OPTICAL PROCESSING OF QUESTIONNAIRES

Pavel Číp, Tomáš Babinec

Doctoral Degree Programme (2), FEEC BUT

E-mail: {xcippa00, xbabin01}@stud.feec.vutbr.cz

Supervised by: Karel Horák

E-mail: horakk@feec.vutbr.cz

Abstract: This paper deals with automation of processing and optical recognition of questionnaires, forms or other paper types filled by people. Questionnaire processing done by human is very time-consuming and after a long time, human brain and eyes are very tired. This leads to misinterpretation and errors. Optical recognition is more effective and faster. But first of all it is important to specify to the algorithms the structure of the evaluated form using basics components - preferably in XML code. This code says to the program, where the main parts of the forms are and which method is usable for its recognition. This paper describes proper ways for form designs and basic steps for detection of answers.

Keywords: computer vision, image processing, forms recognition, test machine, exam forms, GUI

1. INTRODUCTION

The human can't focus on the stereotypic activity for a long time. A long term processing of filled forms, answer sheets, questionnaires or surveys, by single human leads to increasing error rates by eyestrain and fatigue of brain. In some cases it is necessary to evaluate large number of papers in short time. For that it is required to create a system for automatic optical recognition of forms. This system must be precise, smart, reliable, and faster than human and has to utilize methods of computer vision. It should enable flawless processing of all specified (learned) form types, evaluate their values and create report about the results or compare answers with correct solutions.

2. FORM CONCEPT

Questionnaires with information for mass processing have the same design. To simplify the detection it is useful to define basic form sheet parts and inform the application which processing method should be used. In next section we recommend specific parts in forms. In **section 2.6** is the example of a very simple form definition by XML code which is shown in **Figure 2** (right side).

2.1. CALIBRATION MARKS

In automation recognition it is important to have form in right position. Stock of filled forms is putt into automation equipment which can get sheet separately and scanned into an image to the hard drive in PC. No one can guarantee that papers are scanned in the same angle of rotation or position. Therefore it is necessary to put some calibration marks on form's corners. In practice most useful shapes are crosses, rectangles, circles, triangles or lines which describes unique place in form.

2.2. HEADER / FOOTER

Header and footer included some information about faculty, company trademark and logo. It is possible to mention rules about exams here, time interval, sample of right answers' crossing, etc. This part is only information for user. If there is a form with more than one page, it is necessary to mark these pages by numbers (usually by number with its binary code for easier optical recognition).

2.3. PERSONAL DATA AREA

This part is filled by personal data of user, student or interviewee. There are blocks for filling names, gender, actual date, name of subject, version of exam, etc. Some areas from this block could be recognized by OCR algorithms for easier data entries. OCR is the shortcut for Optical recognition of character mechanism in computer vision.

2.4. IDENTIFICATION DATA AREA

Identification part is similar to personal, but there are more blocks for optical recognition. The main reason for using this area is coherence to user. For example, in universities every student has own unique identification number. In final exam, it is very important to know this student's number for correct assignment with the university mark system. User, who filled this exam form, could write this UID number to the special block and it is recognized by OCR after.

If we can guarantee that this form will be received by the appropriate student, we can pre-print his UID here. There are lots of options for coding this number – for example in EAN barcode, QR code or another type of 2D code, or simple print in normal font useful for OCR (See **Figure 1**).

a) b) c) d) e)

Figure 1: Example of number coding: a) 7 segment for filling with examples; b) QR code; c) data matrix; d) EAN barcode; e) code PDF417

2.5. ANSWERING DATA AREA

This part is the most important on the form and has variable size opposed to the other. That size depends on number of questions. Answering part can be formed by table with columns and cells. Checkboxes for answers has shape of rectangle or circle. By crossing this part user chooses his answer. If he makes a mistake, it is able to fill this cell and put correct cross on right position. In this case it is necessary to attach the paper with questions. Another type of form could include questions directly in this answering area. There are questions with options of answers and interviewee (student or user) must cross the answer right here.

