PicFlic : Social Discovery Thru Pictures

Venu Vasudevan, Jehan Wickramasuriya,
Joe Wodka, Myungcheol Doo
Applied Research
Motorola Mobility
{venu.vasudevan, jehan, joe.wodka, myungcheol}@motorola.com

ABSTRACT

The ubiquity of smartphones is driving growth in the use of cameras as visual sensors of the things around us, and towards the sharing of 'photos of things' as a mechanism for the social discovery of physical things. The combination of this user trend with increasing support for device-to-device protocols has the potential to rejuvenate the space of proximity-based social applications. Enabling this future requires solutions to systems issues around creating effortless and efficient sharing of media across large collocated groups. This paper enumerates the platform and user experience challenges in creating such effortless, optimized and anonymity preserving experiences for proximity-based photo sharing, with social discovery as a key user benefit. It presents our efforts building PicFlic, a photo sharing system, based on these principles.

Keywords

User Interfaces, social context, social influence, recommender systems

1. Motivation and Background

As mobile phones have become ubiquitous, they have morphed from merely convenient cameras that are used to take pictures of close friends and loved ones, to visual sensors of our lives. The pictures in our mobile photo galleries now not only encompass the highly personal Christmas gift unwrapping or a funny group picture on the beach, but also photos of our favorite hamburger, the awning of a destination restaurant, the cruise boat we took to our island destination, and the LCD TV we have on our wish list.

A number of marketplace applications are either spurring or riding this trend towards mobile camera as visual sensor. Popular marketplace applications around **visual processing** such as Google *Goggles* for object identification and *WordLens* for language translation, have spurred the use of cameras to take pictures of objects, with some form of visual decision making as the accrued user benefit. Other **personal visual landmark** applications such as *Fooducate*, *Dishpal,* and *Stamped* indicate a trend towards *pictorial* and *social* discovery of commercially valuable 'things' (food, restaurants, books) via consumer-to-consumer taste sharing.

While these cloud-based **social commerce** applications target wide-area social discovery, another opportunity exists in proximity-based social discovery that is distinct from the wide-area opportunity in both user value proposition and associated system challenges in realization. From a user value proposition, proximate & photo based sharing experiences have the social discovery elements of their wide-area variants, but could potentially be higher trust and higher quality, and more conversational. A favorite hors d'oeuvre photo shared at a highly sought after bar or lounge has higher credibility than the same picture shared in a location agnostic manner. The fact that the sharer is at this unique and sought after location boosts the credibility of the sharer, and the shared photograph. The follow on experience to a social photo discovery can also be richer, as it allows for the possibility of range of interactions, from more private texting or instant messaging to face-to-face conversation (assuming the sharer and sharee opt-in to such conversation).

System challenges in an effective realization of such a system comprise of both platform and user experience issues. Creating an engaging **experience** in physically crowded and socially active locales such as piazzas, restaurants, and vacation spots, requires an interface that is *glanceable*, *high cognitive bandwidth*, *agile,* and *elastic*. *Glanceable* interfaces enable users to efficiently flip between physical and digital experiences in socially active locales. *High cognitive bandwidth* enables the system to collect a large and diverse set of photos that are sufficiently representative of the entire collection of people in the vicinity, not just a small and statistically skewed sample. *Agility* implies that the corpus of photos is collected rapidly, so that a user who flips open the application immediately sees a real-time collection of photos that is rich, diverse, engaging, and encourages interaction. The interface needs to be *elastic* in its support of interaction modes – so that a user can cycle from observe to interact to communicate in a very small number of clicks.

On the **platform** front, the challenges are to build a premium experience such as that described above while also ensuring *scalability*, *deployability,* and *energy efficiency*. On *deployability* - Bluetooth is fairly widely deployed on mobile handsets, but has severe limitations as a middleware substrate for proximity-based experiences in terms supporting large groups (tens to hundreds of devices) without complex forms of group federation. Wi-Fi Direct [12] has support for larger and more energy efficient forms of federation, but is an emerging standard that isn't widely deployed. In both cases (Bluetooth and Wi-Fi Direct), tightly coupling the application interface to one of these protocols increases the complexity of the programmer API while also limiting the deployability of PicFlic to only those handsets that support the particular protocol. Multicast DNS (mDNS) is elegant and function rich in terms of device and service discovery, and widely supported on both Apple's iOS and Google's Android platforms, but does not directly support adhoc networking. However, it is our belief that overlay protocols atop mDNS can be an effective approach to both scalability and deployability – supporting low latency and proximity-based experiences in non-adhoc networks while also being able to layer on (and therefore buffer an application programmer from) either Wi-Fi Direct or other adhoc networking substrates. The other platform challenge is in

