

Improving Visual SLAM Accuracy Through Deliberate Camera Oscillations

Mohamed Heshmat*, Mohamed Abdellatif* and Hossam Abbas*[§]

*Mechatronics and Robotics Dept.,

Egypt-Japan University of Science and Technology, Alexandria, Egypt.

Email: {mohamed.heshmat, mohamed.abdellatif}@ejust.edu.eg

[§]Electrical Eng. Dept., Assiut University, Assiut, Egypt.

Email: hossam.abbas@ejust.edu.eg

Abstract—Visual Simultaneous Localization And Mapping, (VSLAM) algorithms exploit the observation of scene naturally-existing distinct features to infer the camera motion and build a map of a static environment. There is an increasing interest towards building efficient VSLAM algorithms mainly from computational perspectives; however, there may be insufficient clues to solve for SLAM parameters efficiently. In this paper, deliberate camera oscillations are superimposed on the camera main motion (robot motion), mostly in a lateral direction to give sufficient physical clues for the solution. Filtering methods exploit correlation to infer the motion parameters, and since oscillation introduces more local changes, it can enhance the estimation by correlation. Simulation results are presented showing the effects of oscillation parameters on visual SLAM performance in different motion scenarios. The results showed significant improvement of accuracy for oscillating camera over the steady camera case, and in several cases errors are reduced to less than half its value. These simulation results can be the basis to design a real experimental system.

Keywords; Visual SLAM, Camera Oscillation, Camera Vibrations.

I. INTRODUCTION

Visual SLAM means the Simultaneous Localization And Mapping using visual information obtained from single [1], [2] or stereo cameras [3], [4]. The SLAM algorithms had gained much popularity due to its importance for mobile robot applications into unstructured environment. Visual SLAM is attractive because it uses the available on-board cameras to complete the SLAM task. The current successful VSLAM implementations use Extended Kalman Filter, (EKF) [5], [6] or Bundle Adjustment, (BA) [7], [8] to solve the problem. There are many differences in computations side, but on the accuracy side, filtering techniques may reach comparable accuracy as that of BA [9]. The general interest in visual SLAM research is towards investigation of the computational aspects of the problem, overlooking the physical aspects which set the bounds for algorithm errors [9], [10]. The accuracy per time cost is a useful measure for VSLAM success for mobile robot applications, and in order to reach the good measure, both physical and computational layers should be considered.

The problem is usually interpreted as follows: given that a camera is attached to a robot and given a set of salient features in the scene; it is required to estimate the robot location and a map for the discovered features [11]. This interpretation forced the researchers to think of the problem

in the computational sense only, while it has both physical and computational aspects. Focusing on computational issues of the SLAM problem helps improving the performance but considering physical aspects can be more effective since it defines the ceiling of the computational improvements. Although SLAM systems are intended to work as measurement systems and mostly will not interact with the robot motion decision making modules, physical clues could be gained from actively superimposing local camera oscillations that do not affect the average motion. This idea stems from two well-known facts. First, the fact that during the human walk the head moves from left to right [12], [13], and this is even more noticeable with aged people [14]. Second, the fact that human eyes oscillate at 90 HZ [15], [16]. Oscillating retina had been used for edge detection [17] and to improve stereo range finding accuracy [18], [19]. Cameras making lateral motions have also been used to help with the correspondence problem [20]. Fusing information from multi cameras was used to add more physical clues to the SLAM system [21].

In this paper, we propose a new acquisition system that uses an oscillating camera to actively increase the estimation clues of camera motion relative to scene features. The oscillation parameters and their effect on the VSLAM performance are investigated through computer simulations. Although camera vibrations may cause the feature extraction to be more difficult, it can be quite beneficial to increase physical clues to extract robot motion and consequently enhance the convergence speed of VSLAM algorithm. It is assumed that features are extracted and presented to the VSLAM algorithm, and this stage will be considered separately since typical methods can be found in the literature [22], [23].

This paper is organized as follows: the next section outlines the basic probabilistic framework for visual SLAM algorithm. Then, in section III, the oscillating camera system design criteria are discussed. Section IV then describes the simulation conditions and the effect of oscillations on the accuracy of VSLAM algorithm. The discussion is presented in section V, and conclusions are finally given in section VI.

