

Novel Interfaces for Digital Cameras and Camera Phones

Christopher McAdam, Craig Pinkerton and Stephen A. Brewster
Glasgow Interactive Systems Group, Department of Computing Science
University of Glasgow, Glasgow, G12 8QQ, UK
+44 141 330 4966

{chrism, stephen}@dcs.gla.ac.uk www.dcs.gla.ac.uk/~stephen

ABSTRACT

Camera phones are now very common but there are some usability issues that affect their use. These can occur because the users look through the LCD to frame the image and can often miss the icons displayed around the edges of the edges that present important information about the status of the camera. This may lead to shots being missed or poorly exposed. Most camera phones do not take full advantage of the features of the underlying phone platform to enhance their interfaces. We created a camera application for the Nokia N95 that featured novel interface elements and made use of the features of the platform to provide a rich variety of information in more usable forms, such as: sonifications of the luminance histogram to ensure better exposure before a picture is taken; phone orientation to give a level indicator to ensure the camera is straight; measuring phone movement to ensure the phone is being held steady; and the detection of image motion to support panning. We also present a scenario for how these features could be used in conjunction with each other during the photo taking process.

Categories and Subject Descriptors

H5.2 [User Interfaces]: Auditory (non-speech) feedback; Haptic I/O.

General Terms

Design, Human Factors.

Keywords

Camera phone, sonification, tactile, luminance histogram, panning, orientation, motion.

1. INTRODUCTION

Mobile phones are ubiquitous devices in our daily lives, and they typically come with a camera as standard. While originally this feature was not very mature, the quality of the camera on today's phones has begun to rival that of dedicated digital cameras. Since the first camera phone was introduced in Japan in 2000 [1] the growth of camera phones has increased significantly throughout the world, with figures stating that in 2007 84% of mobile phones sold in the UK include a camera [2]. The GfK group [3] states that 78 million camera phones were sold in the first half of 2007 in Asia. In the UK, government studies from 2007 [2] showed that 65% of users were aware of the camera in their phone, and that 41% made use of the feature. As phones are upgraded to newer models, this figure is likely to increase. The increase in quality in camera phones has led to them becoming a popular choice for people as an 'always on you' camera. As the technology advances, features from standard digital cameras are finding their

way to camera phones, such as autofocus, optical zoom, 'smile detection' and the ability to record movies. It is likely in the future that other features will continue this migration, making the camera on the phone indistinguishable from a standard digital camera.

There are some usability problems with camera phones, and standard digital cameras, that can cause problems for the user while taking photos. Without a dedicated viewfinder, the LCD is used to compose the image. This screen is also used to present the user with information about the current settings and status of the camera. Most often this information is in the form of small icons presented around the edges of the display, as shown in Figure 1. This is problematic for two reasons. Firstly, when framing an image your attention is often on the scene you are taking a picture of, making it possible to miss these icons. Secondly, every icon displayed on the screen obscures part of the image, so as the number of icons displayed increases it makes it harder for the user to see the entire frame. A common function allows these icons to be hidden, so that the user can use the entire screen to compose the image, but that denies him/her the information that they provide.

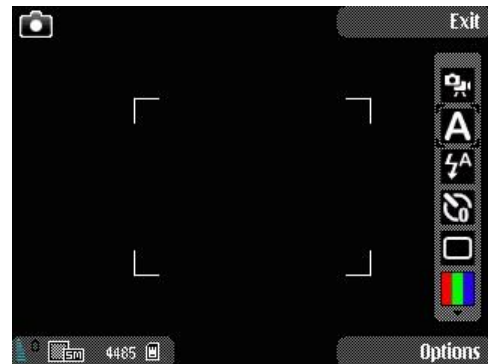


Figure 1 Camera application on the Nokia N95.

The interfaces for camera phones are based upon those of standard digital cameras, an example of which can be seen in Figure 2. These, in turn, are similar to the analogue cameras that preceded them. They do not take full advantage of the capabilities of the phone for providing the user with feedback. For example, many modern phones feature sensors such as accelerometers and magnetometers which could be used to determine a variety of information on the phone's current state, such as the orientation of the device, and could prove to be very useful for improving the picture taking process.

Camera phones are unique amongst the differing camera form factors as they are capable of producing a wide range of audio and

tactile cues. Many phones have high quality audio hardware that is used for playing music and games, which can be used to generate sophisticated audio cues for the user interface. The vibration motor that is used to produce alerts for the phone can be used to generate tactile cues. Combining these outputs with the rich sensors inputs give camera phones a unique potential to determine a range of information that is not typically found in digital cameras and to present that information to users through a variety of different modalities.



Figure 2 Display of a Sony DSC-V3 digital camera while photographing a coloured target, including luminance histogram (white graph shown bottom right) (from www.dpreview.com).

