

Rethinking Camera User Interfaces

Stephen Brewster*, Christopher McAdam, James McDonald and James Maciver
Glasgow Interactive Systems Group, School of Computing Science, University of Glasgow,
Glasgow, G12 8QQ, UK

ABSTRACT

Digital cameras and camera phones are now very widely used but there are some issues that affect their use and the quality of the images captured. Many of these issues are due to problem of interaction or feedback from the camera. Modern smartphones have a wide range of sensors, rich feedback mechanisms and lots of processing power. We have developed and evaluated a range of new interaction techniques for cameras and camera phones that improve the picture taking process and allow people to take better pictures first time.

Keywords: Cameras, camera phones, human-computer interaction, audio feedback, tactile feedback, user interface, luminance histogram

1. INTRODUCTION

Mobile phones are now ubiquitous in our daily lives and typically come with a camera built-in. Over time, the quality of the built-in cameras has increased to a level that many now use them as their primary camera, particularly as most people carry a phone with them all the time.

User interfaces for camera phones are largely based upon those of standard digital cameras, which are themselves based on the film cameras that preceded them. They do not take full advantage of all of the sophisticated capabilities of modern smartphones for providing richer interactions for picture taking. Smartphones feature sensors such as gyroscopes, accelerometers and magnetometers. They also have powerful processors that can be used for real-time image processing. Many have high quality audio hardware for playing music and vibration motors to generate tactile cues. Combining these outputs with the rich sensing means we can change the way pictures are taken.

There are usability problems with cameras that may cause difficulties while taking photos and lead to poor quality images, e.g. poor exposure, blurred pictures, flash switched off, etc. One reason for this is that the LCD is used to compose the image and also to present information about the current settings and status of the camera. Small icons are presented around the edges of the display, see Figure 1. This is problematic for two reasons. Firstly, when framing an image attention is often on the scene you are taking a picture of, making it possible to miss these icons. Secondly, every icon displayed obscures part of the image, so, as the number of icons increases, it becomes harder to see the entire frame. These icons can be hidden but that denies the user the information they provide.



Figure 1. Complex camera user interfaces including luminance histograms (bottom right corners). Left is from a digital camera (www.dpreview.com) and right from an Android smart phone (www.android.com).

*stephen.brewster@glasgow.ac.uk; phone +44 141 330 4966; fax +44 141 330 4913; www.dcs.gla.ac.uk/~stephen

We have developed a range of novel interaction techniques that run on Android smartphones that improve the picture taking process by taking advantage of the capabilities of the phone. The aim of our work has been to understand the problems faced by photographers and then to develop new ways of interacting with the camera and presenting information to address the issues. Some of these reduce the usage of the LCD by giving audio or tactile feedback to allow the photographer to concentrate on the image being taken. Others do use the visual display of the camera but augment it to help the user take better pictures without overloading the display. This paper presents an overview of some of the techniques we have developed.

2. BACKGROUND

Research has shown that users often do little editing of photos after the shot has been taken [1]. Rodden and Wood [2] found that people, when organising their photographs, would separate the good photos from the bad, sometimes discarding the bad ones completely. The ‘bad’ photographs were considered such either because they were uninteresting or because they were “technically poor”. Their users generally did little editing after the photo had been taken. Many cameras and camera phones allow images to be uploaded to social networking sites immediately after shooting, so there is little scope for correcting poor images. We need to help users take the best photos they can first time by rethinking the user interface of the camera to improve the picture taking process.

Kindberg *et al.*[3] have studied the motivations behind camera phone photos. Interviews and grounded discussions with existing camera phone users identified the diverse uses for camera phone images. One of the key areas of dissatisfaction amongst their users was the quality of photos captured. They also studied the content of the images. Around half of the photos randomly selected for discussion contained people. These photos were identified as containing family or friends and, in some cases, the users themselves. This suggests that people use their camera phones to take photos which might benefit from assistance in framing people as part of the shot.

