Markerless Tracking Using Polar Correlation of Camera Optical Flow

Category: Research

ABSTRACT

We present a novel, real-time, markerless vision-based tracking system, employing a rigid orthogonal configuration of two pairs of opposing cameras. Our system uses optical flow over sparse features to overcome the limitation of vision-based systems that require markers or a pre-loaded model of the physical environment. We show how opposing cameras enable cancellation of common components of optical flow leading to an efficient tracking algorithm. Experiments comparing our device with an electromagnetic tracker show that its average tracking accuracy is 80% over 185 frames, and it is able to track large range motions even in outdoor settings.

Keywords: Optical Flow, Polar Correlation, Multi Camera, Markerless

Index Terms: I.4.9 [Image Processing and Computer Vision]: Scene Analysis—Motion, Tracking

1 INTRODUCTION

Motion tracking is a critical aspect of many virtual and augmented reality applications and there is a wide variety of different tracking technologies and approaches. One approach that has gained in popularity in recent years is vision-based tracking. Cameras are inexpensive, have large range, and provide raw images which are rich in information. Vision-based tracking systems can be classified into two genres: Inside-looking-out, in which the optical sensor is placed on the moving object and the scene is stationary, example [11], and Outside-looking-in, in which the optical sensor is stationary and observes the moving object, example [5].

Traditionally, vision-based tracking requires markers as reference points or a pre-loaded model of the physical environment. Unfortunately, markers can clutter the physical environment and preloaded models can be time-consuming to create. In this paper, we present a real-time, markerless, vision-based tracking system employing a rigid orthogonal configuration of two pairs of opposing cameras. We show how opposing cameras enable cancellation of common components of optical flow, which leads to an efficient tracking algorithm. Our prototype is low cost, requires no setup, obtains high accuracy and has large range span.

2 RELATED WORK

There are fewer inside-looking-out systems for user interface applications, even though there are many algorithms for structure from motion (SFM) and simultaneous localization and mapping (SLAM), both used in robotics or general purpose vision applications. SLAM methods include single camera [4] and multiple camera [6] approaches. In [3, 7], SFM and egomotion algorithms are developed for multi camera navigation tasks.

Another approach for computing egomotion using optical flow is described in [10], but the experimental results show that the technique works for only very small motions, which is not practical for user interface application. A multi-camera 6 DOF pose tracking algorithm is presented in [9], but tested only on synthetic data. In [11], LED panels in a room ceiling are used to provide markers for tracking; this cumbersome setup limits its applicability as a convenient tracking system. For a vision based controller to be adopted in user interface applications it must function in real-time, have long range, be accurate, be convenient to use and be low cost. Our approach works in real time and does not require markers, making it a practical tracking approach for virtual and augmented reality applications.

3 TRACKING ALGORITHM

The schematic design of our device is shown in Figure 1(a). Our device is designed as a multi camera rig with four cameras C_k (for k = 1 to 4), placed as a rigid orthogonal configuration of two pairs of opposing cameras. Figure 1(b) shows the position s_k and orientation m_k of each camera with respect the rig coordinate system. Figure 1(c) shows a prototype of the device that we built using off-the-shelf webcams, for testing purposes. This section presents our tracking algorithm to get position information of the device at each time instant.

3.1 Direction of Translation (DOT)

3.1.1 Instantaneous Model

Given two successive images of a scene, the motion of each pixel in the first image to the second image is defined as a vector $[u,\ v]^T$, called Optical Flow, where u and v are velocity components in x and y direction respectively. Using the instantaneous model of optical flow [1], for a camera C_k the optical flow vector $[u^k,\ v^k]^T$ at point $P\ (x,\ y)$ can be written as:

$$u^k = \frac{-t_x^k + x t_z^k}{Z} + \omega_x^k xy - \omega_y^k (x^2+1) + \omega_z^k y, \tag{1}$$

$$v^k = \frac{-t_y^k + y t_z^k}{Z} + \omega_x^k (y^2+1) - \omega_y^k xy - \omega_z^k x, \tag{2}$$

where $t^k = [t_x^k,\ t_y^k,\ t_z^k]^T$ is the translation and $\omega^k = [\omega_x^k,\ \omega_y^k,\ \omega_z^k]^T$ is the angular velocity of camera C_k and Z is the z component (depth) of the 3D point corresponding to the image point $P\ (x,\ y)$.

