

# DIVINEV

**eDoc 5.0**

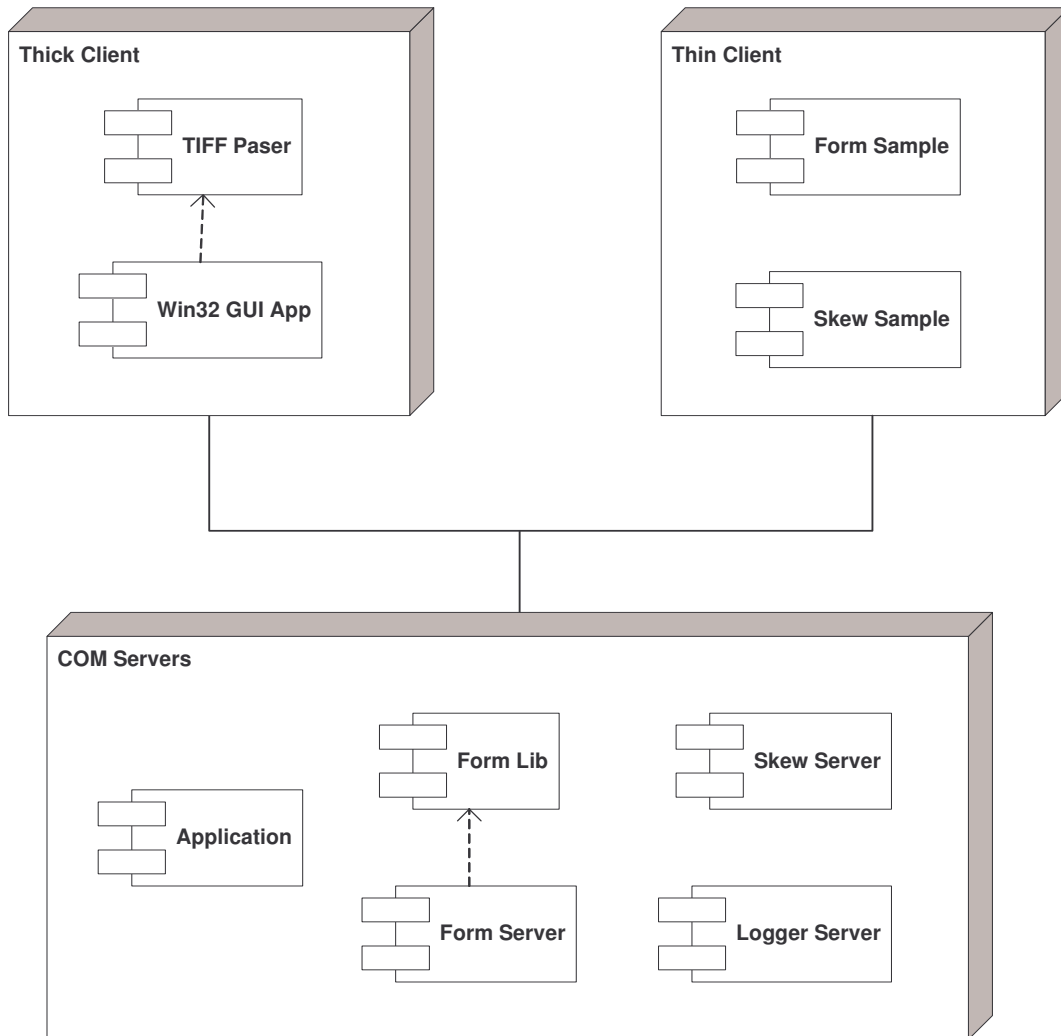
Copyright (c) 2004 Divinev  
All Rights Reserved

Email: [edoc@divinev.com](mailto:edoc@divinev.com)

URL: <http://edoc.divinev.com>

## INTRODUCTION

eDoc is a toolkit for the developers working in the areas of document imaging, OCR, and document management. It includes server components, GUI based client application and SDK sample code. The major functionalities provided by server components are form template training, form identification, form dropout and generic document skew detection.

