

©Copyright 2015

Nicola Dell

Mobile Camera-Based Systems for Low-Resource Environments

Nicola Dell

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Linda Shapiro, Chair

Richard Anderson

Jacob O. Wobbrock

Program Authorized to Offer Degree:
Computer Science and Engineering

University of Washington

Abstract

Mobile Camera-Based Systems for Low-Resource Environments

Nicola Dell

Chair of the Supervisory Committee:
Professor Linda Shapiro
Computer Science and Engineering

The suitability of mobile devices for data collection and decision support in developing countries has been well established. It is now relatively common for field workers to carry devices that help them decide what questions to ask and how to record the resulting data. Many existing data collection tools also allow users to gather rich data from built-in sensors, such as taking photos using the device's camera. Typically, this data is simply stored on the device or uploaded to a server for analysis. However, there are many potential benefits to be gained by analyzing data collected from sensors immediately on the device, such as lowering the cost and bandwidth required for data transmission, enabling immediate analysis in the absence of reliable connectivity, and delivering the results of analysis more quickly than waiting for responses from an external server or off-site human expert. This dissertation describes the creation of systems that run on commercially available mobile devices and that use camera-based input in conjunction with computer vision and machine-learning to improve data collection and disease diagnosis in remote areas. In particular, we:

- Show that commercially available mobile devices are capable of capturing high-quality images and videos that can be processed using computer vision and machine-learning techniques running locally on the device.
- Identify and overcome the technical challenges associated with designing and implementing algorithms to interpret camera-based input on the device and quantify the technical performance of these algorithms.

- Develop an approach to building camera-based systems that allows users to specify how the system should interpret images without needing to recompile the software.
- Demonstrate the viability of our approach by applying our methods to two different problem domains: automatically digitizing data from paper forms, and automatically interpreting diagnostic tests.
- Prove that our systems can be effectively integrated into existing information ecosystems in low-resource settings and demonstrate that they are usable and appropriate under the constraints experienced at all levels of the information hierarchy.

Taken together, these contributions demonstrate that mobile, camera-based systems could alleviate some of the burdens faced by field workers in low-resource settings. Moreover, in addition to identifying and overcoming the technical challenges associated with building systems for low-resource environments, we have also developed a deep understanding of the human challenges - cultural, linguistic, and social - that impact the research process. Through understanding and tackling these challenges, this dissertation contributes a new approach for designing and building technologies for underserved communities and provides evidence for how this approach can be used to strengthen information and health-care systems in developing countries.

TABLE OF CONTENTS

	Page
List of Figures	iii
List of Tables	vi
Chapter 1: Introduction	1
1.1 Automatically Digitizing Data from Paper Documents	3
1.2 Automatically Interpreting Diagnostic Tests for Infectious Diseases	4
1.3 Contributions	6
Chapter 2: Related Work	7
2.1 Paper Forms and Digital Documents	7
2.2 Point-of-Care Diagnostics	15
2.3 Summary	21
Chapter 3: A System for Digitizing Data from Paper Forms	22
3.1 Introduction	22
3.2 Analyzing Paper-Digital Workflows in Global Development	24
3.3 A System for Automatically Digitizing Data from Paper Documents	46

3.4	Techniques to Improve the Transcription of Non-Machine-Readable Data . . .	63
3.5	Field Evaluation of a Mobile System for Digitizing Paper Documents	81
3.6	Summary	101
Chapter 4: A System for Analyzing Diagnostic Tests for Infectious Diseases		102
4.1	Introduction	102
4.2	A System for Automatically Analyzing Rapid Diagnostic Tests	104
4.3	Field Evaluation of a Camera-Based Point-of-Care Diagnostic System	123
4.4	Automated Analysis of Time-Sensitive Diagnostic Test Data	141
4.5	Summary	152
Chapter 5: Conclusion		153
Bibliography		155

LIST OF FIGURES

Figure Number	Page
3.1 Our prototype design probe allows people to create a variety of machine-readable question types and incorporate content generated by other tools (such as images).	26
3.2 The four phases of the paper-digital lifecycle in the context of global development work: design, field work, data entry, analysis.	28
3.3 Two surveys that highlight how some of our participants optimize for clarity (left) and others optimize for density (right).	30
3.4 Part of a survey created for use by rural farmers in Tanzania.	31
3.5 Part of a survey that shows how field workers take advantage of paper's flexibility, which may make data entry challenging.	35
3.6 Paper form used to tally the number of vaccines administered monthly at a health center in Mozambique.	47
3.7 Current paper-based workflow (left) and proposed ODK Scan workflow (right).	49
3.8 Portion of a partially filled bubble form.	50
3.9 Low-cost plastic stand and form tray used to capture images of paper forms.	51
3.10 A marked up image of a processed form.	54
3.11 Sample images showing each lighting condition tested.	56
3.12 Sample images showing the folded and dirty conditions tested.	58
3.13 Sample images showing background conditions tested.	59

3.14	A field coordinator explaining how ODK Scan works to a health worker. . . .	61
3.15	The two methods tested on the desktop platform: (a) Snippets and (b) Baseline.	64
3.16	The two methods tested on the mobile platform: (a) Snippets and (b) Baseline.	65
3.17	The report card paper form that we used for the study.	69
3.18	Average time taken to complete data entry for one paper form in the desktop study. Lower is better. Error bars show ± 1 SD.	72
3.19	Average time taken to complete data entry for one paper form in the mobile study. Lower is better. Error bars show ± 1 SD.	72
3.20	Results of the NASA TLX survey for the desktop study. Lower is better. Error bars show ± 1 SD. ** indicates $p < .05$; * indicates $p < .1$	74
3.21	Results of the NASA TLX survey for the mobile study. Lower is better. Error bars show ± 1 SD. ** indicates $p < .05$; * indicates $p < .1$	74
3.22	The monthly workflow for recording and digitizing CHW commodity consumption data [141].	85
3.23	The two-page ODK Scan compatible CHW commodity consumption form. . .	87
3.24	Percentage of form fields with differences between ODK Scan tallies and groundtruth tallies using the practitioner’s form description and the expert’s form description with additional alignment features enabled.	92
3.25	Excerpt of a form showing extra bubbles drawn in by the CHW.	96
4.1	A portion of the paper-based malaria RDT job aid.	106
4.2	Low-cost 3D-printed platform used to capture images of RDTs.	108
4.3	Screenshots from the digital job aid for the malaria RDT.	111
4.4	Regions of interest on an RDT for HIV.	112

4.5	We tested four different RDTs. From top to bottom: SD Bioline malaria, First Response malaria, SD HIV 1/2 3.0, and SD 1/2 Multi-device HIV.	116
4.6	Two marked-up SD Bioline RDTs: a properly interpreted RDT (top) and an overexposed RDT with two false negative lines (bottom).	120
4.7	Two marked-up SD HIV RDTs: a properly interpreted RDT (top), and an RDT with a false negative test line due to condensation on the inside of the cartridge (bottom).	120
4.8	The system’s user interface showing a processed test for malaria.	126
4.9	Number of malaria tests captured by each study site.	132
4.10	Proportion of each site’s total tests collected each week of the study.	133
4.11	Percentage of each site’s tests that contained each type of user error.	135
4.12	A test that contains too much red background to yield a valid diagnosis.	136
4.13	Proportion of each site’s tests transmitted during each week of the study.	137
4.14	Summary of agreement between (A) the system and visual diagnoses, (B) the review and visual diagnoses and (C) the review and system diagnoses.	138
4.15	Device setup showing the location and scale of the on-card intensity standards, registration marks and immunoassay capture spot.	142
4.16	System architecture and workflow.	144
4.17	Screenshot of a processed video frame.	146
4.18	Normalized intensities over time using (a) MATLAB and (b) a mobile device.	149
4.19	Processing rates on each device and in MATLAB.	150
4.20	Processing rates when calculating locations of the capture spot, registration marks and standards (A) every frame and (B) on the first frame only.	151

LIST OF TABLES

Table Number	Page
3.1 ODK Scan performance under different lighting conditions	56
3.2 ODK Scan performance with folded and dirty forms	58
3.3 ODK Scan performance with different background conditions	59
3.4 Bubble classification accuracy on a sample of 35 consumption forms using the practitioner’s form description and the expert’s form description.	91
4.1 Experimental results for each of the RDTs tested.	119
4.2 Summary of study sites, participants and devices.	129

ACKNOWLEDGMENTS

This document is the culmination of several years of work that would not have been successful without the guidance, support, and collaboration of many wonderful people.

First, I am incredibly grateful to my advisor, Gaetano Borriello, for his guidance, advice, support, encouragement, and friendship. Gaetano was my greatest champion and always came through for me, often at great personal sacrifice to himself. Gaetano's vision showed me the path that I should be on, his wisdom and courage taught me how to choose to work on the right problems for the right reasons, and his humility and sense of humor taught me how to accept failure gracefully. I will forever aspire to be the kind of person that Gaetano was and I hope that my subsequent career will make him proud.

I am also grateful to the other members of my dissertation committee. To Linda Shapiro, thank you for always being there for me from the very beginning to the very end of grad school and for encouraging me to follow my heart. I would not be where I am without you. To Jake Wobbrock, thank you for opening my mind to the wonders of HCI. You broadened my perspective and taught me how to think things through. To Richard Anderson, thank you for your encouragement and for not hesitating to step in when I needed you. To Kate Starbird, thank you for your insight and valuable support.

Thank you also to all of my collaborators on the various parts of this work: Jeff Beorse, Nathan Breit, Jessica Crawford, Timoteo Chaluco, Krittika D'Silva, Ian Francis, Claire Illich, Sarah Jackson, Clarice Larson, Mitchell Lee, Rachel Powers, Munjela Rumanna, Chip Sheppard, Sugandhan Venkatachalam, and Paul Yager. Many others also provided valuable feedback and discussions along the way; it's been a tremendous privilege working with and learning from all of you.

I would like to thank past and current members of the UW CSE ICTD Lab for their collaboration, feedback, advice, insights, humor, and friendship: Yaw Anokwa, Waylon

Brunette, Rohit Chaudhri, Camille Cobb, Brian DeRenzi, Mayank Goel, Carl Hartung, Neha Kumar, Clarice Larson, Trevor Perrier, Fahad Pervais, Sam Sudar, Mitch Sundt, and (last but not least) Aditya Vashistha. Thank you to Lindsay Michimoto for first getting me to UW CSE and then getting me through, and to Julie Svenson for her invaluable help with all administrative things. I'm also grateful to many other past and current members of the department for their feedback and advice. Thanks for making UW CSE my home.

Beyond UW CSE, I am especially grateful to Ed Cutrell, Bill Thies, Desney Tan, and Scott Saponas for their ongoing guidance, support, and mentorship. There is no way to express how much it means to have people like you on my side.

My happiness in Seattle has largely been due to the unwavering love and support of my phenomenal friends. Thank you to Brandon for all the adventures (big and small) and for always being willing to have a glass of wine with me. Your spontaneity and enthusiasm for life make my life better. Thanks for loving me for who I am. I hope you knock 'em dead.

I am especially grateful to my girls: Nell O'Rourke, Franzi Roesner, and Brenna Novak who make everything better, always. Thanks for having my back, for celebrating the good times with me, and for holding me up through the hard times. Your friendship gives me the courage to face life head on. I'd also like to thank the rest of the gang: Greg Akselrod, Tony Fader, Pete Hornyack, Cliff Johnson, and Todd Schiller. Thanks for amazing trips around the world, dinner parties, grilling, games nights, ski days, happy hours, and endless conversations. I look forward to countless more adventures with you all.

Finally, thank you to my parents and sisters. Your unconditional love and belief in me have made me who I am. I hope I make you proud.

The work described in this dissertation were supported in part by funding from a Graduate Facebook Fellowship, a Google Anita Borg Scholarship, several grants from the Bill and Melinda Gates Foundation, Econet Wireless Zimbabwe, NSF Grant 1111433, the Defense Advanced Research Projects Agency (DARPA), and a gift from Google.

DEDICATION

To Gaetano for changing the world and for teaching me that I can do the same.

*All travelers escape the mainland here.
The same geology torn from the stretch
Of hostile homelands is a head of calm,
And the same sea that pounds a foreign beach
Turns strangers here familiar, looses them
Kindly as pebbles shuffled up the shore.*

*Each brings an island in his heart to square
With what he finds, and all is something strange
But most expected. In this innocent air
Thoughts can assume a meaning, island strength
Is outward, inward, each man measures it,
Unrolls his happiness a shining length.*

*And this awareness grows upon itself,
Fastens on minds, is forward, backward, here.*

*The island focuses escape and free
Men on the shore are also islands, steer
Self to knowledge of self in the calm sea,
Seekers who are their own discovery.*

— **Elizabeth Jennings, The Island**

Chapter 1

INTRODUCTION

One of the great challenges in science and engineering today is to develop technologies to improve the lives of people in the poorest regions of the world. To date, many of the world's biggest technological advances have primarily benefitted people living in developed countries, like North America and Europe, that contain only 17% of the world's population. By contrast, Africa and Asia hold 75% of the world's population, but these areas often lack reliable wired technological infrastructure, including electricity or Internet connection [140]. In addition, people living in these regions typically face a higher burden of disease and have less resources to overcome it. For example, Sub-Saharan Africa experiences 25% of the global burden of disease, but has only 3% of the world's health professionals [147].

To address this shortage of trained professionals, government, social, and health organizations are increasingly relying on lightly-trained community health workers to deliver critical health services to communities in rural regions [95]. Health workers are members of the local community that typically visit households to assess, treat, and refer patients, and collect monitoring and evaluation data about the health programs in which they are engaged. Increasingly, health workers are being provided with mobile devices to assist them with their duties. Mobile devices are portable, battery powered, and can work in areas that lack reliable technological infrastructure, including electricity or Internet connection. Many mobile devices have intuitive touch-screen interfaces that make them easier to use than desktop computers for novice users, and a variety of built-in sensors and network interfaces that make them capable of supporting a wide range of social, administrative and health applications. Mobile devices also have ample storage capacity, and are capable of storing data locally on the device and uploading it to a remote database if and when a cellular or data network becomes available. Prior work has shown that mobile devices can be used to remind health workers about their daily tasks or the patients they are due to visit [41],

provide decision support to ensure adherence to health protocols [42], and assist with data collection and transmission [65].

In addition to recording basic text and numbers, many mobile data collection tools provide users with the ability to record rich data, such as audio, GPS coordinates, images, and videos. This rich data is typically stored on the device or uploaded to a remote server for analysis. However, there are many potential benefits to be gained by analyzing the data collected from sensors immediately on the device. For example, uploading video or image data to a server requires a large amount of bandwidth, which increases the time and cost of data transmission. By analyzing the data directly on the device we can lower the amount of data that needs to be transmitted and only upload a textual summary of the analysis or the relevant parts of an image or video clip. Prior work has also explored the potential for images and videos to be transmitted and analyzed by a remote server [103] or off-site human expert [118]. However, both of these approaches require the presence of a reliable Internet connection that may not be available in many remote environments. In addition, obtaining a response from an off-site human expert often requires a health worker to wait for hours or days, whereas processing the data immediately on the phone could make the same information available within seconds or minutes.

Finally, despite the small amount of training that health workers typically receive, the success of health worker programs has resulted in a rapid expansion of the range of tasks that workers are asked to perform. For example, health workers are increasingly being asked to administer diagnostic tests for infectious diseases, and make complex diagnostic decisions for a growing number of diseases and medical conditions [66]. In addition, health workers are often required to collect and report large amounts of monitoring and evaluation data for government and aid organizations, which is an extremely time-consuming process.

This dissertation demonstrates how mobile, camera-based systems can be used to alleviate some of the burdens placed on health workers in low-resource settings. In particular, we have designed, implemented, and evaluated mobile systems that run on cheap, commercially available devices (such as smartphones and tablets) and that use computer vision and machine-learning techniques to automate tasks that were previously tedious or error prone. We targeted two primary problem domains: improving data collection by automati-

cally digitizing data from paper documents [34, 35, 36, 38], and improving disease diagnosis by automatically interpreting diagnostic tests for infectious diseases [33, 37, 40]. We employed a common approach to interpreting camera-based input for both these domains and quantitatively evaluated our algorithms through rigorous laboratory studies. We also spent substantial time working with global development organizations, government ministries, and health workers in the field to ensure that our systems were usable and appropriate under the constraints experienced at all levels of the information hierarchy in low-resource settings. The remainder of this chapter summarizes our work in each of our two main problem domains and describes the high-level contributions made in this dissertation.

1.1 Automatically Digitizing Data from Paper Documents

Global development organizations working in low-resource environments rely on large-scale data collection to measure their impact and control the quality of the services they provide. In regions constrained by poor infrastructure and limited resources, this data must often be collected in paper - not digital - format. Paper is tangible, portable, and does not require batteries. It is a well-understood and trusted medium, and the low cost and ease-of-use of paper suggest it will continue to be extensively utilized throughout the world for years to come. However, data in paper formats is difficult to navigate, process, and store, and many of the complex analyses and visualizations that now routinely help people make sense of the data are feasible only if it is in communicable, searchable, and mutable digital formats.

Chapter 3 describes our work that aims to close the gap between the paper and digital worlds by creating mobile imaging software that automates the capture of digital data from paper. By studying the paper-digital workflows that drive global development organizations, we found that these organizations currently spend many months and thousands of dollars designing paper forms that contain a mixture of data types, including fill-in bubbles, checkboxes, numbers, and text. These forms are printed and distributed to field workers who fill them out by hand. Completed forms are then collected and data entry workers manually transcribe the handwritten data into a computer, often with poor accuracy [38]. To improve these workflows, we designed a mobile, camera-based system that automatically interprets images of paper documents that have been captured with the device's built-in

camera [34]. Although standalone systems that classify individual data types do exist, such as Scantron for fill-in bubbles, these systems typically require specialized hardware and/or large amounts of computational power. Our solution is capable of classifying a mixture of different data types and performs all of the required computation on the mobile device. Digitized data is stored locally on the device and transmitted to a server if and when a network connection is available. The system is robust to variations in lighting and handles dirty, folded, or messy documents by running a multi-stage computer vision algorithm to accurately identify, align, and classify form elements. We also created a lightweight document description language and graphical tools to allow non-technical users to create machine-readable paper documents that can be digitized by the system.

To evaluate our work, we conducted several deployments with health workers in Mozambique. In the first deployment, health workers with minimal training were able to use the system to digitize clinics' monthly vaccine statistics in about 30 seconds with over 99% classification accuracy [34]. Following this, we deployed the system to track 45 health workers' usage of essential medical supplies over a four-month period. We analyzed the impact of the technology with a variety of stakeholders and showed the importance of considering the usability and acceptability of the system at multiple levels of the information hierarchy, such as with NGO workers and Ministry of Health employees, in addition to end-users [36]. We also quantitatively demonstrated that the system affords significant time-savings over manual data entry [35]. This work has resulted in a concrete software artifact, Open Data Kit (ODK) Scan, that we have released as a free, open-source tool for the community.

1.2 Automatically Interpreting Diagnostic Tests for Infectious Diseases

Health workers in remote settings often lack access to convenient, affordable, and usable diagnostic technologies that could help them to quickly diagnose and treat infectious diseases. To address this challenge, low-cost and portable rapid diagnostic tests have been developed and are now routinely used throughout the world to diagnose a variety of diseases including malaria, syphilis, and HIV. However, although the potential benefits of these new diagnostic technologies are immense, little attention has been paid to the challenges faced by the health workers responsible for administering the tests and interpreting their results. Human inter-

pretation is subjective, and health workers often lack confidence in their ability to read tests correctly. In addition, interpretation of test results varies across diseases and test brands, which increases the potential for human error. Moreover, medical researchers are currently developing more complex tests whose results will require quantification or time-sensitive analysis that will be very difficult or impossible to interpret by eye.

Chapter 4 describes our research that aims to help health workers with the process of accurately interpreting diagnostic tests. In particular, we created a mobile system that uses the device's built-in camera to capture a photograph of the test, after which computer vision algorithms running on the device automatically analyze the image and compute the diagnosis [33]. Using the system, analysis of diagnostic tests can become a standardized, auditable, and adjustable process that can be changed without retraining users. Commodity devices can analyze many different tests, negating the need for specialized reader devices, and test manufacturers and clinicians can easily add new tests to the system when they become available. The system also collects and transmits patient and disease data to a central database (if and when a network connection is available) to provide timely, accurate statistics that could aid outbreak detection, supply chain management, evidence-based decision-making, health system evaluation, and global disease monitoring and control efforts.

We evaluated the performance of the system through a laboratory evaluation. We used biological samples collected from patients to create a wide variety of positive and negative results for a range of HIV and malaria tests [33]. After the laboratory evaluation, we deployed the system with 60 health workers at five hospitals and clinics in Zimbabwe [37]. Our findings showed that health workers were successfully able to integrate the system into their clinical workflow after only 60 minutes of training. They used the system to analyze over 1800 malaria tests in a two-month period and devised strategies to overcome poor network connectivity to transmit test data to a centralized database. These results demonstrate the feasibility of using mobile technologies to aid disease diagnosis in remote settings. Finally, we have also conducted preliminary research that analyzes the results of tests with time-sensitive signals that must be continuously monitored and quantified. The system that we built was able to capture and analyze video frames of the tests in real-time and quantify the results, with all of the computation performed locally on the device [40].

1.3 Contributions

This dissertation describes the creation of systems that run on commercially available mobile devices and that use camera-based input in conjunction with computer vision and machine-learning to improve data collection and disease diagnosis in remote areas. We discuss related literature in Chapter 2, present our research in Chapters 3 and 4, and conclude in Chapter 5. In addition to the individual contributions described in Chapters 3 and 4, we also make the following high-level contributions:

- We show that commercially available mobile devices are capable of capturing high-quality images and videos that can be processed using computer vision and machine-learning techniques running locally on the device.
- We identify and overcome the technical challenges associated with designing and implementing algorithms to interpret camera-based input and quantify the technical performance of these algorithms.
- We develop an approach to building camera-based systems that allows users to specify how images should be interpreted without needing to recompile the system.
- We demonstrate the viability of our approach by applying our methods to two different problem domains: automatically digitizing data from paper forms, and automatically interpreting diagnostic tests.
- We prove that our systems can be effectively integrated into existing information ecosystems in low-resource settings and demonstrate that they are usable and appropriate under the constraints faced by users at all levels of the information hierarchy.

Taken together, these contributions demonstrate that mobile, camera-based systems could alleviate some of the burdens placed on health workers in low-resource settings. Moreover, in addition to tackling the technical challenges associated with building systems for low-resource environments, we have also developed a deep understanding of the human challenges - cultural, linguistic, and social - that impact the research process. By understanding and overcoming these challenges, this dissertation provides a new approach that shows how data collected from the built-in sensors on mobile devices can be immediately analyzed and used to strengthen information and healthcare systems in developing countries.

Chapter 2

RELATED WORK

There is a range of research that demonstrates the potential for mobile technologies to assist health workers in the field. Categories of interventions include building tools for data collection [65, 88, 92], providing decision support [5, 42, 98], informing people about health issues [23, 113] and promoting behavior change [57, 108]. We center our discussion of related research around our two primary problem domains: the interaction of paper forms and digital documents and systems for point-of-care diagnostics.

2.1 Paper Forms and Digital Documents

The long-term co-existence of paper forms and electronic documents has resulted in a large amount of research that examines their interaction. We focus on three main areas of prior work: integrating the affordances of paper and digital materials, automated document processing, and digitizing paper in the developing world.

2.1.1 Integrating Paper and Digital Worlds

According to Harper and Sellen, “*Observation of any organizational setting only serves to confirm that the most pervasive, ubiquitous artifact in support of collaborative work is paper*” [63]. The fact that paper and digital materials are both widely used suggests that each medium provides essential affordances that the other does not [10, 121]. Paper is cheap and has inherent material properties that make it easy to use and access in almost any environment. Digital documents are easy to store, search, link and analyze. However, coordinating information across paper and digital materials has proven challenging, and a large body of prior work examines and aims to bridge this paper-digital divide.

Luff et al. [82] examined the relevance of paper in collaborative workflows in an architectural practice, a medical center and a London Underground control room. Harper and

Sellen used ethnographic methods to assess the role of paper in an air traffic control room, a police station and the International Monetary Fund [63]. Other studies examine the use of paper and digital media in healthcare [52], particularly with respect to medical records [51] and information flow [11]. These studies emphasize that, despite efforts to replace paper with digital alternatives, paper remains an integral component of any work environment.

Proposed technical solutions focus on personal document management and new ways of integrating the affordances of paper and digital materials. The Digital Desk [146] was the first system to augment digital documents with the qualities of paper. The digital desk uses a desk, a video camera, a computer and a video projector, and operates by projecting a virtual desktop on to a table to provide some characteristics of a physical desk. A variety of systems have since followed the digital desk. Koike et al. propose an Enhanced Desk [74, 119] that uses computer vision to automatically retrieve and display digital information about real objects on the desk. The system also enables direct manipulation of digital objects by the users' hands and fingers. PaperSpace [128] uses visual codes to track the physical locations on the desk of previously printed academic papers to try and avoid unnecessary reprinting of documents. PaperLink [6] is a system that consists of a camera mounted on a highlighter pen. Users make marks on the paper using the highlighter, and computer vision techniques allow these marks to have meaning in the electronic world.

Many of these systems have contributed substantially to the design space of using camera-based input to enable interaction between the physical and the digital worlds. However, they all target the simultaneous or cyclical use of paper-based data and digital data, and focus on allowing users to easily transition between interacting with paper documents and interacting with digital documents. This interaction model differs from our work, which focuses on capturing bulk amounts of data recorded on paper forms and making it available in a usable digital format for analysis and aggregation. In addition, most of these solutions use a video camera combined with a desktop computer to facilitate interaction in an office environment. By contrast, our work uses mobile devices to facilitate the collection of digital data in remote areas where users may not have access to reliable infrastructure.

Another class of related systems integrate paper and digital documents using printed visual codes, digital pen devices, or a combination of both. For example, Xax [71] and

Protofoil [114] are two systems that together comprise a paper user interface. Each form that is part of the paper user interface is marked with logos, a form ID and registration marks, all of which are machine-readable and serve to identify the form to the system. Paper forms are either scanned or faxed to become integrated with the electronic world. The Paper PDA [67] is a system that allows a paper notebook to be synchronized with an electronic document by periodically scanning the paper and then printing a new copy. To facilitate scanning, each page in the Paper PDA is assigned a unique page identifier and printed with two registration marks. By periodically printing fresh copies, items from the electronic world can be integrated into the paper notebook.

Paper Augmented Digital Documents (PADDs) [61] are documents that can be manipulated either on a computer screen or on paper. In the digital world, PADDs are accessed and edited using a computer. In the paper world, PADDs record marks made on them using a digital pen. The recorded marks can then be merged with the digital document. Intelligent Paper [44] uses a code to identify each sheet of paper and a pointer to specify points of interest. Intelligent Paper pages are standard sheets of paper entirely covered with printed marks that are only visible to the pointer. When the user positions the pointer on a page, the pointer is able to uniquely identify both the page and the location of the pointer, and transmit this information for interpretation or synchronization with digital resources.

The idea to entirely cover the paper with printed, machine-recognizable patterns has been commercialized by the Anoto Digital Pen and Paper technology [110]. Paper forms are digitally authored and then printed on paper with the Anoto microdot pattern. A digital pen that has been integrated with an infrared camera is then able to interpret the microdot data and track the movement of the pen across the paper. Many digital pen and paper based research projects have been built on top of the Anoto technology, such as PaperCraft [80] and Paper++ [83]. Several commercial companies have also licensed the microdot pattern from Anoto and used it in conjunction with new digital pen technology. For example, the LiveScribe Pulse smartpen [96] integrates audio recording into the pen, which allows users to associate audio recordings with pen-marked annotations on the paper. Users can then use the pen and paper to identify and play back the audio that was recorded when an annotation was created [120].

These digital pen systems provide a robust solution to the problem of combining paper and digital data. However, they also have a number of drawbacks when compared with our work, which uses commercially available mobile phones. First, mobile phones have been widely adopted in developing countries. As a result it is easy to purchase and maintain mobile phones in these regions, which cannot be said for specialized digital pen devices. In addition, the digital pen does not incorporate much of the additional functionality, such as built-in network interfaces, touchscreen capabilities and other sensors, that come pre-packaged with most mobile phones. The microdot paper must also be licensed commercially, and the paper forms must be printed with the microdot pattern prior to being used. By contrast, our system does not rely on specially printed paper and can work with paper forms that have been printed using any printer.

2.1.2 Automated Document Processing

Several fields of research focus on using computers to automatically interpret paper documents. Optical mark recognition (OMR) [151] uses reflected infrared light to detect the presence or absence of marks on paper forms. Traditional OMR systems work with a specialized scanning device that shines a beam of light through the form. The device can detect marked areas on the form because they reflect less light than blank areas. Commercial OMR solutions, such as Scantron [27], can operate with over 99% accuracy if the forms are filled out neatly and correctly using an appropriate pencil. However, commercial solutions are typically expensive, costing thousands of dollars for the large, specialized scanner [112] and approximately 10c per page for the printed forms [132]. This cost prohibits Scantron from being a viable option for many organizations in developing countries.

Commercial software, such as the Remark Office OMR solution [116], has also been developed that enables OMR on a desktop computer using a scanner. Although cheaper than the Scantron solutions, this software is still expensive, with the Remark Office OMR software currently costing about USD \$1000. There are also a few open source OMR software solutions, such as the Shared Questionnaire System [75] and the Udai OMR tool [122]. These systems differ from our work in a number of ways. First, they have all been

designed to work with a desktop computer and scanner rather than with a camera-equipped smartphone. In addition, the forms have to be filled in neatly and the form kept clean, and the paper forms must be specially designed for use with the system. By contrast, our solution is capable of operating on either a computer or an android smartphone, is robust to the forms being dirty or folded, and can handle the usage of different pens, pencils, cameras and scanners. In addition, we have developed a graphical tool that assists people with the process of generating scannable forms.

Optical character recognition (OCR) is a well-known topic in the field of pattern recognition. To perform OCR on a document, the first step is to analyze the layout of the document and extract lines of text. Each text line is then segmented into isolated individual character images that are interpreted using a classifier. OCR is a relatively straightforward process for well-printed, clearly-typed documents [16] and there are many commercial and open-source OCR systems available such as ExperVision [46] and Tesseract [72]. However, many of these OCR systems perform poorly if the documents are of low quality or have characters that touch or overlap [16]. For example, a study that focused on recognition of historical newspaper pages found that the accuracy of commercial OCR software varied from 71% to 98% [68]. In addition, OCR of non-roman scripts, such as Arabic or Indian scripts, is still a challenging problem and active area of research [2] [76].

Furthermore, automatically recognizing handwritten text is generally considered to be a much more difficult problem than OCR of printed text due to the substantial variation in appearance of handwriting. Although the postal service has succeeded in creating a system that can accurately recognize postal codes [129], a general solution to the problem of accurate handwriting recognition remains an open research question for both Latin scripts [111] and non-Latin scripts [106]. As a result, the predominant method used to digitize handwritten data from paper forms remains manual data entry, in which people read information on paper and manually type it into a computer.

2.1.3 Digitizing Paper in the Developing World

Paper has long played a central role in the collection and communication of information within governments and other organizations. Gupta [62] and Hull [70] note that, by moving through a bureaucracy as documents, physical materials such as health registers, government notices, or written complaints gain status and trigger actions by officials. Veeraraghavan [143] describes how digitizing paper records can reduce corruption. Singh et al. [124] analyzed the use of paper forms in the context of non-governmental organizations. They investigated the balance between ease-of-use among intended populations and machine readability, and suggested that users with limited education and training tend to prefer numeric and multiple-choice forms since they can fill them more quickly and accurately than forms that require a lot of writing.

CAM [107] is a user-interface toolkit that allows a camera-equipped mobile phone to interact with paper forms. Each form field on the paper is tagged with a two-dimensional visual code (QR code) that serves as a reference to assist the user with transcribing the field. CAM is a powerful tool that can handle a variety of data types, although there are a number of features that distinguish CAM from our work. First, although the visual codes on the form help the user to identify each field correctly, users are still required to manually type all of the data into the phone. By contrast, our system is able to automatically interpret some data types so that users do not have to enter these data types manually. We also eliminate the need for users to possess and refer to the physical paper form by displaying a small image snippet on each form field on the screen.

Shreddr [18], which has been commercialized as Captricity [15], is another system that has contributed substantially to this design space. Users capture images of forms using a camera or scanner and Shreddr segments the image and assigns the recognition of individual fields into tasks that are performed by people via crowdsourcing. Our work adapts and builds on this research in several ways. First, although Shreddr can handle a wide variety of data types, it does not yet leverage the direct machine readability of some data types, although it does have a method of doing machine learning based on a sample of the crowdsourced answers [19]. In addition, a reliable Internet connection and sufficient band-

width are required for the effective use of a crowdsourcing platform, which is problematic for many organizations in developing countries. Although Captricity recently released a mobile version of the system that operates with iPhones running iOS, their mobile client still requires a continuous Internet connection, and users must upload all of the form images to the Internet so that they can be distributed to crowdsourced workers. Neither the mobile nor the desktop version of Captricity is able to provide organizations with the option of distributing the data entry to its own workers instead of crowdsourced workers. In addition, there are a variety of reasons why organizations in developing countries may find it undesirable to use a crowdsourced solution. Some organizations may only need to enter data from a few forms per day, rather than thousands of forms, which might not warrant the overhead required to set up a crowdsourced solution. Furthermore, for a variety of political and security reasons, many organizations would prefer that the data only exists locally within their organization and is not uploaded to the Internet. Many paper-based data collection efforts gather personal health information, which people may be reluctant to provide if they fear that it will be put online for others to see, even anonymously. Finally, employing local data entry workers may be cheaper than outsourcing the data entry, and additionally creates jobs within the local community.

PaperWeb [87] attempts to make it easier for people with limited technical knowledge to perform tasks (such as paying bills) online rather than on paper. The steps required to perform the task are pre-recorded by people who do know how to use a web browser and a computer. Each recorded task is saved in an online task repository along with a template of the corresponding paper document. To use the system, users can take a picture of the paper document using a camera-enabled phone. The captured image is transferred to the cloud and an image registration algorithm [45] automatically identifies the document type and retrieves saved tasks associated with the document. Relevant parameters required for the selected task are automatically identified by aligning the input image with the document template, extracting field sub-images, and interpreting them using OCR. The user can also correct the automatically extracted values if necessary. Once identified, the web task executes in the cloud and returns the results to the user. PaperWeb has a number of commonalities with our approach to digitizing paper. Like our work, PaperWeb requires the definition of

a template for each document that will be processed by the system. However, in contrast to our system, which processes images on the device, all of the processing for PaperWeb is done in the cloud. This means that users must have a reliable Internet connection and sufficient bandwidth to transfer captured images. In addition, PaperWeb relies on OCR of printed text and is unable to process hand-marked data.

Walking Papers [97] and Local Ground [142] are tools that allow 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 stamps. The maps are then scanned, and user markings extracted and overlaid on existing online maps to aid local planning decisions. Local Ground allows people to use special stickers or markings to add meaning to some of their annotations, but the system is currently not able to decipher freeform annotations made by users. Ratan et al. [115] present a financial record management application built on a digital slate device. The solution accepts handwritten input on paper 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. In addition, users preferred the system over a purely electronic solution because they liked having physical evidence of their transactions and understood the voice-based feedback provided by the system. Unfortunately, despite the initial promise shown by the system, the purchase and maintenance of specialized slate devices hindered its scalability and sustainability [93]. Finally, the PartoPen [138] is a system that aims to improve maternal health outcomes through careful labor monitoring. The system uses a LiveScribe digital pen [96] in conjunction with a paper form, called a partograph that has been printed with the Anoto microdot pattern [110] (described in section 2.1.1). The pen runs custom software that is capable of triggering audio alerts to remind health care workers to take routine patient measurements at specified time intervals. The system is also capable of interpreting marks made on the paper and alerting the attendant if conditions arise that require additional observation or intervention. The PartoPen is currently deployed in several hospitals in Kenya [139], although collecting data off the pens and integrating it with patient medical records or other data collection tools remains a significant challenge.

