https://doi.org/10.2352/ISSN.2470-1173.2019.4.PMII-586

Issues reproducing handshake on mobile phone cameras

François-Xavier Bucher, Jae Young Park, Ari Partinen, Paul Hubel; Apple Inc.; Cupertino, California/USA

Abstract

This work was carried out to serve two purposes:

- *Create and share a single motion profile that emulates the handshake of a population of mobile phone users taking still photos under real life conditions.*
- *Describe the validation procedure required to ensure the high fidelity motion platform chosen can correctly reproduce the proposed motion profile.*

By means of psychophysical testing, we examined the frequency and spatial characteristics of human handshake, based on which we created synthetic handshake profile with very similar properties.

We demonstrate how the proposed motion trace correlates extremely well with real handshake and why using a realistic motion profile together with a high fidelity motion platform is crucially important in order to avoid disturbances not present with real users.

Context

- Developing a protocol allowing a fair comparison of stabilization amongst mobile phones.
- Developing a protocol that matches real users, meaning not just a constant profile.

Objectives

This work presents a profile that can correlate with actual human hand-held movement when played on a voice coil motor motion platform.

We started by conducting a user study to record actual hand-held motions and extracted key characteristics that we considered representative of the motion distributions. From those characteristics we generated a synthetic profile that was optimized to match the motions' distributions. This profile was also generated to be periodic so that it could be played repeatedly. We ensured our motion platform was able to play this profile rather accurately, without introducing undesired noise. We finally validated our profile by comparing sharpness from images captured by actual users with the motion platform.

For this work, we used the following hardware:

- A mobile phone that was commercially available to capture images. We defined pitch as being rotation along the device height, yaw along its width and roll along the lens axis. See Figure 1.
- A motion platform equipped with voice coil motor to simulate traces (VCMMP).
- An inertial measurement unit (IMU) for motion recording.

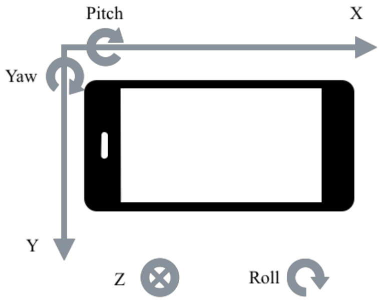

Figure 1: Axis definitions

User hand-held motions collection

We started by collecting an ensemble of hand-held motions through a user study. Our goal here was to be able to better understand typical hand-held movement.

Protocol

The protocol for the user study was the following:

- The IMU is attached to the mobile phone
- Participants were instructed to hold the mobile phone for 45s while trying to frame a chart at 2 meters with the help of the camera app preview. No images were captured.
- During that period, the IMU recorded angular rate.
- We only used the last 30s of the recording, in case the user was still adjusting his or her position at the beginning of the recording.
- For each participant, the angular trajectory was reconstructed by integrating the angular rates using the initial position as the reference coordinates system.

Results analysis

We analyzed two aspects of the trajectories obtained:

- The angular amplitude distribution: In a simple pinhole model, when the subject is far enough, the blur generated by a motion is proportional to the angular amplitude of the motion. This is also compatible with the rule of thumb that states that the exposure speed should be less than the inverse of the focal length, since the amplitude of the image movement is also proportional to the focal length.
- The power spectral distribution: image stabilization systems are dynamic and, similar to most dynamic systems, they are sensitive to the frequencies they have to correct.

Since hand-held movements are not well defined, we considered analyzing these distributions.

Amplitude distributions

We estimated the amplitude of motion between two timestamps as being 6 times the standard deviation of the trajectory. By calculating all the possible amplitudes for all users for a given duration, we can calculate a statistical distribution of the amplitudes. An example of such distribution for 1/10 s duration for the pitch rotations is given in Figure 2.