II. THEORETICAL FRAMEWORK

In this work, Extended Kalman Filter, (EKF) SLAM is used as the core probabilistic framework. The EKF state vector contains robot and landmarks states, modelled as a Gaussian variable using the mean vector and the covariance matrix of

the state vector. This state vector is maintained using EKF through a prediction/correction loop. The state vector can be described as follows:

$$x = \begin{bmatrix} R \\ M \end{bmatrix} = \begin{bmatrix} R \\ L1 \\ \vdots \\ Ln \end{bmatrix} \quad (1)$$

where R is the robot state and M is the map landmarks state. The covariance matrix is defined as:

$$P = \begin{bmatrix} P_{rr} & P_{rm} \\ P_{mr} & P_{mm} \end{bmatrix} \quad (2)$$

The goal of EKF-SLAM is to keep the state vector and the covariance matrix updated. The robot position represented by X-Y-Z position, and the robot orientation represented by quaternion q construct the robot state vector. The landmarks state vector represents all the landmark positions. Each landmark may be represented using different parameterization [24]. In this work, the Euclidean or anchored homogeneous parameterization is used to represent the landmarks [1], [24].

$$R = [X \ Y \ Z \ q1 \ q2 \ q3 \ q4]^T \quad (3)$$

The EKF prediction step:

$$x \leftarrow f(x, u, n) \quad (4)$$

$$P \leftarrow F_x P F_x^T + F_n N F_n^T \quad (5)$$

where u is the control vector, $f()$ is the generic time update functions, F_x , F_n are the Jacobian matrices and N is the covariance matrix of the perturbation n .

The EKF correction step:

$$z = y - h(x) \quad (6)$$

$$Z = H_x P H_x^T + R \quad (7)$$

$$K = P H_x^T Z^{-1} \quad (8)$$

$$x \leftarrow x + K z \quad (9)$$

$$P \leftarrow P - K Z K^T \quad (10)$$

where y is the noisy measurement, x is the full state, $h()$ is the observation function, H_x is Jacobian matrix, R is the covariance matrix of the measurement noise and K is the Kalman gain. Equations 6 and 7 are the innovation mean and covariance matrix.

III. SYSTEM DESIGN

The convergence rate of depth estimation is very important in visual SLAM systems, especially in the monocular camera case [10]. Depth features can be recovered from the camera motion, but its computation depends on the type of motion. It is well known that the lateral motion is the best motion to compute features depth. On the opposite side, forward and wild turning motions are difficult cases for depth computations [9], [25]. To overcome this problem, deliberate camera oscillations are introduced to add small lateral motion beside the basic motion which augments the clues for depth computing. These camera oscillations are independent of the robot motion and in a direction perpendicular to the robot motion. The major objective is that oscillation should be conditioned to provide the maximum clue gain per oscillation change. There are however, a huge number of both design and operational parameters that can be tested. In this paper we will consider a class of such parameters we believe is enough to start realizing a real system, and through hypothesis generation, it can be incrementally improved. It is then clear that the design should benefit from the simulation results that will be described in the next section. However, we can state facts about the trend of “what to do” and “what not to do”

- Camera oscillation will be beneficial for the convergence of the points in the direction of the optical axis since such points have the least clues to derive their depth estimate from camera forward motion.
- Camera oscillation will be beneficial for the fast convergence of the limited visible time features in the curved motion.
- The increase of the camera amplitude of lateral translational oscillation will increase available clues for all scene points convergence.
- Near points should benefit more from the local camera oscillations compared to distant points.
- Regarding the frequency, the camera frame rate, in our case, is 10 fps, which will limit the frequency to about 5 Hz, to have around 2 frames per oscillation cycle.
- Increasing the robot linear velocity in the direction of motion will by itself increase the geometrical clue of features and then in such case, measurements will need no more motion clues as for the case of low velocities.

Intuitively, rotation around the camera focal point will induce no clues for point geometry in terms of triangulation [26], [27]. This is well known in the literature, and hence we are advised not to rotate about the projection center.