2.6. DEFINITION OF FORM

The formal form sheet description is stored is a special XML structure (see Figure 2). For every form part it specifies its size, position, type and recognition methods for successful processing.

```
<?xml version="1.0" encoding="UTF-8"?>
<form id="FE_ACV" version="1.2" size="1000,700" scale=”1”>
    <title text="Final Exam - Application of Computer Vision" />
    <calibrate_points
        <point id="p1" type="circle_hole" position="1,1,50,50" file="" size="30" />
        <point id="p2" type="circle_fill" position="640,1,700,50" file="" size="30" />
        <point id="p3" type="circle_hole" position="1,930,50,1000" file="" size="30" />
        <point id="p4" type="circle_hole" position="640,930,700,1000" file="" size="30" />
    </calibrate_points>
    <areas>
        <area id="header" type="header" method="ignore" position="50,12,640,95" />
        <area id="UID_7digit" type="identification" method="digit_7" position="70,300,250,346" />
        <area id="answer_table" type="answers" method="answer_cross" position="80,431,630,753" props=”column”/>
        <area id="answer_line" type="answers" method="answer_line" position="72,770,650,923">
            <block id="q1" type="question" position="" />
            <block id="a1" type="answer" position="88,832,120,920" method="checkbox" />
        </area>
    </areas>
</form>
```

Figure 2: Example of XML definition with basic parts and methods.

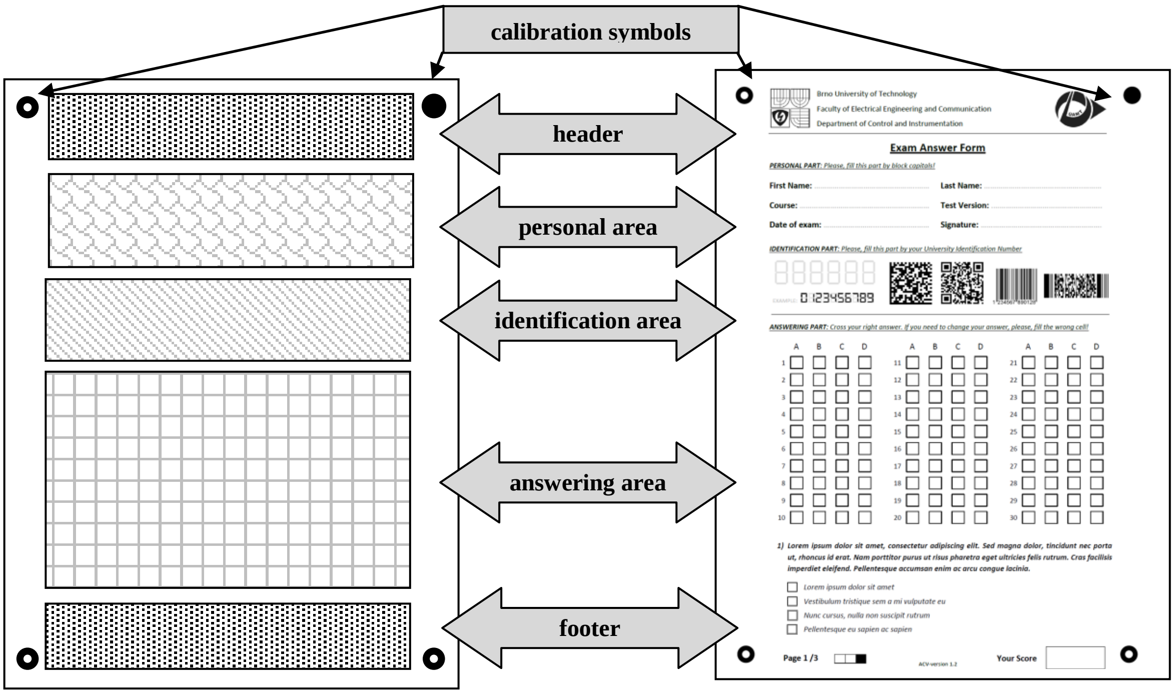

Figure 3: Form concept including basic parts (Right: testing design of exam answer form)

