A Contact Sheet Approach to Searching Untagged Images on Smartphones

Jan Harkes, Mahadev Satyanarayanan,
Ardalan Amiri Sani[†], Benjamin Gilbert

September 2011
CMU-CS-11-132

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

[†]Rice University

Abstract

We describe a cloud-based approach to opportunistic, crowd-sourced, near real-time search of untagged images on smartphones that is sensitive to bandwidth and energy constraints. Our approach is inspired by the long-established practice of photographers using *contact sheets* to rapidly visualize a new collection of photographs, and then selecting a subset on which to focus attention. On behalf of each smartphone, the cloud maintains a *virtual contact sheet* of images that have been captured but not yet uploaded. The virtual contact sheet consists of thumbnails as well as full or partial meta data associated with the image. If search processing on the cloud indicates that a particular thumbnail is relevant, then its full-fidelity image can be obtained from the corresponding smartphone for further search processing or presentation to the user. We identify refinements, design tradeoffs and research questions pertaining to this approach.

This research was supported by the National Science Foundation (NSF) under grant number CNS-0833882, an NSF Graduate Research Fellowship, an IBM Open Collaborative Research grant, and an Intel Science and Technology Center grant. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily represent the views of the NSF, IBM, Intel, or Carnegie Mellon University.

Keywords: smartphones, sensor networks, image search, content search, Diamond, Theia, crowd sourcing, energy-aware adaptation, weakly-connected operation, low bandwidth, 3G, 4G, mobile computing, pervasive computing, wireless networks, cloud computing, cloudlets, real time

1 Background and Problem Statement

Smartphones allow users to easily capture image and video data of real-world events at any time and place. This rich mobile sensing role represents a disruptive technology. "Rich" connotes real-world depth and complexity, implying data of very high dimensionality from the viewpoint of pattern recognition [6]. Rich data capture combined with *opportunism* and *crowd-sourcing* leads to new application metaphors. Specifically, the ability to search image content soon after capture from a large collection of smartphones in the field would be valuable. This is best illustrated with an example [16], summarized below.

Figure 1: Lost Child in Crowd-Sourced Photo

Consider the plight of parents who have lost their child in a large crowd such as the Macy's Thanksgiving Day parade in Manhattan. Searching for the child is a daunting task in these circumstances. Imagine the police geocasting a request to all smartphone owners in the vicinity, asking to search photographs that were taken recently. The background of one of those photographs (such as Figure 1) may reveal the lost child, soon leading to her rescue. Note the role of opportunism here. The users whose pictures are searched were completely unaware of this potential future use. They took the picture for some other reason, such as an interesting building or a funny float in the parade. But because of the richness of the sensed data, there are potentially "uninteresting" aspects of the image (e.g. small child in the corner of the picture) that prove to be very important in hindsight — *it is context that determines importance.* Because of the opportunistic use of image data, it is very unlikely that the images will be manually tagged or automatically indexed for the later search criteria.

Opportunistic use of images is reality today. A recent CNN news article [5] reported the arrest of a thief whose act of stealing appeared in the background of a family picture. Figure 2 shows this photograph, with an outline highlighting the thief caught in the act. Notice, once again, the importance of opportunism and search context. The person in the background of Figure 2 would normally have been ignored by the photographer. It is only the context of a theft that makes this person interesting.

Figure 2: Theft Caught in Background of Family Photo (Source: CNN [5])

These examples show the importance of searching in near real-time. The longer a child is lost, the greater the chances that she will be harmed. The longer a thief is undetected, the more likely that he will escape far from the scene of the crime. These examples also show the significance of near real-time data consistency. In the lost child case, the only relevant photographs are those taken after the child was lost. In the theft example, the real-time constraint is even tighter.