2.2 Point-of-Care Diagnostics

We focus on three categories of research related to point-of-care diagnostics: systems that support health workers in the field, point-of-care diagnostic technologies, and systems for processing point-of-care tests.

2.2.1 Health Worker Support Systems

The lack of trained medical professionals in developing countries means that community health workers are often relied on to provide critical health services in rural regions. This has resulted in a large amount of research that focuses on improving health worker effectiveness. For example, Rowe et al. [117] describe strategies for better health worker training, while DeRenzi et al. [41] target health worker reminder systems. Mobile devices have been used to aid decision making in a number of different scenarios. CommCare [95] is a mobile application that supports health workers as they provide home-based care and social support to patients, and ODK Clinic [5] is a mobile clinical decision support system that helps medical personnel manage the treatment of patients with chronic diseases such as HIV.

The potential for mobile phones to increase health worker adherence to clinical protocols has also been well studied. DeRenzi et al. [42] developed eIMCI, a mobile application that guides health workers through a digital version of the Integrated Management of Childhood Illness (IMCI) protocol. Mitchell et al. [98] developed a PDA-based system that guides health counselors step-by-step through a screening algorithm that determines whether or not a patient should be referred to a doctor. These systems differ from our research in that they provide health workers with a series of simple questions, and the treatment or diagnosis is recommended based on the answers to these questions, while our system uses computer vision algorithms to automatically interpret colored chemical signals displayed on rapid diagnostic tests.

Research has also shown that paper-based job aids can help health care providers perform preventative health tasks and reduce the resources needed for training. For example, Knebel et al. [73] survey medical studies in which the introduction of job aids increased the task performance of health care providers for a variety of tasks, including providing im-

munizations, cancer screening, smoking and cholesterol counseling, and reproductive health care. Harvey et al. [66] conducted a study in Zambia that compared the accuracy of RDT administration among three groups of health workers. The first group received only the RDT package instructions, the second group received only paper-based job aids, and the third group received paper-based job aids and training on RDT administration. The study results show that although the group that received job aids and training performed the best, the group that received job aids alone still administered a significantly higher proportion of RDTs correctly than the group that received the package instructions.

Our research aims to combine the benefits of job aids with the advantages of mobile technologies to create a mobile point-of-care diagnostic system that is appropriate for use in remote settings.

2.2.2 Point-of-Care Diagnostic Technologies

Biomedical engineers have traditionally focused on developing diagnostic technologies that address the needs of patients in the developed world [152]. As a result, many of the diagnostic tests that are routinely administered in well-equipped clinical laboratories are inappropriate for the settings encountered at the point of care in developing countries. Successful point-of-care diagnostic tests must be portable and produce rapid and accurate results without requiring significant user intervention or training. They must also cost little to produce and administer, and incorporate reagents that are stable without refrigeration [84]. To address these challenges, medical researchers have recently developed innovative rapid diagnostic tests (RDTs) that specifically target the needs of patients in developing countries [153]. These low-cost, disposable tests contain all of the elements required to process a biological sample at the point of care, so that the sample does not need to be refrigerated, protected, or transported. Additionally, many of the tests run rapidly, allowing medical personnel to view results and treat patients immediately. This speed is advantageous because many rural patients travel long distances to reach medical facilities and may be unable to return easily to collect test results, delaying treatment. The ease-of-use of RDTs has resulted in a wide variety of companies that market hundreds of different test cartridges,

including commercially available RDTs that can detect pregnancy, malaria, HIV, dengue, tuberculosis, syphilis, influenza, rotavirus, drugs of abuse and many more [55]. Most RDTs for infectious diseases are currently only used in developing countries, although OraQuick [105] is one rapid HIV test that was recently approved for home use in the United States.

The majority of RDTs that are currently used in developing countries are qualitative tests that have a simple positive or negative result. However, several companies are building multiplexed RDTs that test for multiple diseases on a single cartridge. For example, MedMira is developing a set of tests, called Multiplo tests, that can detect HIV and other sexually transmitted infections from a single biological sample [94]. Similarly, Diagnostics for All [3] is developing a cartridge that contains two different malaria tests and one dengue test on the same paper patch. Increasing the number of test results on a single cartridge is likely to increase the complexity of reading the results by eye and the number of errors that health workers make when they interpret the test results.

There are also an increasing number of tests being developed whose results require quantification. Examples include glucose strips [20] that measure sugar concentration, urinalysis tests [29] that measure a variety of different chemical substances in urine samples, and CD4 tests [32] that quantify an HIV patient's CD4 T-cell count. In addition, recent work by Stevens et al. [130, 131] demonstrated that analyzing the changing color of a test signal over time can increase the dynamic range of the test and improve the overall accuracy of the result. There is also growing interest in the potential to perform ratiometric testing [24], in which the quantity of one substance in the sample is divided by the quantity of a second substance, to produce a ratio that provides additional information about the health of a patient.

Manually interpreting the results of tests that require quantification, time-sensitive analysis or ratio-metric analysis is likely to be a complex and error-prone process, particularly for lightly trained health workers attempting to read the test results by eye. It will be difficult for health workers to simultaneously manage the timing sequences for the test, monitor the rates at which test signals are changing, and quantify the test result. These challenges suggest that point-of-care diagnostics could be improved through the development of tools that automatically process and interpret the test results at the point of care.

2.2.3 Processing Diagnostic Tests

A variety of techniques have been developed to overcome the shortage of well-equipped diagnostic laboratories in developing countries. Telemedicine is one approach that relies on existing communications infrastructure to provide clinical health care at a distance. For example, Martinez et al. [89] built a prototype system that uses paper-based microfluidic devices for running multiple tests simultaneously. The color intensity of each test spot is captured using a phone's camera or portable scanner, and established communications infrastructure is then used to transmit the images from the testing site to an off-site laboratory for analysis by a trained medical professional. Sana [118] is another telemedicine platform that consists of a mobile application that runs on the Android platform. Using the Sana application, health workers can run a procedure (such as screening for malaria or diabetes) and collect patient data. The Sana application then uploads the information to OpenMRS (an electronic medical record system) [150] for a doctor to review. After reviewing the data, the doctor can make a decision and notify the health worker of the diagnosis by sending results back to the Sana application. One of the main benefits of this approach is that the test outcome is decided by a properly qualified expert rather than a lightly-trained health worker or nurse. However, a significant disadvantage of the approach is that the trained expert is often an extremely busy doctor, and it may take days or even weeks for them to deliver the diagnosis. As a result, the test outcome is usually not available to health workers within a single patient visit to the clinic, which means that patients will be required to return to the clinic to collect their test results. However, since many rural patients travel long distances to reach medical facilities, they may be unable to return to collect test results, which could delay treatment, while other patients may not return to the clinic at all.

An alternative to remote diagnosis by an expert is automated analysis of tests that is done on a server. Images of the test are captured using a phone camera and transmitted to a cloud-based server for analysis. Mobile Assay [99] and Nexleaf Analytics [103] are two companies that employ this approach, and although analysis on a server is likely to be faster than waiting for analysis by an expert, the systems require reliable connectivity and sufficient bandwidth to be able to send and receive the data, which may not be available.

Another approach for automatically interpreting diagnostic tests is to use specialized reader devices. Automated readers have been shown to increase the sensitivity and overall accuracy of reported test results [136]. Most commercially available readers are designed to read a single test or group of tests from a single manufacturer. For example, many major diagnostic test companies offer some form of laboratory bench-top reader, such as the Aurora Automatic ELISA test workstation [22]. Although bench-top readers typically operate with high accuracy, they are usually large, expensive devices designed for use by trained professionals in high-tech laboratories under controlled conditions. These constraints often make them unsuitable for use in resource limited settings. An alternative to bench-top readers are handheld readers. Many currently available handheld readers have been designed for home use by people in developed countries, and generally target monitoring of metabolic diseases, such as diabetes, rather than detecting infectious diseases [54]. For example, the Clinitek Status Analyzer [4] is a handheld reader that processes Bayer Multistix [133] urine test strips for diabetes monitoring. Handheld readers recognize the value of using mobile technology to automatically process and store the test data, although few readers are currently capable of transmitting the data immediately from the device to a central server. Handheld readers are also usually expensive, costing thousands of dollars each to process one or a few tests from a single manufacturer.

Thus, instead of relying on specialized reader devices, a growing number of projects couple image processing on commercially available smartphones with diagnostic tests [154]. Mudanyali et al. [100] have developed a smartphone-based RDT reader platform that can work with several RDTs. Their platform differs from our research in a number of ways. First, although the system works with a variety of RDTs, each test type requires a custom built holder that fits tightly around the test cartridge and clips it to the rest of the platform. By contrast, our low-cost, 3D-printed stand is capable of fitting a wide variety of RDTs that have different shapes and sizes without the need for any additional parts. Their platform also requires a plano-convex lens, three LED arrays, and two AAA batteries to precisely control the environment in which the test image is captured. Our system does not require any of this additional hardware, which is advantageous since requiring additional parts increases the likelihood that something will get lost or broken and lower the reliability of the system or

render it non-functional. Finally, Mudanyali et al. [101] focus entirely on optically reading the RDT results, and do not consider the human challenges that administering RDTs may represent for lightly trained health workers, whereas our work aims to support the process of administering RDTs in addition to automatically reading the results.

Matthews et al. [90, 91] describe a paper-based dengue fever test that can be imaged and processed by a Windows smartphone. The paper test is small enough to fit on a single finger, and the test is processed by comparing the color of the paper to known color standards. Detected cases of dengue fever are reported to the Center for Disease Control [17] to facilitate outbreak detection. Although the algorithms and workflow presented in this research show promise, there is no evaluation of the accuracy of the color detection algorithm, and no usability or field studies that indicate if the system is appropriate for use by health workers and capable of performing well in the field.

Shen et al. [123] discuss an approach for quantifying colors of colorimetric diagnostic assays with a smartphone. An image of a test strip is captured using a smartphone camera, but instead of using the red, green, and blue color intensities in the image, their approach uses chromaticity values to construct calibration curves of sample concentrations. Their approach yields accurate results for quantifying pH values with a linear response range of 1-12. Wang et al. [145] describe the development of an ELISA test to detect an ovarian cancer biomarker in urine. The test is imaged with a cell phone camera and a mobile application used to calculate the pixel values (red channel) from selected regions of interest on the test. The color intensity is then mapped to a curve to calculate the concentration of biomarker in the sample. Both of these papers are interesting in that they discuss the problem of converting intensity values obtained through image processing to chemical values associated with biological sample concentrations.

There are several commercially available smartphone-based rapid diagnostic test readers. The FIO Corporation [26] has built a suite of tools to support health workers and supervisors with the process of administering RDTs. The FIO Deki reader [48] consists of a custom casing designed to control the environmental and lighting conditions around the test, and a set of proprietary android applications. The central application, called Clinic, guides users through the process of recording patient information and administering RDTs, and then

automatically interprets the RDT results. The test data is then transferred to FIO’s cloud-based server, called AirFio [47], or to a desktop application, called Spiri [49], where the data may be viewed by stakeholders. In addition to not requiring a custom casing to surround the phone and the test, our system differs from the FIO system in that we provide users with the ability to add their own tests and supporting materials to the system, whereas FIO requires that customers choose their RDTs from a set of FIO-supported tests. In addition, any FIO captured data is designed to remain solely within the FIO information ecosystem, whereas our system gives users the ability to export data in a variety of standardized formats, so as to be compatible with a wide range of existing analysis and visualization tools.

Skannex [125] is another company that markets proprietary test reader software called ReaditLateral [126]. The software automatically processes lateral flow tests that have been marked with an identifying barcode. One of their prototypes, SkanSmart [127], is a mobile device that runs a version of their software that has been adapted for smartphone processing. However, the smartphone is again entirely enclosed in custom built casing that precisely controls the environmental and lighting conditions. In addition, Skannex configures and manages the system for its customers, and does not give them the ability to add their own diagnostic tests to the system.

2.3 Summary

This chapter describes the rich body of literature and prior research on which this dissertation builds. We have centered our discussion on research that relates to our two primary problem domains: digitizing data from paper forms, and interpreting diagnostic tests. The following chapters present our original research understanding the challenges that health workers face in these two domains and designing, building, and evaluating mobile camera-based systems that tackle these challenges.

Chapter 3

A SYSTEM FOR DIGITIZING DATA FROM PAPER FORMS**3.1 Introduction**

Global development organizations have gradually but steadily evolved into an established presence across the world, particularly in low- and middle-income countries. Teams within these organizations are frequently present in multiple countries at once, engendering complex and collaborative work environments. Individuals with different backgrounds, perspectives, and responsibilities work together on the design, deployment, and evaluation of initiatives that aim to examine and/or improve the lives of socioeconomically disadvantaged populations around the world. At the same time, these initiatives require them to collaborate with various partner, donor, and government organizations to perform large-scale data collection in the communities they target so that they can measure their impact and control the quality of the services they provide.

The task of gathering accurate and timely data from people living in low-resource environments is challenging. In regions constrained by poor infrastructure and limited resources, where the availability of technology is limited and digital literacy is slowly rising, data must often be collected using paper - not digital - materials. People trust and are familiar with paper. Paper is tangible, portable and does not require batteries. These material affordances make it easy to use and almost universally accessible. As a result, paper has always provided essential support for organizational work processes and will remain an integral and critical component of global development workflows for years to come.

However, although paper has been used to collect data for hundreds of years, the onset of digital data has transformed the organizational workflow. Data in digital formats is easier to navigate, process, manage, and store than paper-based data, and many of the complex analyses and visualizations that now routinely help people to make sense of collected data are only feasible if the data is in communicable, searchable and mutable digital formats.

Despite the recent development of numerous digital tools to aid data collection, direct-to-digital data collection remains challenging for many organizations. The communities in which the data is collected often have insufficient IT infrastructure and support to facilitate effective digital solutions. 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. These challenges suggest a need for new tools that allow organizations to continue to use cheap, familiar paper forms for data collection, and then provide efficient techniques for automatically digitizing and storing the data.

This chapter describes our work designing, building, and evaluating a mobile camera-based system that aims to improve paper-digital workflows by automatically digitizing data from paper forms. In particular we contribute:

1. A qualitative study that analyzes paper-digital workflows in global development, highlights the challenges of transitioning data between paper and digital formats, and identifies opportunities for new tools to bridge the paper-digital divide [38].
2. The design and implementation of ODK Scan, a mobile camera-based system that uses computer vision techniques running on commercially available mobile devices to automatically classify machine-readable data types recorded on paper forms [34].
3. A controlled laboratory analysis that provides quantitative and qualitative data that shows our techniques afford significant time savings without loss of accuracy [35].
4. A four-month deployment that documents the benefits and challenges of using the system as part of an intervention to strengthen the community health worker medical supply chain in two districts in Mozambique [36].

Taken together, these contributions demonstrate the potential for our work to transform paper-digital workflows, make data collection more accurate, efficient, and affordable for global development organizations, and provide numerous insights for other researchers and practitioners who work with diverse communities in low-resource environments.

3.2 Analyzing Paper-Digital Workflows in Global Development

This section examines the tensions between the ubiquitousness of paper and the desirability of digitized data as we analyze the collaborative practices surrounding paper-digital workflows in global development organizations. We use a mixed methods approach to study the paper-digital lifecycle from the perspective of the researchers and practitioners that drive development initiatives, organizing our findings in relation to the different stages of the data lifecycle as it is sought, collected, digitized, and analyzed. We show that workers within these organizations spend considerable time and effort transitioning data between paper and digital materials, while attempting to preserve the correctness and completeness of the data. In addition, we highlight workers' frustrations with the software tools that they use to collect, communicate and synchronize data between remote offices, and discuss design opportunities for new tools to better support their workflows. Finally, we reveal the impact of infrastructural, cultural, and socioeconomic challenges on the paper-digital lifecycle in global development work, many of which can only be understood, not eliminated.

3.2.1 Methodology

This study contributes an in-depth understanding of paper-digital workflows within the global development context by examining the work practices of organizations, particularly as they relate to the use of paper and the move towards digitization. The methods we use to arrive at this understanding include an online survey, a design probe, and a set of interviews. We describe these in detail below.

Online Survey

Our survey consisted of 50 questions that sought information regarding the demographic background of participants and their organizations, their survey design and approvals processes, data collection and entry, and work practices. Our 48 participants from 23 organizations were situated in 16 countries and have been active in a variety of domains, including health, education, agriculture, finance, and logistics. Their responses gave us a global perspective on the relevance of paper and the importance of digitization in these domains.

Design Probe

Based on our survey responses, we prototyped a new survey design tool to expand our understanding of the design process within our target organizations. The tool was designed with the single objective of ease of digitization and allows participants to create a variety of question formats that are optimized for machine-readability (see Figure 3.1). Participants can move and reformat individual questions, and import content from a variety of other sources, including from existing images or Microsoft Word. Our goal was to use the tool as a probe [58] to elicit survey designers' views on the process of converting research questions into material paper surveys and the value of optimizing for digitization and machine-readability.

Interviews

Participants. We recruited nine participants from three global development organizations. All participants were female, aged between 27 and 37 years, and selected based on (1) having participated in our survey; (2) being actively engaged in designing paper-based surveys for global development initiatives; and (3) being willing and able to participate in our study. We interviewed both researchers and practitioners, since both play vital roles in global development workflows but are somewhat differently motivated. Participants who identified themselves as practitioners were usually employed by NGOs concerned with applying existing knowledge and technologies to improve the quality of life of people in resource-poor settings. By contrast, participants who described themselves as researchers usually worked for academic institutions and were motivated by high-level research questions that they hope will be answered through field studies. These participants were concerned about ensuring the validity of the study, enrolling well-defined samples of specific target populations and controlling for external variables. Having both kinds of participants gave us a nuanced understanding of the workflows involved in global development initiatives.

Finally, all our participants are survey designers who work with diverse and geographically dispersed teams, but who are themselves primarily based in the US. Our findings therefore document the priorities and perspectives of these survey designers with regard to the paper-digital lifecycle within their organizations.

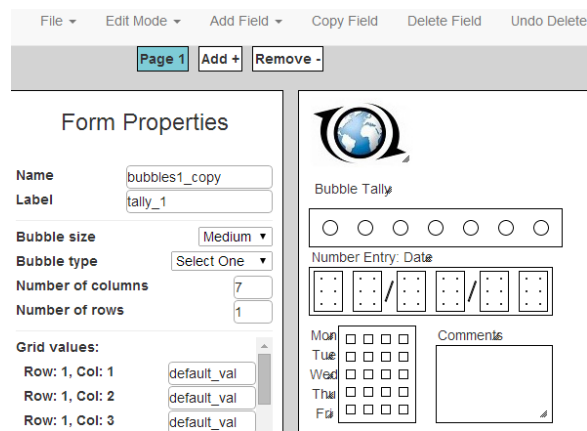


Figure 3.1: Our prototype design probe allows people to create a variety of machine-readable question types and incorporate content generated by other tools (such as images).

Procedure. Participants were introduced to our prototype design tool by watching a 20-minute sequence of tutorial videos that described the tool’s features. After watching the videos, we asked participants to use the tool to create a previously designed paper survey of their choice and observed as they completed this process on their own computers. This phase of the study took approximately 30 minutes and we encouraged participants to articulate their thought processes by requesting them to follow the think-aloud protocol.

We also conducted in-depth 60-minute interviews with the same nine participants. The interviews began with a short design exercise. Participants were provided with a research question and used a paper and pen to design a survey that would collect the data required to answer the question. The research questions were selected at random from the World Health Survey¹ conducted by the World Health Organization. This phase of the interview lasted approximately 15 minutes and allowed us to observe the participants’ design process and how they prioritized data digitization. The next phase of the interview lasted approximately 45 minutes and consisted of an in-depth discussion of the participants’ current work practices, their backgrounds, work responsibilities, current survey design tools and processes, the workflows surrounding the use of paper surveys in the field, and the data entry processes employed by their organization.

¹Available at <http://www.who.int/healthinfo/survey/en/>

Analysis

To analyze our data, we went through our survey results and interview transcripts and organized them using the “*design*”, “*field work*”, “*data entry*”, and “*analysis*” codes to determine which findings were relevant for the distinct phases of the workflows that we examine. For each of these codes, we conducted iterative analyses to ascertain and organize our prominent findings that we present below.

3.2.2 Findings

We study paper-digital workflows from the perspective of researchers and practitioners responsible for driving global development initiatives. The success of these initiatives depends upon effective collaboration across four phases of work between remotely located workers, as depicted in Figure 3.2. In the design phase, researchers and practitioners use digital tools to create surveys. In the data collection phase, workers in the field convert these surveys to paper and use them to collect data. In the data entry phase, workers transcribe data from paper into digital formats. Finally, in the analysis phase, researchers and practitioners process the digitized data to address research or program objectives.

At a high-level, we identify several layers of issues that affect paper-digital workflows in global development organizations, including (1) issues that relate to the material affordances and limitations of paper and digital media; (2) issues that relate to the low-resource nature of the work environments; and (3) issues that relate to the cross-cultural and geographically distributed nature of global development work. In presenting our findings, we limit our focus to how these issues affect the four phases of the paper-digital lifecycle, although we note that each category is worthy of additional study.

a. Design: Creating Paper Materials with Digital Tools

Our analysis reveals that survey design is a highly collaborative process involving multiple stakeholders, although effective collaboration is challenging for geographically dispersed teams that may speak different languages. Survey designers must also manage a number of trade-offs regarding the quantity of data collected from target communities, the layout and

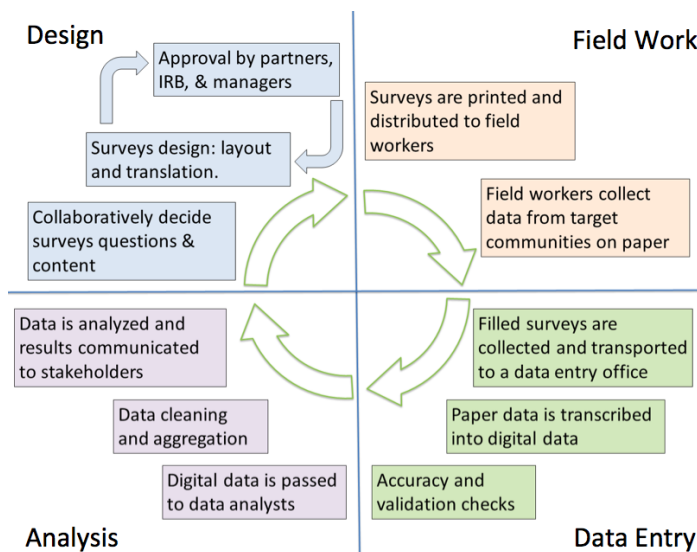


Figure 3.2: The four phases of the paper-digital lifecycle in the context of global development work: design, field work, data entry, analysis.

complexity of surveys, and the amount of work required to complete the survey. To make them easy to fill, surveys are typically highly structured documents that designers struggle to create using currently available software tools. However, the structured nature of the surveys means that, for the most part, they are already optimized for machine-readability, despite not currently being digitized by a machine.

a1: Collaborative survey design is challenging for geographically dispersed teams that may speak different languages.

The process that determines the content to be included on new surveys is collaborative and involves multiple stakeholders, often remotely located, *e.g.*, office staff in the US, field staff in target countries, Ministry officials, and funding agencies. In many cases, time differences between stakeholders in different countries introduce additional complexity to the collaboration, with one of our participants describing how it took her “*four months to go from the idea to the survey.*” Potential delays and constraints for meeting times must be factored in, as must holidays. Feedback from staff in target countries takes at least a day, while feedback from Ministry officials can often take a few weeks. In addition, obtaining the

necessary approvals for survey designs can take anywhere from a few days to many months.

Moreover, surveys designed by English-speaking researchers and practitioners must be translated into local languages if they are to be understood by all interested parties. However, communicating the goals of a survey across languages can be challenging and potentially lossy, particularly if the translator is not a native speaker of both. This could introduce subtle differences between the meanings intended by the designers, and the meanings on the translated survey. In addition, as the survey design evolves, these differences could become compounded by iterative design, editing and translating. Furthermore, the translation can affect the appearance and layout of content on the survey, which may complicate the work processes of both the designers and the translators:

“For most English-Portuguese translations, the stuff that takes three words in English takes seven in Portuguese. So we really have to think about how we fit it all in.”

a2: Survey designers manage several trade-offs regarding the content, layout and complexity of surveys.

Survey designers face a number of challenges deciding the content, layout and complexity of new surveys. Many participants described how they strive to design surveys that collect only the data necessary to answer specific research questions and nothing more, explaining how trying to collect too much data would both complicate their analyses and increase the likelihood of error. However, since collecting data from target communities in low-resource settings requires organizations to invest large amounts of time and effort, it often also makes sense to collect additional data that could be used to answer future research questions:

“It’s always a push-pull of collecting enough information to be useful in the future, even for questions we haven’t thought of yet. I mean there’s stuff in here that I’m not using at all, there’s a lot of redundant data. Looking back on this, it could be a little more streamlined. But there’s a lot of data that isn’t used yet.”

Many participants were also concerned about how field workers would perceive the amount of work required to fill in a survey and they devised strategies to make surveys appear easy to complete. Some participants prioritized density (see Figure 3.3, right),

The image shows two survey forms side-by-side. The left form, titled "Measles Vaccination Study (MVS01) Child Enrollment (CHEN)", has a clean, spaced-out layout. It includes fields for "Visit Date" (Day, Month, Year), "Study Visit", "Completed by", "Participant ID", "Study", "Site", and "Participant #". There is a section for question 7: "Has the child ever received a transfusion (blood, platelets or plasma)?" with "Yes" and "No" checkboxes and a date field. Below this is "Section III: Antiretroviral Therapy (ART) Adherence" with questions 8 and 9, each asking about missed doses in the past 3 days and past 1 month, with checkboxes and a field for "How many doses did the child miss?". The right form, titled "Newborn Health Supplemental Register for HSA", is extremely dense. It contains a table for "ANTENATAL CARE" with columns for "ANC Location", "Date of Visit", "Iron Status", "Maternal Health Conditions", "Hypertension/Pre-eclampsia", "Diabetes", "Underweight", "Carrying twins or triplets", "History of previous delivery", "History of stillbirth or neonatal death", "Registered for CDF", "Phone # to use", "Tb Screen", "Malaria Prophylaxis", "Tetanus", "Activity Level", "Nutrition", "Birth Plan", and "Breastfeeding". It also includes a table for "Home Visits and Education".

Figure 3.3: Two surveys that highlight how some of our participants optimize for clarity (left) and others optimize for density (right).

packing large amounts of data onto a single page to keep the survey as short as possible:

“We don’t want them carrying around too many sheets of paper, so we try to get as much information as possible on one page. And that’s also because of cost.”

Other participants prioritized for clarity and compromised on space (see Figure 3.3, left). Then, they broke each question into multiple parts so that it appeared as if the entire survey consisted of only a small number of questions:

“You can see that there’s question 3a, 3b, 3c. We have stupid numbers of sub-questions. And that was specifically asked for by the team, because psychologically if we’re on question 7 in section 1, that’s much nicer than question 23 of 49.”

In either case, the designer’s expectation and understanding of who will use the survey in the field heavily influenced the complexity of the survey. For example, one participant designed two surveys to collect data regarding how rural farmers in Tanzania spend their time. The first is a household survey designed to be filled by trained field workers employed by the organization. Since the designer expected that these workers would be relatively well educated, the survey consists of many pages and includes complex features like conditional branching. By contrast, the second survey is a diary that will be filled out by rural farmers. The designer does not know the characteristics of these farmers, whether they will be literate,


	7am	7:30am	8am	8:30am	9am	9:30am	10am	10:30am
 Sleeping	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
 Eating/Drinking	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
 Exercising/Sports	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
 Working/Volunteering	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
 Transportation/Commuting	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
 Cooking/Chores	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 3.4: Part of a survey created for use by rural farmers in Tanzania.

or their willingness to engage in the work. As a result, the survey is a single page that consists of images showing common household activities with a simple grid of bubbles for farmers to fill in (see Figure 3.4).

Since the consequences of misjudging the field workers’ abilities may substantially reduce the quality of the data collected, designers use a variety of strategies to increase the likelihood that field workers and target populations will understand the questions. One common strategy is to copy question formats from pre-existing, tried and tested surveys:

“We are trying to write as little original content as possible and use past surveys that have already piloted these questions in a similar context.”

Another popular content layout strategy involved mimicking the format of government registers whose contents were to be transferred onto the survey. Keeping the structure of the two surveys identical may make it easier for field workers to simply copy the content from the register onto the survey.

a3: Designers struggle to create surveys using currently available tools and are willing to try new tools that optimize surveys for machine-readability.

One goal in our study was to probe participants’ reactions to the idea of optimizing their surveys for machine-readability and digitization. Although machine-readable surveys

exist (such as Scantron²), none of the tools that our participants use currently support machine-readability. Instead, almost all of our survey participants reported using Microsoft Word as their primary survey design tool.

However, despite its prevalence, a large number of our interviewees were vocal about how challenging and frustrating they found the process of designing their highly structured surveys in Word. One participant who tried to switch from Word to Adobe InDesign shared that it had taken her several months to learn the new software but *“there is a startup cost to learning any new tool,”* and InDesign allowed her to *“create surveys that look very pretty”*. However, she also found the cost of purchasing the software to be significant (InDesign currently costs US \$240 per year) and, since designing surveys is typically collaborative, her team members would also need to purchase the software to edit the survey. Thus, despite her (time and financial) investment in the new software, and a personal preference for InDesign, she returned to using Word.

Although it is clearly challenging to incorporate new software into their workflows, all of our participants were eager to learn about viable alternative design tools. Moreover, they welcomed the idea of designing surveys that were optimized for machine-readability, with several assuming that automated digitization would be more accurate than human data entry. When participants used our prototype tool to create their surveys, all of the question types that they needed could be represented using features that the tool offered. In addition, almost all data fields used for analysis are already structured as numeric, bubble or checkbox fields, which could be automatically interpreted. Text, which cannot be automatically converted, must be saved in image format and/or manually transcribed by a person. However, several participant surveys contained no text-based data, while some contained a single text field for the person’s name, which was typically not used for analysis:

“Nobody ever looks back at the name. I don’t see how the name field would be relevant, except maybe if you needed to contact the person later.”

These findings suggest that, for the most part, our participants’ surveys are *already* optimized for machine-readability. They simply need to be created using software that enables a

²Available at www.scantron.com

machine to digitize the data. Participants also claimed that they would be willing to spend substantial amounts of extra time designing the survey if it meant that the survey would be machine-readable and ease the data entry process, particularly for large scale surveys:

“There would be 1200 of these. 10 hours upfront definitely outweighs typing in 1200 paper surveys. I think the time trade-off is in favor of [the tool].”

They also argued,

“Using [the tool] might not take longer than doing it in Word, because creating, resizing, moving and aligning so many boxes in Word can be very challenging.”

Finally, we noticed that participants’ thought processes were frequently shaped by the software that they use. For example, participants told us, *“I would do this as a table in Word”* or *“I use boxes because creating bubbles in Word is a nightmare.”* When asked to design a survey using our prototype, they again thought in terms of the tool’s features, *“Yesterday I would have said to use a handwritten number. Now, I would use bubble tallies.”* Participants’ experiences using our tool suggest that using software that encourages them to think in terms of digitization could change survey designers’ perspectives and increase their awareness of digitization.

b. Field work: Collecting Data from Target Communities

In the second phase of the paper-digital lifecycle, paper surveys are used by field workers to collect data from target communities. Our findings highlight both the benefits and challenges of paper-based data collection. Although paper does not require power or Internet connectivity, field workers who use paper still face a variety of social and infrastructural challenges. In addition, despite working hard to design understandable surveys that are appropriate for use by field workers, many participants admitted that they do not fully understand the complexities of conducting field work in low-resource settings.

b1: Field workers conducting surveys in low-resource settings face infrastructural and social challenges.

One of the primary benefits of using paper surveys to collect data in low-resource settings is that recording data on paper does not require electricity or Internet connectivity.

However, the printed paper surveys still need be transported to target communities, and limited transportation infrastructure means that field workers frequently travel on foot, which can make it difficult for them to carry a large number of paper surveys:

“Our clinic assistant takes these binders, puts them in a suitcase and rolls them to our office. And then if something is missing she has to take them back. And they go back and forth. It’s terrible. This poor girl...it’s an enormous suitcase full of binders.”

In addition, field workers may also face social challenges operating within target communities. Participants reported that field workers would sometimes feel unwelcome or uncomfortable asking questions of a personally sensitive nature, and they expressed sympathy for the field workers:

“They are out in the field, having people slam doors in their face and telling them to go away. If they are interacting with somebody who is getting fidgety or bored, I imagine they would want to skip questions or rush through questions that have lots of options.”

However, although the physical properties of paper do make it feasible for field workers to skip questions or sections of the survey, this behavior is not desirable from the perspective of the researchers and practitioners since it would result in incomplete or missing data. Moreover, several participants told us that field workers frequently repurpose parts of the physical paper as they see fit. In one case, a field worker crossed out a particular question, wrote in a different question, and then recorded an answer to the new, handwritten question (see Figure 3.5). Again, although modifications like these are relatively easy to make on paper, they have not been intended or approved by the survey designer, and almost certainly complicate the data entry, since the value recorded on the paper no longer corresponds to the value expected by the digital entry form. To resolve these kinds of issues, participants expressed that it would be preferable to digitize the data in the field so that they could ask field workers clarifying questions if necessary.

b2: Survey designers may not fully understand the complexities of field work in low-resource settings.

13. Make issues to HSAs	DSC	1 Multiple times a day 2 Once a day 3 Once per week 4 <input checked="" type="radio"/> Once per month 5 Twice a month 6 Never
<i>but oversee</i>		
14. Prepare LMIS reports to Districts		1 <input checked="" type="radio"/> Once a month 2 Every other month 3 Never

6

Review reports a same time

Figure 3.5: Part of a survey that shows how field workers take advantage of paper’s flexibility, which may make data entry challenging.

Although our participants spend substantial amounts of time thinking about the work performed in the field, several acknowledged that they had never met the field workers and did not fully understand their backgrounds or perspectives. This lack of understanding sometimes leads to a mismatch between the expectations of the researcher/practitioner and the capabilities of the field workers. For example, when we asked one participant if field workers had trouble filling out her complex survey, she told us, “*They’re used to filling out complex registers,*” although she later revealed that “*60% of [the field workers] couldn’t do it after two days of training.*” In addition, several participants were concerned about the quality of their field workers and the potential for fraudulent data [8] and many expressed a desire to monitor field workers and track their progress as the data is collected.

Finally, participants described how there are frequently events that occur in the field that cannot be anticipated or predicted. Since our study focuses on analyzing paper-digital workflows from the perspective of the survey designers, additional research is necessary to comprehensively analyze the complexities of field work in low-resource settings.

c. Data Entry: Transcribing Data from Paper Surveys

All of our participants hire data entry workers to manually transcribe paper-based data into digital formats and identified data entry as a major bottleneck in their workflows. Participants described how intermittent power and unreliable connectivity complicate data entry and communication. However, despite these challenges, data entry workers receive less attention and training than field workers. Instead, researchers and practitioners attempt

to control the data entry process by constraining the values that may be entered and by employing time intensive quality control techniques.

c1: Unreliable electricity and intermittent connectivity complicate the process of entering and communicating data.

Many participants explained how unreliable electricity and intermittent Internet connectivity can complicate the process of transcribing data into digital formats:

“There is limited electricity and Internet in the data entry offices. They mostly use desktops, so when the power is out, they can’t do data entry, and when the Internet is down, they can’t enter data using REDCap.”

Our participants predominantly use two software tools to store their entered data: Microsoft Access and REDCap³. The benefit of using REDCap is that data is entered directly into an online database. This simplifies the synchronization of digitized data between offices in different locations, and allows stakeholders in other countries to view or analyze the data as soon as it has been entered. However, entering data into REDCap requires continuous access to the Internet, which means that if there is no connectivity, data entry workers are unable to enter any data. By contrast, Access is often chosen because it is capable of running locally on the data entry worker’s computer, which means that (s)he can still enter data in the absence of an Internet connection. However, using Access complicates the process of synchronizing and communicating the data to remote researchers and practitioners:

“After entering the data, [the data entry worker] zips the database and uploads it to Google docs. It’s awful. I wrote out step by step how to zip it. A real issue for us is figuring out how to get data back in a secure way.”

