Body-Mounted Cameras

Claudio Föllmi
Student BSc computer science
Distributed Systems Seminar 2013 report
ETH Zurich
foellmic@student.ethz.ch

ABSTRACT
Digital cameras have become small, light and cheap, which allows them to be worn for an extended amount of time. This makes new use cases feasible and practical. In this report, we take a closer look at three very different approaches to body-mounted camera use.

Keywords: camera, wearable computing, motion capture, eyetap, lifelogging, sensecam, mediated reality, sousveillance

INTRODUCTION
People have been wearing cameras for a long time – the astronauts on the moon had cameras mounted on their torso – but traditionally only to take pictures under difficult circumstances (such as in a spacesuit with restricted movement).
With advances in camera technology, most importantly the switch to digital photography, wearable cameras have become less burdensome. In recent years, putting cameras on helmets has become a staple of reality television and extreme sports. But today's cameras are so small and light that we can even get rid of the helmet, and truly experiment with the placement and use of the camera.
In this seminar report, we will take a look at three systems that attach a camera to the body, and reflect on the implications if these systems actually become adopted.

EYETAP
In the 1980's, wearable computing pioneer Steve Mann developed an electronic eyeglass, inspired by the idea that welding masks could be replaced with a system that does not darken the whole field of vision.
His "EyeTap digital eyeglass"[2] can be worn like a normal pair of glasses and completely replaces the light entering the eye with a projected image. Ever since, Mann has been wearing EyeTap devices of various designs on a regular basis, and in doing so has experienced a wide range of reactions[1, 3].

Approach
The key component of EyeTap is a double sided mirror, which is placed in front of the wearer's eye at an angle of 45 degrees. Incoming light is diverted to one side into a digital camera, which streams the captured images to a wearable computer. This computer can process the images (e.g. applying filters) and forwards them to a projector (called "aremac", which is "camera" backwards). The images are then projected onto the other side of the mirror and into the eye of the user.

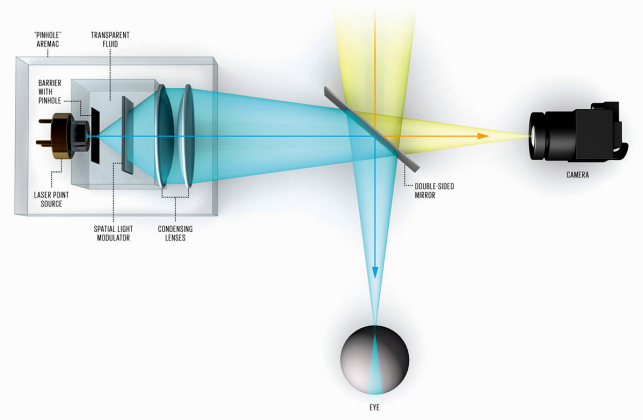

Figure 1: All incoming light goes into the camera, while the aremac generates the light that enters the eye. For each ray of incoming light, there is a collinear ray of projected light, creating the illusion of complete transparency.[3]

If the distances from the mirror to the eye and from the mirror to the camera are equal, then optically the camera has the exact perspective that the eye would have. This is necessary for the projected image to look natural, and it distinguishes EyeTap from other approaches to augmented reality. As a side effect, from the outside it looks as if the user had their eye replaced with a camera, as seen in Figure 2.

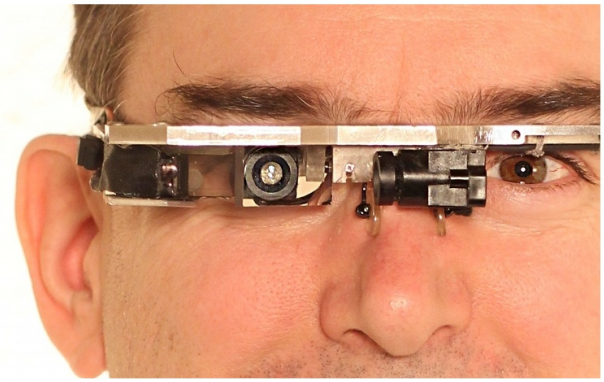

Figure 2: Steve Mann wearing a 1999 EyeTap design.[2]

Use Cases

Since the camera is always recording anyway, the EyeTap can be conveniently used for lifelogging, which is the capturing of our experiences in a visual diary that can be shared with others. Capturing events from the exact perspective of the wearer is especially interesting from an artistic point of view, as the images will be a more authentic reflection of the actual experience than images from any other perspective. But what really sets EyeTap devices apart from other wearable cameras is the ability to completely replace what the user sees.

