Full Scaled 3D Visual Odometry from a Single Wearable Omnidirectional Camera

Daniel Gutiérrez-Gómez, Luis Puig, J.J. Guerrero

Departamento de Informática e Ingeniería de Sistemas (DIIS) -
Instituto de Investigación en Ingeniería de Aragón (I3A)
Universidad de Zaragoza, Spain
{danielgg, lpuig, jguerrer}@unizar.es

Abstract—In the last years monocular SLAM has been widely used to obtain highly accurate maps and trajectory estimations of a moving camera. However, one of the issues of this approach is that, due to the impossibility of the depth being measured in a single image, global scale is not observable and scene and camera motion can only be recovered up to scale. This problem gets aggravated as we deal with larger scenes since it is more likely that scale drift arises between different map portions and their corresponding motion estimates. To compute the absolute scale we need to know some kind of dimension of the scene (*e.g.*, actual size of an element of the scene, velocity of the camera or baseline between two frames) and somehow integrate it in the SLAM estimation. In this paper, we present a method to recover the scale of the scene using an omnidirectional camera mounted on a helmet. The high precision of visual SLAM allows the head vertical oscillation during walking to be perceived in the trajectory estimation. By performing a spectral analysis on the camera vertical displacement, we can measure the step frequency. We relate the step frequency to the speed of the camera by an empirical formula based on biomedical experiments on human walking. This speed measurement is integrated in a particle filter to estimate the current scale factor and the 3D motion estimation with its true scale. We evaluated our approach using image sequences acquired while a person walks. Our experiments show that the proposed approach is able to cope with scale drift.

I. INTRODUCTION

Visual SLAM has become one of the most trending research fields in the last years and has been addressed both by using stereo and monocular systems. The main feature of stereo systems is that, knowing the baseline of the cameras, detected landmarks of the scene can be triangulated and the visual odometry and landmark postions can be completely estimated. SLAM approaches using stereo systems have been presented in [15], [17], [19].

On the other hand, due to the impossibility to extract the depth of a landmark just from one single image, monocular systems only allow the camera motion and scene to be estimated up to an unknown scale. With this in mind, stereo systems may seem more appropriate than monocular ones to perform visual SLAM. However the use of single cameras for visual SLAM is still appealing since they are cheaper, more compact and easier to calibrate than stereo systems.

One of the most important and succesful works on monocular SLAM is the one developed by Davison *et al.* [4]. This approach uses a pattern of known size to initialise some feature locations allowing the SLAM to start. Thus the scale of the map is fixed by the size of this initial pattern. Later, the inverse depth parametrization for the map points [1] allowed the SLAM to start automatically without the need of using an initialisation pattern. In this case the scale is arbitrarily fixed by a depth prior for the map features and the acceleration noise setup parameter.

Althoug the scale can be initialised, *e.g.*, by a pattern of known size or some kind of prior, it is likely that scale drift arises between different portions of the scene as the size of the map gets larger. This drift acts as a source of incremental error in the SLAM estimation, which leads to a deformation of the final map even after applying loop closing techniques. In [23], Strasdat *et al.* propose a loop closing method which corrects the map deformation due to scale drift.

Visual SLAM using omnidirectional cameras has been proposed in [3], [14], [20], [24]. Due to the 360° field of view (FoV) of omnidirectional cameras, features last longer on the image than in the case of conventional cameras, specially during big camera rotations. The increased lifespan of the features on the image translates in a better estimation of the position of the features on the map, a lower need to initialise new features and an increased robustness .

In this work we extend the SLAM approach for catadioptric cameras by Gutierrez *et al.* [9] which derives from state of the art EKF monocular SLAM for conventional cameras [2]. This approach is used to compute the visual odometry from sequences of images acquired with a catadioptric camera mounted on a helmet (Fig. 1).

Fig. 1: (a) Hemlet-camera device used in our experiments. (b) Omnidirectional image captured with our device.

An induced effect of human walking is a head vertical oscillation whose frequency matches up with the step frequency [10]. Our work is based on the premise that if the 6 d.o.f. visual SLAM is accurate enough , this vertical oscillatory motion of the head should be visible. Fig. 2a depicts an example of this behaviour, where the camera trajectory was obtained by performing a visual SLAM algorithm. Hence the step frequency of the camera carrier could be measured by estimating the power spectra of the vertical component of the camera trajectory (Fig. 2b).