Implementing this search capability is not trivial. One major challenge is that smartphones tend to be *weakly-connected* to the Internet. We use the term "weakly-connected" in the sense originally proposed by Mummert et al [10] as "networks with rather unpleasant characteristics: intermittence, low bandwidth, high latency, or high expense." Although smartphones can use Wi-Fi in hotspots, they are more frequently dependent on wireless WAN (WWAN) connectivity which is typically 200 Kbps or so today for uploads. While faster wireless technologies are on the horizon, their rollout and widespread deployment will be a slow process. Further, wireless service providers are moving away from "all you can eat" pricing to volume-sensitive pricing. "High expense" is thus becoming an unpleasant reality on WWANs. For the foreseeable future, WWAN connectivity on smartphones will remain weak relative to improvements in processing power, memory and storage capacity, image resolution, and other smartphone parameters. High-end smartphones already have the flash storage capacity today (32–64GB) to hold many thousands of high-resolution multi-megabyte images. Even as technology evolves, this invariant will endure: ***the ability to capture and store large volumes of image data on smartphones will outstrip wireless transmission of that data from the distant edges of the mobile Internet.*** Video capture worsens this problem.

A second challenge is *battery life.* Computing and wireless transmission both consume energy on a smartphone. An ideal search strategy would trade off a small amount of edge computing for greatly reduced volume of data transmission.

A third challenge is that the features of interest in an image may not be known in advance of a query, because of the opportunistic nature of queries. In Figure 1, for example, it is hard to see how a preprocessing algorithm could be prescient enough to extract and index the unique features of the tiny speck representing the lost child. The problem is easier in hindsight, with the child-specific knowledge possessed by the parents and the context of their search. In Figure 2, the importance of

(a) (b)

Figure 3: Examples of Contact Sheets

the tiny figure in the background is only obvious in the context of the theft. Without that context, one would typically ignore the figure as an irrelevant detail.

More generally, automated indexing of image data has been a long-standing challenge for several reasons. First, automated methods for extracting semantic content from images are still rudimentary. This is referred to as the "semantic gap" [9] in information retrieval. Second, the richness of the data requires a high-dimensional representation that is not amenable to efficient indexing. This is a consequence of the curse of dimensionality [2, 6, 21]. Third, realistic user queries can be very sophisticated, requiring a great deal of domain knowledge and contextual knowledge that is not available in preprocessing. These deep algorithmic issues are further complicated by the energy and bandwidth constraints of smartphones. Our focus on near real-time search suggests that the most valuable images for a search may not yet have been uploaded to the cloud, but are only available on weakly-connected smartphones. The bandwidth and energy constraints imply that neither extensive preprocessing on the smartphones nor uploading of all images for search processing in the cloud is satisfactory.

These considerations frame the problem of interest: *How does one perform opportunistic, crowd-sourced, near real-time search of untagged images on smartphones while respecting their bandwidth and energy constraints?*

2 Solution Strategy

A well-known practice in film-based photography provides the inspiration for our solution strategy. Photographers have long used *contact sheets* to rapidly visualize a new collection of photographs, and to then select a subset on which to focus attention. In the era of film photography, a contact sheet was created exactly as its name suggests: by placing a roll of film in contact with photo paper and then generating a print. Figure 3 shows some example contact sheets. Although photography has moved away from film-based technology, the practice of generating "contact sheets" continues in digital photography. These consist of low-fidelity *thumbnail images* that are much smaller in size and lower in resolution but are otherwise identical to the full-fidelity captured images. Popular software such as Adobe Photoshop and Gimp support creation of contact sheets. The continuing use of contact sheets in digital photography, long past its film-era roots, suggests a solution strategy based on the following insights:

- First, contact sheets represent a good abstraction for humans to rapidly visualize a large collection of images and to then narrow interest to a few images.
- Second, search algorithms that embody image processing may be able to emulate this human ability to rapidly narrow focus from thumbnails.
- Third, a bandwidth-efficient, energy-efficient and near real-time approach to smartphone image search can be obtained by transmitting thumbnails from smartphones to the cloud, executing search algorithms in the cloud to narrow focus, and then using these results to selectively fetch full-fidelity images to the cloud for further search processing.

Elaborating on the above strategy, the cloud maintains on behalf of each smartphone a *virtual contact sheet* of images that have been captured but not yet uploaded to the cloud. The virtual contact sheet consists of thumbnails as well as full or partial meta data associated with the image (such as location and time of capture, camera setting details, and so on). There are many design tradeoffs in maintaining the virtual contact sheet. These include, for example, when thumbnails are generated, their size and resolution, how soon after capture they are uploaded to the cloud, where the cloud is located, the completeness of the meta-data uploaded with the thumbnails, and so on. These tradeoffs are discussed in Section 3.

