Proceedings of the 2015
IEEE Conference on Robotics and Biomimetics
Zhuhai, China, December 6-9, 2015

Inertial Guided Visual Sample Consensus Based Wearable Orientation Estimation for Body Motion Tracking

Yinlong Zhang[1], Jindong Tan[2], *Member, IEEE*, Wei Liang[3,*], *Member, IEEE*, Yang Li[4]

***Abstract*— This paper presents a novel orientation estimate scheme using Inertial Guided Visual SAmple Consensus (IGVSAC) strategy for human body motion tracking. Unlike the traditional visual based orientation estimation methods where outliers among image-pair putative correspondences are removed based on hypothesize-and-verify models such as costly RANSAC, our approach novelly exploits motion prior information (i.e., rotation and translation) deduced from quick-response Inertial Measurement Unit (IMU) as the initial body pose to assist visual sensor in removing hidden outliers, which effectively overcomes the major drawback of those sample-and-consensus models. In addition, our IGVSAC algorithm is able to ensure the estimation accuracy even in the presence of large quantity of outliers among correspondences. Apart from that, the estimated orientation from visual sensor is, in turn, able to correct the IMU estimates using feedback control tactic, which can address IMU inherent long-term drifting issue. Extensive experiments are conducted to verify the effectiveness and robustness of our IGVSAC algorithm. The comparisons with highly accurate VICON Optical Motion Tracking System prove that our orientation estimate system is quite suitable for human body joint capturing.**

I. Introduction

Nowadays, robust wearable orientation estimate is increasingly exerting fundamental roles in a wide range of applications, such as stroke patient rehabilitation, fall detection, indoor localization and navigation, clinical gait analysis, etc [1]–[3].

Generally, inertial-based methods and visual-based methods are the most commonly used in wearable orientation estimate.

The traditional way to estimate the changes in orientation is to take advantage of IMU due to its quick response, compact size, non-invasiveness [3]. Unfortunately, this type of sensor, in effect, is not widely applicable in the case of long-term estimation which boils down to its inherent accumulated drift, its wearable fluctuation, soft tissue artifact, etc. Visual sensor characterizes in stability, accuracy and drift-free features [4]. However, in the process of feature point matching, to distinguish the outliers among the putative correspondences consumes heavy computational time. Worse still, the estimation performance will deteriorate dramatically if mistaken matches exist.

The visual and inertial integration is a reasonable tool for body orientation estimates. The basic principle is to take advantage of their complementary properties in terms of their accuracy and frequency response. Generally, there are two types of Visual-Inertial integration for orientation estimate: (i) camera fixed at specific locations; (ii) camera worn on human body using first of view (FOV) capturing. Researchers [5] have focused on approaches that the selected multiple cameras are fixed on some specific locations in the room combined with the inertial sensors worn on human upper limbs to estimate the subject poses jointly. However, subject self occlusion and delimited effective areas greatly restrict the overall performance of such integration-based techniques in free-living environments. By comparison, the monocular camera (MC) rather than multiple cameras is adopted and combined with IMU [6] in conducting orientation estimates that characterizes first-view visual capturing. This compact MC-IMU system facilitates the wearability near human body joints. However, most of the methods either over-rely on inertial data or neglect to take advantage of inertial information in assisting visual orientation estimate that requires heavy computational burdens.

With respect to these problems in Visual-Inertial integration system for orientation estimate, we propose a novel algorithm called Inertial Guided Visual SAmple and Consensus (IGVSAC) based on our previous works [7]. In IGVSAC, IMU outputs are imposed upon image putative correspondences to distinguish inliers preliminarily. Then the posterior Bayes Rule and Expectation Maximization are subsequently adopted to select the inliers among feature point matches step by step. Afterwards, optimal poses are obtained from these selected image inliers. Eventually, the orientation estimate based on our proposed IGVSAC algorithm is employed for performing body joint capturing. The sub-millimeter accurate VICON motion tracking system validates that our wearable MC-IMU system is quite suitable for robust and accurate human joint capturing.

The main contributions of this paper that are considerably discrepant from others are: (i) IMU motion prior information is exploited as the initial pose estimate to assist visual based orientation estimate, which aims at ensuring the estimation accuracy even in the presence of large quantity of image outliers; (ii) Only a few iterations proceed for orientation estimate given image putative correspondences and IMU prior information; (iii) A consistent parametrization mechanism is

This work was supported by the National Science Foundation of China under contact 61172145 and 61333019.