Walking speed is strictly calculated as the product of step frequency and stride length. However, there exist biomedical studies like the one lead by Grieve [8], which show an empirical relation between step frequency and the walking speed with no dependence on the stride length. Further studies explain this relation as the result of a human tendence to choose a step frequency that minimizes metabolic cost of locomotion at a given walking speed [25], [11].

Based on this, we propose an approach to calculate the scale of the visual odometry from a single omnidirectional camera carried on the head of a person. This is done by first performing spectral analysis on short sections of the trajectory to extract the step frequency. Then we compute the estimated walking speed using the relation between step frequency and walking speed, and finally this estimation is integrated into a particle filter which recursively computes the scale factor.incrementally for each section.

The paper is structured as follows. Related work on scale estimation with monocular vision systems is described in Sec. II. In Sec. III we explain in detail our approach to compute the scale factor. Sec. IV present the experimental results and finally in Sec. V we extract the conclusions.

II. RELATED WORK

The literature on scale estimation in monocular SLAM shows two ways to address this problem.

In some works additional sensors are used to compute the real scale of the SLAM trajectory and map. In [2], Civera *et al.* use GPS information to align and scale the SLAM estimation by a rigid transformation which minimizes the distance between corresponding trajectory points. However this approach was designed for a benchmarking purpose and has little practical utility. Lupton *et al.* [13] and Nützi *et al.* [18] use an IMU to resolve the scale. The former makes the true map scale observable by integrating the visual data and the IMU data within an information filter. This allows the computation of the true map and trajectory estimations with no bias due to acceleration noise and feature depth priors. The latter fuses the SLAM estimation and IMU data in an EKF framework to compute the scale factor. In [5], Eudes *et al.* integrate the odometry and visual information to obtain the scaled map. In [22] Scaramuzza *et al.* use the vehicle speed measurement to compute the distance between the last two frames and recover the 3D structure by triangulation of the common image points.

In other works the scale of the scene is estimated without additional sensors. As mentioned before, it is possible to fix
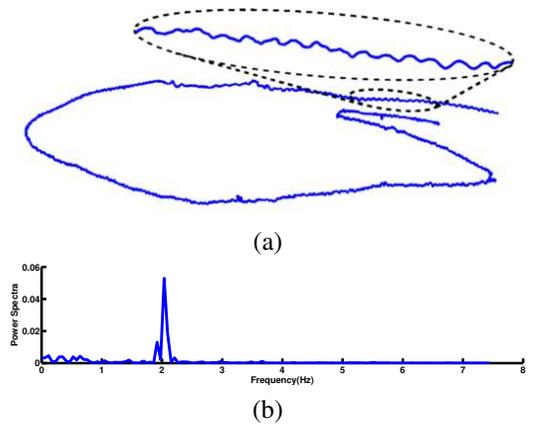

(a)

(b)

Fig. 2: (a) Trajectory estimation of Visual SLAM from a head-mounted catadioptric camera. (b) Power spectra of the vertical component

the scale with the size of the pattern used to initialise the landmarks needed to start the SLAM. However it does not avoid scale drift in large maps. Loethe *et al.* [12] use the prior knowledge of the distance of the camera to the ground plane to compute the scale factor of the scene, which is well suited for camera mounted on vehicles. Scaramuzza *et al.* [21] exploit non-holonomic motion constraints of wheeled vehicles to resolve the absolute scale, although this is only possible when the vehicle turns.

The approach presented in this paper fits in the second category, since no additional sensors are used. It takes advantage of the vertical oscillatory motion of a head-mounted camera to extract the step frequency. The prior information in this case is an empirical formula which allows us to compute the walking speed from the extracted step frequency. The computed walking speed is integrated in a particle filter framework used to smoothly estimate the scale factor rejecting possible spurious estimations of the walking speed.

