Vision: The Case for Cellular Small Cells for Cloudlets

Sharad Agarwal, Matthai Philipose, Paramvir Bahl
{sagarwal, matthaip, bahl}@microsoft.com
Microsoft Corporation, Redmond, USA

Abstract – Today's cellular networks are built with"macro cell" basestations connected to the Internet via a rigid, complicated backhaul. Even with state-of-art technologies like LTE, users get limited throughput and high latency, with high variance. Performance enhancing IP boxes are deployed in the cellular operator's datacenters, far from the user. As a result, the most compelling cloudlet applications are difficult to realize on such networks and cloudlet researchers have thus far focused on Wi-Fi networks only.

We argue that the cloudlet community should consider small cell networks in addition to Wi-Fi networks. Small cells, such as femtocells and picocells, are relatively new additions to the cellular standards. By reducing the cell size compared to the traditional macro cells, they increase spatial reuse of precious licensed frequencies. Users get higher bandwidth and lower latency, with relatively less variance. This architecture, where small cells are deployed simply with power and Ethernet connectivity, lends itself well to cloudlet augmentation. In this position paper, we describe why deployed macro cell basestations are unsuitable for cloudlet deployment. In contrast, we describe why a small cell architecture is amenable for cloudlet deployments. Our experience from operating a small cell testbed in licensed frequencies matches that reported by equipment vendors. The applications we care about require high throughput and low latency. In a cellular network this can be achieved today by augmenting small cells with powerful cloudlets.

1. INTRODUCTION

With the advent of Apple Siri, Google Now, and Microsoft's Cortana, mainstream mobile computing is moving beyond conventional interactive computing. The computer reacts to its users' actions by proactive computing [17], where it continually senses and acts on their behalf. Today, this sensing is based on low data-rate sensors such as time, location, acceleration, and voice snippets. Higher data-rate sensors, especially vision sensors, promise to dramatically increase the semantic richness of the sensed data. But, vision algorithms for analyzing video in real time require a well-provisioned computing node, equivalent to a desktop machine of today [7]. Offloading-based solutions for vision have therefore depended on offloading from the smartphone or wearable to fixed infrastructure.

Researchers have recognized that many applications, which require fast response time for human interactivity, exceed the typical mobile computer's capabilities. The *cloudlet* model [16] addresses this problem by offloading computation to nearby compute nodes that are immobile and plugged into the electrical grid. It is imagined that Wi-Fi access points will be augmented with microprocessors that can perform computation on behalf of the phones associated to them. With the latest Wi-Fi standards such as IEEE 802.11ac, phones can expect upward of 100 mbps throughput and below 10 ms latency to this compute engine. Such network performance is sufficient for transmitting image and video streams to recognition algorithms running on the cloudlet and getting a timely response back to the user [6].

In contrast, cellular networks do not offer such network performance today. On relatively unloaded commercial LTE networks, users can expect roughly 70 ms median RTT latency with 20 ms jitter, and around 12 Mbps downlink throughput and 5 Mbps uplink throughput [9]. Furthermore, augmenting cellular basestations with general purpose compute requires an expensive change to the rigid architecture and infrastructure that supports cellular protocols.

A promising technology within the cellular standards is small cells (such as femtocells or picocells). The physical size and signal coverage of a small cell can be as small as a home Wi-Fi router. However, unlike Wi-Fi, small cells operate in licensed frequencies in conjunction with a pre-existing cellular operator's backend, relying on it for authentication, billing, roaming, and interoperating with the PSTN. The small footprint of a small cell compared to that of a macro cell (traditional cell tower) means that spatial reuse of spectrum can increase dramatically. Requiring only Ethernet connectivity and wall-socket power, and with SON (self-organizing network) support to adapt transmit power and channel selection, the deployment of a new small cell is dramatically cheaper and faster compared to provisioning a new macro cell. Even though the small cell will connect over IP to the cellular operator's backend, the standard allows it to access services on the local LAN directly. Hence, deploying a cloudlet in the small cell architecture is far easier than on the macro cell architecture.