Yinlong Zhang and Yang Li are with the Key Laboratory of Networked Control Systems, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, 110016, China and with University of Chinese Academy of Sciences, Beijing, 100049, China. {e-mail:zhangyinlong@sia.cn, liyang@sia.cn}

Jindong Tan is with the Department of Mechanical, Aerospace and Biomedical Engineering, University of Tennessee, Konxville, TN 37996, USA. {e-mail: tan@utk.edu}

Wei Liang is with the Key Laboratory of Networked Control Systems, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, 110016, China. {corresponding author, e-mail: weiliang@sia.cn}

978-1-4673-9675-2/15/$31.00 © 2015 IEEE

obtained that employs feedback control to compensate for IMU drifts.

II. Methodology

Our approach consists of three parts: (i) initial pose estimate from IMU using twist representation and exponential map; (ii) image outlier removal and optimal pose estimate; (iii) IMU drift correction, as shown in Fig. 1.

A. Initial Pose Estimate from IMU

Rigid body motion can be decomposed into rotation and translation [8]. Let us assume that $\{P_{ab}, R_{ab}\}$ describes the relationship of inertial frame B with respect to frame A; Q_a and Q_b represent the coordinates of point Q in frame A and frame B respectively. Then Q_a and Q_b will satisfy the following homogeneous equation:

$$\begin{bmatrix} Q_b \\ 1 \end{bmatrix} = \begin{bmatrix} R_{ab} & P_{ab} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} Q_a \\ 1 \end{bmatrix} = \bar{g}_{ab} \begin{bmatrix} Q_a \\ 1 \end{bmatrix}, \bar{g}_{ab} \in SE(3) \tag{1}$$

in which P_{ab} represents displacement vector of the origin of frame B from the origin of frame A; $R_{ab} \in SO(3)$ denotes the orientation of frame B relative to frame A. $SE(3)$ represents the Special Euclidean Space.

For each homogeneous matrix $\bar{g}_{ab}$, there always exists a corresponding twist $(\upsilon, \hat{\omega})$ in the tangent space $se(3)$:

$$\hat{\omega} = \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix} \tag{2}$$

$$\upsilon = -\omega \times q \tag{3}$$

where $\hat{\omega}$ is a skew-symmetric matrix corresponding to vector ω and ν is the velocity of body frame. ν is yielded by rotation axis ω and position q.

Let $\hat{\xi} \in se(3)$ represents the twist, then the exponential product of $\hat{\xi}\theta$ is an element that belongs to $SE(3)$ given by:

$$e^{\hat{\xi}\theta} = \begin{cases} \begin{bmatrix} e^{\hat{\omega}\theta} & (I - e^{\hat{\omega}\theta})(\hat{\omega}\upsilon + \omega\omega^T \upsilon\theta) \\ 0 & 1 \end{bmatrix}, \omega \neq 0 \\ e^{\hat{\xi}\theta} = \begin{bmatrix} I & \upsilon\theta \\ 0 & 1 \end{bmatrix}, \omega = 0 \end{cases} \tag{4}$$

If we define $\mathrm{g_{global,n}}(0)$ as the initial configuration of a rigid body in reference n relative to the global frame, then the final configuration is given by:

$$\mathrm{g_{global,n}}(\theta) = \mathrm{e}^{\hat{\xi}\theta} \mathrm{g_{global,n}}(0) \tag{5}$$

B. Preliminary on Image Outlier Removal and Optimal Pose Estimate

1) Camera Model: The camera model which is used in the following is built on the principles of pinhole camera projection [7]: 3D point is projected through the center of camera lens onto the image plane. The specific point P in world coordinate and the corresponding point p in image plane are related through the camera matrix $C_{3\times4}$, i.e.,:

$$\begin{aligned} &p \sim C_{3\times4}P = K_{3\times3}M_{3\times4}P \\ &P = (X,Y,Z,1)^T, p = (x,y,1)^T \end{aligned} \tag{6}$$

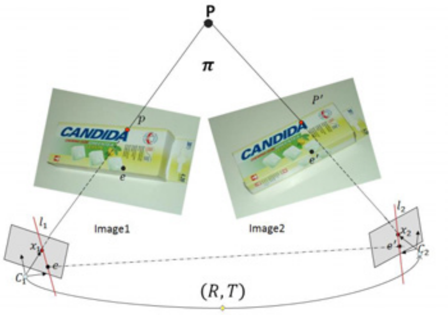


Fig. 2. Epipolar geometry and Epipolar line constraint

2) Two-view Epipolar Geometry: As illustrated in Fig. 2, a world point P is observed from two viewpoints by the moving camera. P is projected onto the consecutive images labeled p and p' respectively in the left and right images. l_1 and l_2 are the epipolar lines that correspond to x_2 and x_1; e is the epipole in $image1$ that all epipolar lines pass through. The same property holds for epipole e' in $image2$.