3. STEPS OF RECOGNITION

3.1. NORMALIZATION TO RIGHT POSITION

Every scanned form is transformed into grayscale image. After that three or four calibration symbols are localized. Methods for finding them are working on principle of transforming image into binary with labeling included objects. From knowledge of view calibration marks found there the best match or shape correlation with the same properties like extent, similarity with convex area or rotation. If we use four symbols in the corners, we must guarantee that one symbol is different from the others. It is useful for detection of rotation around 180°. After finding these symbols, a form is transformed with projective or with simple affine matrix (in case of three symbols). We suppose that forms are scanned in automation machine sheet by sheet with pressure roll, so only rotation and position are deformation which may occur.

For affine transformation, the value of the pixel located at $[\hat{x}, \hat{y}]^T$ in the output image, is determined by the value of the pixel located at $[x, y]^T$in the input image. The relationship between the input and the output point locations is defined by the following equations [2]:

$$\begin{aligned} \hat{x} &= xh_1 + yh_2 + h_3 \\ \hat{y} &= xh_4 + yh_5 + h_6 \end{aligned}, \qquad (1)$$

where $h_1, h_2, \dots, h_6$ are transformation coefficients.

For projective transformation, the relationship between the input and the output points is defined by the following equations [2]:

$$\hat{x} = \frac{xh_1+yh_2+h_3}{xh_7+yh_8+h_9}, \qquad \hat{y} = \frac{xh_4+yh_5+h_6}{xh_7+yh_8+h_9}, \qquad (2)$$

where $h_1, h_2, \dots, h_9$ are transformation coefficients.

3.2. IDENTIFICATION USER

In personal and identification part (mentioned in section 2.4, 2.5) are filled important data. From XML description we know, what data we are finding, and which method we must use for recognition. The identifying of UID is possible by intermediate OCR technique [1], [2]. We use extern methods and algorithms because some of them are protected by patent and their processing (decoding) is not part of this paper. These extern methods give us information in string or integer format for next processing.

3.3. FINDING ANSWERS

Grayscale image of form is converted into binary by adaptive threshold. From definition of form design in XML, we know, where is the answer's area placed. There is a definition which type of answers is there too. We suppose two kinds of that – first is table with columns and cells for crossing, second are questions and answers together in sequence. In first case we take this part and count black pixels (in inverse binary image, white is the pixel of cells) in vertical projection with (3) on each column of the image line by line. [2]

$$SUM_{vertical}(x) = \sum_{y=1}^{H} \begin{cases} 1 & if\ im(x,y) = pixel\ of\ cells \\ 0 & otherwise \end{cases} \qquad (3)$$

Where im is the image, H is the height of image and x is the number of actual column.

If the count of black pixels in a column is equal or similar to zero, the column with checkboxes is segmented. From knowledge that the space between groups of column is twice bigger than space between cells, we can segment this block of answers. After that we have three separately images.

The same projection, but in horizontal direction, is used for these three columns for separating rows. At finally the method of vertical projection is used again for segmentation each checkbox (cell). Result of this projection is obvious on **Figure 4**.

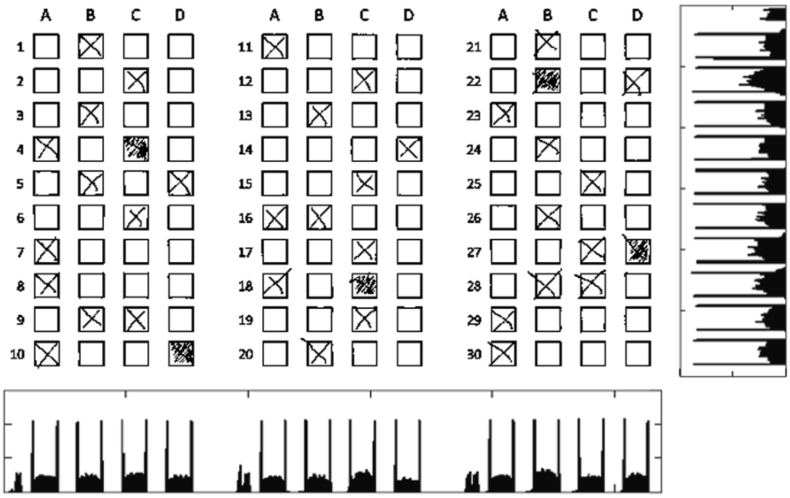

Figure 4: Example of pixels projection of answering area; bottom: vertical projection; right: horizontal projection of last column