Bourke, McCarthy and Smyth [4] attempted to address the lack of support for composing good photos. A novel interface was implemented with a recommendation system to assist users in selecting and framing a shot. GPS information was used to display photos that had been taken by other people nearby. Whilst this may help users to replicate previous photos taken by others, it does not offer interface support for taking good photos in the first place, which is the aim of our work here.

3. SURVEY

To identify some of the common problems people have when taking photos, we conducted an online survey of picture taking with cameras and camera phones. Of our 62 respondents, 53% used a camera phone as their primary camera. Key problems reported were: photos being blurry (80%), under/over exposed images (75%), photos not level (68%), incorrect flash setting (59%) and missing shots because the memory card was full (40%).

We also enquired about the usage of different camera features. 67% of respondents said they used the self-timer for photos (if they had it on their camera), with 51% using the luminance histogram to judge exposure. 46% of people said they took photos with the camera held above their heads and 47% took photos of themselves with the camera held facing them. 12 people reported framing problems when taking pictures in this way. These problems motivated the new interaction techniques we developed to improve the picture taking process.

4. LUMINANCE HISTOGRAM

The first of our new interaction techniques was to make better use of the luminance histogram to improve exposure. Luminance histograms show the tonal range of an image. The x-axis shows the different tones, ranging from black on the far left to white on the far right. The y-axis shows the number of pixels with that value. The histogram is a key feature for determining the exposure of an image and is better than looking at an LCD screen as these often cannot present the full tonal range of an image. Histograms have to be fairly large to make the information clear, obscuring some of the image on the display (see Figure 1), which can result in them being switched off. Some cameras present a limited version of this information by flashing the parts of the image that are over/under exposed. However, this flashing can be distracting to the user so is most often shown after the photo has been taken, which may be too late. We need to help photographers take a correctly exposed image rather than to see what needs to be corrected after it has been taken.

We developed a simple audio cue to represent exposure in sound. Our feedback takes the form of two short notes representing the (0-5) and (250-255) endpoint ranges of the histogram. The pitch of the tones corresponds to the number of pixels at those endpoints; a low number of pixels gives a low note and a high number a high one. For a well exposed image the feedback should be two low notes, as the histogram is mainly centred on the mid-tones. An underexposed image has a high note followed by a low one; the images being high in shadows (the left side of the histogram) and low in highlights (the right of the histogram). Overexposed feedback would be a low followed by a high note. Two high notes indicate that the image has exceeded the dynamic range of the sensor. The aim of this feedback was to give a simple overview of the current exposure. In previous research [5], we developed a complete audio histogram but this was hard to interpret for novice users.

The audio histogram was presented on half-press of the shutter button, replacing the standard focus lock sound (which is two short beeps on many cameras). As it was presented on half-press and before the image was actually captured, the user could 'probe' the scene by tapping the shutter button until a good exposure was heard and the shutter fully pressed to take a correctly exposed picture, all without cluttering the LCD. A user study showed that people could easily interpret the audio histogram presented this way and could correct their exposures with this feedback.

5. STABILITY

Camera movement during picture-taking can cause blur. The icon of a hand shaking often indicates that the image may be blurred (Figure 1). However, there is no real-time feedback that presents actual camera movement to the photographer. If the user has steady hands he/she may be able to take sharp photos even when the icon is being shown. Alternatively, if the user is taking a photo from a bumpy car then the picture may be blurry even before the hand icon is shown. Our approach presents photographers with feedback to indicate how steady they are holding the camera. For them to have confidence that they are holding the camera steady, the feedback must be presented continuously.

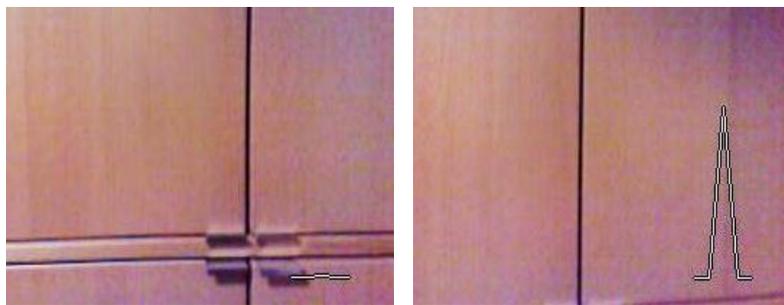


Figure 2: Graph in the bottom right hand corner shows camera movement in real-time, with little movement on the left and more on the right.