These two points, located at x_1 and x_2 in the image planes, are related by the fundamental matrix F. The epipolar constraint is given by

$$x_2^T F x_1 = 0 \tag{7}$$

C. IMU Guided Visual SAmple and Consensus

Our proposed IGVSAC is specifically designed to remove the outliers and derive the accurate body orientations using inertial and visual sensors. It proceeds by firstly taking advantage of IMU outputs as initial values for rotation and translation based upon kinematic model: twist representation and exponential map. Afterwards, these IMU data will be selected to assist distinguishing outliers among image-pair putative correspondences. Finally, these selected inliers will give rise to the accurate body poses and in turn, feedback the accurate optimal orientation estimate to IMU for drift correction.

1) Problem Formulation on IGVSAC: Assume that putative correspondences $S_n = \{(x_1, x_1'), (x_2, x_2'), \cdots, (x_n, x_n')\}$ contain a large amount of correct matches (also called inliers) and a few false matches (also known outliers). We define the set of unknown objective parameters as follows:

$$\begin{aligned} &\theta = \{R, T, O_n\} \\ &O_n = (o_1, o_2, \cdots o_n) \end{aligned} \tag{8}$$

where R and T represent body frame rotation and translation respectively. O_n consist of n point correspondence labels of which $\{o_i | o_i \in \{0,1\}, 1 \leq i \leq n\}$ denotes the $i-th$ point correspondence label: $o_i = 1$ implies the $i-th$ sample belongs to inlier group; $o_i = 0$ implies the $i-th$ sample belongs to outlier group.

Ideally, we would like to find the model that maximizes $P(R,T)$ given image putative correspondences and initial $\{R_0, T_0\}$ calculated from IMU quaternions. However, the algorithm is, in general, largely dependent upon the initial values $P(R_0, T_0)$ in the sense that malicious initial values will result in suboptimal solutions which is sometimes far from our expectation. Thus, the appropriate initial values

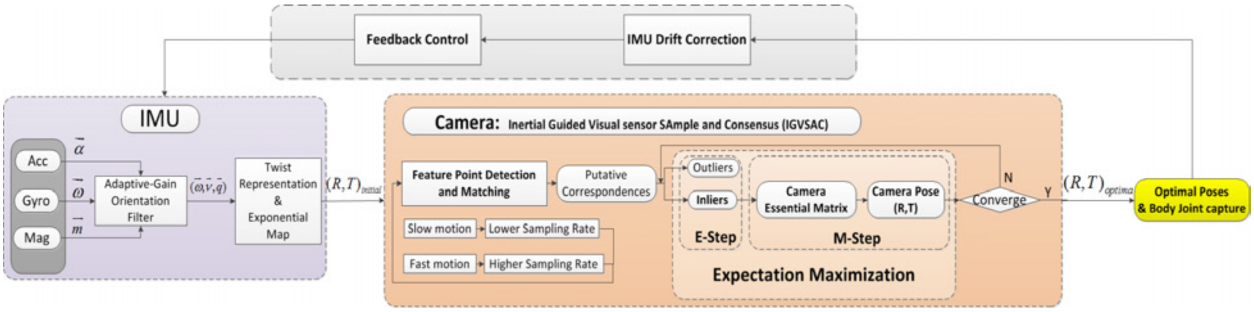


Fig. 1. The framework of the proposed approach for body orientation estimate

count for a great deal in our maximization algorithm. In IGVSAC, the initial values are determined by IMU outputs rather than being randomly selected, which dramatically avoids the suboptimal solution and meanwhile accelerates the optimization process.