In this position paper, we argue that the continuous vision workloads of the future require low latency and high throughput, near 30 Mbps, between a mobile device and a powered cloudlet. In unlicensed frequencies and uncoordinated deployments, Wi-Fi continues to be a viable technology for achieving this performance via deploying cloudlets at the access point. In licensed frequencies and coordinated deployments, we argue that the existing LTE macro cellular network is not sufficient, but LTE small cells are. A small cell can achieve close to 100 Mbps throughput with latencies around 10 ms, with far tighter jitter. Cellular operators are rapidly deploying small cells and the number of small cells overtook the number of macro cells in late 2012 [2]. Several major cellular operators are deploying them, including AT&T, China Mobile, France Telecom/Orange, Telefonica, T-Mobile/Deutsche Telekom,

MCS'14, June 16, 2014, Bretton Woods, New Hampshire, USA.

http://dx.doi.org/10.1145/2609908.2609946.

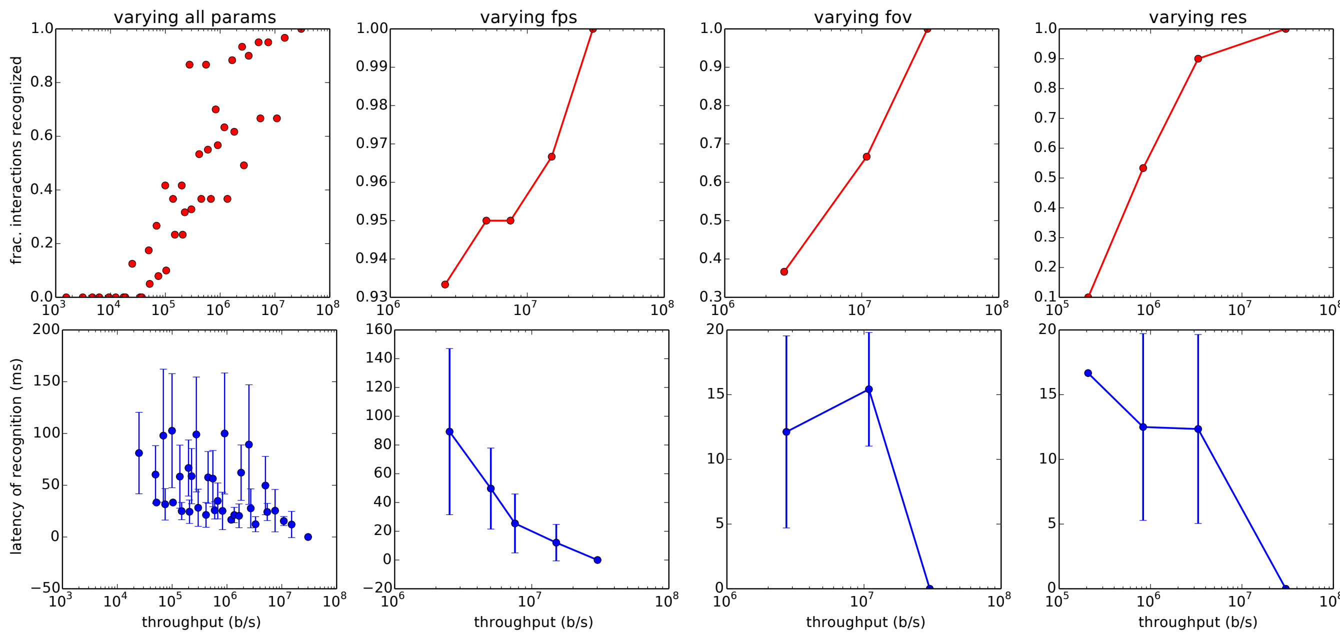

Figure 1: Face recognition performance in fraction of interactions recognized (top) and detection latency (bottom) against throughput (X axis), while varying (a) all parameters, (b) frames per second, (c) field of view, or (d) resolution.

and Vodafone [2]. Thus, while we continue experimenting with cloudlets on Wi-Fi, we need not forsake cellular networks.

2. CONTINUOUS VISION WORKLOADS

There are numerous proactive computing applications based on high-datarate sensing that we want to enable. Others [15, 16, 4, 6] have demonstrated the low latency needs of such applications. We additionally emphasize the need for high throughput between the mobile device and the computation. We do so in the context of one such application – face recognition.

