Discrete Kalman Filter based Sensor Fusion for Robust Accessibility Interfaces

I Ghersi^{1,2,3}, M Mariño², M T Miralles^{1,2}

¹ Pontificia Universidad Católica Argentina (LaBIS-Facultad de Ciencias Fisicomatemáticas e Ingeniería-UCA), Av. A. Moreau de Justo 1500, CABA, Argentina. ² Universidad de Buenos Aires (CIDI-Facultad de Arquitectura Diseño y Urbanismo-UBA), Intendente Guiraldes 2160, Pabellón III, Cuidad Universitaria, CABA, Argentina.

e-mail: igghersi@gmail.com

Abstract. Human-machine interfaces have evolved, benefiting from the growing access to devices with superior, embedded signal-processing capabilities, as well as through new sensors that allow the estimation of movements and gestures, resulting in increasingly intuitive interfaces. In this context, sensor fusion for the estimation of the spatial orientation of body segments allows to achieve more robust solutions, overcoming specific disadvantages derived from the use of isolated sensors, such as the sensitivity of magnetic-field sensors to external influences, when used in uncontrolled environments. In this work, a method for the combination of image-processing data and angular-velocity registers from a 3D MEMS gyroscope, through a Discrete-time Kalman Filter, is proposed and deployed as an alternate user interface for mobile devices, in which an on-screen pointer is controlled with head movements. Results concerning general performance of the method are presented, as well as a comparative analysis, under a dedicated test application, with results from a previous version of this system, in which the relative-orientation information was acquired directly from MEMS sensors (3D magnetometeraccelerometer). These results show an improved response for this new version of the pointer, both in terms of precision and response time, while keeping many of the benefits that were highlighted for its predecessor, giving place to a complementary method for signal acquisition that can be used as an alternative-input device, as well as for accessibility solutions.

1. Introduction

Recent technological advances have converged towards a wave of massively-available consumer devices displaying highly intuitive user interfaces, based on alternate control means, such as face detection, motion and spatial-orientation measurements through 3D MEMS (Micro ElectroMechanical System) sensors [1][2], etc. Within this context, these devices have the potential to be adapted for the provision of advanced accessibility functions, forming customizable environmental control units, as well as to provide new input-means for the control of dedicated medical devices [3], through the integration of these advanced, minimally invasive mechanisms, in the detection of user commands. Aside from their potential as isolated inputs for human-machine interfaces, the group of sensors that are of interest for the present work (MEMS accelerometers, magnetometers and gyroscopes, and cameras) reach their highest impact when combined in real time. Sensor fusion, achieved through algorithms of variable

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution • (cc) of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI. Published under licence by IOP Publishing Ltd 1

³ Corrsponding author

complexity, allows to combine a number of estimations over a variable from isolated, supplementary sensors, in such a way that a joint estimation, more reliable than those obtained by the individual sensors, is obtained. As an example, the detection of the attitude or spatial orientation of body segments requires the combination of information from a 3D-accelerometer with a 3D-magnetometer, since the first is not capable of detecting changes in orientation around the vertical axis. Meanwhile, magnetometers are susceptible to interferences and deviations in their measurements, depending on the environment in which they are implemented [4], resulting in the need to take cautionary measures for their isolated deployment in highly-sensitive systems.

The development of a pointer-type interface, intended for the control of mobile-device applications, and based on face-detection and MEMS orientation sensors, has been presented in [5]. This solution allows to control an on-screen pointer with head movements. However, the potential susceptibility for the response of this system under variable or uncontrolled environments, as well as the specific interest in implementing it as a proposal for accessibility solutions, has fueled the search for alternate means to supplement the functioning of this system, when the reliability of one of its components is significantly degraded.

The present work shows results of the development of an alternate version of this pointer-type interface, taking advantage of and combining signals from different sensors (camera and MEMS gyroscope), with a Discrete Kalman Filter, in order to obtain a new solution, with a low relative computational cost, that displays an improved performance, as well as a higher independence to outer interferences, while keeping many advantages found in the preliminary version of this interface. This new solution can be implemented as a supplement to the first version of this system, and forms a new alternative-input and accessibility controller, valuable on its own for the command of devices and applications.