2. BACKGROUND

There is much work into the use of mobile phones as an interaction device, e.g. [4-6], but there is not much into the actual user interfaces to these devices. In their study, Rodden and Wood [7] found that people, when organising their photographs, would separate the good photos from the bad ones, sometimes discarding the bad ones completely. The ‘bad’ photographs are considered such either because they are uninteresting or because they are “technically poor”. This shows that the quality of photos is important, however users generally did little editing after the photo had been taken. It would be better if the camera could provide assistance to users before they have taken the photo to reduce the likelihood of the photo being poor in the first place.

Håkansson *et al.* [8] developed a prototype camera application on mobile phones to support context photography. In context photography, sensors gather information on the context of where the photo is being taken, which is used to visually alter the photograph. For example, “the chill in the air on a mountain could be made visible as a bluish tint evocative of low temperatures”. This is an interesting example of using the data gathered from sensors to affect the look of the photograph, but this is a different from our approach of using sensor data to improve the quality of the image.

In our own previous work, Brewster & Johnston [9] developed a camera phone interface that provided multimodal feedback to the user on a number of different camera features. Tactile cues were used to indicate the remaining battery level on the device, and audio cues were used to indicate the remaining memory card capacity. A sonification of the luminance histogram was used to inform users of the exposure level of the photo on half-press of the shutter button (i.e. before a photo was taken). We found that sonifying all 256 values of the luminance histogram took too long just to convey information on exposure, and restructured the his-

togram so that it grouped luminance information into 5 ‘bins’ instead. This structure placed greater emphasis on the endpoints of the histogram, as that is where the most relevant information for exposure can be found. It was found that users were able to identify audio luminance histograms at a similar level to visual ones, suggesting that the 5-bin histogram used in the study was sufficient to assess image exposure. The cues for remaining memory and battery life were identified correctly over 80% of the time. This suggests that it is possible to replace traditional visual cues with multimodal ones, and that there is scope for doing this more in camera phone interfaces. In this paper we develop these ideas further to create richer interactions using more sensor inputs from the phone.

3. CAMERA APPLICATION

Standard digital cameras cannot run third-party code, so we designed and built our own camera application in python and c++ on the Nokia N95 phone, one of the most capable phones released in 2007. Our application was designed so that the primary use of the screen was for framing the image and there were no icons or other information displayed visually in its default mode.

Our aim was to deliver as much information as possible multimodally, specifically through audio and tactile feedback, both of which the phone is capable of producing. Visual feedback was also used, but this would only be at the users’ request, and would not be permanently shown on the display. When displayed visually, it was designed in such a way that it would not hinder the user in using the screen to frame the scene.

4. NOVEL CAMERA INTERACTIONS

We investigated a range of novel interactions to improve the picture taking process. The aim was to enable photographers to capture better images for their collections and reduce the number that would have to be deleted or post processed (especially as Rodden and Wood suggest many people do not post process). We chose to work, for example, on sonifying the luminance histograms as these are very useful for assessing exposure before a picture is taken but take up a lot of screen space and are quite hard for novices to interpret [9] (see Figure 2 bottom right hand corner). Providing this information non-visually would allow the user to focus on using the screen for composition. We developed the following new feedback cues: two sonifications of a luminance histogram, a level indicator, steadiness indicator, a motion detector and a ‘who is moving’ indicator.

4.1 Sonification of a Luminance Histogram

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 and look different depending on the angle at which they are viewed.

Histograms have to be fairly large to make the information in them clear visually, leading to them obscuring some of the image on the display (see Figure 2). Some cameras present a limited version of this information by flashing the parts of the image that are over/under exposed. This flashing could be distracting to the user, and also conveys less information than a full histogram would.

Sonification means to use non-speech audio to convey information or to represent data [10]. Our previous work [2] showed that

it is possible to condense a luminance histogram down to 5 values and sonify them, and users were still able to extract enough information to correctly identify image exposure. Our current work extended this in two ways: sonifying all 256 values in the histogram (copying that shown visually in Figures Figure 3 and 4), and simplifying the histogram even further, with the aim of presenting the user with a single feedback cue to indicate the exposure. This method of feedback is likely to be best suited to the novice user, who may have difficulty interpreting the full histogram.



Figure 3 A photo and its corresponding luminance histogram.

4.1.1 HDS Histogram

One aim of this work is to provide the user with a sonification of the entire luminance histogram. In order to play all 256 values using a conventional sonification the duration of the feedback would present a problem. If each value was played for 0.25 sec., for example, then the whole sonification would last 64 seconds. It is not feasible for the user to listen to a sonification of that length before taking a picture. The ideal would be of the order of one or two seconds, so that the user is presented with the feedback and can react to it quickly. A technique that would allow such swift feedback is High Density Sonification (HDS)[11]. HDS was developed to aid visually impaired users in exploring numerical tabular data. HDS works by sonifying the values in a row (or column) in rapid succession, from left to right. The values in the cells of the table are mapped to pitch; higher values to higher pitch notes. This technique takes advantage of the high temporal resolution of sound perception, where a difference of 20ms is sufficient for the user to distinguish which of two sounds is heard first.

