User Manual

ACECAD® Digimemo® based Handwriting Data Collection Tool:

digimemo-dct 1.0

Contents

1   Introduction

LipiTk (Lipi Toolkit) is a generic toolkit for Online HandWriting Recognition (HWR). It provides a set of script-independent shape recognizers, tools and building blocks which can be used by different kinds of users for the recognition of different scripts and shapes. *ACECAD* Digimemo [1] is a portable device that captures user's handwriting on a normal paper. This document describes *digimemo-dct 1.0,* a Digimemo-based data collection tool for supporting handwriting recognition.

1-1   License

This tool is being made available under the MIT license. For license details, refer to **MIT license**.

2 Prerequisites

This section describes the prerequisites for installing and using the *digimemo-dct 1.0.*

2-1 Supported platforms and environment

digimemo-dct 1.0 has been tested on the following platforms:

- Windows XP Professional edition
- Windows 2000 Professional edition

The tool has been tested with the following device models

- **DigiMemo A402**
- **DigiMemo A502**

NOTE: The tool has been tested with forms designed using Adobe® Form Designer 7.0. However, these forms can be created using any form design tool.

2-2 Disk space requirements

The size of *digimemo-dct 1.0* package is 409 KB, and the free disk space required to extract this package is 821 KB.

2-3 Contents of digimemo-dct 1.0 package

The *digimemo-dct 1.0* package contains the following:

1. The executable for the tool, `digimemodct.exe`
2. The configuration file for the tool, `dct.cfg`
3. Template forms and their configuration file
4. `User Manual` and `Release notes` for the tool
5. The device configuration file, `info.cfg`

The contents after extraction will appear as follows:

```
digimemodct1.0/
    docs/
        DigiMemoDCT_1_0_0_Release_Notes.doc
        DigiMemoDCT_1_0_0_User_Manual.doc

    Templateformsandconfigfile/
```

```
templatedct.cfg
templatecalibrationpage.xdp
templatesymbolpage.xdp

dct.cfg
 info.cfg
digimemodct.exe
```

3   Data collection tool - digimemo-dct 1.0

This section discusses in detail the data collection tool and its usage.

3-1 Introduction

ACECAD Digimemo [1] is a portable device that captures user's handwriting on a normal sheet of paper. The sheet is placed over the digitizing pad of the device and while writing, the (digital) pen's position is recorded in the form of X-Y coordinates and stored in device's on-board memory.

The *digimemo-dct 1.0* is a GUI tool that facilitates the creation of handwriting datasets by collecting handwriting data samples from different writers using *ACECAD Digimemo* [1] devices. For a given script, data collection from writers is performed using specialized forms attached to the devices [2]. The output of the tool is a set of *UNIPEN* [3] files organized into a directory structure, containing writer's profiles, collection procedure details and the digital ink data. Please refer to Appendix for a sample *UNIPEN* ink file generated by the tool.

3-2 Steps for data collection

This section describes the steps involved in data collection using the *digimemo-dct 1.0*.

3-2-1 Collecting handwriting samples

The following sections outline the steps involved in collecting handwriting samples using the *digimemo-dct 1.0*.

3-2-1-1   Making forms

Data collection involves collecting handwriting samples from the contributing writers in one or more "trials". Each trial of a user requires a set of forms (A4/A5 size) to be filled using the *ACECAD Digimemo*. This tool assumes the following structure for each trial:

Calibration Page: The first page of the trial which is used for calibration. This page is required only when the calibration is set to **Yes** in the configuration settings of the tool.

Symbol Page(s): The symbol page(s) follow the Calibration Page (if present) in any trial and are used to collect the handwriting samples for the symbols of the script.

