



创享未来 2016 微软开发者峰会



微软认知服务

李京梅

微软中国云计算与企业事业部
首席产品经理

认知服务

让你的应用拥有人的智慧

基于机器学习的人工智能使自然的和基于语境的人机交互变为可能

在你的应用中接入智能，把你的想法变成现实



理解您应用中的数据

认知服务的API帮助您理解并与包括语音文字，图片，及视频数据进行交互



强大的模型

让我们做你的最强大脑

认知服务的模型充分利用了构建微软多项产品线的同样的深度学习与机器学习技术



轻松使用，触手可以

认知服务让您可以跨设备，跨平台轻松调用这些REST API及SDK，让你更加专注你的应用



微软认知服务概览

视觉	语音	语言	知识	搜索
计算机视觉API	自定义智能语音识别服务	必应拼写检查API	学术搜索API	必应网络搜索API
情绪识别API	声纹识别API	语言理解智能服务	实体链接智能服务	必应视频搜索API
人脸识别API	语音识别API	文本分析API	推荐API	必应图片搜索API
视频检测API		网络级语言模型API	知识搜索服务	必应新闻搜索API
		语言分析API		必应自动建议API

视觉



计算机视觉 API

分析图像

基于所输入图像的视觉内容分析出图像的视觉特征，如2000+物体标签及描述功能

识别名人

识别自全球各地涉及商界、政界、体育、娱乐界的名人

读取图片中的文字 (OCR)

在图像中检测并识别文本

生成缩略图

生成高品质和高存储效率的缩略图



分析图像

图片类型：

Clip Art Type	0 Non-clipart
Line Drawing Type	0 Non-Line Drawing
Black & White Image	False

图片内容：

Adult Content	False
Adult Score	0.18533889949321747
Faces	[{"age": 27, "gender": "Male", "faceRectangle": {"left": 472, "top": 258, "width": 199, "height": 199}}]
Tags	[{"name": "water", "confidence": 0.9996442794799805}, {"name": "sport", "confidence": 0.9504992365837097}, {"name": "swimming", "confidence": 0.9062818288803101, "hint": "sport"}, {"name": "pool", "confidence": 0.8787588477134705}, {"name": "water sport", "confidence": 0.631849467754364, "hint": "sport"}]

图片色值：

Dominant Color Background	White
Dominant Color Foreground	Grey
Dominant Colors	White
Accent Color	





识别人人

图片信息：

requestId: "b76c3064-6012-490b-af73-dad660ac42b0",
metadata: "height": 533,
"width": 800,
"format": "Jpeg"

图片名人：

name: "Satya Nadella",
faceRectangle: "width": 134,
"height": 134,
"left": 401,
"top": 103

置信度：

confidence: 0.9576959



CaptionBot.ai 视图机器人

I think it's a woman sitting at a table eating a meal and she seems 😊.

(我觉得是一位女士在桌旁就餐而且她看上去很高兴)





读取图片中的文字

■ 已支持功能:

- 自动识别多语言
- 扫描图片
- 含有文本的图片
- 纹理致密的位置信息

■ 有待增强功能

- 机动车牌照
- 手写文本
- 超大文本识别



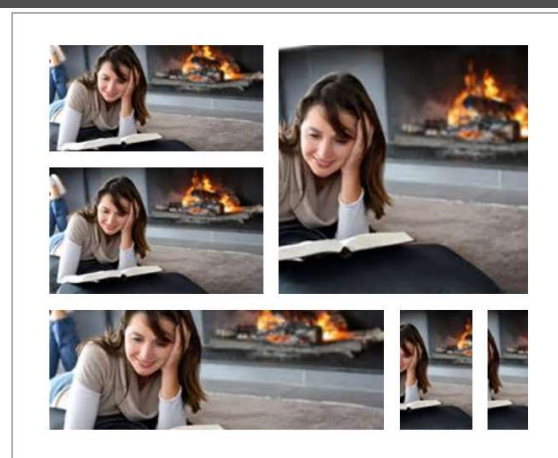
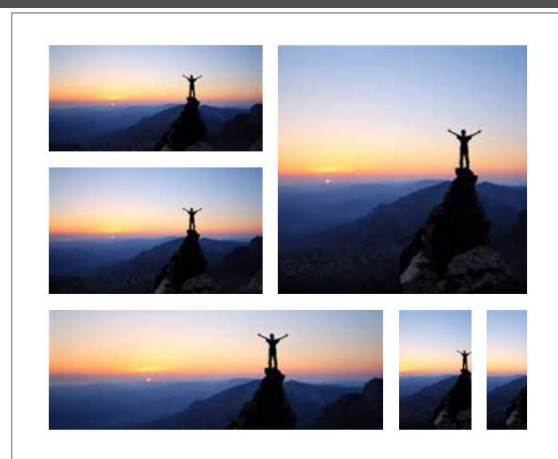
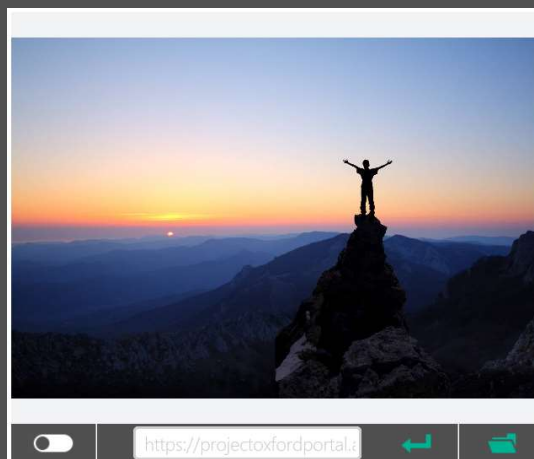
“我从未见过如此卓越的产品。计算机视觉API使我们可以向我们的客户证明数据的可靠性，因此他们可以有信心基于该信息来做出重要的商业决策。”

