



SVMT: A MATLAB toolbox for stereo-vision motion tracking of motor reactivity

M.I. Vousedoukas^{a,*}, P. Perakakis^{b,c}, S. Idrissi^b, J. Vila^b

^a Forschungszentrum Küste, Merkurstraße 11, 30419 Hannover, Germany

^b Department of Personality, Evaluation and Psychological Treatment, Faculty of Psychology, University of Granada, Campus Cartuja, 18071 Granada, Spain

^c Laboratory of Experimental Economics, University Jaume I, Avda. Sos Baynat, 12071 Castellón, Spain

ARTICLE INFO

Article history:

Received 5 May 2011

Received in revised form

16 January 2012

Accepted 17 January 2012

Keywords:

Motor reactivity

Sensory stimulation

Stereo vision

Image processing

Matlab

Computer software

ABSTRACT

This article presents a Matlab-based stereo-vision motion tracking system (SVMT) for the detection of human motor reactivity elicited by sensory stimulation. It is a low-cost, non-intrusive system supported by Graphical User Interface (GUI) software, and has been successfully tested and integrated in a broad array of physiological recording devices at the Human Physiology Laboratory in the University of Granada. The SVMT GUI software handles data in Matlab and ASCII formats. Internal functions perform lens distortion correction, camera geometry definition, feature matching, as well as data clustering and filtering to extract 3D motion paths of specific body areas. System validation showed geo-rectification errors below 0.5 mm, while feature matching and motion paths extraction procedures were successfully validated with manual tracking and RMS errors were typically below 2% of the movement range. The application of the system in a psychophysiological experiment designed to elicit a startle motor response by the presentation of intense and unexpected acoustic stimuli, provided reliable data probing dynamical features of motor responses and habituation to repeated stimulus presentations. The stereo-geolocation and motion tracking performance of the SVMT system were successfully validated through comparisons with surface EMG measurements of eyeblink startle, which clearly demonstrate the ability of SVMT to track subtle body movement, such as those induced by the presentation of intense acoustic stimuli. Finally, SVMT provides an efficient solution for the assessment of motor reactivity not only in controlled laboratory settings, but also in more open, ecological environments.

© 2012 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

In experimental psychophysiology, human motor reactivity to sensory stimulation is usually indexed by surface electromyography (EMG), which requires the placement of several electrodes on the skin surface [1]. This limits the mobility of the subject and reduces the possible contexts in which motor responses can be recorded and evaluated. In

addition, EMG captures only the activity of specific muscle groups, whereas certain motor reflexes to sensory stimuli, such as the startle reflex, involve a global motor response characterized by complex muscle contraction patterns at different body areas [2]. A detailed investigation of these global motor reactivity patterns, in relation to underlying physiological control mechanisms, could result in a more systematic mapping of psychophysiological processes under sensory stimulation.

* Corresponding author. Tel.: +49 511 762 9231; fax: +49 511 762 9219.

E-mail addresses: vousedoukas@fzk-nth.de, vousedoukas@gmail.com (M.I. Vousedoukas).
0169-2607/\$ – see front matter © 2012 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.cmpb.2012.01.006

In this paper, we present a Matlab-based stereo-vision motion tracking (SVMT) system developed at the Human Physiology research group in the University of Granada. SVMT is a low cost, non-intrusive alternative method for the detection of motor reactivity that has so far been tested in controlled laboratory settings. However, due to its portability and flexibility, we envision its future use in more open, ecological environments. The SVMT software presented here combines freely available Matlab toolboxes with a set of specially designed functions, and has been successfully tested/validated in terms of its stereo-geolocation and motion tracking performance, including comparisons with the eyeblink startle measured by surface EMG.

2. Computational methods and theory

The system consists of the following components: (a) 3-D stereo-triangulation, involving accurate description of the stereo-camera system geometry, i.e. estimation of the centres of view and orientations on a common Cartesian coordinate system, in order to reproduce 3-D surfaces; (b) feature matching, related to the robust identification of common features (keypoints) between images; and (c) motion trajectory extraction, focussing on the processing of the keypoints in order to obtain the subject's location and displacement in 3D space and time. These components are described in detail below.

2.1. Lens distortion correction

Most camera lenses impose a certain amount of distortion on the acquired sceneries, the correction of which is fundamental for the precise transformation from image to real world coordinates. Distortion can be radially symmetric about the principal point (radial distortion), or symmetric along a line directed through the principal point (tangential distortion) [3]. Typical types of distortion are the barrel distortion and the pin-cushion distortion, with the former resulting in a tendency of image points to concentrate over the border, while the ones close to the image centre spread on a radial direction. The opposite behaviour is found in the case of pin-cushion distortion, in which images are shrinking towards the centre [4].

Lens distortion correction, or alternatively camera calibration procedure, includes the estimation of the camera's intrinsic parameters operations and can be expressed mathematically through the K matrix:

$$K = \begin{pmatrix} \alpha & 0 & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{pmatrix} \quad (1)$$

where α , β correspond to the focal length measured in u and v pixel units respectively, u_0 and v_0 to the image centre [called also principal point, e.g. see [5]] and I is the identity matrix.

Camera calibration is very well documented in the literature [e.g. see [3,5–8]]. In the present application we used a freely available MATLAB Camera Calibration Toolbox developed by Bouguet [6]. An A3 size checkerboard with vertical and horizontal rectangle dimensions 24 mm and 26 mm respectively, was located in both cameras' Field Of View (FOV) at

several positions, while synchronised image acquisition was active, to obtain pairs of images for each perspective of the pattern. Due to the good contrast of the checkerboard pattern, the corners of each box were extracted for all images using a semi-automatic procedure, resulting in a cloud of points with known coordinates in both image and real world space (relative to each other). The toolbox uses the above as input to iterative solvers so as to define the camera's intrinsic parameters, which were estimated once for each camera and were then assumed constant. According to the results, the mean of the calibrated values differed significantly from nominal values for all variables (with deviations ranging between 2.5% and 5%) and errors >10 pixels (in the corners of the frames) were observed for the original distorted frames.

2.2. The projection matrix

In the present application the pinhole camera model is assumed, according to which the world and image location of each point in the field of view are aligned to the camera Centre Of View (COV). Homogenous coordinates are used in the expression of the camera model, as they provide simplicity of notation and computational generality [4]; they are obtained from geographic coordinates after adding an extra dimension with values equal to one. The transformation from metric system coordinates (X , Y , Z) to image coordinates (U , V) and vice versa, utilizing homogenous coordinates, is performed using the 3×4 perspective transformation matrix P [9]:

$$P \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} U \\ V \\ 1 \end{bmatrix} \quad (2)$$

The P matrix can be decomposed as:

$$P = KR[I - C] \quad (3)$$

where K is the camera calibration matrix expressed by Eq. (1), I is the identity matrix, and C and R are the translation and rotation matrices of the camera's COV respectively [4,10].