A sample symbol page for English uppercase letters is shown below (Figure 1). It contains square boxes for every symbol. The handwriting samples for the symbols must be provided in their respective boxes.
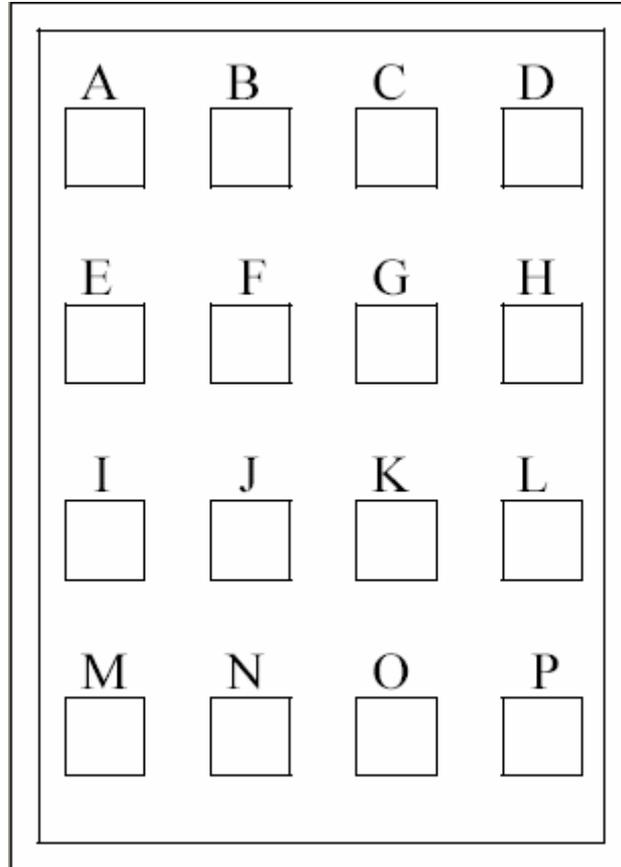

Figure 1: Sample symbol page

The Calibration page has the same format as the symbol page but there are no symbols and each of the boxes contain a smaller box with dotted **X** mark. The writer is required to trace the dotted **X** mark in all of the boxes. A sample calibration page is shown below (Figure 2).
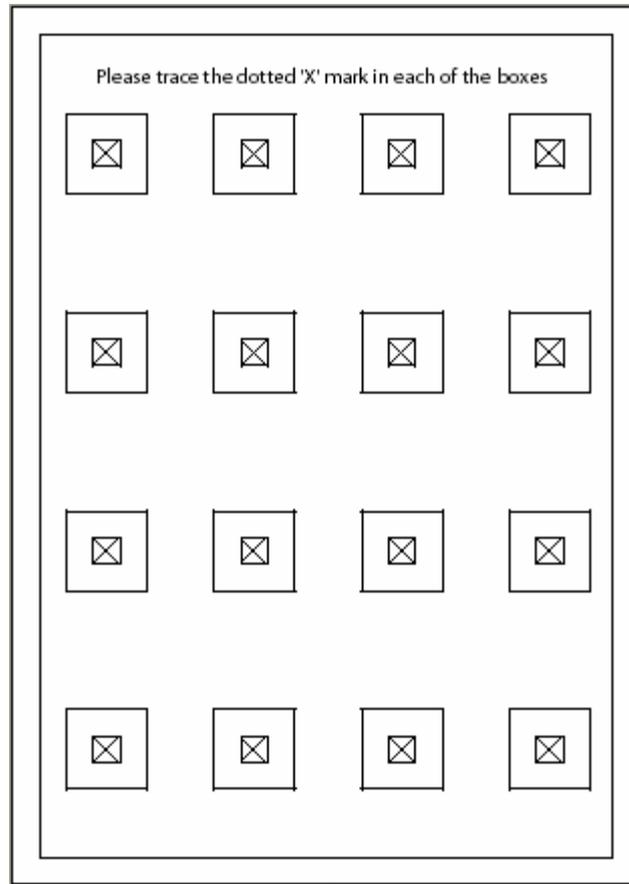
Figure 2 : Sample calibration page

The template forms (designed for A5-sized DigiMemo) can be found under the directory `templateformsandconfigfile/`.

See Appendix for the steps for creating symbol forms from the template form provided, using Adobe® Form Designer 7.0.

3-2-1-2   Printing the forms

Printing the forms must be done carefully. The printing options must be set in such a way that there are no changes in the size, shape, angle or relative positions of the boxes due to printing. See Appendix for printer settings.

3-2-1-3   Setting up and using Digimemo devices

The Supervisor of data collection process should undertake the following steps:

1.   For A5 Digimemo, make a booklet of the blank forms, printed for one or more trails. Attach the booklet to the device as mentioned in the device manual. In the case of A4 Digimemo, clip the blank forms to the writing pad.

2. The *digimemo-dct 1.0* requires the `info.cfg` to be placed in the `DMEMO-M` folder of the device. It contains the following device-specific information.