— Leendert de Voogd, CEO, Vigiglobe



智能缩略图

开启智能缩略功能





人脸识别 API

人脸检测

检测图片中的人脸并识别属性

人脸验证

核对两张人脸是否属于同一个人

相似人脸搜索

从多张人脸中找出与所查人脸相似的人脸

人脸分组

根据识别的人脸进行分组

人脸辨识

查询某张人脸，与提供的数据中的何人相匹配

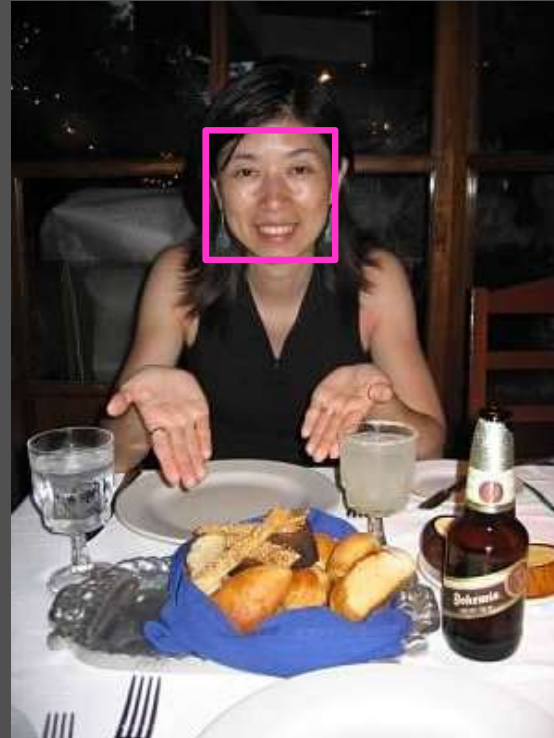
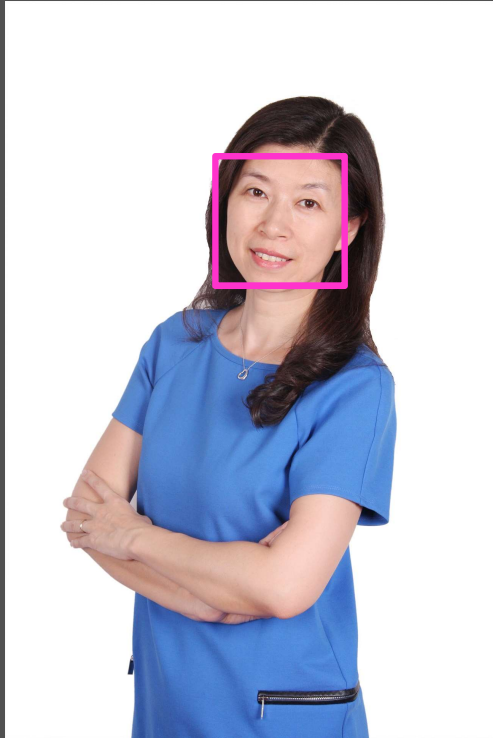


人脸识别 API

```
"faceAttributes": {  
  "age": 43.4,  
  "gender": "male",  
  "headPose": {  
    "roll": 16.5, "yaw": 18.3, "pitch": 0  
  },  
  "smile": 0.009,  
  "facialHair": {  
    "moustache": 0.9,  
    "beard": 0.8,  
    "sideburns": 0.7 },  
  "glasses": "ReadingGlasses" }
```



人脸验证 – 我还是原来的我^_^!



验证结果：

The two faces belong to the same person.

Confidence is 0.803.



情绪识别 API

识别图片中的人脸情绪

基于面部表情来检测感情

识别视频中的人脸情绪

视频情感识别API可识别视频中人脸的表情，并返回不同表情的出现次数的计数总结



情绪识别 API

人脸检测

```
"faceRectangle": {"width": 193, "height": 193, "left": 326, "top": 204}
```

...

情感指数

```
"scores": { "anger": 5.182241e-8,  
  "contempt": 0.0000242813,  
  "disgust": 5.621025e-7,  
  "fear": 0.00115027453,  
  "happiness": 1.06114619e-8,  
  "neutral": 0.003540177,  
  "sadness": 9.30888746e-7,  
  "surprise": 0.9952837}
```