3.1.2 Shifted Cameras

Following [10], for a camera shifted from the origin:

$$t^k = m_k[(\omega \times s_k) + T], \quad \omega^k = m_k \omega, \tag{3}$$

where t^k is the translation and ω^k is the angular velocity of camera C_k, placed at position s_k with orientation m_k, and $T = [T_x,\ T_y,\ T_z]^T$ is the translation and $\omega = [\omega_x,\ \omega_y,\ \omega_z]^T$ is the angular velocity of the rig.

3.1.3 Optical Flow in Each Camera

Substituting values of position and orientation for camera 1 in equation (3), we get:

$$t^1 = \begin{bmatrix} \omega_y + T_x \\ -\omega_x + T_y \\ T_z \end{bmatrix}, \quad \omega^1 = \begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix}. \tag{4}$$

Substituting equation (4) in equations (1) and (2), we get:

$$u^1 = \frac{-\omega_y - T_x + x T_z}{Z} + \omega_x xy - \omega_y (x^2+1) + \omega_z y, \tag{5}$$

$$v^1 = \frac{\omega_x - T_y + y T_z}{Z} + \omega_x (y^2+1) - \omega_y xy - \omega_z x. \tag{6}$$

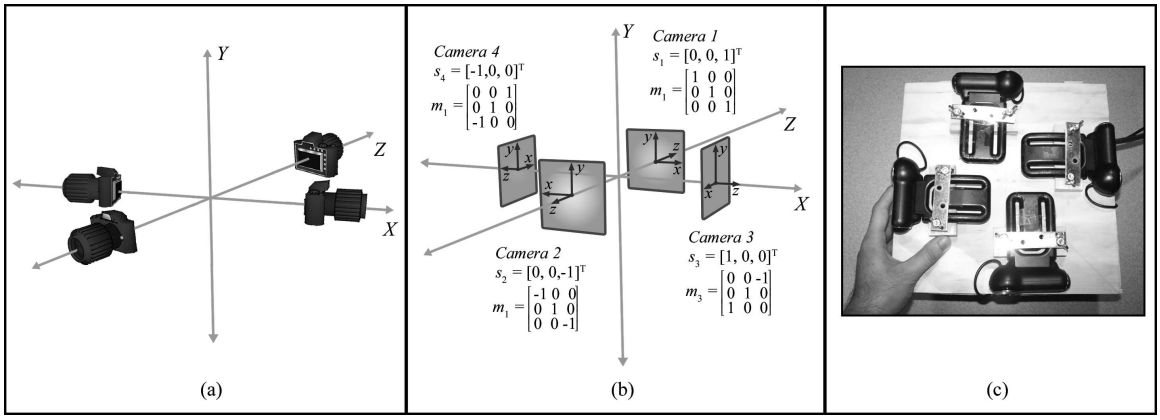

Figure 1: (a) Schematic diagram of the rig, (b) Position and orientation of each camera in the rig, (c) Prototype of the device

Equations (5) and (6) represent the optical flow in camera 1 in terms of the rig motion parameters T and ω. Similarly equations for camera 2, 3 and 4 can also be obtained.