2.1 Application setting

We wish to support applications that analyze continuous video streams produced by mobile devices such as wearables. This analysis should yield detailed information about the wearer such as who they meet, how they interact with them, what they do through the day, how they perform these activities, and where they go including the state of those spaces and the wearer's relative motion within them. Related computer vision algorithms in face recognition [3], object recognition [12] and location [11] are now getting to the point of achieving usably high recognition rates using large-scale machine learning techniques. However, applying them to video today is estimated to require a high-performance desktop-class machine [7].

Two of the most commonly proposed settings for continuous vision applications today are visual augmented reality and assisted services for daily life. Each poses significant challenges. Visual augmented reality requires the smooth visual augmentation of scenes by recognizing entities in the scene and superimposing graphical objects on them. Typically, the wearer makes an effort to focus and stabilize the camera with respect to the object of interest. Given that humans are sensitive to latencies higher than roughly 25-50 ms [5], this setting poses a serious *latency* challenge [15, 16, 4, 6]: all related communication and computation must happen within 20 ms. Such fast computation may require, for example, specialized circuit support [8], posing the question of how to provision such systems as support evolves.

Perhaps less appreciated than requirements on latency are those on throughput. Unlike the augmented reality setting, in assisted services, the wearer is often not focused on relevant entities (e.g., people around them or objects they are handling). The computer vision community has noted the importance of "lucky" frames in this setting, i.e., frames where the entity of interest happens to not move too fast, present a recognizable facet to the camera and be close enough to the camera so that sufficient pixels are available for resolution. Systems that have wide FOV (field of view), to account for lack of spatial focus, and high frame rate (to account for lack of motion synchronization) that maintain adequate resolution tend to be "luckier" than others. However, these requirements add up to high throughput requirements on the video. We examine these requirements below.

2.2 Experiment: throughput vs. recognition

Figure 1 shows the results of an experiment to study the importance of throughput. The goal of the experiment is to recognize *interactions* between the wearer of a camera and those around them. For our purposes, an interaction with a person is simply a sequence of video frames in which that person appears, such that no appearances of that person in the sequence are separated by more than 10s. Such a system may be useful, for instance, in triggering person-based reminders (e.g. "remember to tell Dave about the paper") or simply to name people approaching the wearer (e.g., "that's Dave"). In this setting, even when the interaction partner's face is clearly visible to the wearer, the vision system may not detect the face (or even more commonly, recognize who the face belongs to) because of slight variations in pose, speed of motion, lighting or resolution.

To generate Figure 1, we collected video from a chest-worn GoPro camera at 60 frames/s (fps) at 1940x1080 resolution with a medium-sized (120°) FOV. We used OpenCV to detect faces in the video, label the identity of faces, and cluster these faces per person into interaction segments. Our data includes 34 interactions with 7 people. We made versions of the video stream at various combinations of frame rates (f = 5, 10, 15, 30 and 60 fps), resolutions

(r = 640x360, 320x180, 160x90 pixels) and FOVs (36, 72, 120°, corresponding to fractions v = 0.3, 0.6 and 1 total FOV). We calculate the throughput T for each combination as $T = 24frv^2/100$ bps, accounting for 3 bytes per pixels and a compression factor of 100x via motion-compensated video encoding (our observed factor using H.264 was 99.6).

In each of the lower-throughput configurations, we applied face detection to the correspondingly degraded video. For each interaction, we verified if any face of the person being interacted with was detected (the fraction of interactions where at least one correct face was detected is the y-axis of Figure 1 (top)). When an interaction was correctly detected, we noted the number of frames after the start of the interaction when the first relevant face was detected (y-axis of Figure 1 (bottom) is calculated from this number times the time per frame). The figure shows the detection rates and delays (a) when all parameters are varied and ((b), (c), (d)), when a single parameter is varied keeping all others at maximum.

Even though the data is not very smooth, two trends are clear. First, recognition rates improve significantly with throughput all the way out to 30 mbps. As the upper figures (b)-(d) show, the improvement comes from all three parameters, but especially FOV and resolution. At low FOV and resolution, faces of interaction partners simply happen to not be detectable at lower speed, field of view and resolution in many of the interactions. Second, even with lower throughput, *when faces are detectable*, detection happens within 100ms or so, although high throughput further cuts this number noticeably. Looking at the lower figures (b)-(d) it is clear that frame rate is the dominant factor in detection latency. Not surprisingly, the higher the frame rate, the sooner interaction partners can be detected.