The duration of the sounds has a significant effect on the amount of information that the user is able to extract. For short sound durations (i.e. faster sonification) users are not able to perceive detailed information but they are able to perform comparisons and spot outlying values. If the duration is increased, then the user is able to extract more detailed information, but the overall sonification will take longer. To extract the necessary information about exposure the user must be able to perceive the rough shape of the histogram, the fine details are less important, so a quick sonification should be sufficient. In practice our HDS histograms lasted only slightly longer than the five note version presented by Brewster and Johnston.

We developed three different formats for our HDS histogram;

- *Full Histogram* - Presents the user with all 256 values of the histogram in just over 1 second.
- *Sampled Histogram* – This reduces the size of the histogram by half. It does this by taking the maximum of each consecutive, non-overlapping pair of values in the histogram (e.g. [n, n+1] [n+2, n+3] etc.). This reduces the overall sonification time and the amount of information presented in half, but does so while maintaining the overall shape of the histogram.

The maximum of each pair is taken, as when considering exposure it is the higher values that are important (to avoid clipping).

- *Distorted Histogram* – The third method alters the structure of the histogram in a more significant way. When judging exposure, it is often the endpoints of the histogram that are the most important. Accordingly, the histogram is split into three sections. The first part covers the shadows part of the histogram, and includes the first 25 values in full. The third part covers the highlights part of the histogram, and plays the last 25 values in full. The second part covers the midtones in the image, but it is not presented in full. Instead, it takes a sampling of that section, creating a bin for every consecutive, non-overlapping set of 8 values, where the maximum value in the bin is taken. This presents a significant compression of this section, while still maintaining the significant features of the histogram, such as high values and overall shape. In total this reduces the histogram to 76 values, with the endpoints of the histogram presenting more detail than the midpoint.

4.1.2 Indicate Exposure

Our second approach to sonifying the histogram simplified it to its most relevant information, in this case: is the image correctly exposed or not? Instead of presenting the user with a sonification of the histogram, we examine the structure of the histogram and make a judgement on the exposure of the image (in a similar way to the autoexposure feature of many cameras), and provide the user with an audio cue indicating if the photo is under, over or normally exposed. Whereas the full histogram may be more of a ‘pro’ feature, this is aimed at novices to give only some general information about the likely exposure of their photograph.

The main problem with any categorisation here is that no rules that explicitly state what images can be considered over or under exposed. Any such categorisation is usually based on subjective opinion. As such opinion is not sufficient for our needs; we will define a categorisation scheme based on the endpoints of the luminance histogram. We chose the endpoints as they are the key areas for determining if the image is under or over exposed. If it is determined from the endpoints that the image is neither over nor under exposed, then it can be inferred that the image is normally exposed. Formally, we define underexposed images as those having n or more pixels of luminance value 0; over exposed images as those having n or more pixels of luminance value 255; normal exposed images are those that do not fall into either of these groups. In the case that an image has more than n pixels of both values 0 and 255, then the dynamic range of the image sensor has been exceeded. In our program n had a value of 30.

The feedback itself is designed to take the form of two notes. The tones indicate the two endpoints of the histogram, as these are the significant regions when determining exposure. Example histograms can be seen in Figure 4. The pitch of the tones will correspond to the number of pixels at those endpoints; A low number of pixels will give a low note and a high number of pixels a high one. For a normal, well exposed image, the feedback should be two low notes, as a normally exposed histogram is mainly centred on the mid-tones, and has low values at either end (Figure 4 middle). The feedback for an underexposed image (Figure 4 top) is a high pitched note followed by a low pitched note. This corresponds to underexposed images being high in shadows (the left hand side of the histogram) and low in highlights (the right hand side of the histogram). Conversely, the overexposed feedback should be a low note followed by a high note (Figure 4 bottom).

Two high notes would indicate that the image has gone beyond the dynamic range of the image sensor.

It is hoped that as users continue to use the system, this correspondence between the structure of the histogram and the feedback can help to advance their understanding of histograms and their distributions. However, it is recognised that this may not occur, and as such the feedback is designed in such a way that the user does not need to understand histograms in order to interpret the exposure feedback.