III. PROBLEM FORMULATION

The V-SLAM approach used in this work is based on the Extended Kalman Filter (EKF) which is divided in two parts. In the first part, *Prediction*, the new state of the system is estimated from the previous time step state through the motion model. The second part of the algorithm, *Update*, uses the measurements of the environment to improve the new state prediction. The full state vector, composed of both the map and last camera location, is modelled as a multidimensional Gaussian distribution coded by its mean vector and covariance matrix.

The state of the system is given by the state vector $\mathbf{x}$

$$\mathbf{x} = (\underbrace{\mathbf{r},\ \mathbf{q},\ \mathbf{V},\ \omega}_{\text{Camera state}}, \underbrace{x_i, y_i, z_i, \theta_i, \phi_i, \rho_i, ...}_{\text{3D points (IDP)}}) \qquad (1)$$

where $\mathbf{r}_{(3\times1)}$ is the camera pose, $\mathbf{q}_{(4\times1)}$ is the quaternion of its orientation and $\mathbf{V}_{(3\times1)}$ and $\omega_{(3\times1)}$ are its linear and angular velocities, respectively.

The 3D points are parameterised in inverse depth parametrisation (IDP) [1]. As the depth of the landmarks

cannot be measured from one single image they are initialised with an arbitrary inverse depth prior ρ_{0i} with large uncertainty. This prior is gradually refined in succesive observations. The inability to measure the initial depth of the features involves the unobservability of the absolute scale of the scene. Thus, the scale of the SLAM reconstruction is biased due to the difference between the arbitrary depth prior and the true depth of the first measured landmarks. Moreover, scale is liable to drift due to the lost of old landmarks and the initialisation of new ones.

In this work we perform SLAM with an omnidirectional camera on a helmet carried by a human operator [9]. Since the head moves vertically with the step frequency during walking, SLAM visual odometry can be used to estimate the step frequency by spectral analysis.

To address the problem of determining the real scale dynamically, we propose a method which is performed iteratively on trajectory sections and can be subdivided in four steps:

- Spectral analysis on the SLAM visual odometry for step frequency estimation.
- Empirical estimation of walking speed from step frequency.
- Integration of walking speed in a particle filter for recursive scale factor estimation.
- Scaling of the final visual odometry.

A. Spectral analysis on SLAM visual odometry

In the case of our omnidirectonal camera, the camera frame is oriented with its z-axis pointing approximately to the direction of the normal to the ground plane, so the head vertical oscillation is given by the z-component of the camera position vector. To estimate the step frequency we apply spectral analysis to the z-component of the trajectory. The trajectory is divided in sections of N camera poses and for each section the data sequence $(z_1, z_2, ..., z_N)$ is considered, where z_n is the z-component of the n-th trajectory point of the current analysed section.

Prior to computing the power spectra, the data sequence is filtered to eliminate the low frequency harmonics introduced by the change of the ground height along the path.

The power spectra Γ is calculated by applying the Discrete Fourier Transform (DFT) to the filtered data sequence as follows:

$$\Gamma(f_m) = \frac{1}{F_s N} \| \sum_{n=1}^{N} z_n \exp\left(-j\frac{2\pi f_m(n-1)}{F_s}\right) \|^2 \quad (2)$$

$$f_m = \frac{mF_s}{N} \qquad m = 0, 1, ..., \frac{N}{2} \quad (3)$$

where F_s is the sampling frequency, which in our case is the number of frames per second (fps) of the camera and f_m are the frequencies for which the spectrogram is sampled.

We extract from the power spectra the maximum peak in the interval of feasible human step frequencies, which are assumed to be in the range between 1 and 3 Hz. This frequency is taken as the estimated step frequency f_{step}.

B. Walking speed estimation

To estimate the walking speed we consider the biomedical work by Grieve [8] where a relation between the step frequency (f_{step}) and the walking speed (V_{walk}) normalized with height (H) is presented:

$$V_{walk} = \alpha f_{step}^{\beta} H \quad (4)$$

where V_{walk} is in m/s, f_{step} in Hz, H in m, and α and β are characteristic parameters which differ from one individual to another.

For higher accuracy, we have computed our own α and β parameters for the camera operator. We measured the time t_i it took the operator to walk a distance $s = 100$ m at the times per step ΔT_i given by a metronome ranging from 0.45 to 0.80 seconds in intervals of 0.05 seconds (see Table I). The height of the operator is $H = 1.88$ m.