From time to time, each smartphone performs a full sync with the cloud. This might typically occur many hours or days apart, for example when the smartphone is plugged in for charging its battery and is able to communicate over a wired connection to the cloud. After a full sync, the cloud has copies of all the captured images on the smartphone. The virtual contact sheet for this smartphone is therefore reset to empty on the cloud.

For opportunistic searches such as the examples in Section 1, the features of interest in an image may not be known *a priori.* It is therefore not possible to perform preprocessing for indexing. Instead, the *discard-based search* approach of Diamond [19, 20] can be used. In this approach, search processing is performed by the cloud in response to a specific search query, using code fragments called *searchlets* that are provided as part of the query. If a searchlet indicates that a particular thumbnail in a virtual contact sheet is relevant, one can proceed in a number of ways. For example, the smartphone could be contacted to fetch the full-fidelity image using the *data retriever* mechanism of Diamond, and the searchlet then applied to it. This could be done completely transparently to the user. Alternatively, the user performing the search could be involved in this process by presenting the thumbnail result and letting him decide if it is worth fetching the full-fidelity image from the smartphone. A third possibility is to also involve the owner of the smartphone in the process, either synchronously or by using an asynchronous notification mechanism such as a text message or email. The optimal balance between transparency, search user involvement, and smartphone user involvement may vary depending on factors such as the use case, search context, privacy concerns, incentives for crowd-sourcing, business model, and the technical constraints of the cloud-mobile architecture.

In settings such as participatory sensing [3] and citizen science [13], a complete set of feature extraction and indexing algorithms may be known to be necessary and sufficient *a priori.* Indexed search can then be supported on image content and image meta-data by preprocessing the images in the cloud using these algorithms. This preprocessing can also be applied to a virtual contact sheet, with the results indicating lower confidence because the algorithms can only access thumbnails. As

with discard-based search, various degrees of user involvement and smartphone owner involvement are possible when obtaining a full-fidelity image on demand from a smartphone.

3 Refinements, Design Tradeoffs and Research Questions

Implementing the high-level search architecture described in the previous section requires many details to be specified. Here we discuss the tradeoffs involved in these design choices, as well as optimizations and refinements to the basic architecture. An experimental approach to understanding these tradeoffs follows as a natural research agenda.

One fundamental tradeoff involves the size of thumbnails. Since a larger thumbnail is likely to preserve more relevant detail, the search algorithm to narrow focus is likely to yield precision and recall values close to what it would obtain if executed on the full-fidelity image. In other words, the false positives and false negatives are likely to be the same as on the full-fidelity image. There is an asymmetry here that is worth noting. An *anomalous false positive* on a thumbnail is a false positive that does not exist when the search algorithm is executed on the full-fidelity image. This does not change the precision or recall metrics of the overall search process, but it does waste bandwidth and energy: the corresponding full-fidelity image will be fetched from the smartphone even though it is not really relevant. In contrast, an *anomalous false negative* on a thumbnail is insidious. It worsens the recall metric for the overall search process because the corresponding full-fidelity image will not be obtained from the smartphone. This false negative would not have existed if the search algorithm had been executed on the full-fidelity image. Intuitively, this is a situation where the reduction in fidelity is too severe. There is a clearly a tradeoff here between resource usage (transmission bandwidth and energy) and search quality. Adaptive strategies that dynamically select thumbnail size and fidelity based on available bandwidth, smartphone battery level, and expected time to recharge may be useful.

Another important issue to be investigated is how precision and recall metrics degrade for different search algorithms as fidelity is reduced. To compensate for the lowered fidelity, the algorithms and parameters executed on thumbnails may have to be different from those used on the full-fidelity images. While total compensation may not be feasible, partial compensation for loss of fidelity in the input image may be possible. Since the search algorithms execute in the cloud, higher computational expense may be acceptable if it saves on bytes transmitted from smartphones.