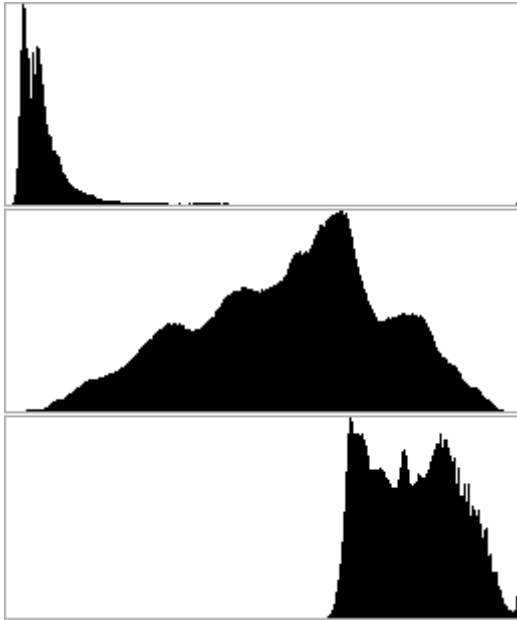


Figure 4 Examples of under, normal and over exposed histograms (from top to bottom).

These two sonification methods provide a means of extracting a wide range of detail from the luminance histogram depending on user knowledge or preferences. The user can opt to receive information on the entire histogram, or just on the probable exposure. As such the system is adaptable to the user's experience, or simply to how much information the user wants at that point in time, making the sonifications useful for both amateur and more knowledgeable photographers alike.

The sonified histograms were presented when the user half-pressed the shutter button, augmenting the standard focus lock sound to provide more information. He/she would then know if the picture would be exposed correctly before the picture was taken and could adjust exposure accordingly, perhaps by tilting the camera to get the sun out of the scene, for example. Thus the user could 'probe' the scene by tapping the shutter button until a good exposure was heard and the shutter button could be fully pressed and a picture taken. This could all take place without getting in the way of the framing of the image in the LCD.

4.1.3 Evaluation

We conducted an evaluation of the sonified histograms to test their effectiveness. We wanted to determine how well users could extract information from them compared to standard visual histograms. Due to the difficulty of ensuring exact consistent images and histograms between the participants when performing the evaluation using the live camera feed on the phone, it was decided

to perform the evaluation on a computer instead. This allowed us to use the exact same images for each participant and greatly improves the consistency. As we are interested in the users' comprehension of the histograms, using a computer instead of a phone does not adversely affect the experiment.

For the Indicate Exposure technique, the experiment took the form of a multiple choice, within-subjects design, presenting one image and three histograms, each indicating a different exposure type, with one being the correct histogram for the image. The user had to identify the histogram that belonged to the image. The user was presented with 30 images (10 underexposed, 10 overexposed and 10 normally exposed) with each person performing all conditions in a counterbalanced order, with the histograms being presented in a standard visual way or with the sonified version. Users received training at the start of the experiment, with examples of the different exposures and histograms presented and explained to them.

For the HDS Histogram feature, the experiment took the form of a three condition (conditions being the three different type of HDS histogram), within-subjects design, presenting one image and three histograms, each indicating a different exposure type, with one being the correct histogram for the image. The user had to identify the histogram that belonged to the image. The user was presented with 30 images (10 under exposed, 10 over exposed and 10 normally exposed) with each person performing all conditions in a counterbalanced order, with the histograms being presented using each of the three HDS histograms. Users received training at the start of the experiment, with examples of the different exposures and histograms presented and explained to them. The training involved 6 images, two of each exposure type, and these images were not present in the main experiment. The participants were allowed as much time as they wanted to work through the training until they were satisfied that they understood the feedback.

4.1.4 Results

Nine participants took part in the experiment for the 'Indicate Exposure' feature. They were 17-24 years old and all were students at the university. The mean score for correct identification of the visual histograms was 25.5 out of 30 (SD 4.4) with 25.25 (SD 3.54) for the audio feedback. This suggests that the users could match the histograms to the images well in both conditions, and that there was little difference between the two methods. There were no significant differences in the recognition of under, over or normally exposed images.

Eight participants took part in the HDS histogram experiment. The participants were aged 18-24 years old and all were students at the university. No participants took part in both experiments. The mean score for correct identification of the full histogram was 21.9 (SD 4.16), with 21.25 (SD 4.3) for the sampled histogram and 22.13 (SD 4.09) for the distorted histogram. The recognition rates for the different exposure types were similar across conditions, and in all three the underexposed images were the hardest to recognise. This was partly expected, as for underexposed images the relevant parts of the histogram are right at the start, so the linear presentation of the sounds means that they are 'overwritten' by the rest of the histogram that follows. Conversely, the overexposed images were usually the best recognised, likely because the relevant details for these images are at the end of the sonification and heard most recently. This technique presents a lot of information so it would not really be for novices (even with training our

participants were still new to the sonifications). We hope performance would increase with more use of the technique.

Comparing back to our previous work, users were able to recognise the 'Indicate Exposure' feature at around the same level as we found for our 5 bin histogram, while the recognition rates for the HDS histograms were only slightly lower. This suggests that these three techniques could form a suite of sonifications for luminance histograms, with users able to choose the one they prefer to use. It also shows that further alteration of the presentation of the HDS histogram, either in presentation or structure, could improve the recognition rates.