This finding illustrates how even digital information can be difficult to store and communicate in low-resource settings. In addition, in some situations, organizations do not have sufficient resources to perform data entry in the target country. Instead, several participants described how they outsource the data entry to external companies, which may introduce additional challenges:

³REDCap is an open-source cloud-based tool designed for research.

“I know that for a paper survey that we worked on in Mexico, we ended up scanning all the surveys to somewhere in Thailand, and these Thai workers were inputting Spanish into their computers.”

c2: Survey designers may not fully appreciate the challenges of performing data entry in low-resource settings.

Our findings reveal a mismatch between the expectations of the survey designers and the constraints experienced by data entry workers. Many survey designers assume that data entry is a relatively simple task that should happen seamlessly, with one participant describing the data entry worker as *“the comfy data person, sitting in an office, sipping a beverage.”* However, in contrast to the assumption that data entry workers simply sit in front of a computer all day, we discovered that in fact data entry workers are often responsible for an array of additional tasks, such as preparing surveys for field workers, collecting and transporting completed surveys to the data entry office, communicating with field workers to resolve discrepancies, filing surveys for safekeeping, and preparing reports.

Moreover, all of our participants prioritized the needs of the field workers over the needs of the data entry workers, with most reporting that their field workers are usually well-trained and informed about the project’s objectives so that they are motivated to collect good quality data:

“We try to set up the environment with high standards. Having a passion for the data they are collecting will help ensure the quality of data is high.”

By contrast, when we asked if data entry workers understood the project’s objectives, several participants told us that *“there really isn’t any need for them to know about the goals of the project.”* In addition, several participants admitted that they were unaware of how long it takes workers to enter data:

“I don’t know how long it takes them to enter one of these forms. At one point I heard some number that was a little bit frightening.”

Given the lack of attention afforded data entry workers, it is perhaps not surprising that organizations in our study find it difficult to recruit and retain good data entry workers:

“We’ve had high turnover with data entry people. Data entry is a tedious job, it

is boring, and so it's tough to get someone who wants to do it and do it well."

c3: Survey designers try to ensure data accuracy by constraining the values that may be entered and by employing time intensive quality control techniques..

Maximizing the accuracy of entered data is a priority for our participants and they described a variety of strategies that they use to ensure entered data is complete and correct. For example, many participants spend a large amount of time designing the data entry interface to look as much like the paper survey as possible. This allows data entry workers to more easily find the correct digital entry box for each value that they see on the paper, although it complicates the process of creating the digital data entry forms:

"After this paper survey was already designed, it took me about four months to recreate the survey in Microsoft Access. It took forever. It's a huge process."

Moreover, many participants constrain the values that can be input into the database using pre-populated dropdown menus and complex validation rules:

"We have data validation, so when she's typing something in, if she says someone is 200 years old, it won't let her move on. But [setting up this validation] takes incredible amounts of work."

However, despite constraining the values that may be entered by data entry workers, many participants still experience significant issues with the accuracy of entered data. To reduce these issues, data entry workers also spend substantial amounts of time and energy doing quality control. Some organizations use double data entry, which requires that each survey be transcribed twice. Alternatively, some participants described using a technique called "*line-listing*," which involves holding the paper next to the screen and visually scanning the paper and digital data for discrepancies. One participant said that this process can add weeks of delay and that, in general, quality control is another major bottleneck.

Finally, in addition to transcribing data from paper surveys, data entry workers are often also responsible for finding and correcting data collection errors that have been made by field workers. In some cases, this can be relatively simple, such as correcting spelling mistakes, while in others, the data entry worker is required to find and communicate with the field worker before they can proceed with the data entry. Several participants also

described data entry workers as being “*at the mercy*” of field workers. Illegible handwriting results in additional challenges for data entry workers who have to decipher what has been written on the paper. Deciding the best way to resolve these issues is subjective and may result in additional stress for designers and/or data entry workers.

d. Analysis and Storage: Making Sense of the Data

The reason for investing so much time and effort to gather data from target communities is so that researchers and practitioners can perform analyses to understand their impact, answer research questions and generate reports for stakeholders. Our analysis of this phase of the workflow highlights that it is challenging for organizations to quickly communicate digitized data to geographically dispersed team members for analysis and visualization. In addition, organizations often expend valuable resources to securely store paper surveys even after the data has been digitized.

d1: Communicating digitized data to geographically dispersed stakeholders is challenging.

Our findings reveal several challenges that affect the synchronization and communication of digitized data between workers located in different countries. Many participants described that slow or unreliable Internet connectivity frequently delays the transmission of data from field offices in target countries to US-based offices where the analysis is performed. One participant shared that her organization uses software from Tableau to analyze data and create reports for stakeholders. However, since Tableau is relatively expensive, her organization purchased only a single license for the software, and, since the license is located in their Seattle office, the data must be communicated from the Mozambique office to the Seattle office before it can be analyzed. In addition, the paper surveys from which the data has been digitized are in Portuguese, and the Seattle-based worker who performs the analysis does not speak or understand Portuguese. However, she described,

“As it stands, you have the Portuguese version that they do data entry in, but the background code is in English. So then when we extract it, we have set up Tableau so that it feeds into the right worksheets in Tableau. So I see only

numbers, and because the code is in English, I see that, oh, they have entered this into the tetanus field.”

Finally, after completing the analysis in Tableau, the Seattle-based worker emails a pdf of the results to her colleagues in Mozambique once a month.

d2: It is difficult for organizations to securely store large quantities of paper-based surveys.

Many organizations keep all their paper materials for reference and/or safekeeping and, although the materiality of paper makes it easy to use in the field, it is unarguably more challenging to store than digital data:

“There are literally rooms where it’s just stacked ceiling to floor with old study surveys. And we have to get grants to get new space to store all the old surveys. It’s tough. But that’s what we do. This is a lot of very personal data. I have to spend hours blacking out people’s names and personal identifiers. These are kids with HIV.”

3.2.3 Discussion

Having presented a close look at paper-digital workflows across the four stages of design, field work, digitization, and analysis, we now synthesize our findings to offer the following takeaways. First, we show that *both* paper and digital materials play vital roles in global development workflows and are necessary if these organizations are to operate effectively in target communities. Second, we show that, in the workflows we examine, the challenges of transitioning data between paper and digital materials reveal opportunities to design new tools to ease the burden of digitization. Third, we aim to increase awareness of the disconnectedness in global development workflows that largely results from geographical, cultural and socioeconomic differences that can only be understood, not eliminated. Finally, we argue that our findings are relevant for development initiatives across domains.

Paper and Computer Supported Cooperative Work

We find that global development work relies on essential affordances provided by both paper and digital materials, and argue that the two *must* coexist if these organizations are to successfully navigate the hurdles posed by poor infrastructure, low connectivity, cultural differences, and other socioeconomic constraints. Paper is cheap, easy to use in almost any environment, and provides stakeholders with visible and material evidence of data collected. However, in contrast to digital data, which can be stored in vast quantities with relative ease, paper materials must be transported by people and stored in warehouses. It is difficult to imagine how any organization would go about storing and navigating giga- or terabytes of paper-based data. In reality, many of the complex analyses and visualizations that help organizations to make sense of the data they collect are only feasible if it is in communicable, searchable and mutable digital formats.

However, our findings also show that global development organizations must pay attention to the material properties of digital data, since *“bits cannot escape the material constraints of the physical devices that manipulate, store and exchange them”* [9]. Organizations must purchase and maintain digital devices for their workers, which may introduce additional challenges. Several of our participants had previously tried providing field workers with laptops or tablets to collect data in digital formats. However, in many socioeconomically disadvantaged communities, possessing an expensive digital device could have social implications for field workers’ relationships with local communities that may be unfamiliar or suspicious of new technologies. The material value of digital devices also makes them a target for theft, as described (in all seriousness) by one participant:

“In a previous survey that I worked on, it was not safe to send field workers out with tablets. It would be like, “please have them - steal them - I’m taking them around your village, I don’t care about them.” We couldn’t send [field workers] out with technology.”

Finally, infrastructural challenges also affect organizations’ abilities to communicate and synchronize digital data. Intermittent electricity, poor Internet connectivity, and costly data transmission often make the communication of digital data expensive, slow and unreliable.

In summary, the reliance on both paper and digital materials, and the challenges of effectively coordinating the paper-digital lifecycle, suggest a need for new tools that better support the complex and highly collaborative workflows that drive global development.

Opportunities for Design

One goal of this study was to probe participants' reactions to the idea of optimizing surveys for machine-readability and digitization. In particular, we wanted to investigate if survey designers at the top of the workflow may be willing to change their design practices, and potentially perform more work, to simplify the data entry process further down the workflow.

Researchers and practitioners currently view the process of transitioning information from paper surveys to digital formats as a major bottleneck in their workflow. They also expressed frustration with the tools that they currently use to design surveys and are willing to try new alternatives, even if they take time to learn or disrupt their existing workflows (several participants have already spent months searching for viable alternatives to Word). This suggests an opportunity to design new tools that ease the pain of converting digital to paper, and paper to digital. In the context of global development organizations, we found that survey design tools need to be affordable, provide creative control over the appearance of surveys, and be easily accessible by team members in multiple countries. Our prototype design probe (see Figure 3.1) appears to meet these criteria and provides a starting point on which we base the design of a new tool for creating surveys. In fact, shortly after we completed our study, we discovered that one of our participants had independently used the tool (which is freely available online) to design surveys for a project in Tanzania.

Optimizing surveys for machine-readability would undoubtedly impact the entire paper-digital workflow. In the design phase, since designers think in terms of the software they are using, a tool that encourages them to optimize for digitization could shape their thought processes and increase their awareness of the digitization process. In the field work phase, filling in machine-readable surveys could create additional work, since field workers may need to complete surveys more neatly to ensure they are accurately interpreted by software. This, in combination with the visual appearance of the survey, could also make field workers

more aware that collected data needs to be digitized. In the data entry phase, the majority of workers' time would no longer be spent typing data from paper into a computer. Instead, workers would be responsible for scanning the surveys, double-checking the interpretation of critical fields, and transcribing the few data items that are not machine-readable. Several participants also suggested that they would teach the data entry workers to perform simple, immediate analyses on the data so that they could monitor the field workers and identify problems. Thus, optimizing surveys for machine-readability could improve the digitization bottleneck in two ways: first, by automatically interpreting machine-readable data, and second, by making *all* workers more aware of the digitization process.

Cross-Cultural Cooperative Work

Successfully coordinating work processes across multiple geographies and cultures poses several challenges for global development organizations. In particular, the four distinct phases of the paper-digital lifecycle that we examined suggest that these workflows are somewhat disconnected. Remote locations, infrastructural challenges, cultural differences and changing time zones all exacerbate this problem. Field workers are trained to understand the importance of collecting data from target communities, but not the challenges associated with digitizing the collected data. Data entry workers are expected to seamlessly digitize large numbers of paper surveys without understanding the broader implications of their work. Survey designers often do not understand the perspectives of the field and data entry workers, and may have misplaced expectations or not fully appreciate the constraints experienced by their remote colleagues.

We also find that the nature of workers' roles within global, cross-cultural workflows are subject to complex social structures. Researchers and practitioners have a higher degree of control than field workers, who have a higher degree of control than data entry workers. This hierarchical structure impacts the resulting attitudes of and relationships between workers in a variety of ways. For example, our findings reveal disparities between the realities of data entry work and the desires and expectations of researchers and practitioners. Our participants frequently expected the data entry process to be seamless, and expressed

frustration and confusion as to why data entry workers found it difficult when “*all they have to do is type this stuff in.*” They frequently attributed delays in data entry to the workers’ backgrounds and attitudes, rather than the challenging nature of the work, with one commenting that “*it turns out data entry in Malawi is not super-accurate.*” Another participant commented that when she performed data entry, she would “*turn on techno and move through it.*” Though this tactic may work for her, it may not necessarily help data entry workers operating in remote locations, who may be struggling with computer safety issues, intermittent electricity, and poor Internet connectivity.

Although we have limited our analysis of cross-cultural issues to those that specifically affect the paper-digital workflow, our findings highlight rich opportunities for future research that focuses on more fully understanding the impact of culture in global development work. In addition, the differences in backgrounds, experiences and attitudes of culturally diverse workers suggest that it may be beneficial to increase workers’ awareness of their cultural and social differences. Although it may not be possible to entirely overcome these differences, our work contributes to a greater awareness of such challenges, in the belief that a greater awareness can strengthen work processes and practices.

Limitations and Generalizability

This study provides a close look at the challenges presented by the relevance of paper and the desirability of digitization in the context of global development workflows. Since our participants were all engaged in development work but drawn from a wide range of organizations and a variety of domains, we argue that our findings are domain-independent, and researchers and practitioners engaged in multi-country, cross-cultural research in any aspect of development could benefit from our work. In addition, by offering a comprehensive examination of the various abstract types of survey questions and how they can be optimized for digitization, we aim to be of help to survey designers as well. In general, we found that the workflow challenges that accompany global, cross-cultural research are exacerbated when low-resource environments are involved. For example, linguistic barriers can impact non-development work as well. However, when they are additionally accompanied by low

literacy and awareness, the situation is much worsened. Finally, we study paper-digital workflows from the perspective of researchers and practitioners responsible for driving global development initiatives. Additional research is necessary to explore these issues from the perspectives of other stakeholders.

3.2.4 Conclusion

This study examined the collaborative practices surrounding paper-digital workflows as enacted by organizations engaged in global development initiatives. We used a mixed methods approach to examine these workflows that span cultures and geographies, organizing our findings according to the different stages of data as it is sought, collected, digitized, and analyzed. We highlight the tensions that arise between the ubiquitousness of paper and the desirability of digitized data, also discussing the inherent affordances of paper and digital materials, and contribute a nuanced understanding of the accompanying challenges and tradeoffs. Taken together, our findings will be useful for survey designers, researchers and practitioners interested in global, cross-cultural research and practice. In addition, our findings could influence the design of new tools that aim to bridge the gap between paper and digital materials in the context of global development. Our work designing and building these new tools is described in detail in the following sections.

3.3 A System for Automatically Digitizing Data from Paper Documents

Based on the design opportunities identified in the previous section of this chapter, we designed and implemented ODK Scan, a mobile system that automates the capture and processing of digital data from paper forms. The phone's camera is used to capture an image of the form and computer vision algorithms automatically extract digital data from the image. Instead of creating a purely digital data collection system, we chose to design a hybrid solution that keeps the cost of deploying and maintaining the system low by continuing to use cheap, familiar paper forms at the lowest levels of the information hierarchy, while also facilitating the automated collection of digital data using computationally powerful smartphones at higher levels of the information hierarchy (such as at the district or provincial levels).

The computational power and intuitive touch-screen interfaces of modern smartphones makes them an attractive computing platform for low-resource environments. Many smartphones have built-in cameras, GPS sensors, and network interfaces that make them capable of supporting a wide variety of social, administrative and health applications. They also have ample storage capacity and are capable of both storing the data locally on the device and uploading it to a remote database when a cellular or data network becomes available.

This section presents the initial design and implementation of the ODK Scan system. It also describes an initial target application and preliminary deployment in which the system was used to digitize paper-based vaccine statistics in rural health centers in Mozambique.

3.3.1 Background and Target Application

We chose one concrete target application on which to focus our initial system design and evaluation: extracting vaccine statistics from paper forms in rural health centers in Mozambique. We chose this application in conjunction with VillageReach [144], an 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.

Figure 3.6: Paper form used to tally the number of vaccines administered monthly at a health center in Mozambique.

The Mozambique Ministry of Health distributes booklets of tally bubble forms, shown in Figure 3.23, 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 3.7 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.

3.3.2 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 will carry the device and digitize data from tally forms during his monthly visit to each health center, and the data collected will be automatically uploaded to the provincial database when s/he returns to the provincial office. This proposed workflow is shown in Figure 3.7.

By targeting provincial-level field coordinators who travel to a variety of health centers within a district, we are 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 are not required to learn a new system 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

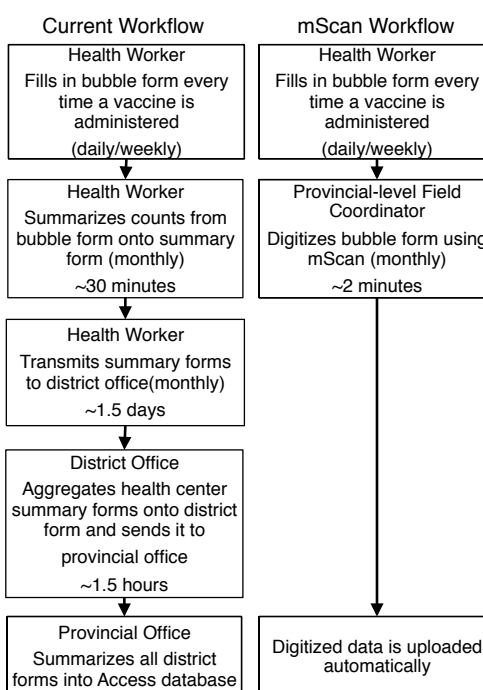


Figure 3.7: Current paper-based workflow (left) and proposed ODK Scan workflow (right).

field coordinators to use the new application than a large number of health workers.

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 system 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 are described later in this section.

Each paper form is filled in over the period of a month, which makes it possible for the form to become folded, dirty or marked accidentally (see Figure 3.23). It is therefore essential that ODK Scan 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



Figure 3.8: Portion of a partially filled bubble form.

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 3.8 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 device to classify bubbles as either filled or empty.

Finally, although the initial ODK Scan prototype targets the task of digitizing vaccine statistics in Mozambique, we have designed the system pipeline to extend easily to other paper forms and applications. We utilize a lightweight, generalizable JSON [31] 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 common data types, including bubbles, checkboxes, handwritten numbers, and text.

3.3.3 System Architecture

We built ODK Scan as an interactive Android system. 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. The image processing components of the system use OpenCV [104], an open source computer vision library, while the user interface components use Android’s Java framework. We use the Java Native Interface (JNI) [60] to facilitate communication between the Java framework and OpenCV’s native image processing algorithms. All of the image processing is performed on the device without requiring an Internet connection.

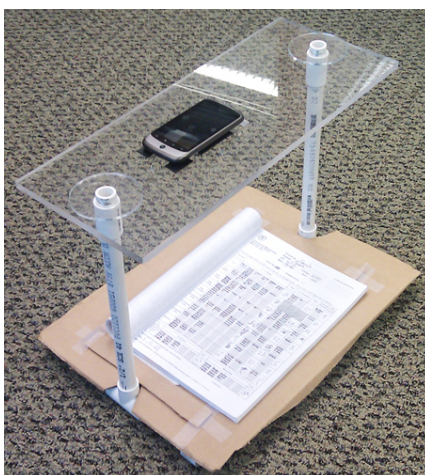


Figure 3.9: Low-cost plastic stand and form tray used to capture images of paper forms.

3.3.4 Implementation

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

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. ODK Scan 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.

Image Capture. The next step in the digitization process is to use the device’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 device in position. As shown in Figure 3.9, the device 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 device 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 device to accept the image, which is saved and passed to ODK Scan for further processing.

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 device. Alignment is necessary to ensure that the entire form has been captured and to determine the locations of the form fields and elements. ODK Scan uses feature detection and matching to perform registration. Features are detected using OpenCV's SURF (Speeded Up Robust Feature) [7] implementation. To decrease the amount of time required for feature matching, ODK Scan 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 [102]. Finally, ODK Scan uses the RANSAC algorithm [50] to compute a transformation that maps the captured image to the reference image, thereby establishing a point-by-point correspondence between the two images.

Segment Detection and Alignment. After image registration, ODK Scan 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.

Classifier Training. Individual bubble classification is performed using a Support Vector Machine (SVM) [28] 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) [109] 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.

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 [109] 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.

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 a bubble from the surrounding image region is left as an area of future work.

Provincia: Nampala Locais de Concentração

Districto: Mualá

US: 7777 código: 12345

Mês: Agosto Ano: 2011

Posto fixo Brigada móvel

Ficha de registo diário do PAV para posto fixo ou brigada móvel - BCG, Pólio, DTP-HepB-Hib e Sarampo

Vacinas	0 - 11 meses	Total	12 - 23 meses	Total	Frascos Abertos	Total Frascos Abertos
1						
BCG	65	65	27	27	9	9
Pólio Primário	90	90			10	10
Pólio 1ª Dose	53	53	50	50	11	11
Pólio 2ª Dose	46	46	27	27	12	12
Pólio 3ª Dose	60	60	22	22	5	5
OP1negH1b 1ª Dose	84	84	25	25	15	15
OP1negH1b 2ª Dose	41	41	25	25	5	5
OP1negH1b 3ª Dose	37	37	10	10	3	3
Sarampo	50	50	30	30	11	11
Femininos		total	Masculinos		total	
53		53	7		7	
65		65	53		53	

Nome registado do responsável pelo preenchimento: _____ Categoria: _____

Data: 9/16/2011

Mkt: Cópia completamente vacinada à saúde que recebeu todas as vacinas antes de completar um ano de vida.

Figure 3.10: A marked up image of a processed form.

Data Output and Integration. After classification, 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 output file that contains the name of each form field (specified in the form description file) and the bubble tally for that field. In addition, the user can manually type in 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 3.10, 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.

3.3.5 *Experimental Evaluation*

To evaluate the technical performance of ODK Scan, 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 were performed using an HTC Nexus One Android device. 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 (TP) or true negatives (TN) while errors are either false positives (FP) or false negatives (FN). 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 form contained 924 filled and 1256 unfilled bubbles, with the same bubbles filled on each form.

Lighting Conditions

Lighting is a critical factor in any image processing system. To evaluate the performance of ODK Scan 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 3.11 depicts sample images used for the experiment. We did not include conditions that

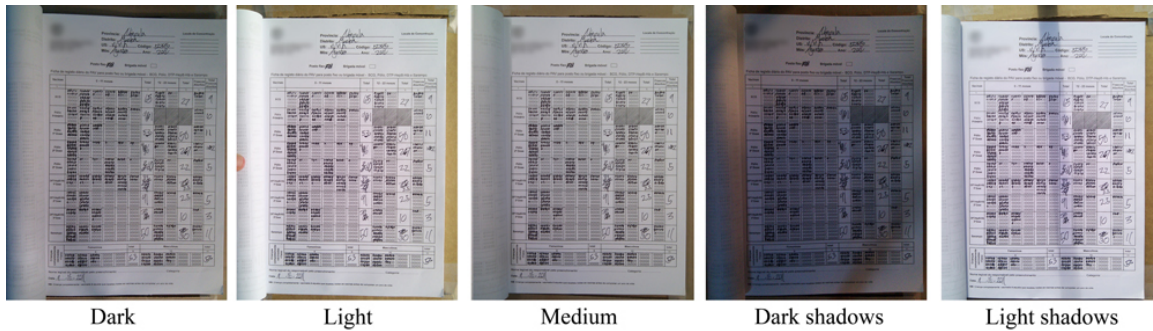


Figure 3.11: Sample images showing each lighting condition tested.

	Condition	Segment Alignment			Bubble Classification					Overall %
		Correct	Error	%	FN	FP	TN	TP	%	
Neat form	Dark	449	1	99.8	5	4	6276	4595	99.9	99.7
	Light	449	1	99.8	12	9	6271	4588	99.8	99.6
	Medium	450	0	100	5	2	6278	4615	99.9	99.9
	Dark shadow	449	1	99.8	5	208	6058	4604	98.0	97.8
	Light shadow	450	0	100	5	5	6275	4615	99.9	99.9
Messy form	Dark	450	0	100	23	10	6270	4597	99.7	99.7
	Light	450	0	100	22	13	6267	4598	99.7	99.7
	Medium	450	0	100	25	10	6270	4595	99.7	99.7
	Dark shadow	449	1	99.8	24	269	6011	4571	97.3	97.9
	Light shadow	448	2	99.6	35	13	6242	4560	99.6	99.1

Table 3.1: ODK Scan performance under different lighting conditions

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.

Results. The experiment results are given in Table 3.1. For each condition, form segments were correctly aligned with overall accuracy above 99%. Additionally, for each lighting condition except dark shadows, ODK Scan 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 97.3% 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 3.11, 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.

Folded or Dirty Forms

Since forms are filled in over a period of one month, they could become folded or dirty and we wanted to ensure that ODK Scan 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 3.12 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.

Results. The experiment results are given in Table 3.2. 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 ODK Scan 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 94.2% with the neat form and 95.1% with the messy form, while bubble classification accuracy dropped to 96.9% with the neat form and 96.1% with the messy form. This resulted in an overall accuracy of 91.3% with the neat form and 91.4% with the messy form. Performance also dropped with a very crumpled form. Under this condition, segment alignment accuracy

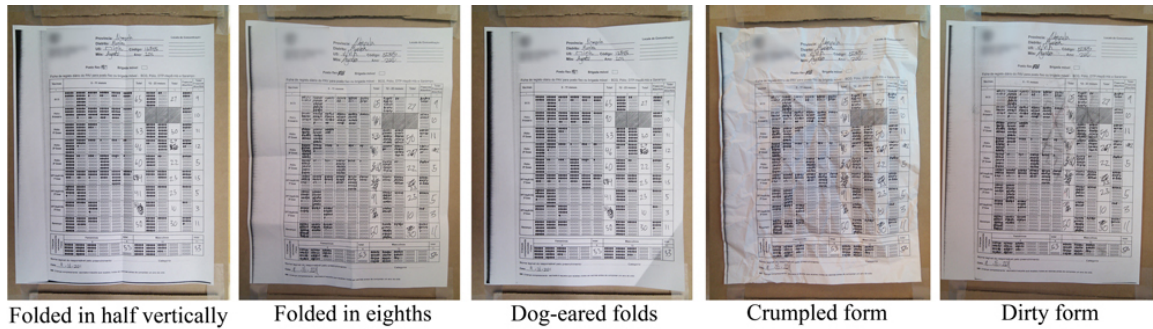


Figure 3.12: Sample images showing the folded and dirty conditions tested.

		Segment Alignment			Bubble Classification					Overall
Condition		Correct	Errors	%	FN	FP	TN	TP	%	%
Neat form	Half	449	1	99.8	9	19	6241	4611	99.7	99.5
	Eighths	424	26	94.2	103	220	5704	4278	96.9	91.3
	Dog-eared	450	0	100	5	22	6258	4615	99.8	99.8
	Crumpled	438	12	97.3	217	310	5827	4251	95.0	92.5
	Dirty	450	0	100	5	12	6268	4615	99.8	99.8
Messy form	Half	449	1	99.8	26	22	6258	4574	99.6	99.3
	Eighths	428	22	95.1	104	301	5718	4287	96.1	91.4
	Dog-eared	449	1	99.8	30	38	6235	4577	99.4	99.2
	Crumpled	406	44	90.2	496	455	5270	3664	90.4	81.5
	Dirty	450	0	100	28	45	6235	4592	99.3	99.3

Table 3.2: ODK Scan performance with folded and dirty forms

was 97.33% with the neat form and 90.2% with the messy form, while bubble classification accuracy was 95.0% with the neat form and 90.38% with the messy form. This resulted in overall accuracy of only 92.5% with the neat form and 81.5% with the messy form, which indicates that severe crumpling and folding of the form should be avoided.

Background Conditions

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 images and we wanted to test whether ODK Scan is able to handle such cases accurately. We defined five different background conditions: brown, black, white, shiny and patterned. Sample images

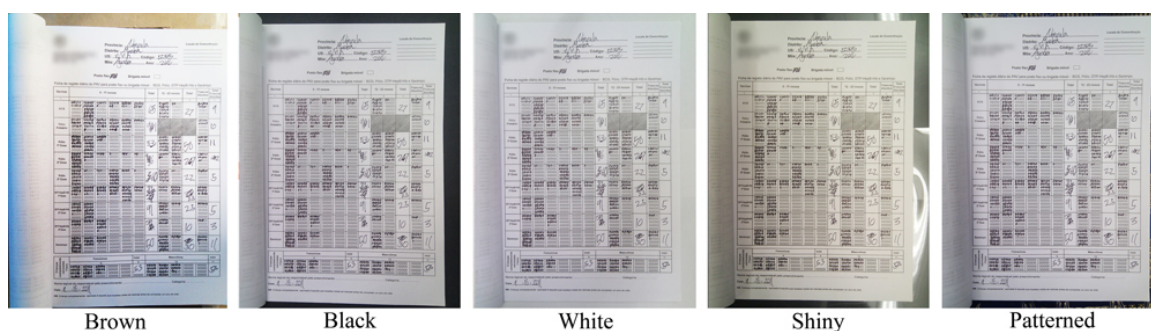


Figure 3.13: Sample images showing background conditions tested.

	Condition	Segment Alignment			Bubble Classification					Overall
		Correct	Errors	%	FN	FP	TN	TP	%	%
Neat form	Brown	450	0	100	5	2	6278	4615	99.9	99.9
	Black	450	0	100	5	4	6276	4615	99.9	99.9
	White	450	0	100	5	0	6280	4615	99.9	99.9
	Shiny	450	0	100	5	8	6272	4615	99.9	99.9
	Patterned	450	0	100	5	3	6277	4615	99.9	99.9
Messy form	Brown	450	0	100	25	10	6270	4595	99.7	99.7
	Black	450	0	100	36	8	6272	4584	99.6	99.6
	White	445	5	98.9	63	85	6139	4508	98.6	97.5
	Shiny	450	0	100	37	10	6270	4583	99.6	99.6
	Patterned	450	0	100	36	11	6269	4584	99.6	99.6

Table 3.3: ODK Scan performance with different background conditions

showing these backgrounds are given in Figure 3.13. 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.

Results. The experiment results are given in Table 3.3. 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 98.9%, while bubble classification accuracy was 98.6%, resulting in an overall accuracy of 97.5%. These results show ODK Scan 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.

Processing Rate

The system processed 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.

3.3.6 Preliminary Field Evaluation

We conducted a preliminary field test of ODK Scan with Ministry of Health field coordinators in the Cabo Delgado province of Mozambique in November 2011. Our objective was to assess the performance of the system in relation to the expertise and technological experience of target users and the environmental conditions experienced in health centers.

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 ODK Scan and explained the importance of environmental factors like good lighting and form position. After demonstrating the application, we observed the field coordinators using ODK Scan 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 they were comfortable interacting with the system.

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 system under the conditions experienced in the health centers. At the first two health centers we visited, we demonstrated the system for the health workers and then observed the field coordinators and health workers using the system to capture data from paper forms. In addition, we explained to the health workers the importance of filling in the



Figure 3.14: A field coordinator explaining how ODK Scan works to a health worker.

bubbles on the form neatly and keeping the form clean. At the next three health centers, the field coordinators demonstrated the system for the health workers and taught them the importance of filling in the form neatly. Figure 3.14 shows a field coordinator teaching a health worker about the system.

Although this preliminary user feedback was encouraging, we also discovered several issues that needed to be addressed. 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 ODK Scan outputs and comparing the results with the filled bubbles on the original form. Since ODK Scan 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. All of these issues have been addressed in subsequent implementations of the system.

3.3.7 Conclusion

This section describes the initial design and implementation of ODK Scan, a mobile system that uses computer vision to automate the capture of digital data from paper forms. Our experimental evaluation and preliminary user feedback show that ODK Scan is an accurate and robust tool that is ready to be extended to a wide variety of different paper forms and use cases. The following sections will describe techniques for digitizing additional data types and a longitudinal field deployment that we conducted to ensure that the application is usable and appropriate given the constraints experienced by low-resource communities.

3.4 *Techniques to Improve the Transcription of Non-Machine-Readable Data*

The previous section described a camera-based mobile system that is capable of automatically interpreting machine-readable data types such as fill-in bubbles. However, many paper forms used by global development organizations contain a variety of different data types, including non-machine-readable handwritten text. Before this information can be aggregated, analyzed, searched or shared, it needs to be transcribed into a digital format. Currently, the predominant method for entering data from paper forms into structured digital content is manual data entry. This is a slow and laborious process in which data entry workers read information written on the paper forms and manually type it into a computer. Moreover, if the data to be entered is in a structured or semi-structured format, the data entry worker is required to look at the screen to navigate to the appropriate digital data entry box, or use the mouse to select a value from a checklist or drop-down menu. Additionally, if the data entry worker is at a less-than-expert level she must look at the screen to monitor the transcription results and correct errors. Any of these scenarios results in the data entry worker repeatedly switching her focus of attention back and forth between the screen and the paper, which reduces the speed of data entry.

To address this challenge, we developed *Snippets*, a new method that facilitates data entry from images of paper forms. Snippets uses computer vision techniques to segment the image into small snippets (see Figures 3.15a and 3.16a) that each contain the content for a single field. The snippets are displayed on the screen so that users can simply look at the snippets and type in the data. This approach negates the need for workers to have access to the physical form. Snippets makes data entry more efficient by (1) using ODK Scan to automatically process machine-readable data types so that they do not have to be manually entered, and (2) displaying non-machine readable data on the screen for efficient entry, so that users do not have to switch their focus of attention between the paper and the screen or keyboard. In addition, all of the segmentation and processing is performed locally, eliminating the need for an Internet connection.

We evaluated desktop and mobile versions of Snippets in a controlled study with 26 participants in Seattle. Our findings show that Snippets increased the speed of data entry over

Student Name	Gender
Alexis Rogers	female Male <input type="checkbox"/> Female <input checked="" type="checkbox"/>
Andrew Kelly	male Male <input checked="" type="checkbox"/> Female <input type="checkbox"/>

(a) Desktop Snippets

Student Name	Gender
<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>

(b) Desktop baseline

Figure 3.15: The two methods tested on the desktop platform: (a) Snippets and (b) Baseline.

a baseline method of looking at the form and manually entering data (see Figures 3.15b and 3.16b) by an average of 28.3% on the desktop platform, and 10.8% on the mobile platform without any detectable loss of accuracy. Participants' subjective comments support our statistical findings and show that participants perceived a performance benefit when entering data using Snippets. Finally, data from a preliminary field study with five participants in Bangalore, India shows similar trends to the data from Seattle, providing further evidence for the success of Snippets.

This section makes two primary contributions: (1) the development of Snippets, a novel method for entering data from paper forms that eliminates the need for workers to switch their focus of attention back and forth between the form and the screen, and (2) an empirical evaluation showing that Snippets significantly increases the speed of data entry over a baseline method in current use on both desktop and mobile platforms.

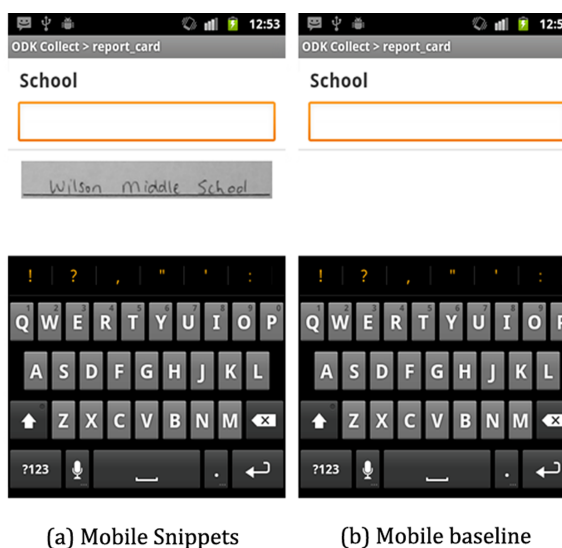


Figure 3.16: The two methods tested on the mobile platform: (a) Snippets and (b) Baseline.

3.4.1 The Snippets System

Snippets makes data entry more efficient by segmenting an image of the form into individual fields that are displayed on the screen for easy entry. Prior to data entry, the image of the paper form is processed by our ODK Scan software (described in the previous section) and data types that are machine-readable are automatically interpreted. We now discuss each of these steps in detail.

Processing an Image of the Paper Form

The previous section of this chapter described the design and implementation of ODK Scan and showed that the software was capable of classifying fill-in bubbles with over 99% accuracy [34]. However, we did not consider the design of user interfaces or methods to aid the entry of non-machine readable data and we did not provide a way for users to check and (if necessary) correct the processed results. We also only implemented the algorithms on a mobile device, and did not consider data entry on desktop computers.