视频检测 API

稳定处理

使抖动的视频变得平滑和稳定

人脸检测和追踪

检测并追踪视频中的人脸

运动检测

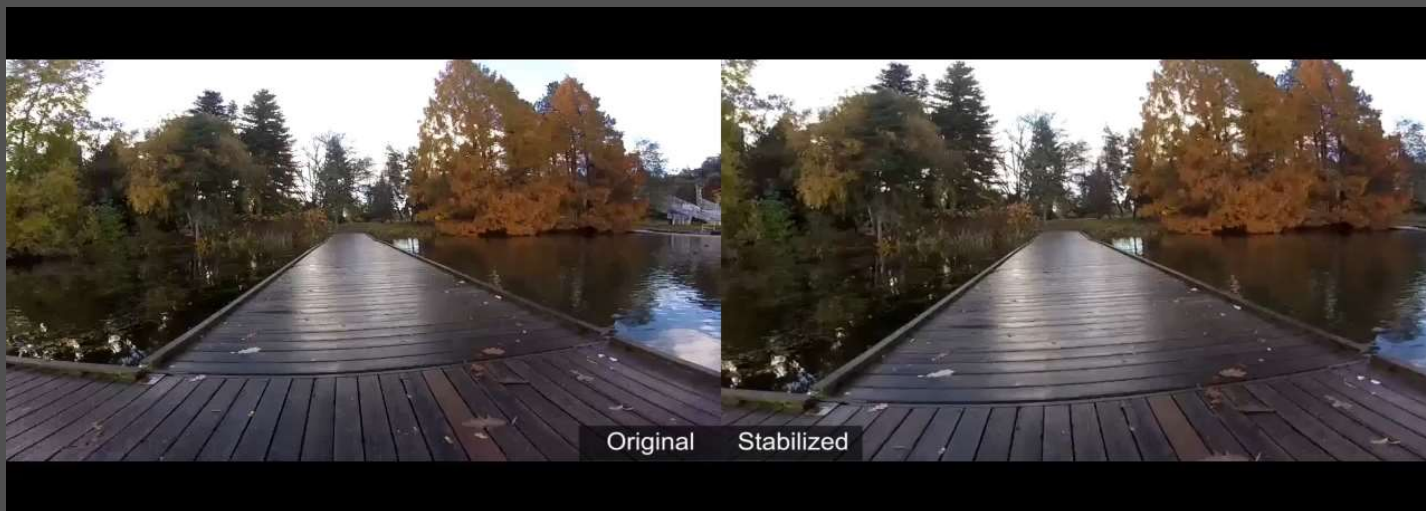
检测视频中发生的运动变化

视频缩略图

从视频中选取一组场景来创建短视频预览

稳定处理

- 提供自动处理，让抖动的视频变得平滑和稳定
- 使用了许多和微软Hyperlapse相同的技术
- 适用于：小幅度的相机移动，无论是否有滚动快门效果（例如，手持静态相机，缓慢行走）





人脸检测及追踪

- 高精度的人脸位置检测和追踪
- 能够检测出视频中64个人脸（不小于24x24像素）
- 返回整个视频中被检测和追踪的人脸坐标和Face ID



Time (sec)	Face ID	x, y	Width, Height
0	0	0.59, 0.23	0.09, 0.16
0	1	0.38, 0.15	0.07, 0.12
1	0	0.54, 0.25	0.09, 0.15
1	1	0.23, 0.18	0.07, 0.12

运动检测

- 当检测视频中以静态背景为参照发生运动变化时发出指示（例如监控视频）
- 训练模型以减少错误警报，例如光线和阴影变化
- 当前限制：
 - 暂不支持夜间视频检测
 - 半透明和小型物体检测效果不佳



Start Time	End Time	In Region
1.9	3.6	0
5.2	15.1	0

视频缩略图

- 提供自动的动作缩略图概要，让用户可以快速看到视频预览或快照。
- 从视频中选取一组场景来创建短视频预览。



12分钟的视频可以做成30秒的短视频
预览

语音

语音 API

语音识别 (语音转文本)

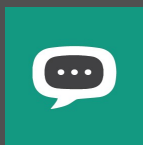
将语音转换为文本

语音输出(文本转语音)

将文本合成语音

语音意图识别

将语音转换为意图



语音识别

	REST 服务	客户端SDK
支持平台	任意	Windows, Android, iOS
数据支持	支持	支持
麦克风支持	不支持	不支持
麦克风沉默识别	不支持	支持
音频长度	短	短与长
响应数结果数量	n-best模式	多个部分结果，较短语音n-best模式以及较长语音的多短语模式

与微软的Xbox , Windows , Cortana, Skype Translator采用相同技术

语音识别

	短模式	长模式
音频时长	< 15秒	< 2分钟
最终结果	n-best模式	Best模式并在语音结束时输出
部分结果	支持	支持

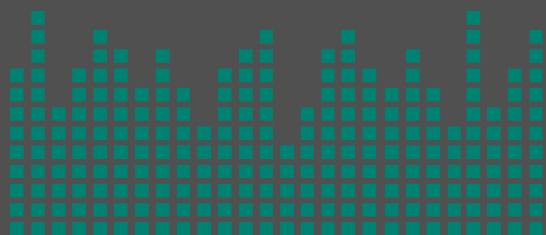


450 6th St.
San
Francisco

***** 最终 N-BEST 结果 *****

[0] Confidence=Normal Text="450 six St San Francisco."
[1] Confidence=Normal Text="For 50 six St San Francisco."
[2] Confidence=Normal Text="456th St San Francisco."
[3] Confidence=Normal Text="450 six St in San Francisco."
[4] Confidence=Normal Text="456 St San Francisco."

语音输出



通过POST请求将文本合成为语音


最长支持15秒

支持17种语言

```
<speak version="1.0"
xmlns="http://www.w3.org/2001/10/synthesis"
xmlns:mstts="http://www.w3.org/2001/mstts"
xml:lang="zh-CN">
<voice name="Microsoft Server Speech Text to Speech
Voice (zh-CN, HuihuiRUS)">
```

