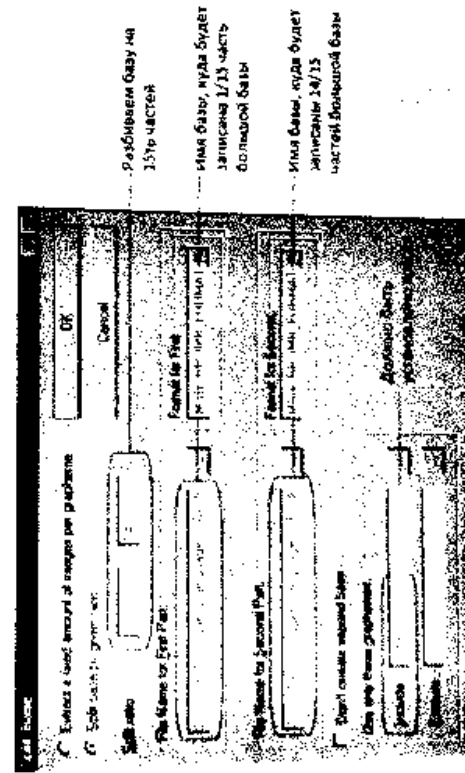
Содержимое файла: ICR_Таблицы_Испр_Испр_UNKNOWN
Например: PT_ICR_Boxed_001_100_UNKNOWN.bas

Даже необходимо разбить получившуюся с одного пакета базу с чистыми символами на несколько мелких, объемом не больше 2 тыс. символов, чтобы верификатор мог проверить целую часть за один раз, не создавая ее по ходу работы, а делая это только один раз, в конце.

Как развить большую базу на маленькие:

19. В меню Edit выбрать команду Split Base.

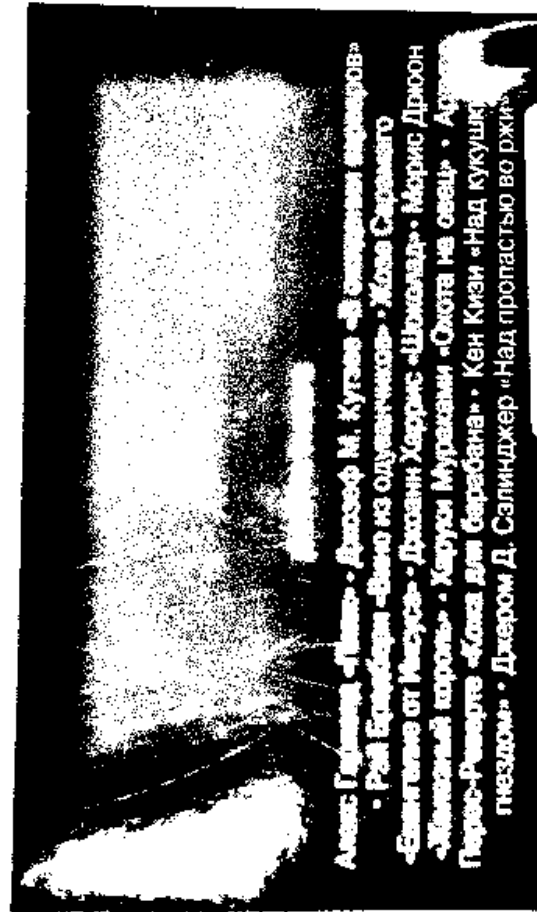
Базу нужно поделить на куски в среднем по 2000 записей. В зависимости от этого рассчитываем количество, в котором будем делить базу. Если база размером около 30 тыс. записей, то делить ее нужно на 15 частей. Примерно настроен диалог SplitBase:



Средства по этому разделу выделяет банк на цели в соответствии с заданными приоритетами. Если банк пожелает финансировать на 15% - 20%, то нужно продать 14% акций.