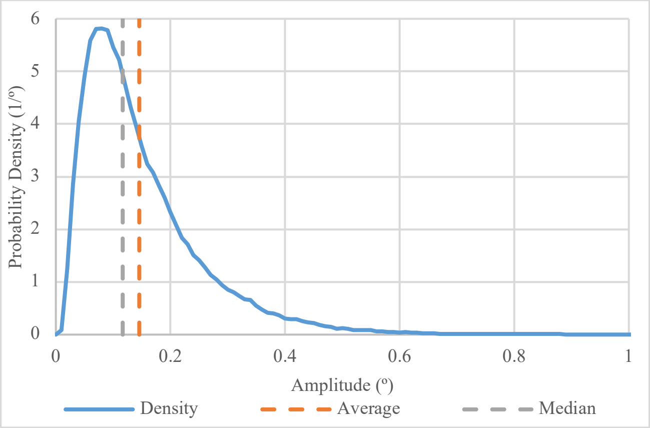

Figure 2: Pitch amplitudes distribution for 1/10s

On this distribution, the median value is about 0.12º. In photographic terms, if a system can only compensate pitch rotation of up to 0.12º, 50% of the captures at 1/10 s will appear blurry.

These measurements also help justify neglecting translations over rotations: for an object located at 1 m, a translation needs to be about 2 mm over 0.10 s, or 2 cm/s, to generate a blur of similar amplitude than a rotation of 0.12º. This equivalent translation is much larger than a normal hand-held motion.

Similar amplitude analysis can be done for any duration along all axes. The following three figures show the evolution of key points of the distribution between 0 and 1 seconds along all three rotational axis.

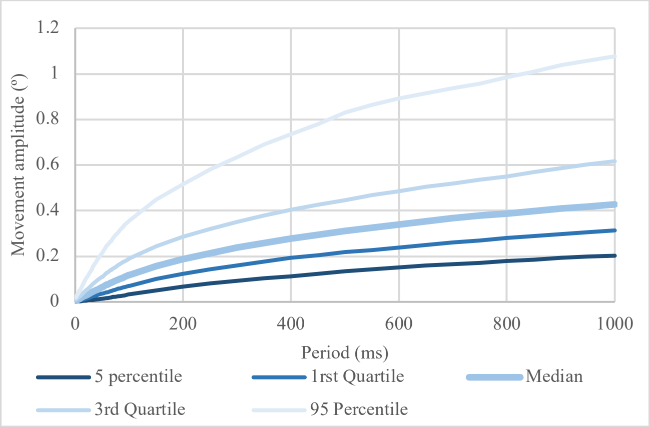

Figure 3: Pitch amplitude distributions

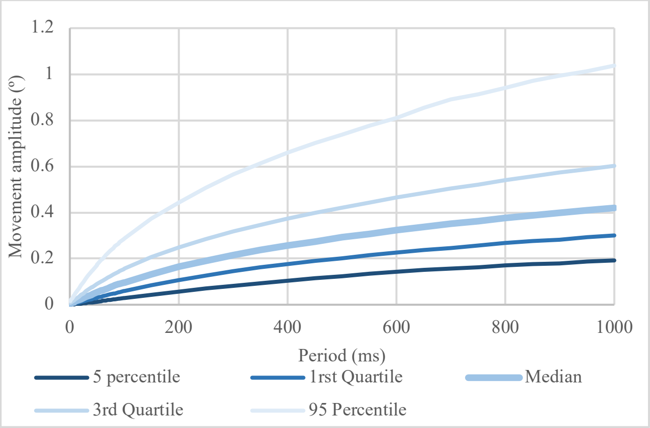

Figure 4:Yaw amplitude distributions

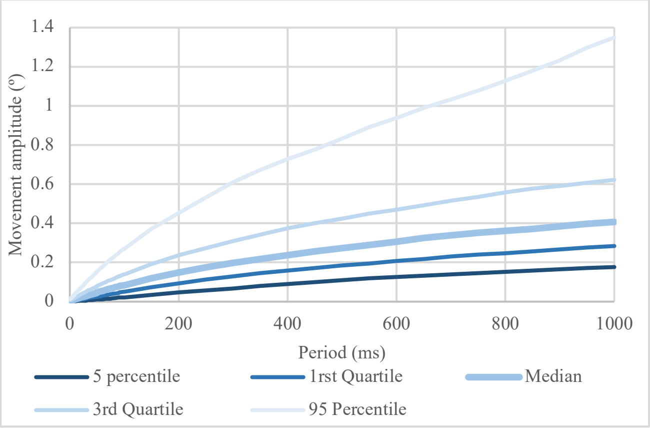

Figure 5:Roll amplitudes distributions

The following conclusions can be drawn from figures 3-6:

- The distribution of amplitudes are quite large for all periods and axis with ratio between the 5 and 95 percentiles being between 5 and 20.
- The distributions are quite similar along all three axes. For instance, median values are of about 0.4º at 1 s for all axes.

Power spectral distributions

In a similar manner we did for the amplitudes, we conducted a statistical analysis of the power spectral distributions. To calculate those distributions, we extracted PSDs over a 2 second period. Since there is no reason for a motion to be cyclical over a 2 second period, we use the 1D periodic profile plus a smooth transformation as described in [1].

In addition, we estimated the noise spectrum of our IMU by calculating its PSD for a static recording.

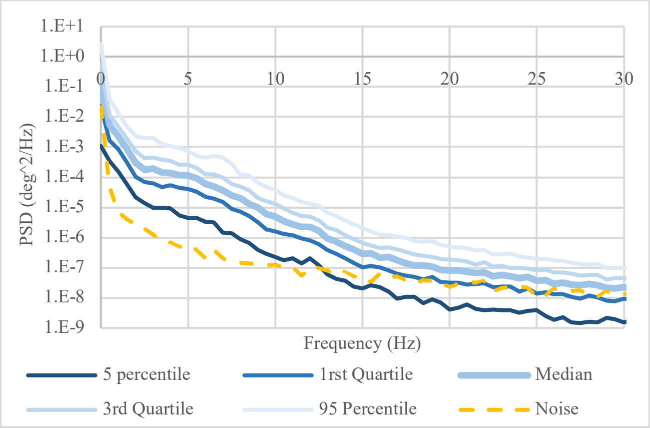

Figure 6:Pitch PSDs distributions

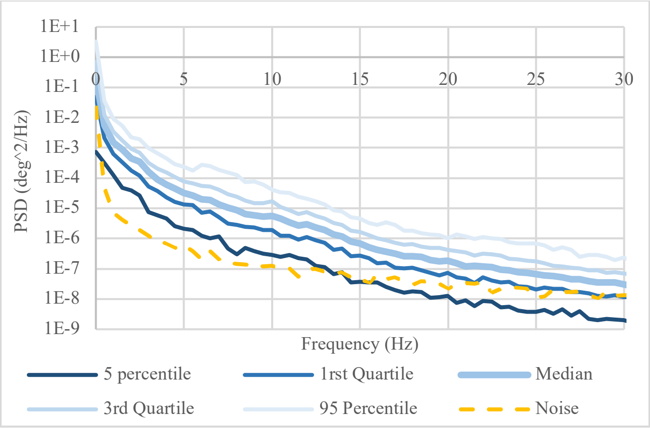

Figure 7:Yaw PSDs distributions

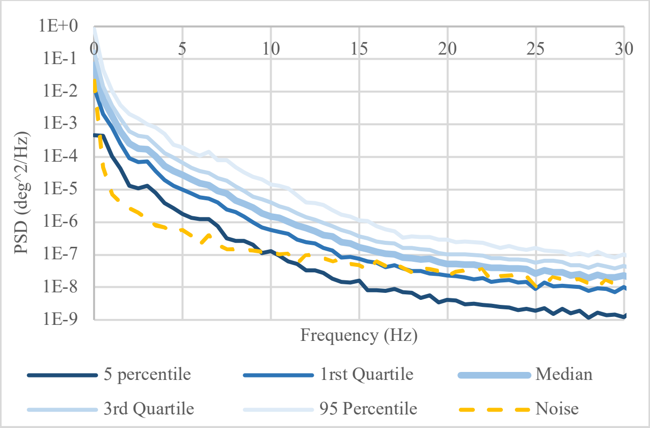

Figure 8:Roll PSDs distributions

From the three above figures we can draw the following conclusions:

- Spectrums are similar along all axis and none seems to be significantly more energetic than the others.
- No particular peak is visible and therefore no frequency seems to be significantly more energetic.
- Our measurement above 15 Hz are of the same order of magnitude than the noise from the IMU, and therefore unreliable. However, the PSD at those frequencies are very low, less than 10^{-6}, and therefore we consider them to be negligible.