4.2 Level Indicator

A common problem that can occur when taking photographs is that it is difficult to align the camera perfectly with the horizon (this problem also occurs when trying to align with the vertical). As such, the resultant photos are often at a slight angle. If, as Rodden and Wood suggest, people do not do much editing of their photos once taken, these might be discarded. To combat this we developed the Levelometer to present the user with feedback on the phone's orientation, using its built-in 3-axis accelerometer, prior to taking a photo.

Orientation was gathered from the accelerometer. Position was determined using simple trigonometry. Once this information was acquired it was necessary to present it to the user. We propose a simple visual feedback cue that acts similarly to those in aircraft artificial horizon displays. A horizontal red line is drawn in the centre of the screen with white cross-hairs. This line indicates the position of the phone. The pitch and roll determine the y-coordinate and the angle that the line is drawn at respectively. An example of this can be seen in Figure 5. While this is a visual display, we designed it to be as non-intrusive as possible. It also actively helps the user in the framing of the photo, instead of distracting the user from performing that task.



Figure 5 Levelometer feature. The red line indicates the current orientation of the phone. The white crosshairs is used as a guide to aid alignment.

When the phone is correctly aligned with the horizon, the red line extends to fit the width of the display (Figure 6). When the phone is correctly aligned with the vertical axis (i.e. is pointing straight ahead, not pitched up or down) then the line will be centred on the crosshairs, and a second red line will be shown that extends from the top to the bottom of the screen.

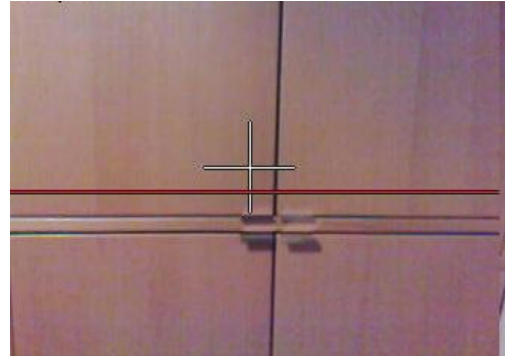


Figure 6 Levelometer with the red line extended across the display, indicating that the phone is correctly aligned with the horizon.

We also developed a second version of this feedback. Instead of the line indicating the orientation of the phone, it showed the orientation of the horizon. The red line was used to indicate where the horizon was. This version behaved in a similar way when the phone was correctly aligned with the horizon, with the red line extending to fill the width of the display.

Informal testing of the two displays found that users preferred to use the version that displayed the orientation of the phone, as they found it easier to make corrections to get the phone aligned.

This type of display is useful when the user is taking a picture with the phone held above head height, for example at a concert. The user is still able to see the display, as the phone would be held slightly forward, and not directly above the head. However, the small screen is harder to see clearly at arms length, so having a clear visual indicator appear when the phone is level would be advantageous in this situation. It is very difficult to tell if the phone is being held level before a photo is taken and it can be hard to see the screen. With our interface the user can look to see the red line extend across the display and then take the shot, knowing it will be straight.¹

4.3 Stability

During the process of taking a photograph the camera is subject to small movements from the users' hands. In many cameras the icon of a hand shaking is used to indicate that the image may be blurred (see the centre of Figure 2 for example). This is only based on the shutter speed of the camera, not the amount that the camera is actually moving. If the user has steady hands they 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.

Currently, there is no feedback presented to the user that indicates whether the camera is really steady or not. Some digital cameras offer an anti-shake feature, but this attempts to compensate for any shake without informing the user if the camera is steady enough to take a sharp photo. Our approach presents users with the appropriate feedback to indicate how steady they are holding the camera (and could be combined with the anti-shake features). In order for them to have confidence that they are holding the camera steady the feedback must be presented in a continuous

¹ Since we completed this work two high-end digital SLR cameras have been released that include a similar feature (Nikon D3 and Canon 7D).

manner, starting from the users request for feedback and finishing with the photograph being taken or the user dismissing the feedback. Again we use the accelerometer in the phone, this time to measure the movements it is being subjected to so that we can then present this to the user. Three different feedback techniques, audio, tactile and visual, were developed for this prototype to see which was the most effective.

The audio feedback consists of a short sound cue. When the phone is being 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 the cues making it sound more arrhythmic, indicating to the user that he/she is shaking the camera too much.

The tactile feedback is similar in nature to the audio. A quick, gentle tactile cue is played in a steady rhythm when the user is holding the phone still. This is done using the internal vibration motor on the phone. If the phone becomes unsteady, then a stronger cue is played with a longer delay to break the rhythm. The tactile cues themselves are not strong enough to affect the steadiness of the camera phone and so cause the image to be blurred.

The visual feedback takes the form of a small line graph in the bottom right corner of the display (small enough to avoid obscuring the image in the LCD). When the phone is being held steady, the line is completely flat (Figure 7). 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 8). 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.



Figure 7 Visual feedback for the steadiness indicator. The near flat line here shows that the phone is being held very steadily.