Based on the previous considerations, we made a simulation of a robot going through unknown environment and equipped with a camera oscillates in a lateral direction of the moving robot. The camera displacement X can be described by:

$$X = A \cdot \sin(2\pi f t) \quad (11)$$

where A is the amplitude, f is the frequency and t is the time.

TABLE I. SIMULATION CONDITIONS

Sensor	Image Size	640 x 480 pixel
	Field of View	90°
	Number of Frames	160 frame
	Frame Rate	10 fps
	Intrinsic Parameters	$[u_0 \ v_0 \ a_u \ a_v]$ [320 240 320 320]
Robot	Motion Increment	0.05 m
	Motion Error	$[X \ Y \ \theta]$ [0.005 0.005 0.05]
Oscillation Parameters	Amplitude	0.05 m
	Frequency	3 Hz
	Sample Time	0.1 sec
General Options	Number of Features	162
	Features for Update	10 features
	Features Initialization	[first frame other] [10 1]

IV. SIMULATION RESULTS

The objectives of the simulation are to answer the following questions:

- How the camera oscillation affects the depth estimate convergence rate for scene features?
- How the oscillation parameters (Amplitude and Frequency) affect the accuracy of visual SLAM?
- How the oscillating camera affects accuracy of robot motion estimation during forward, lateral and curved motion?

We simulate a robot performing different types of motion in an area of 16 m x 16 m. This environment consists of 162 landmarks configured into two layers of landmarks. The same features configuration is used through all experiments to cancel the scene features composition effect [28], [29]. A modified version of 6-DOF EKF-SLAM TOOLBOX [30] is used in this simulation. Table I summarizes the simulation conditions.

A. Effect on Depth estimate Convergence Rate

Two experiments have been made for the robot moving forward for 8 m, exploring scene features. In the first experiment, the normal steady camera was used, while in the second, oscillating camera was used. Figures 1, 2 show the map of landmarks, the localization errors and uncertainty using a steady and oscillating camera. Localization error refers to error in position and orientation. The position error is represented by the Euclidean distance of the position errors E_{euc} :

$$E_{euc} = \sqrt{e_x^2 + e_y^2} \tag{12}$$

where e_x is the error in x direction and e_y is the error in y direction. In oscillating camera case, a fast convergence rate of the features was observed in comparison with the case with no oscillation, therefore, the error was decreased significantly. In steady camera case, there are some features that did not converge even until the end of the robot motion, as a result the errors increased in that case.

Figure 3, 4 show the depth estimate of features selected from regions A and B, which represent the regions that benefit from the oscillation as defined in Fig. 5. It is observed

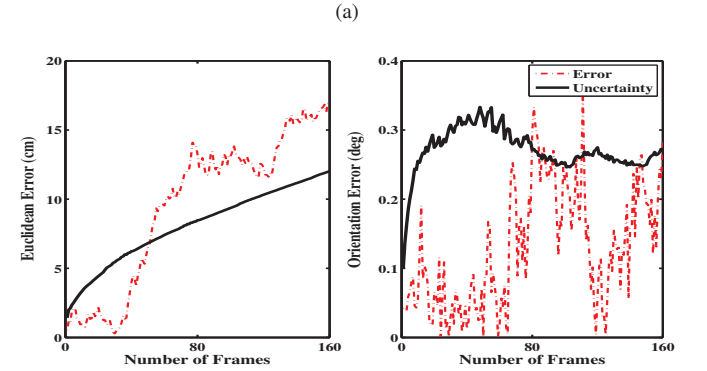
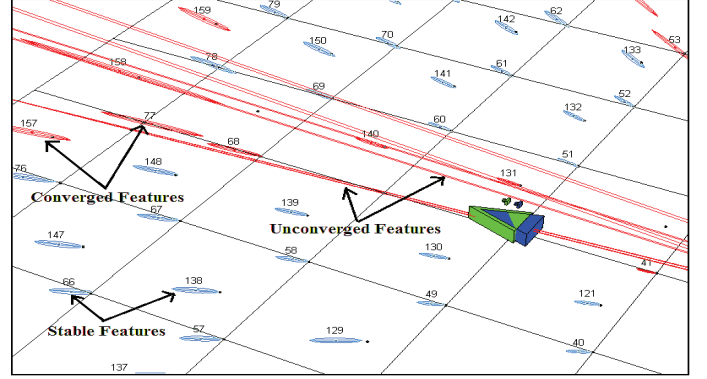


Fig. 1. Steady camera case, (a) Map of landmarks, (b) Localization errors.