In second case (form with questions and answers), previous steps are similar. From form design we know where answers (check boxes) are placed. Again by combination of vertical and horizontal projection we could find the answers.

For recognition, if the box is marked by cross, filled or left blank, we use template matching with model of cross with using knowledge of percent filled area [4]. The result of recognition checkboxes is shown on **Figure 5**. Methods take into account that one question could have more than one right answer.

This result is saved in structure with UID of student, user or with the other information from form for processing by next automatic system. This part is not point of view of this paper.

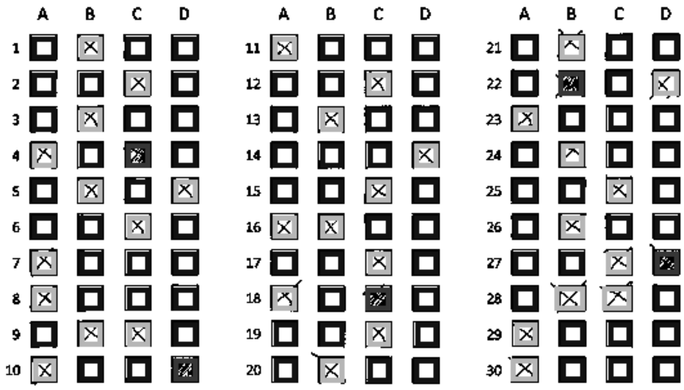

Figure 5: Result of detection crosses answers with filled blocks (grey: cross detected, dark grey: filled box, black: blank box)

4. CONCLUSION AND FUTURE WORK

The proposed method of optical recognition and processing of forms was written in C++ with Computer Vision Library from Intel OpenCV [3]. Stock of forms (design shown on **Figure 3**) with 30 simple questions was filled by students. They had answers and write their UID in 7 segment style (see **Figure 1a**). This part was recognized by Neural Network. Every form was scanned by automatic system and processed in PC (AMD, 1.8 GHz, 1GB RAM). Processing time was about 0.8 seconds with success more than 99.9%. Errors were caused by user that not filled the wrong answer correctly or used a faint color pen.

We defined some methods and principles for processing these forms. It is useful tool for evaluation large amounts of data. These methods are working with most popular type of forms. There is no problem to define basics' parts in unknown forms (by XML). In future work we would like to improve methods for recognition of user data in personal part by OCR methods, recognition of handwriting characters, etc.

5. ACKNOWLEDGEMENT

This work was supported by grant „Modern Methods and Approaches in Automation“ from the Internal Grant Agency of Brno University of Technology (grant No. FEKT-S-10-12), by Ministry of Education, Youth and Sports of the Czech Republic (Research Intent MSM0021630529), Grant Agency of the Czech Republic (102/09/H081 SYNERGY - Mobile Sensoric Systems and Network) and by Brno University of Technology.

REFERENCES

[1] Safraz, M., *Computer-Aided Intelligent Recognition Techniques and Applications*, John Wiley & Sons, Ltd. 2005. 519 p. ISBN 978-0-470-09414-3.

[2] Jan J., *Medical Image Processing Reconstruction and Restoration: Concepts and Methods.* Boca Raton: Francis & Taylor 2006, ISBN 0-8247-5849-8

[3] Bradski G., Kaehler A., *Learning OpenCV*. Sebastopol: O'Reilly Media, Inc. 2008, ISBN: 978-0-596-51613-0

[4] Bishop M. C., *Pattern Recognition and Machine Learning*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.