We use the accelerometer in the phone to measure its movement so that we can then present this to the user. We developed and evaluated several different types of feedback for this. The visual feedback takes the form of a small line graph in the bottom right corner of the display (small enough to avoid obscuring too much of the image in the LCD). When the phone is being held steady, the line is flat (Figure 2, left). As the phone becomes more unsteady, a peak starts to rise from the middle of the line based on the accelerometer data, representing the amount of movement (Figure 2, right). This motion is clear in the users' peripheral vision, and allows them to be informed of the feedback without taking their focus away from composing the image. The feedback is presented on half-press of the shutter button.

We also developed audio and tactile versions of this feedback to help users when holding the camera above their heads to take photos, for example. The audio feedback consisted of a short sound cue. When the phone is held steady, the cue is played repeatedly at a steady rhythm. The cue is played quietly, so as to not distract the user or draw attention from those around. When the phone starts to become unsteady, the cue is changed using the accelerometer data so that it is louder and there is a longer delay between repetitions making it sound more arrhythmic, indicating to the user that he/she is shaking the camera too much. In evaluations, users were easily able to understand the feedback and were able to understand how they needed to steady themselves to reduce blur.

6. MOTION DETECTION AND PANNING

In the previous section we dealt with movement of the camera causing image blur but there is also an issue with movement of the subject of the photograph. The photographer might use this for artistic effect, but it may be just a poor image. The small size of the camera display and the fact that the users will be focused on other things, such as alignment and exposure, can lead to them not noticing subject movement. We use simple image processing techniques to detect movement in the scene and give feedback on it to the user.

Feedback to address this problem must convey two things: location of the moving object in the image and the amount of movement in the image. Our design coloured the areas of the image where motion occurred. As discussed, it is important not to obscure too much of the display so we presented coloured bars every 10 lines with the length of the bars a rough indicator of the amount of movement, with longer bars showing greater movement (see Figure 3).

Now the user can see any movement overlaid on the LCD image. The photographer can decide if the areas that will be blurred are acceptable or not, and if not, try and do something about it. For example, when taking a picture of a racing car, the photographer could pan the camera with the car. The car would have no feedback on it if the user was panning correctly (the car would be still relative to the camera) but the background would be moving relative to the camera and so would be overlaid with the colour feedback. Panning is a difficult skill to master and our display gives feedback on it in real time to aid the photographer in learning the technique. As with the other methods described above, the result will be that a better picture is taken with less work needed in editing afterwards and fewer badly blurred shots.



Figure 3: Green bars indicate the movement of the car in an otherwise static scene.

7. SUPPORTING GROUP SHOTS

Taking group shots with the camera held out in front (see Figure 4) is a common activity. However, framing correctly is difficult as the screen of the camera is pointing away from the photographer. This can lead to shots where people are cut off at the edges. One solution is to add another LCD screen on the front of the camera (e.g. the Samsung DualView cameras, www.samsung.com/us/photography/digital-cameras/EC-TL220ZBPRUS) or a flip-up LCD (e.g. the Canon 60D, www.canon.co.uk/For_Home/Product_Finder/Cameras/Digital_SLR/EOS_60D). These solutions are expensive so we designed an alternative that does not require any additional hardware to be added to the camera or camera phone.



Figure 4: A common group shot scenario.

The Android operating system from Google is used on many smartphones (www.android.com). It includes a face detector class as a standard component. We used this to recognise the number of faces in the viewfinder when the shutter button was half pressed. The number of face detections in view is represented as a number of short pulses of audio or tactile feedback (visual feedback would be no use as the screen is pointing away from the user). This simple encoding allows the user to determine whether the shot is adequately framed by comparing the number of pulses to the number of faces that *should* be in the shot. So, in Figure 4 there would be three pulses if framed correctly.