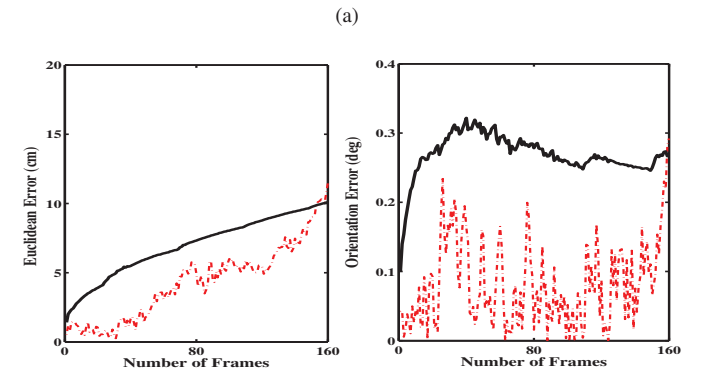
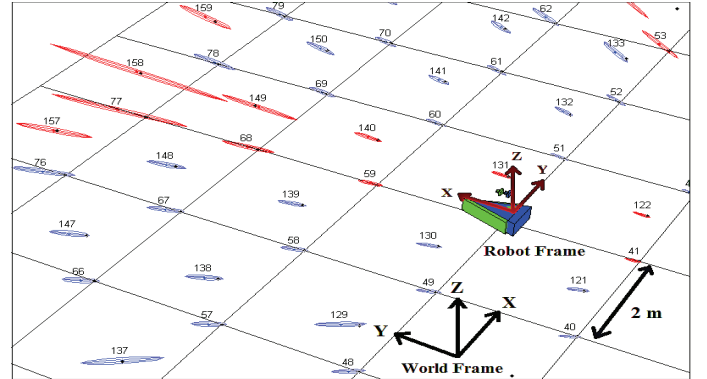


Fig. 2. Oscillating camera case (amplitude = 0.05 m & frequency = 3 Hz), (a) Map of landmarks, (b) Localization errors.

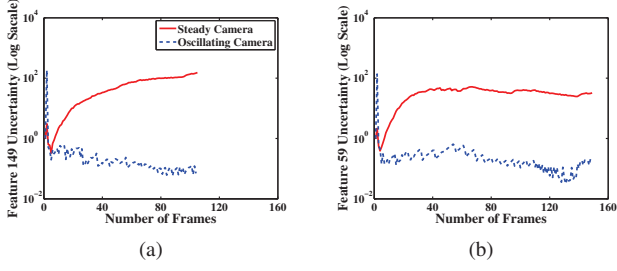


Fig. 3. Depth uncertainty convergence for selected features in region A using steady and oscillating camera (amplitude = 0.05 m & frequency = 3 Hz).

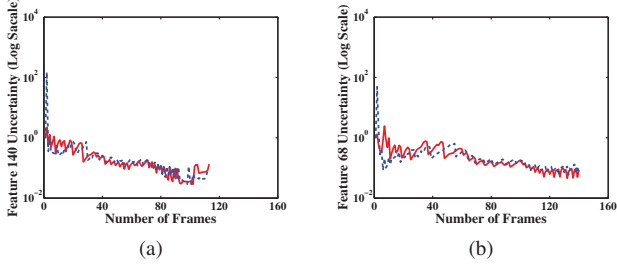


Fig. 4. Depth uncertainty convergence for selected features in region B using steady and oscillating camera (amplitude = 0.05 m & frequency = 3 Hz).

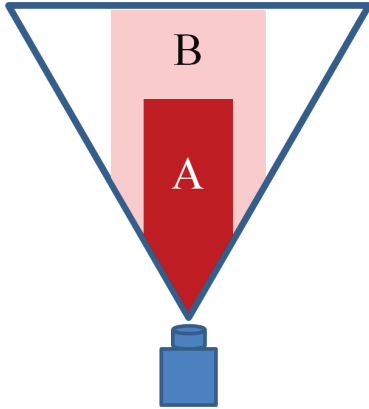


Fig. 5. Schematic diagram showing the regions, that benefit from camera oscillations.

that camera oscillations accelerate the depth convergence rates especially for region A features. This is attributed to the new clues that were added to the system.