Normalized walking speeds V_i' and step frequencies f_i were computed from the raw experimental data. Then a power fitting was applied to obtain the values of $\alpha = 0.329$ and $\beta = 1.534$ (Fig. 3).

C. Particle Filter for scale factor tracking

Having the walking speed estimate, the scale factor for each section k could be easily computed by $d_k = \frac{V_{walk,k}}{V_{SLAM,k}}$, where $V_{SLAM,k}$ would be the mean of the speed at the N camera poses used in the computation of the power spectra. However, given the empirical method for the walking speed estimation and the possible high variability of the SLAM velocity along N frames, we decided to use a probabilistic filter for the computation of the scale factor. This allows us to introduce an uncertainty to the scale factor and at the same time the rejection of spurious estimations of the walking speed.

Thus, we consider a dynamic system whose state $\mathbf{x}_k$ is composed by the magnitude of the SLAM velocity $V_{SLAM,k}$ and the logarithm of the scale factor $\lambda_k = \log_{10}(d_k)$. We

TABLE I: Experimental data used to compute the empirical *Step frequency-Walking speed* relationship for the camera operator.

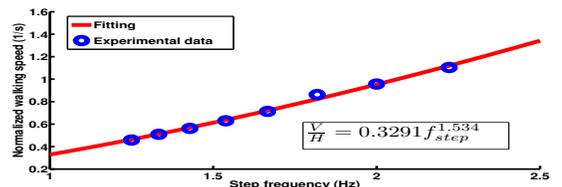| ΔT_i [s]	t_i [s]	$f_i = \frac{1}{T_i}$ [Hz]	$V_i' = \frac{s}{t_i H}$ $\left[\frac{1}{s}\right]$
0.45	48.18	2.22	2.08
0.50	55.60	2	1.80
0.55	61.63	1.82	1.62
0.60	74.54	1.67	1.34
0.65	84.42	1.54	1.19
0.70	94.63	1.43	1.06
0.75	104.42	1.33	0.96
0.80	116.06	1.25	0.86

Fig. 3: Power fitting of the experimental data to compute the relation between walking speed and step frequency ($\mu_{err} = 0.018$, $max_{err} = 0.04$).

take the logarithm instead of directly the scale factor to avoid negative scale values.

$$\mathbf{x}_k^{(L)} = \begin{bmatrix} V_{SLAM,k}^{(L)} \\ \lambda_k^{(L)} \end{bmatrix} \tag{5}$$

To track the scale factor, a particle filter with Sampling Importance Resampling is designed [7]. We use a particle filter rather than an extended Kalman filter (EKF) so that it can deal with high uncertainty priors of the scale factor which would involve a large linearization error in an EKF approach.

Hence the state of the system in each section k is approximated by a set of particles:

$$S_k = \left\{ (\mathbf{x}_k^{(L)}, w_k^{(L)}) \mid L = 1, 2, ..., P \right\} \tag{6}$$

where P is the number of particles and $\mathbf{x}_k^{(L)}$ and $w_k^{(L)}$ are respectively the state vector and the resampling weight of particle L.

The particles are initialised such that the initial values of $\lambda_0^{(L)}$ are drawn from a Gaussian distribution $\lambda_0 \sim \mathcal{N}(0, \sigma_0)$, where σ_0 is a parameter related to the orders of magnitude being scoped out.

In the first step of the particle filter, particles are sampled down by a proposal distribution $p(\mathbf{x}_k | \mathbf{x}_{k-1})$:

$$\mathbf{x}_k^{(L)} \sim p(\mathbf{x}_k | \mathbf{x}_{k-1}^{(L)}) \tag{7}$$

In our system the sampling of the proposal distribution includes both the update of the SLAM velocity, which is taken as a control input coming from the visual odometry, and the possible drift in the scale. This is encoded in the following equations:

$$V_{SLAM,k}^{(L)} = \mu_{V,k} + \nu^{(L)} \tag{8}$$

$$\lambda_k^{(L)} = \lambda_{k-1}^{(L)} + \alpha^{(L)} \tag{9}$$

with $\nu^{(L)} \sim \mathcal{N}(0, \sigma_{V,k})$ and $\alpha^{(L)} \sim \mathcal{N}(0, \sigma_{drift})$, and where $\mu_{V,k}$ and $\sigma_{V,k}$ are the averaged speed and the corresponding standard deviation of the last set of N SLAM camera poses used for spectral analysis, and σ_{drift} is the standard deviation prior of the scale drift between two consecutive sections, which is modelled as Gaussian noise.