3.1.4 Polar Correlation

Consider four symmetric points of the form $Q^0(x,\ y)$, $Q^1(-x,\ y)$, $Q^2(-x,\ -y)$ and $Q^3(x,\ -y)$. Let the flow vector at these symmetric points for camera C_k be $[u^k_{Q^i},\ {}^k_{Q^i}]^T$ (for i = 0 to 3). The equations for flow vectors at these symmetric points in camera 1 can be obtained by substituting the coordinates of these points in terms of x and y in equations (5) and (6) for camera 1. The equations for optical flow at point Q^0 in camera 1 are:

$$u^1_{Q^0} = \frac{-\omega_y - T_x + xT_z}{Z} + \omega_x xy - \omega_y(x^2+1) + \omega_z y, \tag{7}$$

$$v^1_{Q^0} = \frac{\omega_x - T_y + yT_z}{Z} + \omega_x(y^2+1) - \omega_y xy - \omega_z x. \tag{8}$$

Similarly equations for all the four cameras at these four symmetric points Q^0 to Q^3 can be obtained. Next, we compute a quantity $[L^k_x,\ L^k_y]$ for camera C_k as:

$$L^k_x = \frac{\Sigma^3_{i=0} u^k_{Q^i}}{4}, \quad L^k_y = \frac{\Sigma^3_{i=0} v^k_{Q^i}}{4}. \tag{9}$$

Next we compute a quantity $[G_x,\ G_y,\ G_z]$ as:

$$G_x = \frac{-L^1_x + L^2_x}{2},\ G_y = \frac{-L^1_y - L^2_y - L^3_y - L^4_y}{4},\ G_z = \frac{L^3_x - L^4_x}{2}. \tag{10}$$

By substituting equation (9) for all the four cameras in equation (10) we get:

$$G_x = T_x/Z, \quad G_y = T_y/Z, \quad G_z = T_z/Z. \tag{11}$$

$[G_x,\ G_y,\ G_z]$ is the scaled version of translation $T = [T_x,\ T_y,\ T_z]^T$ of the rig. Next we normalize $[G_x,\ G_y,\ G_z]$ to get direction of translation of the rig. The computation of $[G_x,\ G_y,\ G_z]$ cancels all the rotation terms and we are left with only translation terms. This is the concept of Polar Correlation, which says that opposing cameras have common component of optical flow, which we show can be canceled out to get the direction of translation of the rig.

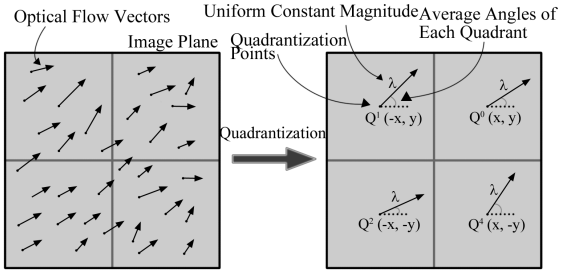

Figure 2: Quadrantization Process

3.1.5 Quadrantization

After computing optical flow in each camera, flow vectors from each frame are passed through the Quadrantization step to get an estimate of optical flow at symmetric points $Q^0(x,\ y)$, $Q^1(-x,\ y)$, $Q^2(-x,\ -y)$ and $Q^3(x,\ -y)$ to use polar correlation. As shown in Figure 2 each frame is divided into 4 quadrants. The center points of each quadrant are called Quadrantization Points Q^i_k (for i = 0 to 3) for camera C_k. Each quadrantization point is associated with a vector with some uniform constant magnitude λ and angle as the average of all flow vectors' angles in that quadrant.

3.2 Angular Velocity

3.2.1 Obtain FOE

When the camera translates in any direction, the translational flow vectors meet at the focus of expansion, and this point might not necessarily lie on the image plane. We show how using the direction of translation yields the focus of expansion. If we consider a sphere surrounding the device, the DOT will intersect the sphere at the FOE. We approximate the sphere by considering a cube centered at the origin and its four faces coinciding with the image planes of the four cameras, as shown in Figure 3. This is a reasonable approximation because of the following reasoning: the field of view of each camera is small (around 30°) and the distance between the rig center and image plane of the shifted cameras is large compared to the practical dimensions of each image capture surface. Thus to find the FOE we find the point where the DOT intersects with the cube surrounding the device. We project this computed FOE on the side faces of the cube to obtain the Local Focus of Expansion (LFOE), which acts as the FOE for the camera whose image plane lies on that face of the cube. Having this local focus of expansion enables us to now compute the angular velocity of the rig.