The C translation matrix accounts for the camera COV position Cartesian coordinates (X_{cam} , Y_{cam} , Z_{cam}) in the world coordinate reference system and can be represented as:

$$C = \begin{bmatrix} X_{cam} \\ Y_{cam} \\ Z_{cam} \end{bmatrix} \quad (4)$$

The R matrix expresses the effective orientation of the COV and is defined by three angles COV-*pan*, COV-*tilt*, COV-*roll*, relative to the Cartesian system of coordinates:

$$R = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \quad (5)$$

2.3. Stereo-triangulation

When the camera's COV orientation and location and thus the P matrix are defined for each point in the field of view, Eqs. (1)–(5) result in a system of 2 equations (one for each image coordinate U, V) with 3 unknowns, the feature's x, y, z world coordinates. Geometrically, this means that the feature can be anywhere along the line defined by the camera's COV and its image coordinates [11]. When the feature is monitored by two cameras, its position is defined by the cross-section of each camera's line. As a result, the system of equations can be solved analytically, as long as the cameras' geometries are defined in the same coordinate system, a process called stereo-triangulation.

The projection matrix R estimation is an essential step for camera calibration. Thus, the dual camera array geometries were defined using the common checkerboard corner features, detected during the two cameras calibration procedure. Synchronous image acquisition allowed association of the identified checkerboard corners between the Camera 1 and Camera 2 frames. These common features, once introduced to Eq. (2), result in an over-specified system, permitting the selection of an optimal solution for the camera angles [for a detailed description see [5,6]] and COV through an iterative solver [Nelder–Mead simplex method, see [12]]. The present application aims to record the subject's amplitude and velocity of displacement. Therefore, an absolute coordinate system (linked to geographic coordinates) is not necessary. Thus, the COV of the left camera was considered as the origin of the Cartesian coordinate system. The above coordinates can be easily transformed to a fixed system by indentifying three or more points of known locations to the desired system and using them as reference to estimate the P matrix [see also [13]].

2.4. Feature matching

As mentioned above, an essential step for motion tracking and geo-location is to identify common features at both cameras' frames at the same instant. Associate pairs of keypoints between images were identified using the approach of Lowe [14], modified in MATLAB accordingly for the needs of the present application. The method was designed to extract distinctive invariant features from images considering rotation, affine distortion, changes in illumination and noise, while small computational times were achieved by following a cascade filtering approach. The four major stages of computation used by the algorithm to generate the set of image features are: (a) space-scale extrema detection, (b) keypoint localization, (c) orientation assignment, and (d) keypoint descriptor. Detailed description of the algorithm is beyond the scope of the present contribution and can be found in Lowe [14].

In the case of the processed imagery, the number of identified features on each image was found to vary from 300 to 500, most of which were located on the subject's body (the homogeneous background did not generate keypoints). Shutter speed was set to its minimum value (0.12 ms) in order to avoid blurring during fast movements. However, in such cases some blurring is inevitable and the number of features can be reduced, but still more than 200 pairs of keypoints are typically produced.

The keypoint identification steps are described below and according to the following image notation; $L-m$ and $R-n$ imply the m th and n th frame (where m and n are frame numbers) of the left (L) and right (R) camera respectively.

1. Features are identified in all images from both cameras.
2. The algorithm propagates through the set of images and matches common keypoints between frames $L-1$ and $L-i$, where $i=2-N$ and N is the total number of acquired images per camera. All matches are indexed according to their occurrence on frame $L-1$ (Fig. 1a).
3. Similarly the whole set of image pairs is processed to identify keypoint matches between frames $L-i$ and $R-i$, keeping the indexing of the previous step (Fig. 1b).
4. Keypoint matches are geo-located according to Eq. (2) and Section 2.2 (see also Fig. 2).

In brief, the above description implies that initially indexed and geo-located keypoint pairs (step 2), are tracked through the frames sequence (step 3) and geo-located (step 4). This results in a cloud of points for each instant (see Fig. 2), which result in keypoints' motion trajectories.

2.5. Motion trajectory extraction

The 'keypoints cloud' shown in Fig. 2 contains the most detailed description of the subject's movement. Individual keypoint motion trajectories are generated after combining all keypoint instantaneous positions. Given that specific body areas (i.e. head, arms, core) are usually of interest for the analysis of typical human motor reactivity, individual keypoints are grouped into clusters, each associated to a predefined body area.

The association of the identified keypoints to the body areas of interest takes place by processing frame $L-1$, following two possible approaches:

1. The feature-matching algorithm identifies specially designed markers attached at the subject's body areas of interest, after processing frame $L-1$ along with pictures of the markers (Fig. 3).
2. Body areas are manually defined on the first frame of each video sequence; this approach is also automatically applied either when the automatic procedure fails (e.g. Fig. 3), or if the user does not provide any marker photos.

The feature clustering and trajectory extraction takes place according to the following steps:

1. Body areas are associated to the identified keypoints, according to the procedure described above.
2. For each temporal step, all the keypoint positions associated to each body area are averaged to generate the 'representative body area position'.

$$\vec{r}_p(t) = \frac{1}{w} \sum_{j=1}^w \vec{r}_{p,j}(t) \quad (6)$$

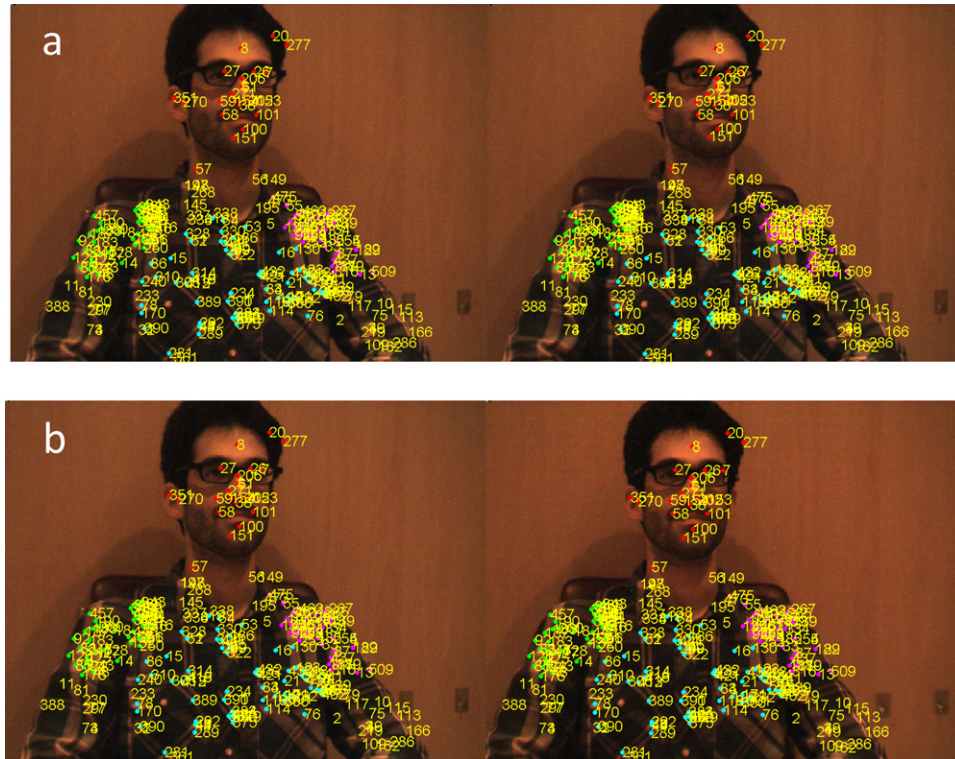


Fig. 1 – Performance of the keypoint matching algorithm between frames L-6/L-1 (a) and between simultaneously acquired frames from both cameras, L-6, R-6 (b). The color of the dots denotes different areas of the subject's body (red, head; magenta, core; yellow, left shoulder; and green, right shoulder) and the numbers the index of each keypoint, which allows tracking of each feature through the image sequence. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where $\vec{r}_p = (x, y, z)$ expresses Cartesian coordinates, p is the cluster index number and w is the number of keypoints per cluster.