2. Materials y Methods

2.1. Materials

The aim of the proposed systems is to sum advantageous features from mobile devices (part of an expanding, increasingly available market), with minimally-invasive external peripherals, in order to enhance their joint interfacing capabilities between user and device. Prototypes and test stages presented in this work were developed on a 7.0 inch tablet, with a dual-core 1 GHz processor, Android 4.1.1 OS, and a 1024x600 pixels screen. Figure 1 shows the architecture of the system: external sensors (3D magnetometer-accelerometer, for the initial version of the system, and 3D gyroscope for the new version presented in this work), were interfaced through a dedicated circuit board, with a 16-bit microcontroller (dsPIC33F, from Microchip) in its core. Information from this board was transmitted to the tablet device through a wireless TCP link, through a PC, for initial test stages of both systems, while a dedicated USB link has also been achieved between the tablet and the external sensors. Both versions of the system take advantage of face-detection functionalities, through the frontal camera of the tablet device, and the native Android library for such purposes [6]. Face-detection is performed over low-resolution frames (160x120 pixels).

2.2. Initial Version: Mouse/pointer interface controlled by head turns

The first pointer interface solution, deployed on this platform, featured the integration between registers from a 3D MEMS orientation sensor (magnetometer-accelerometer), and information provided by the open-source face-detection library. While this MEMS module is subjected to limitations as specified in section 1, its development made it possible to explore the required strategies for the integration of all data sources (camera, orientation sensors in the tablet and external orientation sensor), while serving as a reference for future controlled tests with other versions of the system. Also detailed in figure 1, the proposed strategy for the combination of measurements and estimations between these input-devices, is described as follows:

- The face detector acquires information concerning relative position between the screen and the user's face. The position of the face in the image that is obtained from the frontal camera is indicative of its relative position in two dimensions (x: horizontal, y: vertical), while the size of the face in the image is indicative of the distance between user and screen (z axis).

- The external orientation sensor, fixed to detect head movements (i.e. integrated with a headphone), is useful in the estimation of relative orientation between user and screen, around three rotation axes (referred to as Heading or Yaw, for horizontal head turns, θx , Pitch for vertical head turns, θy , and Roll around the remaining axis). In order to acquire this information, the system obtains the difference between the external sensor (user) and a similar module, fixed to the interface screen (mobile devices usually feature these useful orientation sensors).

In this way, the proposed system is capable of obtaining relative position and orientation information between user and device, through an approximation of the perspective from which the user interacts with it. It has been stated, in [5], that an interface that responds to this information is capable of giving place to a more intuitive and robust interactivity, resulting in an on-screen pointer that remains effective while the user explores the environment surrounding the device, always responding to its movements: by turning its head, the user points to an absolute position on the screen. The coordinates of the pointer in the screen is a result of the two-axis projection of the vector that represents the spatial orientation of the user's head, while taking into consideration relative distance and positions between user and device (see figure 1 and [5] for a detailed description of the algorithm).



Figure 1. System components and pointing-interface strategy [5].

2.3. Updated Version: Camera-Gyroscope Sensor Fusion based on Discrete Kalman Filter

The face detector used for the initial pointer interface acquires information about the *relative position* between user and screen. An alternative interpretation of this information, however, would allow *detecting head turns from an initial position*, since, when turning the head in any direction, its center is displaced over the image acquired by the camera (figure 2). If there are not significant displacements during this interaction, and movements are limited to free head turns, which is considered acceptable for the target population of this work (accessibility interfaces), the face detector can be alternatively used as an input for orientation and distance estimations, without the need to use algorithms with a higher computational cost for such purpose.

These new relative-orientation estimations, acquired from the output of the face detector in both axes, can be realized in many ways. As an example, a basic approximation supposes that the sine of the angle between user and screen (see figure 1), is related to the center of the face in the image, with respect to its origin coordinates (PX-X0, for the x-axis), and the estimated distance between the user and the screen, after adequate projections and unit-conversions between pixels and centimeters. Equations (1) and (2) express these estimations, while it should be stated that, for the expected range of motion

(2)

required for the control of this device in both axes, which falls below $\pm 15^{\circ}$ (7.0-inch screen, at 30 cm to 50 cm distances), other variants, including linear approximations between displacements and angles could be used:

$$\theta x_{FD} = \sin^{-1}((PX_{cm} - X0_{cm})/(Dist_{cm} + 20))$$
(1)

$$\theta y_{FD} = \sin^{-1}((PY_{cm} - Y0_{cm})/(Dist_{cm} + 20))$$



Figure 2. Angle-estimation strategy for horizontal head turns (θx) based on face detection. From an initial position (center, X0), heading angles are approximated from the relative displacements to this point. Image adapted from [7].

