The Case for Context-Aware Compression

Xuan Bao
Duke University

Trevor Narayan
Duke University

Ardalan Amiri Sani
Rice University

Wolfgang Richter
Carnegie Mellon University

Romit Roy Choudhury
Duke University

Lin Zhong
Rice University

Mahadev Satyanarayanan
Carnegie Mellon University

ABSTRACT

The proliferation of pictures and videos in the Internet is imposing heavy demands on mobile data networks. This demand is expected to grow rapidly and a one-fit-all solution is unforeseeable. While researchers are approaching the problem from different directions, we identify a human-centric opportunity to reduce content size. Our intuition is that humans exhibit unequal interest towards different parts of a content, and parts that are less important may be traded off for price/performance benefits. For instance, a picture with the Statue of Liberty against a blue sky may be partitioned into two categories – the semantically important *statue*, and the less important *blue sky*. When the need to minimize bandwidth/energy is acute, only the picture of the statue may be downloaded, along with a meta tag *"background: blue sky"*. Once downloaded, an arbitrary "blue sky" may be suitably inserted behind the statue, reconstructing an approximation of the original picture. As long as the essence of the picture is retained from the human's perspective, such an approximation may be acceptable. This paper attempts to explore the scope and usefulness of this idea, and develop a broader research theme that we call *context-aware compression*.

1. INTRODUCTION

Mobile broadband traffic continues to increase at an overwhelming pace. Predictions for 2014 suggest a 39 fold increase in demand, far exceeding the wireless capacity promised by foreseeable technologies, such as 4G/WiMax/LTE [1]. This dramatic increase is not only attributed to the surge in device density, but also to the eruption of high-resolution pictures and videos in the Internet. In fact, a study reports that by 2012, 3G networks will become saturated if 40% of its subscribers consume video just for 8 minutes a day [2]. Network operators are aware of the impending crisis, and are beginning to adopt precautions. For instance, ATT has already rolled out tiered pricing schemes that require users to operate below a pre-specified download quota. The expectation is that users will forcibly curb their browsing habits and collectively reduce the strain on the wireless spectrum. While pricing is indeed one solution to the problem, it may not be the desirable one.

Several researchers have taken up the challenge to cope with mobile data demands, and are exploring ways to offload cellular data networks. Ongoing approaches are mostly at the PHY/Link layer, including opportunistic migration of 3G traffic to WiFi [3], the use of femto cells [4], smarter antennas [5], etc. We break away from these schemes and explore a complimentary approach that attempts to reduce the volume of content, without significantly reducing the user's satisfaction.

Our observation is that humans are the dominant consumers of online content, and they exhibit an *unequal* degree of interest for different parts of the content. As an example, in a video of a stand-up comedy show, the comedian's actions may need to be preserved as is, however, the backdrop may be amenable to modification. Similarly, in a picture of a child sitting in a garden (Fig. 1), the garden may be altered without compromising the satisfaction of the person who views the picture. If one is able to isolate the subject of the content (the child) from its background (the garden), it may be possible to only download the subject along with a brief description of the background. Once downloaded, the receiver can select a similar background (from its local database of pictures) and carefully insert it behind the subject. The outcome is a variant of the original picture but is expected to preserve its semantic value/context. Some information will obviously be lost, and may reduce the user's satisfaction. However, the cost savings in bandwidth and battery power may adequately compensate for the dissatisfaction. If future network services come with stricter price plans, we believe that context-aware compression may offer a useful knob to cope with the price-performance tradeoff. A person with little left in her download quota may opt for compressed news sites, where the picture of the President is received as is; only the large crowd he is addressing gets locally synthesized.

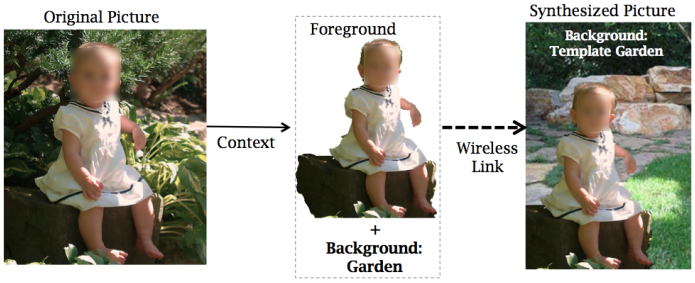

Figure 1: Core idea in context-aware compression

HotMobile '11 Phoenix, Arizona, USA

1.1 Natural Questions

A number of questions arise as one begins to consider the notion of context-aware compression. We touch upon some of them here, and revisit them in subsequent sections through measurements and hypotheses testing.