1 2.14
2 2.13
3 2.12

Иэн Макьюэн (Ian McEwan) — один из самых знаменитых современных британских писателей. Был первым учеником литературного курса, основанного в семидесятых замечательными английскими прозаиками Энгулом Уиндсомлом и Малькольмом Брэдбери. За дебютную книгу «Первая любовь, последнее помазание» получил премию имени Somerset Моэма. Сейчас Иэн Макьюэн — один из «правящего триумvirата» современной британской прозы наряду с Мартином Эммонсом и Джулианом Барнсом, а также с другими выдающимися писателями «Антверпула».



интернет-магазин

OZON.RU



75739806





Charlie Ruby

and live the *loft* life. It was in a munty of sympathetic artists all ing in the same space. It *seemed* her wonderful. But then I realised I don't move in. It's perfect for a *chelor* but not if you're going to e with someone. If you're in a ationship you have to be able to hm doors. Or in this house I can go *another floor*. But in an open plan ice it *would have been a* disaster, so, I've got a *lot* of stuff. I read [a 'ce by] Daniel Liebeskind's wife, 10 wrote something about niwalism. And, as I'r supposedly * *master of minimalism*, it sparked ' interest. She said: "The art of aimalism is cupboard s." But there *minhoard* space fm the

Do you have any houses overseas?

I've got a house in Gascogne, France. I bought it in 1989. It's a stone farmhouse with converted barns, a huge amount of grassland and a swimming pool. It's great for family holidays. If you half close your eyes you can imagine you're in Tuscany but if you open them you know you're not which is sad. I prefer Italy or Spain.

So you haven't found your favourite home yet?

My favourite home might be a case of wishful thinking. My ideal place would be on the sea I stayed in a hotel on the beach in San Sebastian, northern Spain, and thought I could live there - or maybe the borderland of Barcelona, to the right of the Ramblas. It is a very curious mixture of immigrant barber shops and fancy hotels. I can imagine living there and possibly will. You're only one minute away from La Boqueria, the best food market in the world. I don't swim or sail or like sitting on beaches but I like the idea of open expanses. I like to take pictures in that area.

You are a keen photographer and your house is full of photos. Who do you collect?

I've got an Italian publishing company who are about to publish a book of my photos. I'm scared of publishing them. It's like writing a piece of music and not having the nerve to play it. I've been shooting videos over a number of years. I recently had a one-day exhibition in a gallery in Diisseldorf. I liked the brevity of the exhibition. I have prints by Mike Disfarmer, who lived and worked in Heber Springs, Arkansas, in the 1930s and 1940s as well as Miroslav Tichy, who lives in a little town near Kynov, Czech Republic. Their prints face each other in my sitting room. They have a lot in common and a lot not in common. Both are small-town photographers who took pictures of people. Disfarmer took photos in controlled studio settings of people in uniforms or Sunday best. Tichy was the opposite and was an outsider who had been imprisoned by communists. He made his own cameras and took pictures almost exclusively of women. They are insolent photos though not obscene. My aim is to make some kind of musical piece bringing both - Wnoraners together. I find the