In IGVSAC, the goal to estimate the body pose is equivalent to maximize the likelihood of P$\{\theta|(\eta_1,\eta_2,\cdots\eta_n)\}$:

$$P\{\theta|(\eta_1,\eta_2,\cdots\eta_n)\} = P\{\{R,T,O_n\}|(\eta_1,\eta_2,\cdots\eta_n)\} \quad (9)$$

where $(\eta_1,\eta_2,\cdots\eta_n)$ symbolizes the consecutive image putative correspondences:$\{\eta_i|\eta_i = \{x_i,x_i^{'}\}, 1 \leq i \leq n\}$

2) Maximum Posteriori Estimation: Using Bayesian Rule, the posteriori likelihood of parameter set θ (i.e., camera motion and putative correspondence labels) $P\{(\eta_1,\eta_2,\cdots\eta_n)|\theta\}$ is given as:

$$P\{\theta|(\eta_1,\eta_2,\cdots\eta_n)\} = \frac{P\{(\eta_1,\eta_2,\cdots\eta_n)|\theta\}P(\theta)}{P(\eta_1,\eta_2,\cdots\eta_n)} \quad (10)$$

Note that $P(\eta_1,\eta_2,\cdots\eta_n)$ is the prior joint probability of consecutive image point matching outcomes. Each component η_i in $\{\eta_1,\eta_2,\cdots\eta_n\}$ is assumed to be independent and identically distributed (i.i.d). It is merely related to predetermined feature point extraction and description using e.g., SURF algorithm. This assumption allows us to express the likelihood function as the product over all data points of the probability distribution evaluated at each data point.

The posteriori likelihood in Eq. 10 can be extended, i.e.,

$$\begin{aligned}
&\mathrm{P}\{(\eta_1,\eta_2,\cdots\eta_n)|\theta\}P(\theta) \\
&= P\{(\eta_1,\eta_2,\cdots\eta_n)|\theta\}P((o_1,o_2,\cdots o_n)|\{R,T\})P(R,T) \\
&= \prod_{i=1}^{n} P\{\eta_i|\theta\}P((o_1,o_2,\cdots o_n)|\{R,T\})P(R,T)
\end{aligned} \quad (11)$$

We then obtain the optimal solution on θ:

$$\begin{aligned}
\theta^* &= argmax_{\theta}P\{\theta|(\eta_1,\eta_2,\cdots\eta_n)\} \\
&= argmax_{\theta}\prod_{i=1}^{n}P\{\eta_i|\theta\}P((o_1,o_2,\cdots o_n)|\{R,T\})P(R,T)
\end{aligned} \quad (12)$$

Due to the fact that negative logarithm is a monotonically decreasing function with respect to its arguments, minimization of log function is equivalent to maximization of the objective function itself. By minimizing negative log-likelihood of the above objective function, we obtain the final objective function as follows:

$$\begin{aligned}
\theta^* = -argmin_{\theta}\{&\sum_{i=1}^{n} log(P\{\eta_i|\theta\}) + \\
&\sum_{j=1}^{n} log(P\{o_j|\{R,T\}\}) + log(P(R,T))\}
\end{aligned} \quad (13)$$

In Eq. 13, $P\{\eta_i|\theta\}$ is obtained by using the following equations:

$$P\{\eta_i|\theta\} = \begin{cases} \frac{\gamma}{\sqrt{2\pi}\sigma}e^{-\frac{\sum_{n=1}^{2}(y_n-f(x_n))^2}{2\sigma^2}}, & \{\{R,T\},o_i=1\} \\ \frac{1-\gamma}{S}, & \{\{R,T\},o_i=0\} \end{cases} \quad (14)$$

$P\{\eta_i|\theta\}$ is determined based upon the assumption that the error variable for inlier obeys Gaussian distribution while error variable for outlier conforms to uniform distribution. γ refers to the percentage of inliers among the whole putative correspondences. γ changes within a few iterations until convergence. Parameter S is the predefined search window area.

The posteriori of $P\{o_j|\{R,T\}\}$ rests upon the larger likelihood between $P(o_i=1|\{R,T\})$ and $P(o_i=0|\{R,T\})$:

$$\begin{cases} P\{o_i|\{R,T\}\} = max\begin{cases} P(o_i=1|\{R,T\}) \\ P(o_i=0|\{R,T\}) \end{cases} \\ o_j = argmax_{o_j}P\{o_j|\{R,T\}\} \end{cases} \quad (15)$$

$$\begin{cases} P(o_i=0|\{R,T\}) = \dfrac{\frac{1}{S}}{\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{\sum_{n=1}^{2}(y_n-f(x_n))^2}{2\sigma^2}}+\frac{1}{S}} \\ P(o_i=1|\{R,T\}) = \dfrac{\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{\sum_{n=1}^{2}(y_n-f(x_n))^2}{2\sigma^2}}}{\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{\sum_{n=1}^{2}(y_n-f(x_n))^2}{2\sigma^2}}+\frac{1}{S}} \end{cases} \quad (16)$$

where $\sum_{n=1}^{2}(y_n-f(x_n))^2$ refers to the Epipolar Sampson Errors [9].

