Eulerian Video Magnification for Fingerprint Liveness Detection

Nicklas Overgaard[1,2], Ctirad Sousedik[1], Christoph Busch[1]

[1]Norwegian Biometrics Laboratory
Gjøvik University College
Teknologiveien 22
2815 Gjøvik, Norway
ctirad.sousedik@hig.no
christoph.busch@hig.no

[2]Technical University of Denmark
Anker Engelunds Vej 1
2800 Kgs. Lyngby, Denmark
nicklas@isharp.dk

Abstract

The state-of-the art fingerprint biometric systems are potentially vulnerable to spoofing attacks by means of artefact fingerprints that can be fabricated using low-cost, widely-available resources and methods. As most to date biometric applications require cost intensive hardware for the capture device, it would be beneficial to utilize the resources provided by the widely spread Smartphone devices in order to develop a fingerprint capture solution that would include anti-spoofing countermeasures in terms of Presentation Attack Detection (PAD). This article examines the applicability of Eulerian Video Magnification method to emphasize the heartbeat-related color variations of the genuine living fingers as a means of distinguishing between genuine and artefact fingers.

1 Introduction

In this world of ever increasing digitization, there is a strong demand for secure and tamper-proof authentication. Passwords and passphrases have been around for a long time, but have many times proven to be insufficient for providing a secure and convenient authentication factor. The development over the past few years, as regards portable devices, has made the thing even more complicated; recent research has outline that approx. 80 percent of the smartphone users deactivate PIN based authentication due to lacking convenience [1]. One way of providing a more user friendly authentication is to observe a biometric characteristic of the user.

For over a century, the forensic science has used fingerprints to successfully convict criminals, due to the fingerprints' uniqueness property[2] and their stability over the lifetime of the individual [3][4]. The fingerprints have also been used to check physical access to information sensitive facilities and protected areas. Ten years ago, IBM launched the first laptop capable of enrolling the user with his fingerprint. Recently, the Indian Universal ID program has been spawned, in which the Indian

This paper was presented at the NIK-2014 conference; see http://www.nik.no/.

government seeks to provide a unique identity to India's 1.2 billion residents. Most recent Apple has integrated a dedicated fingerprint sensor into its newest product. However, all these applications require additional hardware capable of capturing the fingerprints. Moreover expectations with regards to security and reliability could not always be met. Thus facing the requirement for a convenient, secure and reliable authentication without implications of additional cost we are motivated to investigate the use of widely spread, generic hardware such the camera in state-of-the-art Smartphones [5].

When investigating the security of a Smartphone based fingerprint recognition system we must anticipate that it can potentially be circumvented by an accurate imitation of the scanned fingerprint characteristic, and so the fingerprint capture pipeline needs to be equipped with a Presentation Attack Detection (PAD) module, in order to verify whether the scanned fingerprint stems from a genuine living finger, rather than being an artefact. Currently, the International Standardization work in ISO/IEC develops a framework for describing and testing PAD modules [6], as well as providing a standardized taxonomy and vocabulary for addressing PAD tasks.

This article analyses the applicability of the recently published Eulerian Video Magnification (EVM) approach [7] as a means of developing a reliable Presentation Attack Detection (PAD) module for fingerprint sensing by using generic Smartphone cameras. The EVM was designed to magnify subtle movements and changes of colors in the analysed videos. The material that was provided with the initial publication[7] has demonstrated to reveal changes in colors caused by heartbeat in a generic video of a face. In addition, the EVM method has been demonstrated to be capable of visualizing the subtle movements of the main wrist vein, also caused by the heartbeat. These properties seem to suggest the EVM as an ideal candidate for biometric liveness detection that could be well suited for the fingerprint sensing scenario using video material from generic Smartphone cameras.