Using sound feedback has the benefit that other people in the shot can hear and react to the feedback. For example, if a group of three people were attempting to take a photo together (Figure 4) and only two beeps were heard, the people not directly interacting with the camera could move in closer. Tactile feedback is private, so only the user holding the device benefits from the information. We limited the audio or haptic cues to a maximum of four faces so that we could update the feedback once per second.

We evaluated different audio and tactile cues to present this information back to the photographer. Results from usability tests where we gave users camera phones for a week to take away with them and take pictures as they would in their normal lives showed that the tactile feedback was the most effective and most preferred as users often took pictures in noisy places like bars or clubs where the audio feedback could not be heard.

8. IMAGE STORAGE SPACE REMAINING

Most mobile phones (and standard cameras) show the remaining number of photos that can be stored on the memory card. This is sometimes shown as a number or an icon (Figure 1 shows both a number and an icon top right). Some cameras already give an alert beep when the card is full; others just present a message on screen. We investigated the presentation of this by altering the shutter snap sound. By changing the snap sound to indicate the decreasing space available, the user would get a clearer view of when the card would be full and no more pictures could be taken.

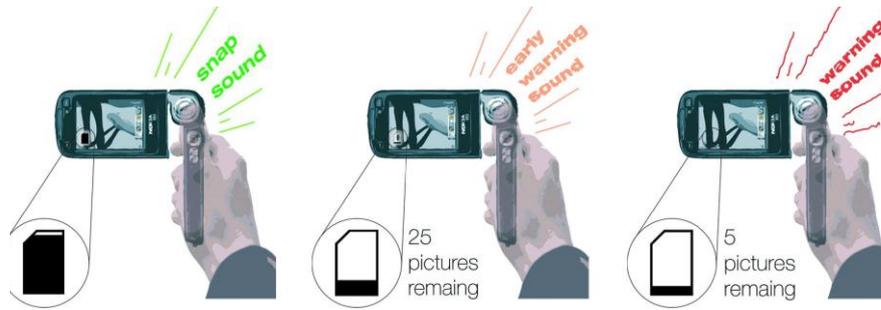


Figure 5: Audio feedback to indicate how much memory space remains.

Our design used the standard shutter snap sound (often just a sound sample played back when the shutter is pressed) until there was space remaining for less than 25-30 more photos. The exact number varies due to the amount of compression of each image and the resolution used. At this point, the standard shutter snap sound was low-pass filtered to create a more ‘muffled’ sound. This early warning sound was played whenever a photo was taken until there was space for less than 5-10 more photos. At that point the sound was changed again. This time the low-pass filtered sound was mixed with the sound of a slide whistle going down. This created the clear impression of something decreasing rapidly, alerting the photographer that something would have to be done or no more pictures could be taken (see Figure 5).

By changing the shutter snap sound we could give a simple warning of the state of the memory card in a way that did not change the behaviour of the camera at all. A finer level of detail could be presented to give a more precise measurement of remaining space but it would be harder to make this perceptible and also it may be unnecessary; the amount of space remaining only becomes really important when the card is nearly full.

9. COMPOSITION AND THE RULE OF THIRDS

The Rule of Thirds is a standard technique for composing visual images and is widely used in photography. Many cameras already provide a 3x3 grid which can be overlaid on the viewfinder (see the yellow grid in Figure 6) to help users compose better images by putting key features like faces at intersection points on the grid. In our survey and discussions with smartphone users, many did not use the grid or know what it should be used for. We developed a technique to help novices compose better images before they were shot, rather than having to crop them afterwards.



Figure 6: Ghost image placed at an intersection point on the Rule of Thirds grid. A red square (left) indicates that the face is not ideally placed. When the image is repositioned and the face located at an intersection point, the square is green (right).