For the following profile generation, we will consider the spectrum above 15 Hz to be null. Assuming frequencies above 15 Hz is quite common in the literature, such as in [2]: we believe this to be incorrect.

Generation of synthetic profile

Requirements on a synthetic profile

Our final goal is to generate a single profile that statistically matches user distribution and that can be played continuously on a motion platform. This translates into the following requirements:

1. Profile amplitude distributions must be similar to the ones we collected during the user study.
2. Profile PSDs must also be similar to the ones we collected.
3. Profile values and derivates at start and end points must match so that it is periodic with minimized discontinuities.

Because of physical limitations, not every amplitude can be played with our platform. Since amplitude is mostly a growing function of the period (see Figure 3 to Figure 5), we decided to limit our amplitude distribution optimizations for periods shorter than 0.5 s.

An example of profile generation from statistics is given in [3]. However, this method assumes that the hand-held frequencies follow a Gaussian distribution, while our measurements suggest otherwise. Also, this generation adds some noise to the signal, which could probably add unrealistic frequencies to the final profile, and hence does not match requirement 2.

Generation algorithm

Each axis is considered to be independent, and therefore a one-axis synthetic profile is independently generated for yaw, pitch and roll following the same procedure.

The procedure to generate a one-axis synthetic profile of duration D is:

- For the considered axis, extract the median PSD over 2 s period from the user study data.
- Set all values above 15 Hz to 0 in the PSD.
- Generate a 2 s profile from this PSD by injecting random phases in the Fourier space and converting it in the real space via a fast Fourier transform (FFT).
- Extend the signal to the desired duration D by adding a constant value at the end of signal.
- Convert to the Fourier space using a fast Fourier transform.
- Add random phases.
- Convert back into the real space with the help of an inverse FFT.
- Adjust signal amplitude to minimize the amplitude distribution error.
- Optimize the amplitude distribution error by adjusting the coefficients of an added high degree polynomial function.

The polynomial function must verify P(0s)=P(D) and P'(0s)=P'(30s). This is to meet requirement 3. Also adding a polynomial function injects rather low frequency movement, preserving requirement 2.

A profile is considered to be a set of three one-axis profiles for all three rotation axes.

Profile file

A file named Handshake_profile_Type1_V1.csv is to be provided. This file contains 6 columns of values separated by commas. The first 3 columns are the translations positions, that are set to 0. The columns following are pitch, yaw and roll values in degrees sampled at 1000 Hz. See Figure 1 for the axis definitions. A graphical representation of this profile is given in Figure 9.

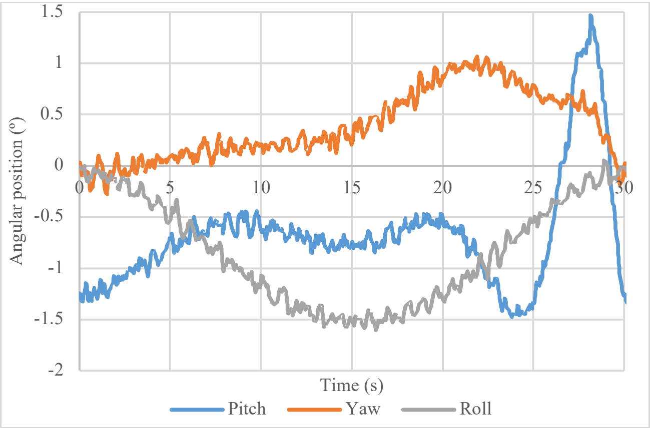

Figure 9:Proposed profile angular positions over time

The rotation center for our profile does not need to be precisely set, as long as it is not more than 10 cm away from the camera. Assuming the center is 10 cm away from the camera with our profile leads to a maximal acceleration of 0.03 g, which is negligible against gravity.

