Digitizing Paper Forms with Mobile Imaging Technologies

Nicola Dell† Nathan Breit† Timóteo Chaluco‡ Jessica Crawford‡ Gaetano Borriello†

†University of Washington
{nixdell,nbnate,gaetano}@cs.washington.edu

‡VillageReach
{timoteo.chaluco,jessica.crawford}@villagereach.org

ABSTRACT

In low-resource settings in developing countries, most records are still captured and maintained using paper forms. Despite a recent proliferation of digital data collection systems, paper forms remain a trusted, low-cost and ubiquitous medium that will continue to be utilized in these communities for years to come. However, it can be challenging to aggregate, share, and analyze the data collected using paper forms. This paper presents mScan, a mobile smartphone application that uses computer vision to capture data from paper forms that use a multiple choice or bubble format. The initial mScan implementation targets the task of digitizing paper forms used to record vaccine statistics in rural health centers in Mozambique. We have evaluated the accuracy and performance of mScan under a variety of different environmental conditions, and our results show that mScan is a robust tool that is capable of accurately capturing and digitizing data from paper forms.

Categories and Subject Descriptors

I.2.10 [**Vision and Scene Understanding**]: Intensity, color, photometry, and thresholding; I.5.4 [**Applications**]: Computer vision

General Terms

Design, Measurement, Human Factors

Keywords

ICTD; computing for development; paper forms; machine readable forms; computer vision; cell phone; smartphone.

1. INTRODUCTION

Organizations working in developing countries depend on large-scale data collection to support the communities in which they work. Government, social and health organizations use such data to measure their impact and control the quality of the services they provide. Many of these organizations rely heavily on paper forms to perform this data collection. Paper forms are a well-understood and trusted medium for data collection in developing communities, and the low cost and ease-of-use of paper forms suggest that paper will continue to be extensively utilized for many years to come. However, the potential benefits of digitizing data from paper forms for the purposes of statistical analysis and aggregation are significant. For example, health centers in Mozambique use paper forms to record the number of vaccines administered to patients each month. Providing the Ministry of Health and other non-governmental organizations with quick access to accurate vaccine delivery statistics could aid critical decision making regarding resource allocation, assessment and planning.

Despite the recent development of numerous digital tools to aid data collection in developing countries, digital data collection in these settings remains challenging. The communities in which the data is collected often have insufficient IT infrastructure and support to facilitate effective digital solutions [21]. In addition, many rural regions in low-income countries are located beyond wireless network access, and in those where wireless access is available, the cost of technological devices and services may still be prohibitively expensive due to limited infrastructure and lack of investment. Furthermore, it can be challenging to collect data using basic mobile phones. SMS can only be used to record a few data points per message, and while Java-enabled phones have been used to create data collection forms, the number and type of data fields that can be included in a form without presenting a challenge to the user limits the potential of these technologies. As a result, collecting and submitting large volumes of data using the limited interface and functionality of basic phones is unrealistic.

The computational power and intuitive touch-screen interfaces of many smartphones suggest that they may be a more effective platform for data collection than basic phones [2]. Many smartphones have built-in cameras, GPS sensors, and network interfaces, making them capable of supporting a wide variety of social, administrative and health applications. Smartphones also have ample storage capacity, and are capable of storing the data locally on the phone and uploading it to a remote database when a cellular or data network becomes available. There are several existing smartphone-based data collection tools that can record a wide variety of data types and form fields [7] [10]. However, the cost of purchasing and maintaining a device for every health worker currently prohibits smartphones from being a viable option in many developing countries. Thus, instead of purely digital data collection tools, what is needed are hybrid solutions that keep the cost of deploying and maintaining the system low, by continuing to use cheap, familiar paper forms at the health-worker level, while also facilitating the collection of digital data on computationally powerful smartphones at the district or provincial level.

DEV '12 March 11-12, Atlanta, GA

Many of the paper forms utilized by social development and health organizations are multiple-choice or bubble forms, in which the user fills in a number of circles or bubbles to record an answer or event. Research suggests that these bubble forms are easier to understand and use than forms that require large amounts of writing, especially by people with little education or low literacy [21]. Bubble forms can also be designed to be machine readable using Optical Mark Recognition (OMR) technology. Traditional OMR systems work with a specialized scanning device that shines a beam of light onto the form. The device can detect marked areas on the form because they reflect less light than the blank areas. However, the cost of the specialized scanner and associated technological equipment prohibit current commercial OMR systems from being an option for most organizations in developing countries. While there are a few open source OMR systems available, all of them require that the forms be specially designed for use with the system, and are unable to digitize data from existing forms.

In this paper we present mScan, a mobile smartphone application that automates the capture and processing of digital data from paper-based bubble forms. The mScan application uses a lightweight form description language so that the paper forms do not have to be specially designed to work with the software. The phone's camera is used to capture an image of the form and computer vision algorithms automatically extract digital data from the image. The extracted data may either be stored locally on the phone or automatically uploaded to a remote database.

Although we plan to develop mScan into a generalizable platform for digitizing data from a wide variety of paper forms, we identified one concrete target application on which to focus our initial design and evaluation. Working in conjunction with Seattle-based NGO VillageReach [24] and the Provincial level Ministry of Health in Cabo Delgado province, we demonstrate the utility of mScan by using it to automate the capture of digital data from paper-based vaccine statistics forms used in rural health centers in Mozambique. The application users will be provincial-level government employees who travel to rural health centers on a monthly basis, a scenario which requires only each provincial-level employee to possess a device, rather than each health worker or health center. This significantly reduces the cost of deploying the system and training users, making mScan an affordable and sustainable solution.

2. RELATED WORK

There is a large body of research that explores the digitization of data from paper forms. We focus on two categories: digitizing data from bubble forms using traditional OMR systems and digitizing data from paper forms in the developing world.