would like to congratulate the Global Mobile Awards

»P

MWOMEN BEST MOBILE PRODUCT OR SERVICE FOR WOMEN IN EMERGING MARKETS

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

BE: Not DE: San

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

BE: App BE: AT&

THE GREEN MOBILE AWARD

Flexenclosure - E-site

BE: KT

BEST MOBILE HEALTH INNOVATION

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

BE: Alca

BEST MOBILE INNOVATION FOR EDUCATION OR LEARNING

OnPoint Digital - CellCast Solution

BE: Vod patl

BEST MOBILE INNOVATION FOR AUTOMOTIVE, TRANSPORT OR UTILITIES

Ford Motor Company Ltd. - Ford SYNC with Emergency Assistance

BE: App

BEST MOBILE MONEY INNOVATION

Etisalat, MasterCard and Oberthur Technologies - Etisalat Commerce

BE: SAI Ctor

BEST MOBILE INNOVATION FOR PUBLISHING

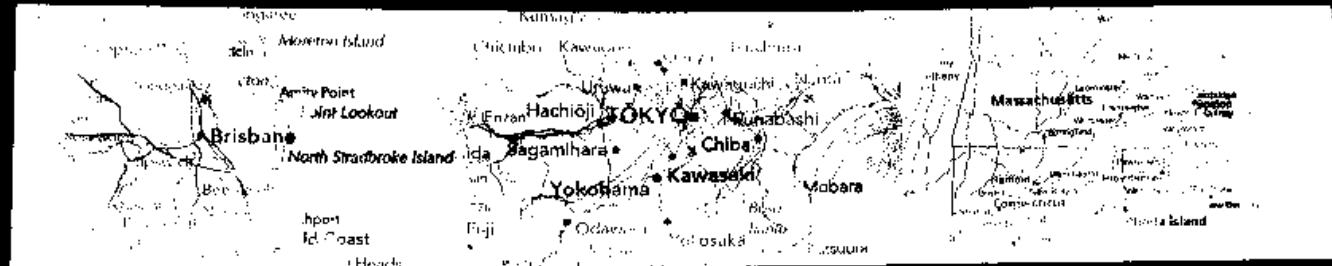
Financial Times and Assanka - The Financial Times Web

R GSI KDI NTT SOF

...

C mobile COVERAGE

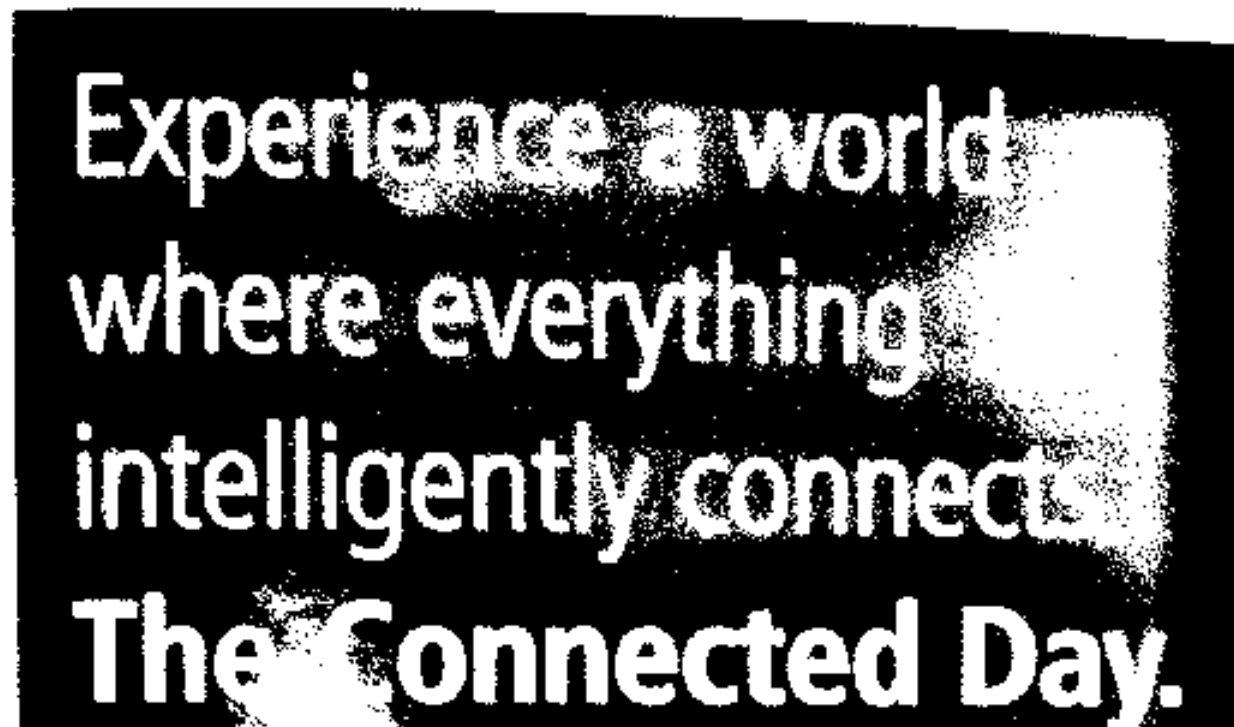
A NETWORK ROAMING COVERAGE MAP SERVICE FOR YOUR VISIT



COME AND VISIT US AT THE GSMA PAVILION IN HALL 8

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 seamless network coverage maps are delivered straight to network operators' corporate web pages to help them show visitors where they can use their phones when abroad.