3. Instantaneous velocities are estimated:

$$\vec{u} = \frac{\partial \vec{r}}{\partial t} \quad (7)$$

4. Following, instantaneous velocities from all points belonging into each cluster are averaged to produce the 'cluster velocities'.

$$\vec{U}_p(t) = \frac{1}{w} \sum_{j=1}^w \vec{u}_j(t) \quad (8)$$

An example of the above procedure is shown in Fig. 4, where the blue lines correspond to the x, y, z positions of all keypoints belonging to specific body areas (defined at step 1) and the black dashed line shows the final motion paths.

3. The stereo-vision motion tracking software

3.1. System installation

After camera installation and connection to the manufacturer's image acquisition software, the individual and stereo

camera calibration described in Section 2 has to be performed. Our software has been designed to use the calibration export files of the Bouguet MATLAB toolbox [6] directly. This procedure is necessary only during the initial system installation and has to be repeated only in case the relative camera positions and orientations change. A directory structure with all the necessary data files has to be prepared following recommendations given in Appendix 1.

The SVMt algorithm has been tested successfully in MATLAB Releases 2009a and 2010a in Windows XP SP2, Mac OS and Ubuntu 10.04 64bit platforms. It was run successfully in commercial laptop and desktop computers of 2008 up to 2011 standards and has not shown to be exceptionally computationally intensive. For some platforms (other than Linux or Windows) the SIFT keypoint identification algorithm has to be compiled to produce a working executable. The SVMt software is freely available at <https://sourceforge.net/projects/svmt>.

3.2. Basic inputs

The processing of each event is performed by the function `event_process`, through the operations described below and in the flow diagram shown in Fig. 5. The function requires two MATLAB data structures as input; one with the processing and visualization parameters (*params*) and another with the paths definition (*paths*). Images of the patterns for automatic area

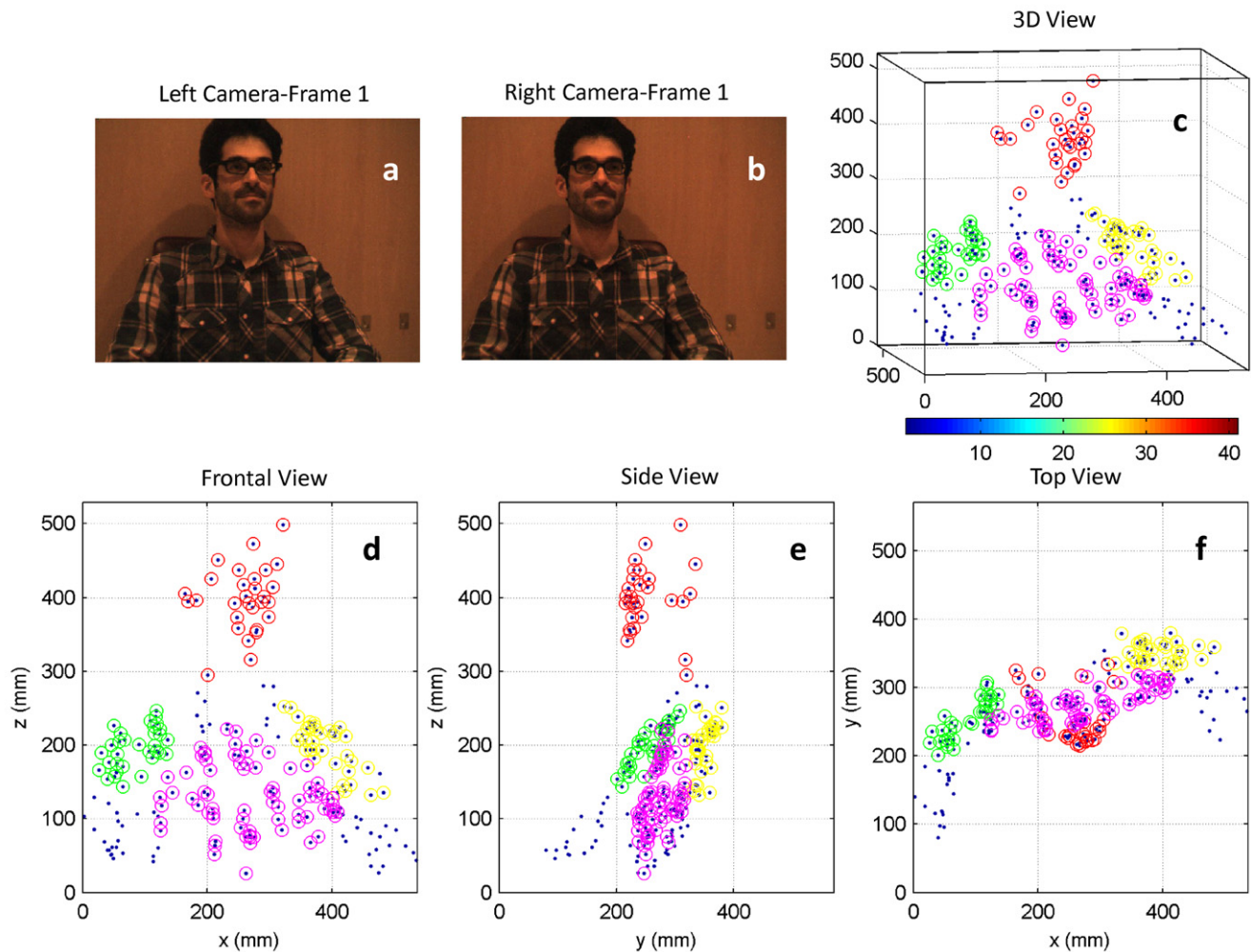


Fig. 2 – Keypoint pairs stereo-triangulation procedure for two instantaneous L-R camera frames (a and b). Blue dots show the detected keypoint locations in 3-D perspective (c), as well as frontal, side and top 2-D views (d, e and f, respectively). Colors denote different areas of the subject's body (red, head; magenta, core; yellow, left shoulder; and green, right shoulder). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

identification can be provided as an optional input. All inputs are described in detail in [Appendix II](#) and can be generated by a Graphical User Interface (SVMT.GUI) developed to facilitate the use of the SVMT toolbox ([Fig. 6](#)). However, users more familiar with programming are encouraged to use the scripts *SVMT_batch* and *SVMT_batch_preparation*, which allow the processing of several events in batch mode.

3.3. Image input

For the present application, images are exported by the image acquisition software in two separate ASCII files, one for each camera, named *CameraLeft.tmp* and *CameraRight.tmp*, containing the complete event image set. They are imported into SVMT in the form of a MATLAB data structure generated by the *read.tmp* function. For a different filename and image storage protocol, only the *read.tmp* function has to be modified accordingly to implement the SVMT toolbox.

3.4. Area definition

The appropriate body areas are defined either manually or automatically through the *define_sections* function. For manual definition, SVMT requests the user to manually define (using the mouse) the perimeter of sections on frame L-1 which contain the body areas of interest. For automatic identification, specially designed markers are attached at the subject's body areas of interest and are identified after comparison with images of the markers alone (see *patterns* data structure, [Appendix II](#), as well as [Section 2.4](#)).