由文本合成语音并传达给你的用户听

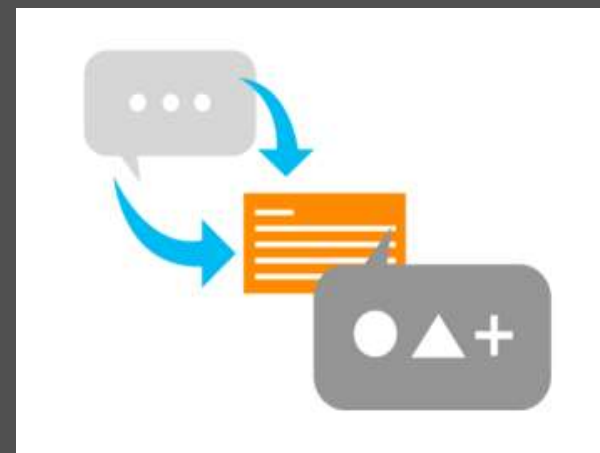
```
</voice></speak>
```



由文本合成为语音
并传达给你的用户
听

语音意图识别

- 将语音转换为意图
- 与语音转换为文本相似，区别在于语音意图识别功能除了可以从语音输入返回识别的文字，服务器还可以返回关于语音输入的结构化信息，这样应用程序便可以轻松地解析说话者的意图，以驱动下一步操作
- 用于意图识别的训练模型是由微软 LUIS 服务提供的



450 6th St.
San Francisco

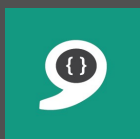
语言



语言理解智能服务

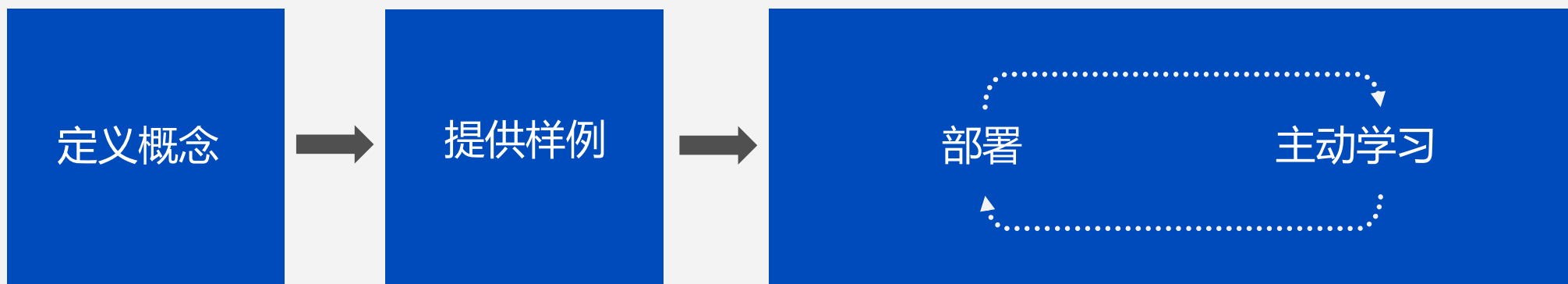
理解用户所说的话

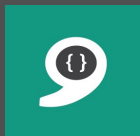
使用来自Bing和Cortana的预构的模型或者自己创建的模型



语言理解智能服务

- 用交互的特征来减少贴标签的工作
- 使用可视化来测量和提升性能
- 与语音识别服务的无缝集成
- 几个样例就足以搭建自己的应用，并自主学习





语言理解智能服务

“关于航班延误的新闻”

```
{
  "entities": [
    {
      "entity": "flight_delays",
      "type": "Topic"
    }
  ],
  "intents": [
    {
      "intent": "FindNews",
      "score": 0.99853384
    },
    {
      "intent": "None",
      "score": 0.07289317
    },
    {
      "intent": "ReadNews",
      "score": 0.0167122427
    },
    {
      "intent": "ShareNews",
      "score": 1.0919299E-06
    }
  ]
}
```


ExerciseTracker

New utterances

Search

Suggest

Review labels

 Publish


Intents 

None

StartActivity

StopActivity

SetHeartRateTarget

Entities 

ActivityType

Pre-built Entities 

number

Regex Features 

No patterns added

Phrase List Features 

ActivityWords 

Suggest utterances that will improve:

Intent: StartActivity

Select highlight to add another entity or click to clear

begin a  now

StartActivity(0.96)

Submit

begin  bike ride

StartActivity(0.94)

Submit

start tracking a  bike ride

StartActivity(0.94)

Submit

Performance analysis

Intents

StartActivity

11 utterances: 11 correctly predicted

StopActivity

5 utterances: 5 correctly predicted

SetHeartRateTarget

5 utterances: 5 correctly predicted

None

17 utterances: 17 correctly predicted

 Correctly predicted

 Error (predicted as other intent)

案例分享

应用案例展示



Zero Keyboard |  Computer Vision API  Speech API
Enable your sales force to quickly record customer information—without typing



FaceHero.co |  Emotion API  Face API
Emoji your face automatically or selecting your favorite expression



Fetch! |  Computer Vision API  Face API
Recognize and classify dogs by breed, or upload a selfie to see what dog breed you are



Mimicker Alarm |  Computer Vision API  Emotion API  Speech API
Wake up and stay up with artificial intelligence



Facial Recognition Door |  Face API
Unlock a door with facial recognition and this IOT hack




Bing.com Image Search |  Bing Image Search API  Face API
See the tech behind Microsoft Cognitive Services in action



WIZR |  Computer Vision API
Analyze pictures on social media and gain valuable, real-time insights for customers



MyMoustache.net |  Face API
Get 'stache-stats with #MyMoustacheRobot and help raise awareness for men's health



How-old.net |  Face API
How old do you look? Ask #HowOldRobot



TwinsOrNot.net |  Face API
Are you twins? Ask #TwinsOrNotRobot to find out for yourself



FAC3 |  Computer Vision API  Emotion API
A generative artwork which turns real emotions into emojis