(1) If the background is not important, why not perform a (very) lossy compression on the background? This will preclude the need to carefully replace the background with a different one. We believe that significantly lowering the fidelity of the background may degrade the user's viewing experience. The user may strive to discern what the background originally was – a blurred garden may not be obvious after heavy lossy compression. Moreover, striking disparity in the resolution of the foreground and background may make the picture look unnatural. With our proposal, the synthesized background remains in high fidelity – the user immediately knows that the child is in the garden. The background can also be watermarked so the user can precisely learn about the synthesized parts, and request the original if desired.

(2) Context-aware compression may not be acceptable always. What kind of use-cases lend themselves to such modifications? Clearly, our ideas become less relevant when the viewer has adequate bandwidth, energy, or can tolerate latency in consuming the content (e.g., going home and downloading the full-fidelity version over WiFi). Certain image-retrieval applications may also be unacceptable because what may be unimportant to one user may not be for another. Further, special-occasion pictures within close social circles (e.g., wedding pictures) will also not be suitable for compression. However, not-so-special content shared with broader audiences may offer opportunity for context-aware compression. Applications may include downloading news/blogs/articles over mobile phones, talk shows on mobile TV, music videos, etc. The content providers may provide both compressed and uncompressed versions; users may choose one based on their position on the price-performance tradeoff.

This paper explores the theme of context-aware content by defining the various hypothesis that needs to hold, and verifying them through small scale experiments. We show that humans indeed exhibit preferential treatment towards different parts of a content, and such preferences are quite correlated across individuals. Encouraged by these findings, we develop a simple heuristic that automatically identifies objects in a picture, and preserves them during transmission. The background to these objects are eliminated and replaced at the receiver using templates created in PhotoShop. Although our prototype is crude at this stage, we believe there is evidence that context-aware compression can be a relevant software primitive for the future. The relevance will not only increase with greater sophistication in image processing, but also with a stricter need to reduce content footprint for overloaded wireless networks and batteries.

The next section formulates and tests the hypotheses that constitute the basis for context-aware compression. Thereafter, we present some preliminary heuristics that demonstrate the promise of this space. We close the paper with discussions on the longer term research agenda, followed by related work and a brief conclusion.

2. HYPOTHESES AND VERIFICATION

We state 3 main hypotheses (in this paper, we focus on images alone and treat videos as a time-sequence of images). These 3 hypotheses are not meant to be exhaustive; they are the critical ones necessary to erect the theme of context-aware compression.

1. **Some parts of images are semantically less valuable than others,** and a user is willing to compromise the fidelity of these parts in exchange for performance gains. Let us call these less-valuable parts *backgrounds*, and the complimentary portions (i.e., the semantically valuable areas), *foregrounds*. Synthesizing the backgrounds (e.g., inserting a template garden behind the child) will not diminish the human's satisfaction excessively.

2. **Human users exhibit high overlap in their description of image foregrounds and backgrounds.** Therefore, a "good" partition of foreground and background will satisfy the majority of users.

3. **Removing the background reduces the content size.** If the background of a picture is naturally amenable to heavy compression (e.g., a clear blue sky), then the gains from context-aware compression will be negligible. We hypothesize that many pictures have a sizable background, and therefore, eliminating them during transmission is gainful.

In an attempt to verify these hypotheses, an experiment was designed and conducted with real users. The experiment methods and findings are described next.

2.1 Experiment Methodology

We implemented a simple image cropping tool in Java. The tool downloads random pictures from Flickr (or any other source), draws a $N \times N$ grid on the picture, and displays it to the user. Users can select the background of the picture by selecting multiple grid boxes – the background can be composed of multiple non-overlapping portions of the picture. We invested effort to make the background selection simple so that operational biases are minimized. Once a user has selected the background, she performs a *crop* operation, which leaves only the foreground on the screen. The user has the option to revert to the original and make changes, if she feels that the context of the picture is not adequately captured. Figure 2 shows a screenshot from our software tool.

We recruited 6 student volunteers, explained our ideas to them, and asked them to partition the foreground and backgrounds in a way that would preserve the context of the picture. We asked them to imagine that the pictures will be viewed over a mobile phone/iPad, and that the viewer is under a $200MB$ data plan (recently launched by ATT). The images were selected randomly from Flickr and covered different genres, including natural scenes, street views, people, paintings, etc. Each participant cropped 50 pictures.