3.5. Keypoint identifier

The keypoint identification algorithm is applied on pairs of images according to the procedure described in [Section 2.4](#), implemented by the *keypoints3* function. The extracted keypoints are associated with the pre-defined body areas



Fig. 3 – Automatic identification of body areas based on predefined picture markers.

(see Sections 2.5 and 3.4) through the `identify_sections_patterns` function and their geo-location is performed by the `geolocate_sections` function based on the theoretical background provided in Sections 2.2 and 2.3.

3.6. Trajectory extraction

For each keypoint, distinct positions tracked in time are combined to generate the final trajectory through the function `post_process`. The individual trajectories are further processed in the `extract_section_trajectories_velocities` function to generate the pre-defined body areas' motion paths and their corresponding velocity and acceleration time series (see also Section 2.5).

3.7. Data export

Once the processing procedure is completed, all the intermediate and final results are summarized in the MATLAB data structure `event` (see also Appendix II), which is exported in a user-defined directory (see also Appendix I). The same data structure is used as input in a set of functions oriented towards easy visualization and validation of the results (e.g. `trajectory_extraction_plot`, `trajectories_3Dplot`, `velocities_plot`, `trajectories_plot`, `make_event_video`). The activation of the above functions is an option during the initiation of the main `event_process` function of the SVMT algorithms and an option on the GUI interface (see also Section 3.2 and Appendix I and Fig. 6).

Table 1 – Camera technical specifications [15].

Specification	FFMV-03M2M/C
Image sensor type	1/3" progressive scan CMOS
Shutter type	Global shutter using Micron TrueSNAP™ technology
Image sensor model	Micron MT9V022
Maximum resolution	752 (H) × 480 (V)
Pixel size	6.0 μm × 6.0 μm
Analog-to-digital converter	On-chip 10-bit ADC
Video data output	8 and 16-bit digital data
Image data formats	8-bit and 16-bit raw Bayer data (color models)
Digital interface	6-pin IEEE 1394a for camera control, video data, power
Transfer rates	400 Mb/s
Maximum frame rate	752 × 480 at 61 FPS 320 × 240 at 112 FPS (region of interest) 320 × 240 at 122 FPS (2 × 2 pixel binning)
Partial image modes	Pixel binning and region of interest modes via Format.7
General purpose I/O ports	7-pin JST GPIO connector, 4 pins for trigger and strobe, 1 pin +3.3 V, 1 Vext pin for external power
Gain control	Automatic/manual, 0–12 dB
Shutter speed	Automatic/manual, 0.12–512 ms
Gamma	0–1 (enables 12-bit to 10-bit companding)
Synchronization	Via external trigger, software trigger, or free-running
External trigger modes	IIDC v1.31 Trigger Modes 0 and 3
Power requirements	8–30 V via IEEE-1394, less than one (1) Watt
Dimensions (L × W × H)	24.4 mm × 44 mm × 34 mm
Mass	37 g (including tripod adapter)
Camera specifications	IIDC 1394-based Digital Camera Specification v1.31
Memory storage	Three memory channels for user configurable power-up settings
Lens mount	CS-mount (5 mm C-mount adapter included) M12 microlens mount2
Compliance	CE, FCC Class B, RoHS
Operating temperature	0–45 °C
Storage temperature	–30° to 60 °C
Warranty	1 year

4. Test implementation and validation

4.1. Equipment and experimental layout

The SVMT method was originally developed to process imagery obtained during experiments designed to elicit the activation of motor reflexes, in the laboratory of the Human Physiology research group at the University of Granada in Spain. The image acquisition system consists of a Pentium IV Personal Computer and two PointGrey FireFly MV video cameras (Table 1), fixed on an aluminium base with dimensions 20 cm × 5 cm × 0.5 cm. The base is constructed in such a way that the relative position and viewing angles of the cameras are not affected during the platform's transport, which can be fixed on a camera tripod with standard 1/4"-20 tpi

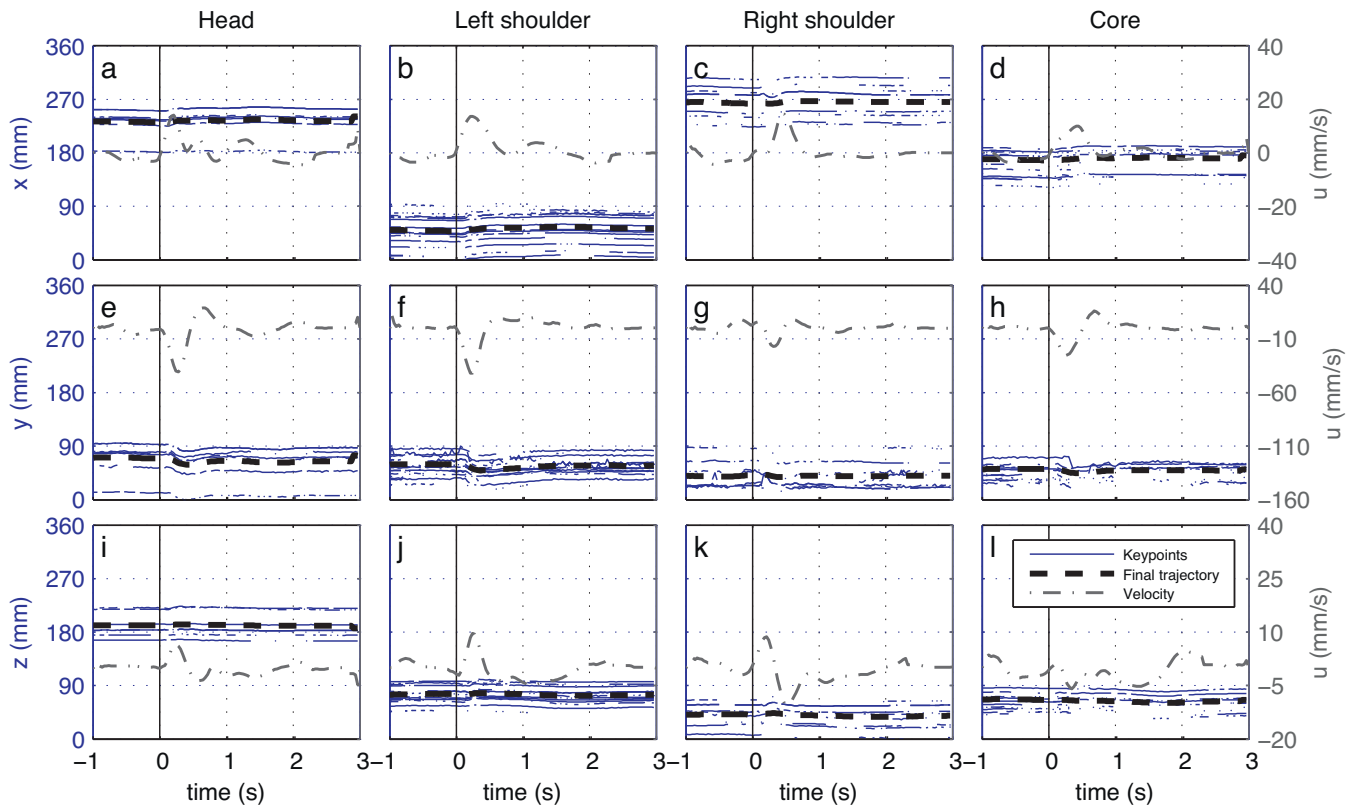


Fig. 4 – Final trajectory extraction procedure. Blue lines show the x-, y- and z-locations (rows) of keypoints along the frame sequence (horizontal axis denotes time), while the black dashed lines show the final motion paths. Columns correspond to different body areas (Head, left-right shoulder and core). The right vertical axis and the gray, dashed-dotted line correspond to the velocity. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(threads per inch) screw dimensions. The cameras (Point-Grey FFMV-03M2C-CS) were selected on the grounds that they combine several desired characteristics for the given application:

- Their small size (44 mm × 34 mm × 24.4 mm) is important in order not to affect the subject's behaviour (thus avoiding bias to the obtained data), and comes in line with the non-intrusive nature of the system.