Key	Description
Digimemold	A unique ID assigned by the supervisor.
XUnitsPerInch	Dots per inch of the device in X-direction.
YUnitsPerInch	Dots per inch of the device in Y-direction.
PointsPerSecond	Device resolution.
XOffset	X co-ordinate of the top left point of the margin in the printed form in device units.
YOffset	Y co-ordinate of the top left point of the margin in the printed form in device units.
InkFileNamePrefix	Prefix string of the ink file name before the page number. For example, if the device generates file names of the form PAGE_001.DHW, prefix is PAGE_
InkFileNameSuffix	Suffix string of the ink file name after the page number. For example, if the device generates file names of the form PAGE_001.DHW, suffix is .DHW

3. The writer can give his/her trial starting with the calibration page followed by the symbol pages. Press the '▽' button at top left corner of the device, after completing each page.

See Appendix for sample configuration file for the Digimemo device.

NOTE:

1. Refer to the device user manual to find the following:
   - Maximum number of pages allowed in a booklet
   - XUnitsPerInch
   - YUnitsPerInch
   - PointsPerSecond
   - InkFileNamePrefix
   - InkFileNameSuffix

2. The XOffset and the YOffset can be measured by placing the pen over the writing pad.

3-2-2 Collecting writer information

On running `digimemodct.exe` the following dialog box (Figure 3) appears.

Figure 3 : Startup screen

Click **Writer Information,** and the Writer Information dialog box (Figure 4) appears. The dialog box enables the supervisor to enter the writer's personal information like Name, Age, Profession and information about his familiarity with the device and the script.

Figure 4 : Writer Information dialog box

1. The Writer Information dialog box essentially collects the following information from the user:

Field	Type	Description
Writer's name	Mandatory	Name of the writer.
Writer's number	Automatically generated	An auto generated number given to every writer who takes the trial.
Age	Mandatory	Writer's age.
Gender	Mandatory	Writer's gender.
Hand	Mandatory	Whether the writer uses left hand or right hand for writing.
Skill with device	Mandatory	If the writer is familiar with the usage of the device with which the data is collected.
Region	Mandatory	Writer's region.
Profession	Mandatory	Writer's profession.
Style of writing	Mandatory	Writer's style. (Normal/Cursive/Mixed/Printed)
Education Level	Mandatory	Writer's education level.
How often do you write in this script?	Mandatory	Writer's frequency of using the particular script.
Does you profession involve writing in the particular script?	Mandatory	Does the writer's profession involve any amount of writing in the particular script in which he/she is asked to give handwriting samples.
Digimemo Number	Mandatory	ID of the Digimemo device found in the device config file "info.cfg".
Start Page Number	Mandatory	The page number of the first digital page of each trial.

2. To reset the writer information in all the edit boxes, click **New Writer**.

3. To delete the user information, click **Delete**.

⚠ **WARNING:** Clicking **Delete** removes the trial information along with the writer's personal information.

4. Setting the **WriterId** loads the dialog box with the information of the corresponding user, if previously entered. This enables the supervisor to edit the writer information any time later.

5. To save the writer information, click **Save**.

6. To enter a trial entry, fill in the ID of the device being used for the trial and the first digital page number of the trial in the edit boxes provided and click **AddTrial**. The total number of trials shown above the edit boxes is incremented on successful entry. Else, an error message appears.

7. Use **NextTrial** and **PreviousTrial** buttons to browse through the trials given by the writer.

8. To remove a trial entry, navigate to the target trial and click **RemoveTrial**.

3-2-3   Setting the configuration to run the tool

3-2-3-1   Configuration file contents

The configuration file, `dct.cfg`, contains the following information:

Key	Description
NumShapes	The total number of symbols.
CollectionUnit	The type of symbols being collected e.g. CHARACTER, WORD etc.
DeviceType	The type of device e.g. Digimemo A502
Language	The language for which handwriting samples are being collected e.g. English , Hindi, Telugu etc.
Script	The script for which handwriting samples are being collected e.g. Devanagari, Tamil
DeltaSize	Difference in size between Real and Virtual box, in millimeters.
Real box co-ordinates	The coordinates of the box within which sample is written, in millimeters.

See [Appendix](#) for a sample configuration file.