After the experiments, we interviewed the participants to understand the reasonings behind their choices. We specifically intended to learn whether they were able to satisfactorily crop out the essence of the pictures, as opposed to selecting only the visually appealing parts. Most people said that the *"context in the pictures heavily overlapped with the visually appealing parts"*. However, they emphasized that in several cases visually unattractive parts were also selected because they were

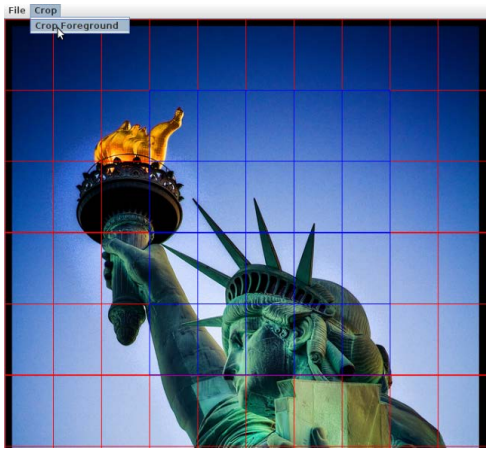

Figure 2: Software tool screenshot

integral to the context. One user, for instance, said that *"the ferry boat in front of the statue of liberty was not visually attractive ... yet, I included it to capture the tourism aspect"*. Based on these interviews, we gained reasonable confidence that the foregrounds reflect the contexts.

2.2 Measurement Results

We verify the hypotheses based on the results of the experiments described above.

(H1) Some parts of images are semantically less valuable than others. Figure 3 shows the CDF of the ratio between the foreground-area and the entire image area. Evidently, for more than 80% of the Flickr pictures, less than 70% of the image areas were cropped out as foreground. For around 50% of these pictures, the foreground covers less than 50% of the area of the entire image. This demonstrates that, on average, the less valuable parts of the picture – the background – makes up a reasonably large area of the picture. Synthesizing them carefully can offer gains.

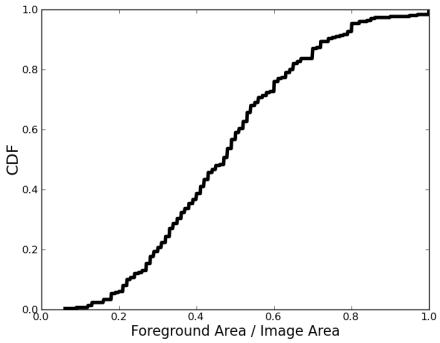

Figure 3: Ratio between the area of the foreground and the area of the entire image

(H2) Human users exhibit high overlap in their description of image foregrounds and backgrounds. To verify whether different users attach importance to similar parts of the image, we computed the *overlap* in foreground for each pair of users. We define overlap as:

$$\frac{Foreground_i \cap Foreground_j}{Foreground_i \cup Foreground_j}$$

$Foreground_i$ denotes the area of the foreground selected by user i. This equation compares the foreground area selected by both users with the foreground area selected by at least one user. Figure 4 shows that in 50% of the cases, the overlap is around 75%. This supports the observation that humans' perception of importance are similar to each other.

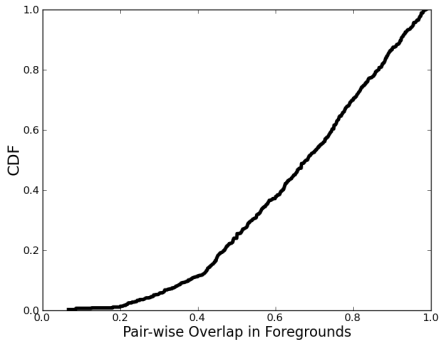

Figure 4: Foreground "overlap" for all pairs of users.

(H3) Removing the background reduces the content size. Figure 5 shows the CDF of uncompressed and compressed background file sizes, as cropped out by the users in our experiments. The uncompressed background is a JPEG file, while the compressed background was produced by subjecting the same JPEG file to the standard Linux-based *bzip* operation. The two curves exhibit a small gap between them implying that backgrounds are not significantly amenable to compression.