A third design tradeoff involves thumbnail creation and upload latency. How soon after image capture should a smartphone create its thumbnail, and when should it upload it to the cloud? A write-through strategy has zero upload latency and best supports near real-time search. Until its thumbnail is uploaded, a captured image is effectively invisible to search processing. Generating the thumbnail may be most efficient soon after image capture. At this point, the full-fidelity image is present in virtual memory or the I/O buffer cache (depending on the details of the smartphone software) and can thus be processed without additional I/O on the smartphone. On the other hand, a case can be made for a write-back strategy in which the computation involved in thumbnail creation is deferred to periods when the smartphone processor is idle. Similarly, under conditions of high bandwidth demand it may be wise to delay uploading thumbnails until network demand from foreground activity is low. There is also an argument to be made for batching thumbnails until sufficient data volume is available for efficient streaming transmission.

In its simplest form, a thumbnail involves uniform reduction of fidelity over the whole image. In some cases, it may be possible to perform relatively cheap computation on the smartphone that reveals areas of the full-fidelity image that are of high importance for a specific application domain. Thumbnails could then be encoded to preserve higher fidelity around these important areas. For example, many cameras today have hardware/firmware support to perform face detection at near real-time speeds with little energy overhead. By preserving the areas around faces at full fidelity, a thumbnail that is shipped to the cloud could have face recognition performed on it without the additional latency of reaching back to the smartphone for the full-fidelity image. This encoding approach assumes that faces occupy only a small fraction of the total image area. In the case of a close-in shot where one or more faces span almost the entire image, the benefit of this approach may be limited. More generally, domain-specific image processing code that detects the presence of specific target classes (such as specific weapons in a military application) could be run on the smartphone to identify areas of importance in the full-fidelity image and to then guide the encoding of the thumbnail at non-uniform fidelity. This leads to many open areas of research to investigate, including the design of the thumbnail encoding format and the design of the domain-specific image preprocessing algorithms. Since the thumbnail of an image encoded in this manner is likely to be larger in size than if encoded at uniform fidelity, there is a complex tradeoff between encoding cost, thumbnail transmission cost, and real-time responsiveness of the overall system. Optimal handling of this tradeoff requires an adaptive thumbnail encoding mechanism that balances low-latency target identification against the bandwidth and energy currently available to the smartphone.

The data representation and transmission of virtual contact sheets involves some tradeoffs. Should the thumbnails in a virtual contact sheet be individual images (typically JPEG or PNG), or should they be merged into a single virtual contact sheet image? Merging involves decoding and re-encoding, and hence involves some energy consumption. However, for a small thumbnail image the storage overhead of the header (e.g., a JPEG or PNG header) can be significant. Since this overhead is amortized when thumbnails are merged into a single image, the total data transmitted from smartphone to cloud may be significantly smaller. The transmission of a single large object rather than many small objects may also be more efficient because of second-order effects involving TCP window size. Some aspects of this tradeoff also extend to storage and use of virtual contact sheets within the cloud.

4 Related Work

Closest in spirit to this work is Theia [1], which addresses the same problem of searching untagged images on a large number of smartphones. However, Theia's approach is fundamentally different in that it performs nontrivial search processing on smartphones in response to a specific search query. Only images that pass this early filtering are transmitted to the cloud and cached there, with further search processing occurring within the cloud. The code to perform this early filtering is shipped from the cloud to the smartphones. Based on results from a sampling of images, attention is focused on a subset of the smartphones. This exploits *relevance locality* to direct search processing to those smartphones that are most likely to have relevant results. Another optimization, called *partitioned search,* dynamically splits search computations between the cloud and smartphones in order to minimize use of the scarcest resource (energy or bandwidth). Theia's approach is more sophisticated and decentralized, but requires more complex software on the

smartphones. This includes the machinery to safely execute possibly untrusted searchlet code on smartphones. In contrast, the contact sheet approach described here is a simpler, centralized and cloud-centric approach that will be easy to deploy across a large number of users with diverse types of smartphones. No open-ended search computations are ever performed on smartphones, only reduction from full-fidelity images to thumbnails. The latter computation can be performed by carefully vetted, pre-installed code with a relatively small footprint in terms of code size and energy usage.