The likelihood of $P(R,T)$ is given by

$$P\left(R,T\right)=\gamma\left\{\frac{1}{\sum_{i=1,o_i=1}^{n}\sum_{m=1}^{2}(y_{im}-f(x_{im}))^2+\sum_{j=1,o_j=0}^{n}{D_{o_j}}^2}\right\} \tag{17}$$

where y_{im} refers to the $i-th$ inlier; D_{o_j} represents the o_j outlier distance.

3) EM Estimation: To find the most suitable body frame motion, we have to seek the maximum likelihood solutions in Eq. 13. It proceeds within two phases: Expectation step and Maximization step. In the first Step, we use the previous motion estimates $\{R,T\}^{old}$ to find the posterior distribution of these latent variables and assign them as inliers or outliers. In the second step, the maximum likelihood function is calculated and the solution θ is obtained during each iteration. In our case, the hidden variable o_i obeys the Bernoulli distribution $\{0,1\}$ and derived via Eq. 15 and Eq. 16.

Without loss of generality, the solution via EM sometimes turns out to be local rather than global maxima in the case of non-convex context. However, in our IGVSAC algorithm, the initial values $\{R_0,T_0\}$ are directly obtained from IMU motion priors, which significantly avoids the solution being involved in the local maxima. Besides, the appropriate initial values will help accelerate the objective function convergence within a few iterative steps in comparison with the random initial value selections.

4) IMU drift compensation: As IMU suffers from severe drift problem in the long term which can be attributed to the fact that MEMS sensor like gyroscope is severely disturbed by random noises; accelerometer reads both body frame accelerations as well as gravitational components; magnetometer readings are sometimes contaminated by magnetic materials nearby, etc.

In the static situation, both body frames (inertial frame and camera frame) experience little or no movements, the corresponding points between the camera consecutive images vary little in image locations. In this context, the more reliable camera data determines that the body frame is whether or not in the static scene. Then the IMU will conduct static drift correction that gyroscope readings will be taken as zero, the accelerometer outputs will be taken as only gravity components and the quaternions $(q_t,\vec{q}_t)$ at time instant will be reset as $(q_{t-1},\vec{q}_{t-1})$ at time instant t-1.

In the dynamic situation, the camera pose $\{R,T\}$ derived from our proposed IGVSAC algorithm would be useful for conducting IMU drift compensation. What we yield iteratively from camera pose estimate is the optimal parameters: rotation and translation $\{\mathrm{R}^*,\mathrm{T}^*\}$ between two consecutive frames at time instant t_1 and t_2. Now the goal is to correct the IMU output quaternion q_{t_2} at time instant t_2, given the previous quaternion q_{t_1} and camera rotation $R_{t_1t_2}$ between t_1 and t_2.

The relationship between quaternion and rotation matrix can be referred to [10].

Given rotation matrix between the initial state to time instant t_1, we will obtain rotation matrix R_{t_2} at time instant t_2 and consequently the quaternion at time instant t_2 will be yielded as depicted in Eq. 20.

$$\begin{aligned} R_{t_2} &= R_{t_1}R_{t_1t_2} \\ q_{t_2} &= T_R^q(R_{t_2}) \end{aligned} \tag{18}$$

III. Experimental Evaluation

To evaluate the performance of our proposed approach for image-pair outlier removal and orientation estimate, we conduct several experiments. In this section, our wearable MC-IMU system setup, the real-experimental results and analysis are given as follows.

A. System Setup

In our wearable MC-IMU system, the monocular camera shown in Fig.3 (a) (Imaging Source product: $DFK23GV024$, Camera Lens: $HV7517$, focal length: $7.5mm-75mm$), samples the data at $5fps$ for static mode, $10fps$ for slow motion mode and $20fps$ for fast motion mode, all with a resolution of 752×480 pixels. The constant focal length is set as $7.5mm$.

We design an IMU board as shown in Fig.3 (b) that includes a hybrid functional module $MPU6500$ integrated with a MEMS three-axis accelerometer (full scale range: $\pm16g$), three-axis gyroscope (full scale range: $\pm2000deg/sec$) and three-axis magnetometer. An embedded ARM processor $STM32FL03$ and a Bluetooth module are also designed on this IMU board. The board size is $35mm\times30mm\times8mm$. The sensor signals are sampled at $40Hz$ and interfaced to the computer via Bluetooth.