978-1-4244-9529-0/13/$31.00 ©2013 IEEE

balancing experience *agility* against platform *energy efficiency*. Here we believe that naïve approaches to content discovery can maximize content availability but at the expense of high network traffic and high battery consumption. We share the fact that some preliminary work on a collaborative caching protocol is efficient in the use of bandwidth, energy, and time to collect an initial photo experience. The protocol integrates diversity measures to ensure that the photo gallery is diverse enough to engage a user who enters a new physical space.

2. Related Work

There has been a large body of work in proximity-based social networking, of which this brief section points to a few that are most pertinent to the topic of this paper. Reagans [7] draws a positive correlation between physical proximity and strength of influence, a point that is relevant to systems such as PicFlic that combine proximity and social discovery. TWIN [5] is a longitudinal user study (250 users, 9 weeks) that demonstrates user acceptance of proximity-based socialization. BlueTunA [3] and Push!Photo [2] show the value of social discovery and social presence across digital media types. Work on Urbanhermes [4] shows the potential of proximity-based social discovery of sellable things in the domain of unique and expensive fashion accessories. More recently, Masubi [13] has presented a social sharing platform enabling users to share and interact with friends in a very secure and infrastructure-independent way. The work here focuses heavily on established secure channels of communication (utilizing trusted group communication protocols) between peers using a UI, which is predicated on the use of "feeds" to present information.

On conversational content - various projects address the need to mix content and communication in a lightweight fashion to build systems that garner adoption. Photochat [8], Lock-on-Chat [10] and Emotipix [9] all point to the value of creating visual conversations around photos, and the importance that the conversational primitives be a micro-grammar that is small and lightweight. Flipper [1] and Emotipix [9] also underscore the need for the photo experience to be lightweight, even ambient to see extended use and avoid user fatigue. Work in TWIN [5] and Digidress [6] show user abandonment in situations where the content experience isn't quick, interesting, and fresh.

Last but not least, there is growing commercial interest in Social Discovery. The notable growth of membership and activity in commercial social networks such as Pinterest and TheFancy [11] points to the increased use of pictures of 'things' for social discovery. TheFancy's business model of taking a 10% cut of the revenue generated from purchases that are inline with social discovery, points to the possibility that money could be an added incentive to drive users to participate on mass scale.

3. The PicFlic System

The design of PicFlic tries to balance a fresh and immediate pictorial experience, with an architecture that is deployable across both infrastructure and adhoc network environments. Below we describe key elements of the user experience, and elements of the system architecture that enable the system to scale up to large numbers of devices, and support flexible mixes of content and communications.

3.1 Design & Experience

The user experience is designed to support varying levels of user engagement with PicFlic by providing an engaging ambient experience, and a richer set of social discovery features accessible within 1 to a few clicks (see Figure 1).

The **0-click** PicFlic *photo zeitgeist* experience is one in which a user invoking PicFlic will automatically see a gallery of diverse pictures from nearby devices. Additionally, there are community activity indicators representing the intensity of click activity from other PicFlic users, and the pictures that are generating interest.

1-click retrieves a larger view of the picture, and supports a pre-conversation mode. The pre-conversation mode enables users to get metadata about the picture such as the camera information, location information, and other associative metadata about the subject in the picture that can be context to a conversation around it.

A small number of added clicks trigger a *visual conversation* with picture taker such as – asynchronous ability to ask bounded vocabulary questions about picture (e.g. cheapest way to get to the vacation spot on the picture, or the best cruise line to an island destination).

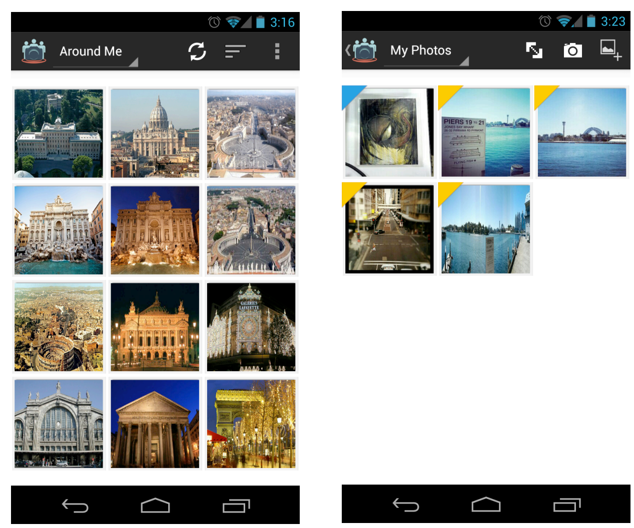


Figure 1: PicFlic UI showing photos in the vicinity ("Around Me") as well as the sharing interface for user photos ("My Photos") with tag indicators