Based on the calibration of device-specific parameters (screen resolution, relative position and camera field of view), the interpupillary distance, in pixels, estimated by the face detector, is used to estimate the distance between user and device, in cm (Dist_{cm}) [8], through a linear estimation [5]. The constant that is added to this variable in equations (1) and (2) (20 cm), represents the distance between the center of the face and its rotation axis. A similar anthropometric variable (distance of the eyes to the back of the head) varies between 14 cm (5th female percentile) and 19 cm (95th male percentile), for adults [9], so the selected value represents a potential maximum among users.

Besides dealing with the issue of being derived from low-resolution images (160x120 pixels), and suffering potential noise, resulting from the image-processing algorithm, this proposal for angle estimations suffers non-linearity, associated to relative displacements between the center of the camera and the screen. However, it has been proposed as a simple estimation, because it is stable, while noisy, and supposes a minimum added computational cost, and its limited response is expected to show improvements when integrated with other sensors.

The sensor-fusion proposal involves the 3D digital gyroscope L3G4200D, with 16-bit outputs, integrated programmable high-pass filter, and used in the range of ±250 °/s, this new MEMS sensor taking the place of the orientation sensor in the previous version (section 2.2). Under this strategy, the camera is capable of providing head-turn estimations, $(\theta x_{FD}, \theta y_{FD})$, while these are limited and do not contemplate variable perspectives. In order to do this, the starting point of the pointing device is calibrated during the initialization of the system. Coming from low-resolution images, face-detection results can be noisy, but represent a stable reference on the user's relative orientation. On the other hand, the gyroscope gathers angular-velocity measurements from these head turns ($\omega = d\theta/dt = \dot{\theta}$), supplementing the noisy information from the camera. However, the direct estimation of angles from this sensor would result in slow drifts in the estimated signals, as a result of noise or zero-level inaccuracies during signal processing stages. As a result, these signals appear to be supplementary, and their fusion is expected to result in a positive joint response.

Sensor fusion between these information sources has been implemented through a Discrete Kalman Filter [10] on each axis, an algorithm that allows estimating variables or states, based on measurements from multiple sources, when these are combined through a known model for the behavior of the system and its estimated uncertainties. A large number of viable models for MEMS sensors have been implemented for research purposes, in the search for more precise position and orientation estimations in the field of navigation systems [11]. In this case, the proposed system model is a simple relationship between angular velocity and heading or pitch angle, depending on the axis, in discrete time. Equations (3) and (4) show the Discrete Kalman Filter model:

$$\theta_{k+1} = \theta_k + \dot{\theta}_k \Delta t + w_k \tag{3}$$

$$z_k = \theta_k + v_k \tag{4}$$

considering:	θ_k : Angle, estimated process state at instant k	
	$\dot{\theta}_k$: Angular Velocity, as measured by the gyroscope	
	Δt : Time interval between measurements, in seconds	
	z_k : Angle measurement, as obtained from the face detector (θx_{FD} , θy_{FD})	
	v_k : Measurement noise w_k : Process noise	

Based on this model, the recursive algorithm of the filter [10], performed over the corresponding signals in each axis, allows updating and projecting head-angle estimations. These estimated angles in each axis (θ_x , θ_y) are later converted to on-screen coordinates in pixels, through an inverse operation from equations (1),(2). During this step, it is possible to amplify or attenuate the sensitivity to these head turns, which, when allowed by the resolution of the system, could be found valuable for accessibility implementations, over users that display a lesser degree of mobility in the neck. According to the specifications of the system, it is possible to tune certain filter parameters (the noise estimations for the model and measurements, Q and R, which affect w_k and v_k), in order to prioritize the gyroscope or the camera, when the other's reliability is compromised.