Motion platform specifications

A crucial aspect to accurately simulate hand-held movements is the quality of the motion platform. For that purpose, we analyzed the response of our platform to make sure it could play our profile without excessive distortion.

Our platform can perform rotation in the range of ±4° on each axis which exceed our profile maximal amplitude of 4° (see Figure 9).

We also checked the dynamic performance of our platform focusing on two aspects:

- The frequency response along an axis. This was done by measuring the actual displacement with the mobile phone mounted on the platform. Measurements were done with laser displacement sensors and results are shown in Figure 10.
- The purity of the actual movement when the platform is requested to perform a simple oscillation. Measurements were done with the IMU attached to the mobile phone with a simple 4 Hz 1° yaw oscillation and results are shown in Figure 11. We also added similar measurements done from a brushless motors motion platform for comparison. Also added is the noise spectrum from the IMU.

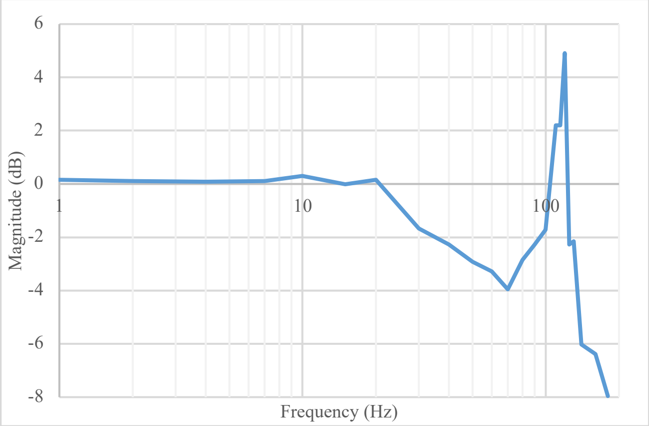

Figure 10:Bode magnitude plot for translations along camera lens axis

From Figure 10, we can see that our motion platform has a 0 dB up to 20 Hz. Since our profile was created with maximal frequencies of 15 Hz, those performances match our requirements.

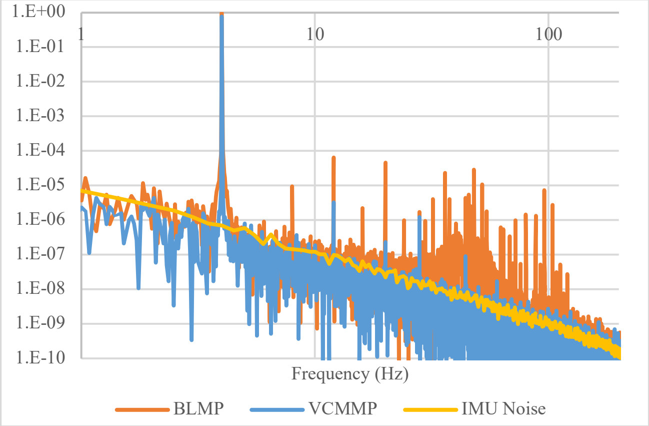

Figure 11:4 Hz 1° yaw oscillation relative power spectrums

From Figure 11, we can see that the voice coil motor motion platform provides much cleaner movements than the brushless motor one. Ideal the resulting spectrum from a motion platform should the noise one with a sharp peak at 4 Hz. Here:

- The peak around 4 Hz is narrower with the VCMMP than it is with the BLMP.
- The harmonics and additional frequencies contain much more energy on the BLMP.

The measurements presented here lead us to conclude that brushless motor motion platforms (BLMP) are not suitable for recreating handshake.