3-2-3-2   Settings

The supervisor of data collection process should undertake the following steps:

1. Click **Configuration Settings** on the main dialog box and the Login dialog box appears.

2. On entering **123** as the password, the Configuration Settings dialog box (Figure 5) appears.
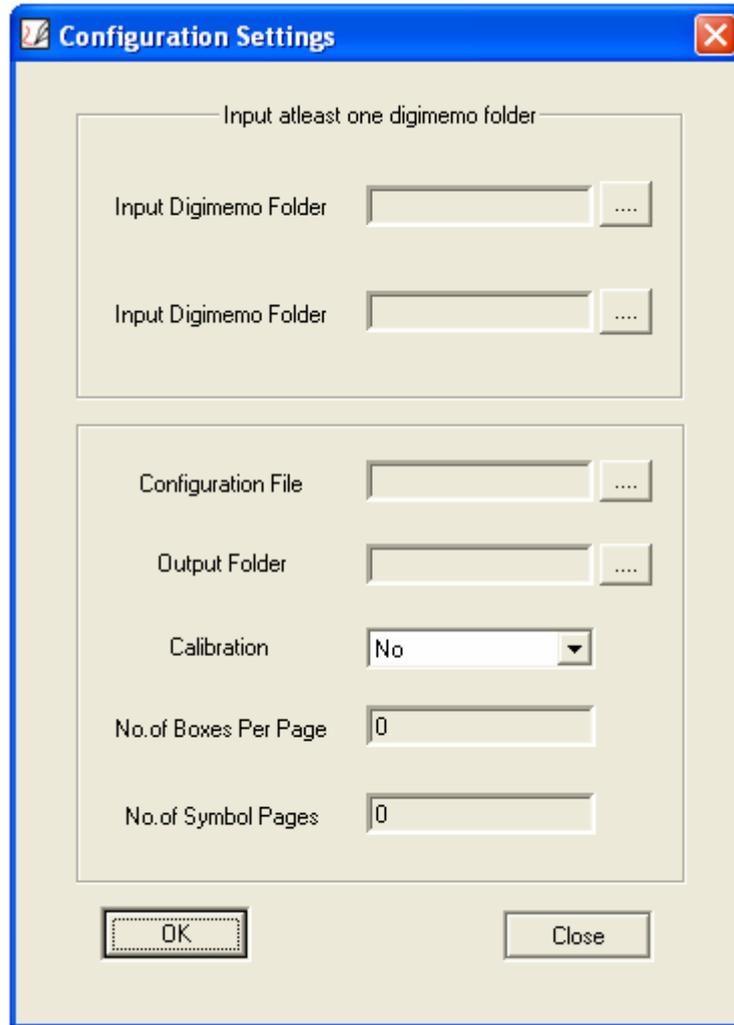
Figure 5: Configuration Settings dialog box

3. In the Configuration Settings dialog box, provide the following information:

a. Select one or two (at least one) digimemo folder(s) which contain the handwriting samples.

Select the configuration file (described in section 3-2-3-1). You may have to edit the sample configuration file, `dct.cfg`, provided under the directory `digimemodct1.0/`. The number of boxes per page and the total number of pages in a trial are calculated from the information provided in this file, and are shown in the edit boxes as read-only fields.

b. Select the folder in which all the collected output data should be stored.

c. Choose **Yes,** if there is a calibration page before the symbol pages, or **No** otherwise.

d. To save all the configuration settings, click **Ok**. The configuration settings are saved in a text file, called `log.txt,` in the current working directory.

e. Finally, to run the tool, click **Run** on the main dialog box.

3-2-4 Verifying the collected data

1. The collected handwriting samples can be found in the output folder specified while configuring the tool.

2. A folder is created for every user, with the name `usr<user ID>`. In this folder, for every sample of a symbol written by the user, a *UNIPEN* file is generated.

3. Every *UNIPEN* file is named using the user number, trial number and the symbol ID. (`OUTPUTFOLDER/usr<user ID>/<symbol ID>t<trial ID>.txt ...`). The order of Symbol IDs is consistent with the order of their corresponding box IDs in the configuration file.

   For example: `D:\data\usr0\000t01.txt`

4. The *UNIPEN* ink file contains the user information, followed by the pen traces captured between a pen-up and pen-down for all the strokes that make up the ink sample.

5. The user information is also stored as a text file named `usr<user ID>.txt`, under the `usr<user ID>` folder.

3-2-5 Troubleshooting

1. The full dialog box of the application cannot be seen - The monitor resolution should be increased to a minimum of 1024x768 pixels.