Applying a digital filter on the images allows for a wide variety of applications, from simulating the effect of lenses (replacing prescription glasses), over adjusting contrast, to a complete recoloring.

Steve Mann himself usually uses a high dynamic range filter[4], which combines frames of different exposure levels to make bright objects less glaring while lightening up faint details at the same time. This is a prime example of an enhancement of vision that can not be achieved using traditional lenses and filters, but only through computation.

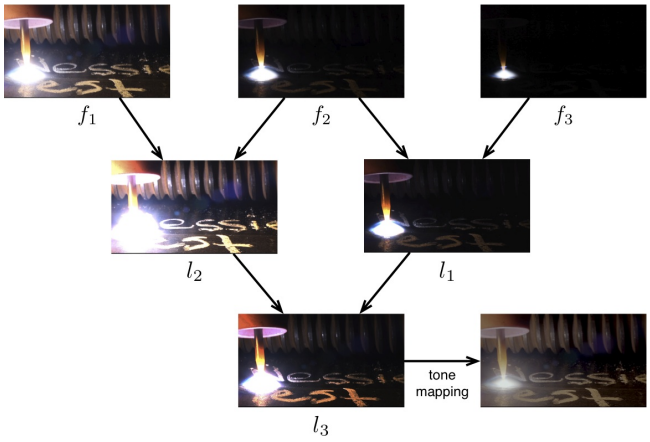

Figure 3: HDR filter in action: dark writing becomes visible next to a very bright light source.[4]

I was personally surprised that the literature did not mention colorblindness, since a dynamic remapping of colors seems like a very useful tool to the many colorblind people in our society. While a remapping of colors will not correct a colorblind person's vision, it would help with color-coded information (such as resistor rings), where the importance lies in distinguishing a specific set of colors at a time.

More complex operations on the images, like the tracking and replacement of objects (e.g. blocking advertisements or highlighting pedestrians) are limited by their latency. If the time between capturing and projection of an image is too long, the user will be irritated to the point of "simulation sickness", leading to nausea, headache and vertigo. (These adverse effects are also the reason why the camera needs to have the perspective of the eye.)

Evaluation

The EyeTap is remarkable in that it completely replaces the real image with a virtual one, while keeping the correct perspective. Other augmented reality systems are usually only able to overlay virtual images onto the real one, whereas a completely virtual image allows a much bigger variety of applications, from correcting eyesight to blocking out advertising. Steve Mann coined the term "mediated reality" to refer to this superset of augmented reality, where information can not only be added but also replaced or subtracted.

As wearable computers become more powerful, we can expect the latency of complex image manipulation operations to drop, making more of them viable.

The biggest question at this moment is not whether the technology is ready (it clearly has been for over a decade now), but whether society is ready. When wearing digital eyeglasses, Mann has repeatedly been the subject of fear, mistrust and abuse. In the long run, camera-based seeing aids seem inevitable, simply because they are so much more powerful than traditional devices. Denying visually impaired people a pair of glasses is obviously cruel, and denial of these more powerful devices will probably be seen the same way – especially because they will be worn by people suffering from more severe afflictions than mere near-sightedness.

To move forward, we will have to become more comfortable with ordinary people casually pointing cameras at us – which at the moment seems very difficult, though that might change if people become more aware of the benefits of digital eyeglasses.

If less radical devices (such as Google Glass) become acceptable, more ambitious systems like EyeTap will hopefully also be tolerated more.

MOTION CAPTURE

Small, unintrusive cameras can not just be used to capture the user's surroundings, they can also be used to capture the users themselves. In many modern motion capture systems, there already is a small camera pointed at the actor's face at all times. This image can later be used as a reference for what the face actually looked like, so errors and missing details in the reconstructed image can be corrected by hand. In their paper[5], Jones et al decided to use this footage for an automatic reconstruction of surface geometry instead, using photometric stereo.

Approach

Instead of surrounding the actor with cameras on a professional lightstage, they decided to only use one camera and multiple light sources at a fixed position relative to the face. The complete device can be seen worn in Figure 4.

Figure 4: The camera lens is surrounded by a ring of LEDs that can be controlled individually.[5]

For every frame of output, they capture four frames of input with different lighting conditions – three different lighting directions using the ring of LEDs and one frame with only ambient light. The ambient light is then subtracted from the other frames, resulting in three images with only one light source at a known position.