2 EVM for video-based fingerprint liveness detection

EVM is a video post processing technique that seeks to emphasize subtle temporal variations in the video that are hard or even impossible to see with the naked eye. As shown by Figure 1, the EVM method analyses the source video in a multi-resolution manner, using the Laplacian pyramid to decompose the original video into different spatial frequency bands. For each of the bands and for every position $\mathbf{x}$, a band image $I_b(\mathbf{x}, t)$ provides for a 1D temporal signal along the length of the original video. For every spatial band image $I_b(\mathbf{x}, t)$, the EVM method computes a filtered band image $B_b(\mathbf{x}, t)$ by means of band-pass filtering of each of the temporal 1D signals in the desired frequency range - frequency of the heartbeat. The amplification is performed by adding the filtered band image $B_b(\mathbf{x}, t)$, multiplied by a factor α, to the band image $I_b(\mathbf{x}, t)$:

$$\hat{I}_b(\mathbf{x}, t) = I_b(\mathbf{x}, t) + \alpha B_b(\mathbf{x}, t)$$

After the magnification has been performed for all the spatial bands $I_b(\mathbf{x}, t)$, the entire pyramid is collapsed back into a single video stream that emphasizes changes that occur within the above mentioned frequency range.

Apart from magnifying color changes in a specific frequency band, the EVM

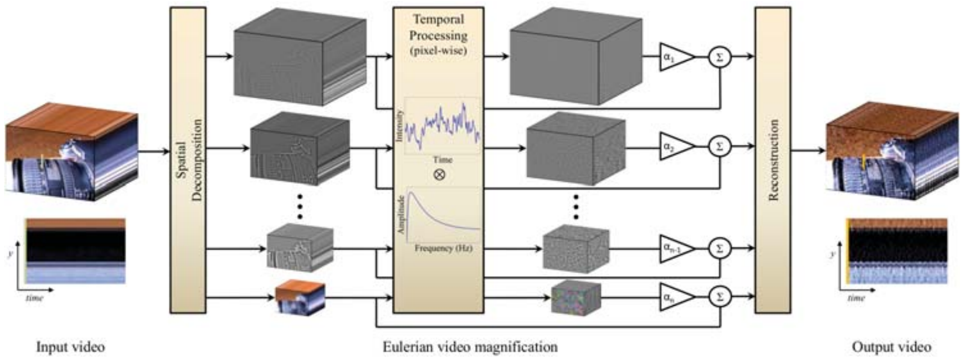


Figure 1: Overview of the EVM framework. The video is spatially decomposed using the Laplacian pyramid, the desired temporal frequencies are amplified and finally the pyramids are collapsed and folded back into the video [7]

method can emphasize subtle movements in the input video stream. The authors of the EVM approach support this by the fact that the filtered band image $B_b(\mathbf{x}, t)$ can be viewed as the term

$\delta(t)\frac{\partial f(x)}{x}$ in the function

$$\tilde{I}(x,t) \approx I(x,t_0) + (1+\alpha)\delta(t)\frac{\partial I(x,t_0)}{\partial x}$$

which is a first order Taylor series approximation of the function

$$I(x,t) = I(x + (1+\alpha)\delta(t), t_0)$$

that represents a model of the pixel value changes in a 1D spatial signal that undergoes a translational motion. The authors claim the idea can be generalized to 2D translational motion.

As the EVM approach has been able to emphasize the pulse in a person's face, the above mentioned property of the EVM method suggests to observe subtle pulse-related changes in colors of a video recording from a genuine living finger. A method that would be capable of automatically assessing the presence of these color variations could be turned into a liveness detection solution for the Smartphone fingerprint sensing scenario.

In order to assess the applicability of the EVM method for our purpose, the publicly available MATLAB source code was used in a feasibility study. All the experiments were performed on a quad core Intel Core i5-3570K processor @ 3.4 GHz, with 16 GB of RAM on a Linux installation, with MATLAB R2012b. For video capturing three different phones were at hand; each model with technical specifications can be seen in Table 1. All of the videos were captured using the highest available quality settings at the Full HD (1080p) resolution.