Once the prediction has been computed, the samples are weighted by integrating the estimation of the walking speed $V_{walk,k}$ from spectral analysis as follows:

$$w_k^{(L)} = p(V_{walk,k} | \mathbf{x}_k^{(L)}) \tag{10}$$

where $p(V_{walk,k} | \mathbf{x}_k^{(L)})$ is the probability density function defined by the measurement model $h(\mathbf{x}_k)$ and the statistics of the sensor noise. By assuming that the speed estimation is affected by Gaussian noise of zero mean and standard deviation σ_{Vwalk} to be set up empirically, weights are computed as:

$$\omega_k^{(L)} = p(V_{walk,k} | \mathbf{x}_k^{(L)}) = \phi \left(\frac{V_{walk,k} - h(\mathbf{x}_k^{(L)})}{\sigma_{Vwalk}} \right) \tag{11}$$

where $\phi(z)$ is the probability density function of the standard normal distribution and the measurement function $h(\mathbf{x}_k^{(L)})$ is given by:

$$h(\mathbf{x}_k^{(L)}) = V_{SLAM,k}^{(L)} 10^{\lambda_k^{(L)}} \tag{12}$$

Finally, the weights are normalized for each particle and the next set of particles S_{k+1} is resampled by drawing P particles from the current set with probability equal to the corresponding weights.

D. Scaling of the trajectory

To track the scale factor, the original SLAM trajectory has been splitted in sections of N camera poses, being each of these sections denoted by their corresponding time-step k. The scale factor to be applied to each section k is obtained by averaging the logarithmic scale values of the particle set S_k and undoing the logarithmic change as follows:

$$\bar{\lambda}_k = \frac{\sum_{i=1}^{P} \lambda_k^{(i)}}{P} \tag{13}$$

$$\bar{d}_k = 10^{\bar{\lambda}_k} \tag{14}$$

This scale factor must be applied to the position and velocity of the N camera states of section k. To simplify, these state variables are encapsulated in a vector:

$$\mathbf{X}_k(n) = (r_x^n, \ r_y^n, \ r_z^n, \ v_x^n, \ v_y^n, \ v_z^n) \quad n = 1, 2, ..., N \tag{15}$$

Finally, to ensure the continuity in position and velocity, each vector $\mathbf{X}_k(n)$ is scaled recursively as follows:

$$\hat{\mathbf{X}}_k(n) = \hat{\mathbf{X}}_{k-1}(N) + \bar{d}_k [\mathbf{X}_k(n) - \mathbf{X}_{k-1}(N)] \quad k = 2, 3, ... \tag{16}$$

$$\hat{\mathbf{X}}_1(n) = \bar{d}_1 \mathbf{X}_1(n) \tag{17}$$

where $\hat{\mathbf{X}}_k(n)$ is the vector which includes the scaled position and velocity of the camera poses contained in section k.

IV. EXPERIMENTS

We use a catadioptric omnidirectional camera with a resolution of 1024x768 and a frame rate of 15 fps. This camera is mounted on a helmet carried by a human operator. We performed two experiments. For the first experiment we acquired three image sequences walking along the same path of 230 meters with three different step frequencies. The ground truth was obtained with a metronome with 0.01 seconds of resolution. It was set up to 0.70, 0.60 and 0.50 seconds per beat for each sequence, which translates in step frequencies of 1.43 Hz, 1.67 Hz and 2 Hz, respectively. The experiment is divided in two parts. Firstly, we test the accuracy of spectral analysis for step frequency estimation with different setups. Secondly, we evaluate our approach to scale the trajectories provided by the SLAM algorithm.

For the second experiment we acquired a sequence in an indoor environment without metronome to evaluate the scaling of the trajectory under a normal gait condition.

The initial unscaled trajectory and map corresponding to each image sequence are constructed with the SLAM approach presented in [9].

