Assistive Technology for Paper Form Digitization in Resource-Limited Environments

Huguens Jean
University of Maryland, Baltimore County, Baltimore, MD, USA, hjean1@umbc.edu

Timothy Oates
University of Maryland, Baltimore County, Baltimore, MD, USA, oates@umbc.edu

ABSTRACT

In developing countries, people are now more likely to have access to a mobile phone than clean water, making cellular based technology the only viable medium for collecting, aggregating, and communicating local data so that it can be turned into useful information. While mobile phones have found broad application in reporting health, financial, and environmental data, many data collection methods still suffer from delays, inefficiency and difficulties maintaining quality. In environments with insufficient IT support and infrastructure, and among populations with limited education and experience with technology, paper forms rather than electronic methods remain the predominant means for data collection. To meet the digitization needs of paper driven data collection practices, this paper presents the development and study of a system that automatically converts unknown paper form images into text and use the SMS channel to transmit the information to a remote server for digital conversion by humans. We discuss our proposed system architecture for dealing with infrastructure constraints and human resources limitations at the local site level and present a novel framework (RLM) that decomposes the form detection task into retrieving, learning, and matching. Our goal is to significantly decrease the effort and cost of data entry, while maintaining a high level of quality.

Keywords: Paper form digitization, image document retrieval, facility-based device

1. INTRODUCTION

Traditionally, the term assistive technology is used to describe adaptive and rehabilitative devices for people with disabilities, but today when developing world communities dealing with enormous healthcare burdens are left to cope with crippled local information systems, this paper presents the term through a different perspective. Assistive technology enables individuals to perform tasks that might otherwise be difficult or impossible by providing enhancements or adjustments of the interaction methods with the technology required to perform such tasks. This concept is no different for health facilities operating in the developing world with broken information systems and responsible for digitizing large-scale health-related data in rural areas. These technology-related drawbacks can be thought of as handicaps, and our research builds on the insight that low-cost mobile devices, when used as facility-based communication devices capable to enhance the expected processes of the facility independently of human intervention, can have a significant impact on the quality and efficiency of service delivery in rural impoverished communities.

According to the International Telecommunications Union's annual report (ITU, 2012), Measuring the Information Society 2012, there are now 6 billion mobile phones subscriptions globally. Nearly 5 billion subscribers are from developing countries with a penetration rate of about 75%, compared to an Internet penetration of 28% at the same time. Thus mobile phones have rapidly become the preferred communication devices worldwide. As a result, there is an expectation that large-scale data collection and digitization in remote health facilities in the developing world will take place on individually owned mobile devices. Various organizations have employed SMS and electronic forms on JAVA-based phones for entering and sending data to a central server (Dell et al., 2011). However, a quantitative evaluation (Patnaik et al., 2009) of data entry accuracy using low-cost phones in resource-constrained environments has shown that errors rates for electronic forms and SMS may be too high to deploy these solutions in a critical application. Providing local community workers with intuitive smartphones or tablets with custom mobile applications is not feasible for most organizations. In some situations substituting paper-based methods for collecting information can be costly and highly ineffective.

To meet the digitization needs of paper driven data collection practices, this paper presents the development and study of a novel system architecture for using a camera-equipped mobile device or a contact imaging sensor (CIS) handheld scanner coupled with a facility-based smart communication device to image paper forms and automatically understand their structure, produce a data schema (if this type of form has not been seen before), extract the raw data for the form, and transmit the information to a remote database via text messages. Through crowdsourcing, humans can be used to digitize raw input images on the server side in more urban regions. We hypothesize that this approach can significantly decrease the effort and cost of large-scale data entry, while maintaining a high level of quality.

The rest of this paper is organized as follows. Section 2 reviews the work related to methods that have been proposed for digitizing paper forms in the developing world. Section 3 provides a description of the research problem. In section 4, we continue by discussing the approach and methodology and pinpoints the intended research contributions. The paper concludes with a brief discussion of the future work.