Device	**Abbreviation**	**Specifications**
Samsung Galaxy Nexus	Nexus	Dual Core 1.2 GHz, 1 GB RAM
Samsung Galaxy S3	SGS3	Quad Core 1.4 GHz, 1 GB RAM
Sony Xperia V	Xperia	Dual Core 1.5 GHz, 1 GB RAM

Table 1: Mobile devices used during experimentation

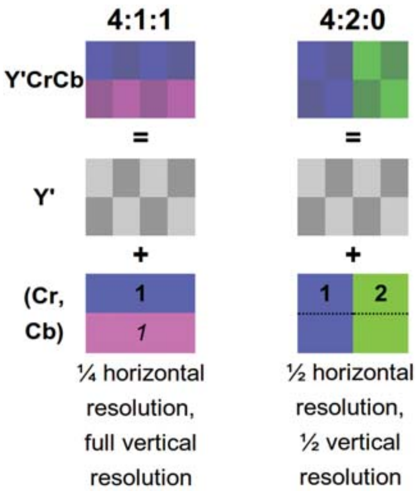


Figure 2: Luma/chroma subsampling. Image taken from [9]

H.264 codec implications

At the resolutions necessary for the fingerprint sensing scenario, the size of the recorded video stream does not allow for a real-time software-based analysis and storage in the raw format using the computational resources provided by the current Smartphones. The existing Smarthone video recording solutions employ a hardware-accelerated video compression in order to decrease the required amount of computational resources as well as the size of the resulting video stream. Nowadays, the H.264 AVC (Advanced Video Coding) is a ISO/IEC coding standard [8] employed for video recording and delivery on a majority of Smartphones and camcorders. Similar to other video encoding solutions, the H.264 seeks to minimize the size of the resulting video stream, while minimizing the degradation of the quality perceived by the human eye. As the information to be emphasized by the EVM approach is below the threshold of the human perception, the details of the AVC codec implementation can have direct influence on the ability of the EVM method to magnify the subtle pulse-related changes in the output video stream.

The basic building blocks of AVC are *macroblocks*, which are block shaped entities containing luma (Luminance, brightness) and chroma (Chrominance, color) samples. Studies during the early days of color televisions showed that the human eye is more sensitive to changes in luma than chroma. This can be directly exploited by codecs to consume less bandwidth, while matching the human perception [10].

In AVC this is achieved by representing colors in the YCrCb space. *Y* is the luma component, representing the brightness of the image. *Cb* and *Cr* represent the deviation from grayscale in the blue and red channel, respectively. When a macroblock is coded, the color sampling is done in such a way that the chroma components have one fourth the number of samples of luma—half the amount of samples in both the vertical and horizontal lane. This is called 4:2:0 sampling and has 8 bits of precision per sample. Luma/chroma subsampling is depicted in Figure 2. Macroblocks are by definition partitioned in a rectangular shape of 16x16 samples of the luma component and 8x8 samples of each of the two chroma components.

Macroblocks are then grouped into *slices*. An image frame can be split into one or several slices and each slice can be encoded using different methods.

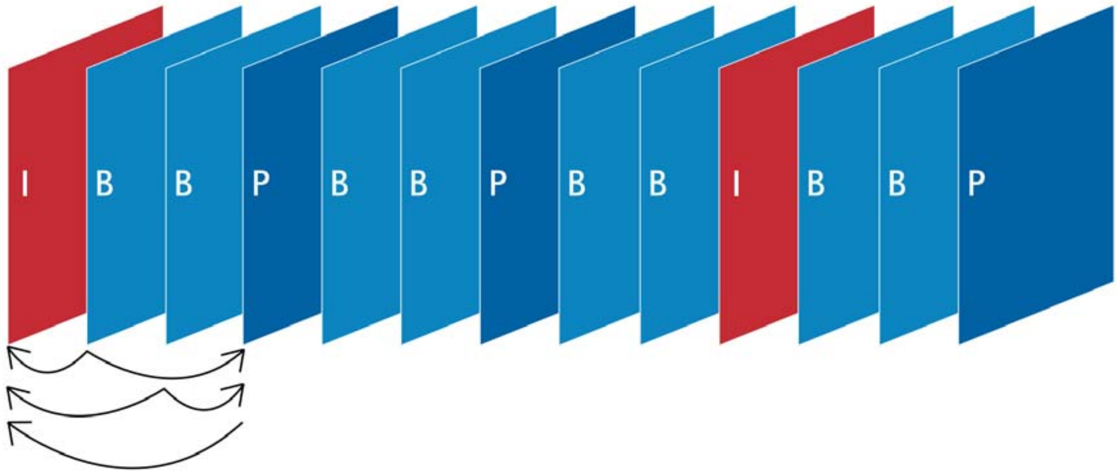


Figure 3: Example of frame sequence in a video coded with AVC. Image taken from [11]