A. Spectral analysis for step frequency estimation

First, we evaluate the feasability of using spectral analysis to measure the step frequency. As stated in III-A, visual odometry is divided in sections of N camera poses and the DFT is carried out on each section.

To compute the DFT we use the FFTW (Fast Fourier Transform West) C library [6]. We compare different section dimensions of $N = 100$ and $N = 200$. As the routines of this library perform faster when the length of the data sequence is a power of 2, data sequences are padded with zeros to a length of N_p to fill this condition. A greater padding involves an increased resolution of the spectrogram. Thus, we compare two zero-padding instances ZP1 and ZP2. ZP1 corresponds to a padding being N_p the power of 2 closest to N. ZP2 corresponds to a padding with $N_p = 1024$.

In Fig. 4 we show the results of the measured step frequency of the three trajectories with different DFT setups.

It can be observed that taking $N = 200$ provides more accurate estimations and less outliers. This is done at the expense of increasing the interval between two consecutive estimations. It is also shown that a greater zero-padding does not provide any improvement in accuracy. Thus we select a setup of $N = 200$ data points and the ZP1 padding instance to compute the DFT for spectral analysis.

B. Scaling of trajectories

In this part we scale the visual odometry from each sequence of images using the approach described in Sec. III. The number of particles used in the particle filter is $P = 5000$. Standard deviations of the gaussians used to model probability distributions have been adjusted.

The standard deviation of the distribution modelling the initial logarithmic scale factor is set to $\sigma_0 = 1$. This setup allows us to consider an initial uncertainty interval for the scale factor between 10^{-2} and 10^2 with a 95% of confidence. The standard deviation of the gaussian noise modeling scale drift was fixed to $\sigma_{drift} = 0.05$. The standard deviation of the walking measurement estimation was set up to a high value of 0.2 m/s to avoid a particle depletion of the particle filter due to a low sensor noise.

After adjusting the priors we tested our approach on the three SLAM trajectories walked at different step frequencies. Fig. 5 shows the final reconstruction of the trajectory compared to the Ground Truth over a satellite view from Google Maps. It can be observed the great improvement respect to the raw visual odometry estimation from the SLAM algorithm. Notice also that our approach provides a better estimation than applying a uniform scale factor. The reason is that the dynamic estimation of the scale factor every N frames (Fig. 6) corrects the scale drift of the raw visual odometry.

(a)
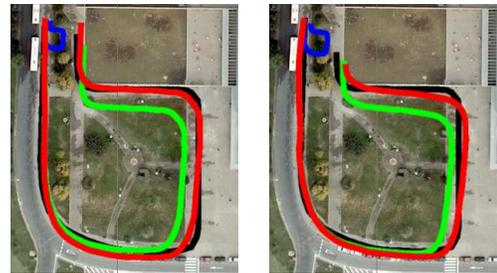
(b)   (c)

Fig. 5: Visual odometry estimations using different approaches on the three trajectories walked at different step frequencies of about 1.43 Hz (a), 1.67 Hz (b) and 2 Hz (c).

Fig. 6: Scale factor estimation corresponding to the trajectory taken at a step frequency of 1.43 Hz. Vertical lines represent the uncertainty with a confidence of 95%. Variation of the scale along the trajectory implies that scale drift is observed.
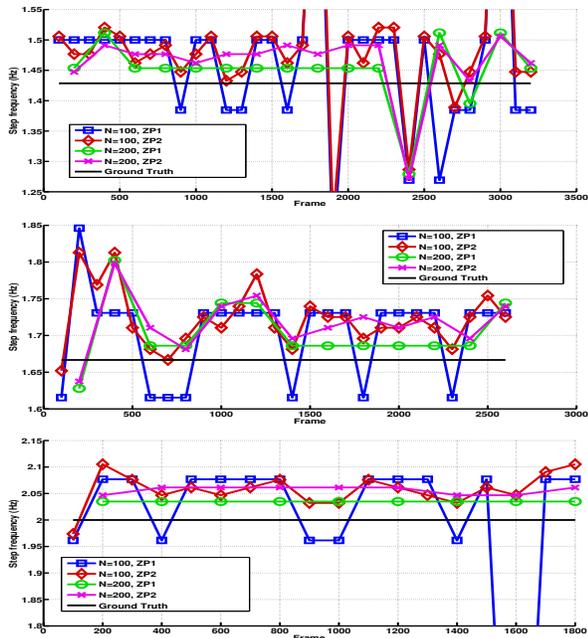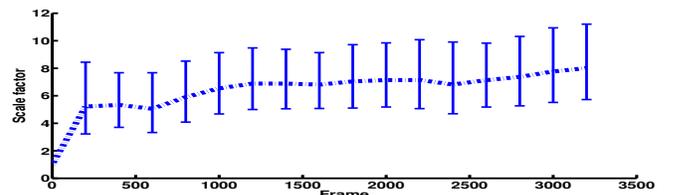