Our approach to saving bandwidth and energy on mobile devices by reducing fidelity and offloading computation builds on extensive prior work on *application-aware adaptation* in the context of Odyssey [7, 11, 12, 18]. In the domain of image processing, fidelity reduction lies at the heart of *perceptual hash algorithms* to define image similarity [8]. These algorithms use fidelity-reduction techniques such as scaling, average hashing, and discrete cosine transforms to generate a tiny image representation that is invariant across visually similar images.

The SLIPstream project [4, 14] explores some ways to distribute and efficiently execute computer vision applications across a data center, while minimizing latency to allow interactive use. The project explores automated techniques to characterize performance of such algorithms as a function of adjustable parameters, including image rescaling, and trade off latency and fidelity [22, 23]. This project has also demonstrated automatic partitioning of a computer vision application between a mobile device and a backend server across slow wireless networks, while maintaining interactive performance [15].

5 Conclusion

Although this document focuses narrowly on image search and smartphones, the ideas described here have broader applicability. For example, they also apply to any collection of networked sensors in which data capture and local storage are cheap and plentiful, but transmission of captured data is expensive. Expressed in abstract form, this approach consists of application-specific lowering of fidelity at the edges of the network, transmission of lower-fidelity representations for compute-intensive and/or data-intensive remote processing, and selective backfetching of full-fidelity representations from the edges based on the results of processing on their low-fidelity representations. This abstract approach applies to any signals that capture information of high dimensionality, images only being the most obvious example.

Another generalization of the ideas expressed in this document involves the term "cloud." All that is intended here by "cloud" is a well-connected network entity with ample storage, processing and energy resources. The range of possible implementations spans anything from a single server to a full-fledged cloud such as Amazon's EC2. Decentralized cloud infrastructure, referred to as *cloudlets*[17], may also be used. These may be particularly relevant in latency-critical applications or in military contexts.

6 Acknowledgements

We thank Glenn Ammons, Kiryong Ha, Lily Mummert, Babu Pillai, Wolgang Richter and Rahul Sukthankar for providing valuable suggestions that have improved the content and presentation of this document. The problem of searching untagged smartphone images was originally formulated in the context of the Theia project. We thank our Theia collaborators (Xuan Bao, Romit Roy Choudhury, Trevor Narayan, Wolfgang Richter, and Lin Zhong) for helping us to understand this problem, to crisply formulate it, and to develop an initial solution strategy in the form of Theia. The contact sheet strategy described here represents an alternative approach to solving the same problem.

This research was supported by the National Science Foundation (NSF) under grant number CNS-0833882, an NSF Graduate Research Fellowship, an IBM Open Collaborative Research grant, and an Intel Science and Technology Center grant. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily represent the views of the NSF, IBM, Intel, or Carnegie Mellon University.

References

[1] AMIRI SANI, A., RICHTER, W., BAO, X., NARAYAN, T., SATYANARAYANAN, M., ZHONG, L., AND CHOUDHURY, R. R. Opportunistic Content Search of Smartphone Photos. Tech. Rep. TR0627-2011, Rice University, June 2011.

[2] BERCHTOLD, S., BOEHM, C., KEIM, D., KRIEGEL, H. A Cost Model for Nearest Neighbor Search in High-Dimensional Data Space. In *Proceedings of the Symposium on Principles of Database Systems* (Tucson, AZ, May 1997).

[3] BURKE, J., ESTRIN, D., HANSEN, M., PARKER, A., RAMANATHAN, N., REDDY, S., AND SRIVASTAVA, M. B. Participatory sensing. In *Workshop on World-Sensor-Web (WSW06): Mobile Device Centric Sensor Networks and Applications* (2006).

[4] CHEN, M.-Y., MUMMERT, L., PILLAI, P., HAUPTMANN, A., AND SUKTHANKAR, R. Exploiting multi-level parallelism for low-latency activity recognition in streaming video. In *Proceedings of ACM Multimedia Systems* (February 2010).

[5] CNN report: New Jersey family's picture catches theft in the making. `http://www.cnn.com/2010/CRIME/08/24/new.jersey.theft.photo/index.html?hpt=C1`, August 2010.