I Slice All macroblocks are encoded using *intra* prediction. These frames are independent frames which can be decoded without the use of other frames.

P Slice In addition to the features of the I slice, some macroblocks of a P slice can be encoded using *inter* prediction, with a maximum of one motion-compensation signal per prediction block. Motion compensation is mentioned later in this section.

B Slice In addition to P slice features, some macroblocks of a B slice can be encoded using two motion compensation blocks per prediction block.

As a consequence, the P slices can only refer back to previously seen I or P slices, whereas B slices can point both backward and forward in the encoded stream. An example of a frame sequence is shown in Figure 3.

The key method employed in the above mentioned scheme in order to save storage space is *motion compensation*. This technique describes differences between frames as transformations that should be applied to previously seen objects (in an I- or P-slice) to make them appear at the correct place in the current B- or P-slice. In AVC motion compensation is applied on a block level with varying accuracy, depending on the quality settings. Storing only transformations in a macroblock drastically decreases the required bandwidth. A visual example of motion compensation is depicted in Figure 4.

The above mentioned properties of the AVC codec could interfere with the information that is aimed to be magnified by the EVM approach. The chroma subsampling is optimized for the sensitivity of the human visual system and could potentially reduce the amount of information regarding subtle color changes in the encoded video stream as compared to the raw video stream.

The motion compensation is attempting to capture movements in the video via motion vectors. However, the EVM is typically applied in order to magnify motions of a very subtle nature that do fall below the perception capabilities of the human vision system. For some settings of AVC, it is possible that the small movements of the pulse in an arm will not be qualified as a motion, effectively resulting in the codec discarding the small changes from the encoded stream.

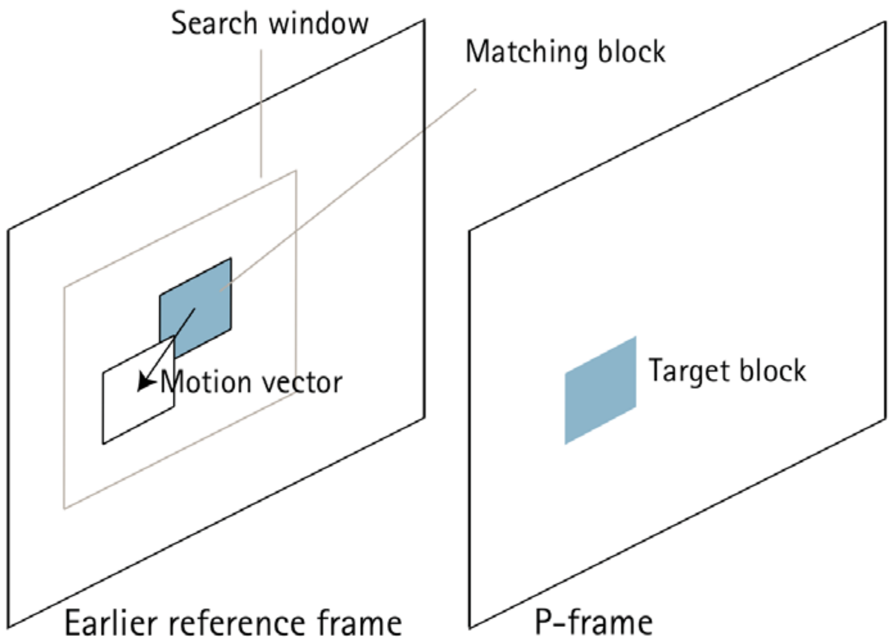


Figure 4: Example of simple motion compensation in the AVC codec. Image taken from [11]

Due to the amount of computational resources provided by the current Smartphones and the hardware-wired video encoding circuitry, it is not possible to discard the video encoding step from the video analysis pipeline. Given the amount of data generated by a raw video stream, such an analysis would be impractical even for other hardware-based solutions, where greater computational power and storage resource are available. To mitigate the issue, all of the videos analysed were recorded using the highest available settings, namely Full HD (1080p) with super-fine settings.