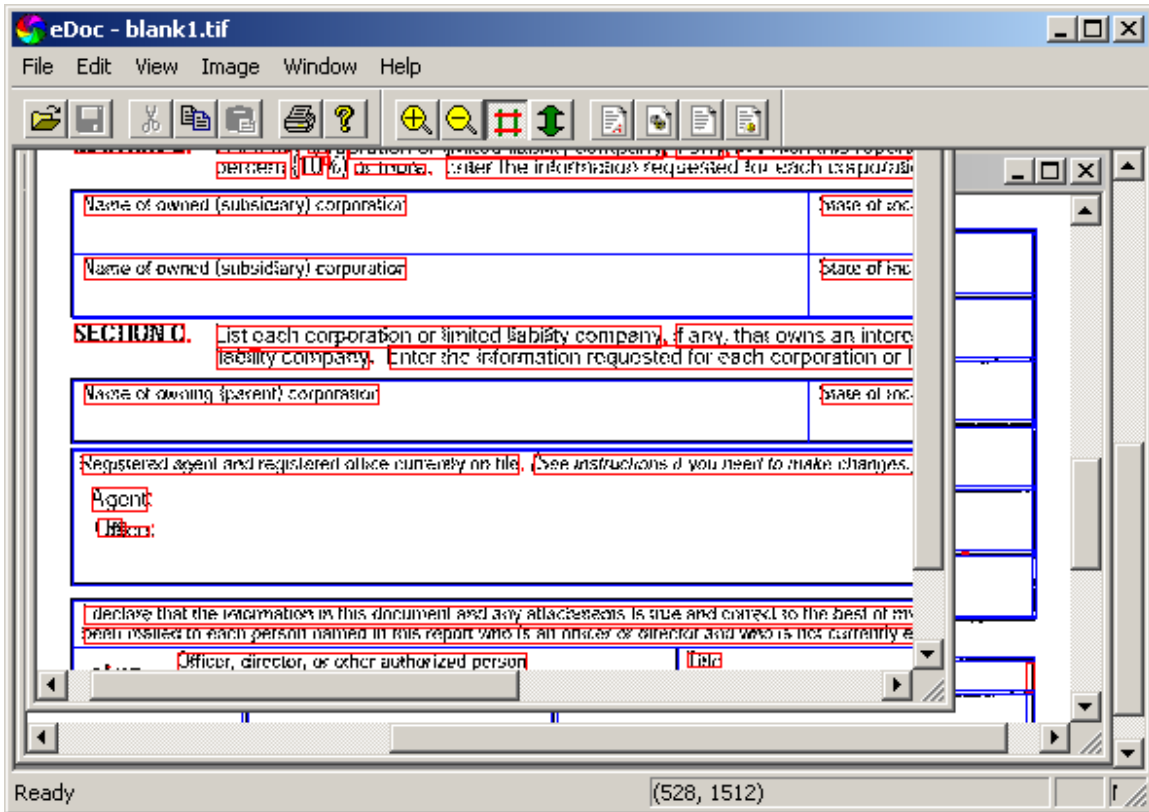


## GUI CLIENT APPLICATION

The eDoc application can be run in either command line mode for batch or intuitive GUI mode. The GUI client is a Windows MFC application, which has the functionality to validate

- Form model from template training
- Form identification
- Form dropout results

- Skew detection output



User can type “eDoc.exe -h” for usage message. The GUI mode will be started if there is no option given.