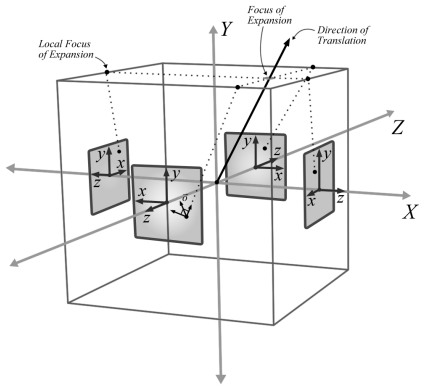

Figure 3: Cube surrounding the rig to find FOE and LFOE

3.2.2 Obtain Angular Velocity

Next, the angular velocity $\omega = [\omega_x,\ \omega_y,\ \omega_z]^T$ of the rig is obtained by using the computed optical flow and the LFOE. After obtaining the LFOE we can obtain the angular velocity using the approach of [8], which shows that the translational component of optical flow at point P always lies on the line connecting the FOE and the point P, and therefore, the component of optical flow perpendicular to the line connecting FOE and the point P has projection of only the rotational component of optical flow. For cameras shifted from the rig center optical flow also has a component of translation due to the rotation of rig. For experimental purposes with small rotation, this component of optical flow has small contribution and therefore can be ignored.

3.3 Tracking 3D Points

To track the 3D position of the device we use the calculated direction of translation and angular velocity. Using the DOT and assuming a constant magnitude of translation we can get the translation vector. After normalizing the angular velocity and multiplying it with a constant magnitude we can obtain the Euler angles of rotation. Rotation matrix R for every time frame can be obtained from the Euler angles. We assume the initial point to be at (0, 0, 0), and from the computed translation vector and rotation matrix for each time frame, we can calculate the next position of the device using the equation:

$$P' = R * [T + P], \tag{12}$$

where P' is the computed current position and P is the previous computed position of the device.

4 Evaluation and Experiments

To evaluate the accuracy of the device prototype, we compared it to a Polhemus PATRIOT tracker, an electromagnetic (EM) tracker that has position and orientation accuracy of $0.1in$ RMS and 0.75° RMS respectively. The readings from the EM tracker are used as ground truth of the tracked motion. The EM tracker and our device prototype provide measurements in different units. To overcome this issue, the position data from the two devices is normalized using a standard normalization technique [2]. For a given trajectory $S = [s_1, \cdots, s_n]$, Norm(S) is defined as:

$$\left[\left(\frac{s_{1,x}-\mu_x}{\sigma_x}, \frac{s_{1,y}-\mu_y}{\sigma_y}, \frac{s_{1,z}-\mu_z}{\sigma_z}\right), \cdots, \left(\frac{s_{n,x}-\mu_x}{\sigma_x}, \frac{s_{n,y}-\mu_y}{\sigma_y}, \frac{s_{n,z}-\mu_z}{\sigma_z}\right)\right], \tag{13}$$

where $s_i = (s_{i,x},\ s_{i,y},\ s_{i,z})$ is 3D position, μ_x, μ_y and μ_z are the means and σ_x, σ_y and σ_z are the standard deviations values in x, y and z coordinates respectively. This normalization makes the distance between two trajectories to be compared invariant to spatial scaling and shifting. For some motion let the trajectory given by the device and the EM tracker be S_n and E_n respectively, for n sample points. The accuracy of our device compared to the EM tracker is computed using the formulation:

$$A = \left(1 - \frac{\sum_{i=0}^{n} d_i}{n}\right) * 100, \tag{14}$$

where d_i is the Euclidean distance between points s_i and e_i for S_n and E_n respectively, obtained after normalization using equation (13).

4.1 Experiments

All experiments were done on real images. The EM tracker was attached to our device and the trajectories formed by both devices were recorded while making motions. The experiment set consists of small motions (around 2 meters) and large motions (around 20 meters). Note that for large motions, the EM tracker did not have enough range so we show the recorded trajectories of our device. The experiments were done with small amounts of rotation.

