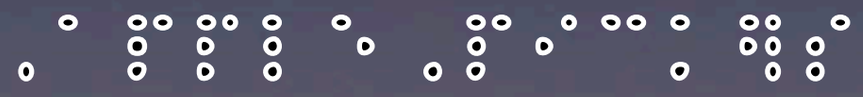
Optical Character Software

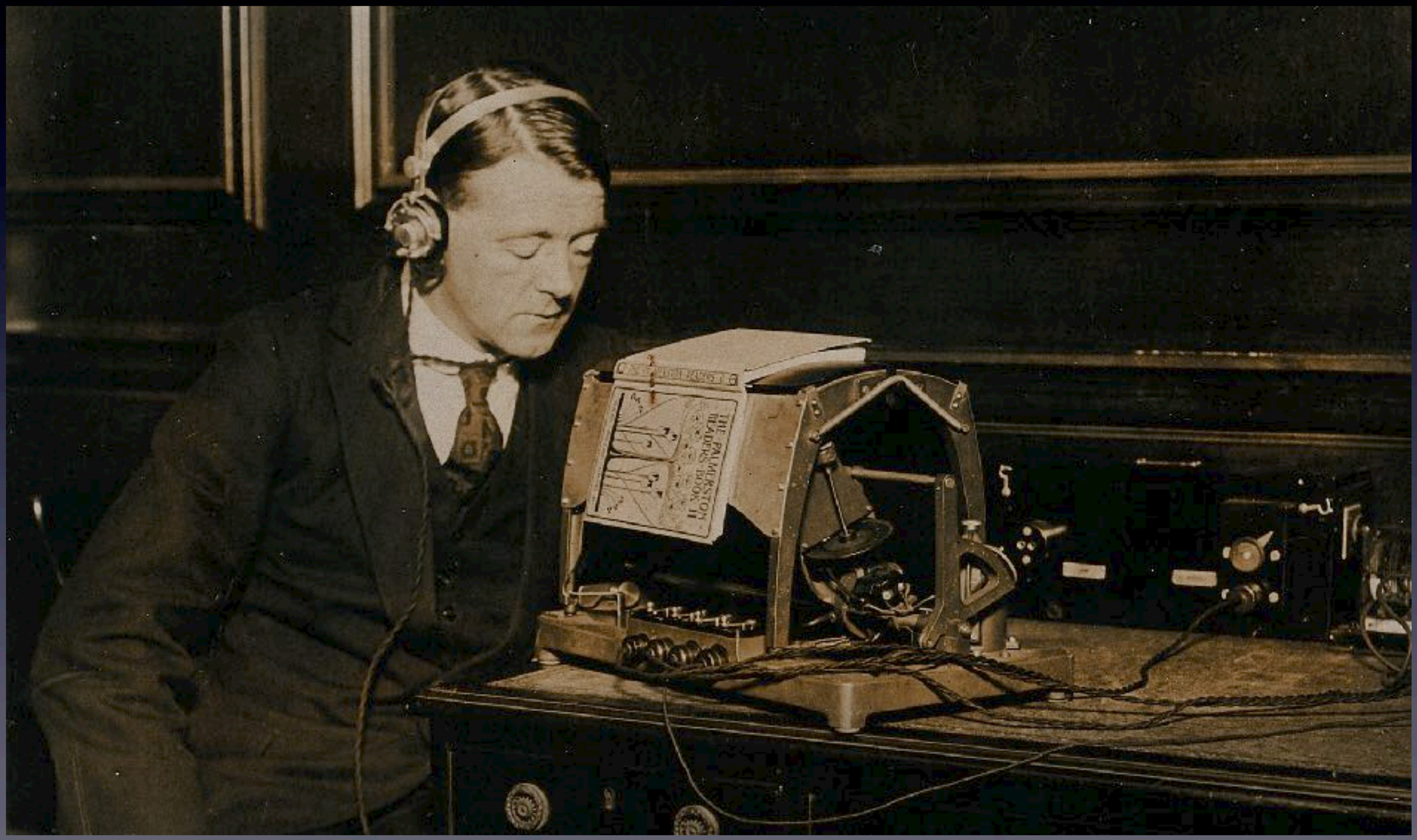
ApplePickers
August 8, 2018

Brief History

- ✓ In 1914, Emanuel Goldberg developed a machine that read characters and converted them into standard telegraph code.
- ✓ Concurrently, Edmund Fournier d'Albe developed the Optophone, a handheld scanner that when moved across a printed page, produced tones that corresponded to specific letters or characters.
- ✓ In the late 1920s and 1930s Goldberg developed what he called a "Statistical Machine" for searching microfilm archives using an optical code recognition system.
- ✓ In 1974, Ray Kurzweil started the company Kurzweil Computer Products, Inc. and continued development of omni-font OCR, which could recognize text printed in virtually any font (Kurzweil is often credited with inventing omni-font OCR, but it was in use by companies, including CompuScan, in the late 1960s and 1970s
- ✓ Kurzweil decided that the best application of this technology would be to create a reading machine for the blind
  - Unveiled in 1976 and commercial product sold in 1978
  - Company sold to Xerox in 1980 which spun unit off as Scansoft which merged with Nuance Corp.
  - OCR to Braille available as OpenSource or BrailleTranslator.org