Two apparent anomalies in the figures are worth explaining. First lines in figure (a) are jagged because there are multiple ways to get a given throughput (i.e. varying FOV vs. fps vs. resolution), and some ways affect recognition rates less than others: lowering throughput by lowering frame rate is preferable to reducing the FOV. Second, in figure (c), detection delay is surprisingly low for the lowest FOV value. It turns out that at very low FOVs, few segments are detectable, but the ones that are detectable are relatively easy to detect.

We have focused in this experiment on detecting social interactions. However, we expect similar results whenever we seek to recognize any events that are not aided (explicitly or implicitly, e.g., via gaze) by the wearer to appear toward the center of the FOV, motion stabilized, or to appear close to the camera relative to its resolution. For instance, recognizing activities via object use and hand manipulation patterns, handled objects, the effect of conversational partners and text on ambient surfaces *especially when the wearer is not explicitly focused on these*, will likely benefit from high throughput analysis.

2.3 System design implications

Such applications that we want to enable can use latencies as low as 20-50 ms (for augmented reality applications) and throughput as high as 30 mbps. Even considering mobile SoC advances that we expect in the next few years, this workload will not be feasible on mobile devices at acceptable battery consumption. Remote computation will be needed. The network will need to support occasional bursts of transfers as high as 30 mbps, and latencies as low as 10 ms to allow the remote computation sufficient time to process video frames.

We do not expect video to be continuously streamed to this computation, but rather the mobile device will have limited computation to identify the short periods of interaction when recognition is needed. Most work on offloading vision algorithms so far has ignored the power cost of offloading. Given the cost of WWAN transmission of 700 mW to 1W, and the average battery budget of 700 mW of a modern phone with a 2000 mAh battery used over 10 hours, transmitting all data is impractical even if the bandwidth were available. It is inevitable that the duty cycle δ of transmission will be limited, e.g., to δ = 10% of the video data if 10% of battery life were budgeted for vision.

At those times, video at high frame rate, high FOV, and high resolution will need to be processed quickly. This remote processing will need to multiplex multiple users and applications. This may require appropriate isolation between the computation, and perhaps use of hardware acceleration. This could come from GPUs or custom circuits, leading us to consider the intriguing possibility that low-latency processing may be attached in a modular fashion.

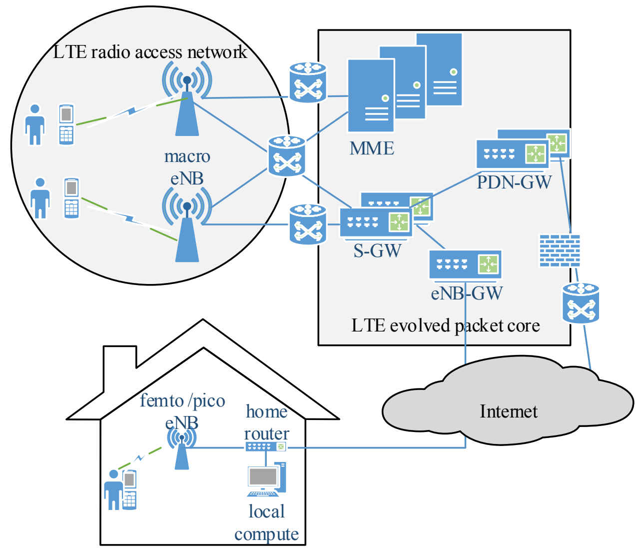

Figure 2: Simplified depiction of the LTE cellular architecture. MME = Mobility Management Entity; PDN-GW = Packet Data Network Gateway; S-GW = Serving Gateway; eNB = E-UTRAN Node B.

3. CELLULAR NETWORKS

Prior work on cloudlets has already identified Wi-Fi as appropriate networks for supporting such applications by augmenting the Wi-Fi AP with cloudlet computation. We now argue that for cellular networks, existing macro cells do not offer the performance today that we need for such cloudlet-enhanced applications.

3.1 LTE network architecture

The top of Figure 2 shows a simplified view of the LTE cellular architecture. Cell towers (macro cell eNBs) are typically connected via dedicated links to the cellular operator's backbone. There, they communicate with a number of other elements of the architecture. The MME helps with control plane signaling. Data plane traffic between the phone and the Internet is encapsulated in a tunnel that terminates at the PDN-GW. Once the user's traffic is decapsulated at the PDN-GW, it may traverse firewalls, NAT boxes, and proxy caches before it reaches the public Internet.