演示- 模仿者闹铃



Mimicker Alarm (模仿者闹铃) | Microsoft |  计算机视觉API  情绪识别API

 语音识别API

让人工智能叫醒你并帮助你保持清醒

64% 11:49

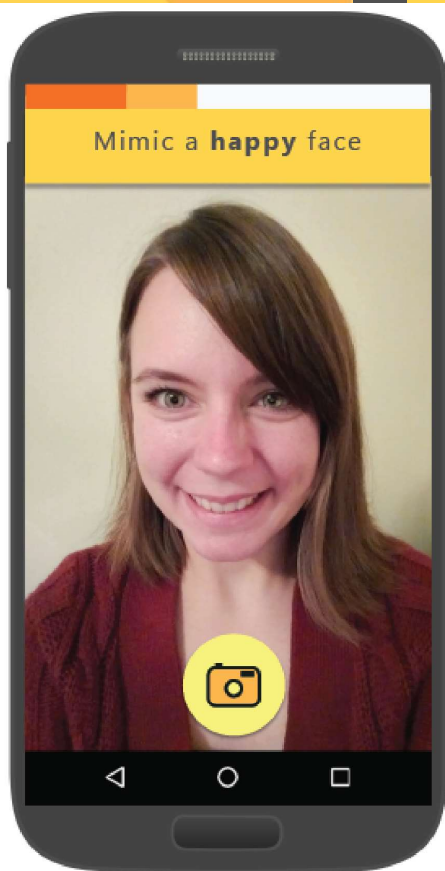
Alarms

09:46

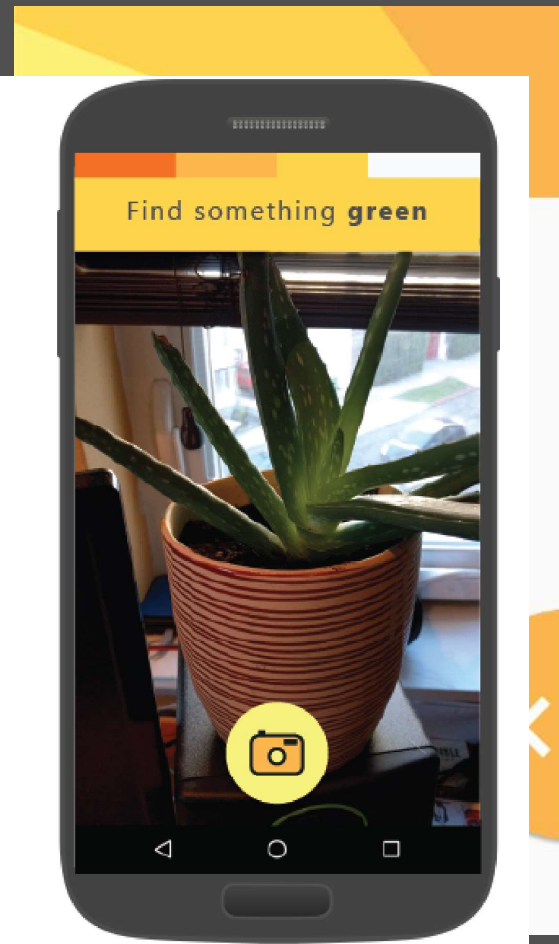
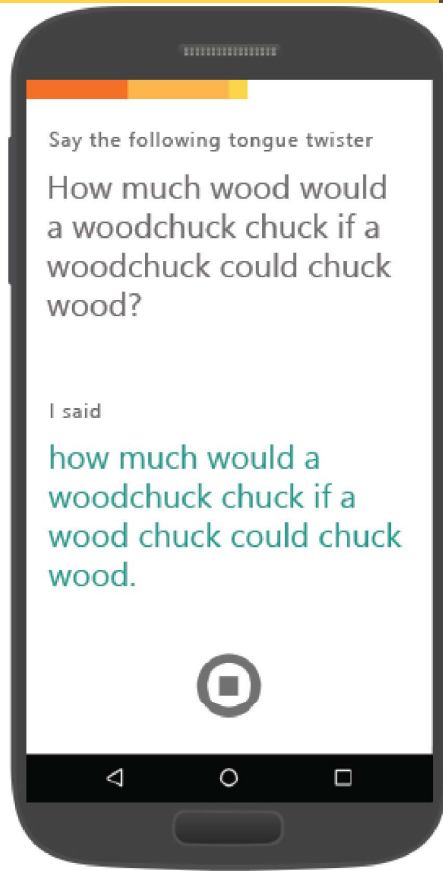
Mimicker Alarm

22:56

Mimicker Alarm



64% 11:49




样例代码

```
((HttpPost) request).setHeader("ocp-apim-subscription-key", this.subscriptionKey);
```

```
public AnalyzeResult analyzeImage(InputStream stream, String[] visualFeatures) throws VisionServiceException, IOException {
    HashMap params = new HashMap();
    String features = TextUtils.join(",", visualFeatures);
    params.put("visualFeatures", features);
    String path = "https://api.projectoxford.ai/vision/v1/analyses";
    String uri = WebServiceRequest.getUrl(path, params);
    params.clear();
    byte[] data = IOUtils.toByteArray(stream);
    params.put("data", data);
    String json = (String) this.restCall.request(uri, "POST", params, "application/octet-stream", false);
    AnalyzeResult visualFeature = (AnalyzeResult) this.gson.fromJson(json, AnalyzeResult.class);
    return visualFeature;
}
```

视频



Facial Recognition Door (会识脸的门) | Windows IoT team |  人脸识别API
使用面部识别和此IoT骇客来解锁一扇门