B. Effect on the Localization Accuracy of Visual SLAM

1) *Amplitude Effect:* In this experiment, the camera oscillates with amplitudes varying from 1 : 10 cm for different velocities (0.25, 0.5, 1) m/Sec. The experiments are performed in the forward and the curved motion. The results are shown in Fig. 6 and Fig. 7, where it is observed that:

- Increasing the amplitude decreases the position error significantly and in particular for low robot velocities.
- The error change in the case of high velocity motion is small, because the motion in the forward direction with high velocity gives strong clues that reduce error in this case and the lateral displacement adds little

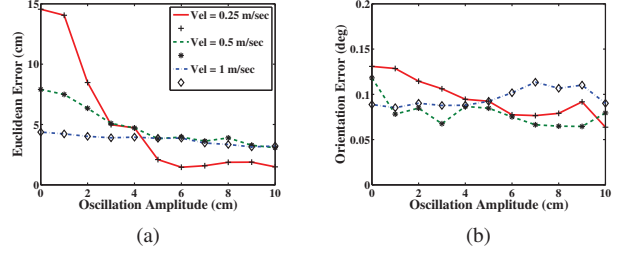


Fig. 6. Effect of oscillation amplitude in the forward motion and using frequency = 3 Hz in case of, (a) Euclidean error of position, (b) Orientation error

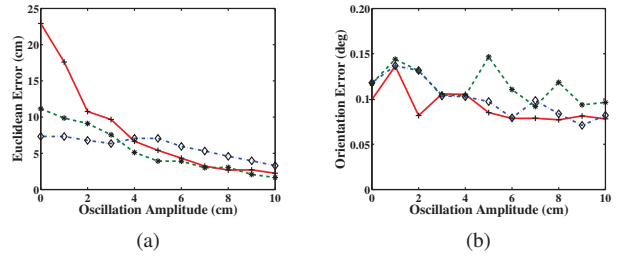


Fig. 7. Effect of oscillation amplitude in the curved motion and using frequency = 3 Hz in case of, (a) Euclidean error of position, (b) Orientation error

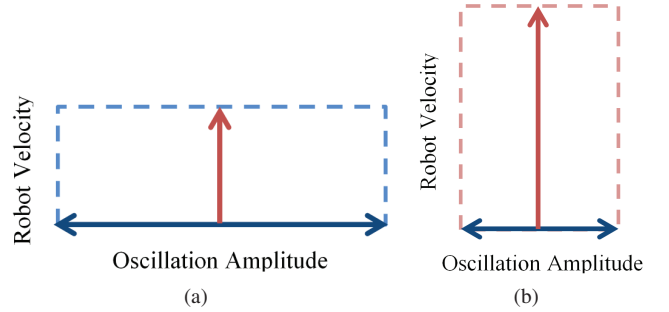


Fig. 8. Schematic diagram showing oscillation amplitude compared to main motion displacement at, (a) Low robot velocity, (b) High robot velocity

clues that are of small effect compared to the forward motion clues. Figure 8 shows a schematic diagram showing oscillation amplitude effect in case of low and high velocity.

- For the oscillation amplitudes from 5 : 10 cm the error is almost the same, which means that there is no need to increase the amplitude for achieving further error reduction.
- For the orientation error, by increasing the amplitude the error decreases in the case of low and medium velocity, but this is not the case for higher speed where the error increases slightly by increasing the amplitude.