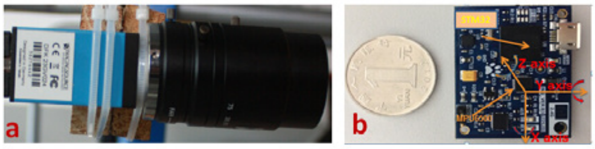
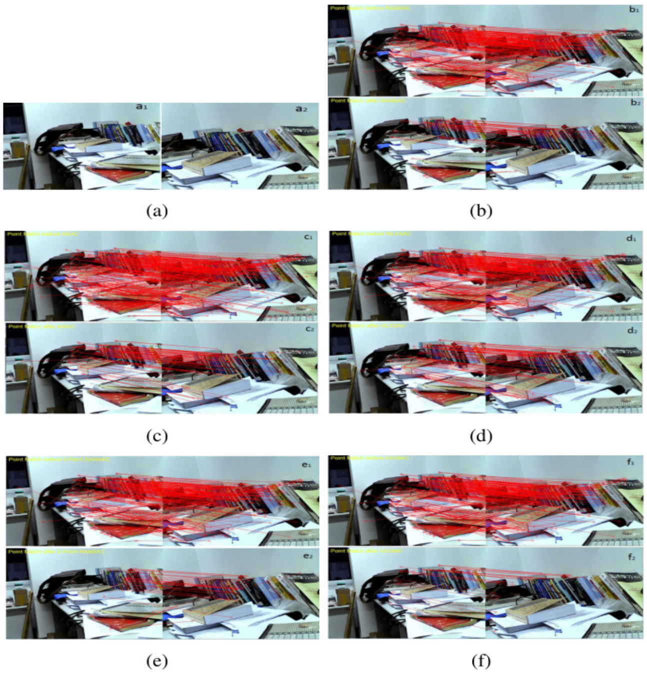


Fig. 3. The system modality: (a) the wearable monocular camera; (b) our designed IMU

B. Experiments and Analysis

1) Identifying Inliers and Rejecting Outliers: In this part, we test the performance of our proposed IGVSAC algorithm and verify the feature-point correspondences on real image pair taken from our laboratory corner. In our experiments, the putative correspondences are computed using SURF feature-point nearest matching method. Then the comparison of our IGVSAC algorithm with the prevailing methods: RANSAC, MSAC, MLESAC and 2-Point-RANSAC are performed as follows.

The putative correspondences are represented with red circles and connected by red lines as shown in Fig. 4. Apparently, before SAC methods, quite a few mismatches exist which deteriorate the calculation of fundamental matrix and will further result in mistaken camera pose estimates. By contrast, SAC methods lowers the percentage of outliers among selected correspondences.



(a) (b) (c) (d) (e) (f)

Fig. 4. Outlier removal comparisons. (a) a_1, a_2 are the original image pairs; (b) b_1 and b_2 show the putative correspondences before and after RANSAC; (c) c_1 and c_2 present the putative correspondences before and after MSAC; (d) d_1 and d_2 portray the putative correspondences before and after MLESAC; (e) e_1 and e_2 denote the putative correspondence before and after IMU-2-Point RANSAC; (f) f_1 and f_2 display the putative correspondences before and after our proposed IGVSAC algorithm.

TABLE I

Comparison of Methods for Image-Pair Outlier Removal

Outlier Removal Methods	No. of Putative Correspon-dences	No. of identified inliers	No. of mis-matches	Time (s)
RANSAC [9]	312	39	5	0.786
MLESAC [9]	297	48	1	1.358
MSAC [9]	309	51	2	0.884
IMU-2-Point RANSAC [6]	395	66	5	0.261
IGVSAC	302	28	0	0.185

Corresponding to Fig. 4, Table I reveals the number of inliers and outliers using SAC methods. The traditional SAC methods: RANSAC, MLESAC, MSAC are capable of alleviating the effect of outliers on pose estimate but fail to eliminate all the outliers for the sake of their randomness in sampling as well as their limited number of iterations. IMU-2-Point RANSAC method consumes less computational time in removing outliers but it still preserves five mismatches since it relies heavily on IMU gyroscopic data for rotation and its inherent randomness still exists though merely 2 points are selected. It is noticeable that our proposed IGVSAC algorithm outperforms these prevailing methods in computational time and in removing outliers. Unlike the other four methods presence of mismatches, the reason why our IGVSAC characterizes no mismatches is that IGVSAC firstly takes advantage of IMU outputs $\{R_0, T_0\}$ as initial values, followed by a few iterations for obtaining the maxima from the predefined cost function using EM in a probabilistic framework. Finally, the optimal pose $\{R^*, T^*\}$ is obtained within the relatively limited time, which is merely 0.185 seconds. Though the number of identified inliers is relatively lower than the other four methods, IGVSAC is able to guarantee the authenticity of the identified inliers.