2.1 Traditional OMR systems

Optical mark recognition (OMR) is the process of capturing human-marked data from paper forms and is used extensively to digitize information from surveys and standardized tests. Traditional OMR systems work with a specialized scanning device that shines a beam of light onto the form. Bubbles that are located in predefined positions on the form are classified as either filled or empty depending on the amount of light they reflect. OMR systems require forms to be filled in according to specific requirements, such as with a dark lead pencil [12]. Assuming that the forms are filled in correctly and carefully, commercial OMR systems are capable of achieving up to 99.5% accuracy. Several commercial OMR systems are available [1] [16] [19], but the cost and maintenance of these specialized systems prevents them from being a viable option for many organizations working in developing countries.

A number of open source OMR systems have recently been developed in response to the high cost of the commercial solutions. The most notable of these is the Udai OMR tool [20]. Unlike mScan, which uses a smartphone camera to capture photographs of forms, Udai requires users to scan forms using a scanner that is connected to a desktop or laptop computer. The scanned image is then loaded into the tool and processed. Udai also requires that the paper forms to be digitized be specially designed for use with the system. Forms are printed with two concentric circles in the top-left and bottom-right corners of the form, which allow the system to determine the angle of rotation and eliminate skew. In contrast to this, mScan can be programmed to capture data from existing paper forms without requiring users to add identifying characteristics or coded marks to the form. Other open source OMR systems include queXF [17] and the Shared Questionnaire System [22]. These systems also require a scanner and computer to read data from specially designed and neatly filled-in multiple-choice forms.

2.2 Paper Forms in the Developing World

The ubiquitous use of paper forms for data collection in the developing world has resulted in a large amount of research that focuses on extracting digital data from these forms. Singh et al. [21] analyze the use of paper forms in the context of non-governmental organizations (NGOs). They investigate the balance between ease-of-use among intended populations and machine readability, and suggest that users in developing countries prefer numeric and multiple-choice forms since they can fill them more quickly and accurately than writing-based forms.

Several solutions have investigated the relationship between paper forms and digital technology. CAM [14] is an interface toolkit that allows a camera equipped mobile phone to interact with paper forms. The forms contain visual codes that serve as references to assist the user with data entry and communication with a remote database. CAM is a powerful tool that can handle a wide variety of data types. However, unlike mScan, which can automatically process the machine-readble portions of the form, CAM users are required to manually enter all of the data into the phone.

Shreddr [4] is another system for digitizing data from paper forms. Shreddr semi-automatically extracts the form schema and locations of the form fields from a scanned form. It then segments the form and assigns the recognition of individual fields into tasks that are performed by people via a crowd-source platform. Although Shreddr can handle a wide variety of data types, the system does not leverage the machine readability of certain data types, like bubbles, and requires people to read and input all of the data. Furthermore, a reliable Internet connection is required for the effective use of a crowd-sourcing platform.

Local Ground [23] is a tool that allows people to document their knowledge of places using barcoded paper maps, computer vision techniques and publicly available mapping tools. Users annotate paper maps using pens, markers and

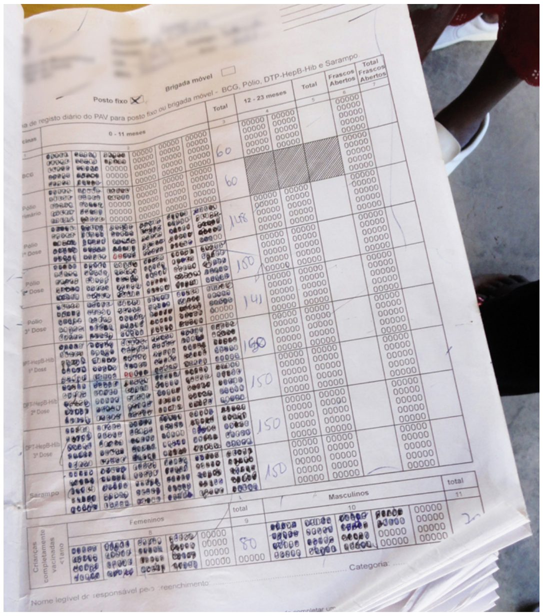


Figure 1: Paper form used to tally the number of vaccines administered at a health center in Mozambique over the period of one month. Identifying characteristics have been blurred out.

stamps. The maps are then scanned, and user markings extracted and overlaid on existing online maps to aid local planning decisions. Unlike mScan, the Local Ground system treats the user markings on the paper as an image layer, and is not capable of reading or making sense of them.

Finally, Ratan et al. [18] present a financial record management application built on a low-cost digital slate device. The solution accepts handwritten input on paper forms and provides electronic feedback. Testing of the paper-pen-slate system showed that data can be collected more quickly with fewer incorrect entries and more complete records using this system. Users preferred the system over a purely electronic solution because they liked having physical evidence of their transactions. However, the purchase and maintenance of specialized digital slate devices hinders the scalability and sustainability of the system.

To the best of our knowledge, mScan is the first solution to make use of computer vision algorithms running on a commercially available smartphone to digitize human-marked data directly from paper forms.

3. TARGET APPLICATION

Although we intend for mScan to be a generalizable platform for digitizing data from paper forms, we identified one concrete target application on which to focus our initial implementation. This target application involves extracting vaccine statistics from paper forms in rural health centers in Mozambique. We chose the application in conjunction with VillageReach [24], a Seattle-based NGO that works with provincial level Ministries of Health in Mozambique to increase vaccine coverage rates through improved delivery logistics. To better understand the problem space and solution requirements, we held several in-depth discussions with VillageReach personnel who have spent significant amounts of time in the field working with the target users.

The Mozambique Ministry of Health distributes booklets of tally bubble forms, shown in Figure 1, to every health center in the nation. Each booklet contains approximately 100 forms. The forms have been designed to record the total number of each type and dose of vaccine administered at the center over the period of one month. At the top of each form are several text fields designed to record the province, district and center information along with the appropriate month and year. The rest of the form contains fields printed with bubbles that are used to tally the number of each type of vaccine administered at the center. Each form has 24 bubble fields separated according to vaccine type and patient age. Within each field, bubbles are located in segments of 20 or 25 bubbles per segment, and one to six segments per field. This results in the number of bubbles per field ranging from 25 to 150, with a total of 90 bubble segments and 2180 bubbles per form. Each individual bubble measures approximately 1 x 1.5 mm.