In this architecture, there exist several potential bottlenecks due to aggregation. Each cell tower typically has long range and multiple sectors, and hence serves many users at peak times. With limited licensed spectrum, the width of radio channels and the number

metric	median	25th %	75th %
DL throughput	12.6 mbps	7.6 mbps	19.7 mbps
UL throughput	5.5 mbps	1.9 mbps	11.2 mbps
RTT	71 ms	50 ms	98 ms

Table 1: Summary of crowd-sourced LTE performance numbers from Figure 5 of the 4GTest paper in ACM MobiSys 2012. The median, 25th percentile, and 75th percentiles are listed.

of resource blocks within those channels that are allocated to each user device can be small. This is typically referred to as the "spectrum crunch" and limits the maximum throughput available to each device. Signaling load on the eNB can exacerbate the problem by slowing down how quickly devices can connect and change radio power states. The long distance backhaul link from the tower to the packet core can also limit the aggregate throughput that users experience.

Prior work [9] has deployed a measurement application called 4GTest to measure speeds that users experience over LTE. We summarize their findings for LTE in Table 1. Despite their data being primarily from 2011 when fewer LTE handsets were in use compared to today and fewer deployed macro cells, users experience high latency to the Internet and wide variance in both throughput and latency. For the continuous vision workloads we want to enable, this performance is grossly inadequate.

To alleviate the problem of limited computation, memory, and/or battery on mobile devices, cloudlets [16] could be used. In this model, the mobile device will *offload* computation to the cloudlet, which has direct access to the electrical grid and hence energy consumption is of relatively minor concern. However, the applications envisioned that need heavy computation on mobile devices also require low latency in interaction with the user. In the LTE cellular architecture, the nearest location where such a compute node can be easily placed today is the cellular operator's data center, between the PDN-GW and the Internet. The performance bottlenecks that users experience on LTE are unfortunately closer to the mobile device than the cellular operator's data center, and hence there is little benefit to deploying a cloudlet there. Equipment vendors [13] are proposing "intelligent base stations" that can break out the traffic at the tower and provide computation there. However, that will require us to wait for these modified towers to be deployed, which may be a costly and time consuming proposition.

3.2 Small cells

To deal with the impending "spectrum crunch" brought on by a plague of faster smartphones and video streaming applications, several strategies are being pursued by cellular operators and standards bodies. Some cellular operators have deployed Wi-Fi hotspots and configured mobile OSes to prefer Wi-Fi whenever available. Wi-Fi networks are much more amenable to the cloudlet model as prior work has demonstrated [6]. Small cells are another strategy that cellular operators are rapidly embracing today. They improve spatial efficiency, are significantly cheaper to deploy, and can be installed trivially by a home owner.

We do not view small cell technology as a replacement for Wi-Fi. Each has its own advantages and disadvantages, and we expect both licensed and unlicensed frequencies to be employed in the foreseeable future for network connectivity. However, we find LTE small cells to have interesting properties distinct from Wi-Fi that warrant deeper investigation at the networking layer as well as at the application layer.

The bottom of Figure 2 shows the small cell architecture. Small cells are being deployed in a variety of situations – "femto" sized ones are installed in homes by users in coordination with the cellular operator, while "pico" sized ones are deployed in public spaces such as malls. Even here, the backhaul throughput can be a bottleneck, where we expect cable modem or DSL throughput to be in the range of 10-50 mbps. However, when augmented with a cloudlet, that bottleneck need not be a concern for our applications.

metric	value
DL throughput	∼110 mbps
UL throughput	∼10 mbps
RTT	∼11 ms

Table 2: Summary of small cell performance numbers reported by Huawei. The network was configured with 20 MHz channel width, 2x2 MIMO for DL, 1x2 MIMO for UL, and a LTE TDD subframe configuration where the vast majority of the transmission subframes were allocated for DL.