We used the real-time face detection feature of Android to identify the largest face in the viewfinder. We then presented a ‘ghost’ version of this face at the closest intersection point of the Rule of Thirds grid, suggesting a better composition (see Figure 6). The photographer can, if he/she chooses, then reposition the face by moving the camera and then take a photo which is better composed.

We developed a second version of this technique to deal with landscape images. A common problem is for the horizon to be positioned so that it runs across the centre of the image, where the Rule of Thirds would suggest that it should be at the top or bottom third of the screen. Using simple image processing, we identify the major horizontal edge in the viewfinder image, we mark this with a red line and arrows indicating the direction of the closest of the upper or lower third of the screen (see Figure 7). When positioned at one of these points the horizontal line turned green to indicate a better composition. No post-processing was then needed to create a more balanced Rule of Thirds composition. Of course, the photographer was free to ignore the feedback if he/she wanted to make an alternative alignment.

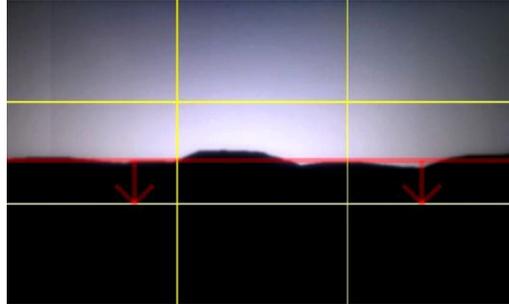


Figure 7: Rules of Thirds feedback for landscape images. The line and arrows suggest where the horizon should be placed for a better composition.

10. PICTURE QUALITY INDEX

In a further attempt to give users more feedback about the image before it was taken, we developed the Picture Quality Index (PQI). This is a simple measure of some aspects of the quality of the image about to be captured. We combined the factors measured above to create this measure. We normalised the values of how level the image was (as measured via the accelerometer), how stable the image was, the current exposure from the luminance histogram, the distance from the closest ideal location in the Rule of Thirds grid and took a simple average of these giving a score in the range of 0 - 1. A score of <0.9 was classed as a poor image, <0.95 an average image and >0.95 was a good image.

A traffic light system was used to display the PQI in the current viewfinder image. Red meant the picture was poor, amber meant it was average and green meant it was good. This was updated once per second when the shutter was half-pressed. By looking at the colour of the traffic light photographers would get some idea of the quality of the shot they were about to take and could adjust accordingly to create a better image. Of course, the quality of an image is far more than the simple combination of these factors. An experienced photographer could ignore the PQI but it provides value to novice users to help them create better images.

11. OFF CAMERA INTERACTION

In our survey a large number of respondents used the self-timer feature of their cameras and camera phones. This is an interaction where the photographer is away from the camera and cannot see the screen, which can lead to framing problems. For example, it can be hard to keep the group of people being photographed in shot when the photographer cannot see what is in frame on the camera. This can result in people being cut off at the sides of the image. It can also be difficult for the photographer to get back in front of the camera when the self-timer is started; he/she may have to run to make it before the shutter fires, which can be stressful.

To solve the problem we gave the photographer control over when the image was taken. This can already be done by using a wireless remote but these have limited functionality. One solution would be to use a second device to allow the photographer to see the camera viewfinder view. We did this using a separate Bluetooth device which can be worn as a watch (Sony Ericsson LiveView micro-display, www.sonyericsson.com/cws/products/accessories/overview/liveview-microdisplay). This connects to the phone (or camera) via Bluetooth. We streamed the camera viewfinder display to this along with a luminance histogram (see Figure 8). Buttons on the watch can then be used to trigger the photo on the camera. This means the photographer can marshal the people he/she is taking a photo of so that they are all in view and then trigger the shutter at the right time. This could also be done with a second phone connected via Bluetooth (see Figure 8).



Figure 8: Displaying the viewfinder image of the camera on a watch display (left), Sony Ericsson LiveView (centre) or another phone (right).

If a secondary device is not available then hand gestures could be used to trigger camera functions. These can then be recognised by the camera and be used to stop, start or change some basic camera settings remotely (Figure 9). This proved to be difficult to implement as the hands are small within the whole image being captured and there may be many hands other than those of the photographer in the viewfinder image, which can lead to false recognitions.