Figure 5: Four frames of input, with different LEDs active as illumination.[5]

As they wanted an output framerate of 30 fps, they needed a camera capable of capturing 120 at fps. While these high-speed cameras are more expensive than conventional cameras, they are not much heavier.
It turned out that when cycling through the LEDs at 120 Hz, the flickering in the lights was noticeable and distracted the actors. To reduce this flickering, Jones et al decided to change the lights at 360 Hz while still capturing at 120 fps, which meant that the exposure time of all frames had to be reduced to a third of their time slot.
Assuming lambertian reflectance (meaning that the surface reflects so diffusely that the exact angle between lights and camera become unimportant), the equation to extract the surface normals is very simple:

$$I = L * N A$$

For each pixel of the image, I is the image intensity (which is what the camera measures), L is the matrix of lighting directions (which is known, because we can measure where the lights are relative to the face), N is the surface normal (a vector orthogonal to the surface) and A is the Albedo (the reflectiveness of the surface, a scaling factor for the surface normal vector).

By simply inverting the matrix of lighting directions (which is different for each pixel, but small and therefore cheap to invert), we get the orientation of the surface and even its albedo as a byproduct. Since all pixels are independent, this is a massively parallel workload, that can make full use of GPU hardware. The surface normals of all pixels can then be integrated over to reconstruct the geometry of the face.
To use this approach in practice, Jones et al had to introduce a correction term to compensate for the deviations from the equation's assumptions. Because both the camera and the lights are so close to the face, the lighting directions for each pixel actually depend on the depth of the surface – using the average distance between the apparatus and the face is simply not good enough. By initializing their system with a very generic smooth face, the authors were able to correct the matrices of lighting directions without the need of complicated measurements.

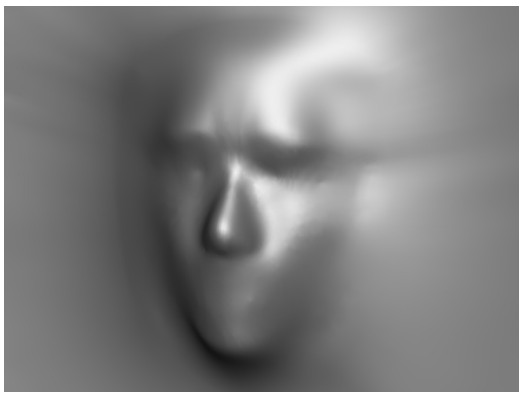

Figure 6: The generic smoothed face used to initialize the system.[5]

Results

Considering that the approach was deliberately kept simple, the results are quite impressive.

Figure 7: From left to right: Albedo, surface normals (color-coded) and face geometry of a single frame of output.[5]

The system was able to capture subtle movements, such as a wiggling of the nose or twitching of an eyebrow, and reconstruct face geometry in real-time. These subtle movements in unpredictable regions are exactly where traditional, marker-based systems break down – as a wiggling of the nose would need to be anticipated by increasing the number of markers covering the nose.

There are some artefacts in the reconstruction, which could be due to a number of factors.

- Very dark shadows turn up as very high albedo (as seen in Figure 7 around the nose). This could be fixed simply by explicitly handling dark shadows.
- If the smoothed face is very different from the face of the user, the matrix of lighting directions will be flawed. Using a more customized initializer could improve this.
- The face might actually not reflect light as diffusely as the equation assumes. This turned out to be the case when using infrared light, but it is unclear how big the effect is when using visible light.
- The device will jiggle slightly, and the implementation does not correct for that effect – which could be done by simply tracking the position of the head as a whole within the image.

Comparison with other approaches

In the same year, Beeler et al published a paper[6], wherein they used multiple cameras, uniform lighting and a novel tracking approach to capture faces. By assuming that some frames (called keyframes) would be typical – meaning that they showed the face in a position it will come back to in the future – they were able to match individual features over time to astonishing accuracy, even to the point of recognising individual pores.