3.2 System Architecture

This section covers two aspects of the PicFlic architecture that have been discovered to be of importance by us and others (see related work) – and the ability to efficiently create a PicFlic experience in widely varying network sizes and differing network dynamics, and the ability to easily transition from ambient experience to visual conversation and back. We optimize for energy by layering service discovery and meta-data exchange on top of mDNS, and by supporting collaborative caching to reduce the average number of service invocations per device to assemble the proximity-based picture gallery. We simplify the transitions between browsing and visual conversations by using small and contextually activated conversational grammars.

The base PicFlic architecture is shown in the Figure (below) and has 3 components – an mDNS based middleware for automated federation of photo services across proximate groups of devices, storage management that supports collaborative caching to enhance system performance, and services that enable and support flexible visual conversations. The choice of mDNS as the middleware substrate is pragmatic, as it is supported by both Apple iOS and Android mobile devices as well as by most Wi-Fi access points. Using mDNS service advertisements enables sizeable populations of devices to effectively find and interface with photo sharing service addresses neighboring devices. Each device autonomously decides the set of mDNS service end points to interface with, and retrieves a representative set of proximate photos.

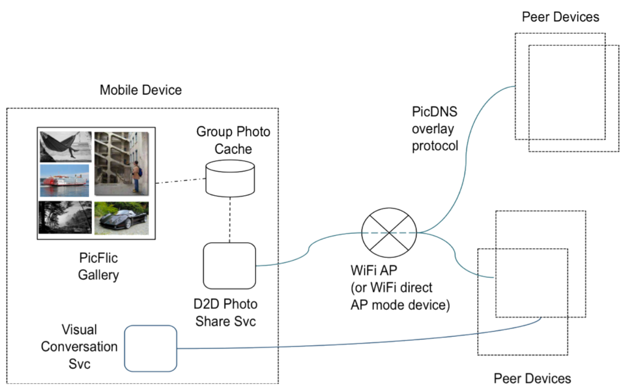


Figure 2: PicFlic architecture

In environments where devices support the mDNS TXT field, we operate PicDNS, an overlay protocol on top of mDNS that *pushes* a collection of associated metadata along with the service advertisement. PicDNS enables device-side filtering of the appropriate group of photos on a client-specific basis. Unfortunately, current Smartphones (both iOS and Android) that do not allow TXT field manipulation even though it is part of the mDNS standard, which then requires clients to *pull* metadata from the surrounding device group on a device-by-device basis. This change from push to pull can negate many of the benefits of layering a proximity-based system over multicast, without added elements in the architecture. In section 3.2.1, we discuss the addition of collaborative caching to enable even pull-based meta-data and picture thumbnail synchronization over large device groups in a manner that grows the execution time for a complete sync in a group of N devices in O(NlogN) rather than O(N^2).

Several past efforts on photo sharing have pointed out the need for conversational overlays as a community formation primitive atop photos. Unlike past work, it is our contention that conversational micro-grammars are likely to be topic-specific (e.g. food vs. vacations vs. favorite book). We believe that the conversational framework needs to be overloadable in a manner consistent with typical discourse around the social object in the picture. Section 3.2.2 provides a brief view into the proposed framework for visual conversation services, and its integration into the user experience.

3.2.1 Collaborative Caching

As mentioned earlier, our design goal for PicFlic is to scale in dimensions of group size as well as group dynamics. We would like to support an engaging PicFlic experience in environments ranging from a café (tens of people), to an office (hundreds) to a conference (potentially thousands). Orthogonally to group size, we would also like to deal with group stability that may vary from stable groups with small and predictable changes in the quorum of proximate devices (e.g. a typical office), to flash mobs (e.g. everyone arriving at the start of a conference). To create interesting visual experiences even in highly dynamic groups, we aim to fill the first PicFlic screen on a tablet with 24 photos relatively quickly, while also maintaining photos from enough different neighbors for the gallery to be representative of the surroundings.