Given that small cells operate in licensed frequencies, only equipment associated with that licensee can operate in a given location. Interference is now limited to only those devices under such coordinated control. Access to the spectrum can be allocated in time and frequency domains, and power control and channel allocation is achieved across all devices that interfere with each other through the SON (self organizing network) part of the standard. In comparison to CSMA, this model can offer different latency and throughput capabilities in different operating conditions. Such coordination also allows for distinct QoS channels for different application classes across the entire network. Handoff from one cell to another can be streamlined. Radio sleep behavior can be made consistent across devices. This can come at the expense of additional coordination protocol overhead and a more complex architecture.

As shown in Figure 2, the core network continues to provide the same set of services that are offered via the macro cell. Voice calls (associated components not shown in the figure) and data traffic are tunneled via the eNB-GW to the same set of network components as with the macro cells. However, small cells can use LIPA [1] (Local IP Access) as defined in the standard to access services and computation local to the subnet in which they are deployed (such as in a home, coffee shop, or enterprise). Any traffic from the mobile device to local IP addresses is not tunneled to the core network but instead exits locally at the small cell's Ethernet interface, while any other traffic is sent over the tunnel to the cellular network core. Clearly, cloudlets can be deployed on the local LAN that the small cell is connected to and can provide an application experience that relies on high data rate processing. This mode of operation does not require any architectural change nor hardware change to existing equipment. While commercial small cells typically have an application processor, memory, and storage, those components are provisioned for running network management code and not the demanding cloudlet workloads such as vision. However, through the use of LIPA, that additional compute can be provisioned via a separate compute server on the same LAN without requiring hardware or software changes on the small cell (other than configuring LIPA).

Due to the small range of the small cell, a similar amount of licensed frequency as in the macro cell is now available to a smaller set of users. The throughput and latency that users can experience is significantly closer to the LTE specification. Qualcomm has demonstrated [14] peak throughput near 150 mbps to a phone from an Ericsson small cell with a 20 MHz LTE channel. Huawei has released more detailed measurements [10] that we summarize in Table 2. Our own research deployment of small cells in licensed frequencies demonstrates similar performance. Most evaluations tend to be focused on download throughput such as in Table 2, under the assumption that future workloads will primarily stream video content from the Internet. However, small cells can be re-configured for our workload – for instance in the case of LTE TDD, it can use subframe configuration 0 instead, where 3 times as many subframes are allocated for uplink as are for downlink. In such a configuration, our vision workload requiring upward of 30mbps can be supported with a wide margin, while still offering DL throughputs exceeding those of commercial macro cell networks.

Beyond the local cloudlet, a phone connected to a small cell has a different path to Internet servers than a phone connected to a macro cell. Internet-bound traffic will traverse the small cell's Internet connection (such as DSL or cable modem or metro Ethernet) to the eNB-GW in the cellular operator's network, and then reach the destination Internet server. Depending on where on the Internet the eNB-GW, Internet server, and small cell are situated and where network bottlenecks are, this latency may be lower or higher. However, we are primarily concerned with latency for continuous vision workloads that will run on a cloudlet, which will experience latencies similar to that listed on Table 2. We expect this interaction with the cloudlet will hide larger latencies to Internet servers.

4. SUMMARY

We predict that the current trend of proactive computing on small form-factor mobile devices will accelerate toward applications that require continuous visual processing. Such applications tend to produce large amounts of data that require significant computing power to process and, in some cases, demand extremely tight end-to-end latency.

Cloudlets are an important piece of the performance puzzle where computing is co-located with wireless access points. While researchers have focused exclusively on combing cloudlets with Wi-Fi, we examine the possibility of implementing cloudlets in cellular networks. We argue that while traditional macro-cell based cellular networks are currently not suitable, the newer small cell based cellular networks lend themselves well to the cloudlet paradigm.

Several avenues of future work are available to the community in this space. For example, continuously uploading 30 mbps of video will help users stay warm in cold climates but not for long. Admission control and gating mechanisms will be needed as the software on the mobile device searches for the lucky image frames, which require further processing at the cloudlet. Consequently, the network workload will consist of bursts of high throughput video uploads. This leads to open problems both in finding the lucky frames and network scheduling to satisfy hard deadlines while minimizing battery impact.

An approach one might purse is as follows: once the video data is available for uploading, the software on the device would wake up the radio and negotiate delivery with the small cell. In response, the small cell scheduler would dynamically adjust the LTE network frame length to accommodate the high upload throughput demand. How to do this in the presence of demands from other devices is an unsolved problem.