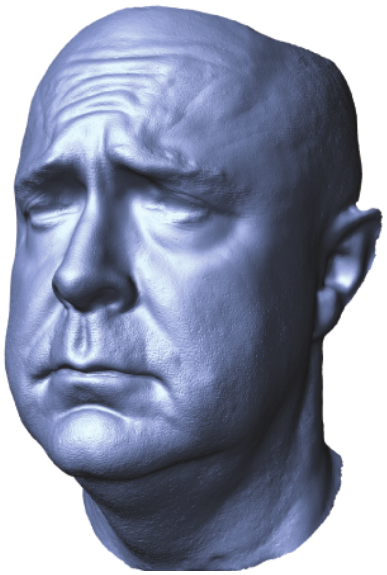

Figure 8: Example of a reconstructed face using Beeler et al's method. [6]

To make full use of these keyframes, the actors were not allowed to move their head, and the lighting had to be held constantly uniform. While these restrictions are acceptable for capturing just faces, they mean the system can not be combined with simultaneous full-body motion capture.

Evaluation

The biggest strength of Jones et al's system is its unintrusiveness – it does not hinder the actors movement, does not block their view and can be combined with any marker-based motion capture system. Comparable modern systems, like the one of Beeler et al, sacrifice these qualities for accuracy, which makes their integration into existing workflows harder. Jones et al's system, on the other hand, can easily replace today's practice of taking reference images for artists, who then manually paint in missing detail into marker-based reconstructions.
As motion capture is a rapidly evolving field of research, we will hopefully see more of these experimental systems, that use innovative hardware configurations and simple algorithms, exploring the possible ways to capture images rather than their processing.
Personally, I would like to see how much more accurate the setup of Jones et al becomes when using more sophisticated algorithms (such as the concept of keyframes). I would also like to see their current system used on different faces – in all examples, we only ever get to see Graham Fyffe, whose face looks very different from the generic initializer and does not have any distinct wrinkles.

SENSECAM

Looking at photographs is a great way to remember past experiences. But actually taking the photographs is a hassle. Handling a camera distracts us from the moment we would actually rather experience than just document, and it is hard to judge if a moment is worth capturing in the first place.
The idea of SenseCam is to have a wearable camera that takes pictures automatically and judges itself whether something interesting is happening or not.

Approach

For their paper[7], Steve Hodges et al designed a small device with a camera chip and multiple sensors, as well as a SD card to store the pictures. An optional bluetooth module allows to connect external sensors, such as a GPS device. The system can be charged over USB, which is also the primary way to transfer captured images to a computer (the SD card is not meant to be removed).
The device uses a fish-eye lens to maximize the field of view. Notably absent is any kind of display, as this would make the whole device much bigger and heavier, drain the battery much faster, and also contradict the idea of not distracting the user.

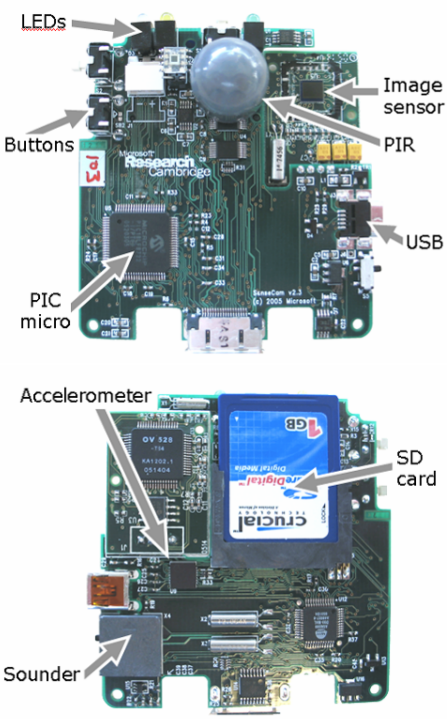

Figure 9: A front and back view of the SenseCam internals. Not labeled is the expansion slot at the bottom, where a bluetooth module can be attached.[7]

While running, the device takes a picture whenever a sensor value exceeds a threshold or a short timer runs out, noting down why this specific picture was taken in a logfile. There is also a button to explicitly trigger a photograph, but no viewfinder – the idea is not to frame the perfect picture but to gather authentic images. The user can wear it on a lanyard around the neck, undistracted from their activity.
Back home, the user then uploads the images to their computer into a library that can be managed and browsed using an application that was written specifically for SenseCam.

Figure 10: A SenseCam device, next to a hand for size comparison.[7]

Use cases

The most obvious application for SenseCam is lifelogging – making a visual diary to review later or share with friends. However, there are several less obvious but more interesting uses for an automatic camera.
It can be used in scientific studies, either to monitor the wearer's behaviour or to look at their typical environment. One interesting idea[8] was to look at the impact of color on our mood – there have been many studies in laboratory settings (where the researcher can control the colors), but using SenseCam, the colors of everyday life can actually be measured. This Replacement of laboratory settings with a measurement of real-world settings could be applied to many known psychological effects.
Reviewing images of experiences can be a useful tool in many forms of therapy, not just to have an objective view of what actually happened (which can help with depression and many forms of delusion), but also to have a basis for discussion that is more meaningful to the patient than hypothetical scenarios.
Probably the most impressive use case however is the treatment of memory disorders, which was explored in the paper with a case study.