Figure 9: Hand gestures that could be used to control camera features.

The combination of multiple devices could be used to aid in other difficult photography situations. For example, when taking an overhead shot at a concert or club it can be difficult to see the screen of the camera. Streaming the viewfinder view to a watch would be a way to solve this problem (see Figure 10); the camera could be held above the head and the watch would show the viewfinder image. This could then include the other feedback discussed above, for example giving motion detection or stability feedback. The combination of multiple devices for photography has great potential and is an area we are investigating in more detail.

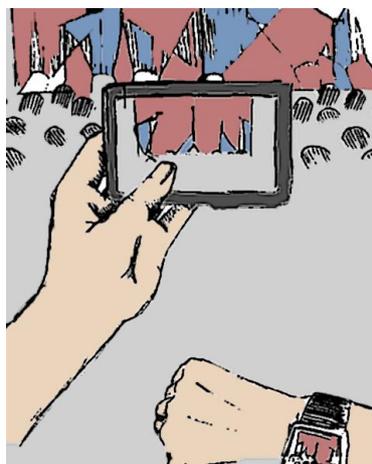


Figure 10: Streaming the viewfinder image from the phone to a watch for overhead shooting.

12. CONCLUSIONS

In this paper we have presented a range of new interaction techniques for cameras and camera phones that can improve the picture taking process. We have evaluated these in lab and real-world settings to ensure they are effective. We identified some of the key issues faced by photographers from a survey and used these to drive new solutions. The advanced sensing, display and processing capabilities available in smartphones allow for new solutions to be applied to some photography problems. Many of these features will migrate to standard digital cameras as the two products converge. For example, some digital cameras now include accelerometers (a common feature in smartphones) which are used to display how level the image is but could also be used for the stability feedback we suggest above.

Some of these features are also useful for videographers. They face many similar issues to still photographers. Feedback is more difficult, however, as audio or tactile feedback may be picked up as noise by the camera. In this case, visual solutions would be the most appropriate (unless the user is wearing headphones). However, that does mean the problem of overloading the LCD screen could cause issues with usability. A further research project is to look into how best to support the needs of those taking videos.

In conclusion, we have developed new interaction techniques for cameras that allow photographers to understand the exposure of a shot before it was taken, know that there is enough space to store image on the camera's memory card and to know when their camera or camera phone was stable enough to avoid a picture being blurred. We have created a tool to help panning and the framing for group shots, along with improving the composition of a shot based around the Rule of Thirds. We have combined these to create an image quality metric, the PQI, which can be used to know if a shot would be acceptable before the shutter is pressed. Finally, we developed a suite of solutions to help with self-timer and 'off camera' interactions so that the photographer can gain control of what the viewfinder sees and when is the best time to capture an image. By rethinking the user interface of cameras we have improved the picture taking process and enabled photographers to get more images right first time.

REFERENCES

- [1] Ames, M., Eckles, D., Naaman, M., Spasojevic, M. and Van House, N. "Requirements for mobile photoware", *Personal Ubiquitous Computing* 14 95-109 (2010).
- [2] Rodden, K. and Wood, K. R. "How Do People Manage Their Digital Photographs?", In *Proceedings of ACM CHI 2003*. ACM Press Addison Wesley, 409-416 (2003).
- [3] Kindberg, T., Spasojevic, M., Fleck, R. and Sellen, A. "The ubiquitous camera: An in-depth study of camera phone use", *IEEE Pervasive Computing*, 4, 42-50 (2005).
- [4] Bourke, S., McCarthy, K. and Smyth, B. "The social camera: a case study in contextual image recommendation", In *Proceedings of ACM IUI 2011*. ACM Press, 13-22 (2011).
- [5] Brewster, S. A. and Johnston, J. "Multimodal Interfaces to Camera Phones", In *Proceedings of MobileHCI 2008*. ACM Press, 387-390 (2008).