For all cues, if the phone becomes too unsteady then a sharp audio cue is played to inform the user that they are moving the phone too much and steps need to be taken to correct this. The feedback also stops at this point, to allow the user to re-steady the phone before trying to take the photo again.

Issues arose in the testing of these different feedback cues. For stability control feedback to be effective it needs to be delivered quickly to users so that they can respond to it, if there is too much lag then the movement of the phone will have changed and their

response may be wrong (they try to steady the phone in one direction but it has started moving in another, for example). The phone and its operating system were not able to play the audio cues rapidly enough in response to the accelerometer data. This lag meant that the cues did not help users steady the phone in the most effective way. This lag does not affect the sonification of the histogram, as in that case the processing of the values was done first and then the sonification could be generated, whereas in this case new cues must be generated in real-time as the values update. Similarly, for the tactile feedback the simple rotational motor in the phone turned out to be a problem. Such motors take some time to spin up to speed and this time again caused a lag in the feedback. In the end, the visual feedback was able to be presented the fastest so provided the most effective cue for keeping the phone steady. In informal evaluations with users they all preferred the visual feedback, even though it could obscure parts of the image they were trying to photograph. However, it is only displayed on screen at the user's request, and while it is displayed the user is responding to the information that it presents. As such, it is different from the icons mentioned previously, which obscure the display and are generally ignored by users. The audio and tactile feedback were limited by the capabilities of the phone we used in this case. Phones with faster processors and different tactile actuators (for example, linear ones with faster response times) could solve some of the issues and further testing would allow them to be more formally compared to the visual display.

4.4 Motion Detection

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. When taking photographs, movement on the part of the subject can lead to the resulting photograph being blurred. The amount of movement will relate to how much blurring occurs. Of course, the photographer might use the blurring for artistic effect, but many times it just leads to a poor image. The small size of the camera display and the fact that the user will be focused on other things, such as alignment and exposure, can lead to the user not noticing subject movement. We developed a method to indicate the movement to the user.

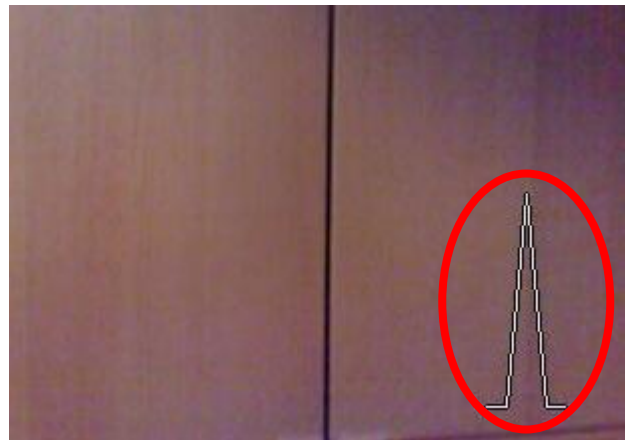


Figure 8 The peak in the line corresponds to the amount of movement.

The feedback that is presented should convey two pieces of information – the location in the image of the movement and how much movement is taking place. Our primary design was to colour the areas of the image where motion was occurring (for ex-

ample, pedestrians walking across the scene), with the colour used to correspond to the degree of movement that is occurring. As discussed above, it is also important that this feedback not obscure too much of the display. Our original version of the feedback worked over the entire display. This introduced a problem where it was possible for too much of the image to become covered. This also led to a slowdown in performance in the application (due to the image processing required, even when the code was written in c++). When performing the motion detection over the entire image, it would take a few seconds to process each new image from the viewfinder. This illustrates the restrictions in processing power that mobile devices have, and how trying to perform computationally intensive tasks can place heavy demands on the available resources. As such, we changed our algorithm from one that worked over every pixel in the viewfinder image to one that worked on a scan line principle (see Figure 9).

For the scan line version of the algorithm, we are now concerned with the pixel rows of the images we use for comparisons. Assume that motion has been detected at pixel ij in row n of the image. It is then likely that corresponding motion has also been detected at pixel ij in either row $n-1$ or $n+1$ (or both rows). As such, we can choose not to examine rows $n-1$ and $n+1$, as row n is likely to provide enough information for our purpose. As such we introduce a step factor s , which says to examine every s th row of cv (current viewfinder image) and ccv (previous viewfinder image). This reduces the number of pixels that need to be compared from $cv.height*cv.width$ to $\text{floor}(cv.height/s)*cv.width$. In practice, our algorithm uses a step of 10, leading to the feedback being generated in real time. The display can be seen in Figure 9. As processing power in phones increases the step size can be decreased.

In this version we no longer use different colours to show motion and all of the feedback is coloured green, with the length of the lines presented act as a rough indicator of the amount of movement that has occurred (longer lines show greater movement). Techniques exist that would allow us to examine every pixel in the image rapidly [12] which would allow us to colour all moving sections of the image. However, we have found that our scanline method is able to convey enough information without obscuring large areas of the image.