4.1.1 Experiment Set 1

The first set of experiments consists of random 3D shapes made in a lab setting by moving the device around in the air. The trajectories formed by the EM tracker and our device are compared using the formulation of equation (14).

4.1.2 Experiment Set 2

The second set of experiments are done on a larger range than the experiments in set 1. The results are compared with the EM tracker, showing how the EM tracker fails when it goes out of range but our device still tracks accurately.

4.1.3 Experiment Set 3

The third set of experiments were done in a hallway and in an outdoor environment, with large range motions to show how the optical device robustly tracks the motion. These open space settings have extreme lighting conditions and sunlight. The EM Tracker fails in such large range scenarios. The specific motions used in this experiment set were rectangles. We chose rectangles in this case to test right angles and how close is the starting point of the rectangle from the ending point. These measures provide a way to evaluate how the tracker is performing in the absence of direct comparisons with the EM tracker.

5 Results and Discussion

Figure 4 shows the trajectories formed by our device in red and the EM Tracker in blue, with total accuracy obtained in each motion instance. Figure 5 shows the change in accuracy over time for these instances. It can be seen that the average accuracy of the our device is around 80%, and it is maintained over 185 frames. The frame rate of our device is $\approx$ 16 Hz, which means that for motions of about 11 seconds the system attains up to 80% accuracy. Figure 6 shows how moving out of the range of the EM tracker cause it to jitter but our device still keeps tracking with good accuracy. Figure 7 shows large range motion instances in the hallway and outdoor settings. It can be seen that the our device tracks with reasonable accuracy, though a drift can be seen. The starting and the ending points do not coincide even though the actual motion was made so that the starting and ending points were approximately the same. However, the drift is small as compared to the total range of the motion and the device is able to track the right angles in the rectangles well.

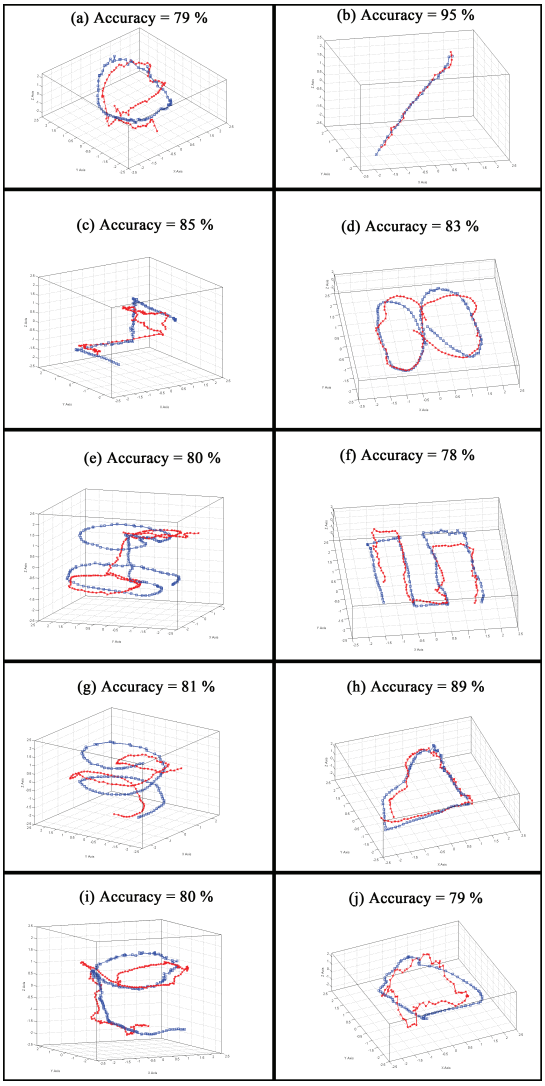

Figure 4: Experiments set 1 with accuracy of the optical device as compared to EM tracker, optical device is shown in red and the EM tracker in blue

6 Conclusion

We have presented a markerless, real time, vision-based tracking system that makes use of the novel concept of Polar Correlation of optical flow. Experiments show that the device has an average accuracy of 80% over 185 frames when compared to an electromagnetic tracker. The prototype of the device is low cost, requires no setup and has a large range span. Future work includes improving the tracking accuracy as well as making the device smaller and wireless.

