Building Text Recognize from Image using Tesseract-ocr-iOS

A wrapper framework from Tesseract-OCR

Tesseract-OCR

Tesseract was originally developed at Hewlett-Packard Laboratories Bristol and at Hewlett-Packard Co, Greeley Colorado between 1985 and 1994, with some more changes made in 1996 to port to Windows, and some C++izing in 1998. In 2005 Tesseract was open sourced by HP. Since 2006 it is developed by Google.The latest (LSTM based) stable version is 4.0.0, released on October 29, 2018.

Optical Character Recognition (OCR) is the process of extracting digital text from images. Once extracted, a user may then use the text for document editing, free-text searches, compression, etc.

Tesseract Limitations

Tesseract OCR is quite powerful, but does have the following limitations:

- Unlike some OCR engines (like those used by the U.S. Postal Service to sort mail), Tesseract is unable to recognize handwriting. In fact, it’s limited to about 64 fonts in total.
- Tesseract’s performance can improve with image pre-processing. You may need to scale images, increase color contrast, and horizontally-align the text for optimal results.
- Finally, Tesseract OCR only works on Linux, Windows, and Mac OS X.

Tesseract-OCR-iOS

Tesseract OCR is Objective-C wrapper for Tesseract OCR written by gali8 which you can use in Swift and iOS.

- Install using pod: pod 'TesseractOCRiOS' (it using Tesseract 3.03-rc1 pls check Tesseract OCR iOS)
- Install using latest Tesseract 4.0.0

pod 'TesseractOCRiOS', :git => 'git://github.com/parallaxe/Tesseract-OCR-iOS.git', :branch => 'macos-support'

- Drag tessdata, i.e. Tesseract language data, from the Finder to the Supporting Files group in the Xcode project navigator. Make sure Copy items if needed is checked, the Added Folders option is set to Create folder references, and target is checked before selecting Finish.

Tesseract-OCR-iOS

- There should be only one file here: **Pods_TextRecognizer.framework**, i.e. the pods you just added. Click the + button below the table then add **libstdc++.dylib**, **CoreImage.framework**, and **TesseractOCR.framework** to your project.
- Enable BitCode should be **NO**

Link to detail for Tesseract-OCR

Tesseract-OCR

Tesseract-OCR-iOS

How to train tessdata

Pod file that working with Khmer tessdata:

pod 'TesseractOCRiOS', :git => 'git://github.com/parallaxe/Tesseract-OCR-iOS.git', :branch => 'macos-support'