Figure 2 depicts the current paper-based workflow. When a patient comes to the center to receive a vaccination, the health worker administers the dose and fills in a bubble on the tally form according to the type of vaccine administered and age of the patient. Filled bubbles accumulate in this way for the period of one month. At the end of each month, the health worker counts the number of bubbles that have been filled in for each vaccine type and dose and writes the totals in another field on the tally form. After all the tallies have been totaled, the health worker records the numbers on a separate summary form. Health workers report that the process of tallying the bubbles and filling in the appropriate forms takes approximately 30 minutes, and that they usually tally the bubbles two or three times to double-check their work. The summary form is then transported from the health center to the district office. Depending on the location of the health center, this process takes about 1.5 days. At the district office, the information from all of the health center summary forms is aggregated and recorded on a district summary form. This aggregation takes approximately 1.5 hours. The district summary form is then transported to the provincial office and the information from each district is summarized and manually entered into a database, where it is subsequently available for analysis by provincial level Ministry of Health personnel.

4. DESIGN CONSIDERATIONS

Our analysis of the problem space revealed several important design considerations. First, we do not attempt to change the method used by health workers to tally the number of vaccines administered. Instead, we target the point at which data from the form is summarized. The target users are therefore provincial-level field coordinators that visit health centers on a monthly basis. The field coordinator would carry the device and digitize data from tally forms during his monthly visit to each health center, and the data collected would be automatically uploaded to the provincial database when he returns to the provincial office. This proposed workflow is shown in Figure 2.

By targeting provincial-level field coordinators who travel to a variety of health centers within a district, we will be able to digitize the forms from many centers using a single device. This significantly reduces the overall cost of deploying the application. Furthermore, our design ensures that health workers in the center are not required to learn a new system

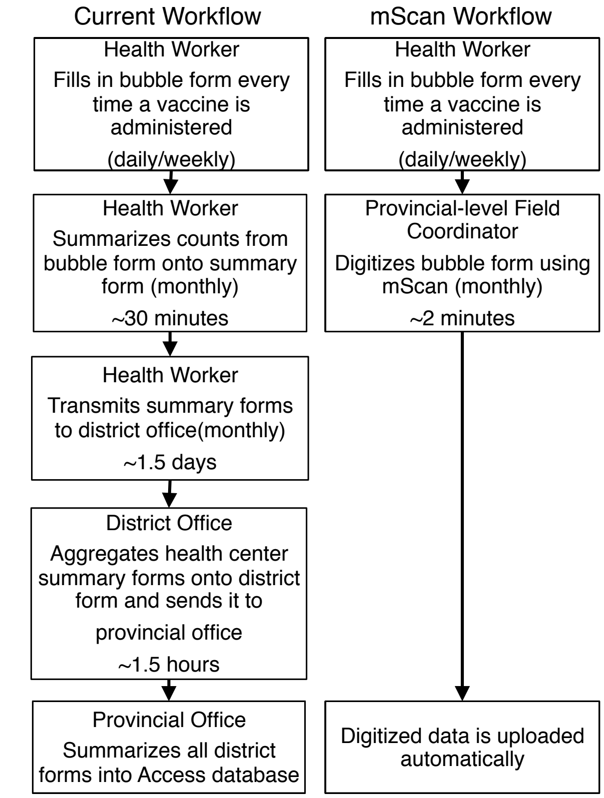


Figure 2: Current paper-based workflow (left) and proposed mScan workflow (right).

and can continue to record vaccine statistics using the paper forms that they are familiar with. This is beneficial since we anticipate that it will require fewer resources to train a small number of provincial-level field coordinators to use the new application than a large number of health workers.

Unlike more traditional OMR solutions, mScan does not require that the paper forms to be digitized be specifically designed to work with the application. This is important because organizations often find it undesirable or infeasible to re-design the paper forms that are already in use in low-resource communities. Our design therefore ensures that mScan is capable of digitizing data from existing paper forms without the need to redesign or add coded marks to the form.

Furthermore, since the paper forms are distributed in booklets, cropping forms out of images by looking for their outline is not a feasible approach. Instead, the application performs feature extraction and matching to align captured form images prior to processing. This approach requires a template form image and schema description be loaded on to the device for each unique form type. The template and form description only need to be created and uploaded onto the device once, and they are subsequently used to align and process all captured images of that form type. The exact details of the techniques we use for creating the form description and for feature extraction and mapping are described in the next section of the paper.

Since a single form is filled in over the period of a month, it is possible for the form to become folded, dirty or marked accidentally (see Figure 1), and it is essential that mScan be robust enough to handle these complicating factors. Furthermore, it is common for a single form to contain markings from several different health workers and a variety of pens and pencils. Thus, the application has been designed to deal with markings made by different types and colors of both ink and pencil. In addition, bubbles are often filled in hurriedly or haphazardly by busy health workers. This results in significant variation in the appearance of filled bubbles. Figure 3 shows a small portion of a partially completed form, which could not be processed with traditional OMR techniques. Our design handles this type of input by running machine learning algorithms on the phone to classify bubbles as either filled or empty.

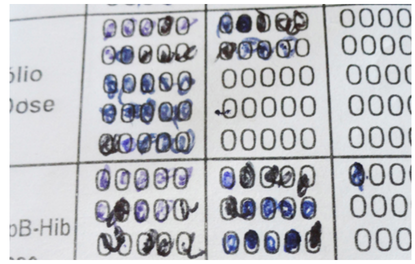


Figure 3: Portion of a partially filled bubble form.