3 Experiments

The initial set of experiments was carried out in order to verify the applicability of the EVM approach to videos recorded by the Smartphone cameras available. A set of videos of still scenes with fixed camera positions were recorded and afterwards the EVM was applied to emphasize temporal changes at the heartbeat frequency range of 1-2 Hz. Even though the still videos were expected not to generate any significant responses after the EVM magnification, the magnified streams from the Nexus device had expressed a constantly appearing periodical brightening/darkening pattern that moved around the entire area of the video in a rather unpredictable fashion. A single representative frame from the magnified video stream is shown by Figure 5b. As the effect was not observed in magnified videos obtained by using the other two Smartphones, the Nexus device was discarded as unfit for the task. However, the clean unmagnified videos captured using this device suggest that the problem might reappear with other devices as the quality requirements on the sensor are much higher if a stable EVM magnified video is to be acquired.

The EVM authors "Pulse in face" and "Pulse in wrist" videos were attempted to be reproduced with the cameras of the remaining Smartphones. The attempt was successful but important limitations of the EVM algorithm were uncovered during

(a) Nexus Snow scene, no EVM (b) Nexus Snow scene, with EVM

Figure 5: Nexus Snow scene frame 143 with and without EVM, full video at http://youtu.be/AYWzZWTLUwk

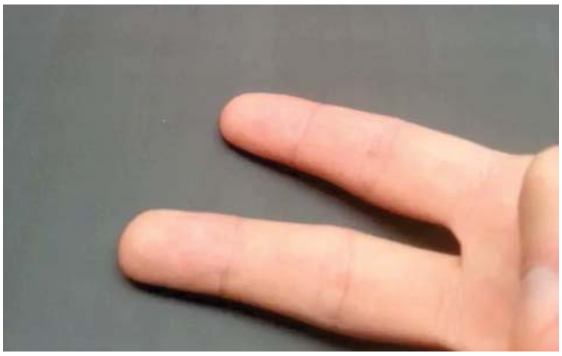
this period.

First and foremost, EVM is highly sensitive to slight movements in the source video. Due to muscle shakes, holding the camera in the hand while recording the wrist adds so much disturbance to the video that the pulse signal is "drowned" when performing the amplification. The EVM emphasizes temporal changes in the source video by amplifying localized temporal signals in a multi-resolution manner. Both the variations in colors and movements cause temporal changes of the pixel values, and they are difficult to distinguish from one another as long as they both occur in the emphasized frequency range. If the camera or the recorded object are not kept completely still, the slightest movements can easily cause strong changes in colors as well as a perception of strongly emphasized movements in the EVM magnified video. Typical muscle shakes are significantly correlated with the persons pulse. This fact along with the way the Smartphones are typically used, pose a serious issue for developing a fingerprint liveness detection solution for Smartphones by means of the EVM approach.

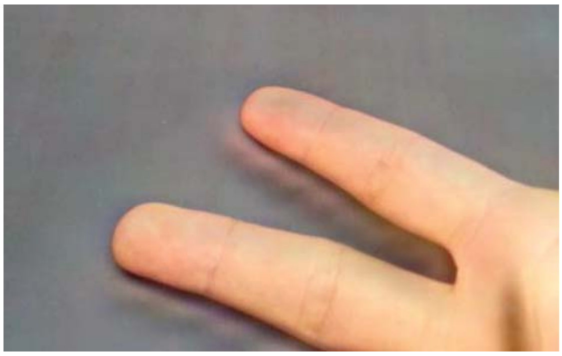
Detecting the pulse in the finger

The above mentioned experiments resulted in the definition of a set of strict constraints in order to verify whether the EVM approach could work under controlled conditions:

1. The Smartphone camera and the recorded finger must be fixed during video capture



(a) SGS3 finger, no EVM



(b) SGS3 finger, with EVM

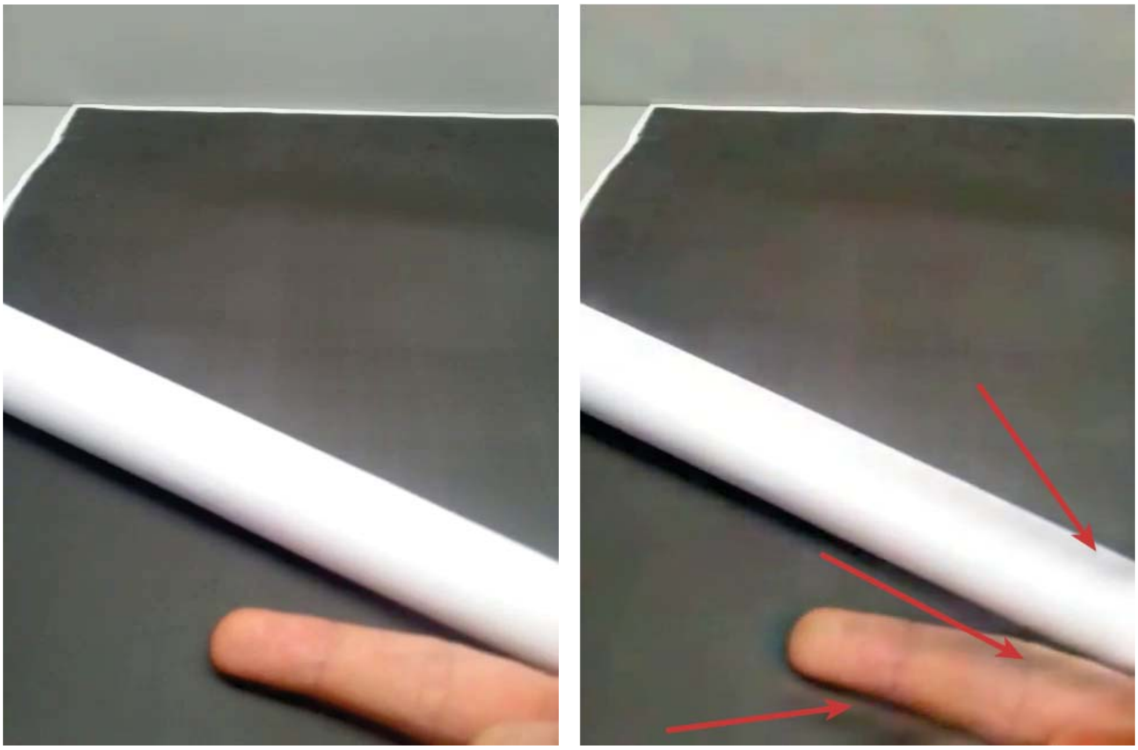
Figure 6: SGS3 video frame of "still" fingers with and without magnification. (b) clearly shows color artifacts along the lower side of the fingers. (full video: http://www.youtube.com/watch?v=POYRtKxKakI)

2. Lighting conditions should be controlled to avoid impact caused by shadows or other changes in the illumination of the scene

These constraints constitute strong limitations that render the EVM method for day-to-day application not suitable. A stable controlled light source that imitates daylight was used, with an additional possibility to add UV backlight. The fingers were held still on a stable surface and the position of the Smartphone was fixed during the recording. Even though the results did reveal periodic patterns in the heartbeat frequency range after the EVM magnification, the further experiments have demonstrated that the patterns are caused by small movements of the fingers that could not be held absolutely still even on a flat stable surface in the controlled scenario. This conclusion is strongly supported by the experiment shown in Figure 7. A piece of paper was tightly wrapped around one of the fingers in order to remove any influence of the pulse-related color changes, while keeping the subtle movements. The resulting magnified video had shown that the wrapped piece of paper generated the same periodic pattern as the not covered finger, which demonstrates that the pulse-related color changes in the fingers are very difficult to visualize even under ideal conditions that are not realistic in many scenarios. The additional mild periodic changes in the static parts of the video further support the hypothesis of the significant influence of the camera sensor noise on the magnified result.