Case Study – treatment of memory disorders

It has already been established that reviewing pictures or reading diary entries can help people with memory disorders. But remembering to take pictures or write a diary is especially hard for these people. An automatic camera allows them to take plenty of photographs with ease and without interrupting their activities.
For their case study, Hodges et al gave a SenseCam unit to a 63-year old married woman, who suffered from memory loss as a result of limbic encephalitis (an inflammation of the brain). She would wear it on special occasions, her husband would filter out unusable images and they would regularly review the pictures together and discuss the events.
Her husband noted down important details (such as where they went, and what people they met) and would compare them to what she remembered.
For comparison, they also tried out reviewing a diary written by the husband.
The results, shown in Figure 11, are astounding.

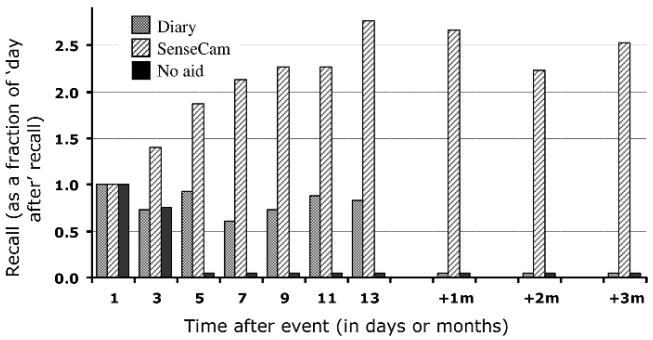

Figure 11: A comparison of SenseCam reviews, written diary reviews and no memory aid at all. The "+3m" means no review for 3 months.[7]

Whereas normally (without any memory aid), she would completely forget an event in less than a week, reviewing a written diary would help her retain her memory, as long as she kept rereading the diary. Reviewing the images on the other hand actually increased the level of details she could remember (including details that were not shown on the images), and she would not lose that memory even when not looking at the pictures for three months.
This suggests that visual memory aids are much more powerful than written diaries, which is convenient as filtering out superfluous SenseCam images is also much easier than writing down detailed accounts of what happened.
Moreover, this result suggests that even some people with severe memory disorders are capable of generating new long-term memories. This is great news not just for them, but also for their partners, who have gotten used to not having any recent shared experiences they both remembered.

Impact

The impressive results of the case study sparked interdisciplinary interest in visual memory aids in general and SenseCam in particular.
The original paper by Hodges et al has been cited over 300 times, and there is a whole conference exclusively about SenseCam-related work (called the SenseCam Symposium).
Microsoft has licensed the design to Vicon, which has been manufacturing SenseCam devices (called the Vicon Revue) commercially for researchers, and is planning to make them available to the general public.
More case studies have been conducted, with people suffering from different kinds of memory disorders, with similarly positive results. It seems that anyone who still benefits from a written diary will also benefit from a visual one.
We can expect automatic cameras (with or without sensors) to become widely adopted as lifelogging becomes more popular. We should also expect more scientific studies to use these cameras to measure the environment the users are actually in, instead of putting them into controlled laboratories.