[6] DUDA, R., HART, P., AND STORK, D. *Pattern Classification*. Wiley, 2001.

[7] FLINN, J., AND SATYANARAYANAN, M. Energy-aware Adaptation for Mobile Applications. In *Proceedings of the Seventeenth ACM Symposium on Operating systems Principles* (1999).

[8] KRAWETZ, N. Looks Like It (in *The Hacker Factor Blog*). `http://www.hackerfactor.com/blog/index.php?/archives/432-Looks-Like-It.html`, May 2011.

[9] MINKA, T., PICARD, R. Interactive Learning Using a Society of Models. *Pattern Recognition 30* (1997).

[10] MUMMERT, L. B., EBLING, M. R., AND SATYANARAYANAN, M. Exploiting Weak Connectivity for Mobile File Access. In *Proceedings of the 15th ACM Symposium on Operating System Principles* (Copper Mountain, CO, December 1995).

[11] NARAYANAN, D., AND SATYANARAYANAN, M. Predictive Resource Management for Wearable Computing. In *Proceedings of the 1st international conference on Mobile systems, applications and services* (San Francisco, CA, 2003).

[12] NOBLE, B. D., SATYANARAYANAN, M., NARAYANAN, D., TILTON, J. E., FLINN, J., AND WALKER, K. R. Agile Application-Aware Adaptation for Mobility. In *Proceedings of the 16th ACM Symposium on Operating Systems Principles* (Saint-Malo, France, October 1997).

[13] PAULOS, E., HONICKY, R., AND HOOKER, B. Citizen Science: Enabling Participatory Urbanism. In *Urban Informatics: Community Integration and Implementation, Information Science Reference*, M. Foth, Ed. IGI Global, 2008.

[14] PILLAI, P., MUMMERT, L., SCHLOSSER, S., SUKTHANKAR, R., AND HELFRICH, C. SLIPstream: Scalable, Low-latency Interactive Perception on Streaming Data. In *Proceedings of the International Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV)* (June 2009).

[15] RA, M. R., SHETH, A., MUMMERT, L., PILLAI, P., WETHERALL, D., AND GOVINDAN, R. Odessa: Enabling Interactive Perception Applications on Mobile Devices. In *Proceedings of MobiSys 2011* (Bethesda, MD, June 2011).

[16] SATYANARAYANAN, M. Mobile computing: the next decade. In *Proceedings of the ACM MobiCloud Workshop* (San Francisco, CA, June 2010).

[17] SATYANARAYANAN, M., BAHL, V., CACERES, R., AND DAVIES, N. The Case for VM-based Cloudlets in Mobile Computing. *IEEE Pervasive Computing 8*, 4 (Oct-Dec 2009).

[18] SATYANARAYANAN, M., AND NARAYANAN, D. Multi-fidelity algorithms for interactive mobile applications. In *Proceedings of the 3rd International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications (DIAL-M 1999)* (Seattle, WA, August 1999).

[19] SATYANARAYANAN, M., SUKTHANKAR, R., GOODE, A., BILA, N., MUMMERT, L., HARKES, J., WOLBACH, A., HUSTON, L., AND DE LARA, E. Searching complex data without an index. *International Journal of Next Generation Computing 1*, 2 (December 2010).

[20] SATYANARAYANAN, M., SUKTHANKAR, R., MUMMERT, L., GOODE, A., HARKES, J., AND SCHLOSSER, S. The unique strengths and storage access characteristics of discard-based search. *Journal of Internet Services and Applications 1*, 1 (2010).

[21] YAO, A., YAO, F. A General Approach to D-Dimensional Geometric Queries. In *Proceedings of the Annual ACM Symposium on Theory of Computing* (May 1985).

[22] YIGITBASI, N., MUMMERT, L. B., PILLAI, P., AND EPEMA, D. H. J. Incremental placement of interactive perception applications. In *Proceedings of HPDC 2011* (2011).

[23] ZHU, Q., KVETON, B., MUMMERT, L., AND PILLAI, P. Automatic Tuning of Interactive Perception Applications. In *Proceedings of the Conference on Uncertainty and Artificial Intelligence (UAI)* (July 2010).