References

[1] A. R. Bruss and B. K. P. Horn. Passive navigation. *Computer Vision, Graphics, and Image Processing*, 21(1):3–20, 1983.

[2] L. Chen, M. T. Özsu, and V. Oria. Robust and fast similarity search for moving object trajectories. In *In Proceedings of ACM SIGMOD International Conference on Management of Data*, pages 491–502. ACM, 2005.

[3] B. Clipp, J.-H. Kim, J.-M. Frahm, M. Pollefeys, and R. Hartley. Robust 6dof motion estimation for non-overlapping, multi-camera systems. *Applications of Computer Vision, IEEE Workshop on*, 0:1–8, 2008.

[4] 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*, 29(6):1052–1067, 2007.

[5] G. Demming. Sony eyetoy ^{TM}: Developing mental models for 3-D interaction in a 2-D gaming environment. In *Computer Human Interaction*, volume 3101, pages 575–582. Springer, 2004.

[6] M. Kaess and F. Dellaert. Visual slam with a multi-camera rig. Technical Report GIT-GVU-06-06, Georgia Institute of Technology, Feb 2006.

[7] H. Li, R. Hartley, and J.-H. Kim. A linear approach to motion estimation using generalized camera models. In *Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[8] H. C. Longuet-Higgins and K. Prazdny. The interpretation of a moving retinal image. In *Proc. Royal Society London. B208*, pages 385–397, 1980.

[9] S. Tariq and F. Dellaert. A multi-camera 6-dof pose tracker. In *In Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 296–297, Washington, DC, USA, 2004. IEEE Computer Society.

[10] A.-T. Tsao, C.-S. Fuh, Y.-P. Hung, and Y.-S. Chen. Ego-motion estimation using optical flow fields observed from multiple cameras. In *Computer Vision and Pattern Recognition*, page 457, Washington, DC, USA, 1997. IEEE Computer Society.

[11] G. Welch, G. Bishop, L. Vicci, S. Brumback, K. Keller, and D. Colucci. High-performance wide-area optical tracking: The hiball tracking system. *Presence: Teleoperators and Virtual Environments*, 10(1):1–21, 2001.

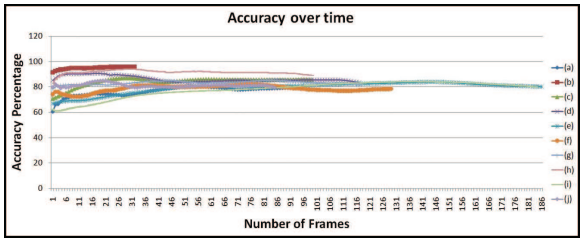

Figure 5: Change in accuracy over time of motions instances from experiment set 1

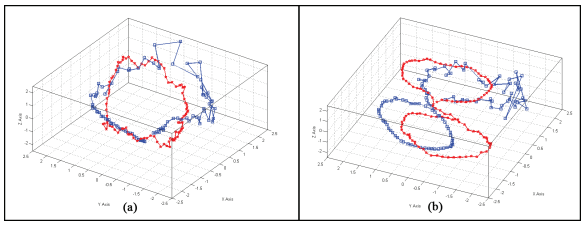

Figure 6: Trajectories from experiment set 2 showing how the optical device tracks accurately, when we move out of the range of the EM tracker, and the EM tracker gives jittery data, optical device is shown in red and EM tracker in blue

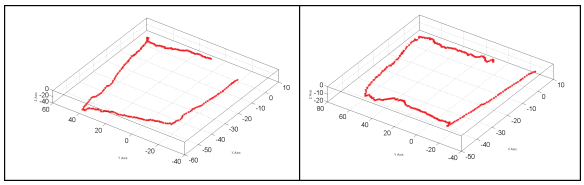

Figure 7: Trajectories of large range motion instances from experiment set 3 in outdoor and hallway settings