Fig. 4: Spectral analysis along the same path at the three step frequencies of 1.43 (top), 1.67 (center) and 2 Hz (bottom) with different setups for the computation of the DFT.
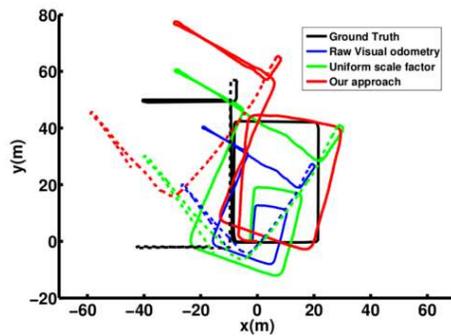
Fig. 7: Visual odometry estimation in an indoor environment with normal gait. Solid line indicates first floor. Dashed line, second floor.
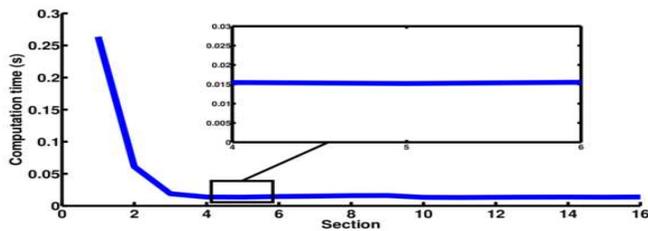

Fig. 8: Computation time used to perform our approach for the different sections.

In the last experiment we test our approach in an indoor environment, with normal gait not set by a metronome [16]. Fig. 7, shows that although turns are not quite well estimated by the SLAM algorithm, posterior scaling with our approach provides a good estimation of the scale of the Ground Truth trajectory, being able to cope with a considerable scale drift present in the raw visual odometry estimation.

The implementation of our approach has been done in MATLAB and it has been performed offline on the estimates provided by a real time monoSLAM application. In Fig. 8 we show the computation time to perform the spectral analysis and the scaling of each section of the trajectory. It can be observed that, after an initial computation time of 0.3 seconds for the first section, it converges to a time of 0.015 seconds. Thus, our approach is able to be used in a real time application as a parallel process, running within the time that takes to fill the data sequence vector with the N elements for the next spectral analysis. As a consequence, there exists a delay in the update of the scaled visual odometry.

V. CONCLUSIONS

In this paper we have presented a new approach to estimate the absolute scale of a visual odometry system using a head-mounted omnidirectional camera. The visual odometry is provided by a monocular SLAM application. The experimental results show an accurate dynamic estimation of the absolute scale which is robust to scale drift. It is shown also the feasability of our method to be implemented in a real time SLAM framework. In future work we will consider the real time implementation as well as the design of active estrategies to cope with sudden variations of the walking speed.