Interactive seamless network roaming coverage maps delivered straight to network operator corporate web pages



C
to
ea
an
th
via

C
A
I
t



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, supporting 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

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 services, 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	China Telecom	
Connections (m)	648.7	199.7	125.3	973.7
Connections, 3G* (m)	51.8	40	125.3	217.1
% 3G*	8%	20%	100%	22%
Market Share	67%	21%	13%	-
Market Share, 3G*	24%	18%	58%	-
Net Additions (m)	15.2	10.6	8.4	34.2
Net Additions, 3G* (m)	8.7	9.8	8.4	26.8
Growth, YoY	11%	19%	38%	16%

Source: Wireless Intelligence • *Includes CDMA2000 1X connections as per ITU classification

lling
the
oms
ch is
IA at
and
both
DMA
s 3G

y of
MIIT)
s had
tage”
rators
\$14.9
nine
nistry

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

Unicom continued to increase its 3G market share throughout 2011 by offering considerable subsidies on its WCDMA handsets, and last month launched a three-year contract plan offering a free iPhone 4S for as little as CNY286 per month. However, the inherent risk in this strategy is borne out by the operator's most

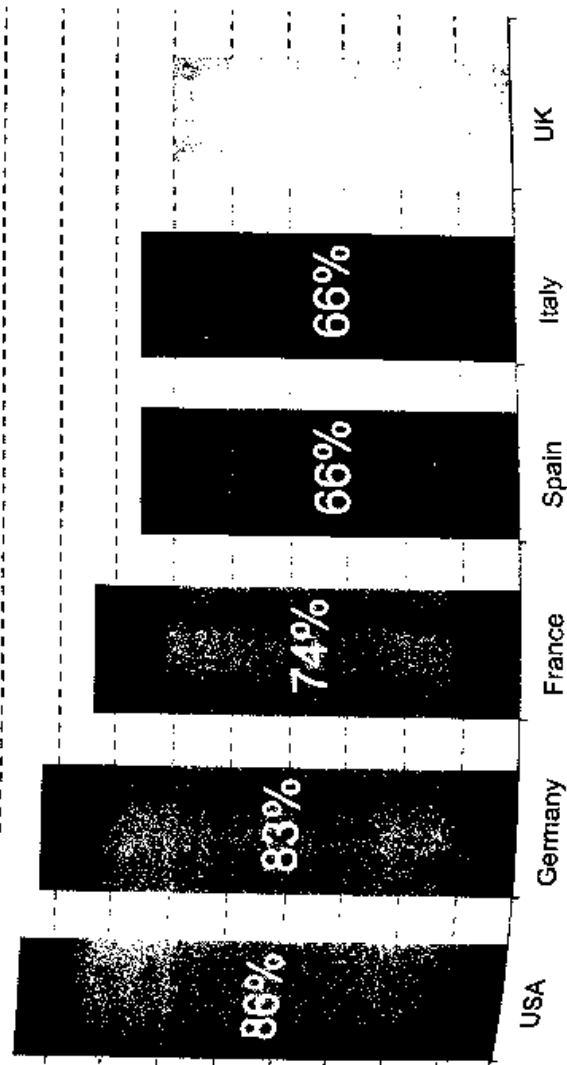
These are the questions we believe will be addressed by mass market rich media campaigns running on mobile. The term 'rich media' is quickly entering the mobile advertising lexicon and is certainly creating a lot of hype. Although potentially quite broad in definition, it generally refers to ads that 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. ORMA (Open Rich Media Advertising), an industry wide open initiative for advertisers, has been standardised by the IAB Mobile Marketing Center of Excellence to create MRAID (Mobile Rich Media Ad Interface Definition) which defines a common API for mobile rich media ads that will run in mobile apps and web sites. This effectively provides the industry standard required for rich media to move mainstream.

