

# Hand Motion and Image Stabilization in Hand-held Devices

Etay Mar Or and Dmitry Pundik

**Abstract** — *The mobile imaging market is a rapidly developing market, and has outgrown the traditional imaging market. This market is dominated by CMOS sensors, with pixels getting small and smaller. As pixel size is reduced, the sensitivity is lowered and must be compensated by longer exposure times. However, in the mobile market, this amount to increased motion blur. We characterize the hand motion with a typical shooting scenario. This data can be used to create an evaluation procedure for image stabilization solutions, and we indeed present one such procedure<sup>1</sup>.*

**Index Terms** — hand motion, image stabilization, mobile imaging.

## I. INTRODUCTION

The digital imaging market is characterized by two conflicting demands: smaller sensor form factors and increased spatial resolution. These trends are reflected in the shrinking of the overall pixel size. The size decrease is accompanied by a sensitivity decrease [6]. The resulting image has lower light sensitivity and reduced signal-to-noise ratio (SNR) [5]. These problems are even more severe in the mobile imaging market, where form-factor considerations are critical.

Offsetting the reduced sensitivity and SNR is easily achieved by increasing the exposure time. However, this option inevitably leads to increased motion blur in the image. The increased blur results in lower spatial resolution [4] as well as unacceptable image quality. Therefore, an image stabilization (IS) mechanism must be used. There are relatively few approaches in this field, the majority of which use moving parts (either lens [2] or the sensor [3]) to compensate for the motion.

This paper presents a study of hand motion, with a focus on implication to mobile imaging. Using these results, we describe the requirements from an IS mechanism aimed at solving the low sensitivity problem. Finally we propose a standard test to assess the quality of different IS mechanisms.

The paper is constructed as follows. In Sec. II we present the experimental setup. We then describe the data analysis method (Sec. III), followed by a detailed mathematical modeling of camera motion and their effect on the image (Sec. IV). We continue by presenting the results (Sec. V) and a discussion on the implication for IS systems (Sec. VI). We end by proposing an evaluation procedure for IS systems based on our results (Sec. VII).

<sup>1</sup> E. Mar Or and D. Pundik are with Advasense Technologies (2004) Ltd., Ra'anana, 43663, Israel (e-mail: {itay.mar-or, dima.pundik}@advasense.com).

## II. EXPERIMENTAL SETUP

We measured hand motion as a function of time, using small board with a CMOS sensor. The board (shown in Fig. 1) form factor was chosen to resemble that of a camera-phone. It is approximately 5cm by 9cm and can be held like a camera-phone. During each measurement, the sensor recorded VGA video at 96 fps. The motion was later extracted from the video sequence.

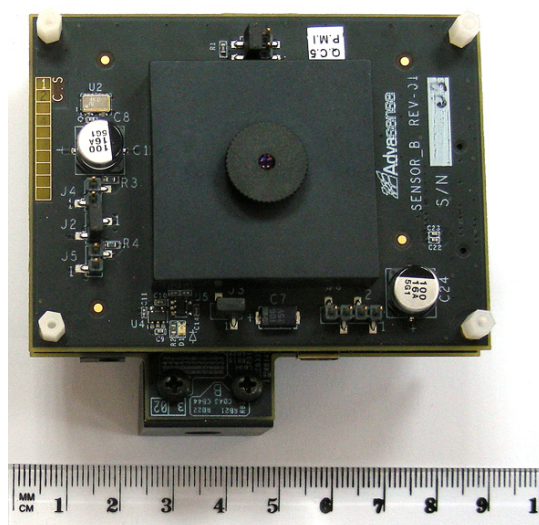


Fig. 1. The sensor board used in the experiments. The ruler on the bottom is in centimeters.

Each subject was asked to hold the board as steady as possible for 2 seconds, after pointing the sensor at a standard resolution chart (using a nearby computer screen). The illumination conditions were regular indoor conditions (around 300 Lux).

25 unpaid subjects (20 males and 5 females, ages from 25 to 50 years old) participated, each making 3 different shots. The total number of different measurements (motions) was 85.