Profile validation

Although we generated our synthetic profile to match real user hand-held motions and evaluated our motion platform, we wanted to ensure that the overall process was able to match an actual user population. For that purpose, we created a reference set of images with a second user study and compared it to our profile. We also compared it to other movement profiles and hand shake creation platforms.

Protocol

The comparison was done using RAW images from a mobile phone. Although this device was equipped with an optical image stabilization mechanism, it was deactivated. Manual exposure was possible with this device which allowed us to test exposure times between 1/1000 s and 1/3 s. The native ISO sensitivity was set on the device and for each exposure time tested we adjusted the light level.

We captured images of a chart at 1m that allows MTF measurements in both vertical and horizontal directions. We considered the average value of the MTF50 at center as the sharpness metric. For each exposure time and experiment, we captured 10 tripod images and used the average MTF50 as a reference to compensate for any slight uncontrolled changes in light conditions. For each movement image (either captured by user or on a motion platform), we calculated the relative sharpness difference defined as follow:

$$Diff = \frac{MTF50_{mvt} - MTF50_{ref}}{MTF50_{ref}} \qquad (1)$$

We built the reference set using images captured by 25 people of various ages and genders. For each tested exposure time, the users were asked to hold the device and frame the chart. During that time the device would automatically capture 10 images. After those captures, the users would put the device on the tripod for the reference image captures. In addition, this delay would allow some rest time to the users and prevent any cramping associated with holding a static position for a long time. Users would then test the next exposure time.

The order of the tested exposure times was randomly generated for each user. That way no tested exposure time would be biased by the order.

We applied a similar process with our profile being played on the voice coil motion platform, except that 30 images captured for each exposure time. We also applied this process to a simple 4 Hz oscillation along the yaw and pitch axes played on the brushless motor motion platform.

Results

We first compared the average sharpness for each exposure time of each test.

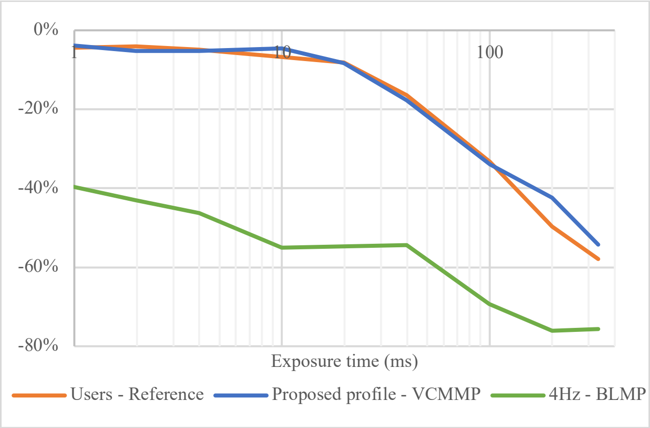

Figure 12:Average MTF50 relative difference

From Figure 12, it is clear that our profile played on a voice coil motor motion platform matches the user reference. Match is very good under 100 ms and stays good above. In comparison the oscillation leads to much blurrier captures. It can also be noted that measurements are plateauing when exposure time is similar to the oscillation period.

Since our profile was created to match not only the average user, but also the variabilities of the user, we also looked at the standard deviation of the relative sharpness difference. We also studied a reduced part of the user population by excluding the top 20% and bottom 20% images in terms of MTF50 relative difference for each exposure time. The average values for this reduced population are unchanged.

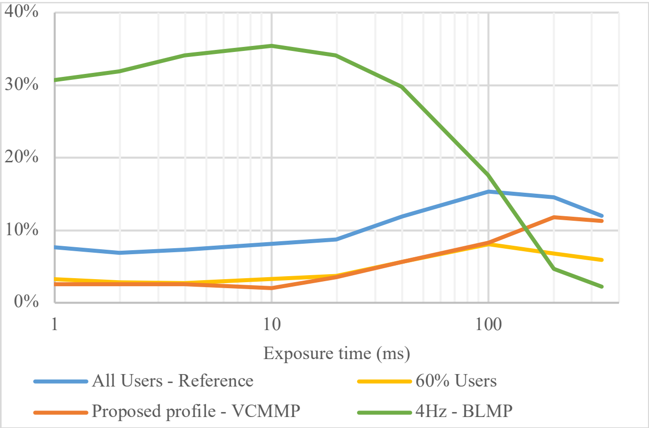

Figure 13: MTF50 relative difference standard deviation

Although our profile exhibits less variability than the real user population, it covers about 60% of the user population. In comparison, the 4 Hz oscillation exhibits a lot more variability except for long exposure times: there, the exposure time is about the same duration as the oscillation period, and there are less variabilities in the movements during this duration.