Finally, although the initial mScan prototype targets the task of digitizing vaccine statistics in Mozambique, we have designed the application pipeline to extend easily to other paper forms and applications. We utilize a lightweight, generalizable JSON [6] form description language to specify the location, size and type of each form field. This form description language is capable of describing a wide variety of data types and is based on the form description language used by other digitization platforms, such as Shreddr [4]. Initial tests performed with other forms show that extending the application to include a new form would require three main components: (1) an image of an empty form to use as a template for feature extraction and matching, (2) a JSON form description that can be used by the application to identify the fields of the form, and (3) a variety of bubble samples from the new form that can be used to train the classifier. The remainder of the application pipeline can remain unchanged and be reused for any bubble form.

5. METHOD

5.1 Architecture

We built mScan as an interactive application running on the Android platform. The decreasing cost of devices and open source nature of the platform, coupled with the fact that our initial target application would require the purchase of only a small number of devices, made Android an attractive choice for the application. The image processing components of the application are implemented using OpenCV [13], an Open Source Computer Vision Library, while the user interface components are implemented using Android's Java framework. We used the Java Native Interface (JNI) [9] to facilitate communication between the Java framework and OpenCV's native image processing algorithms. All of the image processing is performed on the phone without requiring an Internet or cellular connection.

5.2 Implementation

For each digitized form there are eight main processing steps that we describe in detail in the sections below.

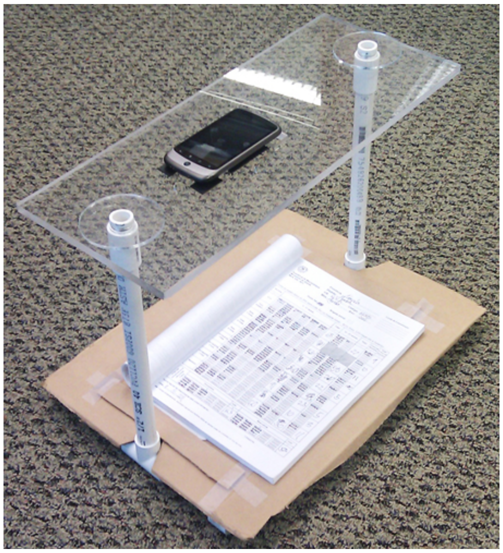


Figure 4: Low-cost plastic stand and form tray used to capture images of paper forms.

5.2.1 Camera Calibration

Camera calibration is the process of determining the extent to which lens distortion affects a captured photograph, and computing the parameters required to correct this distortion. Calibration only needs to be performed once per phone or camera, and the resulting calibration data is saved and used for all the images captured by the camera. mScan uses OpenCV's camera calibration application, which involves processing images of a printed black-and-white chessboard pattern. Data from these images is used to calculate the amount by which the image suffers from lens distortion, so that future images can be appropriately corrected.

5.2.2 Image Capture

The next step in the digitization process is to use the phone's camera to capture an image of the form. To capture fine grained details like bubbles, images should be well focused and taken while the camera is steady. The image should contain all of the form content while maintaining a minimal distance from the form. To make it easy to take photos under optimal conditions, we designed and built a low-cost, plastic stand that may be used to hold the phone in position. As shown in Figure 4, the phone is placed on the stand and the form is placed beneath the camera, thereby ensuring that both the camera and the form are correctly positioned. While using the stand may not be the most convenient method of capturing an image, it increases the chances that the captured image will be of sufficient quality to be accurately processed. To capture an image of the form, the user places the phone on the stand and presses a button to launch the Android camera application. The user can then take and retake photographs of the form. When satisfied with the captured image, the user presses a button on the phone to accept the image, which is saved and passed to mScan for further processing.

5.2.3 Image Registration

After an image of the form has been captured, we perform image registration. This involves spatially transforming the picture of the form to align with a reference image that is stored on the phone. Alignment is necessary to ensure that the entire form has been captured and to determine the locations of the form fields and elements. mScan uses feature detection and matching to perform registration. Features are detected using OpenCV's SURF (Speeded Up Robust Feature) [3] implementation. In order to decrease the amount of time required for feature matching, mScan augments SURF with a grid adapted feature detector that limits the number of features that can be extracted from the image. After features are extracted, matching is performed using the fast approximate nearest neighbor algorithm [11]. Finally, mScan uses the RANSAC algorithm [8] to compute a transformation that maps the captured image to the reference image, thereby establishing a point-by-point correspondence between the two images.

5.2.4 Segment Detection and Alignment

After image registration, mScan performs individual segment alignment. Bubble segments need to be individually aligned because even if the form image has been accurately detected and aligned, the bubble segments are still likely to be slightly warped due to the bend of the form booklet or lens distortion from the camera. Our sample form contains 90 bubble segments. For each segment, the predicted coordinates of the segment bounding box are obtained from the form description file that is stored on the phone. The boundaries of each segment on the form are marked by black lines, and the algorithm searches for these lines as follows: first, it converts the image to a binary format using difference of means and thresholding techniques; then, it calculates the minimum energy lines in the image and considers these to be the lines defining the edges of a form segment. The algorithm locates the corners of the segment by calculating the intersection of the detected lines, and finally performs a transformation to map the detected corner points to the reference corner points.

5.2.5 Classifier Training

Individual bubble alignment and classification is performed using a Support Vector Machine (SVM) [5] classifier. To train the classifier, we used 67 bubble training images and labeled them as filled, partially-filled, barely-filled or unfilled. The training images were normalized and then principal components analysis (PCA) [15] was performed on the training set. The PCA projected training images were then used to train an instance of OpenCV's SVM. The data generated during the training process is cached and saved, so training only needs to be performed once, rather than every time a form is processed.

5.2.6 Bubble Alignment

Within an aligned segment, the approximate location of each individual bubble is known from the provided form description. Using this approximate location as a starting point, the algorithm searches for the most bubble-like region by testing candidate bubble regions. For each candidate bubble region, we calculate the PCA [15] back projection of that region against the bubble data obtained from training the classifier. The sum of squared differences (SSD) between the original candidate region and the PCA back projection is taken to be the objective function that we seek to minimize. We then find a local minimum using a hill-descending search and take it to be the actual bubble location.