UBIQUITOUS RECORDING

One big open question is how society will react to ordinary people using automatic cameras – be it for lifelogging, artistic or medical purposes.
Steve Mann's experience wearing his EyeTap device in public has been rather mixed[1], with negative reactions going as far as physical assault.
The irony is that the groups most strongly opposed to normal people taking images are themselves installing and operating surveillance cameras.
Having not just authorities but also normal people record events can be seen as a democratization of recording power. As a counterpart to surveillance, Mann coined the term "sousveillance", meaning the recording of an activity by a participant. It is important to note that sur- and sousveillance are not actually opposites, but rather completely orthogonal concepts. One can advocate an increase in both, or just one, or be against recording activities in general.
While we might not like the idea of random people recording us in public, we should be aware that we are already being recorded by surveillance cameras whenever we enter a bus, shop, museum or restaurant. And if someone really wants to record us, they can simply use a hidden camera without anyone noticing – only seeing aids like EyeTap actually require the camera to be conspicuous.
The big advantage of capturing events ourselves is that we can actually access the images if we need them as evidence. If a crime has only been recorded by surveillance cameras, we need the cooperation of the owner to see the images – which leads to a massive conflict of interest, if these images might incriminate the owner. Footage may be conveniently lost, accidentally deleted or simply described as useless without ever being seen by the victim. If the perpetrator is a person of authority (such as a police officer), photographic evidence may even be the only evidence strong enough to hold in court. Also, while photographs and video can of course be misleading or even fraudulent, they are a much more reliable account of what really happened than eyewitness testimony.
Considering the usefulness of body-worn cameras, we should expect them to become mainstream – much like surveillance cameras already have for the same reason. This means our attitudes and our laws will need to adapt to this new world we will live in. One plausible legal framework would be that anything that we can legally look at, we must also be allowed to record for strictly personal use (such as a seeing aid).
This reflects the idea that when we are uncomfortable with being recorded, it is not because of the recording itself, but the fear this recording might become accessible to third parties.
As for a change in societal attitudes, this is much more difficult to predict. It should however be noted that attitudes can change quite fast, especially if there is a genuinely useful new technology involved (as has been seen with the adoption of internet access, mobile phones and smartphones).

CONCLUSION

We have seen three very different uses of wearable cameras:

- The EyeTap replaces the eye with a camera and reality with a virtual image.
- A photometric stereo system can capture the face with just one camera.
- SenseCam takes pictures for you completely automatically.

All three systems are already very mature and could easily be used in their current state, but not all of them are.

- EyeTap is used by Steve Mann, but his experience shows that society is currently not very tolerant of people conspicuously wearing cameras. This may change quickly, but until then, very few people will want to actually wear one.
- There is a lot of movement in the field of motion capture, so Jones et al's system might be overtaken soon – and projects that care much about capturing the face might actually prefer a more accurate system, even if it means restraining the actors much more.
- SenseCam is the only system that is already being widely adopted – and we should expect automatic cameras to become mainstream gadgets very soon.

One thing that has become clear is that cameras can be the basis for a wide variety of genuinely useful applications for ordinary people, from documenting our lives to enhancing our vision and even our memory.

While concerns about privacy in a world full of cameras are valid and will need to be discussed, in the long run the benefits of camera-based helpers will simply be too big to miss out on.

We already all have a camera with us all the time.
Soon, we will also have a camera ready all the time.
And soon after, we will be recording all the time.
And we will not want to go back.

REFERENCES

1. Steve Mann. Continuous Lifelong Capture of Personal Experience with EyeTap. CARPE'04. 2004.

2. Steve Mann. Through the Glass, Lightly. IEEE Technology and Society, Vol. 31, No. 3, Pages 10-14. 2012.

3. Steve Mann. My 'Augmediated' Life. IEEE Spectrum, March 1. 2013.

4. Steve Mann, Raymond Chun Hing Lo, Kalin Ovtcharov, Shixiang Gu, David Dai, Calvin Ngan, Tao Ai. Real-time HDR (High Dynamic Range) Video for EyeTap Wearable Computers, FPGA-Based Seeing Aids, and GlassEyes. IEEE CCECE 2012. 2012.

5. Andrew Jones, Graham Fyffe, Xueming Yu, Wan-Chun Ma, Jay Busch, Ryosuke Ichikari, Mark Bolas and Paul Debevec. Head-mounted Photometric Stereo for Performance Capture. 2011 Conference for Visual Media Production, Pages 158-164. 2011.

6. Thabo Beeler, Fabian Hahn, Derek Bradley, Bernd Bickel, Paul Beardsley, Craig Gotsman, Robert W. Sumner, Markus Gross. High-Quality Passive Facial Performance Capture using Anchor Frames. ACM Trans. Graph. 30, 4, Article 75. July 2011.

7. Steve Hodges, Lyndsay Williams, Emma Berry, Shahram Izadi, James Srinivasan, Alex Butler, Gavin Smyth, Narinder Kapur and Ken Wood. SenseCam: A Retrospective Memory Aid. Ubicomp 2006, Pages 177-193. 2006.

8. Aiden R. Doherty, Philip Kelly, Brendan O'Flynn, Padraig Curran, Alan F. Smeaton, Cian O'Mathuna, and Noel E. O'Connor. Effects of environmental colour on mood: a wearable LifeColour capture device. Proceedings of the international conference on Multimedia (MM '10). 2010