For this case, static and non-linearity errors in the estimations of angle from the face-detector have been accepted, because this source complies with the need to rely on a pointer that is stable in the long term, which the gyroscope alone cannot achieve. In this point, it should be noted that, for a user interface such as the one described, the noise that could affect the position of the on-screen pointer is far more detrimental to the experience in its use than the actual accuracy of the orientation estimations, which is why the contribution of the gyroscope is critical in this sense. In case of counting with sufficient motor control, the user could detect and compensate for small inaccuracies in the pointer coordinates, as long as these are not significant, and the pointer can be relied with enough stability in response to its movements. However, there would be no possibility for the user to compensate for a highly changing position, even if its mean value was actually an accurate measurement of the relative angle. Table 1 shows a comparison between properties and algorithms involved in both system versions:

Version Parameter	Initial (section 2.2)	Kalman (section 2.3)	
Sensors	3D Mangetometer-Acelerometer Camera (160x120 pixels)	2D Gyroscope Camera (160x120 pixels)	
Algorithms	FaceDetect (native) 3D-orientation estimation [4] Discrete Signal Filtering User-screen distance estimation [5] Pointer position estimation [5]	FaceDetect (native) User-screen distance estimation [5] User-screen angle estimation Sensor Fusion (Kalman)	
Update Rate	FaceDetect Maximum > MEMS Sensor On-screen Pointer I	10 Hz, Average 7.46 Hz s \geq 100 Hz Position ~ 16 Hz	
Perspective Detection	Distance and relative-positon estimation, between user and screen	Relative distance estimation, between user and screen	

2.4. Test application and analysis strategy

In order to assess the properties and characteristics of both versions of the system, a test application for Android OS was developed with Eclipse IDE. This application shows five options on screen, representing five buttons, and a central, yellow number, which changes randomly, indicating which option the user should point to with the on-screen pointer (crosshair). This application gathers information concerning the requested option, pointer coordinates (x_p : horizontal, y_p : vertical), time stamps, raw sensor data, and an indicator of the detection or absence of a face, as detected by the face-

detector. Figure 3 shows a screenshot of this test application (right), with an xy chart showing results of its use (left), with the corresponding trajectory of the on-screen pointer, concentrated over each option.



Figure 3. Test application screenshot (right), with x_p , y_p (on-screen pointer coordinates) plot (left). Pointer positions are concentrated around each option.

From this application, information about pointer speed and reliability can be obtained, which in turn allows comparing both system versions, estimating their respective performance and potential. The upper side of figure 4 is a time chart displaying the same xy pointer coordinates, through the use of the interface. The bottom of the figure shows another time chart, with the progression of requested options during its use. Based on these signals, pointer stability (range and standard deviation), of special interest for this work, is obtained after changes in the requested option, once the transition begins to stabilize (yellow rectangle, compatible with the yellow circle in figure 3), represented by the definitive entrance of the pointer into a broad proximity of the option. The duration of this stable period is, on the other hand, indicative of the possibility of implementing a selection strategy for the interface, based on holding the pointer over an option for a certain period of time.



Figure 4. Test application data from initial version. Top: screen pointer position, in pixels, over time (x_p : horizontal, y_p : vertical). Bottom: requested option (1-5). The region of interest for stability and reliability assessments is highlighted (yellow).

3. Results

3.1. 1D response: Horizontal Head Turns, sensor fusion behavior and general performance

Figures 5, 6 and 7 display three cases of the estimation of horizontal head turn angles, based on the gyroscope, the face detector, and through the Discrete Kalman Filter, as well as through measurements provided by the orientation sensor of the initial version of the system, as a reference.



Figure 5. Horizontal head-turn angle estimations (θx), based on face detection (purple), angular velocity from the gyroscope (red), and Kalman Filter (blue), together with a reference, obtained from the 3D MEMS magnetometer-accelerometer (green): the sensor fusion signal results in an improved version of the face-detection estimation.



Figure 6. Discrete Kalman Filter response: θ x angle estimations from the gyroscope alone would result in variable drifts (red). Additionally, when the face detector does not provide an update (forced condition between samples 2761 and 3497, shaded in the figure), the output of the filter follows the gyroscope. In sum, the overall solution is more robust than its individual components.



Figure 7. Discrete Kalman Filter – Horizontal head-turn estimation.

As a result of these implementations, the following considerations are highlighted:

• The signal that is derived from the Kalman filter, on the horizontal axis, is less noisy than the one provided by the face-detection algorithm (figure 5), as well as the original reference.