## III. DATA ANALYSIS

The motion data was extracted from the video sequences using the Lucas-Kanade ([7]) algorithm with sub-pixel accuracy. The registration was limited to translation and rotation in the *image plane* (rigid motion), and not a full affine transform. The paths were then translated from pixels to degrees.

The rigid motion assumption was validated by comparing the registration results with measurements taken from a 3-axis gyroscope (carefully mounted in parallel to the sensor board).

The angular motions recorded by the gyroscope were small enough to justify the assumption. We will elaborate on this in the next section.

#### IV. CAMERA MOTION AND INDUCED IMAGE PLANE TRANSFORMATION

It is instructive to follow an exact modeling of effect of different camera motions on the image itself. This allows estimating the significance of each type of motion and its contribution to the image motion blur.

We proceed to describe the camera and image in the affine space. The image plane is assumed to be parallel to the XY plane, lying at  $z = -F$ . The image dimensions are  $(W, H) \equiv (2F \tan(\alpha_x / 2), 2F \tan(\alpha_y / 2))$ , where  $\alpha_x$  and  $\alpha_y$  are the horizontal and vertical angles of view, respectively.

When the optical axis is coincident with the z-axis, the camera is described simply by the following projection matrix ([8]), in homogeneous coordinates:

$$\Pi = \begin{pmatrix} -F & 0 & 0 & 0 \\ 0 & -F & 0 & 0 \\ 0 & 0 & -F & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad (1)$$

$$x' = \Pi x \mapsto (-Fx/z, -Fy/z, -F, 1)^T \quad (2)$$

where  $x = (x, y, z, 1)^T$  are the world homogeneous coordinates of an object and  $x'$  are its homogeneous coordinates in the image plane.

When the optical axis is not coincident with the z-axis, we proceed as follows. Assume the optical axis can be transformed to the z-axis by applying a transformation matrix  $T$ . We apply the inverse transform  $T^{-1}$  on the world coordinates before applying the projection:

$$x' = \Pi T^{-1} x \quad (3)$$

For a rotation around the y-axis (yaw), we have: and the overall projection is:

$$T_{yaw} = \begin{pmatrix} \cos \psi & 0 & -\sin \psi & 0(7) \\ 0 & 1 & 0 & 0(8) \\ \sin \psi & 0 & \cos \psi & 0(9) \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (5)$$

$$\begin{aligned} x'' &= \Pi T_{yaw}^{-1} x \\ &= (-F(x \cos \psi - z \sin \psi), -Fy, \\ &\quad -F(x \sin \psi + z \cos \psi), x \sin \psi + z \cos \psi)^T \end{aligned} \quad (6)$$

Assuming  $\psi \cong 0$  and  $z \gg x, y$  we arrive at (after converting to homogeneous coordinates):

$$x'' \mapsto \left( -F \left( \frac{x}{z} - \psi \right), -F \frac{y}{z}, -F, 1 \right)^T \quad (7)$$

The difference between  $x'$  and  $x''$  is:

$$x'' - x' = (F\psi, 0, 0, 0) \quad (8)$$

Thus, we see that the induced transformation in image plane is translation. The same is true for rotation around the x-axis (pitch). It can easily be shown that a rotation around z-axis (roll) induces a *rotation* in the image plane.

Contrary to intuition, the camera translation (as long as it is small) has little effect on the image. Let's consider the following camera translation:

$$T = \begin{pmatrix} 1 & 0 & 0 & a \\ 0 & 1 & 0 & b \\ 0 & 0 & 1 & c \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (9)$$

$$\begin{aligned} x'' &= \Pi T^{-1} x \\ &= -F(x-a), -F(y-b), -F(z-c), z-c)^T \\ &\mapsto \left( -F(x-a)/(z-c), \right. \\ &\quad \left. -F(y-b)/(z-c), -F, 1 \right)^T \end{aligned} \quad (10)$$

Now, assuming a far enough object ( $z$  being largest than any other length in the system; typical scenario for fixed focus camera-phone) we arrive at:

$$\begin{aligned} x'' &= \left( -F \left( x/z - a/z + O(xc/z^2) \right), \right. \\ &\quad \left. -F \left( y/z - b/z + O(yb/z^2) \right), -F, 1 \right)^T \end{aligned} \quad (11)$$

The difference between  $x'$  and  $x''$  in this scenario is:

$$\begin{aligned} x'' - x' &= -F \left( a/z + O(xc/z^2), \right. \\ &\quad \left. b/y + O(yb/z^2), 0, 0 \right) \end{aligned} \quad (12)$$

Therefore, unless the translations are in the order of  $z$  they are insignificant. Consider a regular shooting scenario, the photographed scene lies a few meters away from camera. In order for this effect to be noticed, the camera motions must be at least tens of centimeters in length. An average user can avoid such movements.