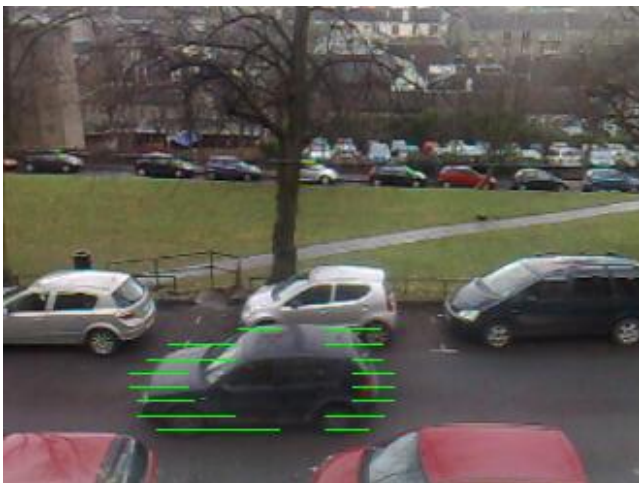


Figure 9 Motion detection. The green lines indicate the movement of the car in an otherwise static scene with the phone being held still.

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 get right and our feedback would help the photographer do it; if he/she was panning correctly the racing car would have no feedback on it, if the car had colour overlaid then he/she is not panning correctly. As with the other methods described above, the result will be that a better picture is taken with less work needed in editing afterwards.

4.5 Who is Moving?

As shown in the previous sections, we now have the means to detect both any instability in the camera and any motion in the scene. These are separate, but are both factors in determining if a photograph is likely to be blurred. A prototype was developed that combined the stability and motion detection features into a single tool. For simplicity, at this stage this tool is only concerned with informing the user as to what is causing the greatest amount of blur in the image – the scene changing or the camera moving. It also indicates to the user if both of these risks are minimised enough that it is okay to take the photograph.

Accordingly, this feature will present the user with a singular point of feedback, which can take one of three values: Scene movement, Camera movement or All OK. The purpose of this initial prototype is geared towards the integration of the two features. Once this integration is established successfully then the feature can be expanded upon to provide the user with more information, but at this point a simple identification of the most likely cause of blurring is sufficient.

There are many ways to present this information to the user. It could be done visually, with a small notification appearing on screen. As this notification would only appear if there was a source of movement, it would not be a permanent distraction to the user and would not interfere with the framing of the scene. An audio cue could be presented to the user to indicate who was moving. A tactile cue could be used in the same way.

Our initial prototype of the feature used a simple visual cue. We used this because it is the quickest to update, as there is no need to wait for sounds to start playing or for vibration motors to spin up or down. We found that the system was quick to identify the current source of movement and to update as that movement was corrected, however it would only run on a very low resolution video stream. The processor load of this application was high as there is a significant computational cost to running the image processing and using the accelerometer data at the same time. Phones have developed a lot since the release of the N95 in 2007 with 2010 devices having 1GHz CPUs and much more memory. These will allow us to develop these ideas further and run them in real time.

Such a feature could be useful if the photographer was trying to arrange a group of people in a photograph. He/she would be primarily focused on arranging the subjects so that they all fit in the image, and are shown relative to an interesting object in the background, they would not be able to focus on movements that are likely to occur in groups of people. As such, they can be focused on the composition, while the system can inform them of any

sources of movement, allowing the user to either steady the camera or tell the group to be still without interrupting their framing of the picture.

5. CONCLUSIONS

We have presented a range of novel interactions to improve the picture taking process on digital cameras and camera phones. By using the sensors built in to modern mobile phones, along with simple image processing techniques, we have demonstrated tools that can help photographers overcome some of the problems that can cause poor quality images to be captured.

The indicate exposure feature was successful in allowing users to identify the likely exposure of the image before a picture was taken. The HDS histograms were successful in providing more details on the structure of the histogram to the user. It was found that manipulating the structure of the histogram did not hinder the user's ability to extract information from it, and that further manipulation of the structure may improve the recognition rate. These sonifications allow the user to test the exposure of a scene before they take the picture, but without obscuring the screen with a large visual histogram.

The levelometer, stability indicator and motion detector showed that it is possible to provide the user with more information than is normally provided in camera phones, and that the underlying capabilities of the phone can be used to provide an enhanced range of useful information that is often not even present in standard digital cameras. Sensors like accelerometers are becoming more common in standard digital cameras, for example as part of image stabilisation systems. So, the techniques we have developed are also possible on these devices to increase their usability.