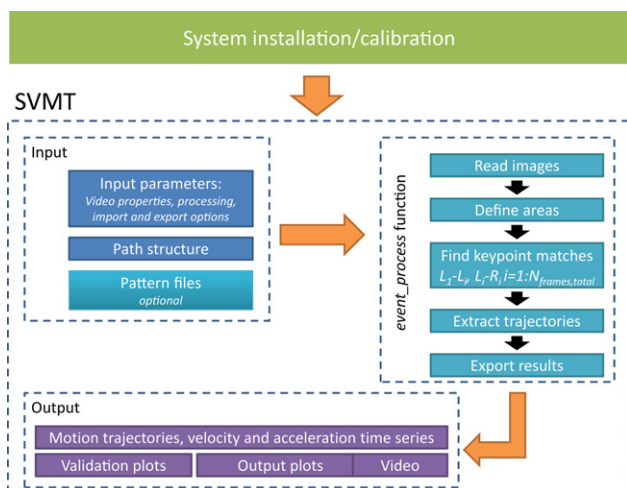


Fig. 5 – The SVMT toolbox flow diagram.

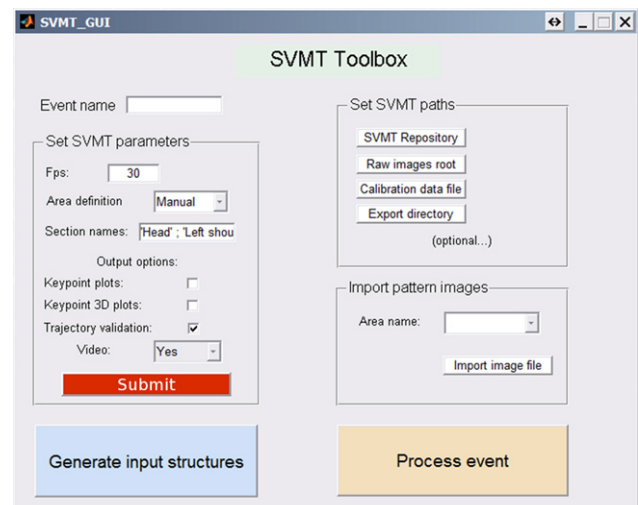


Fig. 6 – The SVMT toolbox graphical user interface.

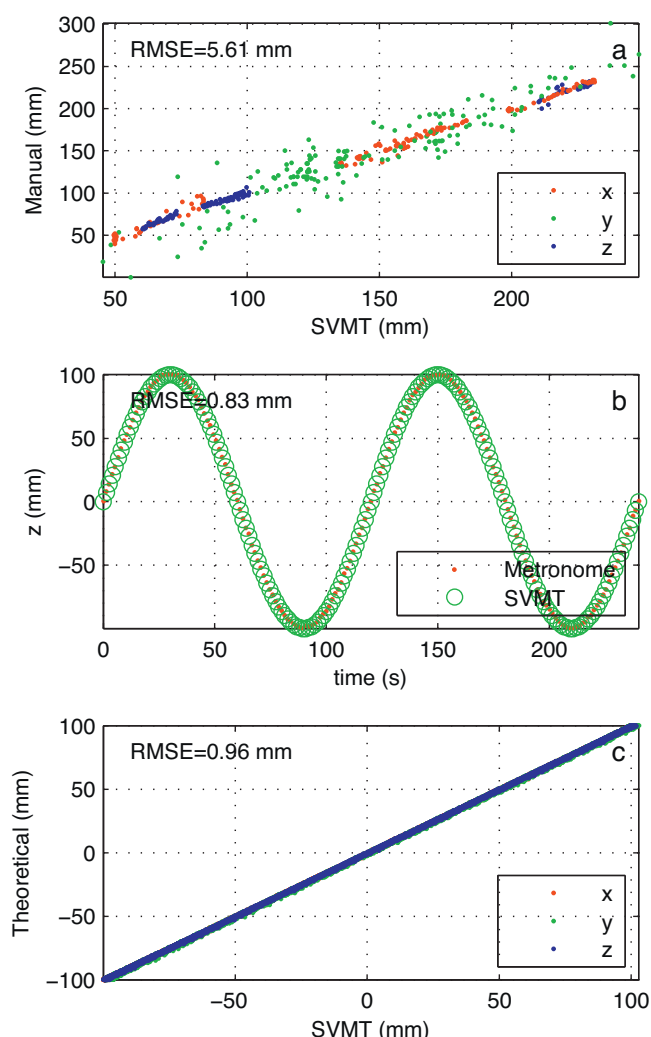


Fig. 7 – System validation comparing the SVMT motion paths with those obtained by manual tracking of the same movement: RMSE for the x, y, and z positions are equal to 1.96, 9.36, and 1.07 mm, respectively (a). System validation comparing the SVMT motion paths of the upper moving part of a mechanical metronome against the theoretical positions: (b) comparison of z-position time series, and (c) scatter plots. Colors red, green, blue, indicate x, y, z coordinates, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

- They provide image resolution and acquisition rate (752×480 pixels at 61 frames per second (fps)), as well as minimum shutter speeds of 0.12 ms that are low enough to reduce blur during fine and rapid movements expected in the present application.
- They allow robust software or external trigger synchronization and ‘time-stamping’ of the camera images, which is vital for a stereo-vision application, as well as to synchronize motion tracking data with other physiological measurements.

The cameras are installed in a room with dimensions of $2.5\text{m} \times 3\text{m} \times 3\text{m}$ and Faraday Caging, and are both connected to the PC through an IEEE-1394a Firewire PCI card, which controls synchronous image acquisition. Since stereo calibration is valid only for specific geometries, the cameras are fixed on a stable base, assuring that their relative positions and orientation remain stable. The PC is located outside the room, integrated in a broad array of physiological recording devices (Coulbourn Instruments, PA; Biopac Systems Inc., CA). A second PC running E-Prime stimulus presentation software (Psychology Software Tools Inc., PA) controls timing precision, transmitting (through the parallel port) an external trigger that is used for the synchronization of all data acquisition devices. In addition, for each image, the PointGrey software records timestamps with accuracy of few milliseconds. Experimental subjects are facing the camera array, which is located at a distance of approximately 1.5 m. The approximate FOV dimensions do not exceed 1 m for all 3 coordinate system dimensions.