Optophone



Options

- ✓ Service companies offer bulk OCR to other companies
  - Conversion of microfiche
  - ExperVision
- ✓ Online free services
  - OCR Space

Name	Founded year	Latest stable version	Release year	License	Online	Windows	Mac OS X	Linux	BSD	Programming language	SDK?	Languages	Fonts	Output Formats	Notes
Tesseract	1985	3.05.02	2018	Apache	No	Yes	Yes	Yes	Yes	C++, C	Yes	100+[1]	Any printed font	Text, hOCR, [2]PDF, others	Created by Hewlett
Readiris	1986	16	?	Proprietary	?	Yes	Yes	?	?	?	Yes	100+[5]	?	?	Owned by Canon
Screenworm	2013	1	2014	Proprietary	No	No	Yes	No	No	Objective-C++	No	57	?	TXT	Product of Funchip. Uses
ExperVision [6]TypeReader & RTK	1987	7.1.170.1125	2010	Proprietary	Yes	Yes	Yes	Yes	Yes	C/C++	Yes	21	2618		Has a Mobile and Embedded System version
ABBYY FineReader	1989	14	1/25/17	Proprietary	Yes	Yes	Yes	Yes	Yes	C/C++	Yes	192[8]	?	DOC, DOCX, XLS, XLSX, PPTX, RTF,	ABBYY also supplies SDKs for
Asprise OCRSDK	1998	15	2015	Proprietary	Yes	Yes	Yes	Yes	Yes	Java, C#,VB.NET, C/C++/Delphi	Yes	20+[11]	?	Plain text, searchable PDF,	Java, C#, VB.NET, C/C++/Delphi
LEADTOOLS[16]	1990[17]	19	2014	Proprietary	Yes	Yes	Yes	Yes	No	C/C++, .NET, Objective-C, Java,	Yes	56[18]	Any printed font	PDF, PDF/A, DOC, DOCX, XLS, XPS,	Supports Latin, Asian, Arabic, and MICR
CuneiForm	1996	1.1	4/19/11	BSDvariant	No	Yes	Yes	Yes	Yes	C/C++	Yes	28	Any printed font	HTML, hOCR, native, RTF, TeX,	Enterprise-class system, can save text
OmniPage	1970s	19.2	2015	Proprietary	Yes	Yes	Yes	Yes	No	C/C++, C#[24]	Yes	125[25]	Machine and handprinted fonts	DOC/DOCX XLS/XLSX PPTX RTF PDF	Product of Nuance
gImageReader[27]	2009	3.2.99	2017-07	GPL	No	Yes	Yes	Yes	No	C++	?	100+	Any printed font	TXT, PDF, hOCR	uses Tesseract OCR engine
GOCR	2000	0.5	2013	GPL	Yes[28]	Yes	Yes	Yes	Yes	C	?	20+	?		
Ocrad	?	0.25[29]	4/16/15	GPL	Yes	Yes	Yes	Yes	Yes	C++	Yes	Latin alphabet	?		Command line
SmartScore	1991	10.5.8	2015-07	Proprietary	No	Yes	Yes	No	No	?	?	?	?		For musical scores
PDF OCR X	2008	2.0.22	2016	Proprietary	No	Yes	Yes	No	No	Java, C++, Objective-C	No	100+	?	TXT, Searchable PDf	Drag and drop UI.
OCRopus	2007	1.3.3	12/16/17	Apache	No	No	Yes	Yes	Yes	Python	?	All languages using Latin	Normal Latin script	TXT, hOCR[30], PDF[31]	Pluggable framework under