- This new integrated estimation compensates for the drift derived from the gyroscope (figure 6)
- In terms of noise and drift, the on-screen angular response, derived from the Kalman filter, proves satisfactory.
- It is necessary to consider that, unlike the angular velocity signal, which is periodically updated, the output of the face detector can provide no updates on its information over an interval, given possible changes in the image acquired by the camera during the control of the pointer. The nature of the parameters from the Kalmn filter, however, made it possible to propose a mechanism to mitigate this possible lack of information, through the real-time change of filter parameters Q and R. Figure 6 exemplifies how, when no face is detected, the filter adapts to prioritize information from the gyroscope (this is achieved by changer filter parameter R). This condition is visible in the shaded section of figure 6.

While it does not mitigate every imperfection from its component signals, the implemented Kalman filter results in an improved version, with a fast response, of the signal generated from the face detector (figure 7), based on information from a completely different sensor than the ones implemented on the initial version of the system (gyroscope).

3.2. 2D-response: Test Application and Comparative Performance

Figure 8 shows results of the implementation of the discrete Kalman Filter, in estimating the on-screen pointer position in two axes.



Figure 8. 2D-screen pointer control results, with discrete Kalman Filter – Control case. Overall, the achieved response matches the one presented by the initial version of the pointer, and is compatible with its expected use, with the benefit of stemming from a different sensing strategy.

For this implementation, no additional pre-processing strategies were applied on the face-detection and angular-velocity signals, only recurring to the programmable high-pass filter, available for the L3G4200D gyroscope, prior to acquiring its outputs. In order to compare the relative performance between the two versions of the pointer, responses acquired from their use with the test application, for a control case (user with training in the use of both systems), were assessed. This test scenario allowed

mitigating the impact of adaptation and learning periods that new users of any interface go through, and is proposed as a preliminary strategy for evaluating relative virtues and flaws between these versions. Results achieved by the control case, in use of the pointer, and under analogous conditions to those shown in [5], show a stable pointer, which can be positioned in any desired spot, and that responds with no perceptible delays to gestures and head movements, in both directions. One of the disadvantages found for the initial version of the pointer consisted in the compromise between the precision of the pointer and its response-time.

Table 2 shows a comparison between characteristic parameters, derived of the analysis of responses from the initial and updated (Kalman) versions of the pointer, for the control scenario. These results confirm the presence of a significant improvement in the performance of the new pointer in the horizontal axis: both standard deviations (± 11 pixels) and ranges of the pointer (46 pixels), once stabilized in the vicinity of the desired option, showed significantly lower values than those achieved by the previous version of the system. Moreover, a similar performance in the vertical axis was observed between both versions. Additionally, it should be noted that the updated version achieved similar performance levels in both axes, unlike the response from the initial version.

In terms of pointer speed and control characteristics, the user is able to hold the pointer in each position, with the detailed precision, for a mean period of stability of 6.15 seconds (minimum 3.73 s, maximum 7.09 s), before a change in the requested option occurred. This result is comparable with the average 5.98 s, achieved in use of the initial version (4.28 s minimum, 7.05 s maximum), while a faster response to changes in position has been verified for this new version, which takes advantage of sensor fusion instead of nested filters, in terms of response delays. This combination of precision and stability, is considered more than sufficient for the application of this strategy in the selection of on-screen options through means of holding the pointer over each position for a certain period of time. Furthermore, the possibility of taking advantage of the high-speed sampling rate of the MEMS gyroscope for the detection of gestures and motion-commands, in the external acquisition board, is still available in this version, as an alternate means for the selection of options that are selected with the pointer.

Table 2. Comparative results nom pointer test application, contro						
	Varsion	Pointer Deviation	Pointer Range	Stable		
	version	(x, y) (pixels)	(x, y) (pixels)	Period (s)		
	Initial	$(\pm 22, \pm 12)$	(91, 55)	5.98 ± 0.61		
	Kalman	$(\pm 11, \pm 12)$	(46, 55)	6.15 ± 0.73		

Table 2. Comparative results from pointer test application, control

Lastly, it is highlighted that the purpose of this scenario was to test the combined response that could be achieved from the raw signals coming from two disjoint measurement strategies, and to assess the exclusive performance of the Kalman filter, with respect to these, as well as to the initial version of the pointer. Therefore, no additional filtering or restrictions have been applied to the outputs of the Kalman filters or to the signals coming from the sensors. This implies that the aforementioned performance could take advantage of basic signal-processing and filtering strategies, prior or subsequent to the sensor fusion algorithm, without posing significant compromises to the response time.

