

Contents

- What is OCR 2
- How does OCR work in ImageCapture Suite? 2
- Use Dynamic OCR in ImageCapture Suite 2
 - Check OCR version and status..... 2
 - Deploy OCR component to your clients..... 2
 - Deploy language data to your clients 3
 - OCR settings 3
- Perform OCR 4
- More Info 4

What is OCR

Optical Character Recognition, usually abbreviated to OCR, is widely adopted by document management solutions nowadays. Dynamsoft offers OCR as an add-on in ImageCapture Suite since version 8.2. Besides searchable PDF and text, our OCR SDK also enables you to extract detailed positioning as well as format information about the extracted text

How does OCR work in ImageCapture Suite?

ImageCapture Suite is designed for web applications. So we provide two options for you to add the OCR feature to your web application:

- Deploy the OCR module on the server side with the help of [Dynamic .NET TWAIN](#). The client machine needs to upload the documents to the server, and then the server performs OCR and returns the result to the client machine. Only Windows server is supported in this case. If you are interested, you can download a working sample [here](#).
NOTE: the sample needs a trial license of its own; you can [contact Dynamsoft Support](#) for it.
- Or you can deploy the OCR module on the client machine together with ImageCapture Suite. In this case, every time the client performs OCR, the result is returned locally. Compared to the last option, it's faster and it doesn't add load to your server which can be Windows, Linux, or even Mac etc.

In this document, we talk about the 2nd option - how to deploy OCR module on the client machine.

Use Dynamic OCR in ImageCapture Suite

Check OCR version and status

Before we use the OCR function, we need to check and see whether the OCR component has been installed on the client machine and whether it is the same version as deployed on the server. To do that, we can use the property [OCRVersion](#):

```
if (DWObject.OCRVersion != '8, 2, 0, 0') {  
    //not installed or not version 8.2.0.0  
}  
else {  
    //installed and version is 8.2.0.0  
}
```

Ok, so what should we do if the OCR component is not installed yet, or need to be updated? We need to deploy the component to the client.

Deploy OCR component to your clients

To do this, we use the method [HTTPDownloadResource](#). It will download a zip file which contains the required files of the OCR component and deploy the files to the client machine. If old files already exist, they will be replaced. After the deployment, the client should be able to use the OCR component locally.

**The zip file contains the OCR DLLs (DynamicOCR.dll & DynamicOCRx64.dll). You can find the zip file in the installation folder of ImageCapture Suite.*

For example, if you place the zip file under <http://yourServer/download/OCR.zip>, you can use the code below to download and deploy OCR:

```
DWObject.HTTPDownloadResource("yourServer", "/download/OCR.zip", "OCR.zip");
```

Note: please make sure the client can access the zip file from the URL.

Deploy language data to your clients

The OCR engine performs its magic with the help of the language data which functions much like a dictionary. Therefore we need to download the language data together with the OCR component. To do this, we use the method [HTTPDownloadOCRLangData](#).

```
DWObject.HTTPDownloadOCRLangData("www.dynamsoft.com", /download/OCR  
Language/English.zip", true);
```

In the sample code above, we download the English language data from Dynamsoft website. Dynamsoft OCR supports more than 40 languages. [Download Other Language Packages>>](#)

If your clients do not have external access, you can download and place the zip file on your own server. For example, if you put the zip file under <http://yourServer/download/English.zip>:

```
DWObject.HTTPDownloadOCRLangData("YourServer", /download/English.zip", true);
```

OCR settings

1. Specify language type and page processing mode.

```
DWObject.OCRLanguage = "eng";  
DWObject.OCRPageSetMode = 3;
```

For [OCRPageSetMode](#), the default value is 3. It is the auto mode which means the OCR component will analyze the page and choose the proper mode which can be one of the following:

```
PSM_OSD_ONLY = 0,  
PSM_AUTO_OSD = 1,  
PSM_AUTO_ONLY = 2,  
PSM_AUTO = 3,  
PSM_SINGLE_COLUMN = 4,  
PSM_SINGLE_BLOCK_VERT_TEXT = 5,  
PSM_SINGLE_BLOCK = 6,  
PSM_SINGLE_LINE = 7,
```

PSM_SINGLE_WORD = 8,
PSM_CIRCLE_WORD = 9,
PSM_SINGLE_CHAR = 10,
PSM_COUNT = 11

2. Specify whether the output format is text or PDF with the property [OCRResultFormat](#):

0 – Plain Text (ASCII).

1 – Plain Text PDF.

2 – Image over text PDF.

```
DWObject.OCRResultFormat =0; //output as text
```

Perform OCR

We provide several ways to perform OCR, you can find the related methods in the blow table.

OCR, OCR1	OCRDirectly	OCRDirectlyEx
OCREx, OCREx1	OCRSelectedImages	OCRSelectedImagesEx

```
DWObject.OCREx(DWObject.CurrentImageIndexInBuffer);
```

More Info

For more OCR-related properties & methods, you can check out:

<http://www.dynamsoft.com/help/ICS/index.htm>

->**Methods**->**OCR Related**

->**Properties**->**OCR Related**