MathOCR	2014	0.0.3	2015	GPL	No	Yes	Yes	Yes	Yes	Java	?	?	?	HTML, LaTeX	Features mathematical formula
Yunmai OCR SDK	2002	1	2013	Proprietary	Yes	Yes	Yes	Yes	Yes	Java, C++, C, object pascal, objective-C	Yes	14	Any printed font	TXT, PDF	Has the advantage of
Anyline SDK	2013[34]	3.5.1[35]	2016[35]	Free non-commercial	No	No*	No*	No*	No*	Java (Android), Objective-C &	Yes[36]	2 (German, English)	Any printed trainable font[38]	Plain text, verification image	*Customizable mobile OCR SDK for
AliusDoc AD-SCI[7]	2005	2.1	2015	Proprietary	No	Yes	No	No	No	VB.Net	For Extensions	All ASCII-compatible languages	?	XML, PlainText, any other thru SDK extensions	Minimal need for post-sale Professional
E-aksharayan	2010					Yes	No	Yes	No			14		RTF, TXT, BRL	
Nicomsoft OCR SDK	1999	5.5	2015	Proprietary	No	Yes	No	Yes	No	C#, VB.NET, C++, Delphi, Java	Yes	25+[14]	?	Searchable PDF, Text, RTF	C#, VB.NET, C++, Delphi, Java OCR
AnyDoc Software	1989	?	?	Proprietary	No	Yes	No	No	No	VBScript	?	?	?		Works with structured, semi-structured,
OCR.space	2015	3.02	2017	GPL	Yes	Yes	No	No	No	C#	Yes	23	Any printed font	TXT	Windows desktop software,
SimpleOCR	2002	3.5	2008	Proprietary	No	Yes	No	No	No	?	?	?	?		
Dynamsoft OCR SDK	2003	8.2	2012	Proprietary	Yes	Yes	No	No	No	C/C++	Yes	40+[23]	?	PDF, TXT	
Microsoft Office	2011	?	2007	Proprietary	No	Yes	No	No	No	?	?	?	?		
FreeOCR	?	4.2	Aug-12	Proprietary	No	Yes	No	No	No	?	?	?	?		[26]
Microsoft Office	?	Office 2007	2007	Proprietary	No	Yes	No	No	No	?	?	?	?		Uses OmniPage[citatio
OCR.net	2016	?	2016	Proprietary	Yes	No	No	No	No	Java, C++, PHP, Objective-c	No	100+	?	TXT, Searchable PDF	Online service powered by PDF OCR X for
Puma.NET	?	?	10/29/09	BSD	No	Yes	No	No	No	C#	Yes	28	Any printed font		.NET OCR SDK based on Cognitive
ReadSoft	?	?	?	Proprietary	No	Yes	No	No	No	?	?	?	?		Scan, capture and classify business
Scantron	?	?	?	Proprietary	No	Yes	No	No	No	?	?	?	?		For working with localized interfaces,
OCRFeeder	2009-03	0.8.1	12/22/14	GPL	No	No	No	Yes	No	Python	?	?	?		Features a full user interface and has a
MeOCR	2012	1.0.0	2012	Freeware	No	Yes	No	No	No	C/C++/C#	Yes	28	Any printed font	HTML, hOCR, native, RTF, TeX,	Windows application.