Therefore, this section focuses on the design of desktop and mobile user interfaces and methods that help *people* to enter data from paper forms. The results of the algorithms

described in the previous section serve as one input to these new data entry methods. The other main input, which we created solely for this study, consists of image “snippets”. Snippets are small fragments of the original form image that correspond to individual form fields, and it is these snippets that we display to users for transcription. In addition, since the ODK Scan algorithms occasionally make classification errors, Snippets affords users the opportunity to check and possibly correct the results of the automated scanning.

Data Entry from Image Snippets

Traditional data entry requires users to look at the form and find a field to enter, then look at the screen and find the corresponding entry box, and then type in the value. One major disadvantage of this approach is that the user has to switch her *focus of attention* (FOA) back and forth between the paper and the screen or keyboard. The burden of switching FOA and the benefits of reducing the FOA required for a task is a well-studied phenomenon [85]. For example, if the worker is an expert touch typist and is merely copying unstructured text, she may not need to look at the screen or keyboard at all, and can instead remain entirely focused on the paper form. However, if the data to be entered is in a structured or semi-structured format, even touch typists will need to look at the screen to ensure that they are typing data into the correct entry box or selecting the correct value from a list or drop-down menu. This increases the FOA required for the task and will reduce the rate at which data can be entered.

Snippets improves data entry by eliminating the need for users to refer to the physical form. Instead, an image of the form is captured and segmented into image snippets, with each snippet containing the portion of the image that corresponds to a single field. Each snippet is displayed on the screen with a data entry box so that users can simply look at the snippet and type in the value. An additional benefit of our approach is that it makes it clear how users should progress with data entry by guiding them from snippet to snippet. This reduces the likelihood that users will accidentally skip fields. Snippets also incorporates the output of the ODK Scan software, which consists of processed data values for fields that are machine-readable. Snippets uses these values to pre-populate the entry boxes that

correspond to these fields. This allows users to quickly check the results of the automated processing and, if necessary, correct any errors.

For this study, we implemented two versions of Snippets: a desktop version, shown in Figure 3.15a, that runs within a custom-built web application, and a mobile version, shown in Figure 3.16a, that runs on Android devices. In the desktop version, all of the snippets for the form are displayed simultaneously on the screen in a table. Multiple forms can be displayed at the same time, with the snippets for each form represented as a row in the table. Users can enter data horizontally or vertically, and columns of fields that have already been entered can be hidden by clicking on the column heading. In addition, users are able to save partially entered forms so that the transcription can be completed at a later time.

We implemented the mobile version of Snippets using the Android-based data collection tool ODK Collect [65] that we modified and extended to include the display of the image snippets. The small screen size of many mobile devices makes it impractical to display all of the image snippets for a whole form on one screen. Instead, as shown in Figure 3.16a, we display each image snippet individually on the screen along with the corresponding data entry box. Users type in the data value using the standard android touchscreen keyboard, and then swipe to progress to the next snippet. As with the desktop version, users are able to save partially entered forms and return to them at a later time.

3.4.2 Laboratory Evaluation

We evaluated Snippets through controlled user studies. We wanted to see if the new technique would improve the rate and accuracy at which people enter data compared to the traditional method of reading a paper form and typing into the device. We conducted two studies: one in which participants entered data into a desktop computer, and another in which participants entered data using a touchscreen smartphone. We chose not to compare the user experience and performance of the desktop and mobile platforms since these platforms target different usage scenarios and most organizations only use mobile phones for data entry if using a desktop platform is inappropriate or impossible.

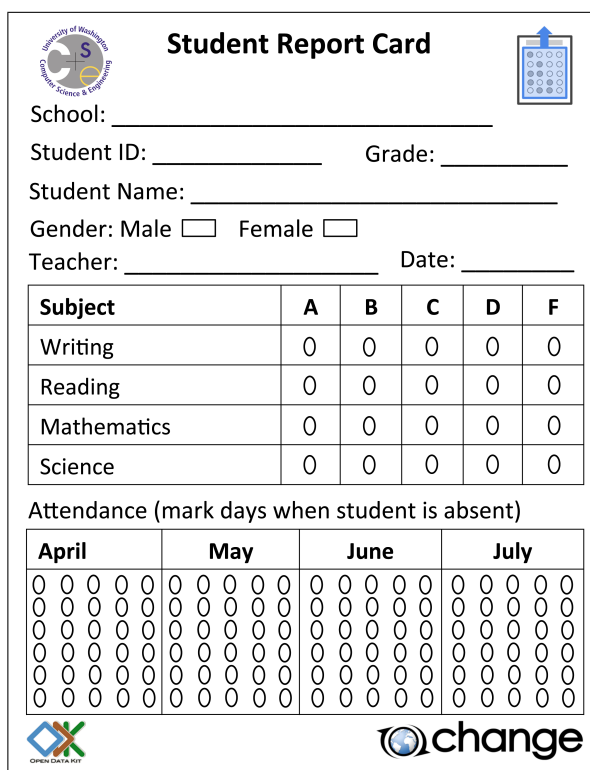
Participants

Our user studies took place in Seattle, WA. We recruited 26 participants (13 female) ranging in age from 18 to 49 years ($M = 27.0$, $SD = 5.9$). All participants had over 10 years experience with computers and self-rated as intermediate to expert computer users. None of the participants were professional data entry workers or typists. Ten participants were near-expert touchscreen users with approximately three years of touchscreen use. Five participants did not own a touchscreen device and had little experience with them.

Apparatus

For the desktop study, participants entered data using a custom web application running in the Chrome browser on a Dell desktop computer with a 2.5 GHz processor and 2 GB of memory. A 17" monitor, standard optical mouse, and keyboard were connected to the computer, and the application recorded all of the user's screen interactions and timing information. For the mobile study, participants entered data using an HTC Nexus One Android smartphone with 512 MB memory and a 1 GHz processor. The phone had a 3.7 inch capacitive touchscreen. Participants entered data using the ODK Collect software [65] that we modified considerably to include the display of image snippets. In addition, we recorded all the touch and keyboard events and timing data.

We created a paper report card to use for the forms in the study (see Figure 3.17). The form contained 15 fields: 3 text, 3 number, one checkbox, 4 multiple-choice, and 4 tally fields. We created enough test forms so that no participant would enter data from the same form more than once and filled the forms with realistic, fictitious values. Since we wanted to use the same forms for the laboratory study in Seattle, WA and the field study in Bangalore, India (described later in this section), half the forms contained common American names and the other half common Indian names. All the forms were filled in English. An image of each form was captured using a 5 mega-pixel smartphone camera and processed using our ODK Scan software.



Student Report Card

School: _____

Student ID: _____ Grade: _____

Student Name: _____

Gender: Male Female

Teacher: _____ Date: _____

Subject	A	B	C	D	F
Writing	0	0	0	0	0
Reading	0	0	0	0	0
Mathematics	0	0	0	0	0
Science	0	0	0	0	0

Attendance (mark days when student is absent)

April	May	June	July
0 0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0 0
0 0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0 0
0 0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0 0
0 0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0 0
0 0 0 0 0	0 0 0 0 0	0 0 0 0 0	0 0 0 0 0

OPEN DATA KIT

change

Figure 3.17: The report card paper form that we used for the study.

Procedure

The experimental procedure was designed to fit into a single 60-minute session. Each session began with an introduction to the data entry platforms and techniques, and a description of the tasks that participants would perform. Participants were asked to practice entering data using both entry methods. They were guided through the practice session by a short tutorial that explained how to navigate the software, transcribe data and save the results. This practice phase lasted about 15 minutes, until the user was comfortable with the different techniques and had completed the entry of data from an entire paper form using both methods without assistance. Since many of the form fields were similar in nature and the data entry was somewhat repetitive, we felt that transcribing data from one entire form for each technique was sufficient practice.

For both the desktop and mobile platforms, participants entered data using each of two entry methods: Snippets and a baseline method in which participants looked at the paper

form and manually typed in the data. On the desktop platform, participants sat at a desk with a computer, mouse and keyboard. For the baseline method, the forms were placed on the desk to the left of the keyboard. On the mobile platform, participants sat at a table and entered data holding the device without setting it down on the table. For the baseline method, the forms were placed on the desk in front of the participant. Participants completed four trials for each method. A trial was defined as entering all of the data for a single form. For the baseline, participants typed in values for all 15 form fields per trial. For Snippets, participants typed in values for the 6 text and number fields. For the other 9 fields, participants checked the automatically processed values and corrected any that had been incorrectly classified. We asked participants to enter data quickly and accurately, and to fix errors in the scanned data.

The order of presentation for levels of *Entry Method* was counterbalanced on both platforms to avoid carryover effects. A test of entry method *Order* on task time was significant on the desktop ($F_{1,170.1} = 7.23, p < .01$). However, there was no significant *Order x Entry Method* interaction ($F_{1,24} = 0.21, n.s.$), indicating that although the method completed first was slower than the method completed second, this order effect was symmetrical for both methods, which were presented first an equal number of times. We did not observe a significant effect of entry method *Order* on task time for the mobile platform ($F_{1,6.1} = 0.08, n.s.$), and no *Order x Entry Method* interaction ($F_{1,24} = 1.87, n.s.$), indicating that counterbalancing was effective. After completing each method, participants filled out a NASA task load index (TLX) questionnaire [64] to rate their subjective experience with that method. Participants also completed a questionnaire at the end of the session that collected some demographic data, previous computer and touchscreen experience, and subjective comments related to the experimental techniques.

Design and Analysis

We conducted two single-factor studies to evaluate Snippets: one on a desktop computer and the other on a mobile touchscreen device. Each study was a within-subjects single-factor study with two levels. The single factor was *Entry Method* and the levels were Baseline and

Snippets. Participants completed 4 trials with each entry method, where a trial was the entry of all the data on a single form. This resulted in 8 trials each on the desktop and mobile platforms, for a total of 208 trials in each study. Overall, we collected 36,509 key presses from 26 participants.

Our time measurements do not include the time required to capture form images. We made this decision because different image capture methods may result in vastly different capture times (e.g., document feeder vs. scanner vs. camera) independent of data entry method. For example, an organization might use an automated document feeder to quickly scan a large number of forms and then distribute the data entry to clerks with mobile devices. In addition, capturing an image takes a fixed amount of time for each capture method, but the time for data entry depends on the amount of data in the form. The form we used contains a relatively small amount of data, but many forms contain much larger amounts of data.

As is typical, time measures violated the normality assumption for ANOVA as they were log-normally distributed. Therefore, form completion times were log-transformed to restore normality before analysis. For readability, however, graphs and averages are shown as raw times, not logarithms of times. Statistical analyses of log-time were carried out using a mixed-effects model analysis of variance with *Entry Method* as a fixed effect and *Form* and *Participant* as random effects [56, 81].⁴ Statistical analyses of error counts and NASA TLX scores were conducted with nonparametric Wilcoxon signed-rank tests [149].

3.4.3 Results

Speed

As shown in Figure 3.18, the average time to enter a form on the desktop was 60.8 seconds ($SD = 19.6$) for the baseline and 42.8 seconds ($SD = 14.1$) for Snippets. This difference was statistically significant ($F_{1,172} = 811.73, p < .0001$), indicating that Snippets was significantly

⁴The levels of random effects—in our case, the specific forms and human participants—are not of interest and were drawn randomly from larger populations over which results are meant to generalize. Mixed-effects models preserve larger denominator degrees of freedom than traditional fixed-effects ANOVAs but compensate by using wider confidence intervals, making significance no easier to detect (and often harder). They can also result in fractional denominator degrees-of-freedom for unbalanced designs.

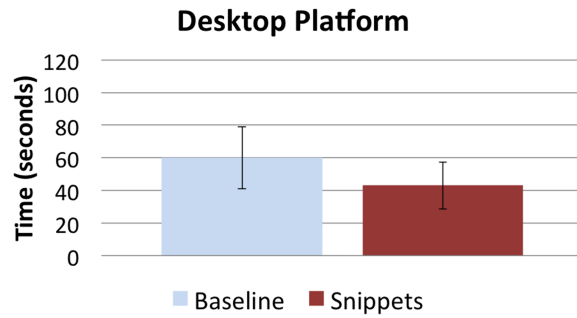


Figure 3.18: Average time taken to complete data entry for one paper form in the desktop study. Lower is better. Error bars show ± 1 SD.

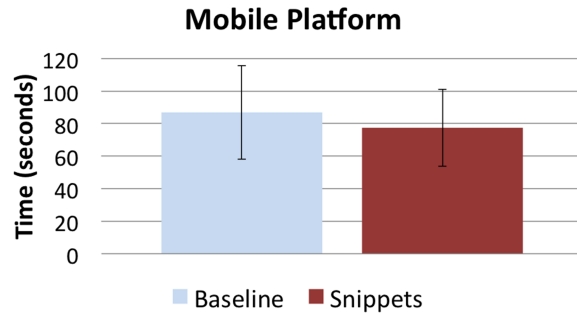


Figure 3.19: Average time taken to complete data entry for one paper form in the mobile study. Lower is better. Error bars show ± 1 SD.

faster than the baseline on the desktop. On the mobile platform, the average time to enter a form using the baseline method was 86.8 seconds ($SD = 28.7$); using Snippets it was 77.4 seconds ($SD = 23.7$) (see Figure 3.19). This difference was statistically significant ($F_{1,5.9} = 10.31$, $p < .02$), indicating that Snippets was significantly faster on the mobile platform.

Accuracy

A measure of accuracy in our tasks can be obtained by using the number of incorrect fields per form (out of a total of 15 fields per form). On the desktop, the average number of incorrect fields per form was 0.26 ($SD = 0.49$) for the baseline method and 0.35 ($SD = 0.41$) for Snippets. The number of incorrect fields was tabulated for each participant for

each entry method and a nonparametric Wilcoxon signed-rank test was run. The test was nonsignificant for *Entry Method on Errors* ($Z = -1.50, n.s.$). On the mobile platform, the average number of incorrect fields per form was 0.31 ($SD = 0.59$) for the baseline and 0.33 ($SD = 0.57$) for Snippets. The number of incorrect fields was again summed for each participant for each entry method and a nonparametric Wilcoxon signed-rank test was run. The test was also nonsignificant for *Entry Method on Errors* ($Z = 0.29, n.s.$). Taken together, these results indicate that there was no detectable difference in accuracy between the two methods on either the desktop or mobile platform.

Subjective Responses

The NASA task load index (TLX) survey [64] was completed by participants after each data entry method, allowing for 1-20 ratings on six scales: mental demand, physical demand, temporal demand, perceived performance, perceived effort, and perceived frustration. Except for perceived performance, which ranges from “perfect” (1) to “failure” (20), all scales range from “very low” (1) to “very high” (20). Lower is therefore better on all scales.

Results for the desktop platform are shown in Figure 3.20. Nonparametric Wilcoxon signed-rank tests were conducted on the desktop ratings with significant outcomes in favor of Snippets on *all* scales: mental demand ($Z = -2.74, p = .006$), physical demand ($Z = -3.09, p = .002$), temporal demand ($Z = -2.45, p = .014$), perceived performance ($Z = -2.97, p = .003$), perceived effort ($Z = -3.67, p < .0001$), and perceived frustration ($Z = -3.22, p = .001$). These findings show that participants perceived a substantial performance benefit of Snippets over the baseline.

NASA TLX results for the mobile platform are shown in Figure 3.21. Nonparametric Wilcoxon signed-rank tests were conducted on these ratings with a significant outcome only for mental demand, with Snippets being significantly less mentally demanding than the baseline method ($Z = 2.49, p < .02$). However, there were trend-level results in favor of Snippets over the baseline method for temporal demand ($Z = 1.69, p = .092$), perceived performance ($Z = 1.85, p = .065$), and perceived frustration ($Z = 1.86, p = .063$), providing further evidence for the success of Snippets.

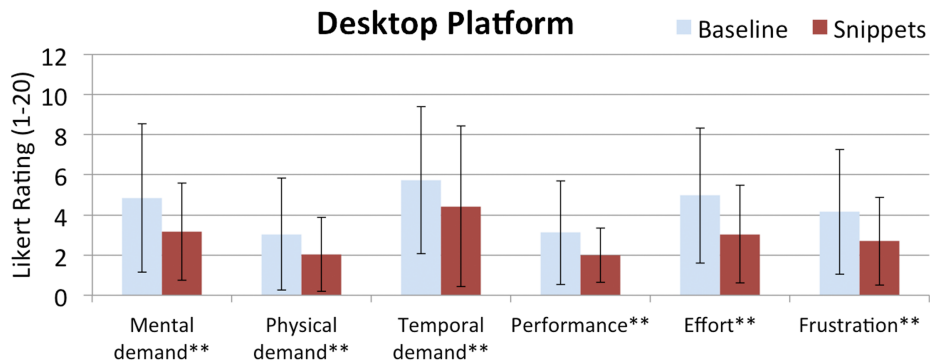


Figure 3.20: Results of the NASA TLX survey for the desktop study. Lower is better. Error bars show ± 1 SD. ** indicates $p < .05$; * indicates $p < .1$.

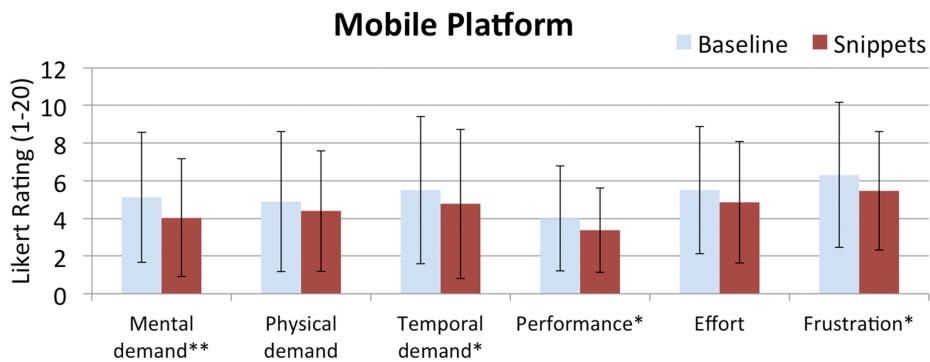


Figure 3.21: Results of the NASA TLX survey for the mobile study. Lower is better. Error bars show ± 1 SD. ** indicates $p < .05$; * indicates $p < .1$.

Interpretation of Results

Snippets improves the speed of data entry over the baseline method by an average of 28.3% on the desktop platform and 10.8% on the mobile platform without any detectable loss of accuracy. Since entering data with Snippets still requires participants to read and type in many of the data values, a large proportion of the time savings is likely due to improving the users' focus of attention. Snippets eliminates the need for users to look at the form to find the next data item and then find the entry box on the screen before typing in the value. Thus, it is likely that Snippets moves us closer to the lower bound on data entry time.

However, since our Snippets method incorporates the output of the ODK Scan software, it is impossible to tell what proportion of the speed-up is due to displaying image snippets on the screen and what proportion is due to the pre-populated values computed by ODK Scan. Since we asked participants to check the pre-processed values from ODK Scan and correct any mistakes, we are confident that participants at least read and interpreted all of the form fields even if they did not manually enter values that had been processed correctly.

It is also interesting that we did *not* see any substantial differences in the accuracy of data entry between the baseline method and Snippets on either the desktop or the mobile platforms. Data entry tasks typically result in a speed-accuracy tradeoff, in which an improvement in the speed of data entry comes at the cost of decreased accuracy. It is therefore highly encouraging that the improvement in speed that we saw with Snippets did not detectably affect the accuracy of data entry. Instead, both methods exhibited similar errors rates. Observation of the data entry process revealed that although we made an effort to ensure that the handwriting on the test forms was clearly readable, a large proportion of errors resulted from participants being unable to correctly interpret the handwriting. For example, participants often purposely typed the letter ‘v’ when the handwritten letter was actually ‘u’. Since participants were reading the same handwritten values on both the paper form and the image snippet, these kinds of errors were equally likely to occur with both methods. Unfortunately for data entry workers, accurately transcribing human handwriting remains an inherently difficult task.

The NASA TLX questionnaires for the desktop platform show that participants rated Snippets as being significantly better than the baseline method on *all* six scales tested: mental demand, physical demand, temporal demand, perceived performance, perceived effort, and perceived frustration. These results clearly show that participants perceived the performance benefit of Snippets and preferred it over the baseline method. Participants’ subjective comments overwhelmingly support this preference. Participant 7 told us, “*My favorite technique was [Snippets]. I felt the fastest with this technique and it was also easy to check that my responses were correct, since the images were right next to the entry fields.*” In addition, several participants liked being able to transcribe the data by column, rather than by row. Participant 4 said, “*The best entry technique was [Snippets]; I was able to enter the*

same column for all four forms, then move to the next column for all four forms, minimizing the amount of mental context-switching to process each particular column.” When combined with our statistical results, these findings suggest that Snippets is a simple, practical technique that improves the process of entering data from paper forms.

On the mobile platform, Snippets improved the speed of data entry by an average of 10.8% over the baseline. This is substantially less of an improvement than we saw with the desktop, which could be explained by a number of reasons. First, since participants held the device in their hands, they were able to position the device relatively close to the paper form for the baseline method, which decreased the amount of head and eye movement required to switch focus of attention between the device and the paper. In addition, for both Snippets and the baseline, it was generally more difficult for participants to type using the small, soft mobile keyboard than the desktop keyboard. Participant 4 commented, *“The worst entry techniques were the two on the phone; I’m an experienced mobile touchscreen user, but it’s still a pain to enter any non-trivial amount of data using just my two thumbs.”* The general difficulty that participants had with data entry on the mobile device suggests that organizations who choose to use mobile devices for data collection should try to minimize the amount of text that must be typed. This could be achieved by auto-filling fields with likely responses, providing pre-programmed options that can be chosen with a single touch, and maximizing the use of drop-down menus.

However, although participants found data entry on the mobile platform to be generally slow and laborious, the results of the NASA TLX surveys show a trend in favor of Snippets, and participants’ subjective comments support this trend. Participant 21 told us, *“I [preferred] the phone with Snippets over the [baseline] due to having the form data right on the screen. This was especially useful for data entry that was largely just confirming that the system had already made the correct selection.”* Again, when combined with our statistical results, these findings suggest that Snippets could aid data entry on mobile devices.

3.4.4 Field Evaluation in Bangalore, India

Our primary research objective is to develop data entry techniques that will be useful in low-resource environments. As such, we felt that it was important to collect preliminary data from a low-resource setting. Thus, in addition to the laboratory evaluation in Seattle, we also performed a small field evaluation at the data entry offices of the Akshara Foundation in Bangalore, India [135]. This field evaluation was intended to be an additional data gathering exercise rather than a formal, quantitative evaluation. The Akshara Foundation is an NGO working to universalize equitable access to quality preschool and elementary education for all children in India. The Foundation relies on large scale paper-based data collection to measure its impact and control the quality of the services it provides. At the beginning, middle, and end of each school term, teachers and evaluators administer academic assessments to over 200,000 students, and record the performance of each student on a paper form. The hundreds of thousands of paper forms are collected and transported back to the Akshara offices in Bangalore where they are manually entered twice by different people to avoid errors. To perform the data entry, Akshara employs a number of full time workers who spend their entire day entering data from paper into computers.

Participants

We recruited five data entry workers (all female) ranging in age from 22 to 40 years ($M = 28.0$, $SD = 2.6$). Four participants had about 3 years of computer experience, and the fifth had about a year of computer experience. None of the participants owned a touchscreen device, and all self-rated as beginner touchscreen users. Three participants said this was their first time using a touchscreen.

Apparatus and Procedure

The field evaluation took place in the data entry offices of the Akshara Foundation in Bangalore, India. Participants experienced the same procedure as those in Seattle, completing the practice session and the data entry of four paper forms (trials) for each technique using a desktop and an HTC Nexus One touchscreen device. The forms used were the same as

those in the Seattle study, and participants in India entered the same amount of data per method as the participants in Seattle. In addition, participants again completed a NASA TLX questionnaire after each entry method.

Results

Due to the relatively small number of participants in the field study, we do not have sufficient data for statistical significance. Instead, we report descriptive statistics and qualitative data in an effort to characterize our users' experiences. For speed, Snippets took the lowest average time on the desktop, averaging 80.9 seconds ($SD = 20.0$) to transcribe a form compared to 84.9 seconds ($SD = 18.8$) with the baseline method. Snippets also took the lowest average time on the mobile platform, averaging 110.8 seconds ($SD = 23.2$) compared to 155.3 seconds ($SD = 46.4$) with the baseline method. For accuracy, Snippets resulted in the fewest errors on the desktop, averaging 0.80 ($SD = 0.89$) incorrect fields per form, while the baseline method averaged 0.95 ($SD = 0.82$) incorrect fields per form. On the mobile platform, Snippets also resulted in the fewest errors, averaging 1.05 ($SD = 0.82$) incorrect fields per form, while the baseline averaged 1.20 ($SD = 0.89$) incorrect fields per form.

As with the laboratory study, the NASA TLX questionnaire was given to participants after each method. On average, the NASA TLX ratings for Snippets were all lower than the baseline method, indicating that participants seemed to perceive a performance benefit with Snippets. Participant comments overwhelmingly support this perception. Participant 31 told us, *“Entering data from the [paper form] is more difficult, the scanned image [snippets] and especially the fields where I can select the answer, like the [multiple choice] marks or [bubble] tallies, make it easier to enter data.”*

However, participants generally struggled to use the small, soft keyboard on the mobile device. Participant 29 said, *“The typing is hard because the keyboard is very tiny,”* while Participant 31 told us, *“It is quite stressful to do data entry on the phone because I have to continuously bend my head to look at the screen and [navigate] back and forth when I make a mistake.”* These findings mirror our observations in Seattle, and suggest that using mobile devices for large-scale data entry presents challenges that make them inferior to desktops.

3.4.5 Discussion

We observed a number of interesting differences between the participants in India and the participants in Seattle. First, data entry workers at the Akshara Foundation currently enter data from paper forms directly into Excel spreadsheets and are thus familiar with moving to the next field on the screen using the arrow keys on the keyboard. By contrast, the participants in Seattle were familiar with using the tab key to navigate their way around the screen. Since our interfaces supported the use of the tab key but not the arrow keys, the Seattle participants may have been at an advantage. In addition, the Indian participants were generally more cautious with the data entry than the Seattle participants and took time to check the values that they entered. This is probably due to the fact that they have been trained to recognize the importance of accurate data entry. Furthermore, rather than just entering whatever values had been recorded on the form, the Indian participants thought about whether these values made sense. For example, participant 28 asked us, *“This name could be a girl’s name, but the form says that it is a boy. Should I correct the gender to female?”* She recognized that whoever filled out the form may also have made mistakes and that she could correct some of these mistakes at data entry.

For many of the Indian participants, data entry from paper forms has been their primary interaction with computers, and they were visibly more comfortable when they had the physical paper form to refer to. Participant 30 commented, *“I am familiar with the [paper form] so it was easy, but the scanned [snippets] were as easy even though it was my first time using it. If I used the scanned [snippets] for some time then I think it would be faster than the [paper form].”* Furthermore, when entering data for the baseline method, the majority of Indian participants kept one finger physically on the paper form so as to be able to easily remember which form field to enter next. Additional observation of the current data entry process at Akshara revealed that this appears to be standard practice for these workers. By keeping one finger on the paper, participants were only able to type using one hand (and in several cases, only one finger), and were unfamiliar with how to type using two hands. This unfamiliarity may have affected the speed at which data was entered for Snippets, which could explain why we saw less of an improvement with Snippets on the desktop in India

than in Seattle. We would expect that as these workers become familiar with typing using two hands, their rate of data entry will likely increase.

3.4.6 Conclusion

Manually transcribing the data collected on paper forms into a digital format is a slow and laborious process. To address this problem, we developed Snippets, a novel data entry method that improves data entry by (1) automatically processing machine-readable data types so that they do not have to be manually entered, and (2) displaying non-machine readable data on the screen to eliminate the need for users to refer to the physical paper form. Findings from an empirical study with 26 participants in Seattle, WA show that Snippets increases the speed of data entry over the baseline method of reading the paper form and manually entering the data by an average of 28.3% on the desktop platform and 10.8% on the mobile platform without any detectable loss of accuracy. Findings from a preliminary field study with five participants in Bangalore, India support these empirical results. We conclude that Snippets is a simple and efficient technique that could be widely used to improve data entry from paper forms. The next section will describe a longitudinal field evaluation of the system with workers in rural health centers.

3.5 Field Evaluation of a Mobile System for Digitizing Paper Documents

The previous sections in this chapter describe the design, implementation, and laboratory evaluation of our system for automatically digitizing data from paper documents. This section describes a field evaluation of the system with health workers in Mozambique. More specifically, we describe a deployment in which the system was used to strengthen the community health worker (CHW) supply chain. CHWs are members of the community who typically receive a small amount of training and who then assess, treat and refer patients in their community according to established health protocols. In addition to seeing patients, community health workers are typically also responsible for collecting monitoring and evaluation data about the health programs in which they are engaged.

For a treatment-based community health program to be successful, it is essential that CHWs receive a reliable and uninterrupted supply of health commodities, including diagnostic tests, basic medicines and contraceptives, so that they possess the medical supplies required to treat patients effectively [25]. Securing this uninterrupted supply of commodities requires a strong supply chain to ensure that the right quantities of the right products are available at the right time, place and condition, and for the right cost. However, limited communications and transport infrastructure make it difficult to ensure that medical supplies are ordered, received and distributed on time. In addition, as in Mozambique, there is often no standardized system for collecting and reporting of logistics data, particularly regarding the consumption and delivery of CHW commodities. The lack of such a reporting system leads to CHWs running out of supplies and being unable to treat patients, and supervisors and partners not knowing the consumption patterns or stock status of CHWs.

To address these challenges, we partnered with the non-governmental organization (NGO) Village Reach to create an intervention that allows CHWs to collect data regarding the commodities that they dispense using cheap, familiar and easy-to-use paper forms. CHWs bring these paper forms to their supervisors during their monthly supervisory meeting. The supervisors then use our mobile application, ODK Scan [34], to digitize the CHW consumption data and transmit it to a central server where it is available for immediate viewing and analysis by other stakeholders.

This section describes our experiences integrating ODK Scan into the CHW supply chain in two districts Mozambique. We document the impact of the technology at multiple levels of the information hierarchy: the NGO practitioners who designed the ODK Scan compatible forms, and set up and maintained the system; the NGO field staff who trained and supervised the users of the application; the CHWs who recorded their consumption of medical supplies on the paper forms; and the health district supervisors who digitized and reported the data using ODK Scan.

This section makes three contributions:

- (1) We detail our experiences integrating ODK Scan into the community health worker supply chain in Mozambique;
- (2) We evaluate the impact of the technology at multiple levels of the information hierarchy, providing quantitative and qualitative data that exposes the benefits, challenges and limitations of the technology; and
- (3) We share lessons learned and provide actionable guidance to researchers and practitioners interested in deploying ODK Scan or other systems that bridge the gap between paper-based and digital data collection.

3.5.1 Background

To better understand the current state of the CHW supply chain in Mozambique we conducted interviews and site visits to clarify and validate descriptions of the current program, understand data collection procedures and identify strengths and weaknesses of the program.

In 1978, Mozambique's Ministry of Health began the CHW program to extend coverage of the national public health system to include underserved rural populations. During Mozambique's civil war this program eventually became defunct and is currently undergoing revitalization. In the new program, each CHW is responsible for a population of between 500-2000 people and performs routine tasks such as developing strong ties with the community, health promotion and education, family planning counseling and prevention and treatment of common ailments [148]. CHWs receive a monthly stipend in exchange for the services that they provide, although the stipend is not a full-time salary and the CHWs

often have to balance their health worker duties with other work. To support the curative and preventative services they provide, CHWs receive a monthly kit of health supplies, including medicines, rapid diagnostic tests for malaria, and male condoms. Each CHW is given medicines and supplies designed to treat 250 patients per month, though depending on their catchment area and the season they may treat more or fewer. Supplies reach CHWs through Mozambique's network of public pharmacies. District pharmacies obtain supply kits from the provincial pharmacy and distribute them to health centers on a monthly basis. CHWs then visit their assigned health center once a month to receive their supply kit.

Each district operating the CHW program has one Ministry of Health staff member, known as the district supervisor, dedicated to supervising approximately 25 CHWs in the district. The tasks assigned to the district supervisor include training new CHWs, conducting regular supervision of the CHW program, assessing the availability of medicines and supplies, and monitoring the quality of case management and proper completion of reports.

3.5.2 Current Challenges

Our exploration of the current supply chain revealed several key challenges faced by the CHW program. One of the biggest challenges is that there is currently no standardized resupply process or system for reporting of logistics data [141]. For example, the quantity of commodities included in the CHW supply kit is based on forecasts made at a central level according to predicted disease burden and estimated monthly consumption of each commodity. However, when asked about the monthly consumption patterns of the commodities in their supply kits, different CHWs reported varying rates of consumption. This situation highlights that differences in consumption patterns between CHWs often result in CHWs either stocking out or being overstocked. However, there is currently no reporting mechanism that allows supervisors to monitor stock status of CHWs or national-level partners to validate kit quantities.

In addition, although supplies are ordered regularly, they are not necessarily delivered regularly resulting in CHWs experiencing stockouts. However, there is no process for documenting the frequency or severity of these stockouts. This results in CHWs not having

supplies and district supervisors not knowing the stock status of the CHWs. To address this issue, several district supervisors and pharmacy staff expressed a desire for higher quality commodity consumption data from CHWs. In addition, CHWs expressed that it is challenging for them to track their own health supplies each month. Many CHWs in Mozambique have only attended primary school and so find it difficult to understand logistics concepts and track their usage of supplies over time.

These challenges suggest that it would be beneficial to design a standardized data collection and reporting system to allow CHWs to collect and share logistics data with their supervisors, and make it easy for district supervisors to aggregate and analyze this data and report it to higher levels of the information hierarchy. Moreover, it is important to do this quickly to maximize the time available to respond to stockouts or over-stocks.

3.5.3 Intervention Design

In response to the challenges identified above, we designed an intervention to introduce a standardized system for collecting and reporting CHW commodity consumption data. Our design focuses on developing a workflow that will be cheap, familiar and easy-to-use for CHWs and that will also provide district supervisors with an easy way to digitize, analyze and disseminate CHW logistics data.

Figure 3.22 shows the intervention workflow. To track CHW consumption of health commodities, practitioners at the non-profit organization VillageReach [144] designed a paper form that will be distributed and filled out by CHWs. The choice to use a paper form at the CHW level was made for several reasons. First, paper forms are familiar to the CHWs and well-suited to their education level and prior experience. In addition, paper forms are cheap to produce and each form can store and display a relatively large amount of data, such as the number of each of 16 commodities consumed. Furthermore, the paper form will serve as a visual and tangible record that will allow CHWs to see their supply usage from day to day. Finally, in this scenario, the paper forms will be easy to distribute because the CHWs already visit their supervisor on a monthly basis to collect supplies and they will collect the paper form during this visit.

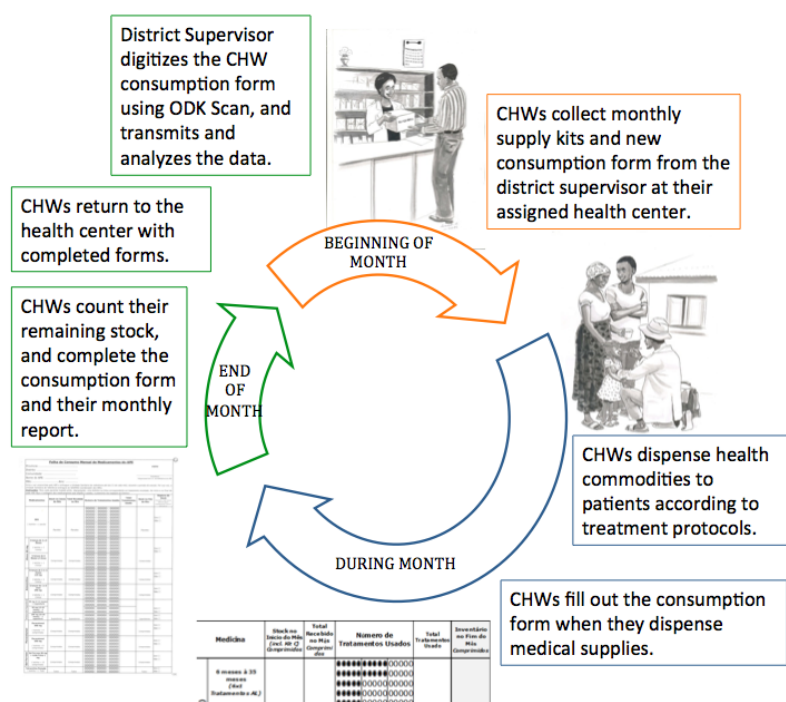


Figure 3.22: The monthly workflow for recording and digitizing CHW commodity consumption data [141].