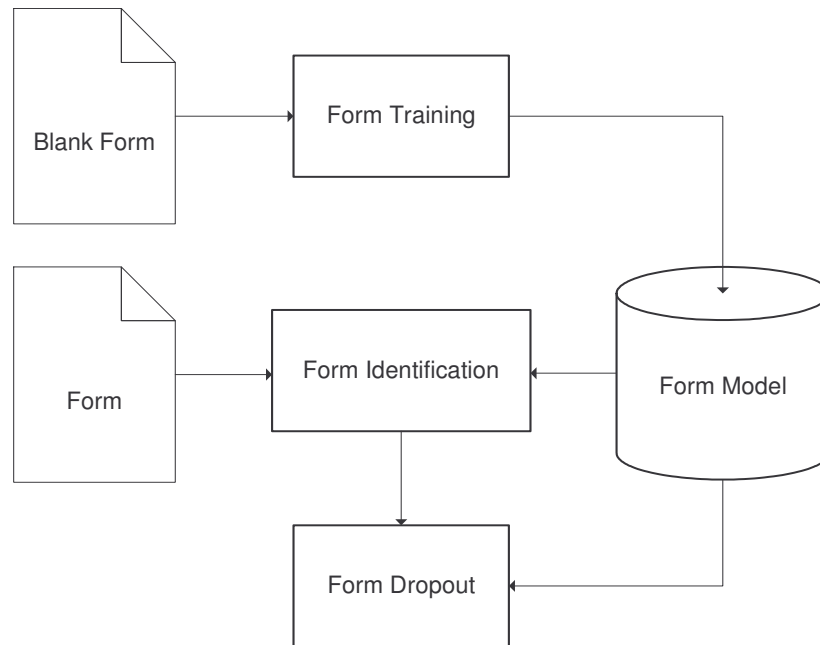
## SERVER COMPONENTS

All the server components are built in COM compliance. The thin client sample code tells how to easily use these COM server components. One major server component is form server that provides the services of form template model training, form identification against trained form models and form dropout. Another major server component is skew detection server that can detect skew angle for any type of documents.

- It can detect skew angle within an arbitrary range up to 180 degrees.
- It can work for various document images at any resolution.
- It has a flexible accuracy from 5 to 0.01 degrees.
- It has a high throughput.



- High tolerance of difference between form template and filled form caused by printer, digitizer, or other factors:
  - Horizontal and vertical scale:  $\pm 5\%$
  - Horizontal and vertical shift:  $\pm 1$  inch
  - Skew:  $\pm 10^\circ$
  - Scanning resolution required:  $> 150$  DPI



## 1. Form Template Model Training

If a blank form template is available, it can be used in form template model training. The training process is one hundred percent automatic without any interaction required. A trained form model contains the location information of form frames, static form text blocks, checkboxes, and barcodes. The variation between two form templates caused by different form types, versions, deformation or other factors has been taken into account.

## 2. Form Identification

The form identification module can automatically find out a best form template model from a number of candidates for an input form. It will return nothing, if there is no good match found.

## 3. Form Dropout

A best-matched form template that has a match score higher than a threshold will be used to remove form frames and static form text. At the same time, checkmarks and barcodes will be located and characters or strokes broken by form frame removal will be reconstructed. The results will be saved into an image file with user data only along with a ROI text file containing location and type information.



#### 4. Sample Output – User Data

0407021001

3-20092-6185-3

2004

|||||  
GFI PROPERTIES LLC  
16414 PEMCANYON  
SAN ANTONIO TX 78240-5601

08001432-24 7



JAMES W. GIBSON

MANAGER

16414 PEMCANYON SAN ANTONIO, TX 78240

JAMES W. GIBSON  
16414 PEMCANYON  
SAN ANTONIO, TX 78240



MANAGER 16 MAR 04 210 262-3741

0004382

## 5. Sample Output – ROI Text File

1356 452 343 119 barcode  
1290 92 189 46 text  
1116 184 227 29 text  
1530 186 62 21 text  
163 260 616 27 text  
217 296 351 17 text  
217 329 292 17 text  
217 362 495 17 text  
1431 384 178 27 text  
1664 387 12 24 checkmark  
114 596 41 32 text  
316 919 614 28 text  
1245 1246 14 34 text  
276 2036 205 77 text  
789 2078 200 38 text  
229 2042 75 96 text  
1031 855 115 20 text  
146 856 246 21 text  
216 1858 246 21 text  
219 1889 243 24 text  
216 1921 344 28 text  
1067 2060 202 41 text  
1347 2076 289 27 text  
1583 2120 115 29 text