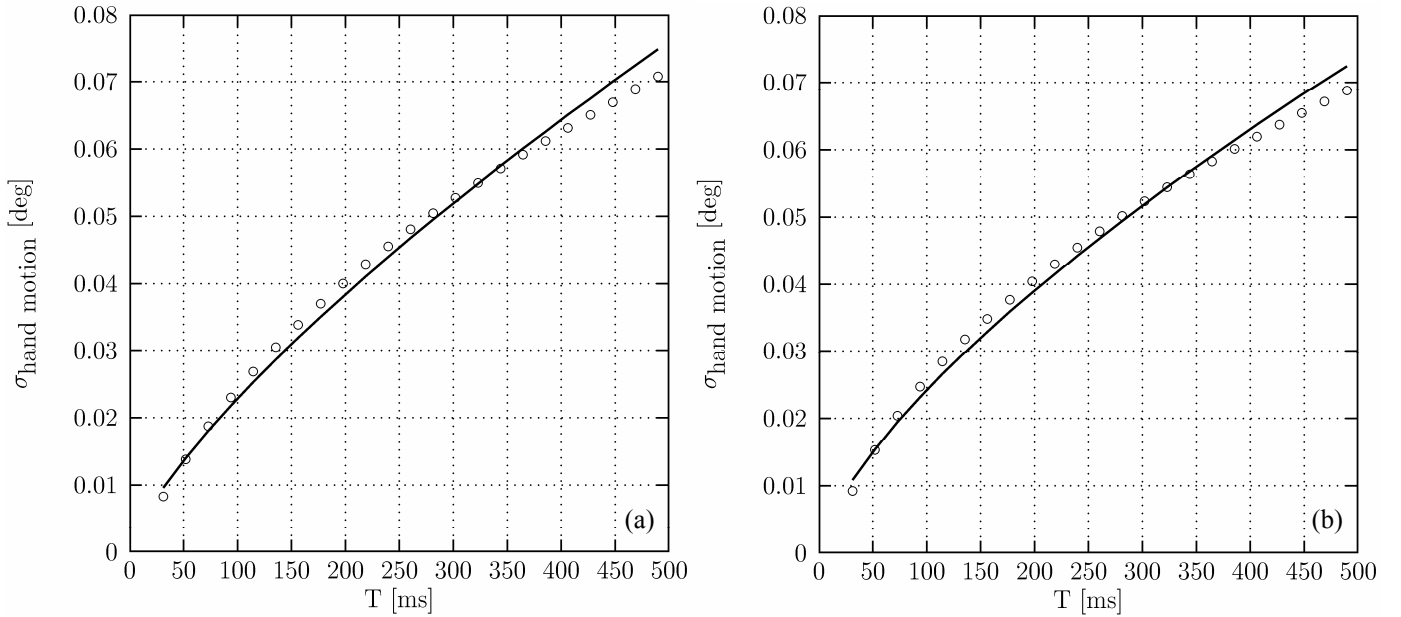


Fig. 2. Standard deviation of motion as a function of exposure time for (a) x-axis and (b) y-axis. The solid lines are the exponential fits.

In order to translate the translation in image plane from pixels to angles, we use the following conversion factor:

$$\psi = \frac{1}{F} \frac{\Delta_x}{N_x} W = 2 \frac{\Delta_x}{N_x} \tan(\alpha_x / 2) \quad (13)$$

$$\theta = \frac{1}{F} \frac{\Delta_y}{N_y} H = 2 \frac{\Delta_y}{N_y} \tan(\alpha_y / 2) \quad (14)$$

where  $\Delta_x$  and  $\Delta_y$  are the translation in pixels in x and y axes, respectively;  $N_x$  and  $N_y$  are the number of pixels in x and y axes.