2) *Frequency Effect:* Through these experiments the frequency effect is explored. Oscillation frequency takes values from 1 : 4 Hz by 0.5 Hz steps. The simulations are done in two motion types, forward motion and curved motion. It is well known that frequency values are bounded by the sampling time which in our case is $T_{sample} = 0.1$ Sec. From Nyquist-

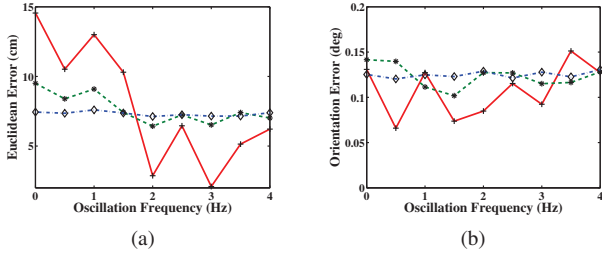


Fig. 9. Effect of oscillation frequency in the forward motion and using amplitude = 0.05 m in case of, (a) Euclidean error of position, (b) Orientation error

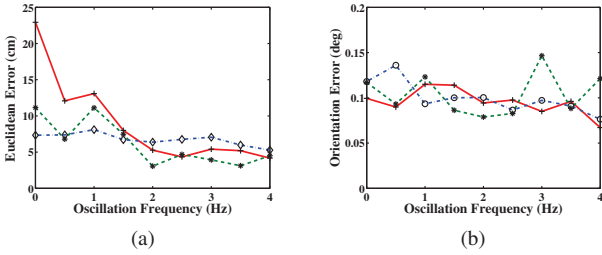


Fig. 10. Effect of oscillation frequency in the curved motion and using amplitude = 0.05 m in case of, (a) Euclidean error of position, (b) Orientation error

Shannon sampling theorem:

$$T_{sample} = \frac{1}{2f} \quad (13)$$

therefore the maximum theoretical frequency is $f = 5$ Hz, but it has been observed that it is $f = 4$ Hz, and this can be attributed to the motion model noise.

The results are shown in Fig. 9 and 10, from which we observe the following:

- Oscillation frequency has pronounced effect in low velocity case rather than high velocity, because the cycle's number increases with frequency, but the samples in each cycle will decrease and this makes the lateral motion dominant based on the samples place.
- The frequency should be selected based on robot velocity, so we can choose a reasonable frequency based on the velocity.

For the optimal parameters of oscillation corresponding to the least localization error, we choose oscillation amplitude = 0.05 m and oscillation frequency = 3 Hz, for our particular system.

C. Effect of different robot motion types

Experiments were made for forward, lateral, and curved robot motion. Table II shows the RMSE errors for localization and orientation errors for each type of motion. The curved motion and the forward motion enjoyed a significant reduction in the localization error, while minor effect is observed for the lateral motion. Figure 11 shows the curved robot path, as an example, using steady and oscillating camera against ground truth.

TABLE II. THE EUCLIDEAN ERROR OF POSITION AND ORIENTATION ERROR FOR DIFFERENT MOTION TYPES

Type of Motion	Error Types	Steady Camera	Oscillating Camera
Forward Motion	Euclidean Error	11.0293 cm	5.0937 cm
	Orientation Error	0.1553 deg	0.1026 deg
Curved Motion	Euclidean Error	11.1495 cm	3.9308 cm
	Orientation Error	0.1170 deg	0.1465 deg
Lateral Motion	Euclidean Error	3.6709 cm	3.6163 cm
	Orientation Error	0.2119 deg	0.1766 deg

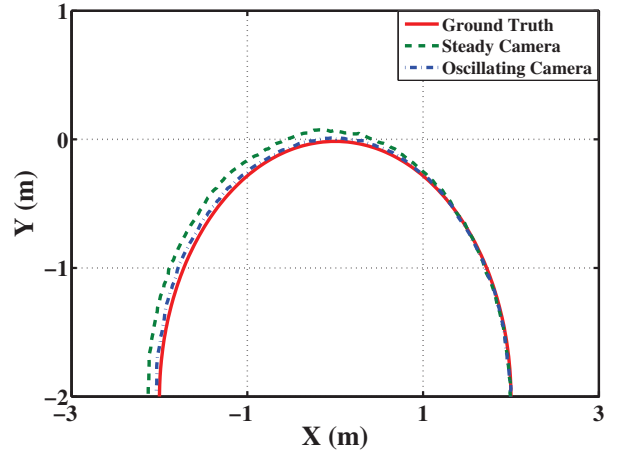


Fig. 11. Robot path using steady and oscillating camera (amplitude = 0.05 m & frequency = 3 Hz) against ground truth in case of curved motion

V. DISCUSSION

The idea of this paper has been built on the facts of the head motion during the human walk and the biological retina oscillations, which were believed to have a functional benefit. Despite, the need to have a robust method to detect distinctive feature, from vibrating blurred retina, there is a significant increase in the physical clues that can help the visual SLAM methods to recover feature depth faster and more accurately. Our simulation experiments confirmed the effectiveness of the camera oscillations in particular for curved routes where features are only visible for a limited time. Benefits were also confirmed for the difficult forward motion and this is useful in practice, since it is common and expected during the course of robot motion.