4. Conclusions

This work presented an updated signal-processing strategy, based on sensor fusion (between a camera and a MEMS gyroscope) achieved through a Discrete Kalman Filter. Based on this new signal processing strategy, a supplementary version for a pointing interface aimed at accessibility solutions, developed in previous work, has been achieved, presenting relative advantages under a similar working principle. Advantages found for this updated version, with respect to the original pointer proposal, include the following:

• Given its composition, its performance is less susceptible to external interferences and to its environment. This version only requires visibility of the face of the subject, and the correct placement of the sensor for its use, without needing further adjustments.

- The precision of the system has been greatly improved on the horizontal axis, and a comparable response has been achieved in the vertical axis, for the fixation of the pointer over an option, even when this version takes one less reliable source for the estimation of the angle, derived from the face detector.
- The delay in the response of the pointer has been greatly reduced, while holding this stability and precision, making the user experience of the product a more intuitive one.
- This version holds the robustness of the original version, through the combination of supplementary information sources, as well as the possibility to incorporate gesture detection (made possible by the high-speed update rates of the gyroscope signals) for the activation of options selected with the pointer.

In sum, a robust strategy for the real-time control of an on-screen pointer, directed to accessibility interfaces over small screens has been achieved, with an increased independence in its performance to its environment of deployment. The proposed combination between image-processing and external sensors, based on simple estimation models, has significantly increased the potential for the systems that are currently in development. These diverse sources of information supplement each other, and make for a more robust solution overall, bringing forward the possibility to add embedded gesture detection based on these sensors for the comprehensive control of the pointer, for future versions of the system.

5. References

- [1] Vittorio S A 2001 Micro-Electromechanical Systems, Cambridge Scientific Abstracts.
- [2] Sim N Gavriel C Abbott W and Faisal A 2013 The Head Mouse-Head Gaze Estimation 'in-thewild' with Low-Cost Inertial Sensors for bmi use 6th International IEEE/EMBS Conference on Neural Engineering.
- [3] World Health Organization 2010 *Medical Devices: Managing the Mismatch-an Outcome of the Priority Medical Devices Project*, WHO Press, Geneva.
- [4] ST Microelectronics 2010 Application Note 3192, Using LSM303DLH for a tilt compensated electronic compass, Rev. 1.
- [5] Ghersi I Miralles M T Mariño M 2014 Augmented-Reality and Alternative-Input System based on Simultaneous Head Position and Pose Estimation, *IFMBE Proceedings*, Vol 49, 2015, pp 222-225, DOI 10.1007/978-3-319-13117-7_58.
- [6] Apache License Agreement at http://www.apache.org/licenses/
- Bachman T 2006 Patient Positioning: More than Just "Turn Every 2h", *Barrow Quarterly*, Vol 22, N° 3, Barrow Neurological Institute.
- [8] Dodgson N A 2004 Variation and extrema of human interpupillary distance. SPIE Proc. vol. 5291 Stereoscopic displays and Virtual Reality Systems XI, California, USA, pp 36-46.
- [9] Peebles L Norris B 1998 Adultdata: The Handbook of Adult Anthropometric and Strength Measurements, Department of Trade and Industry, Great Britain.
- [10] Grover Brown R Hwang P Y C 1997 Introduction to Random Signals and Applied Kalman Filtering, 3rd Edition, Wiley, United States.
- [11] Sabatini A 2011 Kalman-Filter-Based Orientation Determination Using Inertial/Magnetic Sensors: Observability Analysis and Performance Evaluation, *Sensors*, Vol. 11, 9182:9206.

Acknowledgements

This work was carried out under –and supported in part by- UBACyT 2062120100011BA 2013-2016 project, Facultad de Arquitectura, Diseño y Urbanismo, Universidad de Buenos Aires (FADU-UBA), and by Facultad de Ciencias Fisicomatemáticas e Ingeniería, Pontificia Universidad Católica Argentina (UCA).