## V. RESULTS

We first noticed that during one second of motion, the amount of rotation was negligible. The maximal rotation during a  $1/96$  second period is  $0.12^\circ$ , for a full 1 second it is  $0.88^\circ$ . In the rest of the analysis we ignore the rotation, as its effect is considerably smaller than that of the translation.

The results are summarized in Figs. 2 and 3. In Fig. 2, we present the mean standard deviation of the motion path as a function of time. The mean is over the 85 motions. The standard deviation is a good estimate for motion blur in most cases. Fitting the results to a power law:

$$\sigma_{\text{handmotion}}(t) = At^a \quad (15)$$

yields for x-axis  $a = 0.75$  and for y-axis  $a = 0.69$ . This result is close, yet higher, than that of a random walk. This is expected as the motions are not fully random, and contain some inertial part. At this point, we cannot explain the difference between the two axes. It is most likely the result of the USB cable (standard cable with 0.5 cm

diameter) that connected the board to the PC. Another factor could be the way the camera is held. Most subjects gripped the camera board by encircling the base of the board using one hand.

Fig. 3 presents the mean spectrum of the hand motions. This plot demonstrates that most of the energy of the motion lies in low frequencies (99% lies below 10Hz). This result is in agreement with previous studies of involuntary hand motion measuring hand motions in eye surgeons ([1]).

It should be noted that due to the sensor's electronic rolling shutter (rather than the traditional mechanical shutter), our measurements are not completely accurate. The effect of the rolling shutter is that of a low pass filter, that of a rectangular function of width  $T_{\text{shutter}}$  in the time domain. In frequency domain this transforms to  $\text{sinc}(\omega \cdot T_{\text{shutter}})$ . At low frequencies ( $< 50$  Hz), the effect is negligible.

## VI. IMPLICATION FOR IS SYSTEMS

A well-known rule-of-thumb among photographers states that the exposure should not exceed the inverse of the focal length (given in millimeters for a 35mm camera). It has been shown that motion blur is affected (among other factors) by camera mass ([4]), and so the situation in camera-phones and other hand-held devices might be different. However, it is still interesting to quantify this rule.

We can use Fig. 5, to translate this common wisdom into a quantitative criterion. For  $F = 35\text{mm}$  (angle of view is  $\alpha_x = 54.4^\circ$ ;  $\alpha_y = 37.8^\circ$ ), the exposure time is  $t_{\text{max}} = 1/35 \text{ sec}$ . From the figure, we see that this amount to a standard deviation of approximately  $0.009^\circ$  per axis.

As mentioned above, this rule-of-thumb was set for film photography. This number for a typical consumer digital