2. RELATED WORK

Over the past few years, interest in coupling mobile technology and paper in the developing world has grown significantly. Numerous research efforts are trying to leverage the ubiquitous nature of paper and mobile devices to bring about innovative approaches for extracting digital information from paper forms in spite of infrastructure limitations. mScan (Dell et al., 2012) is an Android powered smartphone application that uses computer vision to capture digital data from bubble forms. The application employs a series of computer vision algorithms to preprocess bubble segments and make use of support vector machines (SVMs) to classify them. Currently, mScan can only recognize bubble marks on specific forms. All information remains on the phone and cannot be remotely accessed or verified. CAM (Parikh et al., 2005) is a human-information interface toolkit that bridges the display properties of paper with the interactivity and functionalities of a camera-equipped mobile phone. The framework uses quick response (QR) codes printed on predefined paper forms to facilitate data entry and communication tasks between the phone and a remote server. While the system withdraws some of the complexities with dealing with smartphones, the data entry requirement put on the mobile device may introduce some serious latency in large-scale data collection situations. Shreddr (Chen et al., 2012) is an online crowdsourcing platform architecture that semi-automatically extracts form fields from a scanned form and assigns the task of recognition to online users. The system uses a form image template and a schema description file with location information of input fields to segment individual field regions. Although Shreddr can handle a wide variety of data types, the system does not concern itself with the problems of semi-supervised form retrieval and feature-based input recognition.

3. PROBLEM STATEMENT

Consider the digital image of a paper form with variable field spaces (dynamic region) in which to write, mark or select. Assuming forms of the same kind contain unchanging texts, lines or graphics (static region), and a set of bounding boxes enclosing each dynamic field region can be obtained from labeling a single image of the form, our goal is to automatically detect the form in subsequent images, extract its input fields' images and transfer them to a remote server. Images of forms can be captured with a digital camera, a mobile phone or a handheld scanning device. Form images are to be detected and transferred from an area with poor power and IT infrastructure, and the presence of at least a 2G cellular network can be assumed. Additionally, human interaction with the system should require limited education and experience with IT. Once images have been succesfully transferred, input field values for a given form can be effectively digitize through crowdsourcing. In this paper we focus on describing the architecture and framework necessary for sucessfully detecting and transmitting paper form images in areas with low wireless bandwidth.

4. Solution Approach and Methodology

To convert paper form images to text, there are a number of sub-problems that need to be tackled. The bold circles in figure 1 indicate our research contributions. They are briefly discussed in the subsequent sections.

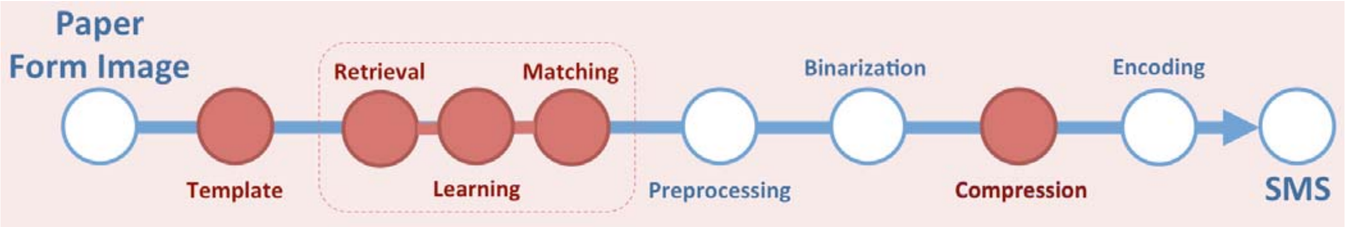

Figure 1: Paper Form Image to Text Pipeline

4.1 Template Building

Adaptive templates have been used extensively in object and motion tracking (Kalal et al., 2012), but they have not been applied for the purpose of identifying paper forms. Our aim is to find and track a combined feature profile that exposes the intrinsic manual input method. The assumption here is while the visual appearance of input fields might change, the manual input method will always remain the same. In other words, fields that require bubble marks will tend to have a similar feature profile, and fields that require handwritten input will tend to have a shared feature profile different from that of bubble marks. The overall detection of dynamic form regions will follow an approach similar to Multiple Component Learning (MCL) (Dollár et al., 2008). MCL is a discriminative learning approach for detection that is inspired by part-based recognition approaches. Our intention is to learn individual component classifiers for input fields and combines them into an overall classifier for the dynamic region.