The design goals are to accomplish such scalability on both WiFi networks that support mDNS as well as emerging device-to-device adhoc networks that are constructed over WiFi Direct. In the former, there are packet traffic limits for mDNS packets and in the latter, there are both power and complexity constraints in federating large groups of WiFi Direct devices.

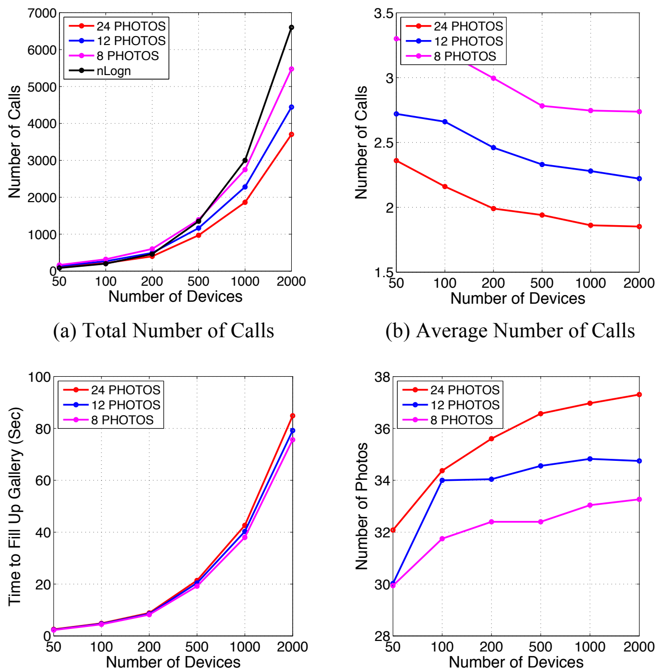
To determine the efficacy of collaborative caching in supporting energy efficient photo retrieval, we've conducted simulations of collaborative caching of both photos and photo meta-data with different synthetic device arrival configurations and varying payload sizes in the data exchanged between devices.

For the purposes of simulating an access point based mDNS environment (e.g. an office), we assume that each place has one 54Mbps WiFi access point and that a maximum of 30% of this bandwidth can be used for mDNS traffic. Each device shares at most *M* number of photo including two of its own photos, to keep one or a few devices from dominating the photo experience. We assume service invocations in the collaborative caching system to be symmetric (i.e. there is a bidirectional exchange between caller and callee of any pictures they possess). A device randomly selects a target and exchanges photos from proximate devices, with the target up to the maximum number of photos that can be transferred as shown in Table I. For example, given a device, *P,* who has two of its own photos and a device, *Q*, who has two of its own photos and 50 photos of others, *P* sends one of her photos to *Q*. Then *Q* sends 24 photos back to *P*, if the maximum number of photos transferred is set to 24. Like *P* and *Q*, each device repeatedly invokes the photo service on a diversity of nodes, until every device has filled the initial gallery screen with 24 photos.

Parameter	Values
Num. of Devices	50, 100, 200, 500, 1000, 2000
Max. Num. of Photos Transferred	8, 12, 24
Max. Network Bandwidth	30% of 54Mbps
Size of Photo Thumbnail	(2K + 10% Overheads) bytes

Table I. Simulation parameters and their values

Figure 3 shows the simulation results while varying the number of devices and the maximum number of photos transferred. Our modeling shows that collaborative caching can be done less than linear logarithmical calls while increasing the number of devices as shown in Fig 3(a). Initially every device has two photos only. As photo exchange proceeds, some devices have enough photos to fill the initial gallery. If a device, *R,* calls these devices, then *R* may finish filling the initial gallery up with only one photo service call. Also one call of exchange increases the number of photos in two devices. Therefore, on average, two or more calls, which is approximately (Number of Slots in Home Screen / Max. Number of Photos Transferred), are enough for fill up the initial gallery as shown Fig. 3(b). Even in the extreme case of flash mobs of thousands of devices, every device will have enough photos to populate the PicFlic tablet screen of 24 pictures with sufficient diversity (no more than 1 picture per owner) in about 80 seconds as shown in figures Fig. 3(c). When a device calls another, we do not specify how many photos are needed. Called devices send back up to the maximum number of photos transferred. Although devices no longer call others once it has 24 photos or more, unfinished devices may feed finished devices by calling them. As a result, devices may have more photos to be needed for filling the gallery as shown in Figure 3(d).