Pros and Cons of Mac-based applications

- ✓ Some have been around for a very long time
- ✓ Most are automatic with little capability to tweak output
- ✓ Generally, you get what you pay for
  - None do a mistake-free job
  - Output should be reviewed
- ✓ Image resolution, font size, line straightness can dramatically affect quality
- ✓ All can handle multiple languages
- ✓ Most all use either OpenSource or proprietary 3rd party OCR engines such as Tessaract which is now Google's ML OCR Kit

Search Results for 'OCR' 1-120 of 172

Sort By: Relevance

PDFScanner - Scan a...
Productivity
★★★★★ 185 Ratings
$14.99

HP Easy Scan
Utilities
★★★☆☆ 541 Ratings
OPEN

PDF Office: Edit Text ...
Business
$0.99

PDF OCR X Communit...
Productivity
★★★☆☆ 42 Ratings
GET

Presenter
Utilities
OPEN

PDF Converter Pro
Productivity
★★★☆☆ 14 Ratings
$39.99

OCR App by I FANTO
Business
★★★★★ 5 Ratings
GET

PDF-OCR-Free
Productivity
★★☆☆☆ 16 Ratings
GET
In-App Purchases

iText - OCR & Translat...
Utilities
★★★★★ 15 Ratings
GET
In-App Purchases

PDF-to-Word-Free
Productivity
★★☆☆☆ 24 Ratings
GET
In-App Purchases

PDF OCR X Enterprise...
Productivity
$29.99

PDF to Word with OCR
Productivity
★★★☆☆ 37 Ratings
$12.99

PDF Converter with O...
Business
★★★★☆ 40 Ratings
$24.99

Mathpix Snipping Tool
Productivity
★★★★★ 6 Ratings
GET

FineReader OCR Pro
Productivity
★★★★☆ 3 Ratings
$119.99

OCRKit
Productivity
★★★★☆ 8 Ratings
$29.99

Prizmo 3 - Scanning &...
Business
★★☆☆☆ 6 Ratings
$39.99
In-App Purchases

PDFelement 8 Pro - P...
Productivity
★★☆☆☆ 4 Ratings
GET
In-App Purchases

Condense
Utilities
★★★★☆ 20 Ratings
$2.99

Tessa - OCR
Productivity
GET
In-App Purchases

PDF-Converter-Free
Productivity
★☆☆☆☆ 8 Ratings
GET
In-App Purchases

OCRTOOLS
Productivity
$4.99

Sign Master - PDF Ass...
Business
★☆☆☆☆ 4 Ratings
GET
In-App Purchases

PDF-to-Exce -Free
Productivity
★★☆☆☆ 17 Ratings
GET
In-App Purchases

PDF to Excel OCR Con...
Business
★★☆☆☆ 10 Ratings
GET

PDF Converter OCR
Business
★★★★★ 10 Ratings
$69.99

DocuScan Plus
Productivity
GET

PDF Text Extractor - E...
Productivity
★★★☆☆ 31 Ratings
$14.99

PDF Editor Pro
Business
★★☆☆☆ 9 Ratings
$69.99

PDF-to-Excel-Pro
Business
★☆☆☆☆ 5 Ratings
GET
In-App Purchases

PDF to Excel with OCR
Business
★★★☆☆ 11 Ratings
$12.99

PDF Scanner Pro
Business
★★☆☆☆ 8 Ratings
GET
In-App Purchases

Text Extractor - Extra...
Business
★★★★☆ 30 Ratings
$9.99

Paperless
Productivity
★★★☆☆ 7 Ratings
$19.99

PDF to Word - +
Business
$19.99
In-App Purchases

PDF-to-ePub-Free
Productivity
★☆☆☆☆ 8 Ratings
GET
In-App Purchases

PDF-to-HTML-Free
Productivity
★☆☆☆☆ 8 Ratings
GET
In-App Purchases

Aristocrat
Utilities
★★★★☆ 5 Ratings
GET

PDF & Image Text Extr...
Business
★★★★☆ 1 Ratings
$2.99

Visualizer
Utilities
GET

PocketDrone [Service]
Business
GET

PDF-to-PowerPoint-Fr...
Productivity
GET
In-App Purchases

DocScanner
Productivity
★★☆☆☆ 18 Ratings
$19.99

Image-to-PDF-Free
Productivity
★★☆☆☆ 8 Ratings
GET
In-App Purchases

PDF to Text with OCR
Productivity
$1.99

eDoc Organizer Cloud...
Business
GET

PDF-to-Text-Free
Productivity
GET
In-App Purchases

Shimo Scanner - PDF ...
Productivity
$49.99

ExactScan Pro
Productivity
★★★☆☆ 8 Ratings
$99.99

+ PDF Converter OCR
Business
$19.99

PDF Editor Plus - PDF ...
Productivity
$19.99

PDF Toolbox +
Productivity
★★★☆☆ 5 Ratings
$4.99

PDF Edit Express - PD...
Business
$0.99

Easy Screen OCR
Productivity
GET

Scanner Pro -PDF Doc...
Productivity
★★★★☆ 35 Ratings
$0.99

OCRWizard - Convert ...
Utilities
★★★☆☆ 10 Ratings
$69.99

WorldCard
Business
GET

Subtitle Extractor
Video
★★★☆☆ 17 Ratings
$49.99

PDF to HTML with OCR
Utilities
$7.99

PDF Converter + +
Productivity
$29.99

Two Major Groups

✓ Those that come with scanning software or scanner itself, and those that do not

- Example: HP Easy Scan

  ✦ Get from Mac App Store

✓ Generally those that focus solely on OCR do a better job

Stand-alone Applications/ Predominantly OCR

	Fine Reader Pro	Prizmo	Cisdem OCR Wizard	PDF OCR
				
MSRP	$120	$50	$60	$0
Pros	Best accuracy, handles tabular data, scans	Multi-column text, Business cards	Easy to use	Free, simple interface
Cons	Cost, erratic support	Can't handle tabular data	Poor with tabular data	Output only PDF or .txt files

Combination Products

	Acrobat	PDFPen	VueScan
			
Cost	$15/mo	$75/$125	$40
Pros	Best quality, many other features	Mac Centric, other features	Works with almost any scanner
Cons	Beyond expensive	Good OCR, but exports file	Average OCR, complicated layout

OCR Everywhere

- ✓ There are several websites that offer OCR services
  - Minimal to no tweaking of scan
- ✓ OCR is built in to several iOS apps including Adobe Acrobat, MS-OneNote, Text Scanner and Office Lens