Dynamics CRM - Social Hub 应云而生

产品背景

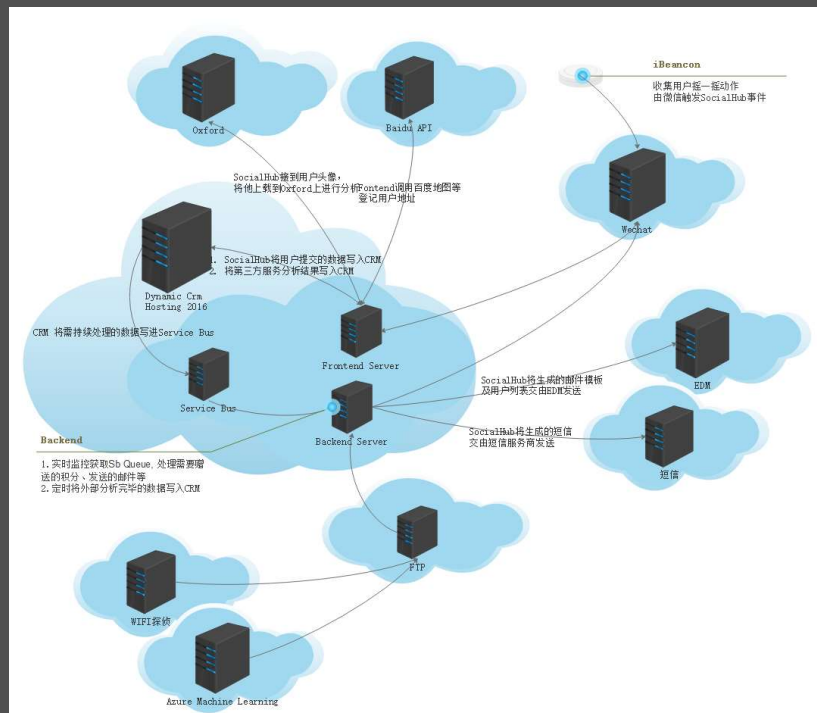
- 1、CRM 产业向Social CRM 演化
- 2、中国拥有独特的Social 环境
- 3、大量的PaaS 服务成熟催生营销场景创新

现状

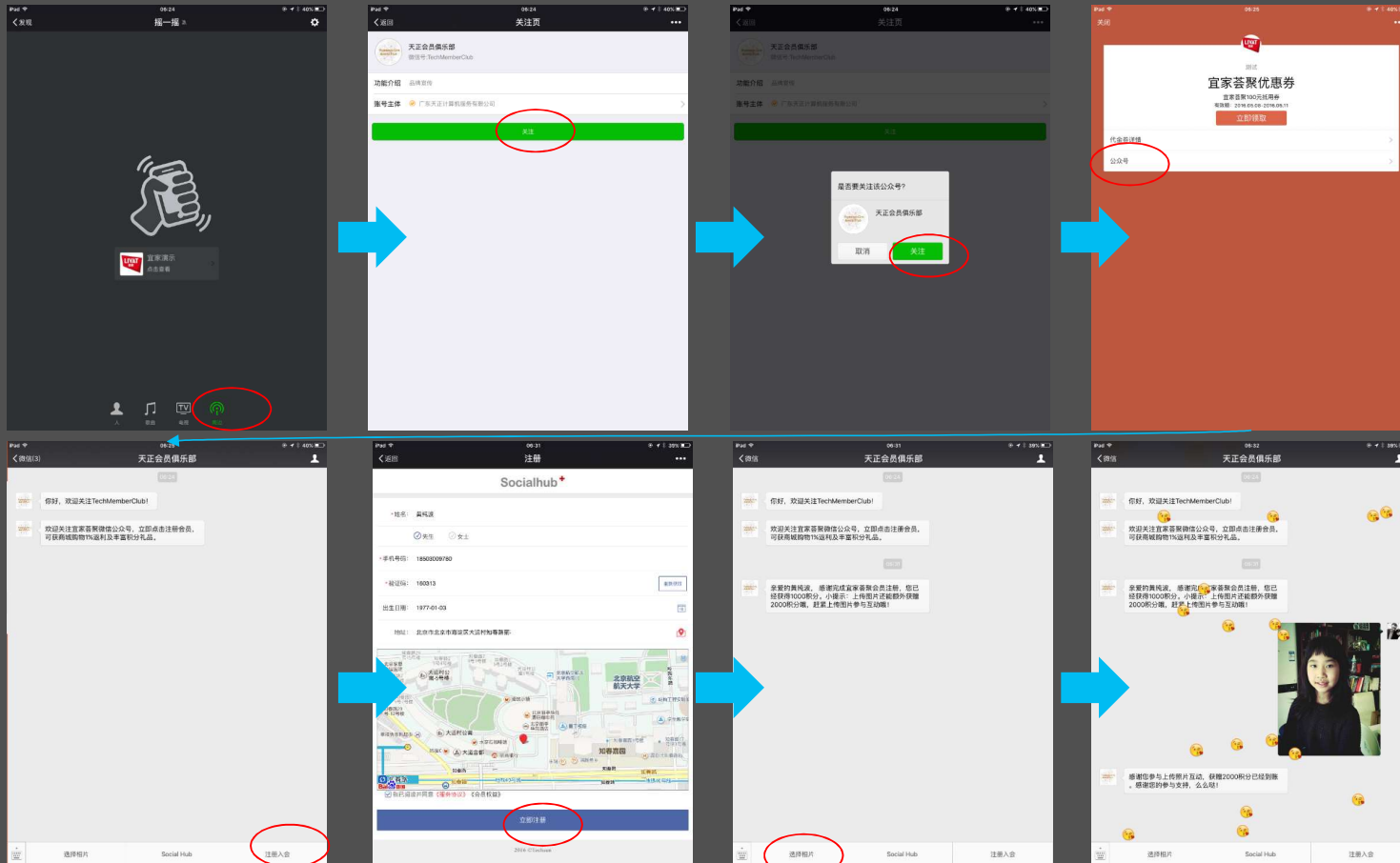
- 1、已经接入15+企业级的PaaS服务/互联网API ,年底将达50+
- 2、众多重量级企业开始应用：Adidas、Microsoft、国家大剧院.....
- 3、获Azure云伴计划SaaS创业大赛一等奖