To complete the form, CHWs record their initial stock levels at the beginning of the month and any extra stock received during the month. They then fill in one bubble on the form for each treatment provided to patients. CHWs will also use the form to record their remaining stock at the end of the month and whether they faced stockouts that month. They will then present the completed form to their supervisor when they go to pick up the next month's supply kit. The supervisor will use ODK Scan [34] to digitize, aggregate, and analyze the CHW consumption data and transmit the data to a cloud-based server running ODK Aggregate [65]. The data can be used by the Ministry of Health or partner organizations to keep track of CHW activities and recognize when CHWs experience stock problems so that they can be provided with supplementary stock if necessary.

3.5.4 Implementation

We now describe the work required to incorporate ODK Scan as part of a new system for collecting and reporting of CHW commodity consumption data in Mozambique. We focus on documenting the experiences of three key stakeholders: the NGO practitioners responsible for implementing the technology, the CHWs who use the paper forms in the field, and the district supervisors who use ODK Scan to digitize the paper forms.

Practitioners

Form Design and Description. The first step in the implementation was to design an ODK Scan compatible paper form to be used by CHWs to record the commodities that they consume. The form, shown in Figure 3.23, was designed by NGO staff with input from district and provincial health staff and other partner organizations. The form incorporated several features to facilitate digitization by ODK Scan. For example, the form contained three pictures located in the corners of the form to assist ODK Scan with feature extraction for form alignment. In addition, the form fields that record the number of supplies consumed consist of machine-readable bubble tallies that can be automatically processed by the application.

In addition to designing the paper form, NGO practitioners also created the ODK Scan form description file to tell the application the name, size, and data type of each form field. Creating this form description file was the most technically challenging aspect of the entire deployment. To make it easier to specify form fields, we developed a Javascript web application called the “template maker” that allows users to drag and drop a form image into the browser window and mark up different form fields by drawing boxes on the image. The software then automatically creates a JSON object that describes the size and location of the boxes. To specify a bubble, the user clicks on a point within an already defined form field. If the location of the bubble is not exactly correct, the user may refine the bubble coordinates by editing the generated JSON in a side panel on the user interface. All newly created fields are assigned a default “bubble” classifier. By editing the JSON that specifies the classifier, users may customize the size or type of the bubbles on the form.

Figure 3.23: The two-page ODK Scan compatible CHW commodity consumption form.

Creating the form description file required significant communication between the ODK Scan developers and the NGO practitioners designing the form. Over several rounds of iterative design, the developers worked to make the template maker easier to use and to add features requested by the practitioners, and the practitioners used the tool to create a description file for the CHW consumption form. The challenges faced and lessons learned from this process and the design decisions that resulted from the inherent technical nature of the form descriptions are discussed in Section 3.5.7.

Training Materials and Documentation. In low-resource settings, many practitioners or end-users have limited experience with technology. As a result, it is crucial to develop comprehensive documentation and training materials that do not assume prior knowledge of the application or device. To aid other practitioners who want to setup, configure, and deploy ODK Scan we developed a “setup and technical guide”. This guide provides instructions on basic functionality such as how to power up and navigate the device and

access the SD card. It then covers how to install and configure ODK Scan, Collect, and Aggregate. Finally, the guide covers how to handle a variety of common mistakes and error messages. We also created a user manual for district supervisors that demonstrates how to place the phone's camera above the form and take good quality pictures. Step-by-step instructions and screenshots guide the user through capturing and processing form images using ODK Scan, entering non-machine readable data using image snippets in ODK Collect, and transmitting data to ODK Aggregate. The manual also addresses common errors or problems. Finally, we also designed a CHW help booklet with pictures and instructions on how to correctly fill out the paper consumption forms.

CHWs

In addition to providing CHWs with help booklets, we also held a one day training workshop (from 8am-7pm) to teach CHWs how to correctly use the consumption forms. The day began with classroom sessions that explained how to fill out the forms, after which CHWs completed a series of practical exercises in small groups that included a large amount of practice filling out forms. Since many CHWs in Mozambique only have primary school education, the groups were specifically designed to allow better educated CHWs to help less educated CHWs understand the concepts. To evaluate the efficacy of the workshop, CHWs completed a short quiz before and after the training sessions that assessed their knowledge of how to use the forms. CHWs who had trouble with the quiz received additional help until they demonstrated sufficient confidence with the form. In addition, since tracking commodities and recording stock-outs were new concepts for many CHWs, we also conducted a half-day follow-up training session one month after the initial training that provided CHWs with additional practice and assessment on their already filled out forms.

District Supervisors

In addition to the CHW training workshops, each district supervisor received a full-day one-on-one training session on how to use the ODK tools to digitize forms and transmit data. To begin, an NGO staff member demonstrated how to digitize several consumption

forms. Then the district supervisor digitized approximately 15 forms under the supervision of the field staff member. In addition to using the software, the supervisor also received advice on how to maintain the phone and keep it secure, and on how to assess the quality of form completion and give advice to CHWs if the forms were not filled out satisfactorily.

3.5.5 Quantitative Evaluation

To evaluate the impact of integrating ODK Scan into the CHW supply chain, we conducted a four month study in two districts in Maputo province: the Manhiça and Marracuene districts. A total of 45 CHWs (22 from Manhiça and 23 from Marracuene) were trained to fill out the consumption forms and two district supervisors were trained to digitize the forms using ODK Scan. Training was completed in December 2012 using the methods and training manuals described in the previous section. After the training sessions, the application was deployed for a period of four months between January and April 2013.

Data Collection

To measure the amount of CHW consumption data collected, we counted the number of forms that CHWs submitted to district supervisors and the number of forms that the supervisors digitized using ODK Scan, Collect, and Aggregate. We also evaluated the quality of the consumption forms with respect to folding, warping and cleanliness.

Results. Over the four month study period, the 45 CHWs submitted a total of 140 paper forms (out of a possible 180) to the district supervisors. There were a variety of reasons why CHWs sometimes did not submit forms. Several lived very far away and were unable to submit forms because of transport difficulties. Others did not fill out forms because they did not receive their monthly kit of supplies. Some were simply not working or had left the area. Of the 140 forms received, district supervisors scanned, processed and uploaded a total of 122 forms. The supervisors said that the remaining 18 forms were not scanned because they contained too many errors to yield useful data. A large proportion of the unscanned forms were from February and subsequent investigation revealed that many CHWs did not fill out the form at the beginning of the month because they had not received their monthly stipend

from the Ministry of Health. In addition, many of the 122 forms that were digitized and transmitted were in poor condition: dirty, warped or folded. These artifacts contributed to a higher than expected ODK Scan error rate that we will now discuss in detail.

Accuracy

Each form contains 6 text fields, 68 numeric fields, 21 yes/no checkbox fields and 1395 bubbles in 16 tally fields. To measure the accuracy of the data collected, we selected a random sample of 35 consumption forms and manually created groundtruth data for these forms. Then we evaluated the accuracy of ODK Scan by comparing ODK Scan's bubble and checkbox classifications to the groundtruth data. We evaluated the combined accuracy of ODK Scan and Collect by comparing the data submitted to ODK Aggregate with the groundtruth data.

To tease apart the proportion of scanning errors that occurred as a result of the ODK Scan algorithms and the proportion that occurred due to errors in the form description or the poor condition of the forms, we asked an expert ODK Scan developer to create a new form description file for the consumption form and compared the results obtained using the expert's form description with the results obtained using the original NGO practitioner's form description. The expert's form description also enabled several additional form alignment features. These features were not enabled in the deployment because initial testing of the form design revealed that they had a slightly negative effect on the overall classification accuracy. The forms we used to test the form design were in relatively good condition compared to the real consumption forms submitted, and the extra alignment features appeared to slightly overcorrect the form alignment when corrections were not necessary. However, since many of the real consumption forms were in very poor condition, we wanted to see how the extra alignment features might affect the accuracy of processing.

Results. Table 3.4 shows the results of comparing the automated ODK Scan classifications to the groundtruth data. Overall, we saw a classification accuracy of 91.3% with the practitioner's form description and 98.9% with the expert's form description and extra alignment features. A large number of the errors that occurred with the practitioner's form description

	Practitioner's Form Description	Expert's Form Description
True Negatives	39387	42699
False Negatives	867	354
True Positives	6537	7050
False Positives	3504	192
Overall Accuracy	91.3%	98.9%

Table 3.4: Bubble classification accuracy on a sample of 35 consumption forms using the practitioner's form description and the expert's form description.

resulted from false positive classifications and a closer examination of the form description revealed that many specified bubble locations were in fact a few pixels away from the true center of the bubble. Thus, about half of the improvement in accuracy that we saw with the expert's form description resulted from correctly specified bubble locations and the other half resulted from the extra alignment features.

To assess the impact that the incorrectly classified bubbles had on the number of fields that required correction in ODK Collect, we calculated the percentage of fields in which the ODK Scan tally differed from the groundtruth tally. Using the practitioner's form description, 39.4% of fields differed from the groundtruth, whereas with the expert's form description and extra alignment features, 12.3% of fields differed from the groundtruth. In both cases, the values of most of these fields differed from the groundtruth values by a small number of bubbles, as shown in Figure 3.24. The few cases where the ODK Scan values differed from the groundtruth values by a large number (more than 10 bubbles) was likely due to alignment errors in which an entire form field was misaligned. Fortunately, this kind of error occurred infrequently.

Despite the relatively high error rate that occurred with the practitioner's form description, we observed a high level of overall accuracy in the data that the district supervisors submitted to ODK Aggregate. The accuracy of manually entered numeric and text fields was 98.5%, and the overall accuracy of all data types submitted to the server was 98.1%. This indicates that users were correcting the errors made by ODK Scan. The implications of the differences in accuracy between the practitioner's form description and the expert's form description are discussed in Section 3.5.6.

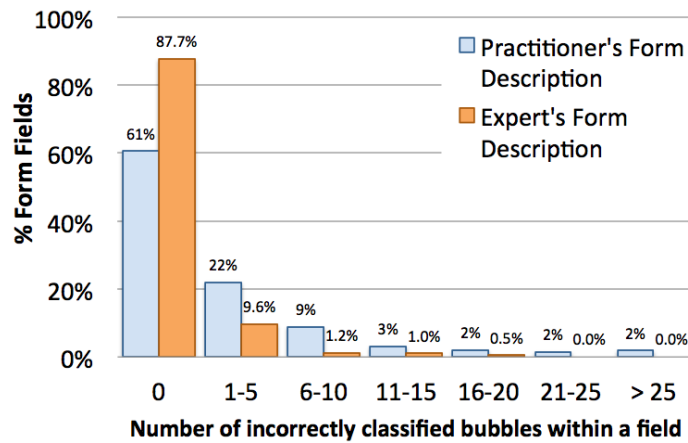


Figure 3.24: Percentage of form fields with differences between ODK Scan tallies and groundtruth tallies using the practitioner’s form description and the expert’s form description with additional alignment features enabled.

Time

To measure how long it took to digitize data, we timed the district supervisors as they digitized several forms every month. We also measured how long it took for the supervisors to complete the ODK Scan portion of the process followed by ODK Collect with image snippets. On average it took district supervisors 12 minutes to complete the scanning and data entry for each two-page consumption form. The ODK Scan portion took about 2 minutes and the data collection in ODK Collect took an additional 10 minutes. Interestingly, the amount of time required for data entry did not change substantially from month to month, with supervisors needing about the same amount of time to digitize the data at the end of the study as at the beginning. Feedback obtained from the supervisors and reasons for the amount of time required to digitize the data are discussed further in Section 3.5.7.

Cost

We assessed the total cost of deploying and sustaining the intervention by summing the total costs of the devices, paper forms, data transmission, data storage, and personnel. The two district supervisors in the study were each issued an HTC Nexus One Android device that

cost \$320. Obtaining and printing all the paper forms cost less than \$25 per month. The airtime for transmitting the digitized data cost \$0.05 per form (or roughly \$1.50 per month) and storing the data in the cloud cost \$8.50 per month. Additional costs included the transportation and per diem costs for the 45 CHWs and 2 district supervisors to attend the training sessions, and the salaries of the practitioner staff who implemented and supported the deployment. We estimate that monitoring the CHWs and district supervisors required 10-15 days worth of field staff time per month, while designing the form and creating the form description cost another 10 days worth of technical staff time at the start of the study. Thus, as is common in ICTD interventions, we found that personnel costs far exceeded the costs associated with the technology and we recommend that other organizations pay close attention to this factor for any deployment.

3.5.6 Qualitative Evaluation

To understand the benefits, challenges, and limitations associated with deploying and sustaining the intervention we collected and analyzed qualitative data from three different user groups: NGO practitioners, CHWs, and district supervisors. Prior studies have shown that qualitative responses may suffer from participant response bias [39] [78]. To mitigate the potential for response bias, we conducted multiple interviews and observations over the entire study period and compared how users' experiences and attitudes changed over time. In addition, to elicit varied responses from participants, the interviews at the end of the study used different language and wording to the interviews conducted during the study. Finally, we also compared the qualitative data obtained between user groups to better understand the impact across the different levels of the information hierarchy.

Practitioner Experiences

We conducted semi-structured interviews with the practitioner staff involved in the intervention. Each interview lasted roughly one hour and involved discussing all of the tasks performed by the practitioner, their opinions and experiences with the technology, and any challenges or issues that arose while supporting the CHWs and district supervisors.

Findings. Practitioners reported that designing the consumption form was time-consuming because many stakeholders (NGO and Ministry of Health staff) wanted different data points included on the form. This resulted in a large amount of design iteration and the form layout and fields changed every time the design changed. In addition, each new design had to go through a review and approvals process and a new form description had to be created.

Creating the ODK Scan form description was also time-consuming and challenging. Although the template-maker software provided a GUI to reduce the amount of JSON that had to be written, the form designer still had to manually edit JSON to adjust bubble locations and provide labels and identifiers for form fields. The form contained 21 checkboxes and 1395 bubbles in 16 tally fields. To specify a bubble location, the form designer clicked on the center of the bubble in the template maker. However, it was easy to miss the center by a few pixels, and adjusting the bubble location required manually editing coordinates in a side panel on the user interface. In addition, every time the form changed, a large amount of work was required to adjust the form description, including updating the locations of each bubble. It is likely that the large number of changes to the form design and description introduced some systematic alignment issues that led to the relatively high error rate encountered in the field. This is discussed further in Section 3.5.7.

Practitioner field staff also reported facing several challenges while supporting CHWs in the field. Many CHWs lived and worked in very remote places, and in some cases it took over 4 hours to reach a CHW. On several occasions the practitioner was unable to reach a CHW because the road was washed away. An important implication of this finding is that it must have also been challenging or impossible for the CHW to travel to the health center to receive a supply kit.

CHW Experiences

To evaluate the experiences of the CHWs, we chose a random sample of 8-9 CHWs each month (4-5 from each district) and visited each CHW in their community to observe and assess their performance. In total, we visited 33 CHWs over the four month study period. Each visit lasted between 30 mins and 2 hours. To evaluate the accuracy of the CHW's

consumption form, we first checked their stock level for each health commodity and compared their real stock level with the stock level on the form. We also checked the CHW's patient register, calculated the number of each commodity that had been distributed and compared that number to the amount indicated on the consumption form. Finally, we performed a semi-structured interview with the CHW to understand their experiences with the consumption form and other issues that may have affected their performance.

Findings. The majority of CHWs visited were using the form correctly to record their consumption of medicines, and the amount of stock they possessed matched the stock recorded on the form. However, several CHWs experienced problems with the form. The most common problem was that CHWs would forget to record their initial stock levels at the beginning of the month and would later try to guess how much stock they had started with. This problem occurred with 7 of the 33 CHWs that we visited and was exacerbated in one district during the month of February because the CHWs did not receive their monthly stipends from the Ministry of Health and so lacked the motivation to record their data. For 2 of the 33 CHW visits we found the data on the consumption form did not match the physical stock levels or patient register, and for another 2 we found the CHWs were not using the form because it was too complicated and they did not know how to fill it correctly.

There were also several occasions where delivery delays resulted in CHWs not receiving their full kit of health supplies on time. As a result, some CHWs were forced to improvise to meet the needs of their patients. For example, when some CHWs did not receive 250mg paracetamol tablets, they would break 500mg tablets in half and distribute them as 250mg tablets. This behavior affected the data on the forms since the CHWs reported distributing 250mg tablets despite not having received any stock that month. In addition, one of the medicines was delivered to CHWs in packages of nine tablets. As a result, to record the number of tablets distributed, CHWs had to multiply the number of packages by nine. Several CHWs found it difficult to do this calculation correctly and made mistakes on the consumption forms. We also observed some interesting additions made to forms. For example, as shown in Figure 3.25, one CHW ran out of bubbles to fill in and improvised by drawing additional bubbles on the form to record the commodities distributed to patients.

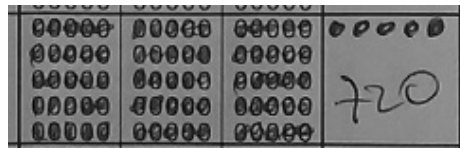


Figure 3.25: Excerpt of a form showing extra bubbles drawn in by the CHW.

District Supervisor Experiences

To assess the experiences of the district supervisors, we visited each supervisor once a month for the four months of the study. Each visit lasted several hours, in which we checked on the state of the phone and watched the supervisors digitize data from several consumption forms. We also conducted semi-structured interviews to monitor the supervisors' experiences with the application and document any issues that came up during the month. At the end of the study, we performed an additional semi-structured interview with each district supervisor to further understand the impact that the new technology had on their work.

Findings. At the start of the study, each district supervisor reported owning a feature phone, although neither had used a touchscreen device before. As a result, they were somewhat hesitant to use the phone at first and expressed concern that they might break the phone or the application. However, their proficiency with the application increased substantially after the first month and they became visibly more confident and comfortable with the technology. Both supervisors were very careful with the phones and kept them locked up when they were not in use. None of the phones were broken, lost, or stolen during the study. The supervisors were able to charge the phones and did not experience any hardware or software malfunctions. In general, the supervisors reported having sufficient network connectivity to send data to the server. This is supported by the fact that the server successfully received regular data transmissions from the supervisors throughout the study. However, the supervisors also said that sometimes they needed to go outside to find a spot with good connectivity before being able to successfully transmit data. This confirms the importance of being able to store data locally on the phone and transmit it when a connection is available.

Supervisors described data entry with ODK Scan as being more efficient than paper based summaries because they were able to process the form from anywhere since the images of each data field were displayed on the phone. However, the supervisors also found verifying the data in ODK Collect and manually entering non-machine readable data types to be time-consuming. One reported, *“In the beginning I took the whole day or hours to do the job. I had all my focus on it, and when I stopped, I lost my concentration.”* One reason that it took a long time to enter data may be the decreased level of ODK Scan accuracy discussed in Section 3.5.5. Another reason could be that the form contained 68 numerical fields that needed to be typed into the phone. Form designers should therefore try to minimize the number of fields that must be manually entered and should also provide likely default values for as many fields as possible.

Although it took a long time to digitize the forms, the supervisors both felt that using the application was more efficient and saved them time over a purely paper-based reporting system. One supervisor told us, *“Using ODK Scan it is quicker to send the data to anywhere. When I have written data, I need to write everything down, then take it for [an approval] signature, then find transport to send that information to [the provincial office]. It can take days to do that.”* In addition, both supervisors noted that the forms made it easier to track health supplies in a standardized way. One supervisor said, *“Before [this intervention], we used to control the medicines in a rudimentary way because there were different pieces of information in different sources. This form provides the information all in one place. It is very useful for me. I can just see the form and know what happened. Before, I would have to go to different data sources to confirm everything.”* In addition they liked no longer needing to prepare handwritten reports: *“The system facilitates the sending of reports, so there is no need for me to move.”*

Finally, the supervisors cared about the accuracy of the data. We observed that they sometimes used white-out to correct errors on CHW consumption forms, and they also took time to use ODK Collect to correct classification errors made by ODK Scan. The supervisors also recognized that the data collected would be beneficial for other stakeholders: *“This information is very useful for the pharmacy to see the consumption patterns and to know how much [stock] should be sent to the district”* and they recommended that the

intervention be scaled up: *“This is very important [for other districts] because we can now track the consumption of medicines. It was something that we weren’t doing before this was started. It is important for us to know and it helps us to know how to use the medicines in the right way.”*

3.5.7 Discussion

Study Size

The success of any intervention depends on the input and buy-in of all partners, including public health staff, NGO and technical assistance staff, and end users. Although our study targets only two district supervisors as the end-users of ODK Scan, a much larger number of people were impacted by the choice to integrate the technology into the CHW supply chain. To support the district supervisors, several practitioner staff learned ODK Scan, created training and trouble-shooting materials, organized training sessions, and supported the district supervisors throughout the study period. Several other technical and managerial NGO staff designed the consumption form, consulted with other stakeholders and ensured that the form was compatible with ODK Scan. Finally, 45 CHWs had to learn how to use the consumption forms. This study therefore impacted at least 2 district supervisors, 5 NGO practitioner staff and 45 CHWs. Understanding the impact that integrating ODK Scan has at this relatively small scale, but at multiple levels of the information hierarchy, is a critical step in understanding how the technology might be successfully scaled up.

Form Design and Description

It is encouraging that the NGO practitioners in this study were able to design an ODK Scan compatible form, create the corresponding form description file and successfully deploy the application. However, creating the form description file was technically challenging and required substantial time and effort. In addition, the impact that using a slightly inferior form description had on the overall accuracy of the scanning and subsequently on the number of fields that had to be corrected in ODK Collect suggests that generating a highly accurate form description file should be a priority for any organization wishing to use ODK Scan. In

addition, rigorously testing the form description with forms that are in very poor condition is essential to ensure that the extra ODK Scan alignment features are enabled if necessary.

The biggest challenge in creating the form description file was specifying hundreds of the bubble locations with pixel-perfect accuracy. When specifying exact bubble locations by clicking on points in a form field, it was easy to miss by a few pixels and this led to a higher number of incorrect bubble classifications. In addition, whenever the form image changed slightly, a large amount of work was required to adjust the corresponding form description. Although we added a number of transformation controls to address this problem, we found that generally the bubble locations did not match up perfectly after a transformation due to rounding errors and other minor form differences. Finally, the current version of the software requires that the user know how to write and edit JSON. A more polished GUI that hides the JSON from the user might have decreased the amount of JSON that had to be written by the user, although we saw that the ability to copy and paste sections of JSON generally made editing form descriptions easier.

Lessons Learned and Recommendations

Many published ICTD deployments are conducted by developers or researchers, not by practitioners. Although practitioners are frequently able to imagine the potential benefits afforded by a new technology, they often do not understand the technical expertise or effort required to deploy the technology successfully, which may lead to false expectations. One of the major lessons learned in this work is that it required a substantial amount of time, effort, and patience on the part of the practitioners to successfully deploy and sustain the intervention. Practitioners had to learn how to use the technology themselves and also how to provide support to CHWs and supervisors through regular visits and monitoring. Without this effort it is unlikely that the deployment would have been successful.

One of the major benefits of ODK Scan is that it allows people to continue to use cheap and familiar paper forms at the bottom level of the information hierarchy. However, the ease of use of paper forms also means that people are likely to use them in unexpected ways. For example, we observed district supervisors using white-out to correct errors on consumption

forms prior to using ODK Scan. We also found that CHWs sometimes drew additional bubbles on the forms (see Figure 3.25). These findings highlight an interesting trade-off between human-readable and machine-readable data. On one hand, it is clearly beneficial to have a person verify all of the processed data values. On the other hand, requiring that a person verify all of the data increases the time required for data entry. Ultimately it will be up to the user to manage this trade-off and decide the level of human verification required for a given application.

We also experienced a number of implementation issues that affected the deployment. For example, district supervisors did not always receive consumption forms from some CHWs for a variety of reasons: some CHWs lived very far away and were unable to submit forms, some did not fill out the forms because they did not receive their kit of supplies, while others were simply not working or had left the area for personal reasons. When asked about the missing forms, one supervisor told us, *“The program will only work as much as the people work”*. In addition, as noted by another supervisor, *“In February we had a lot of mistakes [on the forms] due to lack of CHW motivation because they did not receive their stipends.”* However, despite these issues the supervisors were happy to receive data from the CHWs that did submit forms and frequently stated that having access to the data improved their knowledge and decision-making ability regarding the CHWs’ medical commodity needs and helped them to calculate the resupply amounts required.

3.5.8 Conclusion

This section describes an in-depth analysis of the work required to integrate ODK Scan into the CHW supply chain in Mozambique. ODK Scan was successfully deployed in two districts for four months and the application proved to be a usable and useful tool for recording CHW commodity consumption data. We found the process of identifying and documenting the benefits and challenges associated with deploying and sustaining the technology to be extremely valuable and have learned many lessons that we hope will be useful to other researchers and practitioners who want to deploy ODK Scan or other applications that bridge the gap between paper-based and digital data collection in low-resource settings.

3.6 Summary

This chapter describes our work designing, building, and evaluating a mobile camera-based system that aims to improve paper-digital workflows by automatically digitizing data from paper forms. In particular we contribute:

1. A qualitative study that analyzes paper-digital workflows in global development settings, highlights the challenges that organizations face transitioning data between paper and digital formats, and identifies design opportunities for new tools that could bridge the paper-digital divide [38].
2. The design and implementation of ODK Scan, a mobile camera-based system that uses computer vision techniques running on commercially available mobile devices to automatically classify machine-readable data types recorded on paper forms [34].
3. A controlled laboratory analysis that provides quantitative and qualitative data demonstrating that our digitization techniques afford significant time savings without loss of accuracy [35].
4. A four-month deployment that documents the benefits and challenges of using the system as part of an intervention to strengthen the CHW medical commodity supply chain in two districts in Mozambique [36].

Taken together, these contributions demonstrate how a mobile, camera-based system could be used to improve global development workflows for organizations operating in low-resource environments. ODK Scan is a free and open-source tool that we have made publicly available as part of the Open Data Kit platform.

Chapter 4

**A SYSTEM FOR ANALYZING DIAGNOSTIC TESTS FOR
INFECTIOUS DISEASES****4.1 Introduction**

Frequently, populations living in developing regions face a disproportionately high burden of disease. For example, Sub-Saharan Africa experiences 25% of the global burden of disease, but has only 3% of the world's health professionals [147]. Providing communities living in these regions with access to accurate, convenient and affordable diagnostic tests could facilitate the control of many infectious diseases. However, many of the diagnostic tests that are routinely administered in clinical laboratories in developed countries are inappropriate for point-of-care settings in developing countries, where low-income patients may be most accessible. As a result, these patients do not have access to adequate medical testing and suffer a high mortality rate from diseases that could be quickly diagnosed and cured.

Recent medical research has led to the emergence of innovative rapid diagnostic tests (RDTs) that specifically target the needs of patients in developing countries. These low-cost, disposable tests are capable of diagnosing a variety of common diseases, such as malaria, syphilis, and HIV. RDTs contain all of the elements required to process a biological sample at the point of care, so that the sample does not need to be refrigerated, protected, or transported. Additionally, many of these tests run rapidly, allowing health workers to view results and treat patients immediately. This speed is advantageous because many patients travel long distances to reach medical facilities and may be unable to return easily to collect test results, delaying treatment.

However, although the potential benefits of these new diagnostic technologies are immense, research has shown that health workers often make mistakes when administering the test or interpreting the results by eye [66]. The number and variety of RDTs that are available is also expanding rapidly, and different tests often require slightly different pro-

cedures or interpretation. In addition, medical researchers are in the process of developing more sophisticated tests whose results require quantification or time-sensitive analysis. For example, researchers at the Burnet Institute are developing a rapid test to quantify a patient's CD4 count [13], while Stevens et al. demonstrated that time-sensitive analysis of a microfluidic assay test can substantially increase the dynamic range of the test [130]. These tests will require health workers to simultaneously manage the timing sequences for the test, monitor the rates at which test signals are changing, and quantify the overall result. Furthermore, manually recording data about the RDTs administered, and updating patients' records with the test results can be a complex and time-consuming process.

This chapter describes our research designing, implementing, and evaluating a mobile camera-based system that aims to support health workers in the field by automatically interpreting diagnostic tests for infectious diseases using image processing performed on a mobile device, and transmitting data regarding the test, results, and the patient to a remote server to improve disease surveillance and tracking. In particular, we contribute:

1. The design and implementation of ODK Diagnostics, a mobile camera-based system capable of automatically interpreting a variety of commercially available rapid diagnostic tests using image processing performed entirely on the device [33].
2. An eight-week field deployment of the ODK Diagnostics system with 60 health workers at five hospitals and clinics in Zimbabwe that demonstrates the benefits and challenges of integrating the system into daily clinical workflows [37].
3. The design, implementation, and technical evaluation of a mobile, camera-based system that automatically quantifies time-sensitive diagnostic test data for an experimental microfluidic test that is currently under development [40].

Taken together, these contributions demonstrate the potential for mobile, camera-based systems to assist health workers with the process of quickly and accurately diagnosing a variety of infectious diseases. Our findings could inform the design of future systems that aim to support health workers in the field and help to guide ministries of health and other stakeholders working to deploy mobile health systems in similar environments.

4.2 A System for Automatically Analyzing Rapid Diagnostic Tests

This section seeks to answer the following research question: *How can we support health workers as they are required to make increasingly complex diagnostic decisions for an increasing number of infectious diseases?* To address this question, we created ODK Diagnostics, a mobile, camera-based system that supports health workers in three ways: (1) by facilitating the creation of digital job aids that provide in-context assistance to users administering RDTs, (2) by automatically interpreting the RDT results using computer vision algorithms running on the phone, and (3) by automating the process of collecting data regarding the test administered and its outcome.

The main contribution of this section is the design and implementation of an end-to-end solution for administering and analyzing a variety of commercially available RDTs. Our solution provides several key benefits over existing solutions, including:

- Digital job aids can provide in-context, on-demand assistance to workers in the field.
- Commercially available smartphones can be used to analyze a wide variety of different tests, negating the need for specialized reader devices.
- The system can handle both simple, binary tests and more complex tests that may require quantification or time-sensitive analysis.
- The system can easily keep a record of all the tests administered and their outcomes for quality control monitoring and evaluation.
- Analyzing the test results can become a standardized, auditable and adjustable process without needing to retrain users.
- New diagnostic tests can easily be added to the system as they become available.

4.2.1 System Design

We designed ODK Diagnostics to assist health workers with administering and interpreting a variety of RDTs. We wanted to create a platform that can be easily extended to incorporate many RDTs of different types and formats and so we have tried to minimize the technical knowledge required to add a diagnostic test to the system. However, we anticipate that the process of adding an RDT to the system will not be done by the health worker who

will use the system, but rather by someone higher up in the information hierarchy and then distributed to health workers for use in the field. ODK Diagnostics aids health workers in three ways: (1) by providing a variety of digital job aids to give in-context assistance to users administering RDTs, (2) by automatically analyzing the RDT and delivering the diagnosis, and (3) by automating the data collection regarding the type and outcome of the test. We have designed these components so that they can be installed and operated independently of each other, since some programs may want to make use of digital job aids but not automatic interpretation of RDTs, while some may elect to skip the job aids and use only the component for automatically interpreting RDTs.

Digital Job Aids for RDTs

One set of tools that has been developed to assist health workers with the process of administering and interpreting RDTs is paper-based job aids. A job aid is typically a one or two-page instruction sheet that provides a step-by-step explanation of how to administer an RDT along with clearly worded instructions and visual depictions that guide the health worker through the procedure. Figure 4.1 shows part of a paper-based job aid. Job aids have been shown to increase the proportion of tests that are administered correctly [66].

However, paper-based job aids have a number of limitations. They need to be printed, transported and distributed to health workers, which makes it difficult to easily update the materials or change the type of RDT being used. This could be problematic since health workers might inadvertently be referring to out-of-date or incorrect information in a paper-based job aid. Furthermore, as the variety and complexity of the RDTs used at the point of care increases, health workers will be required to carry around and refer to a large number of different paper-based job aids. Most importantly, the procedure might be complex and require calculations or decision-making that are difficult to follow on paper, particularly when a health worker is juggling other tasks and interacting with one or more patients.

As an alternative to paper-based aids, prior research has demonstrated the potential for mobile technologies to strengthen healthcare systems [42] [95]. In particular, the growing adoption of smartphones in developing countries suggests that they may be an effective

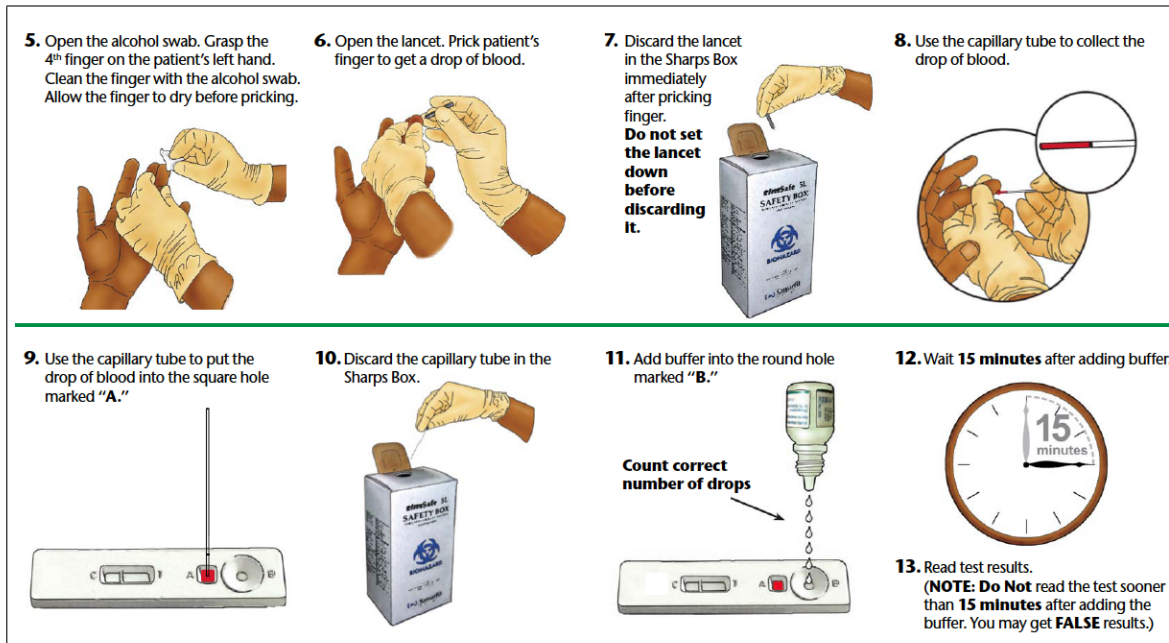


Figure 4.1: A portion of the paper-based malaria RDT job aid.

platform for a point-of-care diagnostic system. Using a smartphone to provide digital job aids that assist health workers with point-of-care diagnostic tests will eliminate the need for them to transport and protect paper-based job aids. The digital job aid can be programmed to automatically guide the health worker through the correct flow of steps that apply to the particular RDT she is administering, including providing accurate timing for events such as when to read the RDT results, which will increase the likelihood that the test will be administered and interpreted correctly. Digital job aids will also be easier to distribute, update and translate than paper-based job aids, provided each health worker with a device is able to periodically visit an area with Internet connectivity and download the requisite materials, a now common situation in most developing countries. In addition, smartphones are portable and have the capacity to store a large number of job aids, so the device also serves as an instantly accessible RDT instruction manual.

In this work, we focus on digitally replicating several widely used and WHO approved paper-based job aids, such as the malaria RDT job aid (partially shown in Figure 4.1). We

rely on medical experts to create the appropriate job aid content and focus on building a system that simplifies the process of creating and distributing the job aids. However, the capabilities of the technology mean that, in the future, more sophisticated job aids could be created that use a variety of additional media, like voice recordings or short videos that demonstrate how to correctly complete tasks.

ODK Diagnostics also leverages the fact that many RDTs share a number of common steps, such as asking the health worker to put on latex gloves, or waiting a specific amount of time after starting the test before reading the results. We allow users to reuse and customize a set of basic job aid template files, rather than requiring these common steps to be reprogrammed every time a new job aid is created. Our design also allows the digital job aids to be used either as a standalone platform or in conjunction with our algorithm for automatically interpreting the RDTs. The advantage of this separation is that the system is not limited to the creation of job aids for RDTs, but could also be used to create and distribute supportive material for other medical procedures.

Optical Processing of RDTs

Although interpreting RDT results may appear to be simple, research has shown that incorrect interpretation of test results accounts for a large proportion of the errors made by health workers when administering RDTs [66]. In addition, as more RDTs are developed for a wider variety of diseases, the number of different RDTs that health workers are required to use is expanding. Many of these RDTs vary slightly in their interpretation, such as having a different number of test lines that need to be read. Moreover, several companies are starting to manufacture multi-RDTs that test for more than one disease on a single cartridge [94], which increases the complexity of correctly interpreting all of the test results. Furthermore, more sophisticated diagnostic tests are currently being developed that will require the user to quantify the test result [30] or record the rate of change of the test signal [130], which are likely to be challenging or impossible to interpret by eye.

To address these challenges, we designed an algorithm that automatically interprets the RDT results using computer vision algorithms running on the phone. Our initial implemen-



Figure 4.2: Low-cost 3D-printed platform used to capture images of RDTs.

tation described here focuses on interpreting simple RDTs that are already commercially available. However, we have designed the system to be easily extended to more complex tests by using a lightweight test description language to specify the regions of interest on the RDT and how they should be interpreted. Our work analyzing more sophisticated, time-sensitive diagnostic test signals is described later in this chapter.

To make it easy for health workers to process RDTs correctly, we designed and 3D-printed a small stand, shown in Figure 4.2, that holds the phone in position above the RDT. This stand ensures that users can interact with the system without needing to pick up the phone, and we anticipate that this will be advantageous for health workers who are busy interacting with patients and handling RDTs and biological samples. The stand design leaves the upper portion of the phone exposed so that the system will work correctly with different phone models and varying camera positions. In addition, the position at which the RDT should be placed below the camera is indicated by a rectangular ridge that allows a variety of RDTs of different shapes and sizes to be correctly positioned without needing to adjust the stand or camera placement. In our mechanical design, we sought to develop an approach that requires minimal handling (and thus lowers risk of contamination of both the RDTs as well as the stand and phone). Finally, our design does not require any additional

lighting, sensors or batteries to be attached to the phone. Instead, we normalize the lighting of the RDT using a portion of the test strip that does not contain any reactive agent and can therefore be reliably compared to the reactive regions of the test.

Data Collection

Collecting timely and accurate data regarding the number, type, and results of RDTs administered by health workers is extremely important for a number of reasons. First, decision-makers need to monitor consumption of RDTs so that they can ensure health workers are supplied with the correct number of tests to cover the patient population. Oversupply of RDTs will result in expired tests that should be discarded since they might produce inaccurate results if inadvertently used to diagnose patients, while undersupply will result in health workers being unable to diagnose patients since they do not possess the relevant test. RDT consumption data is also important for generating reports for donor and aid agencies regarding the use of funds that they provided. Additionally, providing quick access to accurate statistics regarding the location and results of RDTs could help decision-makers to analyze the spread of infectious diseases and facilitate the early detection of outbreaks.

RDT data is currently predominantly collected on paper forms. A plethora of prior research discusses the limitations of paper-based data collection [65] [107], and the benefits provided by digital data collection solutions [95] [88]. Our design replaces paper-based collection of RDT data with a digital data collection toolkit running on the device. The type, location, and outcome of the test can be collected automatically by the software, stored locally on the phone, and uploaded to a remote database if and when an Internet connection is available. This will substantially decrease the time that it takes to make the data available to decision-makers and will also eliminate the potential for transcription and aggregation errors.

4.2.2 Implementation

We built ODK Diagnostics using the Android platform. The decreasing cost of devices and open source nature of Android, coupled with the growing adoption of the platform

in developing countries, made Android an attractive choice. Although we designed ODK Diagnostics as a generalizable tool to assist users with a variety of diagnostic tests, our initial implementation focuses on a set of commercially available and widely used RDTs for malaria and HIV, which will allow us to deploy and evaluate the system with health workers who currently use these RDTs in the field.

Digital Job Aids for RDTs

We chose to implement the digital job aids using HTML and JavaScript rather than native Android code since we anticipate that it will be easier for users with limited technical experience to work with HTML rather than native Android code. There are also a vast number of tools available to assist users with the creation of HTML content. Each job aid consists of a set of simple HTML pages with a small amount of common JavaScript. To construct a new job aid, users create or modify template HTML and JavaScript files. We implemented the navigation and user interface components using jQuery Mobile, and the job aid is packaged as an Android application using PhoneGap [1], which allows it to be installed locally and additionally provides access to native functionality.

We created digital versions of several paper-based job aids that are used by health workers to administer malaria and HIV RDTs. Figure 4.1 shows part of the paper based job aid for the malaria RDT. For the digital job aid, each step in the process is shown individually on the screen and numbered to provide context of where the user is in the process (e.g. Step 5 of 16). Figure 4.3 depicts screenshots of the digital job aid for the malaria RDT. Users navigate the system by pressing buttons or swiping to go to the next or previous step. When the user reaches the point in the RDT at which s/he has to wait before reading the results, the application automatically starts a timer, and alerts the user when the correct amount of time has passed and the results are ready for interpretation. To ensure that health workers can use the digital job aids as a standalone tool, we have also implemented the decision tree that guides users through the process of manually interpreting the RDT results. Users can then choose to read the results visually or automatically using the algorithm described next.

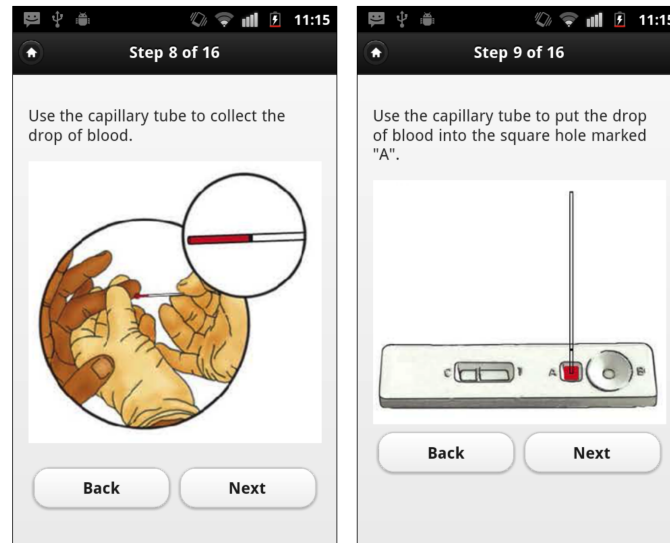


Figure 4.3: Screenshots from the digital job aid for the malaria RDT.

Optical Processing of RDTs

After the health worker has been guided through the process of administering the RDT, the next task is to obtain the test result. Our algorithm for automatically interpreting the RDT results is implemented as a native Android application, with the image processing components making use of a Java implementation of OpenCV [104], an open source computer vision library. All of the image processing is performed on the phone without requiring an Internet connection. There are five main processing steps required to automatically read and interpret an RDT that we describe below.

a. Create RDT Description and Reference Image. Our system design attempts to minimize the complexity of adding new RDTs to the system. For each type of RDT to be processed, the user must provide two files: (1) an image of an unused RDT that will be used as a template for aligning captured images and (2) a test description file that will be used to locate and identify the regions of interest. We utilize a generalizable JSON [31] format to specify the location, size, and type of each RDT region of interest, such as the control lines, test lines, and sample wells (see Figure 4.4). In addition, users must specify a

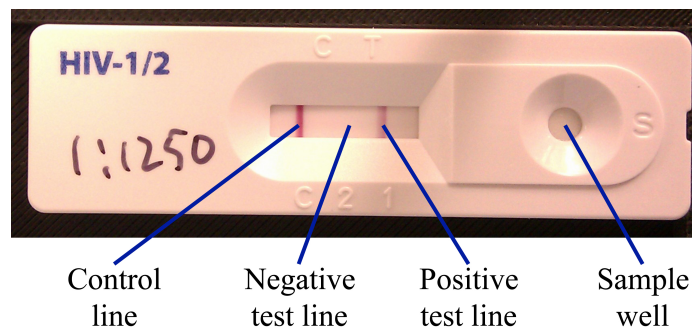


Figure 4.4: Regions of interest on an RDT for HIV.

region of the RDT strip that is *empty*, in which no test lines or control lines are located. On average, each test description file required roughly 10 lines of simple JSON. Creating the test description file is a preliminary step that only needs to be performed once for each type of RDT. We anticipate that the health workers who use the application in the field will not create the test description file. Instead, the description file will be created by a supervisor or technician higher up in the information hierarchy and distributed to health workers.

b. Image Capture. Once the test description file and reference image have been created and loaded on to the phone, the application is ready to capture and process RDTs. The first step in the processing pipeline is to use the phone’s camera to capture an image of the RDT. To ensure accurate processing, images should be well focused and taken while the camera is steady. The captured image should contain all of the RDT content while maintaining a minimal distance from the test. To make it easy to take photos under optimal conditions, we designed and 3D-printed a small stand that holds the phone in position above the test. As shown in Figure 4.2, the phone is placed on the stand and the RDT is placed beneath the camera, thereby ensuring that both the camera and the RDT 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. Additionally, making use of the stand may actually be advantageous, since the user is not required to pick up the phone and can instead focus on handling the RDT correctly. To capture an image, the user places the RDT in position below the phone and

launches the Android camera application. The user can then take and retake photographs of the test. When satisfied with the captured image, the user presses a button to accept the image, which is saved and passed to ODK Diagnostics for further processing.

c. Image Registration. After an image of the RDT has been captured, the next step is to locate the portion of the image that contains the test content. This involves locating and spatially transforming the picture of the RDT to align with the reference image that is stored on the phone. Alignment is necessary to ensure that the entire RDT has been captured and to determine the locations of each specific RDT element. To perform the image registration, we first convert the captured color image to grayscale, and then use a k-means clustering algorithm [86] to separate the image into the portion that contains the RDT and the portion that contains the background. We then use contour-finding to isolate the set of points that represents the RDT, and compute the minimum-area bounding rotated rectangle that contains this set of points. We rotate the rectangle that contains the RDT to be horizontal, and crop the image to contain only this rectangle. Finally, we compute a transformation that maps the cropped RDT image to the reference image, thereby establishing a point-by-point correspondence between the two images. We then save an image of the aligned RDT and display it on the screen so that users can check the alignment prior to processing the test results.

d. Locate RDT Elements and Compute Results. Once the RDT has been successfully aligned, the next step is to locate and process those parts of the RDT that depict the test results. To do this, we normalize the color intensities of the captured image and use a Gaussian blur to remove noise. Then, the description file for the RDT is loaded and parsed, and the locations of each RDT element extracted. The algorithm then looks for a field that has been labeled as *empty*. This *empty* region corresponds to a portion of the test in which no test lines or control lines are located. We compute the average pixel intensity of this *empty* region and use it to determine a threshold intensity that will be used to determine whether a line is present in the fields that are labeled as *control* and *test* regions.

After the threshold value has been calculated, each RDT field that is labeled as being either a *control* line or a *test* line is processed. To do this, we perform an image subtraction

to determine the difference in pixel intensities between the *empty* region and the *test* or *control* region. If the region of the image that we are testing does not contain a line, the results of the image subtraction will be close to zero, since the pixel intensities in the *test* region will be equal to the pixel intensities in the *empty* region. However, if the region that we are testing does contain a line, the results of the subtraction will differ from zero.

To determine if a line is present, we analyze the results of the image subtraction for every column of pixels in the region being tested. For each pixel in the column, we see if the absolute value of the image subtraction for that pixel is greater than the threshold value we calculated from the *empty* region. To increase the chances that the algorithm will successfully process faint test lines that represent low-positive results, we set the threshold value to be small, but require that the majority of pixels in the column exceed the threshold before we determine the presence of a line.

Once we have determined whether the region being tested contains a line, we use the label of the region to compute the final outcome. If the region is labeled as a *control* region, a line indicates the test is valid and no line indicates the test is invalid. If the test is invalid no further results are displayed. The system instead alerts the user that the test is invalid and recommends that the user repeat the test using a new RDT. If the test is valid, then for each region labeled as a *test* region, the presence of a line indicates a positive result, and no line indicates a negative result.

e. Save Output and Display Results. After all the *control* and *test* regions of the RDT have been processed, the outcome of the test needs to be saved and displayed. To do this, the system creates a JSON file that contains the name and type of each RDT region and the corresponding result of processing for that region. In addition, the system also saves the original captured image and a marked-up image depicting the results of alignment and processing. The system draws a colored box around each region that indicates the result: a green box indicates that a line was detected, while a red box indicates that no line was detected. These saved images can be used to resolve any discrepancies that may arise after the RDT is processed and will ensure that a record of the test is preserved and archived. The test results are also displayed textually and visually on the screen.

Data Collection

After the RDT has been interpreted and the diagnosis displayed to the health worker, the next step is to collect and store data to keep a record of the test and its outcome. As discussed in Section 4.2.1, it is important to collect timely and accurate data regarding the numbers, types, and results of RDTs administered and to make this data available to decision-makers quickly and in a usable format. Since there are already a number of open-source data collection tools available that have been designed and tested in developing countries, we decided not to build our own data collection software. Instead, we chose to integrate ODK Diagnostics with the ODK toolkit, and allow users to export the results of RDT processing to ODK Collect. Integrating ODK Diagnostics with ODK Collect has a number of advantages. First, users of ODK Diagnostics will be able to easily combine the RDT data with other data that might be relevant, including GPS coordinates of the test location or additional patient-specific data. Second, through ODK Collect we allow users of ODK Diagnostics to gain access to a number of other useful ODK tools, such as ODK Aggregate and ODK Tables [69], which will facilitate the analysis and visualization of the data collected. In addition to the ODK tools, an increasing number of other platforms, such as Formhub [53] and DHIS [43], have also chosen to integrate their tools with ODK to provide additional services for hosting and data analysis.

4.2.3 Evaluation

We now describe a controlled laboratory study that we conducted to evaluate the technical performance of our algorithm for automatically interpreting RDTs.

Apparatus

The experiments described in this section were conducted in collaboration with researchers at the Global Solutions for Infectious Diseases (GSID) [59] laboratory in San Francisco, USA. GSID is a non-profit global health organization engaged in the development of products to help prevent the spread of infectious diseases, especially in the developing world. Our experiments were designed to test the robustness and accuracy of the interpretation

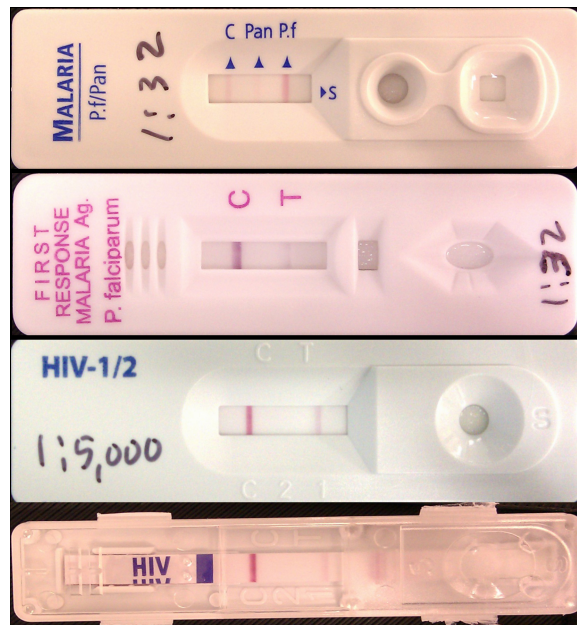


Figure 4.5: We tested four different RDTs. From top to bottom: SD Bioline malaria, First Response malaria, SD HIV 1/2 3.0, and SD 1/2 Multi-device HIV.

algorithm using four sets of commercially available RDTs that were administered with varying sample concentrations (see Figure 4.5). We tested two malaria RDTs (First Response Malaria and SD Bioline Malaria) and two HIV RDTs (SD 1/2 3.0 Multi-device HIV and SD 1/2 3.0 HIV). We selected these tests for a number of reasons. They are all used extensively at the point of care and target diseases that place a significant burden on many developing countries. Additionally, we had access to enough tests to be able to run a dilution series for each test using different sample concentrations and also possessed the necessary malaria whole blood and HIV+ serum for administering the tests in the laboratory. The experiments were performed using an HTC Nexus One Android device that was placed in the 3D-printed stand depicted in Figure 4.2. Since we wanted to minimize the complexity of future user training, we set all of the camera parameters to automatic.

Procedure

Before we could evaluate the performance of our algorithm, we needed to create the appropriate diluted samples and run each set of RDTs. For the malaria RDTs, the dilutions we used were: *undiluted*, *1:8 (strong positive)*, *1:32 (weak positive)* and *negative*. These dilutions were selected based on the researchers' prior experience running RDT dilutions and knowledge of which concentrations would produce strong and weak positive results. Each First Response malaria RDT has one control line and one test line for *Plasmodium falciparum*, while each SD Bioline malaria RDT has one control line and two test lines, one for *Plasmodium falciparum* (Pf) and another for *Falciparum/Pan* (Pan). The malaria whole blood that we used to run the tests contained both Pf and Pan, and so each positive SD malaria test contains three lines (one control and two test lines), while each positive First Response test has two lines (one control and one test). Each negative First Response or SD malaria test has one control line and no test lines. We included a *negative* test series to ensure that the algorithm could correctly determine the absence of a line.

For the HIV RDTs, the dilutions we used were: *1:50 (very strong positive)*, *1:250 (strong positive)*, *1:1250 (weak positive)*, *1:5000 (very weak positive)* and *negative*. Each HIV RDT contains one control line and two test lines. However, as shown in Figure 4.5, the HIV positive serum that we used produces a positive result for only one of the test lines, and so each positive test that we ran contained two lines (one control line and one test line), while each negative test contained one control line and no test line.

Lighting conditions in rural health centers are likely to vary, so in addition to testing the algorithm on RDTs with varying sample concentrations, we also ran tests in both indoor and outdoor lighting conditions. For each RDT, five test images were captured indoors under the lighting in the laboratory, and another five were captured outdoors in direct sunlight. Our final set of test images for First Response Malaria therefore consisted of 4 dilutions x 5 images of each dilution x 2 lighting conditions x 2 lines per test: 80 lines, while the SD Bioline Malaria tests consisted of 4 dilutions x 5 images of each dilution x 2 lighting conditions x 3 lines per test: 120 lines. For each of the SD 1/2 multi-device HIV and the SD 1/2 3.0 HIV, the experiment consisted of 5 dilutions x 5 images of each dilution x 2

lighting conditions x 3 lines per test: 150 lines each. This gives a total of 500 lines for all of the experiments: 350 positive and 150 negative.

Analysis

We ran our algorithm on all of the test images. For each control line or test line, we categorize a correct result (C) as being a true positive (TP) if a line is correctly detected, or true negative (TN) if the absence of a line is correctly detected. We categorize an incorrect result as being a false positive (FP) if a nonexistent line is detected, or false negative (FN) if a line that should be detected is missed. It is interesting to consider the relative importance of each error type. For example, the consequences of a false negative, which may result in an infected patient not receiving treatment, are often more severe than a false positive, which may result in a healthy patient receiving unnecessary treatment.

Results

The experiment results are given in Table 4.1. Out of a total of 140 valid control lines, 138 were correctly interpreted. Two were incorrectly interpreted, and both of these errors were false negatives. One error resulted from a highly overexposed image, depicted in Figure 4.6. This suggests that it would be beneficial to create a way to automatically check the quality of the captured image to ensure that images are not overexposed, underexposed or blurry. However, it is preferable that any errors that do occur are false negatives rather than false positives, since a false negative would simply result in the RDT being repeated for the patient, rather than a potentially dangerous misdiagnosis.

Out of a total of 150 negative results, 147 were correctly interpreted. The three errors were all false positives: one with a First Response malaria negative test and two with an SD Bioline malaria negative test. Out of a total of 210 positive test lines, 204 were correctly interpreted. One of these errors occurred due to the same overexposed image described previously and depicted in Figure 4.6. The other five errors were false negatives that occurred with the SD 1/2 multi-device HIV RDT at a dilution of 1:5000 under the outdoor lighting condition. As shown in Figure 4.7, we saw that placing the SD 1/2 multi-

	Dilution	Indoor lighting						Outdoor lighting					
		Control Lines			Test Lines			Control Lines			Test Lines		
		C	FN	FP	C	FN	FP	C	FN	FP	C	FN	FP
SD	Undiluted	4	1	0	9	1	0	5	0	0	10	0	0
Bioline	1:8	5	0	0	10	0	0	5	0	0	10	0	0
Malaria	1:32	5	0	0	10	0	0	5	0	0	10	0	0
	Negative	5	0	0	10	0	0	4	1	0	8	0	2
First	Undiluted	5	0	0	5	0	0	5	0	0	5	0	0
Response	1:8	5	0	0	5	0	0	5	0	0	5	0	0
Malaria	1:32	5	0	0	5	0	0	5	0	0	5	0	0
	Negative	5	0	0	4	0	1	5	0	0	5	0	0
	1:50	5	0	0	10	0	0	5	0	0	10	0	0
SD	1:250	5	0	0	10	0	0	5	0	0	10	0	0
1/2	1:1250	5	0	0	10	0	0	5	0	0	10	0	0
HIV	1:5000	5	0	0	10	0	0	5	0	0	10	0	0
	Negative	5	0	0	10	0	0	5	0	0	10	0	0
	1:50	5	0	0	10	0	0	5	0	0	10	0	0
SD	1:250	5	0	0	10	0	0	5	0	0	10	0	0
Multi	1:1250	5	0	0	10	0	0	5	0	0	10	0	0
HIV	1:5000	5	0	0	10	0	0	5	0	0	5	5	0
	Negative	5	0	0	10	0	0	5	0	0	10	0	0

Table 4.1: Experimental results for each of the RDTs tested.

device RDT in direct sunlight for a few minutes resulted in condensation on the inside of the cartridge that made it difficult to interpret the low-positive lines. However, the condensation also made it impossible to read the test result by eye, which suggests that this particular RDT should not be used if the test is to be performed outside or in a hot environment.

We also measured the time that it took to process each RDT. On average, it took 2.42 seconds to align the captured image, and 5.72 seconds to process the results, giving a total of 8.14 seconds for the entire RDT. This time could be reduced by using a native implementation (rather than a Java implementation) of the image processing code. However, since most RDTs require that the user wait 15 to 30 minutes before reading the test results, we feel that 8.14 seconds is sufficiently fast for our initial field tests.

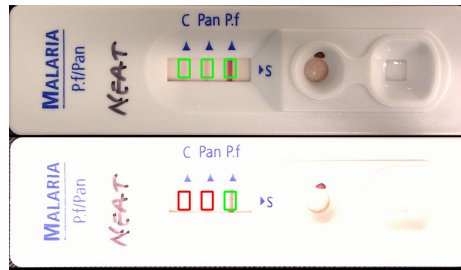


Figure 4.6: Two marked-up SD Bioline RDTs: a properly interpreted RDT (top) and an overexposed RDT with two false negative lines (bottom).

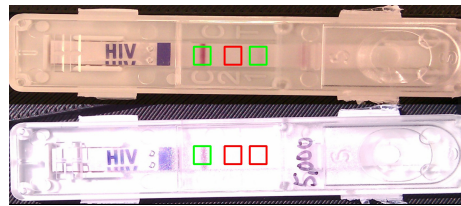


Figure 4.7: Two marked-up SD HIV RDTs: a properly interpreted RDT (top), and an RDT with a false negative test line due to condensation on the inside of the cartridge (bottom).

4.2.4 Discussion

This section describes our first steps towards the goal of creating an electronic point-of-care diagnostic system and our initial technical evaluation of the system has raised a number of interesting issues. First, although running dilution series' may be a good method for initially evaluating the system, the results obtained from diluted samples might not be quite the same as real weak positive results. In addition, we did not analyze which sample concentrations represent the limit of what can be visually interpreted by humans or how many errors health workers make when manually interpreting the test results. To evaluate the system more effectively, the next section in this chapter describes a field evaluation of the system in which we use our system to automatically process RDTs and compare the resulting diagnoses with the visual diagnoses of trained health workers.

Second, the sensitivity of RDTs manufactured in different batches is known to vary slightly. In addition, for some diseases, like malaria or syphilis, the consequences of a

false negative test result are severe, since an infected patient may not receive the necessary treatment. In other cases, a false positive test result might cause a patient to be treated with unnecessary drugs that may have adverse side-effects. These issues suggest that it may be beneficial to create a mechanism for calibrating the sensitivity of the system for each RDT to account for these variations.

Our current design makes use of a low-cost plastic stand to position the phone correctly above the RDT. We anticipate that it will be advantageous for users to not have to pick up and handle the phone, since this will leave their hands free to focus on handling the RDTs and biological samples correctly. However, as with many medical procedures, health workers wear latex gloves when administering RDTs, and although the capacitive touchscreen of the device still works with latex gloves, the gloves may hamper the usability of the interface if they are not tight fitting. Additionally, if a health worker is handling infectious material, such as HIV positive blood, it is likely to be undesirable for her to touch the device at all, since she may accidentally transfer some of the infectious material on to the device, and it will then be dangerous for anyone who is not latex gloves to touch the device. This is an issue that is likely to be relevant for many different mHealth applications, since health workers typically wear gloves when they interact with patients, but may not always wear gloves when they handle the device. One solution is to create a touch-free method of interacting with the system, possibly using voice or gesture recognition, and we believe that this topic holds rich potential for future research.

We integrated ODK Diagnostics with the rest of the ODK toolkit to make it easy for users to collect and analyze data regarding the number, type, and results of RDTs. However, it will also be important to integrate the data with an individual patient's medical record so that patient specific data can be recorded and tracked over time. There are a number of medical record systems for Android that are currently in use in developing countries, and it would be advantageous to integrate ODK Diagnostics with one or more of these systems. ODK Diagnostics provides digital job aids to assist health workers with the process of administering RDTs. In some cases, it may also be appropriate to supply health workers with recommendations for treatment if the test result is positive. Alternatively, the system could suggest additional tests or procedures to try if the test result is negative. For example,

if a child has a fever, but the malaria RDT is negative, the system could provide suggestions for other tests that might help to diagnose the patient.

4.2.5 Conclusion

This section describes the design, implementation, and technical evaluation of ODK Diagnostics, an end-to-end solution for administering and analyzing a variety of existing RDTs. ODK Diagnostics has been designed to support health workers in three ways: (1) by facilitating the creation of digital job aids that provide in-context assistance to users administering RDTs, (2) by automatically interpreting the RDT results using computer vision algorithms running on the phone, and (3) by automating the process of collecting data regarding the test administered and its outcome. Results from our technical evaluation of the system suggest that it is sensitive and accurate enough to be field tested. The next section of this chapter describes a field evaluation of the system with health workers in Zimbabwe.

4.3 Field Evaluation of a Camera-Based Point-of-Care Diagnostic System

This section analyzes the user interaction and deployment challenges that arise when health workers in low-resource settings integrate our mobile camera-based system for interpreting RDTs into their clinical workflow. We conducted an extensive field evaluation of the system with sixty health workers at five hospitals and clinics in Zimbabwe that explores how health workers are able to integrate the system into their daily patient care routines, both at small, rural clinics and at large, urban hospitals. In addition, we expose a variety of issues that are generally applicable to a range of mobile sensor-based systems (such as data collection issues, user errors, infrastructure challenges, etc.).

The field evaluation described in this section focuses on three important research questions: (i) the impact of the mobile system on health workers' patient care routines, (ii) the impact of poor infrastructure on system usage and data collection, and (iii) the quality of the mobile system's automatically computed diagnoses. Our findings show that health workers were able to integrate the system into their clinical workflow and successfully collect extensive amounts of test data after only sixty minutes of training. Furthermore, health workers used our system to capture and analyze diagnostic tests consistently over an eight-week period. In addition, health workers developed a variety of strategies to overcome poor network connectivity and transmit data to a centralized database. Finally, we show strong agreement between the system's computed diagnoses and the visual diagnoses provided by well-trained health workers, which suggests that the system could assist with disease diagnosis in a variety of scenarios. Taken together, our findings will help ministries of health and other stakeholders to assess the viability and acceptability of deploying mobile sensor-based systems to assist health workers in the field. In addition, our insights will guide future researchers working to deploy mobile health systems in similar environments.

4.3.1 Usage Scenario and Design Principles

We target the following usage scenario: a health worker in a rural clinic has been issued a mobile device to assist with the analysis of RDTs. The health worker enters a patient exam room to assess a patient. After spending time interacting with the patient and using the

device to collect patient data, the health worker decides to perform a diagnostic test for malaria. S/he obtains a blood sample from the patient and administers the RDT. Then, s/he uses the device's built-in camera to capture an image of the test. The system processes the image, displays the test result on the screen, and transmits the data to a database if and when there is sufficient connectivity. Finally, based on the test result, the health worker recommends the appropriate treatment or selects another test to run.

In addition to being appropriate for our target problem of interpreting diagnostic tests, many of our design principles are also applicable to a range of other mobile sensor-based systems. For example, similar usage scenarios might apply to capturing images for cell-phone microscopy [12] or sound to analyze lung function [77].

Commercially available devices. Requiring users to import, configure, and maintain custom hardware systems may present a significant barrier for many users in low-resource settings. Therefore, our system runs on commercially available Android devices and only uses sensors that come built into these devices. The system does not use any additional hardware or specialized reader device and analyzes rapid diagnostic tests using only images captured from the built-in camera. The variety of Android devices available in low-resource settings will allow users with varying needs and budgets to choose devices that best fit their requirements. In addition, smartphones are multi-purpose devices that can also provide utility beyond diagnostics (such as voice and text communications, patient management or data reporting).

Local computation. Many low-resource settings lack sufficient Internet connectivity to reliably transmit high-resolution images. Thus, our system performs all of the computation and image processing on the device, which allows the system to be fully functional in the absence of a network connection.

Asynchronous transmission. The quality of the network connection might vary with network traffic, time of day and power outages. Thus, our system supports asynchronous data transmission in which collected images and data are stored locally on the device until a network connection or sufficient bandwidth becomes available for transmission.

Data collection. In addition to capturing and processing images, our system allows users to collect and report data (including text, barcodes, GPS data, etc.) about the patients seen and diseases detected. Rather than building a new data collection platform, we integrated the image processing components of the system with CommCare [95], a widely used data collection platform based on Open Data Kit [65].

Easily configurable. We wanted to make it easy for test manufacturers and clinical experts to add new rapid diagnostic tests to the system and to control the sensitivity of the image-processing algorithm for their specific test. To add a new test, a clinical expert creates a simple text file, called a test description file, which specifies the size and location of the regions that contain the test results. The test description file is uploaded to the device and used to configure the mobile system for use by health workers in the field. The health workers will not adjust the system parameters. Instead, health workers will simply use the smartphone's built-in camera to capture images of tests that have previously been added to the system by clinical experts. A typical test description file consists of about ten lines of text and may be created using any text editor or using a graphical web tool that we developed. Thus far, clinical experts in the laboratory have created and tested description files a variety of rapid diagnostic tests for different diseases and test brands, including tests for malaria, HIV and syphilis [33].

Interaction Design and System Workflow. The previous section describes the system architecture and image processing algorithms in detail [33]. Here, we provide an overview of the steps required for a health worker to interact with the system to capture and process a diagnostic test. A health worker begins by launching the system and using the mobile application to record any relevant patient data (such as age, gender, etc.) They then select the type of diagnostic test from a menu on the screen that contains a list of all the tests that have been added to the system. Selecting a test automatically launches the Android camera application, which allows users to take (and, if necessary, re-take) an image of the relevant diagnostic test. When the user is happy with the image that they captured, they press a button to save and process the image. The next step in the workflow is to align the captured image, which involves separating the part of the image that contains the diagnostic



Figure 4.8: The system’s user interface showing a processed test for malaria.

test cartridge from the background. After aligning the image, the system displays the test on the screen, and provides users with an opportunity to re-take the image should the test appear to be misaligned. If the test has aligned correctly, the user presses a button to compute the diagnosis. The system uses test description file that corresponds to the chosen diagnostic test to locate and process each region on the test that the clinical expert specified as showing a result. For each region, the system uses a thresholding algorithm (described in detail in [33]) to determine if a line is present in that region. Finally, the system uses the combination of detected lines to determine the final test result. After the result has been computed, it is displayed on the screen with an image that shows the user the results of processing so that they can verify the diagnosis. Figure 4.8 shows the system’s user interface for a processed malaria test with a positive result. After checking the result, the user saves and exits the application. The system then stores all of the data associated with the captured test locally on the smartphone, and transmits this data, along with an image of the captured test, to a central database if and when a network connection is available.

4.3.2 *Field Evaluation*

To explore the impact that using our mobile camera-based system had on health workers' patient care routines we conducted an eight-week field study with sixty users at five hospitals and clinics in Zimbabwe. Although we have extensively tested the system with rapid diagnostic tests for multiple diseases in the laboratory, including tests for HIV and syphilis [33], our field evaluation in Zimbabwe targeted only diagnostic tests for malaria. The practical constraints of integrating a new system into government public health facilities limited our deployment to diagnostic tests that are already routinely purchased and distributed by the Ministry of Health and Child Care. Rapid diagnostic tests for malaria are available at all hospitals and clinics in Zimbabwe and are used to test and treat patients in a variety of clinical departments, including the maternal and child health, outpatient, opportunistic infections and maternity departments. Health workers in Zimbabwe have been trained to administer rapid diagnostic tests for malaria and use them on a daily basis, and the Ministry of Health allowed us to analyze these tests. We acknowledge that only evaluating the mobile system with malaria tests is a limitation of our study. However, since the process of administering rapid diagnostic tests for other diseases is the same as administering tests for malaria, the workflow to analyze tests for other diseases would be identical.

Research Questions

This study was the first time that the system has been evaluated with health workers in low-resource settings. We formalize our study into three research questions:

Q1: To what extent are health workers willing and able to integrate a mobile camera-based system into their daily patient care routine?

Although many health workers in Zimbabwe own basic mobile phones, the majority have never used a touchscreen device. We wanted to assess how easily health workers learned to interact with these new devices, how easily groups of health workers within a clinic were able to use a single device as a shared resource, and how system usage might vary across health facilities of different sizes. Finally, we wanted to explore the range of user errors that occurred when health workers used the system in the field.

Q2: What infrastructure challenges are faced by health facilities using the system and what strategies might be used to overcome these challenges?

The rapid growth of mobile networks in developing regions has led to the widespread deployment of 2G/3G networks. However, data speeds are usually erratic and availability can be intermittent. We wanted to examine how connectivity varied at different sites and see if the sites possessed sufficient bandwidth to transmit collected data and images. In addition, we were interested in what strategies were used to keep the devices safe and charged.

Q3: To what extent do the automatically processed system diagnoses agree with the visual diagnoses made by trained health workers?

One of our long-term goals is for the system to help health workers who may not have much medical background to correctly diagnose patients. However, since this study is the first field evaluation of the system, we did not want to affect the standard of patient care at this stage. Thus, health workers who participated in our study were trained nurses whose interpretation of tests is currently used to guide patient care. All patient treatment was based on their visual diagnoses collected prior to image processing. This enabled a blinded comparison of the visual diagnoses and the system diagnoses to assess if the system might be used to direct patient care in the future.

Study Sites

Following approval of the study protocols by the IRB, we identified five clinical study sites in the Manicaland province: the Mutare provincial hospital, the Hauna and Nyanga district hospitals, and the Tombo and Zindi rural health centers. We selected sites that typically have a high prevalence of malaria testing and deliberately targeted facilities that ranged in size from large urban hospitals to small rural clinics to explore how system usage varied at different levels of the healthcare hierarchy. Mutare hospital, located in Zimbabwe's third largest city, is the biggest hospital in Manicaland province. Nyanga district hospital is in a town roughly 100km from Mutare. Hauna hospital is located in a more rural and difficult to access valley roughly 100km from Nyanga. Finally, the Tombo and Zindi clinics are in small, rural villages roughly 80km from Nyanga and 60km from Hauna respectively.

Site	Participants	Devices
Mutare hospital	27	4
Nyanga hospital	17	4
Hauna hospital	11	4
Zindi clinic	3	2
Tombo clinic	2	1
All sites	60	15

Table 4.2: Summary of study sites, participants and devices.

Participants

We recruited 60 health workers (47 female) across the five study sites (see Table 4.2). Participation was limited to trained health workers who conduct malaria diagnostic tests regularly as part of their job. Participants ranged in age from 24 to 64 years ($M = 35$) and had between 1 and 42 years ($M = 6$) of experience employed as health workers. Most participants owned basic mobile phones, but 32 had never used a touchscreen device and another 15 had less than six months experience using a touchscreen device.

Apparatus

We configured our software to collect patient data and capture diagnostic tests for malaria. Since our study deals with sensitive personal health information, all the data we collected was anonymized. We recorded the patient’s gender, age, location, test type, date, time, the health worker’s visual diagnosis, and the system’s computed diagnosis. The software was also instrumented to record the times at which the test started and ended and the time at which the captured data was transmitted to the server.

We deployed the mobile system on Samsung Galaxy XCover 2 Android devices because they were readily available and moderately priced in Zimbabwe. The XCover 2 has a 1 GHz dual core processor, a 5 mega-pixel built-in camera, a 4-inch capacitive touchscreen and a ruggedized plastic cover to protect the device from dust and moisture. Each device was loaded with a 2GB external SD card and a local SIM card with a 300MB prepaid data bundle. As shown in Table 1, the provincial and district hospitals received four devices each,

with one device placed in each of the clinical departments that perform the most malaria diagnostic tests (the maternal and child health department, the outpatient department, the maternity department and the opportunistic infections department). Since rural clinics typically employ fewer health workers than hospitals, each clinic was issued with one or two devices to cover their patient load. In total, we deployed 15 devices across five study sites. Each device was accompanied by a simple plastic stand designed to provide stability and a fixed focal length for the device's camera. In addition, we distributed user manuals to provide support and guidance for users in the absence of the researchers.

Procedure

Preliminary Site Assessment. Before beginning the study, we visited Zimbabwe to assess the selected study sites, understand the health workers' current routines, and gain insight into how a mobile system might be successfully integrated into the clinical workflow. We also tested the network connectivity and data transmission rates at each site.

Adding Diagnostic Tests to the System. During our site assessment, we obtained the details of the three brands of malaria tests currently in use at the sites. Following this, a clinical expert at Global Solutions for Infectious Diseases (San Francisco, USA) [59] created the test description files to add the malaria tests to the system. Encouragingly, creating new test description files was relatively quick, taking roughly 30 minutes per test. Rigorously testing the system parameters for each test took longer, typically several days, since it was necessary to prepare and administer series' of tests with differing dilutions of malaria infected blood under a variety of lighting conditions, before empirically determining the parameters that resulted in the optimal sensitivity and specificity. To make the optimization phase easier, we created a version of the system that could batch process a variety of different test images and parameters at once, which allowed the clinical expert to more quickly narrow in on the optimal algorithm parameters for each test.

Training and Deployment. After adding the malaria tests to the system, we returned to Zimbabwe and conducted participant training sessions at each study site. Each training

session lasted approximately 60 minutes and began by demonstrating the system to a group of participants. Participants were then divided into pairs and given time to read user manuals, ask questions and practice capturing and reporting diagnostic test data. The amount of practice that participants required varied depending on their familiarity with technology, although all participants mastered the system within the training session. During each session, we explained that this was the first time the mobile system was being field tested and that all patient care should continue to be based on participants' visual diagnoses. At the end of each session we collected demographic data regarding each participant's age, work experience and familiarity with technology.

After the training sessions, participants were asked to integrate the system into their daily patient care routine and to capture and transmit data every time that they performed a malaria test. At the hospitals, we explained that each of the main clinical departments that conducted malaria tests would receive one device and that participants working in these departments would share the device accordingly. We conducted follow-up visits at each site one week and six weeks after the training sessions to assess how easily participants were able to integrate the technology into their workflow. During these visits, we observed participants using the system and conducted semi-structured interviews to collect data about participants' opinions and experiences.

4.3.3 Results and Discussion

Q1: Integrating the System into the Clinical Workflow

Our first research question explores the extent to which participants were able to integrate the system into their clinical workflow, including how they used the system and what errors they made when interacting with the system.

Analyzing System Usage and Data Collection.

Health workers reacted positively to the introduction of the system into their clinical workflow. Most participants were eager to learn the new technology and excited that the system represented progress for Zimbabwe's national health system. Figure 4.9 shows the number of tests captured by each site. In total, participants captured and transmitted 1828 malaria

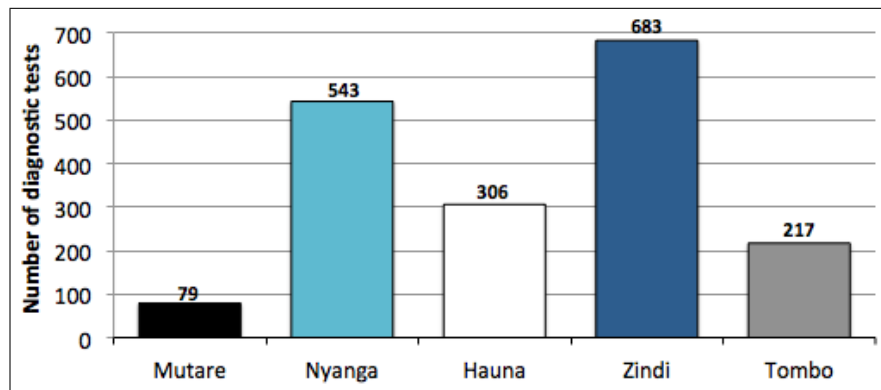


Figure 4.9: Number of malaria tests captured by each study site.

tests over the two-month study period. The median time that it took to capture and process a test and record patient data was 1.5 minutes (ranging from an average of 2.5 minutes at Mutare to 1.3 minutes at Nyanga).

In addition, a large proportion of participants immediately identified the potential for the system to reduce their paperwork burden by automatically generating aggregated monthly reports. Participants at the small clinics told us that they were currently responsible for manually maintaining a large number of paper-based patient registers, several of which required the same data to be repeated and most of which required a written monthly report. They were eager to show us which registers could be replaced by collecting data using the mobile system and they requested that the system be extended to enable them to collect a broader range of health information beyond diagnostic test data.

Figure 4.10 shows the proportion of each site's captured tests that were collected during each week of the study. Encouragingly, participants used the system regularly to collect data for the duration of the study, which suggests that they were generally able to integrate the system into their daily clinical routines. In addition, several participants mentioned that patients responded favorably to the new system, believing that it was important for their diagnostic test data to be reported directly to the Ministry of Health.

Somewhat surprisingly Zindi, one of the small rural clinics, captured the largest number of tests (see Figure 4.9). This could be explained by several factors. First, Zindi is located

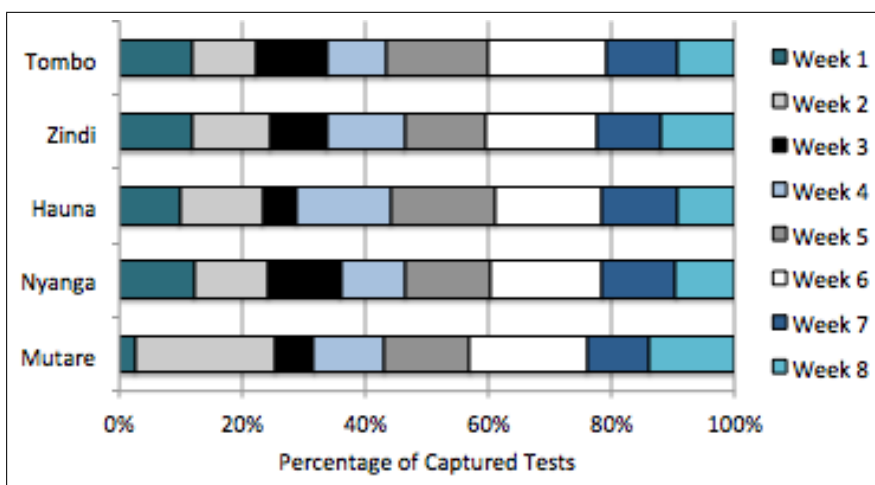


Figure 4.10: Proportion of each site’s total tests collected each week of the study.

in the most malaria endemic region of all the study sites and typically performs a large number of diagnostic tests during the malaria season. In addition, small clinics are usually a patient’s first point of contact with the health system and represent an ideal place at which to quickly diagnose and treat malaria. Finally, we found that the health workers at Zindi set up a dedicated desk that they used exclusively for performing malaria tests. Placing the mobile device on the same desk made it easy for them to integrate the system into their current workflow and capture tests as necessary.

By contrast, Mutare, the largest hospital in the study, captured the fewest tests (see Figure 4.9). There are several possible explanations for this. First, since visiting the hospital is substantially more expensive than visiting a clinic, many patients only visit the hospital if they are referred there by a clinic. In these cases, patients often already received a diagnostic test at the clinic and there is no need to perform another test at the hospital. In addition, the clinical departments at Mutare hospital are much larger than those at the district hospitals and have many patient exam rooms, and it was challenging for participants to coordinate usage of a single device across multiple rooms. Several participants explained that sometimes when they needed to use the device, it was already being used in another room. Thus, for large hospitals, a better model may be to provide one device per room.

We also discovered that the large number of patients at the hospital increased participants' fear that the devices would get stolen if left in a room unattended. As a result, participants frequently locked the device in a cabinet when they left the room. Other participants were then unable to access the device until they found the health worker with the key. During a follow-up visit, we explained to participants that it would be better to use the system and risk theft of the device than keep it locked away. This encouragement may have been responsible for the larger number of tests collected in Mutare in the second week of the study (see Figure 4.10) although the number of tests decreased again in subsequent weeks. At the end of the study several participants suggested that it would be better for each health worker to have a device and carry it around the hospital. However, this would substantially increase the number of devices and overall cost of deploying the system.

Analyzing User Errors.

To analyze the errors that participants made when they interacted with the system, we manually examined all of the captured images and coded any anomalies. User errors fell into five categories: (1) no image captured - users collected patient data but failed to capture an image of a test; (2) captured image is not of a diagnostic test - in these cases the image was usually of the desk or stand and probably captured accidentally; (3) incorrect test selected - the type of test shown in the image is different to the test type selected by the user; (4) test placed upside-down - the test in the image was placed the wrong way up; and (5) unusable image - the captured image is not usable for analysis because it is out of focus, overexposed or the device was not positioned in the stand correctly.

Figure 4.11 shows the percentage of each site's tests that contained each type of error. Overall, we identified errors in 114 out of 1828 tests (6.2%). The most common error was placing the test upside-down (53/1828 or 2.9%), followed by incorrectly selecting the test type (31/1828 or 1.7%). The other error types were more infrequent, with users failing to capture an image in 11/1828 or 0.6%, capturing an image of something other than a test in 6/1828 or 0.3%, and placing the device incorrectly in 13/1828 or 0.7%. Zindi had the lowest error rate at 2.2%, while Tombo had the highest error rate at 11.5%. One reason for the error rate at Tombo was one participant repeatedly positioning the device incorrectly.

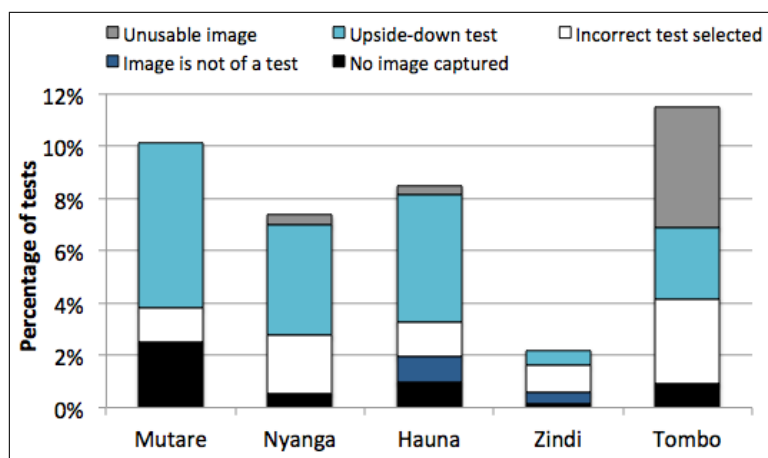


Figure 4.11: Percentage of each site’s tests that contained each type of user error.

We could address several error types through additions to the algorithm. We have already made it impossible for users to submit data without capturing an image. In addition, for tests that contain any identifying markings (such as a brand name), it would be relatively straightforward to automatically detect the test type and/or if the test is upside-down. Finally, additional or repeated training sessions would likely also decrease user errors.

Using the System for Quality Control.

Our findings also show that the system has the potential to be a valuable quality control tool. Early in the study, our analysis revealed a series of transmitted images in which the tests had been incorrectly administered since the test strip had a dark red background (see Figure 4.12). This could be the result of health workers either using too much blood on the test or of reading the test result too early. In either case, the dark red background could easily obscure a weak positive result and the test should be discarded. Instead, health workers were incorrectly reporting these tests as valid, usually negative, diagnoses. This problem occurred in a total of 58 tests (3.2%), of which 34 were from Tombo clinic. Based on this data, we alerted the relevant supervisor who reviewed the test manufacturer’s guidelines with the health workers. In the future, it will be relatively simple to automatically identify tests that have a dark red background and instruct the user to run another test.



Figure 4.12: A test that contains too much red background to yield a valid diagnosis.

Q2: Infrastructure Challenges

Our second research question explores the infrastructure challenges faced by participants using the system. None of the devices were broken, lost, or stolen during the study. In addition, participants at each study site were able to keep the devices sufficiently charged despite experiencing frequent electricity outages, which confirms the benefits of deploying portable, battery-charged devices.

We observed substantial variations in network connectivity between the study sites. Mutare and Nyanga had fairly reliable 3G connections, Hauna and Zindi had much slower, unreliable 2G connections, and although Tombo appeared to have a 2G connection, the distance from the clinic to the cell tower made transmitting any data extremely challenging. Thus, although we initially planned to transmit relatively high quality images (roughly 300KB), we decided to instead transmit low-quality, compressed images, which decreased the amount of data to about 70KB per test.

Our findings show several interesting differences in data transmission between the sites. Figure 4.13 shows the proportion of each site's total tests that were transmitted during each week of the study. Mutare and Nyanga were able to reliably transmit data to the server for the duration of the study. By contrast, Tombo was unable to transmit any data for the first two weeks of the study. To overcome this issue, a health worker from Tombo traveled by bus every few weeks to a town roughly 40 minutes away to transmit data. This behavior is illustrated by the high percentages of data transmitted by Tombo in weeks 3 and 6 of the study (see Figure 4.13). Zindi was also unable to transmit data for the first few weeks of the study. We are unsure what caused the changes in connectivity that enabled Zindi to transmit large amounts of data in weeks 6-8 of the study, although we hypothesize that the

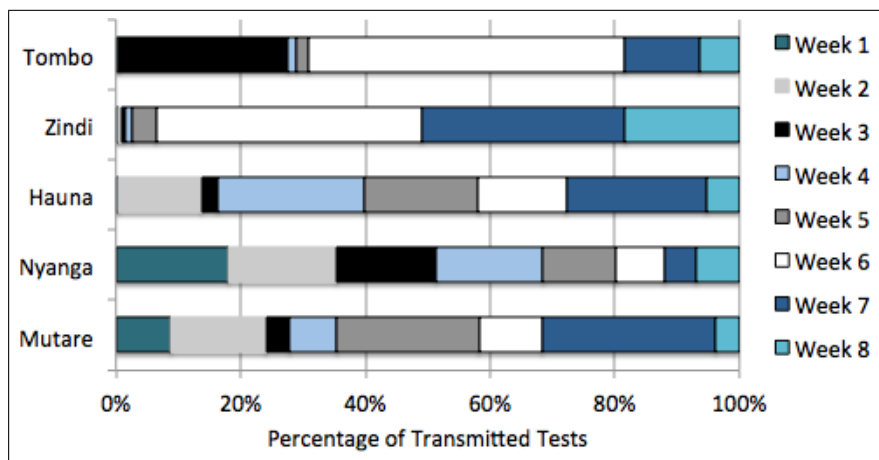


Figure 4.13: Proportion of each site's tests transmitted during each week of the study.

network provider installed a new cell tower or signal-boosting hardware in the area. Finally, health workers at Hauna also struggled initially to transmit data. However, after noticing that connectivity seemed to be more reliable at night, health workers began leaving the devices on to transmit overnight. In total, Hauna transmitted 34% of their tests between 10pm and 4am, compared to 4% by Mutare, 2% by Tombo and 0% by Zindi and Nyanga during the same time frame.

Our findings also show large variations between sites in the delays between capturing tests and transmitting them to the server. Mutare and Nyanga experienced the smallest delays in transmission, with geometric means of 33 seconds ($SD = 23$ hrs) and 45 seconds ($SD = 17$ hrs) respectively, compared to mean delays of 2 hrs ($SD = 129$ hrs) at Hauna, 2 hrs ($SD = 250$ hrs) at Zindi and 6 hrs ($SD = 140$ hrs) at Tombo.

These findings can inform the design of other mobile health systems. Many systems that target low-resource settings face challenges relating to network connectivity, and our study highlights the importance of supporting asynchronous data transmission. Depending on the site, it took anywhere from a few seconds to a few weeks for a captured test to be transmitted to the server, which also highlights the need for mobile systems to process data locally rather than transmit it for off-site analysis. Furthermore, several of the strategies that our participants developed for transmitting data could be generalized to other contexts.

		A System		
		Positive	Negative	Kappa (95% CI)
Visual	Positive	345	26	0.84 (0.81-0.87)
	Negative	70	1273	

		B Review		
		Positive	Negative	Kappa (95% CI)
Visual	Positive	352	19	0.93 (0.91-0.95)
	Negative	22	1321	

		C Review		
		Positive	Negative	Kappa (95% CI)
System	Positive	367	50	0.91 (0.88-0.93)
	Negative	7	1290	

Figure 4.14: Summary of agreement between (A) the system and visual diagnoses, (B) the review and visual diagnoses and (C) the review and system diagnoses.

Q3: Comparison of System and Human Diagnoses

Our final research question explores the extent to which the system’s diagnoses agreed with the health workers’ diagnoses. Although the Ministry of Health informed us prior to the study that the sites used three brands of malaria tests, we found that in practice all the tests they performed were of a single brand (Paracheck Pf). Our analysis of agreement did not include tests that contained user errors (shown in Figure 4.11). Thus, our dataset consisted of 1714 Paracheck tests. As shown in Figure 4.14(A), the system and visual diagnoses agreed in 1618 tests (94.3%). We used Cohen’s Kappa [21] to compute a statistical level of agreement between the two diagnoses. The measured Cohen’s Kappa for these results was 0.84 (95% CI:[0.81,0.87]), indicating strong agreement.

Further analysis revealed several sources of disagreement. In some cases, the correct diagnosis was clear and either the system or the health worker made an obvious error. Unfortunately, we cannot determine if visual errors were the result of data entry mistakes or if the participants in fact diagnosed a patient incorrectly. In addition, a large portion of tests (430 or 25%) contained at least some amount of red background coloring, which caused a substantial number of the errors made by the system (often these were false positive results). In testing the system prior to the study, we always used the correct amount of

blood and waited the right amount of time before reading the result. Thus, we did not test the system on images with colored backgrounds. Further adjustments to the algorithm may be required to deal with this issue. Finally, there were tests with equivocal results at the limit of detection. In these cases, either the system or the health worker reported a positive result. However, since our study did not include microscopy-based analysis of blood samples by a clinician (considered the gold standard of malaria diagnostics), we do not know which diagnosis was correct.

To further analyze differences between the system and visual diagnoses, we created a review diagnosis for each test. We asked three researchers to examine each test image and provide a diagnosis, and we recorded their majority opinion as the review diagnosis. As shown in Figure 4.14(B), the review and visual diagnoses matched in 1673 tests (97.6%). The measured Cohen's Kappa for these results was 0.93 (95% CI: [0.91, 0.95]), indicating a stronger agreement than observed between the system and visual diagnoses (Figure 4.14(A)). Finally, the review and system diagnoses agreed in 1657 tests (96.7%) (see Figure 4.14(C)). The measured Cohen's Kappa for these results was 0.91 (95% CI: [0.88, 0.93]), which is only slightly lower than observed between the review and visual diagnoses (Figure 4.14(B)).

The strength of our findings is highly encouraging and will could support several future deployment scenarios. For example, if health workers are well trained, the system could simply focus quality control efforts on tests in which the human diagnosis differs from the system diagnosis. Alternatively, in situations where health workers are less experienced, the system could provide a second opinion and instruct health workers to run another test if there is disagreement. Finally, in situations where health workers are untrained, the system could perform the diagnosis and inform the health worker of the outcome.

Deployment Costs

The cost of deploying and sustaining a new technology is an important consideration in low-resource settings. The XCover devices that we deployed cost roughly USD \$250 each in Zimbabwe. The devices were loaded with 2GB SD cards that cost \$5 each and 300MB data bundles that cost \$15 each. The plastic stands to hold the devices cost \$10 each.

Transmitting a single test required roughly 70KB of data, which equates 0.35 cents per test. Storing data from all 15 devices in an online database cost less than \$10 per month. Health workers were trained during regular work hours and did not receive additional compensation for participating in the study. Thus, the entire deployment cost approximately \$4200, or \$280 per device. The travel, transport and lodging costs for four researchers and two ministry of health staff were 4 to 5 times more than the costs of the deployment.

Limitations

Our study has several limitations. We only analyze tests for malaria and would like to field test the system with tests for a variety of other diseases. In addition, participants were well trained and we want to explore the potential for the system to aid workers that have less medical background. Our system also uses a stand to hold the device in position above the test. Systems that require users to capture images using a handheld device may be more convenient to use, but will likely experience additional image capture issues. Finally, we focus on a camera-based mobile system. Although some of our findings apply to sensor-based systems in general (such as data transmission issues), future research could expose additional challenges for systems that use external sensors attached to the device.

4.3.4 Conclusion

This section describes an eight-week field deployment of our mobile system for capturing and analyzing rapid diagnostic tests. Health workers in Zimbabwe used the system for the duration of the study to analyze thousands of tests for malaria. In addition, they employed a variety of strategies to overcome poor network connectivity and transmit test data to a server. Finally, the system's computed diagnoses strongly agreed with the visual diagnoses provided by trained health workers. Taken together, our findings highlight the potential for mobile systems like ours to aid the delivery of healthcare in low-resource settings and provide valuable insights for researchers and practitioners working in similar environments.

4.4 Automated Analysis of Time-Sensitive Diagnostic Test Data

The previous sections in this chapter describe the design, implementation, and field evaluation of a mobile system that automatically interprets rapid diagnostic tests for infectious diseases. We chose to analyze rapid diagnostic tests in part because they are already commercially available and widely-used all over the world, and so we would be able to deploy the system with real health workers and evaluate how well it fits into clinical workflows. However, the majority of rapid diagnostic tests have qualitative results (positive or negative) that are relatively simple to interpret, and bioengineers are in the process of developing more sophisticated tests whose results will require quantification or time-sensitive analysis that will be extremely difficult or impossible to read by eye [130]. We wanted to investigate how easily our mobile system could be extended to interpret these more sophisticated tests, thereby ensuring that a platform capable of interpreting the tests already exists if and when the tests become commercially available for use by health workers in the field.

This section describes our work designing and building a mobile, camera-based system that automatically quantifies time-sensitive diagnostic test data using image processing performed entirely on the device. We show that the mobile device is capable of processing test data rapidly enough to deliver a diagnosis within a single patient visit to the clinic and confirm the potential for a mobile camera-based system to accurately interpret sophisticated (currently experimental) tests that would be difficult or impossible to interpret by eye.

4.4.1 Background and Prior Work

Our methods for quantifying time-sensitive diagnostic test data build on prior work by Stevens et al. [130] at the University of Washington who developed a low-cost, paper-based microfluidic flow-through membrane immunoassay (FMIA) diagnostic test. On-card rehydration of reagents enables unrefrigerated storage of the test and provides several months of reagent stability. The test captures the progress of a reaction that detects a malarial biomarker. The presence of the biomarker causes the reagents on the test card to change the color of a capture spot, and the rate and degree of color change indicate the quantity of biomarker present in the sample.

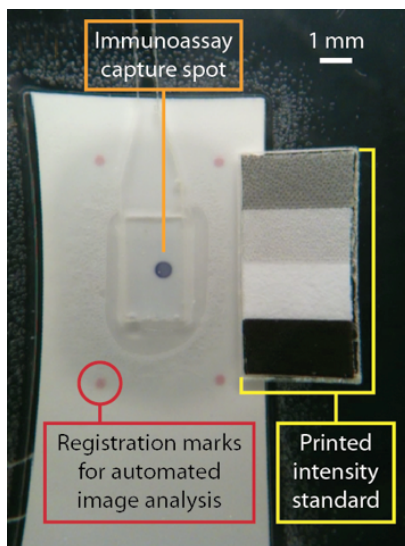


Figure 4.15: Device setup showing the location and scale of the on-card intensity standards, registration marks and immunoassay capture spot.

To measure the progress of the test, a number of features were added to the test card that facilitate video capture by a mobile device. The small flow-through area of the test and smaller dimensions of the capture spots (less than 1 mm in diameter) require magnification to produce images of sufficient resolution for quantification. The magnification is provided using an off-the-shelf 15x simple lens (approximately \$30). Additionally, an intensity standard printed with a conventional inkjet printer is attached to the card, which allows for normalized quantification of signals under variable lighting conditions. Finally, the test card is patterned with registration marks that may be used to locate the capture spot. The on-card intensity standard and registration marks, depicted in figure 4.15, simplify the task of automatically locating and processing the test result.

To record the progress of the test, Stevens et al. [130] captured video frames of the test using an Apple iPhone 3GS with a resolution of 480x640 pixels per frame. Quantification of the test results was performed by determining the intensity of the capture spot using pixel color intensities that are normalized to on-card intensity standards. The locations and intensities of the standards, registration marks and capture spots were identified using image processing MATLAB software written specifically for this task. The iPhone 3GS was

used solely as an image and video capture device, and while this proved that the phone was capable of capturing images of sufficient quality, none of the image processing was performed on the device.

4.4.2 Method

The main contribution presented in this section is the creation of software capable of automatically quantifying time-sensitive test data using image processing performed on a mobile device. We now discuss the system architecture, the technologies used in the development of the system, the image processing algorithms implemented on the device to quantify recorded test signals, and the user interface that facilitates selection of tests for processing and displays the results. The methods described here closely follow those presented in Steven's et al.'s prior work, but leverage a number of existing optimized computer vision algorithms such as Canny edge detection [14] and contour-finding [134].

Architecture

The software for quantifying test data was developed on the Android platform. We chose Android because of the variety of different devices that run that operating system including a variety of mobile phones and tablets, as well as the open-source nature of the development environment. Figure 4.16 shows the system architecture. The user interface and visual display of the quantified test signals are implemented in Java using the standard Android Software Development Kit (SDK). To facilitate the use of optimized computer vision algorithms, the image processing components of our system make use of OpenCV, an open source computer vision library [104]. Since OpenCV is written in native code, we used the Android Native Development Kit (NDK) to compile the OpenCV library and our native code for inclusion in the system. We then use the Java Native Interface (JNI) to access the native code. This allows us to take advantage of the convenience of the Android framework for the user interface and graphical display, but still make use of the optimized computer vision algorithms provided by OpenCV.

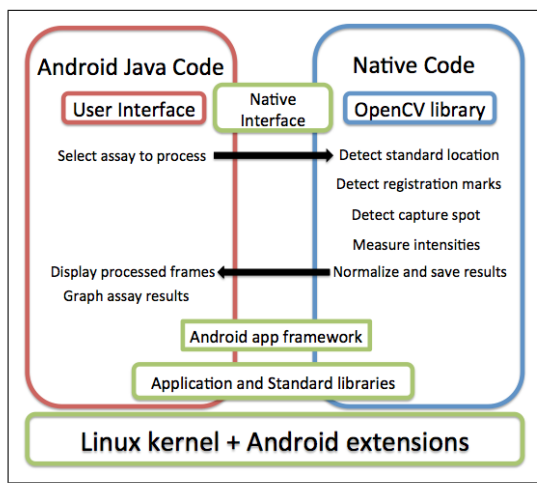


Figure 4.16: System architecture and workflow.

Image Processing

We now describe the algorithm used to automatically quantify video frames of a test. Processing every frame was found to be unnecessary since accurate quantification may be achieved using a small fraction of the total number of frames. Instead, frames were sampled at evenly-spaced intervals for the duration of the video. For each sampled frame, there were four main steps in the image processing algorithm: locating the on-card standards, locating the registration marks, locating the capture spot, and measuring the pixel intensities of the standards and capture spot.

Detecting the Intensity Standard. The first step in processing the test is the identification of the on-card intensity standard. This determines the location of the standard measurement areas, and provides a means of estimating the location of the registration marks. A long, vertical strip through a portion of the image likely to contain the intensity standards is defined, and Canny edge detection [14] used to identify the horizontal edges of the standard. A similar process is used to identify the left and right edges of the standard. The identified image region is then subdivided into the four color standards, which are easily labeled as white, light gray, dark gray and black using simple color analysis. The measurement regions, used to calculate the intensity of the standard during the quantifi-

cation process, are defined using a rectangular region that is slightly smaller than each intensity standard.

Detecting the Registration Marks. The next step in processing the test is to identify the locations of the registration marks. The test card design places the marks at known distances from the capture spot, and at specific locations relative to the intensity standard. The location of the standard, determined by the previous step, is used to estimate the four regions of the image that are likely to contain a registration mark. This is advantageous since it decreases the search space that must be considered to locate the registration marks. Within these estimated regions, the registration marks are red in color and highly saturated. This allows them to be segmented from the background image by thresholding on the hue, saturation, and value channels. The thresholded channels are then combined to create a binary map of the image region, and contour detection performed to identify shapes that are likely to be the registration mark. The moments of each detected contour are used to calculate the area and centroid of the contour, and the registration mark is identified as the contour whose area falls within a known range. Finally, the centroid of the identified contour is taken to be the location of the registration mark.

Detecting the Capture Spot. The capture spot is located at a known distance from the registration marks and is found as follows. First, half the length of the diagonal between two opposite registration marks is calculated. This length is used as the radius of a circle that is drawn around each registration mark, with the center of the circle corresponding to the centroid of the registration mark. The intersection of the four circles is then calculated, and the centroid of the intersecting area is taken as the location of the capture spot. The measurement area is defined as a circle of pixels within a fixed radius from the calculated center. The radius of the measurement area is chosen to be slightly smaller than the capture spot to avoid the possibility of taking measurements at the edges of the spot, which may incorrectly influence the intensity value. Our approach has the benefit that, if a registration mark fails to be detected, the capture spot location may still be correctly identified by the other three registration marks.

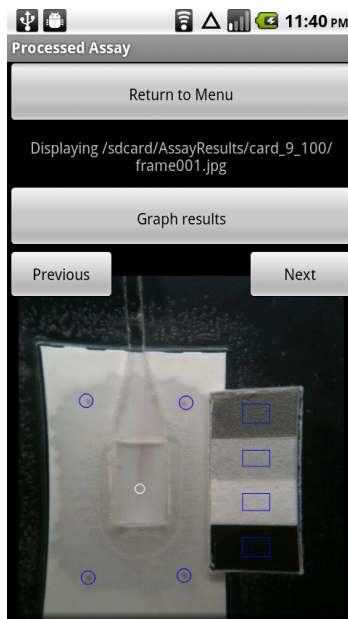


Figure 4.17: Screenshot of a processed video frame.

Quantifying the Test Signal. The final part of the image processing is to measure and record the results of the biochemical reaction. The mean pixel color intensity of each standard region (white, light gray, dark gray and black) is calculated, as well as the mean pixel color intensity for the capture spot area. The intensity of the capture spot is then normalized linearly to the white and dark gray standards, with the white standard corresponding to 0 and the dark gray standard corresponding to 1.

An image of each processed frame, as well as a file containing the locations of the standards, registration marks and capture spot, and both the original and normalized intensity values, are created and saved on the SD card of the phone. The results of the test may then be displayed to the user, sent away for further analysis or incorporated into data collection systems such as an electronic medical record.

User Interface

The user interface component of the system has been built using the standard Android framework. When the system starts up, the user is presented with a list of image stacks

from tests stored on the phone, and may choose a specific test to process. While the selected test is being processed by the underlying native code, the user is presented with a “processing” icon to ensure that the user knows the system is busy and should not be interrupted. When processing is complete, the first frame of the processed image stack is displayed on the phone’s screen, shown in Figure 4.17, and the user can check the detected locations of the standards, registration marks and capture spot. The user can also step through the processed test on a frame-by-frame basis to observe changes in capture spot intensity. Alternatively, a “graph results” button is provided that, when clicked, displays a graph showing changes in the normalized test signal and standard intensities for the duration of the test.

4.4.3 Evaluation

We evaluated the performance of the mobile system by comparing the results of processing to those of the previous MATLAB software [130]. The MATLAB implementation was run on an Apple MacBook Pro with a 2.53 GHz dual core processor and 4GB RAM, while the mobile system was run on two different Android devices, a Motorola Droid and an HTC Nexus One, both running Android 2.2.1. The Droid has a 550MHz processor, 256 MB RAM, a 16GB SD card and a 5.0MP camera, while the Nexus One has a 1GHz processor, 512 MB RAM, a 4GB SD card and a 5.0MP camera.

Data Set

To compare the image processing on the Android phone with the MATLAB software on a computer, we analyzed the same data set used by Stevens et al. [130]. At this stage we have not captured new data using the camera on the Android phone, but have rather processed the existing video data captured previously using an Apple iPhone 3GS.

The data captured was from an immunoassay test that models a test used to test for malaria, dengue, measles, and rickettsia [153]. In these tests, the presence of certain antibodies in the sample causes the capture spot to turn red, and then to become darker and bluer with the addition of a gold-enhancement solution. The darkness of the capture spot

increases with increasing concentration of antibodies in the sample. The changing color of the capture spot is recorded over time, and the rate at which this change occurs is calculated and used to quantify the test signal.

Measuring the rate of change in test signals over time is a commonly accepted improvement over single endpoint measurements for several reasons [137]. Firstly, since the concentration of analyte present in a given sample is unknown, it can be difficult to determine the appropriate endpoint at which to record the result. Measuring the result too early may prevent low-concentration samples, whose signals have not yet risen to the appropriate level, from being accurately recorded. However, measuring the result too late can cause inaccuracies due to a variety of factors, such as a build up of reaction products or depletion of substrate. Additionally, automatically detecting and identifying predictable events during the course of a test can provide timing cues for operation and analysis, such as when to record a measurement [79]. Finally, kinetic data allows user errors to be identified early and the resulting test data discarded.

The data set we analyze consisted of 20 videos, each of which recorded the progress of a single test. The videos ranged in length from 13:14 minutes to 13:52 minutes and were captured at 30 frames per second with a resolution of 640 by 480 pixels per frame. Processed frames were typically sampled at a rate of one out of every 200 frames evenly spaced throughout the duration of the video, which resulted in processing an average of 122 frames per test.