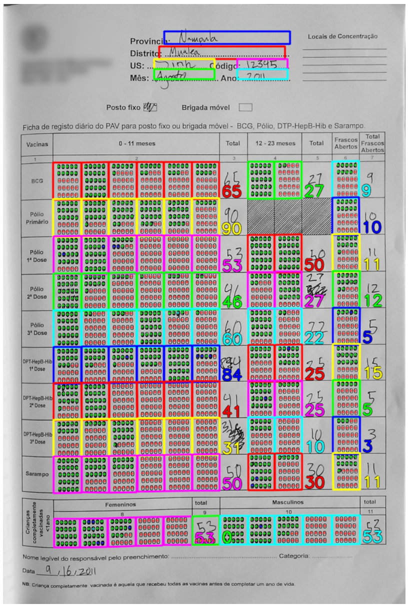


Figure 5: A marked up image of a processed form. The form contains fictitious data and identifying characteristics have been blurred out.

5.2.7 *Bubble Classification*

Individual bubbles are classified by running the SVM's predict function with a normalized PCA-projected image of the aligned bubble. In the current implementation, bubbles that are classified as filled, partially-filled and barely-filled are all added to the final tally. However, by providing several classification categories in addition to simply filled or unfilled, we create the potential for future implementations to infer the likelihood of a bubble being filled from the surrounding form region. For example, if a bubble is classified as being barely-filled, but all of the surrounding bubbles are classified as being unfilled, it is likely that the barely-filled bubble is not really filled but rather that the result is due to noise in the captured image. Inferring the likely classification of an individual bubble from the surrounding image region is an area of the application that is still under development.

5.2.8 *Data Output and Integration*

After all of the individual bubbles have been classified, the final tally count for each form field needs to be saved and output in a usable format. To do this, the application constructs a JSON [6] output file that contains the name of each form field (specified in the form description file) and the appropriate bubble tally for that field. In addition, the user must manually enter a small number of text fields from the form, including the center name, district and province. This information is combined with the bubble tally data to create a single digital record for each processed form. This digital record can be automatically integrated with existing digital systems, uploaded to remote databases or stored and retrieved locally on the phone. In addition to the digital record, a marked-up image of the form, shown in Figure 5, that depicts the results of the form alignment and bubble classification is also saved and made available to the user for inspection and validation. This image can be used to resolve any discrepancies that may arise after the form is processed, and will also ensure that a record of the original paper form is preserved and archived.

6. EXPERIMENTAL EVALUATION

To evaluate the technical performance of mScan, we conducted experiments designed to test the robustness and accuracy of the application under different environmental conditions. Our objective was to evaluate our algorithm and determine the ideal environmental conditions that should be targeted for field testing. The experiments described in this section were conducted by researchers in Seattle, USA. We have not yet rigorously evaluated the performance of the application with users in rural health centers.

The experiments were performed using an HTC Nexus One Android device. Since we wanted to minimize the complexity of future user training, we set all of the camera parameters to automatic. To analyze the performance of the application, we defined two types of processing errors: segment errors, in which individual segments are incorrectly aligned, and bubble errors, in which individual bubbles are incorrectly classified. We identified segment errors in two ways. First we see if a segment is convex and non-self-intersecting. Then, we see if the area of the segment differs from the expected area by more than 15%. If either of these conditions are detected, the segment is deemed to be incorrectly aligned. Bubbles classified correctly are either true positives (T-P) or true negatives (T-N) while errors are either false positives (F-P) or false negatives (F-N). Two different test forms were used for the experiments. A 'neat' form in which the bubbles were filled in relatively neatly, and a 'messy' form, in which the bubbles were filled in haphazardly in a manner consistent with that observed on photographs of forms obtained from the target community. Each test form contained 924 filled bubbles and 1256 unfilled bubbles, with the same bubbles filled in on each form.

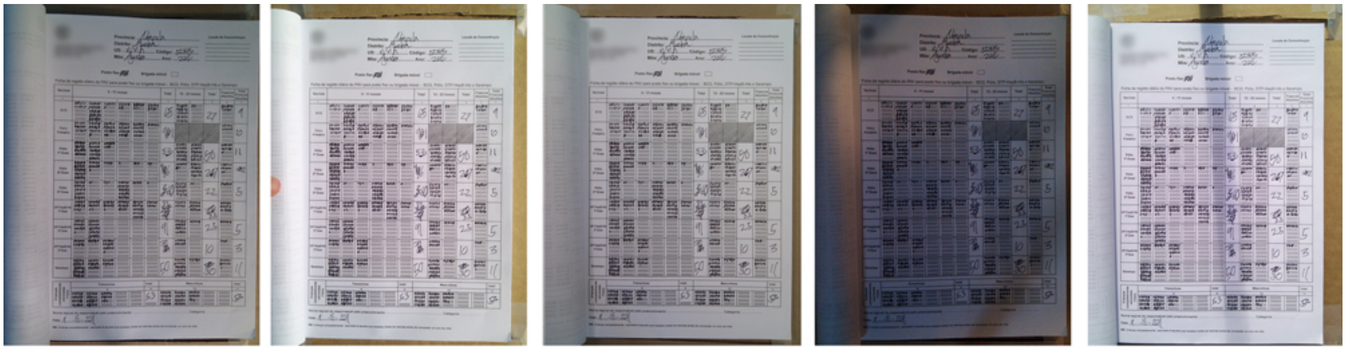
6.1 Lighting Conditions

6.1.1 *Experiment Design*

Lighting is a critical factor in any image processing application. To evaluate the performance of mScan under a variety of lighting conditions, we first defined five different lighting conditions that might reasonably be expected in a rural setting: (1) Dark: test images were captured in a dark room, (2) Light: test images were captured in a very light room with the form placed in a sunny spot, (3) Medium: test images were captured in a well-lit room, (4) Dark shadows: test images were captured with dark shadows falling on to the form, and (5) Light shadows: test images were captured with light shadows falling on the form. Figure 6 depicts sample images used for the experiment. We did not include conditions that required use of the camera's flash for two reasons: not all phone models are guaranteed to possess a flash, and using the flash will deplete the battery life of the phone. In addition to placing the phone in a plastic stand to minimize positional variation, the form was also placed against a brown background that was kept constant across lighting conditions. For each lighting condition, five images of each test form were captured and processed.

6.1.2 *Results*

The experiment results are given in Table 1. For each condition, form segments were correctly aligned with



Dark | Light | Medium | Dark shadows | Light shadows

Figure 6: Sample images showing each lighting condition tested. (The forms contain fictitious data and identifying characteristics have been blurred out.)

Table 1: mScan performance under different lighting conditions

		Segment Alignment			Bubble Classification					Overall
	Condition	Correct	Errors	Accuracy	F-N	F-P	T-N	T-P	Accuracy	Accuracy
Neat form	Dark	449	1	99.78%	5	4	6276	4595	99.92%	99.70%
	Light	449	1	99.78%	12	9	6271	4588	99.81%	99.59%
	Medium	450	0	100%	5	2	6278	4615	99.94%	99.94%
	Dark shadows	449	1	99.78%	5	208	6058	4604	98.04%	97.82%
	Light shadows	450	0	100%	5	5	6275	4615	99.91%	99.91%
Messy form	Dark	450	0	100%	23	10	6270	4597	99.70%	99.70%
	Light	450	0	100%	22	13	6267	4598	99.68%	99.68%
	Medium	450	0	100%	25	10	6270	4595	99.68%	99.68%
	Dark shadows	449	1	99.78%	24	269	6011	4571	97.31%	97.09%
	Light shadows	448	2	99.56%	35	13	6242	4560	99.56%	99.12%

overall accuracy above 99%. Additionally, for each lighting condition except dark shadows, mScan was able to correctly classify more than 99% of the bubbles on the form. Dark shadows caused the accuracy to drop to 98.04% for the neat form and 97.31% for the messy form. These results indicate that while conditions that result in dark shadows should be avoided, the application is robust to moderate variations in lighting. Finally, as depicted in Figure 6, the images captured for the dark, medium and light conditions appear to be relatively similar to each other. Since we did not want to burden users with specific camera adjustments, we capture all images using fully automatic settings on the camera. This results in some lighting autocorrection being performed when the image is captured, making the images appear similar despite being captured under different environmental conditions.

6.2 Folded or Dirty Forms

6.2.1 Experiment Design

Since forms are filled in over a period of one month, they could become folded or dirty and we wanted to ensure that mScan is robust to these kinds of effects. We tested five different folded or dirty conditions: (1) a form folded in half vertically, (2) a form folded in eighths with both vertical and horizontal folds, (3) a form with dog-eared folds, (4) a crumpled form and (5) a dirty form. Since we wanted to compare the same filled forms across different folded and dirty conditions, we used photocopies of filled-in forms for this experiment. Figure 7 shows some sample images used for the experiment. We captured the images in a setting with medium lighting and a uniformly brown background, and we maintained as much lighting consistency as possible across all conditions. For each test condition, we again captured five images of each test form and analyzed the results.

6.2.2 Results

The experiment results are given in Table 2. Segment alignment, bubble classification and overall accuracy were above 99% for three of the five conditions tested: folded in half, dog-eared folds and dirty form. These results show that mScan is robust to moderate folding and dirtying of the form. However, when the form contained many folds, such as being folded in eighths, segment alignment accuracy dropped to 94.22% with the neat form and 95.11% with the messy form, while bubble classification accuracy dropped to 96.87% with the neat form and 96.11% with the messy form. This resulted in an overall accuracy of 91.27% with the neat form and 91.41% with the messy form. Application performance also dropped with a very crumpled form. Under this condition, segment alignment accuracy was 97.33% with the neat form and 90.22% with the messy form, while bubble classification accuracy was 95.03% with the neat form and 90.38% with the messy form. This resulted in overall accuracy of only 92.50% with the neat form and 81.54% with the messy form. The results indicate that severe crumpling and folding of the form should be avoided.

6.3 Background Conditions

6.3.1 Experiment Design

The previous experiments all used a uniform brown background. However, we expect that users may place the form against other types of backgrounds to capture

Figure 7: Sample images showing the folded and dirty conditions tested. (The forms contain fictitious data and identifying characteristics have been blurred out.)

Table 2: mScan performance with folded and dirty forms

		Segment Alignment			Bubble Classification					Overall
	Condition	Correct	Errors	Accuracy	F-N	F-P	T-N	T-P	Accuracy	Accuracy
Neat form	Folded in half	449	1	99.78%	9	19	6241	4611	99.74%	99.52%
	Folded in eighths	424	26	94.22%	103	220	5704	4278	96.87%	91.27%
	Dog-eared folds	450	0	100%	5	22	6258	4615	99.75%	99.75%
	Crumpled form	438	12	97.33%	217	310	5827	4251	95.03%	92.50%
	Dirty form	450	0	100%	5	12	6268	4615	99.84%	99.84%
Messy form	Folded in half	449	1	99.78%	26	22	6258	4574	99.56%	99.34%
	Folded in eighths	428	22	95.11%	104	301	5718	4287	96.11%	91.41%
	Dog-eared folds	449	1	99.78%	30	38	6235	4577	99.38%	99.15%
	Crumpled form	406	44	90.22%	496	455	5270	3664	90.38%	81.54%
	Dirty form	450	0	100%	28	45	6235	4592	99.33%	99.33%

images and we wanted to test whether mScan is able to handle such cases accurately. We defined five different background conditions: brown, black, white, shiny and patterned. Sample images showing these backgrounds are given in Figure 8. Once again, the form was placed in the plastic stand and we attempted to keep a consistent medium lighting across all images. For each condition, five images of each test form were captured and the results analyzed.

6.3.2 *Results*

The experiment results are given in Table 3. Segment alignment, bubble classification and overall accuracy were above 99% in all cases except one: a messy form on a white background. Under this condition, segment alignment accuracy was 98.89%, while bubble classification accuracy was 98.63%, resulting in an overall accuracy of 97.53%. These results show mScan to be a robust tool that is capable of accurately processing forms placed against a variety of backgrounds, but if possible, the user should try to place the form against a dark background.

6.4 Processing Rate

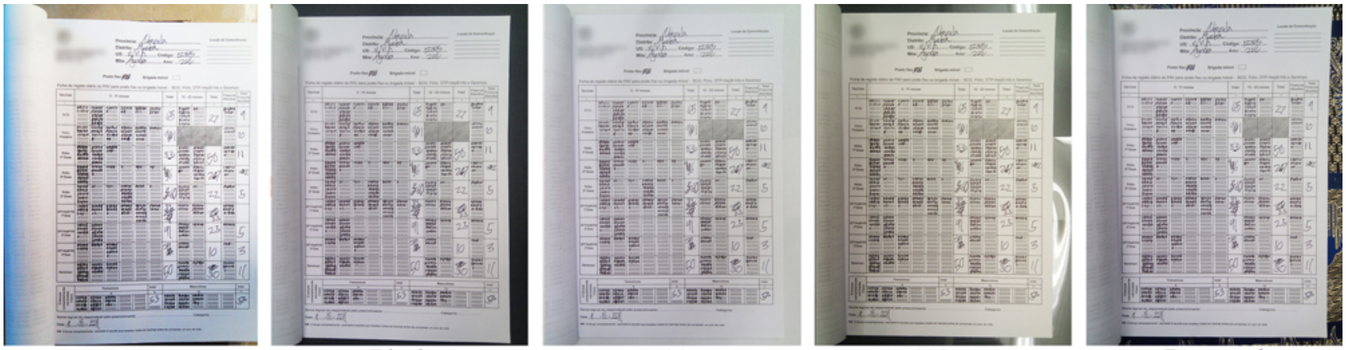
The application processes each form in approximately 25 seconds. Removing distortion from the captured image takes about 4 seconds. Image registration then takes about 15 seconds, including feature extraction and matching. Individual segment alignment requires about 1.7 seconds, bubble alignment 3.4 seconds, and bubble classification 0.9 seconds. Processing the entire form image in 25 seconds provides a significant improvement over the 30 minutes that it takes for health workers to aggregate the data by hand.

7. USER EVALUATION

We conducted a preliminary field test of mScan with Provincial Ministry of Health field coordinators in the Cabo Delgado province of Mozambique in November 2011. Our objective was to assess the application in relation to the expertise and technological experience of the target users and the environmental conditions experienced in rural health centers. We did not perform a rigorous usability evaluation at this time, but rather gathered general feedback to inform the design of a more complete user interface.

We held a training session with three provincial-level field coordinators, the Expanded Program on Immunization (EPI) Chief, and a VillageReach field officer at the provincial health offices in Cabo Delgado. During the session, we demonstrated mScan and explained the importance of environmental factors like good lighting and form position. After demonstrating the application, we observed the field coordinators using mScan to capture data from paper forms. The field coordinators understood quickly how to use the application and were able to successfully digitize data without being prompted. During the training session, we discovered that several field coordinators owned laptop computers and Android smartphones, and were completely comfortable interacting with the application. In addition, the plastic stand simplified the process of positioning the form, and is easy to fold up and carry in a backpack.

After the training session, we visited five rural health centers in the province over a period of several days. During these visits we were accompanied by two field coordinators. The purpose of the visits was to see if the field coordinators could successfully use the application under the conditions



Brown Black White Shiny Patterned

Figure 8: Sample images showing background conditions tested. (The forms contain fictitious data and identifying characteristics have been blurred out.)

Table 3: mScan performance with different background conditions

		Segment Alignment			Bubble Classification					Overall
	Condition	Correct	Errors	Accuracy	F-N	F-P	T-N	T-P	Accuracy	Accuracy
Neat form	Brown	450	0	100%	5	2	6278	4615	99.94%	99.94%
	Black	450	0	100%	5	4	6276	4615	99.92%	99.92%
	White	450	0	100%	5	0	6280	4615	99.95%	99.95%
	Shiny	450	0	100%	5	8	6272	4615	99.88%	99.88%
	Patterned	450	0	100%	5	3	6277	4615	99.92%	99.92%
Messy form	Brown	450	0	100%	25	10	6270	4595	99.68%	99.68%
	Black	450	0	100%	36	8	6272	4584	99.60%	99.60%
	White	445	5	98.89%	63	85	6139	4508	98.63%	97.53%
	Shiny	450	0	100%	37	10	6270	4583	99.57%	99.57%
	Patterned	450	0	100%	36	11	6269	4584	99.57%	99.57%

experienced in the health centers. At the first two health centers we visited, we demonstrated the application for the health workers and then observed the field coordinators and health workers using the application to capture data from paper forms. In addition, we explained to the health workers the importance of filling in the bubbles on the form neatly and keeping the form clean. At the next three health centers, the field coordinators demonstrated the application for the health workers and taught them the importance of filling in the form neatly. Figure 7 shows a field coordinator teaching a health worker about the application.

Although this preliminary user feedback is very encouraging, we also discovered several issues that need to be addressed as we develop mScan further. First, the tally forms in several health clinics had a number of subtle differences to the template form that we used. As a result, we were unable to process the forms without first modifying the template image stored on the phone. The field coordinators informed us that the forms in the clinics were an old format and would soon be replaced by the newer forms that we were using. However, this complication highlighted the need for a quick and easy way to add new form templates to the application. Furthermore, several of the health centers we visited used more than one form per month, so it would be useful to add functionality that allows the data from multiple forms to be aggregated. Finally, the field coordinators enjoyed examining the marked-up form image that mScan outputs and comparing the results with the filled bubbles on the original form. Since mScan currently achieves around 99% accuracy, the field coordinators occasionally found a bubble that had been read incorrectly, and expressed an interest in being able to manually correct the errors.