(a) Total Number of Calls (b) Average Number of Calls

(c) Time to Fill the Initial Gallery (d) Average Number of Photos

Figure 3. Simulation Results

3.2.2 Visual Conversation Services

A number of previous efforts [8, 9, 10] have alluded to the value proposition of photos as social objects with associated conversational overlays. They have proposed specific conversational primitives around commenting, annotation and chat. In the context of social discovery (as opposed to generic photo browsing), it is our belief that the conversation has to be focused on making *decisions* around the topic of the photo (e.g. 'what cruise line works best' applies only to tag 'vacations' not 'food'), as opposed to being free from and informational. Scoping the conversational micro-grammar to the focus object enables the conversation to be lightweight (even 1-click) and optimized (e.g. use of cached answers to repetitive questions). PicFlic currently provides a very limited set of tags to influence users to publish photos of specific kinds of things, and to support conversational micro-grammars around them. We are cognizant of user reluctance to tag photos (a prerequisite for conversation scoping) and are looking into heuristics short of perfect topic recognition as ways to partially automate the process of associative tagging.

Another feature to maintain lightness of the PicFlic experience is to provide mechanisms to decouple the discovery of an interesting photo from the triggering of a conversation around it. *Rubberbanding* extends a proximity-based encounter over time, thus enabling a user with an interest in a photo to 'pin' it into his gallery (along with contact metadata), but frame questions to the owner of the photo at a later point. This expands the conversation to when the asker and the answerer are not collocated.

4. Prototype Evaluation

We conducted a prototype evaluation where we solicited feedback from 18 users on the utility, design and privacy aspects from both a personal and social perspective. While our narrow focus from a user needs perspective was picture-based social discovery, we were also looking to mobile photo sharing of objects that didn't necessarily (or immediately) lead to a product decision.

From a *utility* perspective, we were looking to understand the bounds of both sharing and exploration. For sharing, we wanted to understand the key sharing moments and locations, along with what would make a user more (or less) comfortable with sharing more expansively. In terms of exploration, we wanted to examine whether users would be likely to use this as an "always-on" application and subsequently whether it would keep users engaged in consistent use scenarios.

From a *design* perspective, we were trying to gauge the balance between simplicity and expressive power as it pertained to both browsing and sharing. Rather than ask direct questions about *privacy* (which often leads to visceral answers), we couched privacy in terms of user reaction to the opposing forces of *ego casting* (wishing to see their photos acknowledged) and paranoia (concern that the photos would be misused). Key insights from the study are described below.

Utility. Serendipity, or Rejuvenation?

In our efforts to understand how social discovery maps to people, places and things, we asked users both free form and multiple choices questions on the most (and least) likely scenarios of use. Users gave us strong positive feedback on the use of PicFlic in *unfamiliar and unique places* where the user could leverage *pictures* to make *decisions*. In particular, users thought PicFlic would be very useful in restaurants, art fairs, and sports venues. These were new locations where there was time pressure to make decisions, and pictures of things purchased or noticed by the people around you could be helpful both indirectly (as Zeitgeist) and directly (an "I'll have one of these" food choice) as social discovery cues. User feedback on *rejuvenation* of a familiar place (i.e. noticing the unfamiliar in a familiar place) was mildly but not overwhelmingly positive. Users felt that they would likely follow their routine in a familiar place, and less likely to be in an exploratory mode.

Users were very enthusiastic about the ability to ask discreet questions about pictures of interest. However, they felt that the ability to have a conversation around a picture should not be restricted to real-time collocation. This feedback provides support for visual conversations with *rubberbanding*, whereby a photo shared at a location could become the topic of a more persistent conversation.

Design. Balancing Glanceability and Expressive Power

We asked users a number of questions to understand dimensions of passive content consumption vs. active conversation, as well as the extent to which community cues might be of value in creating and retaining engagement. User feedback pointed to the need for a layered interaction design with both ambient and conversational interfaces. Users expected a PicFlic to support an effortless and ambient photo consumption interface, from which a user could launch a conversational interface if anything caught his or her attention. While community activity indicators got some positive feedback, users wanted a prominent egocasting interface that showed hits or shares of pictures they had taken.