We tested the stereo motion tracking software in a psychophysiological experimental context by analysing a series of images obtained at a sampling rate of 30 frames per second during the elicitation of a startle response to intense acoustic stimulation. Image acquisition initiated automatically 1 s before the presentation of the acoustic stimulus and continued for 3 s post-stimulus. The acoustic stimulus was a burst of white noise containing frequencies in the 20 Hz to 20 kHz range, 105 dB intensity and instantaneous rise-time. The noise was generated by a Coulbourn auditory stimulus generator and presented binaurally through headphones (Telephonic TDH Model-49, Telephonic Corporation, NY).

The eyeblink component of the startle response was measured by recording the electromyographic activity from the orbicularis oculi muscle beneath the left eye using Ag/AgCl miniature electrodes. The raw EMG signal, sampled at 1 kHz, was amplified (30,000), and frequencies below 90 Hz and above 300 Hz were filtered, using a Coulbourn S75-01 bioamplifier. The signal was subsequently rectified and integrated using a Coulbourn S76-01 contour-following integrator, with a time constant of 100 ms. The psychophysiological test was performed on ten undergraduate psychology students (three male and seven female), who participated in the experimental session for course credit. Each subject received three stimuli with an inter-stimulus interval of 2 min. A resting period of 10 min for adaptation to laboratory conditions preceded the presentation of the first stimulus.

4.2. Results

During the development stage and prior to its use in an experimental context, substantial validation efforts took place in order to assess the overall accuracy of the SVMT method. This accuracy is mostly affected by the combination of stereo-triangulation errors, the feature matching algorithm performance, and artefacts of the final motion paths extraction procedure. All the above are discussed independently in the following sections.

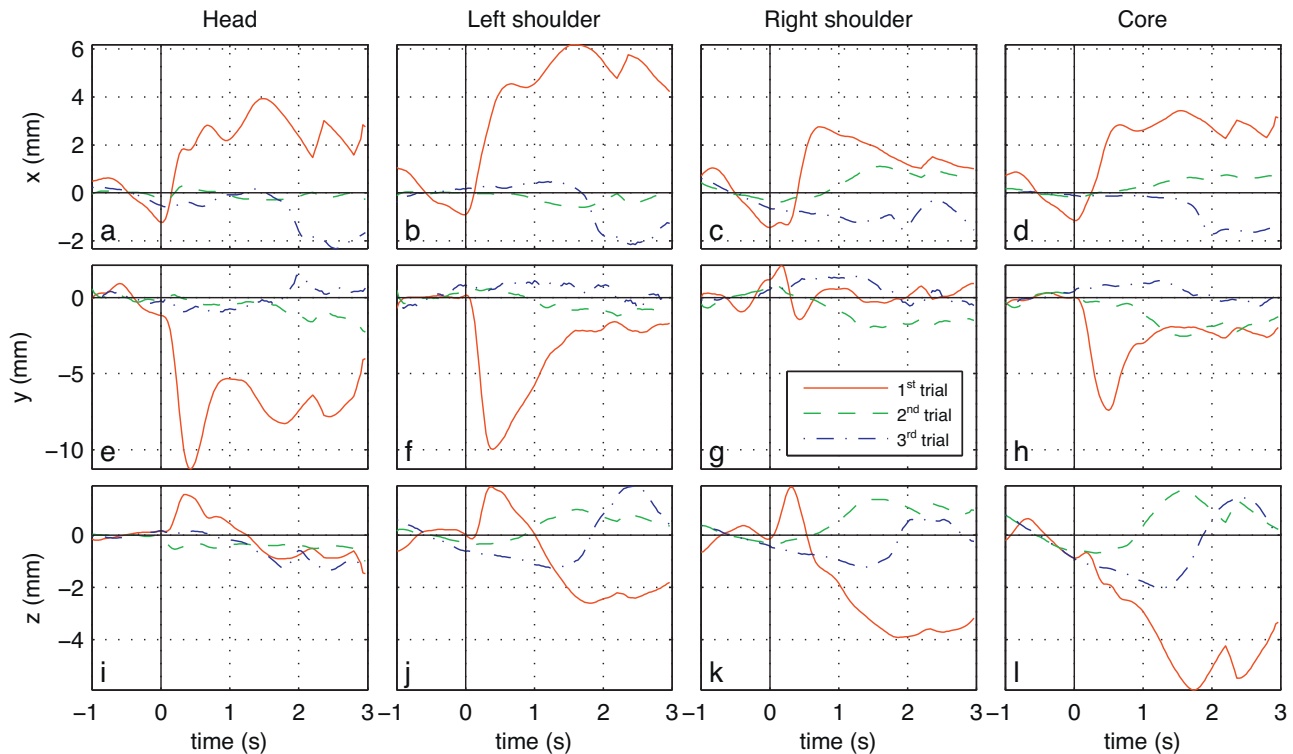


Fig. 8 – Motion trajectories extracted from the three consecutive presentations of an intense acoustic stimulus to a representative subject. Rows indicate Cartesian coordinate system axes and columns represent different body areas. In each graph, the first presentation of the stimulus is represented by a red continuous line, the second by a green dashed line and the third by a blue, dash-dotted line. Specific dynamical features of motor responses are observed, while habituation to repeated stimulus presentations is also evident. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.2.1. Stereo-triangulation accuracy

Stereo-triangulation errors were estimated during the camera calibration procedure and the camera geometry definition steps (see also Sections 2.1–2.3), as the checkerboard features provide detailed ground truth information to assess performance in three dimensions. For the present application, we define the lateral, anteroposterior and superior-inferior axes as the x -, y -, and z -axes, respectively. The stereo-triangulation accuracy is limited by the pixel footprint (distance per pixel ratio), which is directly controlled by the distance from the object, the distance of the cameras, and the image pixel resolution. Validation results showed that pixel footprints for the presently used stereoscopic system are around 0.5 mm. They were also found to be slightly higher for the anteroposterior axis (y). As a result, the estimated stereo-triangulation RMS error for the x - z and y - z planes was 0.3 mm, which is slightly lower than the error for the x - y plane (RMS error of 0.5 mm).

4.2.2. Feature matching performance

The SIFT algorithm has been extensively validated and applied for computer vision [14] and numerous other applications including commercial image stitching software (<http://www.kolor.com/panorama-software-autopano-pro.html>) and even coastal video monitoring [9,10], and has been proved very robust. The patterns used for the present application were selected after extensive testing,

thus ensuring that they consistently produce a satisfactory amount of extracted keypoints (>10).

In addition, the feature matching and tracking procedure was further validated by comparing the SVMT results with those of manually extracted motion paths for 4 specific features on the image. The features were manually identified (clicked with a mouse) along the entire image set (from both cameras) and stereo-triangulation followed. Such validation took place for two different events, with 3 different users performing manual tracking, in order to control for bias introduced by individual human performance. The overall RMS error was 5.6 mm, which corresponds to less than 3% of the movement range (Fig. 7a), which is well below the acceptable maximum for the present application. A closer look showed that the above error corresponds to the anteroposterior axis where the stereo-triangulation errors are higher (see previous section), while for the other two axes RMS errors were less than 2 mm (Fig. 7a). Moreover, it is expected that manual matching is less reliable than automatic, due to manual feature selection artefacts.