4.2 Retrieve-Learn-Match

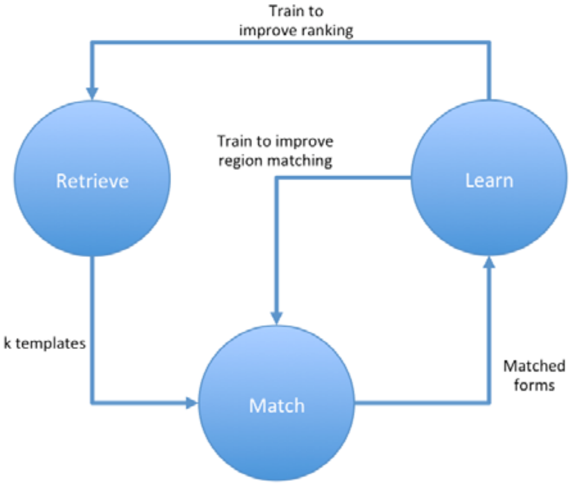

Figure 2: RLM Framework

The second contribution of this research is the design of a novel framework that decomposes the task of paper form identification into three sub-tasks: retrieving, learning, and matching (RLM). This configuration is inspired by the predator framework (Kalal et al., 2009) used for long term tracking of unknown objects in video streams.

We start by assuming that neither image retrieval nor duplicate image detection alone can fulfill the requirements of form matching, but a marriage where ideas from both techniques can cooperate and feed off of each other might provide significant benefits. Given an input document image, the retriever finds and ranks its k (where k≈3) most related templates and feeds them to a matchmaker. The matchmaker localizes the static and dynamic regions of each template on the input image and decides which of the three templates most correspond to the input form. Based on the matchmaker's answer, the learning component adjusts the retriever so that it can provide better rankings. The learning also analyzes the matchmaker's response, estimates its errors in matching dynamic and static regions, and trains the matchmaker for better performance in the future. Figure 2 illustrates the algorithm's architecture.

4.3 COMPRESSION

The third contribution of this research is a compression technique in which an image patch representing the ensemble of dynamic input regions of a known form is compressed along with the meta information that describes the form structure and is sent to a remote server for recognition. It is not within the scope of this research to define a new document compression scheme. Our aim is to leverage our form matching technique in order to minimize the amount of information necessary to reproduce on a remote computer a form with identical input data. The visual appearance of input field regions on the remote server must not alter their meaning when compared to the fields on the original form. The static region of an unknown form is sent to the server only once after generating the form's template. In subsequent detections, only the dynamic portion of the form is transmitted. Prior to transmission, document images are preprocessed, binarized, and compressed. In our method, compression happens in two ways. If an Internet connection can be established through a wireless technology higher than 2G, a binary image representing the entire dynamic region is compressed and uploaded to the server. If only a 2G cellular network is available, the binary image of the dynamic region is broken down into smaller image patches. Patches are compressed, encoded, and sent via text messages. For this research we use the DjVu [32] compression technique that is specifically geared towards the compression of high-resolution, high-quality images of scanned documents.

4.4 IMPLEMENTATION

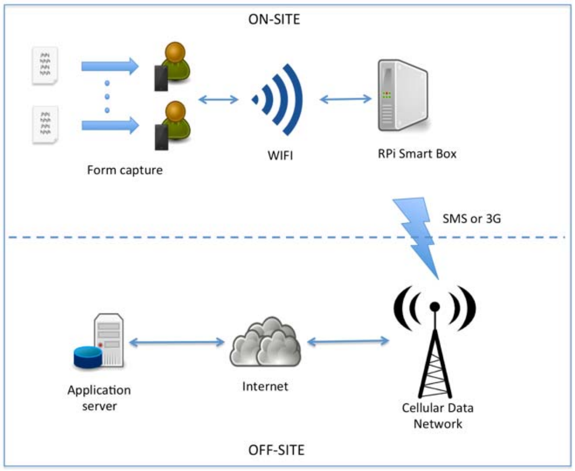

Figure 1: Paper Form Image to Text Pipeline

Though we rely on cellular technology, our physical implementation uses a custom facility-based communication device similar to Smart Connect (Anderson et al., 2010). The device is not meant to be a replacement for mobile phones; in fact, in our architecture, mobile phones are used as hybrid peripherals for imaging paper forms, controlling the processing capabilities of our smart device, recording text information and visualizing progress and results. In other words, the mobile phone will play the role of a keyboard, mouse, camera, and monitor all at the same time while directing the core image processing, storage and communication tasks to our smart device.