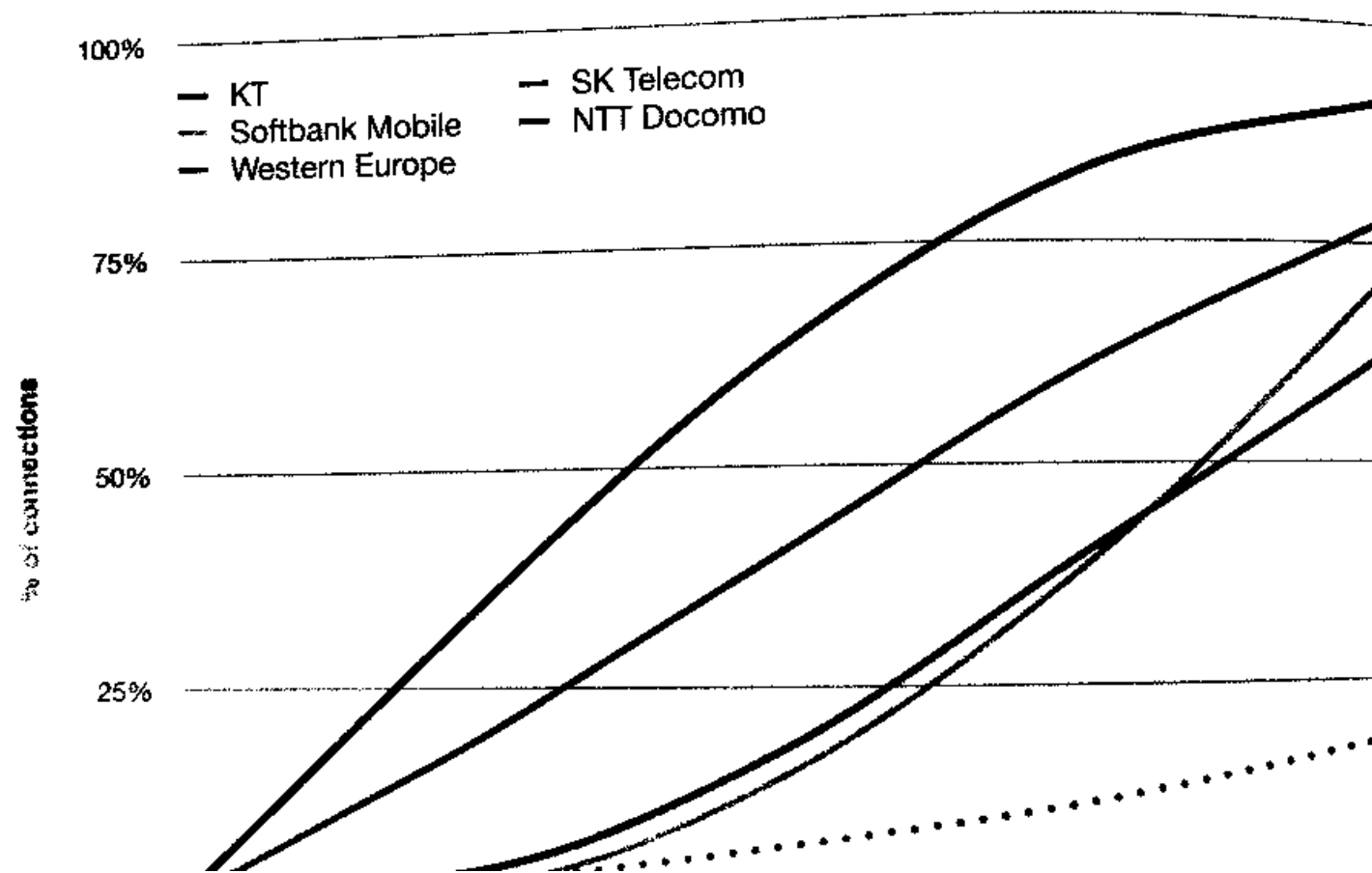
Third, from a creative perspective, rich media gives advertisers complete design flexibility to deliver messages and experiences. Expandable banners are the most common rich media format, whereby consumers tap on a banner and get presented with countless options and experiences such as interactive mobile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