The goal of these synthetic experiments is to validate the idea. We realize that it may not be practical to shake the camera for 5 cm amplitude and this will be considered in the future. The amplitude was shown through experiments to be the most significant factor to benefit from camera oscillation. Issues like resolution, focal length and possible use of wide field lenses will be considered in the future. However, it was also shown that only the position errors were reduced, while orientation errors did not benefit from our oscillation pattern.

VI. CONCLUSIONS

This paper introduced a novel acquisition system for visual SLAM system in which the camera oscillates to gain more

clues about the scene feature depth and hence improve robot localization. This is done in contrast to passive cameras watching scene features and employs whatever motion clues that may be passively available to solve the SLAM problem. Camera oscillations, and in particular in the lateral direction of heading orientation, improved the feature depth estimation very fast and hence increased the SLAM accuracy significantly. Various oscillation parameters of the camera together with motion variables of the mobile robot were investigated and the following conclusions can be made:

- Oscillations increase feature depth convergence rate for some near features while it has little or no effect elsewhere.
- Oscillations improved the self-localization errors in terms of positioning, and amplitude has the most significant effect while frequency has only slight effect at low robot velocities.
- The improvement in robot positioning was high for curved and forward robot motion, while it has a slight effect on lateral motion which already enjoys low errors due to rich clues generated by its motion.
- Orientation errors were only slightly affected by the oscillation.

The results will help to design a full experimental system and to tune its performance.

ACKNOWLEDGMENT

The first author is supported by a scholarship from the Ministry of Higher Education, Government of Egypt which is gratefully acknowledged.