REFERENCES

[1] J. Civera, A. J. Davison, and J. M. M. Montiel. Inverse depth parametrization for monocular SLAM. *IEEE Transactions on Robotics*, 24(5):932–945, 2008.
[2] J. Civera, O. G. Grasa, A. J. Davison, and J. M. M. Montiel. 1-Point RANSAC for EKF Filtering: application to real-time structure from motion and visual odometry. *Journal of Field Robotics*, 27(5):609–631, 2010.
[3] P. Corke, D. Strelow, and S. Singh. Omnidirectional visual odometry for a planetary rover. In *IROS*, 2004.
[4] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse. MonoSLAM: Real-time single camera slam. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 29:1052–1067, 2007.
[5] A. Eudes, M. Lhuillier, S. Naudet-Collette, and M. Dhome. Fast odometry integration in local bundle adjustment-based visual slam. In *ICPR*, 2010.
[6] M. Frigo and S. G. Johnson. The design and implementation of FFTW3. *Proceedings of the IEEE*, 93(2):216–231, 2005. Special issue on "Program Generation, Optimization, and Platform Adaptation".
[7] N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *Radar and Signal Processing, IEE Proceedings F*, 140(2):107–113, 1993.
[8] D. W. Grieve. Gait patterns and the speed of walking. *Biomedical Engineering*, 3(March):119–122, 1968.
[9] D. Gutierrez, A. Rituerto, J. M. M. Montiel, and J. J. Guerrero. Adapting a real-time monocular visual SLAM from conventional to omnidirectional cameras. In *11th OMNIVIS, held with 13th ICCV*, 2011.
[10] E. Hirasaki, S. T. Moore, T. Raphan, and B. Cohen. Effects of walking velocity on vertical head and body movements during locomotion. *Experimental Brain Research*, 127(2):117–130, 1999.
[11] A. D. Kuo. A simple model of bipedal walking predicts the preferred speed-step length relationship. *Journal of Biomechanical Engineering*, 123:264–269, 2001.
[12] P. Lothe, S. Bourgeois, E. Royer, M. Dhome, and S. Naudet-Collette. Real-time vehicle global localisation with a single camera in dense urban areas: Exploitation of coarse 3d city models. In *CVPR*, 2010.
[13] T. Lupton and S. Sukkarieh. Removing scale biases and ambiguity from 6DoF monocular SLAM using inertial. In *ICRA*, 2008.
[14] C. Mei. *Laser-Augmented Omnidirectional Vision for 3D Localisation and Mapping*. PhD thesis, INRIA Sophia Antipolis, Project-team ARobAS, 2007.
[15] C. Mei, G. Sibley, M. Cummins, P. Newman, and I. Reid. RSLAM: A system for large-scale mapping in constant-time using stereo. *International Journal of Computer Vision*, 94(2):198–214, 2010.
[16] A. C. Murillo, D. Gutiérrez-Gómez, A. Rituerto, L. Puig, and J. J. Guerrero. Wearable omnidirectional vision system for personal localization and guidance. In *2nd Workshop on Egocentric Vision, held with CVPR*, 2012.
[17] D. Nistér, O. Naroditsky, and J. Bergen. Visual odometry for ground vehicle applications. *Journal of Field Robotics*, 23:3–20, 2006.
[18] G. Nützi, S. Weiss, D. Scaramuzza, and R. Siegwart. Fusion of IMU and vision for absolute scale estimation in monocular SLAM. *Journal of Intelligent Robotic Systems*, 61(1-4):287–299, 2010.
[19] L. M. Paz, P. Piniés, J. D. Tardós, and J. Neira. Large scale 6DoF SLAM with stereo-in-hand. *IEEE Transactions on Robotics*, 24(5):946–957, 2008.
[20] A. Rituerto, L. Puig, and J. J. Guerrero. Visual SLAM with an omnidirectional camera. In *ICPR*, 2010.
[21] D. Scaramuzza, F. Fraundorfer, M. Pollefeys, and R. Siegwart. Absolute scale in structure from motion from a single vehicle mounted camera by exploiting nonholonomic constraints. In *ICCV*, 2009.
[22] D. Scaramuzza, F. Fraundorfer, and R. Siegwart. Real-time monocular visual odometry for on- road vehicles with 1-point RANSAC. In *ICRA*, 2009.
[23] H. Strasdat, J. M. M. Montiel, and A. Davison. Scale drift-aware large scale monocular SLAM. In *Robotics: Science and Systems*, 2010.
[24] J.-P. Tardif, Y. Pavlidis, and K. Daniilidis. Monocular visual odometry in urban environments using an omnidirectional camera. In *IROS*, 2008.
[25] M. Zarrugh, F. Todd, and H. Ralston. Optimization of energy expenditure during level walking. *European Journal of Applied Physiology and Occupational Physiology*, 33:293–306, 1974.