4.2.3. Validation using a metronome

Additional testing was performed by processing image sequences from an analogue metronome with a spherical pattern attached on its higher moving part. The metronome was set to perform a periodic motion with a period of 2 s and

imagery was obtained from different viewing angles and COVs. SVMT processing was applied to all obtained datasets and the estimated motion paths were compared according to the following procedure: (a) rigid transformation was applied to align the motion paths to the same coordinate system, since due to the varying angle and centre of view, they were originally expressed in different coordinate systems; (b) time lags were estimated and removed using cross-correlation technique; (c) since after the above pre-processing the extracted motion paths should be identical, the RMS error of all instantaneous positions were estimated for each axis, against the theoretical values of the motion paths (typical periodical oscillation with 2 s period); (d) an additional RMS error was estimated after dividing the original RMS errors per axis with the equivalent average amplitude of motion.

The extracted SVMT motion trajectories were in very good agreement with the metronome's periodic movement (Fig. 7b). Comparisons of the results between image sequences from different points of view showed RMS errors around 0.97 mm (c), which again were higher for the anteroposterior axis (RMS error around 1.15 mm), which is the one expressing distance from the camera towards the subject (see also Section 4.2.1). Estimated errors were found to be less than 1% of the metronome's movement amplitude.

4.2.4. Psychophysiological results

System application showed robust performance and processing times (2010 PC standards) around 120 s for each experimental trial tested. The modified Lowe [14] algorithm proved to be robust, and the small shutter times ensured image sharpness during the subjects' rapid movements, which is essential to maintain a sufficient number of identified keypoint matches. For the events tested, the minimum number of keypoint pairs per body area was always above 10 for each frame, which proved to be sufficient for effective motion tracking.

Fig. 8 depicts motion trajectories extracted from three consecutive presentations of the acoustic stimulus for a representative subject. The trajectories clearly illustrate the motor reactivity pattern of the subject to the acoustic stimulus. For example, during the first trial (red continuous line), the reaction of the subject was an upward (increasing z-position) and posterior (increasing y-position) movement, which was more enhanced for the head. At the same time, the right shoulder appeared to move forward (decreasing y-position), in the opposite direction of the other body areas of interest. This body movement is substantially more pronounced in response to the first presentation of the stimulus.

For the same subject, in Fig. 9 we observe the close correspondence between the EMG signal indicating the eyeblink startle response (black continuous line) and the velocity signal extracted by the SVMT (blue dotted line). The EMG signal indicates the expected eyeblink habituation to the intense noise; the response amplitude is progressively reduced as we move from trial 1 to trial 3. Interestingly, the velocity signal depicts a significantly more pronounced habituation of the global motor response to the acoustic stimulus (note the difference in the scale of the vertical axes in the three trials). Taken together, these observations confirm that the SVMT method can detect fine motor responses to sensory

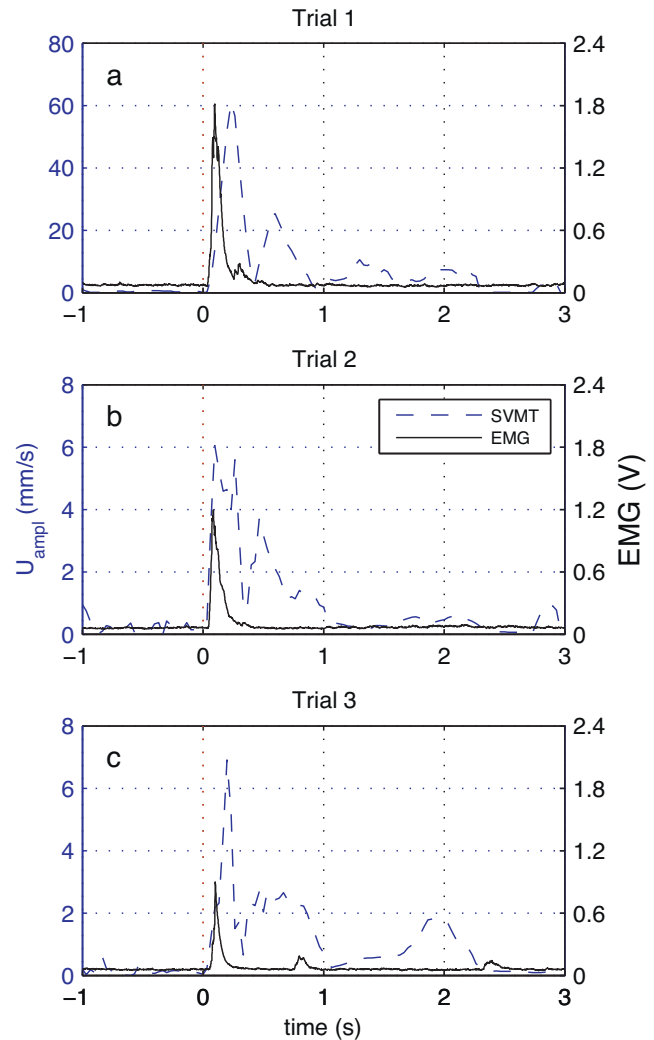


Fig. 9 – Comparison of EMG derived eyeblink startle and SVMT velocity signals in three consecutive presentations of an intense acoustic stimulus in one representative subject. On each plot left vertical axis and blue dashed line express velocity amplitude of the subject's head (spherical coordinates), while right vertical axis and solid black line the EMG signal in Volts. Horizontal axes express time in s, and the vertical red dotted line at $t = 1$ s is the time of the intense acoustic stimulus presentation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

stimulation and provide important complementary information to that obtained by measuring the activity of individual muscle groups.

4.3. Discussion

The SVMT system showed robust performance, while combining several other advantages, such as low cost, non-intrusive data acquisition, accurate quantitative information and bootstrap manner operation. Installation is not logistically or technically demanding and the stereo camera rig calibration is straightforward and can be performed in less than an

hour. The output format of the results facilitates processing, while several functions for analysis and plotting are already included in the toolbox.

Importantly, the presently tested set-up proved capable of monitoring very subtle movements, with amplitude of few mm (e.g. Fig. 8). In fact, the resolution of the measurement is directly related to the stereo-triangulation resolution, expressed through the pixel footprint (mm per pixel), which for the used stereo rig was around 0.5 mm. The noise of the measurements is also ranging at the pixel footprint dimensions and was not significant since the measured movements were typically of amplitudes reaching/exceeding 1 cm. The resolution can be further improved without any changes in the software, benefiting from the constantly increasing pixel resolution provided by commercial cameras.

Regarding the other possible error factors, the SIFT key-point detector was selected after testing several feature matching algorithms, and it was proven that it is not affected by light, color and contrast changes, as well as noise, at least to the extent found in the processed imagery. On the other hand, other feature matching procedures could be easily incorporated to SVMT. Furthermore, comparison of the velocity signal extracted by the SVMT system with the eyeblink startle measured by surface EMG (Fig. 9), clearly demonstrates the ability of SVMT to track subtle body movement as the one induced by the presentation of an intense acoustic stimulus.

The movements tracked in the present application were short in duration and did not involve full subject rotation. Thus, the point clustering/filtering procedure described in Section 2.5 was optimized for features which remain in the FOV. Even though modifying the processing algorithms to track subjects that fully rotate, or leave the FOV sporadically is beyond the present scope, such a feature could be included with reasonable effort and without significantly increasing computational times.

Finally, although in this paper we describe an application integrated in a psychophysiological laboratory, it is easy to imagine a similar setup in a more ecological environment, such as in a classroom, where, for example, the motor activity of students could be evaluated for the diagnosis of hyperactivity disorders. Another intriguing possibility is the use of such a system for the detection and patterning of facial expressions that can be used either in research, or even for the engineering of human-machine communication interfaces.