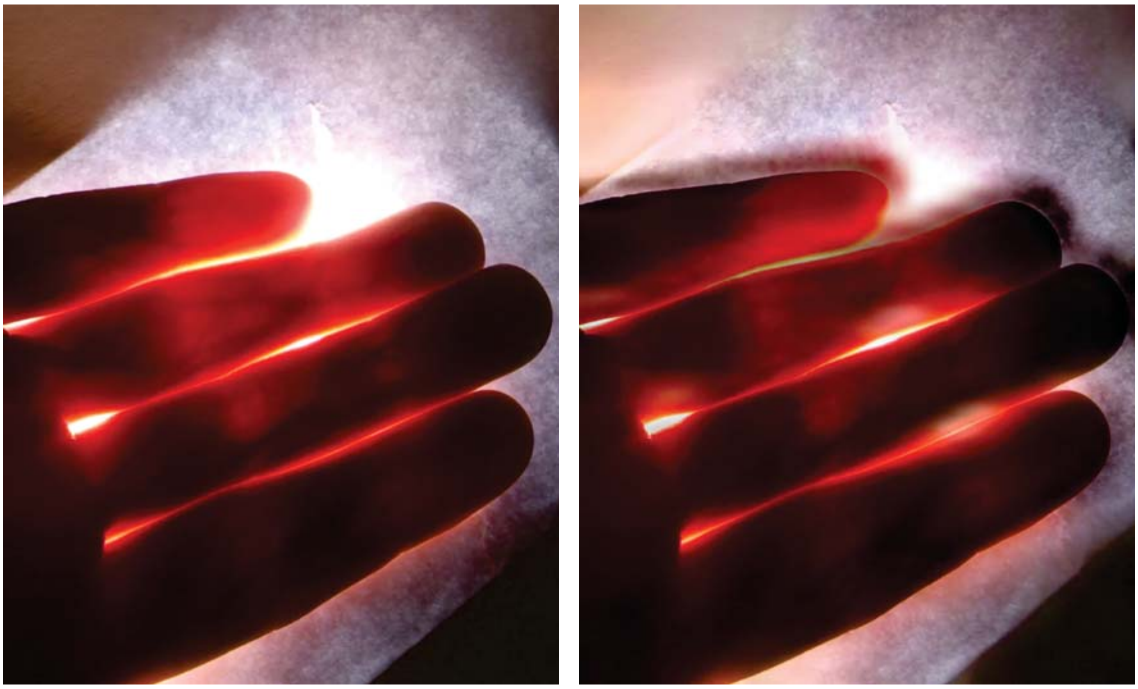
In the second experiment, the hand was exposed to a strong controlled backlight in order to make the vein structure partly visible in the recorded video. A frame from this capture scenario can be seen in Figure 8. The magnified video did not reveal any observable patterns that would suggest that the pulse was visualized. The periodic pattern in the magnified video exhibited the same properties as in the previous experiment, suggesting that it is caused by the slight movements of the hand held still, rather than by the pulse-related variations.

In the next experiment, a fingertip was directly pressed against the lens of the camera with a strong light source in the background. This setting was supposed to suppress the movement related variations, while keeping the pulse-related color variations as a part of the out-of-focus video stream. Nevertheless, the magnified video did not reveal any positive result.