We hypothesize that if we restrict the tasks of the mobile phone to that of an I/O device, it will make it easier to work with smartphones with different operating systems, integrate CIS devices and run them concurrently. Most importantly, our architecture ensures that organization based information is stored on a dedicated organization owned device that cannot be used for any other purpose but its intended administrative functions.

5. CONCLUSION AND FUTURE WORK

In this paper, we have outlined a framework for using mobile phones with a facility based intelligent communication device for transmitting large-scale data being collected on paper forms in rural areas of developing countries. We have described our paper form detection architecture and our work in designing the initial components of the system. A considerable amount of work remains to be done in terms of technical development. However, the preliminary arguments and efforts highlighted in this paper show promising results for using a smart communication tool to enhance the information systems of rural facilities. The essence of this research is to understand how artificial intelligence can complement human effort in the developing world to improve the quality of local information systems and service delivery, all within a single framework.

REFERENCES

Anderson, R., Blantz, E., Lubinski, D., O'ourke, E., Summer, M., and Yousoufian, K. (2010). "Smart Connect: Last mile data connectivity for rural health facilities”. *NSDR'10*, San Francisco, CA, USA.

Chen, K., Kannan, A., Yano, Y., Hellerstein, J. M., and Parikh, T. S. (2012). “Shreddr: pipelined paper digitization for low-resource organizations”. *Proceedings of the 2nd ACM Symposium on Computing for Development (ACM DEV '12)*. ACM, New York, NY, USA, Article 3 , 10 pages.

Dell, N., Breit, N., Chaluco, T., Crawford, J., and Borriello, G. (2012). “Digitizing paper forms with mobile imaging technologies”. *Proceedings of the 2nd ACM Symposium on Computing for Development (ACM DEV '12)*. ACM, New York, NY, USA, Article 2, 10 pages.

Dell, N., Venkatachalam, S., Stevens, D., Yager, P., and Borriello, G. (2011). “Towards a point-of-care diagnostic system: automated analysis of immunoassay test data on a cell phone”. *Proceedings of the 5th ACM workshop on Networked systems for developing regions (NSDR '11)*, ACM, New York, NY, USA, pp. 3-8.

Dollar, P., Babenko, B., Belongie, S., Perona, P., and Tu, Z. (2008). “Multiple Component Learning for Object Detection”. *Proceedings of the 10th European Conference on Computer Vision: Part II (ECCV '08)*, Forsyth, D., Torr, P., and Zisserman, A. (Eds.). Springer-Verlag, Berlin, Heidelberg, pp. 211-224.

ITU. (2012). “Measuring the Information Society 2012”. Retrieved from: http://www.itu.int/ITU-D/ict/publications/idi/material/2012/MIS2012_without_Annex_4.pdf, 01/02/13.

Parikh, T. S. (2005). “Using Mobile Phones for Secure, Distributed Document Processing in the Developing World”. IEEE Pervasive Computing 4, 2, pp. 74-81.

Patnaik, S., Brunskill, E., and Thies, W. (2009). “Evaluating the accuracy of data collection on mobile phones: a study of forms, sms, and voice”. *Proceedings of the 3rd international conference on Information and communication technologies and development (ICTD'09)*, IEEE Press, Piscataway, NJ, USA, pp.74-84.

Kalal, Z., Mikolajczyk, K., and Matas, J. (2012). “Tracking-Learning-Detection”. IEEE Trans. Pattern Anal. Mach. Intell. 34, 7, pp. 1409-1422.

Kalal, Z., Mata, J., and Mikolajczyk, K. (2009). “Online learning of robust object detectors during unstable tracking”. On-line Learning for Computer Vision Workshop.

Authorization and Disclaimer

Authors authorize LACCEI to publish the paper in the conference proceedings. Neither LACCEI nor the editors are responsible either for the content or for the implications of what is expressed in the paper.