...y's
...rs.
...to
...are

TE-
roof
the
ble
1980
1981





EUROPE

ASIA-PACIFIC



CONNECTIONS
(million)

IN THIS ISSUE

UK operator
hails the Spotify
effect

UK SAYS MUSIC
SERVICE IS DRIVING
REVENUE PAGE 4

Alcatel jumps on
Android
bandwagon

ANOTHER HANDSET
MAKER ON BOARD THE
GOOGLE BUS PAGE 8

Verizon Wireless
signals new
approach to
VoIP threat

US MARKET LEADER IN
GROUND-BREAKING DEAL
WITH SKYPE PAGE 13

LiMo aligns
itself with apps
fightback

RIVAL, ANDROID OS PREPS
LAUNCH OF LATEST
VERSION OF LINUX
PLATFORM PAGE 13



Telstra CTO proposes alternative approach to app stores

by Andrew O'Connell

A SENIOR EXECUTIVE of Australia's largest operator put forward an alternative proposal at yesterday's conference keynote session dedicated to mobile applications. Dr Hugh Bradlow, CTO of Telstra, said that customer demands would force the company to support numerous app stores and handset operating systems (OS), regardless of the various industry initiatives to reduce fragmentation.

"For example, Apple and Samsung's bada won't go away, and we'll have to work with them," said Bradlow. "We don't see the growing number of available app stores or mobile operating systems as fragmentation, but as competition. Before there will still be at least six to 10 mobile operating system offerings on the market in the coming years."

The problem with supporting numerous operating systems, according to Bradlow, falls away

as the operators develop common ways by forcing them into testing their software on up to 100 different handsets to ensure compatibility. "Developers are still struggling with the handset environment."

Turning to how Telstra would offer applications to its customer base, Bradlow said the company planned to position itself as a "shopping centre" where subscribers browse through a selection of store fronts and select apps which are supported on their handset. "We will build the shopping centre environment, which means we won't be bypassed in the value chain."

Bradlow's comments are topical in light of Monday's announcement whereby at least 21 operators around the world are to join forces to launch an open international applications platform, marking the largest unified move to date by the operator community into the mobile apps space. Osim is not believed to have signed up to the Wholesale Applications Community so far. (Cont. on P14)

Huawei claims LTE device world-first



Huawei is set to pre-empt a market in the device space by yesterday unveiling what it claims the world's first triple-LTE modem. The Chinese vendor said the P9 smartphone is backward compatible with 3G, 2G and 2G+ GSM technology. (Cont. on P14)

BBC pushes mobile app strategy

by Andrew O'Connell

UK BROADCASTER BBC has set to launch two applications in the next month that will make it easier for smartphone users in the country to access the broadcaster's content on their handsets, said Erik Haggard, the BBC's director of creative media and technology, at yesterday's keynote conference session on mobile content, content