(a) SGS3 finger, no EVM (b) SGS3 finger, with EVM

Figure 7: SGS3 video frame of fingers having one finger covered. (b) clearly shows color artifacts along the finger and covered finger. Red arrows are pointing to areas of interest (full video: http://youtu.be/nS7Z5QFHlmA)



(a) No magnification (b) Color magnified

Figure 8: Video frame of fingers with strong backlight. In (b) color skew is visible, but no real color amplification (Full video: http://youtu.be/nVnVsG_yCIs)

Detecting pulse in the wrist

The authors of the EVM method have demonstrated the capability of their method to visualize the pulse-related motion of the main wrist vein. Motivated by their work a further experiment was designed in order to analyse whether this magnification and visualization is feasible for a video that was recorded using the camera of a hand-held Smartphone. The results have again demonstrated the importance of a fixed camera position. Unless the camera position was fixed, the other periodic patterns and variations made the pulse-related signal impossible to observe. In order to remove the negative influence of the muscle shakes of the hand that was holding the Smartphone camera, the video was processed by the Deshaker [12] video stabilization plugin for VirtualDub [13]. However, as the Deshaker only handles vertical shakes in the X-Y directions, and not dislocations in the Z direction, the deshaked videos still contained some movement, causing the resulting videos to include magnification artifacts and not the sought pulse.

4 Conclusion

The EVM approach is by definition extremely sensitive to the slightest changes in the source video. This fact requires the high quality videos that are recorded using a fixed camera position and makes the EVM method very difficult to apply with current hand-held Smartphones. In addition, even in a scenario when the fingers are held still on a flat surface and the camera position is fixed, the slight muscle shakes in the fingers still cause strong responses in the resulting EVM magnified video. Even in the ideal controlled conditions, it was not possible to observe patterns that would be clearly related to the color variations caused by the pulse, and thus even validate the presence and the sufficient strength of this signal. Despite the fact that the videos were recorded using the highest possible quality settings, the AVC codec might be partly responsible for the negative result. However, the amount of data generated when recording a raw uncompressed video stream makes it unfeasible to analyse the raw stream using the computational power of current Smartphones. Thus we conclude that despite the promising results that the inventors of the EVM method have published, the approach is unlike to be used in the near future for liveness detection of camera based fingerprint recognition on Smartphone. We do expect future research to address this relevant security topic by reflection analysis or multispectral analysis of the captured scene.

References

[1] C. Nickel, *Accelerometer-based Biometric Gait Recognition for Authentication on Smartphones.* Phdthesis, Technische Universität Darmstadt, 2012.

[2] Y. Han, C. Ryu, J. Moon, H. Kim, and H. Choi, "A study on evaluating the uniqueness of fingerprints using statistical analysis," in *Information Security and Cryptology â ICISC 2004* (C.-s. Park and S. Chee, eds.), vol. 3506 of *Lecture Notes in Computer Science*, pp. 467–477, Springer Berlin Heidelberg, 2005.

[3] R. Müller, *Fingerprint Verification with Microprocessor Security Tokens.* PhD thesis, München Techincal University, 2001.

[4] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, *Handbook of Fingerprint Recognition.* 2nd ed., 2009.

[5] C. Stein, V. Bouatou, and C. Busch, "Video-based fingerphoto recognition with anti-spoofing techniques with smartphone cameras," in *IEEE 12th International Conference of the Biometrics Special Interest Group (BIOSIG)*, (Darmstadt), pp. 1–12, Sept. 2013.

[6] ISO/IEC, "Information technology – biometrics – presentation attack detection – 30107." http://www.iso.org/iso/catalogue_detail.htm?csnumber=53227.

[7] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. T. Freeman, "Eulerian video magnification for revealing subtle changes in the world," *ACM Trans. Graph. (Proceedings SIGGRAPH 2012)*, vol. 31, no. 4, 2012.

[8] ISO/IEC, "Information technology – coding of audio-visual objects – part 10: Advanced video coding – 14496." http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=61490.

[9] Wikipedia, "Chroma subsampling." https://en.wikipedia.org/wiki/Chroma_subsampling.

[10] D. A. Kerr, "Chrominance subsampling in digital images." http://dougkerr.net/pumpkin/articles/Subsampling.pdf, January 2012.

[11] A. Communications, "H.264 video compression standard.." http://www.axis.com/files/whitepaper/wp_h264_31669_en_0803_lo.pdf, 2008.

[12] G. Thalin, "Deshaker." http://www.guthspot.se/video/deshaker.htm, August 2011.

[13] A. Lee, "virtualdub.org." http://www.virtualdub.org/.