2) Upper Limb Motion Tracking: During the test, two IMUs are fixed on the subject right arm planar region near the elbow and ulnar region near the wrist respectively. The monocular camera is fixed under the IMU for elbow. In order to seek a reliable estimation for the position and orientation of human arm movements, we resort to numerical algorithms for solving inverse kinematic problems. As previously mentioned in section II, we adopt twist representation and exponential map product characteristic of mapping (ν, ω) to the corresponding (R, T). It is assumed that IMU and camera are rigidly fixed and attached to the planar region near the elbow joint. The transformation between IMU frame $\{I\}$ and the camera frame $\{C\}$ is computed using [11]. To validate the performance of our IGVSAC method for motion capture system, we resort to VICON motion capture system with sub-millimeter accuracy as ground truth data.

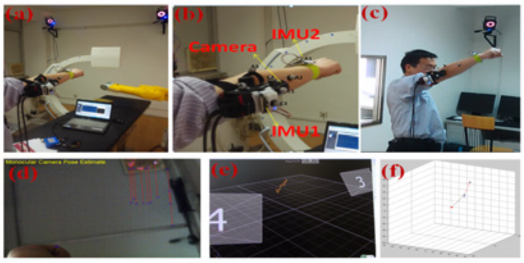


Fig. 5. VICON and wearable MC-IMU arm motion tracking. (a) Wearable MC-IMU and VICON System; (b) Camera, two IMUs and seven labeled VICON markers on arm; (c) Arm waving up and down; (d) Adjacent frame feature point correspondences; (e) Elbow and wrist position estimated using VICON system; (f) Motion tracking using MC-IMU system.

Fig. 5 shows a screenshot of the arm motion tracking. The participant is required to perform a series of gentle and smooth upper arm movements, i.e., arm lifting up and down, arm waving left and right and forearm bending over. Seven reflective markers are affixed in proximity to the elbow and wrist. The positions of the reflective marker for elbow and wrist, which are derived from VICON, serve as the benchmark.

In the tests, elbow angles are calculated from IMU measurements, MC-IMU measurements and are compared with ground truth from VICON datasets. In Fig. 6, the red curve represents the angle from VICON truth. The blue one and the green one show the angles estimated from IMU and MC-IMU systems respectively. Three types of movements are conducted in arm motion tracking tests, i.e., up-down movement, left-right movement and flexion-extension movement. The first two peak-and-trough pairs reveal the two times of arm moving up and down. Followed by arm waving left and right once. The third movement is flexion-extension (forearm bending over). As can be seen, in the first and second stage, the two curves both from IMU and MC-IMU match well with VICON data curve. In the third stage, however, there apparently exists severe discrepancy between MC-IMU curve and VICON curve, not to mention IMU curve and VICON curve. The reason for these differences is that soft tissue artifacts affects the motion tracking accuracy, which is inevitable.

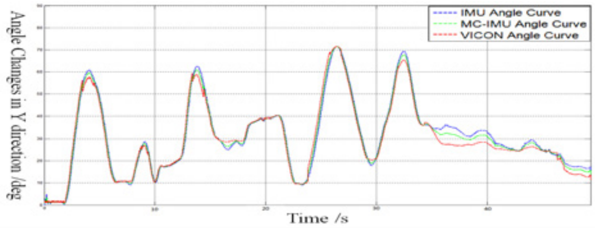


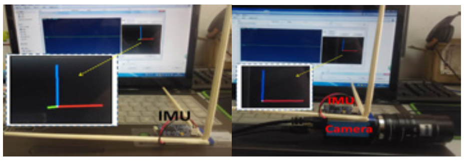
Fig. 6. Elbow angle curve comparison

TABLE II
ANGLE CURVE COMPARISONS OF IMU AND MC-IMU WITH VICON

Angle Estimate	Average Difference		Corrrelation	
	Test I	Test II	Test I	Test II
IMU [3]	3.2732°	2.3503°	0.8527	0.8993
MC-IMU	2.745°	1.562°	0.8943	0.9345

These tests are repeated two times, the corresponding angle curve average differences and correlations are shown in Table II. As can be seen that using monocular camera, to a large extent, assists IMU in improving angle estimate accuracy.