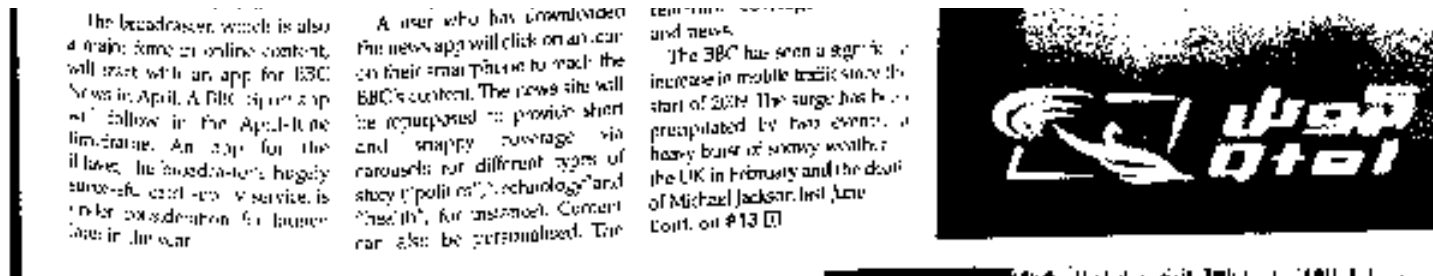
The broadcaster will develop the apps first for Apple's iPhone followed soon after for Blackberry and Android-based handsets. It also intends to develop versions of the apps for other smartphone operating systems. Yesterday Mobile 22 announced it is developing the BBC's iPhone apps, claiming to have won a competitive pitch involving more than twenty mobile application vendors.



sports and will be talking focused on the BBC's popular football coverage and its extensive coverage of results

Join us in Zone 3

Join us in Zone 3, the new mobile app for the BBC's popular football coverage and its extensive coverage of results.



Время работы

Время Работы кода на C++:

```
image: Dataset/01.JPG - sec/megapixel: 0.180898
image: Dataset/02.JPG - sec/megapixel: 0.181851
image: Dataset/03.JPG - sec/megapixel: 0.182748
image: Dataset/04.JPG - sec/megapixel: 0.180665
image: Dataset/05.JPG - sec/megapixel: 0.182378
image: Dataset/06.JPG - sec/megapixel: 0.180547
image: Dataset/07.JPG - sec/megapixel: 0.180336
image: Dataset/08.JPG - sec/megapixel: 0.184976
image: Dataset/09.JPG - sec/megapixel: 0.182872
image: Dataset/10.JPG - sec/megapixel: 0.189801
```

```
image: Dataset/11.JPG - sec/megapixel: 0.183123
image: Dataset/12.JPG - sec/megapixel: 0.182579
image: Dataset/13.JPG - sec/megapixel: 0.189904
image: Dataset/14.JPG - sec/megapixel: 0.182623
image: Dataset/15.JPG - sec/megapixel: 0.180961
```

```
In [41]: times = np.array([0.181851, 0.182748, 0.180665, 0.182378, 0.182378, 0.180547, 0.180336,
                           0.183123, 0.182579, 0.189904, 0.182623, 0.182623])
```

```
In [42]: len(times)
```

```
Out[42]: 15
```

```
In [43]: np.mean(times)
```

```
Out[43]: 0.18329359999999997
```

Среднее время работы - 0.1833 сек/мегапиксель, что во много раз быстрее, чем на питоне. (Код на питоне обрабатывал 15 фото 24 минуты, то есть больше минуты на фото, а если учесть, что в тестовых фото примерно 3 мегапикселя, значит скорость питона примерно 20 сек/мегапиксель.)

Результаты.

Практически на всех изображениях осталась большая часть текста, которую может прочесть человек (отлично распознаны изображения 1, 5, 7, 10, 11, 12, 13, 14). Но где-то из-за бликов (изображения 4, 6, 8, 9, 15) и других проблем со светом (изображения 2, 3) пропала часть текста, что можно настроить в конкретных случаях разным порогом `t_gros`, но не совсем понятно, как это делать автоматически. Хуже всего были обработаны фото с большим количеством бликов и низкоконтрастные изображения. Также, в некоторых частях изображения текст не черный на белом, а белый на черном, что наверно можно инвентировать, если уметь находить контур текстового блока и смотреть на его фон. Если он черный, то инвентировать цвета внутри.

Достоинствами метода Бредли, как отмечалось в самом начале, является простота реализации и высокая скорость выполнения, а также для большинства случаев нет нужды подбирать параметры. Кроме того, метод хорошо работает с неоднородным фоном (например, 4, 6, 8, 14, 15).

Таким образом, метод Бредли представляет из себя отличный метод биноризации изображения без какого-либо машинного обучения.

In []: