Mixed-Initiative Approach to Extract Data from Pictures of Medical Invoice

Seokweon Jung Kiroong Choe Seokhyeon Park Hyung-Kwon Ko Youngtaek Kim * Jinwook Seo †

Department of Computer Science and Engineering, Seoul National University, Seoul, Republic of Korea

ABSTRACT

Extracting data from pictures of medical records is a common task in the insurance industry as the patients often send their medical invoices taken by smartphone cameras. However, the overall process is still challenging to be fully automated because of low image quality and variation of templates that exist in the status quo. In this paper, we propose a mixed-initiative pipeline for extracting data from pictures of medical invoices, where deep-learning-based automatic prediction models and task-specific heuristics work together under the mediation of a user. In the user study with 12 participants, we confirmed our mixed-initiative approach can supplement the drawbacks of a fully automated approach within an acceptable completion time. We further discuss the findings, limitations, and future works for designing a mixed-initiative system to extract data from pictures of a complicated table.

Index Terms: Human-centered computing—Human computer interaction—Interactive systems and tools—User interface toolkits;

1 INTRODUCTION

Supplier invoices are important data for insurance companies when they decide how much they should reimburse the customer for medical expenses. Accounts Payable teams are in charge of managing invoice data. They review, record, and back up those invoices transmitted to the companies [14]. Customers usually submit invoices in pictures taken using their smartphones, and thus extracting data from images is a crucial task for the team. The extracted data should be exact in amount and contain no errors.

As manual transcription of the invoice is a very time-consuming and labor-intensive task, many researchers have tried to automate the procedure [3, 13, 14]. One of the most successful methods uses coordinates information explicitly. This method requires the image to be straight and aligned, and this is usually the case for scanned or faxed images. A growing number of customers, however, are submitting smartphone pictures of invoices. Since such pictures do not satisfy the key assumptions made for scanned or fax images, existing invoice processing methods are not feasible for extracting data from them.

In this paper, we propose a mixed-initiative approach to extract data from pictures of medical invoices. First, we developed a fully automated pipeline that tackles the core difficulty in recognizing medical invoices, such as finding a data table from images and recognizing template variation. Then, we observed that the majority of errors contained in the pipeline could be easily corrected by human interactions. We developed a mixed-initiative pipeline, in which users can discover errors of each stage with visualizations and correct them with interactions. A user study involving 12 participants confirmed that our interactive pipeline can capture about 97% of relations in invoice images within comparable time and with less human efforts.

In summary, our study makes the following contributions:

- We investigated the characteristics of the real-world dataset containing smartphone pictures of a medical invoice and defined core difficulties in recognizing them.
- We enhanced the existing table structure recognition algorithm with heuristics and devised the fully automated pipeline.
- We introduce the mixed-initiative approach as an alternative solution to the problem by inviting users to correct intermediate machine output at a low cost.

*e-mail: {swjung, krchoe, shpark, hkko, ytaek.kim}@hcil.snu.ac.kr

†e-mail: jseo@snu.ac.kr

2 RELATED WORK

When automating invoice processing, invoice recognition should come first. According to Taylor et al. by analyzing the templates of a document we can improve the performance of data extraction from the document [15]. Therefore, most studies on invoice recognition have utilized the fact that most invoices have a proper template according to their purpose. Hamza et al. used CBRDIA (Case-based Reasoning for Document Invoice Analysis) based on case-based reasoning to check if a similar document had already been processed [8]. Schulz et al. proposed the use of the self-teaching mechanisms of the state-of-the-art invoice analysis software. They used transfer learning to catch key features of templates, which made a significant improvement to the initial recognition rates [13]. Sun et al. detected contour boxes in an image and used them to match the template with a pre-trained template dictionary of Chinese invoice [14]. Previous methods showed successful performances, however, they were not applicable for our dataset, Korean medical invoice.

Conventionally, computer vision techniques are used to detect and analyze tables [12]. Recently, models using machine-learning algorithms have made huge performance improvements. CascadeTabNet [11] proposed a Cascade mask Region-based R-CNN model to detect tables and found intersections with computer vision method to manage structure recognition. Zhong et al. proposed the EDD(Encoder-Decoder-Decoder) model to handle the full process of table recognition [17]. Some researchers tried to adopt interactions and mixed-initiative approaches to handle errors of fully automated approaches. Koc et al. proposed the use of XLindy [10], a novel method for layout inference and table recognition in a spreadsheet. Hoffswell et al. [9] proposed a lightweight interaction technique to repair tables extracted from PDF documents on mobile devices.

Although these studies on table recognition have shown good performance, their coverage is limited to data of computer-generated images with perfect alignment and zero distortion, such as spreadsheet, TableBank, and PubTabNet[1]. Also, the structure of tables included in the dataset is limited to a very simple, single table. Our study differs in that it targets real-world smartphone pictures of invoices, including tables with complex structures, misalignment, and distortions.

3 BACKGROUND

Many companies are interested in lowering an unnecessary burden on labor and material resources needed for invoice identification [14]. Therefore, we interviewed a manager affiliated with the invoice processing team and personnel belong to the big data team in a major insurance company. Next, we formulated tasks and the main obstacles hindering them.

3.1 Domain Expert Interview

According to the interviewee, there are three types of invoice images in the company's database; 65% of images are directly scanned by employees within the company, 20% of images are faxed, and the remaining 15 % are smartphone pictures. Every submitted image follows two stages: automated OCR system and human review. Because scanned and faxed images contain well-aligned invoices, the existing OCR system utilizing the coordinates of the target information shows moderate accuracy for these two types of documents (80% and 50%, respectively). For smartphone pictures, however, it fails with almost 0% accuracy. Therefore, employees should type all misread letters in the human review stage.

[1] github.com/{doc-analysis/TableBank, ibm-aur-nlp/PubTabNet}

Figure 1: The standard invoice template consists of 4 different sub-tables: a1) Header with personal information, a2) Mid-left with price details, a3) Mid-right with a summary of the mid-left table, a4) Footer with hospital information and manual.

Every recognition result should be checked by staff as no errors are tolerated due to the nature of the task; The invoice contains information about the amount of money the company has to pay. Simply adding one more zero digits at the end of the amount can cause a big loss to the company. Currently, the company was using a custom spreadsheet format to manage the data extracted from invoice images. Due to the variations of Korean medical invoices, about 10% of the entries were missing and being accumulated in the "Other than Standard Items" field. For smartphone pictures of invoices that do not have initial recognition results, staffs enter all entries in the spreadsheet by typing manually and calculate the sum of non-standard price items. This is a very time-consuming and error-prone way of working. When the authors replicated the procedure to make ground truths of our dataset, it took three or more minutes on average to handle a single invoice image.

3.2 Problem Statement

Korean medical invoices follow the official standard template(Figure 1), which is occasionally updated every few years and is declared by the government. While the standard template is used among most of the hospitals, some hospitals have their own customized format for the following reasons. First, the previous template might not have been replaced yet. Second, hospitals might have reordered the items for their convenience. Last but not least, they might have attached other irrelevant documents on the same page. Examples of customized templates are given in supplementary material.

As Figure 1 shows, the standard template has a complicated structure that consists of four simple sub-table sections: the header at the top, price details on the middle-left, the summary on the right, and the footer at the bottom. The primary target the company wanted to recognize was the price details section since it contains the key information. The section includes a single two-dimensional table whose numeric values are explained by corresponding row and column header cells. Each row header cell defines the individual charge item, such as "Hospitalization" and "Medical Treatment", and each column header cell describes whether the patient or the health insurance should pay the fee.

After receiving 400 pictures of medical invoices: 200 scanned, 100 faxed and 100 smartphone-taken, which are collected and anonymized by the staff of the company, we set up tasks to solve the problem after iterative design process with domain experts of the company.

- **T1:** Detect the location of invoices from smartphone pictures.
- **T2:** Identify the target area where target information is located.
- **T3:** Classify templates of each detected invoice.
- **T4:** Extract data from the target area without a critical loss of information.

Then, we analyzed the main issues to complete each task.

- **O1: Poor image quality.** Some images have a low resolution, which impairs the overall recognition procedure (T1–4).
- **O2: Poor alignment.** Depending on the camera perspective, invoices in pictures might not be aligned vertically or horizontally and can be seen as distorted. Furthermore, invoices might be randomly positioned inside the background (T1, T2).
- **O3: Physical distortion.** In some cases, invoices are folded or crumpled, leading to curved or folded line segments that disrupt conventional line detection algorithms. The light condition also affects the image; making certain parts of the image darker or brighter (T1, T2).
- **O4: Dense template with distractors.** Because there are several independent sections on the same page, the primary section is condensed in the middle area and surrounded by other similar templates (T3, T4).

4 Automated Pipeline

First, we describe the fully automated pipeline, which resolves the core difficulties in recognizing pictures of medical invoices. Based on the stages of table recognition by Gobel et al. [7] and tasks we defined in the previous section, we designed a five-stage pipeline to recognize a table and extract data from it: invoice detection, target area identification, structure recognition, optical character recognition, and template matching. The whole process is done after the preprocessing stage, which includes binarization and dilation. Some images dropped out in this preprocessing stage, because of their poor quality resulting from strong light reflection or low resolution (O1).

4.1 Invoice Detection

This stage is dedicated to locating the invoice in the target image (T1). We tested the trained models of ICDAR 2019 Competitions on Table Detection and Recognition [6], but the results were not satisfactory due to the following obstacles (O1, O2, O3). Instead, we decided to use a conventional computer vision approach, which can benefit from the dense structure of our dataset (O4). By dilation, dense lines become a single connected chunk, the bounding box of which can be regarded as the location of the invoice.

4.2 Target Area Identification

After detecting the invoice from the image, identifying the target area became our next goal (T2). In our case, the primary interest is on the price detail section that lies at the center of the invoices, which has about 27 rows and 5 columns by default (Figure 12) We have attempted to address this problem by exploiting the characteristics of the document that we pointed out as an obstacle: the dense structure of the table (O4).

According to the previous research, CNN is superior in capturing visual features such as line segments for intersections [4]. Therefore, we attempted to make use of CNN to find out four lines of a quadrangle. In the same way, we attempted to find out one horizontal line and one vertical line which distinguish header cells and value cells.

We extracted horizontal and vertical line segments from the image. Hough line detection [1] is a common method used to detect lines from images. However, it has a limitation that it did not work properly on distorted images (O3), because it only detects straight lines. So we used a heuristic computer vision method to detect possibly curved line segments, by detecting paths of continuous black pixels. We selected paths longer than a certain threshold and unioned them. We also unioned the result of the Hough line detection to cope with small gaps as our algorithm does not tolerate any disconnection.

Stage	Initial state	Invoice detection	Area identification	Structure recognition	Template matching	OCR
Error	0	13	14	6	14	31
Pass	100	87	73	67	53	22

Table 1: Errors counts occurred during the pipeline.

After detecting horizontal and vertical lines, we trained two types of CNN models; The first model is a binary classifier that determines if a small patch of an image contains a certain line, and the second is a regression model that estimates the exact position of the line. For each of the six lines, we trained two models respectively. The random sample consensus algorithm (RANSAC) was used for its robustness to outliers [5]. We trained these models using 100 manually labeled faxed images. Faxed invoice images had a better alignment of invoices on images than smartphone pictures, which helped models to capture important features with fewer images.

4.3 Structure Recognition

This stage aims to split the price details section into two-dimensional grid cells (T3). Because of the dense structure of our dataset (O4), previous table structure recognition methods were not suitable for it. Therefore, we tried to recognize the table structure with a combination of heuristics originates from our observations about invoice template and the classical computer vision approach of previous research [12]. First, the area inside a bounding quadrangle was transformed into a rectangular image using the homography transformation. Then, we detected all the bounding boxes of texts. Finally, we calculated the histogram of x and y coordinates of bounding boxes to determine proper horizontal and vertical borderlines at each local minimum of the histogram. We weighted more on the text boxes in the headers section to ignore the effects of text alignments in cells (O3). Although invoice templates have a few merged header cells, we decided to ignore them because merged header cells always have the same pattern. Thus, the output of this pipeline is the image of all individual cells inside the table and the metadata to identify the type and position of cell images.

4.4 Optical Character Recognition

We used the Clova Deep Text Recognition Toolkit [2], one of the best public optical character recognition (OCR) models for the Korean language, to recognize texts in each cell (T4). Using frequent letters in Korean invoice templates, we generated a synthetic dataset under the transformation of color, position, truncation, and addition of line noises. Then it was used to train two separate models for recognizing Korean and numeric characters, respectively.

4.5 Template Matching

Classify templates of each detected invoice (T3). Since there are 20- to 30-row header cells in a template, it is difficult to read the exact text value from all header cells at the same time. As a practice of case-based reasoning (CBR) [8], we collected over 300-row header templates from other invoice data and used them to select the most similar template with the OCR result. Here, intersection over union (IOU) of length-3 sub-words was used to compare imperfect OCR results with template texts.

4.6 Performance

We investigated the performance of the automated pipeline using 100 smartphone pictures of Korean medical invoices. None of the pictures in the dataset were used during the training of models in the pipeline. Table 1 shows the errors that occurred in each stage.

The quality of the recognition process was evaluated by the proportion of correct entries among all entries. However, naively counting the number of correct cells can be misleading, as falsely adding or removing a header affects all cells and dramatically decreases the overall accuracy. Instead, we defined an entry as a combination of the row header category, the column header category, and a corresponding numeric value. This measure considers a table as a relation rather than a table itself and is less sensitive to structural changes [16]. Also, there are many cells with zeroes in typical Korean medical invoices, so the measure can be unreasonably high by predicting every cell as 0, which is not proper behavior. Thus, we excluded all relations whose numerical value is zero and used the F1 measure to compare two sets. The average F1 score was 0.7, and 32 out of 100 images were perfectly-recognized. We also observed that errors cascaded so that errors made in previous stages affected subsequent stages (Table 1).

5 A Mixed-Initiative Approach

With heuristic algorithms and CNN, we achieved about 70% accuracy in our fully automated pipeline. However, the introduction of heuristic algorithms creates side effects that increase the sensitivity of the system because some important assumptions, such as the size of the threshold used for detecting line segments, must always be satisfied for the heuristic algorithm to work properly. This is where human mediation can be introduced. Once the machine learning algorithm shows moderate or better performance, humans can correct the output of the pipeline to satisfy the assumptions that heuristic algorithms need through only a little interaction.

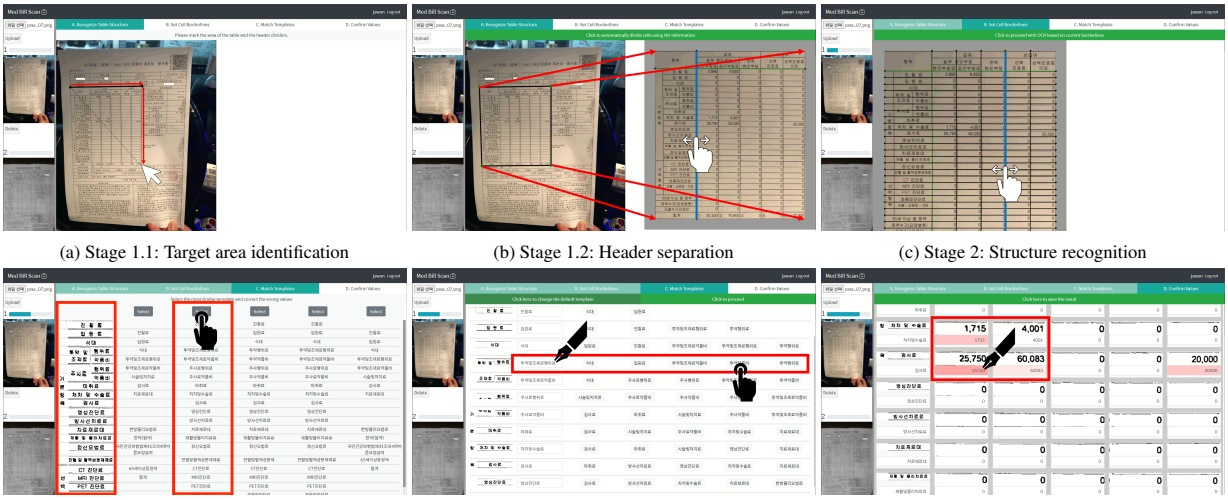
Therefore, we decided to adopt a mixed-initiative approach for our fully automated pipeline to improve performance. The overall flow of the mixed-initiative pipeline (Figure 2) is based on our fully automated pipeline. We tried to supplement the drawbacks of a fully automated pipeline by allowing users to check the intermediate output of each stage with simple visualizations and correct them through simple interactions.

Table detection If users can directly identify the target area, there is no need to detect invoices from images (T1, T2). Unlike CNN models that interpret it as six straight lines, we reverted the concept of a quadrangle (Figure 2a) and lines (Figure 2b) to retain the original meaning and lessen the number of necessary interactions, from 12 clicks for drawing six lines to eight clicks for a quadrangle and two lines. While drawing quadrangle, we visualized connections between clicked points with solid lines and dotted lines for the connection between clicked points and current mouse position. For both stages, we offer zoom and pan for users' convenience.

Structure recognition In the table structure recognition stage (Figure 2c), borderlines detected by the automated module are visualized on the image of the identified target area. Users can click the left button of the mouse to add missing lines or the right button to delete incorrect lines. Users can also drag misplaced lines to modify their location or angle. To minimize users' effort, we provide a multi-line dragging interaction with which users can select multiple lines and move them simultaneously by dragging them.

Template matching A caveat of our template matching algorithm is revealed when the exact template does not exist in the previous cases and there are many similar candidates. Thus, we had users select from among the top five candidates ranked by the number of identical row header items (Figure 2d) by comparing it with the original row header images located at the left. Wrong cells in the template can be modified in the next stage (Figure 2e). Since keyboard typing is a relatively expensive operation, the interface displays the candidate words as buttons after the header cells of each row so that the users can click to modify the values. Candidate words are defined as words that occurred in the same position among similar templates. Users should type raw text only when none of the candidate words match the real text on the image (T3, T4).

Numeric cell correction After revising the header cells, users can also correct numeric cells (T4). To compensate for the case where the number of numeric cells exceeds the perceptual capacity



(a) Stage 1.1: Target area identification (b) Stage 1.2: Header separation (c) Stage 2: Structure recognition

(d) Stage 3.1: Template matching (e) Stage 3.2: Template revision (f) Stage 4: OCR result revision

Figure 2: Figures describing stages of the mixed-initiative pipeline. (a) Users can select the target image at the left component. The selected image appears at the right component. With a mouse click, users can draw a quadrangle surrounding the target area. (b) Users can add one horizontal line and one vertical line respectively to distinguish header cells from others. (c) Based on lines that the automated model predicted, users can modify each line. Also, users can select multiple lines with drag & drop interaction and manipulate them simultaneously. (d) The automated model extracts data from the divided cells with the OCR module. With Case-based reasoning, five candidates of the template are given. Users can choose the most similar candidate among the candidates. (e) After selecting the candidate template, users can modify minor errors of header values. (f) Finally, users modify money amount values and submit the result.

of the users, the system highlights cells that have low prediction confidence (Figure 2f).

6 User Study

We conducted a user study to confirm that our mixed-initiative pipeline is efficient and accurate enough to replace the fully automated method. We recruited 15 participants (6 females, 9 males) who had prior experience with reading and editing data tables in spreadsheet applications. We excluded three participants who had a system issue during the experiment. As a result, we got 12 valid participants (4 females, 8 males) and their ages ranged from 20 to 34 ($\mu = 24.92, \sigma = 3.64$). The participants were informed about our interface and practiced it for about 10 minutes, which was followed by a 40-minute main session and exit interview.

For the main session, we selected 20 pictures of medical invoices from our dataset. We balanced the 20 images from easy instances to hard ones, where the difficulty was assessed by the number of non-zero entries and the level of distortion. Participants were asked to complete as many images in the image recognition pipeline as possible within the time limit of 40 minutes, considering both accuracy and completion time. Participants could choose which image to complete first by using the navigation menu in the interface. The order of provided images was randomized to prevent the users from completing them in order of difficulty.

We monitored the behavior of participants during the main session. In the exit interview, we asked participants about the reasons for their unusual behavior such as using certain interactions very frequently. Finally, we asked participants about the most and least useful features of the proposed interfaces.

Overall, the average precision, recall, and F1 score of the interactive pipeline were 0.97, 0.95, and 0.96, respectively. Also, Figure 3a shows that our interactive pipeline clearly improved the recognition accuracy from that of automated pipelines under statistical significance ($p < .05$), requiring additional 80 seconds per invoice on average(42.94 : 125.73). Figure 3b shows the average time taken to complete each stage and the entire process. On average, each stage took 30 seconds, and the entire process 120 seconds. Measurements of the System Usability Scale (SUS) of our system showed relatively

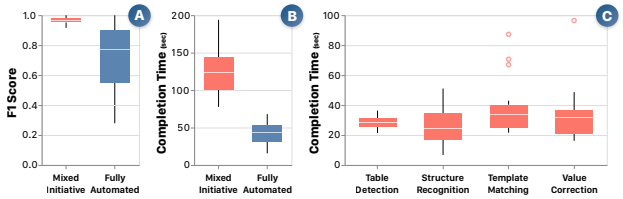


Figure 3: Results of the user study. (a) F1 scores of the mixed-initiative pipeline and the fully-automated pipeline (left). (b) Completion time of each stage of the mixed-initiative pipeline (right).

good usability with an average score of 79.2 points.

7 Discussion

In this section, we discuss findings from the user study and the feedback of the exit interviews. We believe that this discussion will help researchers designing a mixed-initiative data extraction system that targets images including complicated administrative documents, especially in tabular forms.

Usability Although most of the participants agreed that the interaction design of our system was useful, some pointed out several usability issues during the user study. Our system provided a convenient feature that automatically extends a small line segment into a straight borderline. P11 tried to use this feature but stopped using it as the extended lines were often different from what he expected. This could be prevented by showing an auxiliary line that extends a line segment before the user confirms it. P2 pointed out that in the template matching stage, there were some cases where the correct text of a cell image was not in its candidate words but was in those of other cell images. He suggested adding a drag and drop interaction will help in such situations. P4 said that candidate words were so far apart from each other that it caused unnecessary mouse movement. P1 and P7 wanted additional keyboard shortcuts, such as pressing the ESC key to erase the selected line or pressing $(ctrl + z)$ to undo the previous interactions.

Excessive amount of texts Due to the nature of our dataset, interfaces can involve a large number of texts to be displayed on the

screen. In the template matching stage, a large number of texts were displayed to show multiple candidates. P4, P11, and P12 pointed out that it was not an intuitive job to check multiple candidate templates as they were very similar and contained many texts. "I read the leftmost candidate first from the top to the bottom. When I realized it was a wrong template at the end, I had to read another one from the top again, which was tedious for me,", P11 said. Similar feedback was also given for correcting the values of numeric cells. Providing sufficient filtering procedures in advance will alleviate this problem.

Responsibility of each stage In our pipeline, we divided the whole image recognition process into multiple stages. When an exceptional problem occurs, it was ambiguous to determine which stage was responsible for the problem. Some participants made conflicting comments about the OCR result in the value correction section. P7 and P12 said that the OCR model for numeric values showed satisfactory performance. However, P9 and P11 said the exact opposite opinion. These differences are assumed to be attributable to differences of completeness in previous steps. Although our system provides a way for users to retrace previous steps, additional features such as bookmark will be needed to easily identify which previous stage is affecting the current stage.

Model performance The performance of automated models has been pointed out as a limitation of our system. P1, P7, and P12 said that it required too many additional interactions when the prediction models fail. P7 said that "In the template matching stage, default templates worked fine for most cases, but for once it was so inaccurate that it was hard to correct all the mispredictions.". This feedback implies the necessity of additional fallback interaction to cope with untypical situations. Also, because every stage transition requires time for the server to calculate and communicate with the client, repeatedly switching stages increases the waiting time of the user. Long waiting time can cause degradation of efficiency. Some participants became frustrated when they tried the trial & error approach by repeatedly setting borderlines and checking OCR results. This implies that it is recommended to reduce the calculation time of each model to maximize the efficiency of the pipeline.

Towards more general application In this paper, we focused on the specific category of invoices and their variations. Our target template was a Korean medical invoice, which has a table with a complicated and dense structure. As a result, our system has the limitation that it cannot handle other documents with different templates. However, our investigation and deep understanding of documents with these characteristics will help us when we meet similar cases in the future. For future work, we can expand our research to improve it to be applicable to documents with similar templates.

Experimental setting In our experiment, we compared the performance of our two different approaches and proved suggested mixed-initiative approach can improve the recognition performance with fair usability. However, for better comparison, it would be recommended to set the current practice as a baseline. Therefore, in future work, we can design a new experiment with employees of the company as participants, and the current practice as a baseline. Quantifying the result will give us clear evidence that our approach is well designed.

8 Conclusion

In this paper, we proposed a mixed-initiative approach to better perform the data extraction from smartphone pictures of invoices. Through an expert interview, we postulated the problematic situations in an invoice processing task where we constructed an automated invoice processing pipeline by adopting a mixed-initiative approach to improve the performance. Our work provided a practical implication for the mixed-initiative approach and showed that a small human effort via system interaction can improve the performance of data extraction task for pictures taken with smartphones, compared to the fully automated way.

Acknowledgments

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. NRF-2019R1A2C2089062). Jinwook Seo is the corresponding author.

References

[1] N. Aggarwal and W. C. Karl. Line detection in images through regularized hough transform. *IEEE transactions on image processing*, 15(3):582–591, 2006.

[2] J. Baek, G. Kim, J. Lee, S. Park, D. Han, S. Yun, S. J. Oh, and H. Lee. What is wrong with scene text recognition model comparisons? dataset and model analysis. In *International Conference on Computer Vision (ICCV)*, 2019.

[3] S. Bayar. Performance analysis of e-archive invoice processing on different embedded platforms. In *2016 IEEE 10th International Conference on Application of Information and Communication Technologies (AICT)*, pp. 1–4. IEEE, 2016.

[4] M. Dusmanu, I. Rocco, T. Pajdla, M. Pollefeys, J. Sivic, A. Torii, and T. Sattler. D2-net: A trainable cnn for joint detection and description of local features. *arXiv preprint arXiv:1905.03561*, 2019.

[5] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[6] L. Gao, Y. Huang, H. Déjean, J.-L. Meunier, Q. Yan, Y. Fang, F. Kleber, and E. Lang. Icdar 2019 competition on table detection and recognition (ctdar). In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pp. 1510–1515. IEEE, 2019.

[7] M. Göbel, T. Hassan, E. Oro, and G. Orsi. A methodology for evaluating algorithms for table understanding in pdf documents. In *Proceedings of the 2012 ACM symposium on Document engineering*, pp. 45–48, 2012.

[8] H. Hamza, Y. Belaïd, and A. Belaïd. Case-based reasoning for invoice analysis and recognition. In *International conference on case-based reasoning*, pp. 404–418. Springer, 2007.

[9] J. Hoffswell and Z. Liu. Interactive repair of tables extracted from pdf documents on mobile devices. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2019.

[10] E. Koci, D. Kuban, N. Luettig, D. Olwig, M. Thiele, J. Gonsior, W. Lehner, and O. Romero. Xlindy: interactive recognition and information extraction in spreadsheets. In *Proceedings of the ACM Symposium on Document Engineering 2019*, pp. 1–4, 2019.

[11] D. Prasad, A. Gadpal, K. Kapadni, M. Visave, and K. Sultanpure. Cascadetabnet: An approach for end to end table detection and structure recognition from image-based documents. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 572–573, 2020.

[12] V. Ranka, S. Patil, S. Patni, T. Raut, K. Mehrotra, and M. K. Gupta. Automatic table detection and retention from scanned document images via analysis of structural information. In *2017 Fourth International Conference on Image Information Processing (ICIIP)*, pp. 1–6. IEEE, 2017.

[13] F. Schulz, M. Ebbecke, M. Gillmann, B. Adrian, S. Agne, and A. Dengel. Seizing the treasure: Transferring knowledge in invoice analysis. In *2009 10th International Conference on Document Analysis and Recognition*, pp. 848–852. IEEE, 2009.

[14] Y. Sun, X. Mao, S. Hong, W. Xu, and G. Gui. Template matching-based method for intelligent invoice information identification. *IEEE Access*, 7:28392–28401, 2019.

[15] S. L. Taylor, R. Fritzson, and J. A. Pastor. Extraction of data from preprinted forms. *Machine Vision and Applications*, 5(3):211–222, 1992.

[16] R. O. Weber, K. D. Ashley, and S. Brüninghaus. Textual case-based reasoning. *The Knowledge Engineering Review*, 20(3):255–260, 2005.

[17] X. Zhong, E. ShafieiBavani, and A. J. Yepes. Image-based table recognition: data, model, and evaluation. *arXiv preprint arXiv:1911.10683*, 2019.