Accuracy

The mobile system correctly located the intensity standards, registration marks and capture spot on all 20 test videos attempted, and successfully extracted signal intensity measurements from every test. We compared the quantitative measurements obtained by the mobile system, depicted in Figure 4.18(b), to those obtained by the MATLAB implementation, depicted in Figure 4.18(a). The shape of the curve representing the capture spot describes the progress of the test. The initial spike is due to the addition of a colored labeling reagent. This is then washed out, leaving behind the spot intensity. The subsequent increase in dark-

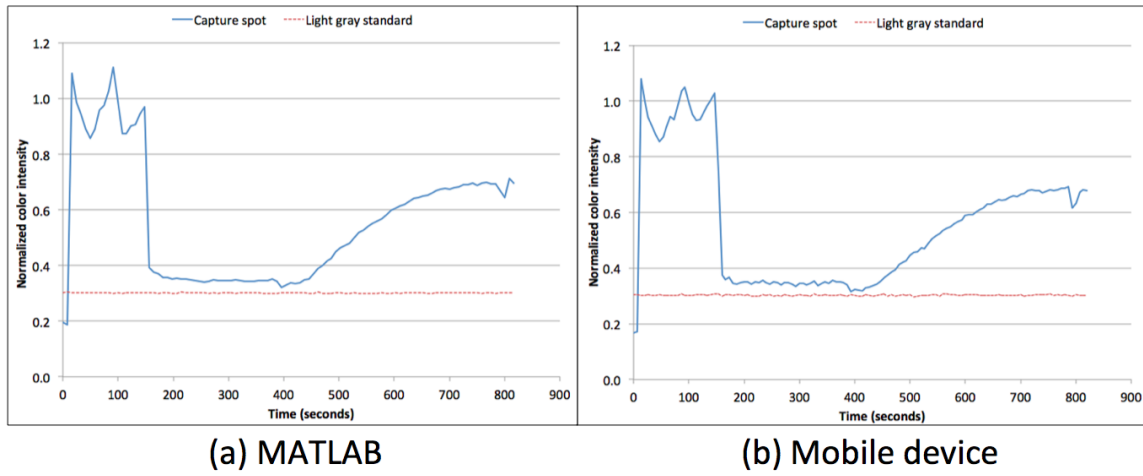


Figure 4.18: Normalized intensities over time using (a) MATLAB and (b) a mobile device.

ness of the spot is due to the addition of gold-enhancement solution. The line representing the intensity of the light gray standard is provided for comparison, and correctly remains constant for the duration of the test. The small differences between the two graphs may be attributed to the different methods used to identify the pixels corresponding to the capture spot area. Additionally, the exact frames that were sampled by each implementation varied slightly. However, the general similarity of the graphs indicates that the test results obtained by the mobile system are comparable to those of the previous method, and that the system is capable of accurately quantifying test results on at least two Android phones.

Processing Rate

We evaluate the rate at which test data is processed with respect to several different factors. First, we compare the rates of processing test data on the Android devices with the computer-based MATLAB software. For these tests, the locations of the standards, registration marks and capture spot were calculated only for the first frame and applied to each subsequent frame. The results of the comparison, depicted in Figure 4.19, show that the Nexus One achieved the fastest processing rate, averaging 5.95 frames per second (fps). The Droid achieved 3.81 fps and the MATLAB algorithm 1.03 fps. The superior performance of the mobile system may be attributed to the efficiency of implementing the image processing

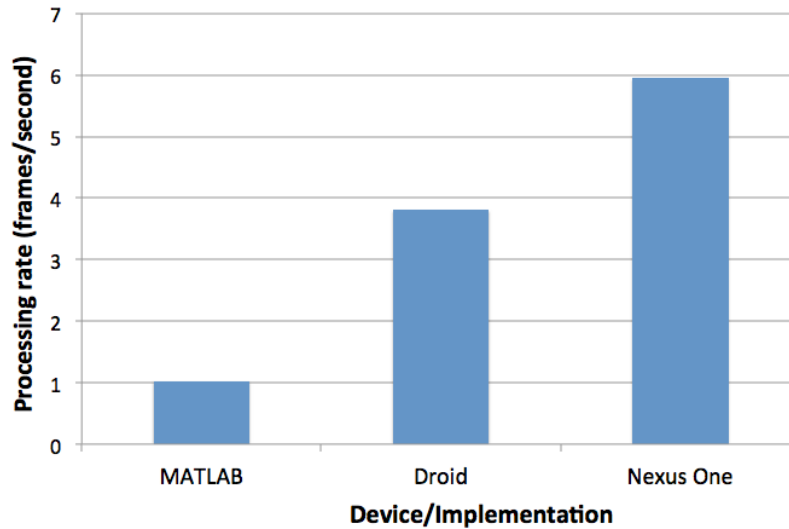


Figure 4.19: Processing rates on each device and in MATLAB.

in native code and indicates that the system would be capable of processing test data fast enough for the results to be available to within a single patient visit to the clinic.

Second, we compared the processing rate on the Nexus One and the Droid when the calculations to determine the location of the capture spot, registration marks and standards were performed every frame, figure 4.20(A), rather than only on the first frame, figure 4.20(B). Performing all of the location calculations for every frame resulted in a processing rate of 0.92 fps on the Nexus One and 0.51 fps on the Droid, whereas calculating the locations for only the first frame resulted in an overall processing rate of 5.95 fps on the Nexus One and 3.81 fps on the Droid. The benefit of performing the location calculations for every frame is that the algorithm may accommodate slight movements of the phone during video capture. However, the length of the test necessitates the use of a mount to hold the phone securely, which makes such movements unlikely. As a result, performing these calculations only once per test is a valuable optimization and yielded signal curves that were visually indistinguishable from those generated by calculating the locations for every frame.

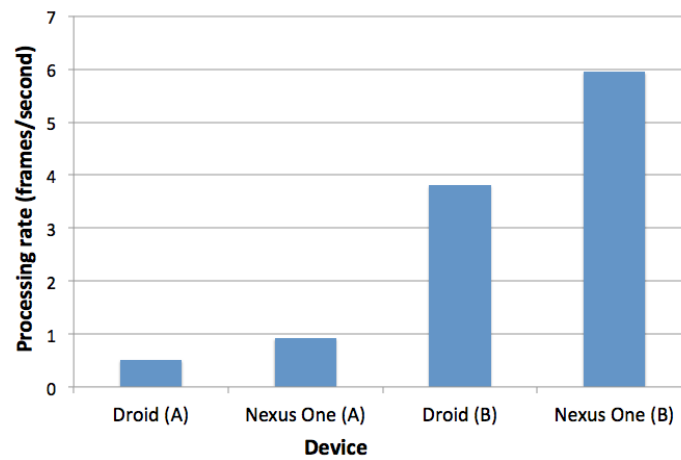


Figure 4.20: Processing rates when calculating locations of the capture spot, registration marks and standards (A) every frame and (B) on the first frame only.

4.4.4 Conclusion

This section describes the design, implementation, and technical evaluation of a mobile system that automatically quantifies time-sensitive diagnostic test data using image processing performed entirely on the device. The speed and accuracy exhibited by the system suggest that it would be capable of delivering accurate diagnostic results immediately to patients and health workers in low resource environments. Although these time-sensitive diagnostic tests are not yet ready for deployment with health workers in the field, our work has demonstrated the potential for a mobile, camera-based system to be used as a viable computing platform for analyzing these sophisticated tests if and when they do become available.

4.5 Summary

Many of the diagnostic tests administered in well-funded clinical laboratories are inappropriate for point-of-care testing in low-resource environments. As a result, inexpensive, portable rapid diagnostic tests have been developed to facilitate the diagnosis of many diseases common to developing countries. However, manually analyzing the test results at the point of care may be complex and error-prone for untrained users reading test results by eye, and providing methods for automatically processing these tests could significantly increase their utility. This chapter describes our research designing, implementing, and evaluating a mobile camera-based system that automatically interprets diagnostic tests for infectious diseases using image processing performed on a mobile device. Specifically, we contribute:

1. The design and implementation of ODK Diagnostics, a mobile camera-based system capable of automatically interpreting a variety of commercially available rapid diagnostic tests using image processing performed entirely on the device [33].
2. An eight-week field deployment of the ODK Diagnostics system with 60 health workers at five hospitals and clinics in Zimbabwe that demonstrates the benefits and challenges of integrating the system into daily clinical workflows [37].
3. The design, implementation, and technical evaluation of a mobile, camera-based system that automatically quantifies time-sensitive diagnostic test data for an experimental microfluidic test that is currently under development [40].

Taken together, these contributions demonstrate the potential for mobile, camera-based systems to assist health workers in low-resource environments with the process of quickly and accurately diagnosing a variety of infectious diseases. Our findings could also inform the design of future systems that aim to support health workers in the field and help to guide ministries of health and other stakeholders working to deploy mobile health systems in similar environments.

Chapter 5

CONCLUSION

Mobile devices have been shown to be an effective platform for collecting data to support a wide range of social, administrative, and health applications in low-resource settings. Many of these applications also allow users to gather rich data from sensors that come built-in to the device, such as taking photos using the device's camera. However, in addition to simply gathering data, there are many potential benefits to be gained by analyzing the data collected from sensors immediately on the device, such as lowering the cost and bandwidth required for data transmission, enabling immediate analysis in the absence of reliable connectivity, and delivering the results of analysis to workers more quickly than waiting for responses from an external server or off-site human expert.

This dissertation describes the design, implementation, and evaluation of mobile systems that run on cheap, commercially available devices (such as smartphones and tablets) and that use computer vision and machine-learning techniques to automate tasks that were previously tedious or error prone. We discussed the rich body of literature on which this dissertation builds in Chapter 2. We then presented our original research that targets two primary problem domains: improving data collection by automatically digitizing data from paper documents (Chapter 3) and improving disease diagnosis by automatically interpreting diagnostic tests for infectious diseases (Chapter 4). We employed a common approach to interpreting camera-based input for both these domains and quantitatively evaluated our algorithms through rigorous laboratory studies. We also spent substantial time working with global development organizations, government ministries, and health workers in the field to ensure that our systems were usable and appropriate under the constraints experienced at all levels of the information hierarchy in low-resource settings.

In addition to the individual contributions described in Chapters 3 and 4, this work also makes the following high-level contributions:

- We show that commercially available mobile devices are capable of capturing high-quality images and videos that can be processed using computer vision and machine-learning techniques running locally on the device.
- We identify and overcome the technical challenges associated with designing and implementing algorithms to interpret camera-based input and quantify the technical performance of these algorithms.
- We develop an approach to building camera-based systems that allows users to specify how images should be interpreted without needing to recompile the system.
- We demonstrate the viability of our approach by applying our methods to two different problem domains: automatically digitizing data from paper forms, and automatically interpreting diagnostic tests.
- We prove that our systems can be effectively integrated into existing information ecosystems in low-resource settings and demonstrate that they are usable and appropriate under the constraints faced by users at all levels of the information hierarchy.

By identifying, understanding, and overcoming both the technical challenges and the human challenges associated with designing and building camera-based systems for low-resource environments, this work provides a new approach that shows how data gathered using the built-in sensors on mobile devices can be analyzed immediately on the device and used to strengthen information and healthcare systems in developing countries.

BIBLIOGRAPHY

- [1] Adobe PhoneGap. <http://phonegap.com/>.
- [2] H Al-Muhtaseb, S Mahmoud, and R Qahwaji. Recognition of off-line printed arabic text using hidden markov models. *Signal Process.*, 88(12):2902–2912, 2008.
- [3] Diagnostics For All. <http://www.dfa.org>.
- [4] Clinitek Status Analyzer. <http://healthcare.siemens.com/point-of-care/urinalysis/clinitek-status-analyzer>.
- [5] Y Anokwa, N Ribeka, T Parikh, G Borriello, and M Were. Design of a Phone-Based Clinical Decision Support System for Resource-Limited Settings. In *International Conference on Information and Communication Technologies and Development (ICTD)*, 2012.
- [6] T Arai, D Aust, and S Hudson. Paperlink: A technique for hyperlinking from real paper to electronic content. In *ACM SIGCHI Conference on Human factors in computing systems*, pages 327–334, 1997.
- [7] H Bay, A Ess, T Tuytelaars, and L Van Gool. SURF: Speeded Up Robust Features. *Computer Vision and Image Understanding*, 110(3), 2008.
- [8] B Birnbaum, G Borriello, A Flaxman, B DeRenzi, and A Karlin. Using behavioral data to identify interviewer fabrication in surveys. In *SIGCHI Conference on Human Factors in Computing Systems, CHI '13*, pages 2911–2920, 2013.
- [9] J Blanchette. A material history of bits. *J. Am. Soc. Inf. Sci. Technol.*, 62(6):1042–1057, June 2011.
- [10] O Bondarenko and R Janssen. Documents at hand: Learning from paper to improve digital technologies. In *SIGCHI Conference on Human Factors in Computing Systems, CHI '05*, pages 121–130, 2005.
- [11] C Bossen and L Jensen. How physicians 'achieve overview': A case-based study in a hospital ward. In *ACM Conference on Computer Supported Cooperative Work & Social Computing, CSCW '14*, pages 257–268, 2014.
- [12] D Breslauer, R Maamari, N Switz, W Lam, and D Fletcher. Mobile phone based clinical microscopy for global health applications. *Plos One*, 4(7), 2009.

- [13] Burnet Institute. <http://www.burnet.edu.au/>.
- [14] J Canny. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.
- [15] Captricity. <http://captricity.com/>.
- [16] R Casey and E Lecolinet. A survey of methods and strategies in character segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(7):690–706, 1996.
- [17] Center for Disease Control. <http://www.cdc.gov/>.
- [18] K Chen, J Hellerstein, and T Parikh. Data in the First Mile. In *Conference on Innovative Data Systems Research (CIDR)*, pages 203–206, 2010.
- [19] K Chen, A Kannan, Y Yano, J Hellerstein, and T Parikh. Shreddr: pipelined paper digitization for low-resource organizations. In *ACM Conference on Computing for Development*, 2012.
- [20] W Clarke, D Cox, L Gonder-Frederick, W Carter, and S Pohl. Evaluating clinical accuracy of systems for self-monitoring of blood glucose. *Diabetes Care*, 10:622–628, 1987.
- [21] J Cohen. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20:37–46, 1960.
- [22] Aurora Company. Automatic elisa test workstation. <http://www.medicalexpo.com/prod/aurora-instruments/automatic-elisa-test-workstations-80210-504874.html>.
- [23] V W Consulting. mHealth for Development: The Opportunity of Mobile Technology for Healthcare in the Developing World. *United Nations Report*, pages 1–70, 2009.
- [24] D Cooper, B Callahan, P Callahan, and L Burnett. Mobile image ratiometry: A new method for instantaneous analysis of rapid test strips. *Nature Precedings*, 2012.
- [25] CORE Group, Save the Children, BASICS, and MCHIP. Community Case Management Essentials: Treating Common Childhood Illnesses in the Community: A Guide for Program Managers. Washington, DC: CORE Group, 2010.
- [26] FIO Corporation. <http://www.fio.com>.
- [27] Scantron Corporation. <https://scantron.com>.
- [28] C Cortes and V Vapnik. Support-Vector Networks. *Machine Learning*, 20(3), 1995.

- [29] B Croal, D Finlay, E Davidson, W Mutch, D Rothnie, A Dickie, and R Newall. Evaluation of the bayer multistix pro point-of-care urine test. *Journal of Near-Patient Testing and Technology*, 2:144–148, 2003.
- [30] B Croal, D Finlay, E Davidson, W Mutch, Duncan S, I Rothnie, A Dickie, and R Newall. Evaluation of the Bayer Multistix PRO 10LS Point-of-Care Urine Test. *Point of Care: The Journal of Near-Patient Testing and Technology*, 2:144–148, 2003.
- [31] D Crockford. Introducing JSON. json.org, 2009.
- [32] S Crowe, D Anderson, A Landay, T Denny, M Garcia, R Lloyd, J Diaz, S van de Waarsenburg, and N Barnes. Development of a Rapid, Point-of-Care Immunochromatographic Test for Measurement of CD4 T-cells. <http://www1.imperial.ac.uk/resources/>.
- [33] N Dell and G Borriello. Mobile Tools for Point-of-Care Diagnostics in the Developing World. In *ACM Conference on Computing for Development*, 2013.
- [34] N Dell, N Breit, T Chaluco, J Crawford, and G Borriello. Digitizing paper forms with mobile imaging technologies. In *ACM Conference on Computing for Development*, 2012.
- [35] N Dell, N Breit, J Wobbrock, and G Borriello. Improving Form-Based Data Entry with Image Snippets. In *ACM Graphics Interface Conference (submitted)*, 2013.
- [36] N Dell, J Crawford, N Breit, T Chaluco, A Coelho, J McCord, and G Borriello. Integrating ODK Scan into the Health Worker Supply Chain in Mozambique. In *International Conference on Information and Communication Technologies and Development, ICTD '13*, 2013.
- [37] N Dell, I Francis, H Sheppard, R Simbi, and G Borriello. Field evaluation of a camera-based mobile health system in low-resource settings. In *Proceedings of the 16th International Conference on Human-computer Interaction with Mobile Devices & Services, MobileHCI '14*, pages 33–42, 2014.
- [38] N Dell, T Perrier, N Kumar, M Lee, R Powers, and G Borriello. Paper-Digital Workflows in Global Development Organizations. In *ACM Conference on Computer Supported Cooperative Work, CSCW '15*, 2015.
- [39] N Dell, V Vaidyanathan, I Medhi, E Cutrell, and W Thies. “Yours is better!” Participant Response Bias in HCI. In *ACM Conference on Human Factors in Computing Systems*, 2012.
- [40] N Dell, S Venkatachalam, D Stevens, P Yager, , and G Borriello. Towards a Point-of-Care Diagnostic System: Automated Analysis of Immunoassay Test Data on a Cell

- Phone. In *NSDR*, 2011.
- [41] B DeRenzi, L Findlater, J Payne, B Birnbaum, J Mangilima, T Parikh, G Borriello, and N Lesh. Improving Community Health Worker Performance Through Automated SMS. In *International Conference on Information and Communication Technologies and Development (ICTD)*, 2012.
- [42] B DeRenzi, N Lesh, T Parikh, C Sims, M Mitchell, W Maokola, M Chemba, Y Hamisi, D Schellenberg, and G Borriello. e-IMCI: Improving Pediatric Health Care in Low-Income Countries. In *ACM SIGCHI Conference on Human factors in computing systems*, 2008.
- [43] District Health Information System. <http://dhis2.org>.
- [44] M. Dymetman and M. Copperman. Intelligent paper. *Electronic Publishing, Artistic Imaging, and Digital Typography*, pages 392–406, 1998.
- [45] V Edupuganti, V Agarwal, and S Kompalli. Registration of camera captured documents under non-rigid deformation. In *Proceedings of the 24th IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 385–392, 2011.
- [46] ExperVision. <http://www.expervision.com/>.
- [47] FIO. AirFio. <http://www.fio.com/airFio.html>.
- [48] FIO. Deki Reader. <http://www.fio.com/howdekiworks.html>.
- [49] FIO. Spiri. <http://www.fio.com/healthmanagers.html>.
- [50] 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.
- [51] G Fitzpatrick. Integrated care and the working record. *Health Informatics Journal*, 10(4):291–302, 2004.
- [52] G Fitzpatrick and G Ellingsen. A review of 25 years of cscw research in healthcare: Contributions, challenges and future agendas. *Computer Supported Cooperative Work (CSCW)*, 22(4-6):609–665, 2013.
- [53] Formhub. <http://formhub.org/>.
- [54] I Francis. Diagnostic Test Reader Landscape Analysis. *Global Solutions for Infectious Diseases, White Paper*, 2011.

- [55] I Francis. Rapid Diagnostic Test Landscape Analysis. *Global Solutions for Infectious Diseases, White Paper*, 2011.
- [56] B Frederick. Fixed-, random-, and mixed-effects ANOVA models: A user-friendly guide for increasing the generalizability of ANOVA results. *Advances in Social Science Methodology*, B. Thompson (ed). Stamford, Connecticut: JAI Press, pages 111–122, 1999.
- [57] R. Gandhi, R. Veeraraghavan, K. Toyama, and V. Ramprasad. Digital green: Participatory video for agricultural extension. In *International Conference on Information and Communication Technologies and Development*, pages 1 –10, 2007.
- [58] W Gaver, A Dunne, and E Pacenti. Design: Cultural probes. *interactions*, 6(1):21–29, January 1999.
- [59] Global Solutions for Infectious Diseases. <http://www.gsid.org/>.
- [60] R Gordon. *Essential JNI*. Prentice Hall, 1998.
- [61] F Guimbretière. Paper augmented digital documents. In *ACM Symposium on User Interface Software and Technology (UIST)*, pages 51–60, 2003.
- [62] A Gupta. *Red Tape: Bureaucracy, Structural Violence, and Poverty in India*. Duke University Press, Durham, NC, 2012.
- [63] R Harper and A Sellen. Paper-supported collaborative work. *Rank Xerox Research Centre, Cambridge Laboratory*, 1995.
- [64] S Hart and L Staveland. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *P. Hancock and N. Meshkati (Eds.), Human mental workload*, pages 139–183, 1988.
- [65] C Hartung, Y Anokwa, W Brunette, A Lerer, C Tseng, and G Borriello. Open Data Kit: Building Information Services for Developing Regions. In *International Conference on Information and Communication Technologies and Development (ICTD)*, 2010.
- [66] S Harvey, L Jennings, M Chinyama, F Masaninga, K Mulholland, and D Bell. Improving community health worker use of malaria rapid diagnostic tests in Zambia: package instructions, job aid and job aid-plus-training. *Malaria Journal*, 7:160, 2008.
- [67] J Heiner, S Hudson, and K Tanaka. Linking and messaging from real paper in the paper pda. In *ACM Symposium on User Interface Software and Technology (UIST)*, pages 179–186, 1999.
- [68] R Holley. How good can it get? analysing and improving ocr accuracy in large scale

- historic newspaper digitization programs. *D-Lib Magazine*, 15(3/4), 2009.
- [69] Y Hong, H Worden, and G Borriello. ODK Tables: Data Organization and Information Services on a Smartphone. In *NSDR*, 2011.
- [70] M Hull. Documents and Bureaucracy. *Annual Review of Anthropology*, 41:251–267, 2012.
- [71] W Johnson, H Jellinek, J Leigh-Klotz, R Rao, and S Card. Bridging the paper and electronic worlds: the paper user interface. In *ACM SIGCHI Conference on Human factors in computing systems*, pages 507–512, 1993.
- [72] A Kay. Tesseract: an open-source optical character recognition engine. *Linux Journal*, 2007.
- [73] E Knebel, S Lundahl, A Raj, H Abdallah, J Ashton, and N Wilson. The use of manual job aids by health care providers: What do we know? *Operations Research Results. Published for USAID by the Quality Assurance Project*, 2000.
- [74] H Koike, Y Sato, and Y Kobayashi. Integrating paper and digital information on enhanceddesk: a method for realtime finger tracking on an augmented desk system. *ACM Transactions on Computer-Human Interaction*, 8(4):307–322, 2001.
- [75] H Kubo, H Ohashi, M Tamamura, T Kowata, and I Kaneko. Shared questionnaire system for school community management. In *Proceedings of the 2004 Symposium on Applications and the Internet-Workshops (SAINT 2004 Workshops)*, pages 439–445, 2004.
- [76] Vasantha L, R Jain, and C Patvardhan. Ocr of printed telugu text with high recognition accuracies. In *Proceedings of the 5th Indian conference on Computer Vision, Graphics and Image Processing*, pages 786–795, 2006.
- [77] E Larson, M Goel, G Boriello, S Heltshe, M Rosenfeld, and S Patel. Spirosmart: Using a microphone to measure lung function on a mobile phone. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing, UbiComp '12*, pages 280–289, 2012.
- [78] J Ledlie. Huzzah for my Thing: Evaluating a Pilot of a Mobile Service in Kenya. *Qual Meets Quant, ICTD*, 2010.
- [79] K Lewandrowski. Three Wishes for POCT Testing: A Compendium of Unmet Needs from the Prespective of Practitioners in the Field. *Point of Care: The Journal of Near-Patient Testing and Technology*, 7(3):86–88, 2008.
- [80] C Liao, F Guimbretière, and K Hinckley. Papiercraft: a command system for inter-

- active paper. In *Proceedings of the 18th annual ACM symposium on User interface software and technology*, pages 241–244, 2005.
- [81] R Littell, P Henry, and C Ammerman. Statistical analysis of repeated measures data using SAS procedures. *Journal of Animal Science*, 76:1216–1231, 1998.
- [82] P Luff, C Heath, and D Greatbatch. Tasks-in-interaction: Paper and screen based documentation in collaborative activity. In *ACM Conference on Computer-supported Cooperative Work, CSCW '92*, pages 163–170, 1992.
- [83] P Luff, C Heath, M Norrie, B Signer, and P Herdman. Only touching the surface: creating affinities between digital content and paper. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 523–532, 2004.
- [84] D Mabey, R Peeling, A Ustianowski, and M Perkins. Tropical infectious diseases: Diagnostics for the developing world. *Nature Reviews Microbiology*, 2:231–240, 2004.
- [85] S MacKenzie and W Soukoreff. Text Entry for Mobile Computing: Models and Methods, Theory and Practice. *Human Computer Interaction*, 17:147–198, 2002.
- [86] J MacQueen. Some Methods for classification and Analysis of Multivariate Observations. *5th Berkeley Symposium on Mathematical Statistics and Probability*, 1:281–297, 1967.
- [87] S Madhvanath, G Manjunath, S Kompalli, S Banerjee, S Ramachandrupa, and S Godavari. PaperWeb: Paper-triggered Web Interactions. In *ACM Conference on Computing for Development*, 2013.
- [88] Magpi. (formerly episurveyor). <http://magpi.com>.
- [89] A Martinez, S Phillips, G Whitesides, and E Carrilho. Diagnostics for the Developing World: Microfluidic Paper-based Analytical Devices. *Anal Chem*, 82:3–10, 2010.
- [90] J Matthews, R Kulkarni, M Gerla, and T Massey. Rapid dengue and outbreak detection with mobile systems and social networks. *Mobile Networks and Applications*, 17(2):178–191, 2012.
- [91] J Matthews, R Kulkarni, G Whitesides, M Sarrafzadeh, M Gerla, and T Massey. A Light-Weight Solution for Real-Time Dengue Detection using Mobile Phones. In *International Conference on Mobile Computing, Applications, and Services (MobiCase)*, 2009.
- [92] I Medhi, M Jain, A Tewari, M Bhavsar, M Matheke-Fischer, and E Cutrell. Combating rural child malnutrition through inexpensive mobile phones. In *Nordic Conference on Human-Computer Interaction*, pages 635–644, 2012.

- [93] I Medhi, A Tewari, M Jain, and E Cutrell. The Fate of a Digital Slate: Unexpected Issues with Deployment in Rural India. *User Experience Magazine*, 11(2), 2012.
- [94] Medmira. Introducing Multiplo - A Single Test That Gives Multiple Diagnoses in Minutes. *Executive Healthcare Magazine*, 2012.
- [95] G Mhila, B DeRenzi, C Mushi, T Wakabi, M Steele, P Dhaldialla, D Roos, C Sims, J Jackson, and N Lesh. Using Mobile Applications for Community-based Social Support for Chronic Patients. *HELINA*, 2009.
- [96] J Miele. The live scribe smartpen: A revolutionary new platform for audio/tactile graphics. In *International Cartographic Conference*, 2007.
- [97] M Migurski. Walking Papers. <http://walking-papers.org/>.
- [98] M Mitchell, N Lesh, H Cranmer, H Fraser, I Haivas, and K Wolf. Improving careimproving access: the use of electronic decision support with AIDS patients in South Africa. *International Journal of Healthcare Technology and Management*, 10(2):156–168, 2009.
- [99] Mobile Assay. <http://www.mobileassay.com/>.
- [100] O Mudanyali, S Dimitrov, U Sikora, S Padmanabhan, I Navruz, and A Ozcan. Integrated Rapid-Diagnostic-Test Reader Platform on a Cellphone. *Lab on a Chip*, pages 2678–86, 2012.
- [101] O Mudanyali, S Padmanabhan, S Dimitrov, I Navruz, U Sikora, and A Ozcan. Smart rapid diagnostics test reader running on a cell-phone for real-time mapping of epidemics. In *Proceedings of the Second ACM Workshop on Mobile Systems, Applications, and Services for HealthCare*, pages 1:1–1:6, 2012.
- [102] M Muja and D Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. In *International Conference on Computer Vision Theory and Applications*, 2009.
- [103] Nexleaf Analytics. <http://nexleaf.org>.
- [104] OpenCV. <http://opencv.willowgarage.com/wiki/>.
- [105] OraQuick. The first in-home oral HIV test. <http://www.oraquick.com>.
- [106] U Pal, R Jayadevan, and N Sharma. Handwriting recognition in indian regional scripts: A survey of offline techniques. *ACM Transactions on Asian Language Information Processing*, 11(1):1:1–1:35, 2012.
- [107] T Parikh, P Javid, K Sasikumar, K Dhosh, and K Toyama. Mobile Phones and Paper

- Documents: Evaluating a New Approach for Capturing Microfinance Data in Rural India. In *ACM SIGCHI Conference on Human factors in computing systems*, pages 551–560, 2006.
- [108] V Parmar, D Keyson, and C Debont. Persuasive technology for shaping social beliefs of rural women in india: An approach based on the theory of planned behaviour. In *Proceedings of the 3rd international conference on Persuasive Technology*, pages 104–115, 2008.
- [109] K Pearson. On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine*, 2(6), 1901.
- [110] Anoto Digital Pen and Paper Technology. <http://www.anoto.com/>.
- [111] T Plotz and G Fink. Markov models for offline handwriting recognition: a survey. *International Journal on Document Analysis and Recognition*, 12(4):269–298, 2009.
- [112] Scantron Scanner Pricelist. www.peppm.org/Products/scantron/price.pdf.
- [113] D Ramachandran, J Canny, P Dutta, and E Cutrell. Mobile-izing health workers in rural india. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1889–1898, 2010.
- [114] R Rao, S Card, W Johnson, L Klotz, and R Trigg. Protofoil: storing and finding the information worker’s paper documents in an electronic file cabinet. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 180–185, 1994.
- [115] A Ratan, S Chakraborty, P Chitnis, K Toyama, K Ooi, M Phiong, and M Koenig. Managing Microfinance with Paper, Pen and Digital Slate. In *International Conference on Information and Communication Technologies and Development (ICTD)*, 2010.
- [116] Remark Office OMR software. <http://www.gravic.com/remark/officeomr/index.html>.
- [117] A Rowe, D de Savigny, C Lanata, and C Victora. How can we achieve and maintain high-quality performance of health workers in low-resource settings? *Lancet*, 366:1026–35, 2005.
- [118] Sana. <http://sana.mit.edu/>.
- [119] Y Sato, Y Kobayashi, and H Koike. Fast tracking of hands and fingertips in infrared images for augmented desk interface. In *IEEE Automatic Face and Gesture Recognition*, pages 462–467, 2000.
- [120] K Schreiner. Uniting the paper and digital worlds. *IEEE Comput. Graph. Appl.*,

28(6):6–10, 2008.

- [121] A Sellen and R Harper. *The Myth of the Paperless Office*. MIT Press, Cambridge, MA, USA, 2003.
- [122] A Seth. Uдай OMR tool. <http://www.cse.iitd.ernet.in/~aseth/udai/OMRProj/README.html>.
- [123] L Shen, J Hagen, and I Papautsky. Point-of-care colorimetric detection with a smart-phone. *Lab on a Chip*, 12(21):4240–4243, 2012.
- [124] G Singh, L Findlater, K Toyama, S Helmer, R Gandhi, and R Balakrishnan. Numeric Paper Forms for NGOs. In *International Conference on Information and Communication Technologies and Development (ICTD)*, 2009.
- [125] Skannex. Lateral Flow and Elisa Reader Systems. <http://www.skannex.com/>.
- [126] Skannex. Readitlateral image processing software. <http://www.skannex.com/products/technology>.
- [127] Skannex. SkanSmart Mobile Reader. <http://www.skannex.com/products/skansmart>.
- [128] J Smith, J Long, T Lung, M Anwar, and S Subramanian. Paperspace: a system for managing digital and paper documents. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems*, pages 1343–1348, 2006.
- [129] S Srihari and E Kuebert. Integration of hand-written address interpretation technology into the united states postal service remote computer reader system. In *Proceedings of the 4th International Conference on Document Analysis and Recognition*, pages 892–896, 1997.
- [130] D Stevens. *Development and Optical Analysis of a Microfluidic Point-of-Care Diagnostic Device*. PhD dissertation, University of Washington, Department of Bioengineering, 2010.
- [131] D Stevens, C Petri, J Osborn, P Spicar-Mihalic, K McKenzie, and P Yager. Enabling a microfluidic immunoassay for the developing world by integration of on-card dry reagent storage. *Lab on a Chip*, 8:2038–2045, 2008.
- [132] Scantron Corporation Store. <https://store.scantron.com>.
- [133] Bayer Multistix Pro Urinalysis Strips. <http://www.bayermultistix.com>.
- [134] S Suzuki and K Abe. Topological Structural Analysis of Digitized Binary Images by Border Following. *Computer Vision, Graphics, and Image Processing*, 30(1):32–46, 1985.

- [135] The Akshara Foundation. <http://www.akshara.org.in/>.
- [136] P Tighe. Improving the quality of urine strip testing: The Clinitek 50 urine chemistry analyser. *Euro Clin Lab*, pages 16–20, 1997.
- [137] V Tsang, B Wilson, and S Maddison. Kinetic Studies of a Quantitative Single-tube Enzyme-linked Immunosorbent Assay. *Clin Chem*, 26:1255–1260, 1980.
- [138] H Underwood, R Sterling, and J Bennett. The Design and Implementation of the PartoPen Maternal Health Monitoring System. In *ACM Conference on Computing for Development*, 2013.
- [139] H Underwood, R Sterling, and J Bennett. The PartoPen in Training and Clinical Use: Two Preliminary Studies in Kenya. In *Proc. 6th International Conference on Health Informatics*, 2013.
- [140] United Nations Statistics Division. Population and Vital Statistics Report. <http://unstats.un.org/unsd/demographic/products/vitstats/default.htm>, 2012.
- [141] USAID — DELIVER PROJECT. Task Order 4. 2012. Mozambique: Strengthening the Community Health Worker Supply Chain. Arlington, VA.
- [142] S Van Wart, K Tsai, and T Parikh. Local Ground: A Paper-Based Toolkit for Documenting Local Geo-spatial Knowledge. In *ACM Conference on Computing for Development*, 2010.
- [143] R Veeraraghavan. Dealing with the digital panopticon: The use and subversion of ict in an indian bureaucracy. In *International Conference on Information and Communication Technologies and Development*, ICTD '13, pages 248–255, 2013.
- [144] VillageReach. <http://villagereach.org/>.
- [145] S Wang, X Zhao, I Khimji, R Akbas, W Qiu, D Edwards, D Cramer, B Ye, and U Demirci. Integration of cell phone imaging with microchip elisa to detect ovarian cancer he4 biomarker in urine at the point-of-care. *Lab Chip*, 11:3411–3418, 2011.
- [146] P Wellner. Interacting with paper on the DigitalDesk. *Communications of the ACM*, 36:87–96, 1993.
- [147] WHO. The global shortage of health workers and its impact. *Fact sheet*, 302, 2006.
- [148] WHO. Global Experience of Community Health Workers for Delivery of Health Related Millennium Development Goals, 2010.
- [149] F Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1:80–83, 1945.

- [150] B Wolfe, B Mamlin, P Biondich, H Fraser, D Jazayeri, C Allen, J Miranda, and W Tierney. The OpenMRS system: collaborating toward an open source EMR for developing countries. *AMIA Annu Symp Proc*, page 1146, 2006.
- [151] R Wynn. Optical mark recognition. *Data Processing*, 26(9):26 – 27, 1984.
- [152] P Yager, G Domingo, and J Gerdes. Point-of-Care Diagnostics for Global Health. *Annual Review of Biomedical Engineering*, 10:107–144, 2008.
- [153] P Yager, T Edwards, E Fu, K Helton, K Nelson, M Tam, and B Weigl. Microfluidic Diagnostic Technologies for Global Public Health. *Nature*, 442(7101):412–418, 2006.
- [154] H Zhu, S Isikman, O Mudanyali, A Greenbaum, and A Ozcan. Optical imaging techniques for point-of-care diagnostics. *Lab Chip*, 13, 2013.