From this figure, one may estimate the 95% confidence interval of the average MTF50 relative difference with the formula:

$$CI = \pm 2 \times \sigma / \sqrt{N_{Captures}} \qquad (2)$$

Applying this formula with data from Figure 13, one may estimate the 95% confidence interval for the proposed profile in Figure 11 is between ±1% below 10 ms and up to ±4% at 1/3 s, based on 30 captures per exposure time.

Ensuring the accuracy of our metric is empirically sufficient is an important point with our proposed profile: Since it is by design non-stationary, using not enough captures could lead to very inaccurate results. When using this profile to evaluate a camera, one should ensure enough measurements are made to estimate performance with suitable confidence.

Discussion, known limitations and conclusion

In this paper we presented and shared a new trace to simulate human hand-held movement. This profile was generated by matching statistics from actual user hand held recordings in order to be representative of a large population. This profile was made cyclical so that it can be looped over several cycles. To validate this profile, we compared actual captures from a mobile phone mounted on a motion platform playing this profile with captures taken in similar conditions by actual users. Our work shows that our profile can accurately simulate a population of users when played on a proper motion platform.

Some limitations are however known:

- A key goal of this effort was to represent a population of users with a single trace. A consequence of this is that the statistics of the profile are not stationary. We do not expect this trace to be used with a single capture, but rather with multiple asynchronous captures spanning the period of the profile. Care must be taken that enough captures are made to faithfully represent the statistics of the population.
- This profile was validated for still captures with exposure times and capture periods up to 1/3 s. Future work is planned to test for broader use cases.
- Captures on longer than 1/3 s periods, whether those are still captures or videos, may not match real user experience.
- This profile was only validated using the referenced VCM motion platform. Many off-the-shelf motion platforms cannot faithfully reproduce either this profile or human handshake in the context of mobile phone camera testing.
- This profile is only aimed at matching users holding their mobile phone in a landscape orientation. Results might be different for portrait orientation, or with larger devices.

We think the profile being shared along with this publication can be a valuable tool to evaluate and benchmark stabilization systems that are designed to correct human handshake. Further work to broaden the utility of this profile is on-going and thus we plan to release future versions.

References

[1] L. Moisan, “Periodic Plus Smooth Image Decomposition”, Journal of Mathematical Imaging and Vision, vol. 39, issue 2, pp. 161-179, February 2011

[2] E. Mar Or, D. Pundik, “Hand Motion and Image Stabilization in Hand-held Devices”, IEEE Transactions on Consumer Electronics, vol. 53, no. 4, November 2007

[3] S-P. Lu, J. You, M. Wang, A. Munteanu, “Synthesis of Shaking Video Using Motion Capture Data and Dynamic 3D Scene Modeling”, in IEEE International Conference on Image Processing, Athens, Greece, 2018.

Author Biography

François-Xavier Bucher graduate from the École Centrale Paris (class of 2007). He joined the camera hardware group at Apple Inc. in 2015.

Jae Young Park received his BSc in photographic and electronic imaging sciences (2000), MSc in digital and photographic imaging (2006), and PhD in image science from University of Westminster (2014). In 2014, Jae joined camera hardware group at Apple Inc. in CA

Ari Partinen graduated from Helsinki University of Technology, and has since been working with many different aspects of digital imaging technologies and image quality.

Paul Hubel received a BSc in Optics from University of Rochester and a D.Phil. in Engineering Science from Oxford University. He has been working on camera technology at Apple since 2008.