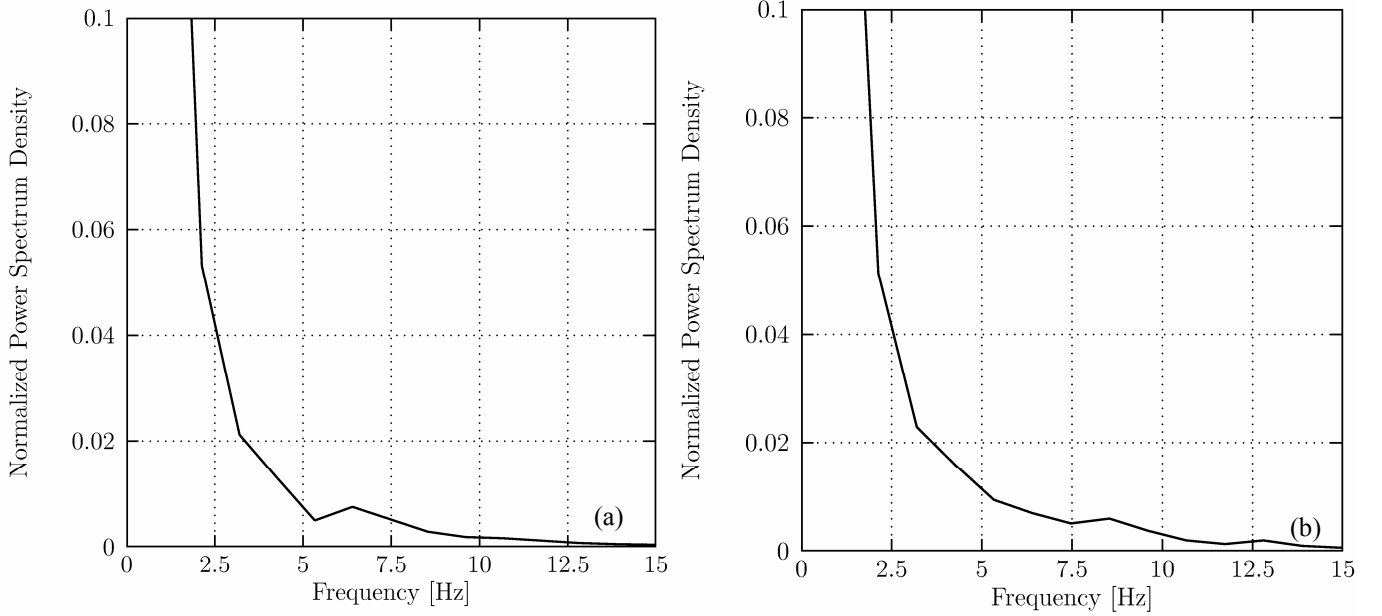


Fig. 3. Power spectrum of hand motion for (a) x-axis and (b) y-axis.

camera (5 mega-pixels camera with 2592x1944 resolution) translates into a total standard deviation of 0.7 pixels. It seems that this value is too severe, especially in view of the demosaicing process which introduces (at minimum) a blur of 1 pixel in the image data.

It is obvious that a maximal acceptable blur must be defined in relation to the sensor resolution. The current trend is to increase the sensor resolution while maintaining the same field of view (by decreasing the form factor and adjusting the lens accordingly). This means that the *same* motion translates to a *larger* blur in systems with identical field of view.

## VII. IS EVALUATION

IS system are typically evaluated by performing two manual shots of the same scene — with IS enabled and with IS disabled. There are several drawbacks to such a method:

- *Motion Repeatability* The motion during the two shots is different and may differ greatly. Thus, the comparison between the two images might be misleading.
- *Limited Meaningfulness* The motion affecting the two shots is unknown. Therefore, we cannot judge if the test conditions represent a typical case.
- *Test Reproducibility* The motions are very subject dependant, and it is hard to reproduce the test elsewhere.

In order to address these issues we propose a *standard, meaningful* and *reproducible* benchmark for IS systems.

The benchmark system consists of:

- Motion reproduction system (“Mechanical Hand”), capable of reconstructing a motion with band-limited frequencies of up to 10Hz.
- Standard scene to be photographed: test chart (ISO-12233).

- A standard corpus of motions with average spectrum similar to Fig. 3. The variations between the motions should also be representative of the expected variations in the population (i.e. similar to those in the figure).

The benchmark itself is performed as follows:

- For each motion, capture images at increasing exposure times (from 1/250 to 1 second). If the IS is sensitive to the illumination conditions, repeat this at appropriate conditions.
- For each image, perform the following, measure the modulation transfer function (MTF) using horizontal edges and vertical edges.
- Compute  $k_0$  s.t.:

$$MTF(k_0) = e^{-1/2} MTF(k=0) \quad (16)$$

- $1/k_0$  is used as a measure of the image blur caused by misalignments of the IS system.<sup>2</sup>
- To obtain the benchmark, we average for a given exposure time  $t$ :

$$\sigma_{\text{benchmark}(t)} = \frac{1}{N_{\text{motions}}} \sum_{\text{all motions}} 1/k_0 \quad (17)$$

where  $N_{\text{motions}}$  is the number of motions.

