Aquaforest OCR SDK for .NET Release Notes

Version 2.20
April 2017

Aquaforest

1 Version 2.20

1.1 .NET Framework

The Aquaforest engine has been upgraded to use .NET Framework 4.5.2 instead of .NET Framework 3.5. Consequently, there is no need now to specify "`useLegacyV2RuntimeActivationPolicy`" in the `app.config` file of your application.

However, if you are using Visual Studio 2012 or 2013 to build your application(s), you need to ensure that you also have the .NET Framework 4.5.2 Multi-targeting pack installed so as to be able to build applications that target .NET Framework 4.5.2.

1.2 Additional Language Support

The Extended engine now supports Arabic and Farsi languages.

1.3 VC++ Redistributables

This version requires VC++ 2013 instead of VC++ 2012 for the Aquaforest engine.

1.4 PDFBox

The SDK now uses PDFBox instead of iTextSharp.

1.5 Process PDF files with vector objects in native mode

PDF documents that contain only vector images (e.g. CAD drawings) can now be OCRed natively. In previous versions, the PDF needed to be re-imaged (`ConvertToTiff`) before OCRing.

By default, pages that contain only vector objects are rasterized. Pages that do not have any images but contain vector objects as well as electronic text are skipped from rasterization. However, sometimes there can be a page that contains vector objects (CAD drawings) but its title may be in electronic text. To force rasterizing pages like these, there is a property called "`PdfToImageForceVectorCheck`" which needs to be set to true. This property can be set through the `OCR` object or the Properties.xml file.

1.6 Font sizing

The sizing of OCRed text added to PDF documents has been improved in both OCR engines. This can be tested by selecting the OCRed text in a PDF reader.

Previous versions	New version
(12) **United States Patent** Gutterman	(12) **United States Patent** Gutterman (12) **United States Patent** Gutterman (12) **United States Patent** Gutterman

1.7   Retain Viewer Preferences

A new property (`ocr.RetainViewerPreferences`) has been added to enable retaining the PDF viewer preferences (Page Layout, Page Mode) when OCRing PDF and using "`ConvertToTiff`". The viewer preferences are automatically retained when processing in "Native" mode.

1.8   PDF/A

To generate PDF/A files, "`ocr.EnabledPDFAOutput`" needs to be set instead of "`ocr.EnabledPdfOutput`".

If the source file is a PDF file, "`ocr.ConverToTiff`" needs to be set to `true` because conversion to PDF/A is no longer available in "Native" mode.

Additionally, the "`ConvertPDFToPDFA`" method which previously enabled creating PDF/A files without going through the OCR process, is no longer available as there were lots of instances where invalid PDF/A outputs were generated.

1.9   PdfToImageEngine

This version of the SDK only has 1 engine that converts PDF pages to images. All occurrences of "`PdfToImageEngine`" have been removed from the API and the Properties.xml.

1.10 Stamps

Stamps have been removed in this release. To use stamps, you need to use the PDF Toolkit, which is available to users with Advanced and Extended License.

1.11 FIPS Compliancy

The SDK is now FIPS compliant.

1.12 Additional settings in the Extended engine

The following settings have been added to the Extended engine:
- BinarizationMode
- TextSpacing
- TextType
- CharactersPerInch
- LimitCharsetCharacters
- UserLexicon

Refer to the **OCRSDK 2.2 (Extended engine).chm** file found in the folder "[SDK installation path]\docs\help" to view description of these new settings.

Aquaforest

2 Version 2.10

2.1 Licensing

There are a couple of changes in the way this release is licensed; this is to offer buyers a higher flexibility. The table below shows a breakdown of the licensing.

Function	Basic Edition	Standard Edition	Advanced Edition	Extended Edition
OCR from Bitmap or TIFF	✓	✓	✓	✓
Image Pre-Processing and Auto-Rotation	✓	✓	✓	✓
Support for 23 Languages	✓	✓	✓	✓
.NET Programmatic / Zonal Access to results	✓	✓	✓	✓
Txt / RTF Output	✓	✓	✓	✓
1 Thread	✓	✓	✓	✓
Blank Page Removal	✓	✓	✓	✓
PDF Merging	✓	✓	✓	✓
Barcode Decoding	✓	✓	✓	✓
PDF Input		✓	✓	✓
Searchable PDF Output		✓	✓	✓
2 Threads		✓	✓	✓
Stamps on PDF Output		✓	✓	✓
Unlimited Threads			✓	✓
Advanced MRC Compressed PDF Output			✓	✓
Advanced Pre-Processing			✓	✓
Support for 127 languages				✓
Asian language support				✓
Support for multiple languages within a single document from the same character set				✓
Multiple document output formats: CSV, DOCX, EPUB, EXCELML, HTML, OPEN DOCUMENT TEXT, PDF, RTF, TXT, WORDML, XLSX and XPS				✓
Multiple PDF version output support				✓
Intelligent High Quality Compression				✓

Aquaforest

2.2 Barcode Decoding

SDK 2.10 now supports decoding barcodes from images and PDF documents. The following barcode formats are currently supported:

- Aztec 2D barcode format.
- CODABAR 1D format.
- Code 39 1D format.
- Code 93 1D format.
- Code 128 1D format.
- Data Matrix 2D barcode format.
- EAN-8 1D format.
- EAN-13 1D format.
- ITF (Interleaved Two of Five) 1D format.
- MaxiCode 2D barcode format.
- PDF417 format.
- QR Code 2D barcode format.
- RSS 14
- RSS EXPANDED
- UPC-A 1D format.
- UPC-E 1D format.
- UPC/EAN extension format.
- MSI
- Plessey

2.3 New iDRS engine (Extended OCR Module)

Aquaforest OCR SDK 2.10 now has the latest version of the iDRS engine (iDRS 15) in the Extended OCR module. It provides the following new features:

- Improved character recognition engine
- Additional output formats:
  - PDF/A-1a
  - EPUB (short for electronic publication) is an e-book standard by the International Digital Publishing Forum (IDPF). Files have the extension .epub.
- New Asian OCR engine. It now has two Asian OCR engines to process documents with Asian languages (Japanese, Korean, Simplified Chinese and Traditional Chinese). The engine to use can be set using the `AsianOcrEngine` property in the `Ocr` class.

2.4 System.IO.Stream Input & Output

SDK 2.10 now provides the option of reading input source from `System.IO.Stream` and saving output results to `System.IO.Stream`.

2.5 Retain Bookmarks

Added support for retaining bookmarks in the output PDF document when OCRing a PDF source and using `ConvertToTiff`.

2.6 Retain Metadata

Added support for retaining metadata in the output PDF document when OCRing a PDF source and using `ConvertToTiff`.

2.7   Merging PDFs

Merging certain PDF files was causing the resulting merged PDF to be very large. This has now been fixed.

2.8   Hanging

The Aquaforest OCR SDK was hanging on certain machines when subjected to extreme load such as running in multi-core. This has now been fixed.

2.9   SMask support

Added support for PDF files with SMask images.

2.10 LogFilePath

A new `LogFilePath` property has been added to enable users to get the location of the debug log file path when `EnableDebugOutput` is set to `DebugLevels.LOG_FILE_PATH`

Δquaforest