2. The error message "Error in Calibration page. Ignoring a trial of User<num>" appears on running the tool.

   a. Make sure the first page of the trial is the calibration page

   b. Check if the cross marks in the calibration page is recorded properly. In case of aberrations, the writer has to fill the calibration page again.

4 Appendix

4-1 Steps for creating custom symbol forms from the template form

To create a custom symbol form from the template form, follow these steps:

1. Open the template form (sampleformsandconfigfile\templatesymbolpage.xdp) in Adobe® Form Designer 7.0.

2. Select appropriate font and size in the toolbar and fill in the symbols in all the text controls.

3. Use the **'PDF Preview tab'** to see the preview. This is necessary especially if you are using Indian language fonts to make sure they will be printed properly.

4. Save as PDF.

Make as many copies of the symbol template as needed.

Alternatively, user can design his own forms using Adobe® Form Designer or any other form design tool.

4-2 Printer settings for printing a form

For an HP LaserJet printer, do the following:

1. Uncheck the following options:
   - **Collate**
   - **Shrink Over sized pages to paper size**
   - **Expand small pages to paper size**

2. Check **Auto rotate and center pages**

3. Click **'Properties'**

4. Set paper size to A5 (or A4 for A4 sized Digimemo). Select **Effects** tab.

5. Set Print Document on 'A5' (or A4 for A4 sized Digimemo) and Uncheck **Scale to Fit**

4-3 Sample configuration file for the digimemo-dct 1.0

A sample configuration file, `dct.cfg`, looks as follows:

NumShapes = 150

CollectionUnit = CHARACTER

DeviceType = Digimemo A502

Language = Telugu

Script = Indic

 # Format: DeltaSize = {deltaX,deltaY}

DeltaSize = {7.5, 7.5}

 # Format: BoxId[id] = {topLeftX, topLeftY, X_width, Y_width}

BoxId0 = {6.35,19.05,19.05,19.05}

BoxId1 = {41.274, 19.05, 19.05, 19.05}

.

.

.

BoxIdN = {111.125, 161.925, 19.05, 19.05}

NOTE:

1. The values for DeltaSize and the real box co-ordinates must be in millimeters.

2. The real box co-ordinates are with reference to the origin at the top-left point of the margin (not the form)

The configuration file `templatedct.cfg` for the template forms can be found under the folder `templateformsandconfigfile/`.

4-4 Sample configuration file for Digimemo device

Typically, `info.cfg` looks like this:

DigimemoId = 0

XUnitsPerInch = 1000

YUnitsPerInch = 1000

PointsPerSecond = 125

XOffset = 310

YOffset = 8220

*InkFileNamePrefix =PAGE_*

InkFileNameSuffix =.DHW

4-5 Sample output ink file (UNIPEN format)

.VERSION 1.0

.HIERARCHY CHARACTER

.COORD X Y T

.SEGMENT CHARACTER 0 OK "0"

.H_LINE 749 1499

.V_LINE 249 999

.X_DIM 749

.Y_DIM 749

.X_POINTS_PER_INCH 1000

.Y_POINTS_PER_INCH 1000

.POINTS_PER_SECOND 125

.COMMENT CALIB X: -11 Y: -104

.WRITER_INFO

Name: V JAGADEESH BABU

WriterId: 0

Hand: right

EducationLevel: post graduate

Gender: male

Profession: student

Region: visakhapatnam

Age: 23

DeviceType: digimemo

SkillDevice: good

Language: telugu

Script: indic

Native: true

Style: discrete

UsageFreq: once a month

Proficiency: good

Date: 2006-11-21

WritingInProfession: no

.PEN_DOWN

621 981 0

613 982 0

605 983 0

597 985 0

590 986 0

583 988 0

576 989 0

569 990 0

563 991 0

557 994 0

551 997 0

.PEN_UP

4-6 References

[1] Acecad Digimemo: **http://www.acecad.com.tw**

[2] HMM-based Online Handwriting Recognition System for Telugu Symbols, Jagadeesh Babu V., Prashanth L., Raghunath Sharma R., Prabhakara Rao G.V., Bharath A, 9th International Conference on Document Analysis and Recognition (ICDAR 2007), Curitiba, Brazil, Sept 23-26, 2007 **http://www.hpl.hp.com/india/documents/papers/TeluguSymbolRec_ICDAR_2007.pdf**

[3] A standard format from the International Unipen Foundation (**www.unipen.org**) to store on-line handwriting data (as digital ink) and its annotations.