Figure 9: A field coordinator explaining how mScan works to a health worker.

8. DISCUSSION AND FUTURE WORK

This paper presents an initial implementation of mScan. There are still many challenges that need to be addressed to create an end-to-end pipeline for digitizing data from paper forms. Our evaluation thus far has focused primarily on the technical aspects of the application, investigating the feasibility and optimal environmental conditions for a mobile image recognition solution. Although the preliminary user feedback that we have obtained is encouraging, we have not yet rigorously evaluated the system with respect to user satisfaction and task performance. Working extensively with users in the field will be critical to ensure that the application is usable and appropriate given the constraints experienced by low-resource communities. We expect to complete several rounds of iterative interface design as we work with users to create a usable and robust application.

Furthermore, although most of the information collected on the vaccine tally form is captured through bubbles that can be digitized automatically, there are a small number of text fields that must currently be entered into the phone manually, such as the health center name, district and province. However, since the provincial-level employee is likely to visit the same health centers many times, we plan to experiment with storing the health center information permanently on the phone and, instead of typing in the relevant details every time, the user could simply select the appropriate health center from a list of stored centers.

Finally, although mScan has been designed to generalize easily to other paper forms, the implementation presented in this paper focuses on digitizing data from one particular bubble form. We plan to extend the application to include a variety of different bubble forms, as well as other data types, focusing initially on automatic recognition of checkboxes and coded form fields, and progressing to handwritten numbers. We will also investigate integrating mScan with the Shreddr platform [4], since both solutions use the same generalizable JSON form description language. We anticipate building a solution in which the machine-readable parts of the form are processed automatically using mScan and the human-readable parts are processed with the aid of a crowd-sourcing solution like Shreddr.

9. CONCLUSION

Government, social and health organizations working in developing countries rely heavily on paper forms to collect information from the communities in which they work. Digitizing, aggregating and sharing this information is crucial to help these organizations provide services to low-resource populations. We have designed and implemented mScan, a mobile application running on the Android platform that uses computer vision to automate the capture of digital data from paper forms that use a multiple choice or bubble format. Our experimental evaluation and preliminary user feedback show that mScan is an accurate and robust tool that is ready to be extended to a wide variety of different paper forms and use cases. Beyond aiding vaccine delivery, mScan has the potential to play a useful role in bridging the gaps that exist between the paper and digital worlds in resource-constrained environments.

10. ACKNOWLEDGMENTS

This work was funded in part by a Bill and Melinda Gates Foundation Grand Challenges Explorations grant. We would also like to thank Ruth Anderson, Yaw Anokwa, Kuang Chen, James Dailey, Brian DeRenzi, Steve Geluso, Leah Hasselback, Jeremy Keeton, Beth Kolko, Vincent Lung, Eleanor O'Rourke, Ron Pankiewicz, Baron Von Oldenburg, Aijia Yan, VillageReach and the Provincial Ministry of Health in Cabo Delgado, Mozambique.

11. REFERENCES

[1] ABBYY Formreader 6.5. http://www.abbyy.com/formreader/.

[2] P. Baudisch and C. Holz. My new PC is a mobile phone. *XRDS*, 16:36–41, 2010.

[3] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. SURF: Speeded Up Robust Features. *CVIU*, 110(3), 2008.

[4] K. Chen, J. Hellerstein, and T. Parikh. Data in the First Mile. In *CIDR*, 2011.

[5] C. Cortes and V. Vapnik. Support-Vector Networks. *Machine Learning*, 20(3), 1995.

[6] D. Crockford. Introducing JSON. json.org, 2009.

[7] eMOCHA. Electronic Mobile Open-source Comprehensive Health Application. http://emocha.org.

[8] M. Fischler and R. Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, 4, 1981.

[9] R. Gordon. *Essential JNI*. Prentice Hall, 1998.

[10] C. Hartung, Y. Anokwa, W. Brunette, A. Lerer, C. Tseng, and G. Borriello. Open Data Kit: Building Information Services for Developing Regions. In *ICTD 2010*.

[11] M. Muja and D. Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. In *VISSAPP'09*.

[12] NCS Pearson, Inc. ICR, OCR, and OMR: A Comparison of Technologies. www.syrex.com.ph/Product/Download?filename=WhatIsOMR.

[13] OpenCV (Open Source Computer Vision). http://opencv.willowgarage.com/wiki/.

[14] T. Parikh, P. Javid, S. K, K. Dhosh, and K. Toyama. Mobile Phones and Paper Documents: Evaluating a New Approach for Capturing Microfinance Data in Rural India. In *CHI 2006*.

[15] K. Pearson. On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine*, 2(6), 1901.

[16] Pegasus Software Development Toolkit. http://www.accusoft.com/smartscanicr.htm.

[17] queXF: An Open Source, web based paper form verification and data entry system. http://quexf.sourceforge.net/.

[18] A. Ratan, S. Chakraborty, P. Chitnis, K. Toyama, K. Ooi, M. Phiong, and M. Koenig. Managing Microfinance with Paper, Pen and Digital Slate. In *ICTD 2010*.

[19] Remark Office OMR software. http://www.gravic.com/remark/officeomr/index.html.

[20] A. Seth. Udai OMR tool. http://www.cse.iitd.ernet.in/ aseth/udai/OMRProj/README.html.

[21] G. Singh, L. Findlater, K. Toyama, S. Helmer, R. Gandhi, and R. Balakrishnan. Numeric Paper Forms for NGOs. In *ICTD 2009*.

[22] SQS: Shared Questionnaire System. http://dev.sqs2.net/projects/show/sqs.

[23] S. Van Wart, K. Tsai, and T. Parikh. Local Ground: A Paper-Based Toolkit for Documenting Local Geo-spatial Knowledge. In *DEV 2010*.

[24] VillageReach. http://villagereach.org/.