认知服务 & IoT

- 1、人脸的识别与分析是线下营销场景构建/消费者视图构建的重要环节；
- 2、IoT 让消费者与企业在线下的互动/营销/行为分析成为可能



Social Hub IoT&认知服务社交化营销场景



Social Hub IoT&认知服务社交化营销场景

Microsoft Dynamics CRM | 会员管理 | 联系人 | ct2016051000006 | RAdmin LoyaltyOnline

保存 保存并关闭 新建 添加到市场营销列表 分派 删除 窗体 共享 追随

联系人
ct2016051000006

行为分析

行为

名称	行为类型	联系人	传感器
黄纯波在2016年05月10日 06:24:...	微信摇一摇	ct2016051000006	5781018

照片分析

名称	联系人	性别	年龄
黄纯波在2016年05月10日 06:31:49进行了一次照片...	ct2016051000...	女	

创建者: RAdmin | 创建时间: 2016/5/10 6:24 | 修改者: RAdmin | 修改时间: 2016/5/10 6:31

可用

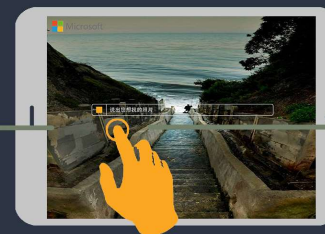
案例分享- 编程之美 大赛获奖作品

- Percepicture智能相框
- 学术搜索超新星

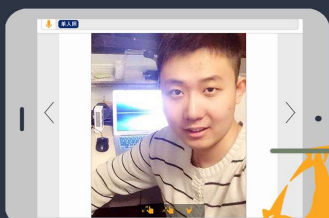
PERCEPICTURE PROTOTYPE



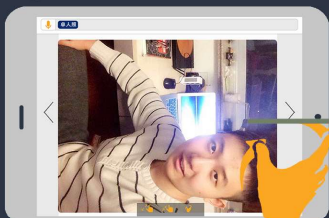
1 进入登陆页面，点选手势点
选语音按钮可以开始说话



2 开始语音后，原语音按钮变
为停止键，点选手势即结束
语音



5 收放手势可放大缩小图片



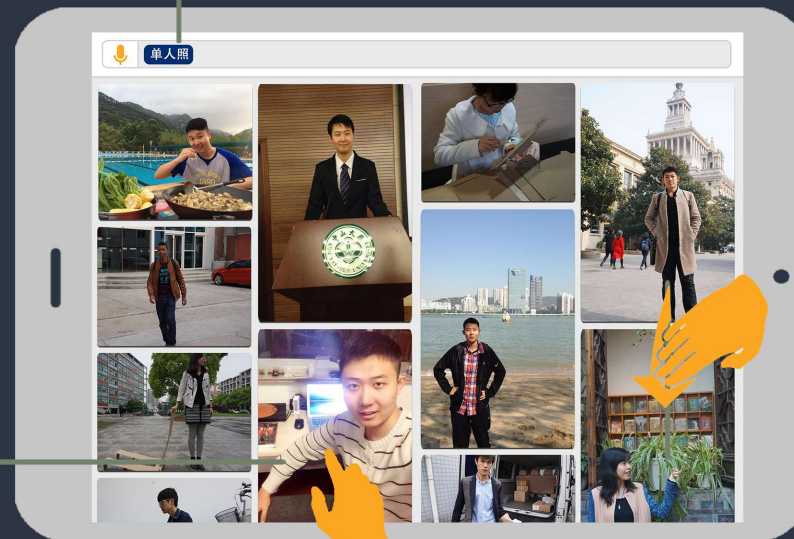
6 旋转手势可将图片进行旋转



4 点选手势可放大查看该图

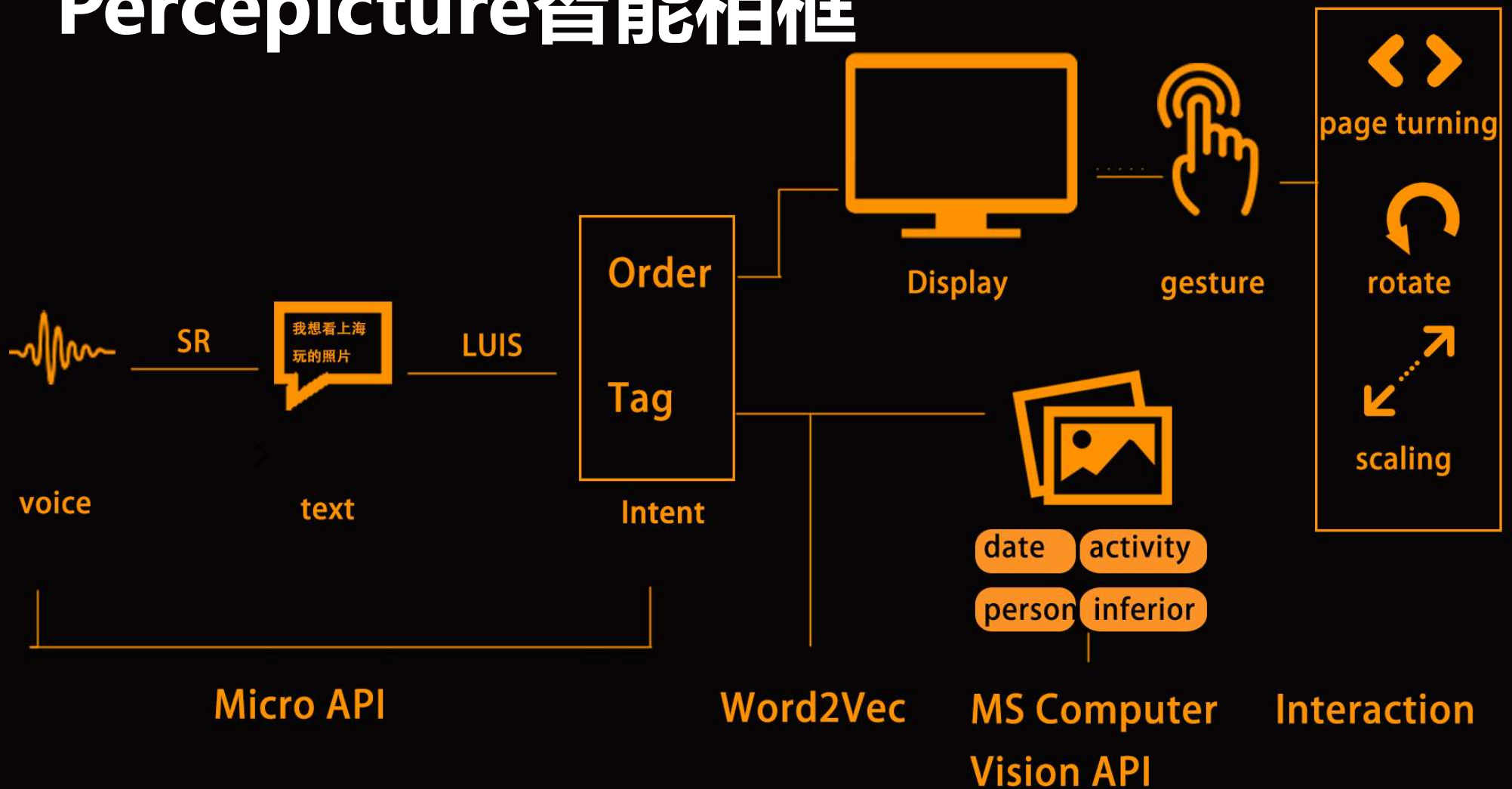


7 挥手可切换下一张



3 上方框内语音中的关键词，图片显示在主页面中。若语音分
析有误，用户可重新语音。另外，用户可通过手势进行上下
翻页等操作。

Percepicture智能相框



学术搜索超新星

- 学术搜索API
- 学术质量，关系网络
- 数据分析
- Kmeans无监督聚类
- 应用场景：
 - 学生寻找导师
 - 企业寻求学术合作



视频 Seeing AI

(看见人工智能)

微软认知服务概览

视觉	语音	语言	知识	搜索
计算机视觉API	自定义智能语音识别服务	必应拼写检查API	学术搜索API	必应网络搜索API
情绪识别API	声纹识别API	语言理解智能服务	实体链接智能服务	必应视频搜索API
人脸识别API	语音识别API	文本分析API	推荐API	必应图片搜索API
视频检测API		网络级语言模型API	知识搜索服务	必应新闻搜索API
		语言分析API		必应自动建议API



认知服务商用计划

全球版Azure.com – 公开预览版

(涵盖8个API付费计划)

世纪互联运营的中国
Azure.cn - 未来上线

Why Azure Products Documentation [Pricing](#) Partners Blog Resources Support [FREE ACCOUNT](#)