Conflict of interest

None of the authors has any financial or personal relationships with other people or organisations that could inappropriately influence (bias) their work.

Acknowledgements

We wish to thank the editor and the anonymous reviewers of *Computer Methods and Programs in Biomedicine*, whose insightful comments and suggestions significantly improved the quality of our manuscript. We also wish to express our gratitude to Kyr Giannis for the design and construction of the camera base for

the stereo rig and to Dagmara Wziatek and all the volunteers who participated in the laboratory experiments. The project was partially supported by the Spanish Ministry of Science and Innovation (grant PSI2008-04372).

Appendix I. Recommended directory structure

All the operations are taking place inside a root path that will be referred hereinafter as /SVMT. The root folder includes the following subdirectories:

~/Calibrations.Geometries: Calibration results inside a folder with name 'Calib_dd.mm.yyyy', including the calibration images and the export files.

~/Patterns: In case pattern templates are attached on the subject's body to automatically define desired areas to track, it is recommended that all pattern images are stored in this directory. An example of how to implement them is also provided. However, this is just a recommendation, and the algorithm will not be affected if the directory does not exist.

~/SVMT.Repository: The directory which contains the main MATLAB toolbox, with functions separated in categories: Camera_geometries, Keypoints, Plotting, Sections, Misc_functions, Trajectory_extraction. It should be added to the MATLAB path.

~/Test_Images: The directory where the original raw images are stored, separated in folders dedicated for each processed event. It is recommended that the names of the folders are unique for each event.

~/Exports: A directory automatically generated by the SVMT software where the data files of each processed event are stored in separate directories according to the unique event name mentioned in the previous paragraph. The most important data files are the MATLAB data structure files containing the raw images (~/Images/[test.name].mat) and the event summary including all intermediate and final results (event.summary.mat).

~/Workspace: The working directory. This directory should also contain the keypoints identifier executable used (SIFT, in this case), compiled for the corresponding operating system. Inside the Workspace directory, the following directories are generated automatically:

- 3D_trajectories – containing plots of the three-dimensional body area motion trajectories.
- Trajectory_extraction – containing plots per body area, of all the individual keypoint motion trajectories, as well as the final ones (see also Fig. 4).
- Velocities – containing velocity plots for each body area
- Trajectories – containing plots of the final motion trajectories for each body area
- Videos – containing the .avi video files giving an overview of each processed event
- Sections – containing plots showing the defined areas on the first image of each processed test

Appendix II. Basic input/output variables

The *params* structure

params.asects – sets the approach to define the preferred body area: 1 for manual definition
params.fps – the video acquisition frequency in frames per second
params.kplot – if set to 1, the frames with all identified keypoints are plotted at the /SVMT/Exports/Keypoints directory (recommended value: 0)
params.tdplot – if set to 1, a 3D scatter plot is generated for each frame showing the keypoints cloud. The plots are stored at the /SVMT/Exports/Keypoints directory (recommended value: 0)
params.trajplot – if set to 1, the trajectory extraction validation plot is generated at /SVMT/Workspace/ Trajectory_extraction (recommended value: 1)
params.mvideo – option to generate the event summary video file at /SVMT/Workspace/Videos (recommended value: 0; 1 to make video without compression, for compression input the compression string instead)
params.sc – a cell containing the names of the desired body areas as strings (e.g. 'Head'; 'Left shoulder'; 'Right shoulder'; 'Core')

The *paths* structure

paths.repos – the path of the toolbox directory SVMT.Repository
paths.raw – the path of the Raw image directory
paths.exports – the Exports path location
paths.calibration – the Calibration export files path

The *patterns* structure

A $n \times 1$ structure, where n is the number of preferred body areas.
patterns.im – the pattern image as imported in MATLAB with the *imread* function
patterns.name – the name of the preferred body areas (equivalent to *params.sc*)

The *event* structure

event.keypoints – all the information related to the identified keypoints
event.geops – stereo-triangulation results of all the identified keypoints, also separated in body areas
event.sections – perimetry of the areas containing each body area in the first frame of the left camera (L-1)
event.ntraj – trajectories of all the identified keypoints, as well as their corresponding image coordinates (not to be confused with *event.traj*, which is an intermediate result)
event.strajs – trajectories of the identified keypoints grouped per body areas, as well as their corresponding image coordinates

event.trajectories – the final estimated trajectories of the body areas of interest
event.velocities – the final estimated velocities of the body areas of interest
event.time – the time vector (in seconds) of the processed event
event.nlims – provides the range of the x, y, z coordinates of the identified keypoints for better visualization of the outputs

REFERENCES

- [1] L. Tassinari, J. Cacioppo, The skeletomotor system: surface electromyography, in: Handbook of Psychophysiology, 2000, pp. 163–199.
- [2] T.D. Blumenthal, B.N. Cuthbert, D.L. Filion, S. Hackley, O.V. Lipp, A. van Boxtel, Committee report: guidelines for human startle eyeblink electromyographic studies, *Psychophysiology* 42 (2005) 1–15.
- [3] R.Y. Tsai, A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses, *IEEE Journal of Robotics and Automation* RA-3 (1987) 323–344.
- [4] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, 2006.
- [5] J. Heikkilä, O. Silvén, A four-step camera calibration procedure with implicit image correction, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, 1997, pp. 1106–1112.
- [6] J.-Y. Bouguet, Camera Calibration Toolbox for Matlab. <http://www.vision.caltech.edu/bouguetj/calib.doc/> (2007).
- [7] Z. Zhang, A flexible new technique for camera calibration, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (2000) 1330–1334.
- [8] J. Weng, P. Cohen, M. Herniou, Camera calibration with distortion models and accuracy evaluation, *IEEE Transaction on Pattern Analysis and Machine Intelligence* 14 (1992) 965–980.
- [9] M.I. Voutsoukas, G. Pennucci, R.A. Holman, D.C. Conley, A semi automatic technique for Rapid Environmental Assessment in the coastal zone using Small Unmanned Aerial Vehicles (SUAV), *J. Coast. Res. SI64* (2011) 1755–1759.
- [10] M.I. Voutsoukas, P.M. Ferreira, L.P. Almeida, G. Dodet, U. Andriolo, F. Psaros, R. Taborda, A.N. Silva, A.E. Ruano, Ó. Ferreira, Performance of intertidal topography video monitoring of a meso-tidal reflective beach in South Portugal, *Ocean Dynam.* 61 (2011) 1521–1540.
- [11] P.R. Wolf, Elements of Photogrammetry (With Air Photo Interpretation and Remote Sensing), McGraw-Hill Book Company, St. Louis, MO, 1974.
- [12] J.C. Lagarias, J.A. Reeds, M.H. Wright, P.E. Wright, Convergence properties of the Nelder–Mead simplex method in low dimensions, *SIAM J. Optimiz.* 9 (1998) 112–147.
- [13] M.I. Voutsoukas, A.F. Velegrakis, K. Dimou, V. Zervakis, D.C. Conley, Wave run-up observations in microtidal, sediment-starved pocket beaches of the Eastern Mediterranean, *Journal of Marine Systems* 78 (2009) S37–S47.
- [14] D.G. Lowe, Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision* 60 (2004) 91–110.
- [15] Point Grey, Firefly MV Technical Datasheet, 2011.