The resulting plot,  $\sigma_{\text{benchmark}}(t)$ , can be used as a detailed benchmark for the IS system. By comparing it

<sup>2</sup> For an ideal imager (point spread function being a  $\delta$ -function) with Gaussian distributed misalignments  $1/k_0$  is equal to the standard deviation of the misalignments. For a non ideal imager,  $(1/k_0)^2 = \sigma_{\text{IS}}^2 + \sigma_{\text{PSF}}^2$ , where  $\sigma_{\text{IS}}$  originates from the IS misalignments and  $\sigma_{\text{PSF}}$  originates from the imager non-ideal PSF.

to  $\sigma_{\text{handmotion}}(t)$ , one can measure the reduction in blur due to the IS. One could also extract from the plot the maximal exposure time ( $\tau$ ) that yields a 2 pixel blur:

$$\tilde{\sigma}(\tau) \leq \sqrt{2} \text{ [pixels]} \quad (18)$$

The  $\sqrt{2}$  pixel blur is twice the blur due to demosaicing.  $\tau$  can be used as a scalar benchmark number for the IS system.

### VIII. CONCLUSIONS

We have presented a detailed time-domain measurement of hand motion in conditions resembling camera-phone photography. The results indicate the major source for the blur is yaw and pitch in the camera plane, leading to translation in the image plane. Motions in other axis were negligible.

We were able to use these results to quantify the well-known “one over F” rule-of-thumb from film photography. The results proved to be too severe for typical consumer cameras. It seems that this rule-of-thumb should be modified to fit digital photography.

Using these measurements we were able to define a standard, reproducible and meaningful benchmark for IS systems. This benchmark can also be used to estimate the maximal exposure time for a given IS system.

The future generation of camera-phone, assuming the current trend of reduced pixel size continues, will inevitably require IS systems. Currently, there are no metrics for comparing such systems. We hope that our work may expose the requirements from IS systems, as well as help create a metric to evaluate them.

### ACKNOWLEDGMENT

The authors would like to thank Ron Soferman, for his help in the initial stages of the research, and Haim Dassa for his help in acquiring the experimental data.

### REFERENCES

- [1] C. N. Riviere, R. S. Rader, P. K. Khosla, “Characteristics of Hand Motion of Eye Surgeons” in *Proc. 19th Annu. Int. Conf. of the IEEE*, 1997, vol. 4, pp 1690-1693.
- [2] T. Otani, K. Washisu, “Image Stabilizing System,” U.S. Patent 5774266, June 30, 1998.
- [3] T. Hirota, Y. Tanaka, “Camera with shake correction mechanism,” U.S. Patent Application num. 20060056829, March 16, 2006.
- [4] F. Xiao, A. Silverstein, J. Farrell, “Camera-Motion and Effective Spatial Resolution” in *Proc. Int. Congr. of Imaging Science*, 2006, pp. 33-36.
- [5] F. Xiao, J. Farrell and B. Wandell, “Psychophysical Thersholds and Digital Camera Sensitivity: The Thousand Photon Limit,” *Proc. SPIE*, Vol. 5678, (2005).
- [6] J. Farrell, F. Xiao and S. Kavusi, “Resolution and Light Sensitivity Tradeoff with Pixel Size” in *Proc. SPIE Electronic Imaging Conf.*, 2006, vol. 6069.
- [7] B. D. Lucas, T. Kanade, “An Iterative Image Registration Technique with an Application to Stereo Vision” in *Proc. 7th Int. Joint Conf. on Artificial Intelligence*, 1981, pp. 674-679.
- [8] R. Hartley, A. Zisserman, *Multiple View Geometry in Compute Vision*, 2nd ed. Cambridge Press, 2003

**Etay Mar Or** was born in 1978 in Nahariya, Israel. He received his B.S. degree in Computer Science from Hebrew University in Jerusalem in 2002 and M.S. degree in Physics from Tel Aviv University in 2005. He has been working as an algorithm developer for CMOS image sensors since 2005, at Advasense Technologies (2004) Ltd.

**Dima Pundik** was born in 1980 in Charkov, Ukraine. He received his B.S. degree in Computer Engineering from Technion Israeli Institute of Technology in 2003. He is currently working toward M.S. degree in Computer Science from the Interdisciplinary Center Herzeliya. He has been working as an algorithm developer for CMOS image sensors since 2005, at Advasense Image Sensors since 2005.