REFERENCES

- [1] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "MonoSLAM: Real-time single camera SLAM," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1052–1067, 2007.
- [2] E. Eade and T. Drummond, "Scalable monocular SLAM," in *Computer Vision and Pattern Recognition, CVPR*, vol. 1, 2006, pp. 469–476.
- [3] J. Diebel, K. Reuterswärd, S. Thrun, J. Davis, and R. Gupta, "Simultaneous localization and mapping with active stereo vision," in *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, vol. 4, 2004, pp. 3436–3443.
- [4] 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*, vol. 94, no. 2, pp. 198–214, 2011.
- [5] E. Eade and T. Drummond, "Monocular SLAM as a graph of coalesced observations," in *IEEE 11th International Conference on Computer Vision, ICCV*, 2007, pp. 1–8.
- [6] J. Civera, O. G. Grasa, A. J. Davison, and J. Montiel, "1-point RANSAC for EKF filtering: Application to real-time structure from motion and visual odometry," *Journal of Field Robotics*, vol. 27, no. 5, pp. 609–631, 2010.
- [7] G. Klein and D. Murray, "Parallel tracking and mapping on a camera phone," in *8th IEEE International Symposium on Mixed and Augmented Reality, ISMAR*, 2009, pp. 83–86.
- [8] K. Konolige and M. Agrawal, "FrameSLAM: From bundle adjustment to real-time visual mapping," *IEEE Transactions on Robotics*, vol. 24, no. 5, pp. 1066–1077, 2008.
- [9] H. Strasdat, J. Montiel, and A. J. Davison, "Visual SLAM: Why filter?" *Image and Vision Computing*, vol. 30, pp. 65–77, 2012.
- [10] S. Huang and G. Dissanayake, "Convergence and consistency analysis for extended Kalman filter based SLAM," *IEEE Transactions on Robotics*, vol. 23, no. 5, pp. 1036–1049, 2007.
- [11] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping (SLAM): part I," *IEEE Robotics & Automation Magazine*, vol. 13, no. 2, pp. 99–110, 2006.
- [12] J. D. McDonald, A. T. Bahill, and M. B. Friedman, "An adaptive control model for human head and eye movements while walking," *Systems, Man and Cybernetics, IEEE Transactions on*, no. 2, pp. 167–174, 1983.
- [13] W. Einhäuser, F. Schumann, S. Bardins, K. Bartl, G. Böning, E. Schneider, and P. König, "Human eye-head co-ordination in natural exploration," *Network: Computation in Neural Systems*, vol. 18, no. 3, pp. 267–297, 2007.
- [14] J. A. Honaker and N. T. Shepard, "Age effect on the gaze stabilization test," *Journal of Vestibular Research*, vol. 20, no. 5, pp. 357–362, 2010.
- [15] A. L. Yarbus and L. A. Riggs, *Eye movements and vision*. Plenum press, 1967, vol. 2.
- [16] B. W. Tatler, N. J. Wade, H. Kwan, J. M. Findlay, and B. M. Velichkovsky, "Yarbus, eye movements, and vision," *I-Perception*, vol. 1, no. 1, pp. 7–27, 2010.
- [17] M. O. Hongler, Y. L. de Meneses, A. Beyeler, and J. Jacot, "The resonant retina: exploiting vibration noise to optimally detect edges in an image," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 1051–1062, 2003.
- [18] D. Geiger and A. Yuille, "Stereo and eye movement," *Biological cybernetics*, vol. 62, no. 2, pp. 117–128, 1989.
- [19] N. Maru, A. Nishikawa, F. Miyazaki, and S. Arimoto, "Active detection of binocular disparities," in *IEEE/RSJ International Workshop on Intelligence for Mechanical Systems, Proceedings IROS*, 1991, pp. 263–268.
- [20] M. Okutomi and T. Kanade, "A multiple-baseline stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 4, pp. 353–363, 1993.
- [21] J. Sola, A. Monin, M. Devy, and T. Vidal-Calleja, "Fusing monocular information in multicamera SLAM," *IEEE Transactions on Robotics*, vol. 24, no. 5, pp. 958–968, 2008.
- [22] L. A. Vese and S. J. Osher, "Modeling textures with total variation minimization and oscillating patterns in image processing," *Journal of Scientific Computing*, vol. 19, no. 1-3, pp. 553–572, 2003.
- [23] Q. Tian, N. Sebe, M. S. Lew, E. Loupias, and T. S. Huang, "Image retrieval using wavelet-based salient points," *Journal of Electronic Imaging*, vol. 10, no. 4, pp. 835–849, 2001.
- [24] J. Sola, T. Vidal-Calleja, J. Civera, and J. M. M. Montiel, "Impact of landmark parametrization on monocular EKF-SLAM with points and lines," *International journal of computer vision*, vol. 97, no. 3, pp. 339–368, 2012.
- [25] M. Hasan and M. Abdellatif, "Experimental verification of direct depth computing technique for monocular visual SLAM systems," in *International Conference on Innovative Engineering Systems*, 2012, pp. 142–147.
- [26] L. Wang, S. B. Kang, H.-Y. Shum, and G. Xu, "Error analysis of pure rotation-based self-calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 2, pp. 275–280, 2004.
- [27] P. H. Torr, A. W. Fitzgibbon, and A. Zisserman, "The problem of degeneracy in structure and motion recovery from uncalibrated image sequences," *International Journal of Computer Vision*, vol. 32, no. 1, pp. 27–44, 1999.
- [28] M. Heshmat and M. Abdellatif, "The effect of feature composition on the localization accuracy of visual SLAM systems," in *the International Conference on Computer Vision Theory and Applications, VISAPP*, 2012, pp. 419–424.
- [29] M. Abdellatif, "Monocular visual SLAM: Improvements from the object side," in *The 13th Mechatronics Forum International Conference*, 2012, pp. 393–398.
- [30] J. Sola, D. Marquez, J. Codol, and T. Vidal-Calleja, "EKF-SLAM toolbox for matlab," [Online]. Available: <http://www.joansola.eu/JoanSola/eng/toolbox.html>, Accessed: April 2013.