Users strongly supported a conversational interface over content, and were interested in a diversity of (anonymous) conversational interfaces for different kinds of photos. Support for persistence of both photos and the conversation about them was of high interest. There was feedback and discussion on conversational interfaces needing to be almost invisible until called upon, so as not to come in the way of an ambient and always-on 'idle screen' pictorial experience. Our general takeaways here were that the layered design (separation of 0, 1 and 2-click experiences) resonated with users, but that the community interfaces needed to be rethought with ego casting and simplicity in mind.

Privacy & Sandboxed Automation

Depending on the sharing context – pictures can be evocative and gratifying to share, or incriminating and a privacy concern. Given that perfect privacy and zero sharing friction are opposing needs – we wanted to understand the nuances of automation that would reduce the effort in sharing to where PicFlic could be an always-on experience, while at the same time avoiding a privacy faux pas. In particular we wanted to understand the relative value of restricting sharing based on content metadata, place (e.g. Geo-fencing policies for sharing) or time (e.g. temporal sharing policies and automated expiration of access rights to photos).

Users were very open to sharing pictures of things (as opposed to people). As long as they could trust the system to filter out pictures of people, users found it entertaining and gratifying to have their pictures transitively shared by other users without complicated spatial and temporal constraints. In addition to automated filters, users favored the notion of share folders as explicit sandboxes – so that contextual sharing still took place within a photo corpus that the user explicitly designated as shareable.

5. Conclusion

This paper introduced the potential of proximity-based photo sharing for social discovery, and identified the challenges associated with it. Our prototyping demonstrated the feasibility of these architectural ideas, and raised new issues in their pragmatic realization. On the latter, we've prototyped PicFlic on Android devices running ICS and above, tested it on Galaxy Nexus/Nexus 7 & Motorola devices running Android ICS (XOOM-WiFi, RAZR) and performed limited trials with users on the application experience and efficacy. Ongoing and future work includes further longitudinal users trials, and follow-on redesign based on the user feedback outlined herein. On the systems side, we plan to examine energy efficient realizations that can support an always-on experience, and refinements to the layered UI design proposed in this paper. Additionally we are looking into more intelligent and automated forms of sharing that enable opt-in control from the user but can potentially automate operations that currently require manual effort.

6. References

[1] Counts, S. and Fellheimer, E., Supporting Social Presence through Lightweight Photo Sharing On and Off the Desktop, In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* CHI '04, pp.599-606.

[2] Rost, M., Jacobsson, M., and Holmquist, L. E. (2006) *Push!Photo: Informal Photo Sharing in Ad-Hoc Networks*. In *Adjunct proceedings of UbiComp 2006.*

[3] Baumann, S. et al., BluetunA: Let your neighbor know what music you like, In CHI '07 Extended Abstracts on Human Factors in Computing, pp.1941-1946.

[4] Christine M. Liu and Judith S. Donath. 2006. Urbanhermes: social signaling with electronic fashion. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (CHI '06), pp.885-888.

[5] Kaisa V., et al. User experience of social ad hoc networking: findings from a large-scale field trial of TWIN. In *Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia* (MUM '10).

[6] Per Persson, Jan Blom, and Younghee Jung. DigiDress: a field trial of an expressive social proximity application. In *Proceedings of the 7th international conference on Ubiquitous Computing* (UbiComp'05).

[7] Ray Reagans. 2011. Close Encounters: Analyzing How Social Similarity and Propinquity Contribute to Strong Network Connections. *Organization Science* 22, 4 (July 2011), 835-849

[8] Yasuyuki Sumi, Jun Ito, and Toyoaki Nishida. 2008. Photochat: communication support system based on sharing photos and notes. In *CHI '08 extended abstracts on Human factors in computing systems*(CHI EA '08), pp.3237-3242

[9] Lisa Cowan, William G. Griswold, Louise Barkhuus, and James D. Hollan. 2010. Engaging the Periphery for Visual Communication on Mobile Phones. In *Proceedings of the 2010 43rd Hawaii International Conference on System Sciences* (HICSS '10)

[10] Nishida, T and Igarashi, T. Lock-on-Chat : Boosting Anchored Conversation and its operation at a Technical Conference, INTERACT 2005, Springer LNC3583. pp970-973.

[11] Apple Fancies TheFancy -- but can it make it work?, http://bit.ly/QxDiLB

[12] Wi-Fi Direct Alliance, http://www.wi-fi.org/discover-and-learn/wi-fi-direct

[13] Dodson B., Vo I., Purtell J., Cannon A., and Lam M.S. Musubi: Disintermediated Interactive Social Feeds for Mobile Devices. In WWW2012 Lyon, France, April 2012