We also showed that it was possible to combine some of these features to complement each other. While we have only examined each feature in isolation so far, looking at how they would be used together is interesting, and where we see the benefits of using three different modalities (visual, audio and tactile). Johnston and Brewster's previous work presented two different cues to the user at the same time in the audio and tactile channels, which the users were still able to understand. The following possible scenario of camera usage shows how this would work, including the features described here and those from Brewster & Johnston's paper:

Graham is on holiday with friends in London. He wants to take a photo of his friends in front of the London Eye using his camera phone. He gets them to line up in front of the landmark and uses the LCD on the camera to frame the image. As he frames the image to include both his friends and the landmark, an audio cue plays indicating that there is a lot of movement happening in the scene. Graham tells them to stay still while he continues to frame the image. As he is framing the image, using the levelometer to ensure that the photo is not at an angle, the green lines on the display tell him that there is some movement happening in the background (the London Eye itself going around). As none of this is covering the intended focal points of the image, Graham decides to continue with the photo. As he half depresses the shutter button to set the focus, a sonified histogram is played, informing Graham of the likely exposure of the photo. Following this, an audio rhythm plays to indicate the steadiness. Happy that the camera is steady, that the exposure is correct and that his friends are posed correctly graham takes the photo. After the photo is taken, an audio cue plays to indicate the remaining capacity on the memory

card, and a tactile cue plays to indicate the remaining battery life. Graham, while reviewing the photo he has just taken, knows that the battery is nearly full but he only has space remaining for a few more photos. Satisfied with the photo, Graham puts his phone away and he and his friends continue with their holiday.

In conclusion, we have presented a series of prototypes of novel interaction techniques to improve the picture taking process for camera phones and digital cameras. These allow users to receive feedback in a way that does not distract them from the process of framing the image and capturing the picture. We presented a scenario of how these feedback cues could work together in a real-world setting, and have shown that multimodal feedback can add to the photo taking process and that camera phones are a good platform for investigating new interfaces and interactions for cameras.

6. ACKNOWLEDGMENTS

This work was funded by the Nokia university donation programme.

7. REFERENCES

- [1] Okabe, D. Emergent social practices, situations and relations through everyday camera phone use. In Proceedings of the Conference on Mobile Communication in Seoul (Seoul, 2004).
- [2] OfCom The Communications Market 2007: 1 Converging communications markets. 2007.
- [3] Asian Mobile Sales Top \$22B as Color Camera Phones Continue to Dominate, <http://www.wirelessdesignasia.com/article-6628-asianmobilesales-top22bascolorcameraphonescontinuetodominat-Asia.html>, Last Accessed on 29/01/2010
- [4] Hardy, R. and Rukzio, E. Touch & Interact: Touch-based Interaction with a Tourist Application. In Proceedings of the MobileHCI '08: Proceedings of the 10th international conference on Human computer interaction with mobile devices and services (Amsterdam, The Netherlands, 2008). ACM.
- [5] Rohs, M. and Oulasvirta, A. Target acquisition with camera phones when used as magic lenses. In Proceedings of the Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems (Florence, Italy, 2008). ACM Press.
- [6] Toye, E., Sharp, R., Madhavapeddy, A., Scott, D., Upton, E. and Blackwell, A. Interacting with mobile services: an evaluation of camera-phones and visual tags. *Personal and Ubiquitous Computing*, 11(2006), 97-106.
- [7] Rodden, K. and Wood, K. R. How do people manage their digital photographs? In Proceedings of the Proceedings of the conference on Human factors in computing systems - CHI '03 (New York, New York, USA, 2003). ACM Press.
- [8] Håkansson, M., Gaye, L., Ljungblad, S. and Holmquist, L. E. More than meets the eye: an exploratory study of context photography. In Proceedings of the NordiCHI '06: Proceedings of the 4th Nordic conference on Human-computer interaction (Oslo, Norway, 2006). ACM Press.
- [9] Brewster, S. A. and Johnston, J. Multimodal interfaces for camera phones. In Proceedings of the Proceedings of the

10th international conference on Human computer interaction with mobile devices and services - MobileHCI '08 (New York, New York, USA, 2008). ACM Press.

- [10] Kramer, G., Walker, B., Bonebright, T., Cook, P., Flowers, J., Miner, N., Neuhoff, J., Bargar, R., Barrass, S., Berger, J. and Others Sonification report: Status of the field and research agenda. Prepared for the NSF by members of the ICAD.(Available on the World Wide Web at <http://www.icad.org/websiteV2.0/References/nsf.html>)1997).
- [11] Kildal, J. and Brewster, S. A. Explore the Matrix: Browsing numerical data tables using sound. In Proceedings of the Proceedings of the 11th International Conference on Auditory

Display (ICAD2005) (Limerick, Ireland, 2005). Department of Computer Science and Information Systems, University of Limerick.

- [12] Wang, J., Zhai, S. and Canny, J. Camera phone based motion sensing: interaction techniques, applications and performance study. In Proceedings of the Proceedings of the 19th annual ACM symposium on User interface software and technology (Montreux, Switzerland, 2006). ACM.