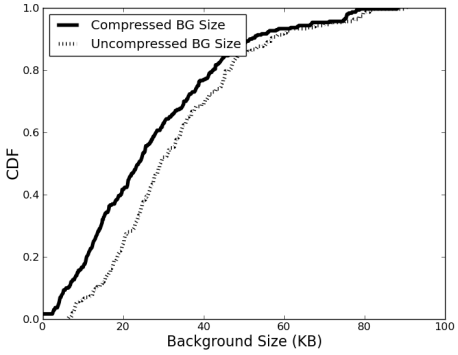

Figure 5: CDF of compressed and uncompressed background file sizes

We translate the above findings into expected performance gains. We assume that the background of each picture can be substituted by a high level meta-tag (e.g., garden, cloudy-sky, etc.), and a similar background can be inserted at the receiver. Thus, the performance gain G is the ratio of the background size to the entire size of the picture. If $G = 0.4$, it implies that context-aware compression reduces the size of the picture by 40%. We also compute G_z, which is the gain if all pictures were zipped at the 3G tower, and unzipped at the mobile device. We define G_z as $\frac{bzip(background)}{bzip(background)+bzip(foreground)}$. Figure 6 plots the distribution of G and G_z for all backgrounds cropped out by the human users. On average, picture sizes can be reduced by 50% if one is willing to unzip the content at the receiver (hence, pay an energy cost). Otherwise, the savings are around

40%. We believe this order of savings justifies further research in context-aware compression.

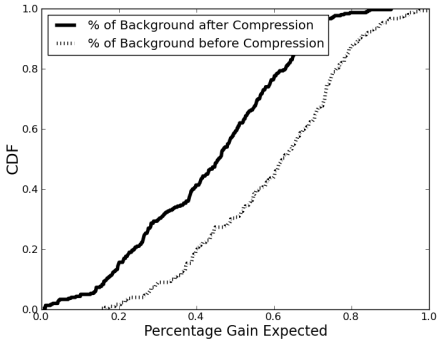

Figure 6: Potential gain for compressed (using bzip) and uncompressed images

3. FRAMEWORK DESIGN

We build a preliminary framework to demonstrate the feasibility of context-aware compression. We target both the image and the video domains.

3.1 Context-Aware Image Compression

We have made an early attempt at context-aware image compression. An application that requires human assistance on a per-image basis is not quite feasible. Towards an automatic means of extracting the foreground, we employ a simple heuristic. We assume that the context of an image is typically captured through objects in the image, and that these objects are often located near the center of the image. Thus, we first employ object recognition methods in image processing and accordingly identify the foreground.

We borrow object identification techniques from [6]. Authors in [6] use a combination of multiscale saliency, color contrast, edge density, and super-pixel straddling, to identify square shaped windows in an image. Our heuristic selects large sized windows, located near the center, as the foreground. Figure 7 shows the operation on a few randomly chosen Flickr images, used in our experiments. The union of the red-boxed areas is assumed to be the foreground. Our results show that the computer selected foregrounds reasonably overlap with the human-selected foregrounds (Figure 8). More than 50% of the cases, the overlap between two selected foregrounds is more than 55%. We concede that our heuristic is not accurate – the selected foreground may not include all important objects and the boundary of the foreground may not align precisely with objects' edges.

3.2 Video Transmission

For many video content, the video-recording is performed in indoor environments, using cameras from multiple vantage points. Examples include talk shows, news broadcast, interviews, stand-up comedies, etc. In these programs, the background environment is typically (well-decorated) studio walls, stage backdrops, or perhaps a sitting audience behind the speaker. When a video plays, the backgrounds across different video-frames are likely to be different views of similar environment. Therefore, it may be possible to use templates for the backgrounds for different room settings and different camera