If we look beyond small cells and into cloudlets, there are several other problems we must solve. For example, for continuous vision recognition applications, we need to thoroughly understand the limits of today's image sensors and battery capacities. In addition to the high bandwidth demand issue, the constantly-streaming model implicit in today's proposed systems is impractical from a energy consumption and associated battery lifetime perspective. For augmented-reality applications, driving latency down to the sub-50 ms level will require careful system tuning and provisioning. For assistance-style applications, it is important to identify and validate scenarios where the constraints of today's recognition systems are acceptable. More broadly, it is poorly understood how vision-based systems need to be architected to scale to recognize subsets of the millions of places, people and objects that users may experience.

Note, we do not take on the question of LTE small cells for cloudlets versus Wi-Fi for cloudlets. Nonetheless, it is important to ask the question which is a better solution? Most modern smartphones come equipped with both Wi-FI and LTE capabilities and most indoor locations have Wi-Fi available. So the answer perhaps lies in evaluating performance against metrics such as which of the two provides better quality of service (latency and bandwidth) and better battery lifetime under a variety of workloads.

There is ongoing activity in standards bodies on using the LTE protocol in unlicensed frequencies, such as those that Wi-Fi operates in. In the future, will users rely on Wi-Fi for high speed wireless connectivity, or licensed LTE small cells, or unlicensed LTE small cells, or some combination? Are there inherent advantages to any one model such as power consumption on the phone or robustness to flash crowds that will make the difference in the end, or will it boil down to spectrum ownership and technology licenses? As a community we need to answer these and related questions before we can take a position on which of the two cloudlet solutions is better.

5. REFERENCES

[1] Architecture aspects of Home NodeB and Home eNodeB (Release 9). 3GPP TR 23.830 V0.50 Specification.

[2] Small cell market status, Feb 2013. Informa.

[3] DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In *CVPR*, 2014.

[4] E. Cuervo, A. Balasubramanian, D. Cho, A. Wolman, S. Saroiu, R. Chandra, and P. Bahl. MAUI: making smartphones last longer with code offload. In *ACM MobiSys*, 2010.

[5] S. Ellis, K. Mania, B. Adelstein, and M. Hill. Generalizeability of latency detection in a variety of virtual environments. In *Human Factors and Ergonomics Society 48th Annual meeting*, 2004.

[6] K. Ha, Z. Chen, W. Hu, W. Richter, P. Pillai, and M. Satyanarayanan. Towards wearable cognitive assistance. In *ACM MobiSys*, 2014.

[7] S. Han and M. Philipose. The case for onloading continuous high-datarate perception to the phone. In *HotOS*, 2013.

[8] Y. Hanai, Y. Hori, J. Nishimura, and T. Kuroda. A versatile recognition processor employing Haar-like feature and cascaded classifier. In *Solid-State Circuits IEEE International Conference*, 2009.

[9] J. Huang, F. Qian, A. Gerber, Z. M. Mao, S. Sen, and O. Spatscheck. A Close Examination of Performance and Power Characteristics of 4G LTE Networks. In *ACM MobiSys*, 2012.

[10] Huawei. LTE Small Cell v.s. WiFi User Experience. http://www.huawei.com/ilink/en/download/HW_323974.

[11] K. Konolige, J. Bowman, et al. View-based maps. In *Proceedings of Robotics: Science and Systems*, 2009.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, 2012.

[13] NokiaSiemensNetworks. Intelligent base stations. white paper.

[14] Qualcomm. Qualcomm Demonstrates 150 Mbps Category 4 LTE on MSM8974/Snapdragon 800. https://www.youtube.com/watch?v=vphn48XZqgA.

[15] M.-R. Ra, A. Sheth, L. Mummert, P. Pillai, D. Wetherall, and R. Govindan. Odessa: Enabling interactive perception applications on mobile devices. In *ACM MobiSys*, 2011.

[16] M. Satyanarayanan, P. Bahl, R. Caceres, and N. Davies. The Case for VM-based Cloudlets in Mobile Computing. In *IEEE Pervasive*, 2009.

[17] D. Tennenhouse. Proactive computing. *Commun. ACM*, 43(5):43–50, May 2000.