Use intelligence APIs to enable vision, speech, language, and knowledge capabilities

[Try for free](#)

[Cognitive Services Overview](#) [Documentation](#)

How Azure pricing works

- ✓ No upfront cost
- ✓ No termination fees
- ✓ Pay only for what you use
- ✓ Per minute billing

[Learn more about Azure pricing](#)

Earn discounts on Azure services with a prepaid 12-month subscription. [Learn more.](#)

[Emotion API](#) [Face API](#) [Language Understanding Intelligent Service](#) [Recommendations API](#) [Bing Speech API](#) [Text Analytics API](#)
[Web Language Model API](#) [Speaker Recognition API](#)

REGION: CURRENCY:

Emotion API

Microsoft's cutting edge cloud-based emotion recognition algorithms let you build more personalized apps. The API identifies emotions in the wild based on facial expressions that are universal.

Pricing Details

The below pricing reflects a preview discount.

TIER	FEATURES	UNIT	PRICE
Emotion API Usage - Basic	Up to 10 transactions per second Face Rectangle specified	Transactions	\$0.10 per 1000 transactions
Emotion API Usage - Standard	Up to 10 transactions per second	Transactions	\$0.25 per 1000 transactions



还等什么？马上行动起来！

➤ 快来我们的网站，体验demo

<https://www.azure.cn/cognitive-services> 线上demo

➤ 开始免费试用

<https://www.azure.cn/cognitive-services/zh-CN/subscriptions> 免费API密钥

➤ 打造你的智能应用

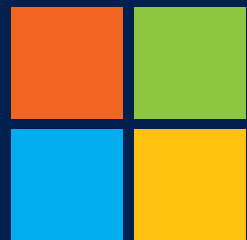
<https://www.azure.cn/cognitive-services/en-us/documentation> 文档，SDK，论坛等着你

➤ 分享你的应用，我们为你提供展示窗口

<https://www.azure.cn/cognitive-services/zh-cn/applications> 应用展示

关注微信公众号-Microsoft认知服务





Microsoft