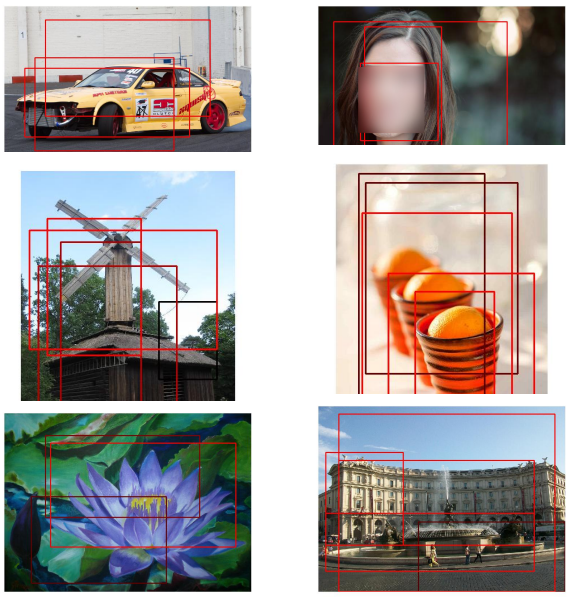

Figure 7: Images with selected foreground

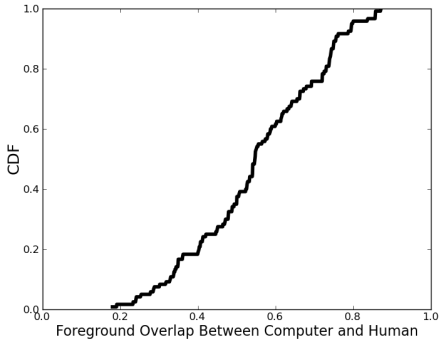

Figure 8: Overlap between foregrounds selected by computer heuristics and human users

angles. For instance, we can have templates for the broadcasting room for different camera angles and ranges. These templates can be pre-loaded at the users' mobile phones. Later, the video server can transfer the foreground with pointers to which templates may be used to synthesize the appropriate background. The insertion of templates can be done at users' devices.

We have made an early attempt with a talk-show video (from YouTube). Figure 9 shows how the background can be constructed by employing the "content aware deletion" tool in Photoshop. Specifically, a human crops the foreground – the speaker in this video frame – and Photoshop deletes this foreground to generate the frame on the right. This right frame is now carefully inserted to subsequent frames. Figure 10 shows the outcome – the left side frames are the originals, and the corresponding right side ones are synthesized. **We found that these synthesized frames are quite viewable.** We asked 6 people to rate the quality of synthesized images. The average rating was 4 out of 5, demonstrating that the user's satisfaction does not degrade excessively. Of course, this is only a toy test – extensive experimentation is necessary to quantify the deeper tradeoffs between bandwidth and user satisfaction.

Figure 9: (a) The original frame from a talk show. (b) Extracting the background.

Figure 10: Synthesizing backgrounds for videos.

4. LIMITATION AND FUTURE WORK

This section discusses some additional research questions that would require future research attention.

Template backgrounds. We assumed that the background of the original picture can be concisely summarized (e.g., cloudy blue sky). This may require sophisticated image processing, such as Google Goggles, or some form of crowdsourcing to label pictures reasonably well. These kind of techniques are gaining traction [7], but may not be fully mature in the near future.

Template substitution. Even if an appropriate background template is available, we assumed that it can be inserted in a way that does not affect the foreground. For instance, when a Ferrari car is the foreground, and the background is a view of a street, it is important to superimpose the Ferrari in a way that is meaningful. If the image operation inserts the Ferrari on the sidewalk, or on top of other cars in the street, the image will be distorted. Focused research would be necessary to synthesize the picture with respect to placement and proportion. Some image processing techniques are already making progress in this direction [8]. Admittedly, insertion of a visually satisfying background without any human assistance is a difficult task for computer vision. We hope content providers may have the incentive to invest effort in creating synthesized backgrounds. Moreover, under certain circumstances (e.g., extremely bad connection or very limited download quota), a simple solution may be to transmit the foreground only, and delay the transmission of the background (perhaps until the user is in WiFi range). Thus, the user would be able to view the picture quickly, and in the few cases in which she cares about the picture's precise background, she would need to tolerate some latency.

Potential gain for video compression. Advanced video compression techniques such as MPEG-7 [9] have already eliminated most of information redundancy by exploiting similarity between successive frames. So a natural concern is what is the gain of our technique compared to the established ones. Our viewpoint is that our technique can always be applied on top of any of these conventional techniques since even the background of the "key" frames can be substituted in our case. One may view this as a lossy technique, where the notion of loss is influenced by human psychology (as opposed to human perception). Clearly, significant future work is necessary to translate this notion to an acceptable/usable system.

Potential applications and requirements. Besides saving bandwidth, context-aware compression can also be used towards *variable-fidelity storage* and information distillation. One may envision a surveillance camera recording videos in variable fidelity (backgrounds of older videos proportionally reduced in fidelity). Future research will need to explore the variety of applications in context-aware compression.

5. RELATED WORK

The idea of context-aware compression draws from multiple threads of research, including compression, image processing, and application-awareness.