3) IMU Drift Correction: As shown in Fig. 7, the IMU drift correction is conducted. Originally, IMU is being fixed with its Z-axis vertical to the ground marked with blue line using Opengl; its X-axis vertical to the laptop screen marked with red line; its Y-axis vertical to X- and Z- axes marked with green line. Fig. 7 (a) depicts the IMU drifts over time along Z-axis. By comparison, when we add the camera to assist IMU, its drift turns out to be dramatically constrained, as demonstrated in Fig. 7 (b).



(a) (b)

Fig. 7. IMU drift correction. (a) IMU drifts over Time; (b) Visual assisted IMU for drift correction.

Fig. 8 shows the IMU drifts over time and its corrections using visual assistance. The upper plot shows the angle fluctuates about 5 degrees within 30 minutes. Whereas the lower plot displays the waveform of angles after visual correction that reduces the fluctuation of IMU random drift correction within 0.5 degrees.

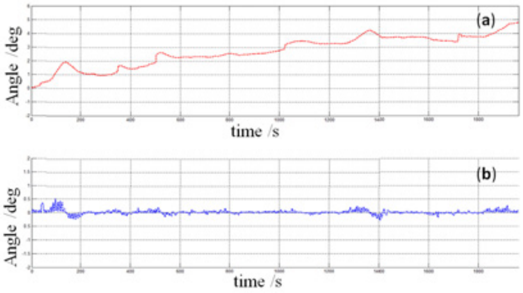


Fig. 8. IMU drifts and correction. (a) IMU drifts over time; (b) Visual assisted IMU drift correction.

IV. CONCLUSION AND FUTURE WORK

A novel orientation estimate approach called IGVSAC, which is specifically applied to human body joint capturing, is presented in this paper. Our proposed method effectively resolves the problems of visual-based major limitations: inaccuracy and heavy computations, which inherently exist in the process of distinguishing inliers and outliers among putative correspondences, by means of taking advantage of inertial data as motion prior knowledge.

Apart from that, the optimal estimate jointly calculated from visual and inertial data is able to feedback to IMU to correct its drift, which means that the inertial sensor estimation stability could be guaranteed in the long term. Its application on elbow capturing proves that our visual-inertial integration approach is quite suitable for wearable human motion tracking.

REFERENCES

[1] F. Raudies and H. Neumann, “A review and evaluation of methods estimating ego-motion,” *Computer Vision and Image Understanding*, vol. 116, no. 5, pp. 606–633, 2012.

[2] L. Pan, A. Song, G. Xu, H. Li, and B. Xu, “Upper-limb rehabilitation robot motion control based on dynamic interpolation,” *Robot*, vol. 5, pp. 549–535, 2012.

[3] X. Chen, J. Zhang, W. R. Hamel, and J. Tan, “An inertial-based human motion tracking system with twists and exponential maps,” in *IEEE International Conference on Robotics and Automation*. IEEE, 2014, pp. 5665–5670.

[4] D. Scaramuzza and F. Fraundorfer, “Visual odometry [tutorial],” *IEEE Robotics & Automation Magazine*, vol. 18, no. 4, pp. 80–92, 2011.

[5] Y. Tao and H. Hu, “A novel sensing and data fusion system for 3-d arm motion tracking in telerehabilitation,” *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 5, pp. 1029–1040, 2008.

[6] C. Troiani, A. Martinelli, C. Laugier, and D. Scaramuzza, “2-point-based outlier rejection for camera-imu systems with applications to micro aerial vehicles,” in *IEEE International Conference on Robotics and Automation*. IEEE, 2014, pp. 5530–5536.

[7] Y. Zhang, W. Liang, Y. Li, H. An, and J. Tan, “Orientation estimation using visual and inertial sensors,” in *2015 IEEE International Conference on Information and Automation*. IEEE, 2015, pp. 1871–1876.

[8] X. Zhao, Q. Huang, Z. Peng, L. Zhang, and K. Li, “Kinematics mapping of humanoid motion based on human motion,” *Robot*, vol. 27, no. 4, pp. 358–361, 2005.

[9] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[10] R. M. Murray, Z. Li, S. S. Sastry, and S. S. Sastry, *A mathematical introduction to robotic manipulation*. CRC press, 1994.

[11] J. Lobo and J. Dias, “Relative pose calibration between visual and inertial sensors,” *The International Journal of Robotics Research*, vol. 26, no. 6, pp. 561–575, 2007.