Lossy Compression. Lossy compression [10] dedicates to search for encodingds that can compress file size significantly while Lossy compression [10] pertains to encoding algorithms that tradeoff information loss for reduced content volume. Although a mature field, there is renewed interest here in light of the pressing need to reduce content size. Very recently, Google developed an image compression format "WebP" [11] that reduces content size without affecting the viewing experience too much. Authors in [12, 13] have also looked into human factors. They have observed that human attention is usually drawn to certain visual features, and hence, images can be rendered with more details to such objects. In contrast, our proposal is to exploit the human interest at the higher, *semantic* level – we extract and preserve the main context as is.

Image processing. Our proposal relies heavily on image processing (especially object recognition) for tasks such as selecting foreground/background, constructing and replacing backgrounds with templates. Current effort in this field has made significant progress in identifying objects [6], recognizing the characteristics [14], and even tagging them [15]. Further advancements in such algorithms will only facilitate context-aware compression.

Application-awareness. The notion of application-awareness is broad and has been employed in various domains, such as operating systems and image search [16, 17]. Authors in [16] introduced the notion of fidelity and the value of lowering fidelity for bandwidth savings. A related paper [18] showed the benefits of application-awareness in energy saving. This paper draws from a variety of these ideas.

6. CONCLUSION

This paper proposes the notion of context-aware compression. The key observation is that sizable portions of human-consumed content are not critical towards preserving the semantic value of the content. One may leverage this slack by

transmitting only the contextually relevant portions – the foreground – in full fidelity, and concisely summarizing the background. Upon receiving the information, the receiver may be able to approximate the original content by "stitching" the foreground with a similar template background, drawn from its local cache. Such an approximation may be a "smaller price to pay" in comparison to the rising cost of wireless bandwidth. Thus, although our ideas and results in this paper are preliminary, we find evidence to believe that context-aware compression can be a promising tool for a variety of future applications.

7. REFERENCES

[1] Rysavy Research, "Mobile Broadband Capacity Constraints And the Need for Optimization," 2010.

[2] Nokia Siemens Networks, "Unite: Trends and insights 2009," 2009.

[3] A. Balasubramanian, R. Mahajan, and A. Venkataramani, "Augmenting Mobile 3G Using WiFi," in *ACM Mobisys*, 2010.

[4] J.H. Yun and K.G. Shin, "CTRL: A Self-Organizing Femtocell Management Architecture for Co-Channel Deployment," *ACM MobiCom*, 2010.

[5] A.A. Sani, L. Zhong, and A. Sabharwal, "Directional Antenna Diversity for Mobile Devices: Characterizations and Solutions," in *ACM MobiCom*, 2010.

[6] B. Alexe, T. Deselaers, and V. Ferrari, "What is an object," *CVPR*, 2010.

[7] T. Yan, V. Kumar, and D. Ganesan, "CrowdSearch: exploiting crowds for accurate real-time image search on mobile phones," in *ACM MobiSys*, 2010.

[8] P. Debevec, "Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography," in *ACM SIGGRAPH*, 2008.

[9] ISO, "MPEG-7 Overview," *http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm*.

[10] G.K. Wallace, "The JPEG still picture compression standard," *IEEE TCE*, 2002.

[11] Google, "Introduction to WebP on Google code," *http://code.google.com/speed/webp/*.

[12] C. Oï£¡Sullivan and et. al., "Perceptually adaptive graphics," *Eurographics State of the Art Reports*, 2004.

[13] Seung-Hyun Lee, Sang-Bok Choi, and et. al., "Non-uniform image compression using a biologically motivated selective attention model," *Neurocomputing*.

[14] D.G. Lowe, "Object recognition from local scale-invariant features," in *ICCV*, 1999.

[15] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE TPAMI*, 2002.

[16] B.D. Noble, M. Satyanarayanan, D. Narayanan, J.E. Tilton, J. Flinn, and K.R. Walker, "Agile application-aware adaptation for mobility," *ACM SOSP*, 1997.

[17] M. Satyanarayanan, R. Sukthankar, L. Mummert, A. Goode, J. Harkes, and S. Schlosser, "The unique strengths and storage access characteristics of discard-based search," *Journal of Internet Services and Applications*, vol. 1(1), 2010.

[18] J. Flinn and M. Satyanarayanan, "Energy-aware adaptation for mobile applications," in *ACM SOSP*, 1999.