FingertipCubes: An Inexpensive D.I.Y Wearable for 6-DoF per Fingertip Pose Estimation using a Single RGB Camera

Ojaswi Gupta
TCS Research
New Delhi, India
ojaswi13071@iiitd.ac.in

Ramya Hebbalaguppe
TCS Research
New Delhi, India
ramya.hebbalaguppe@tcs.com

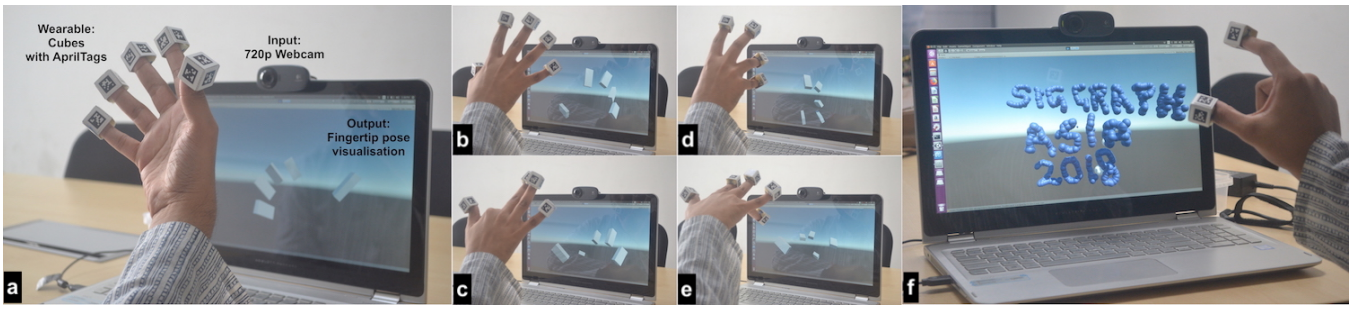

Figure 1: (a) Our setup consists of a D.I.Y. wearable, and a laptop with a webcam (b,c,d,e) Pose estimation results of fingertips, visualized in Unity for different hand poses (f) Example application of in-air writing

CCS CONCEPTS

• **Human-centered computing** → **Interaction devices**; **Graphics input devices**; *Pointing devices*; *Gestural input*;

KEYWORDS

3D User Interface, Hand Tracking, Input Device, Pose Estimation

ACM Reference Format:
Ojaswi Gupta and Ramya Hebbalaguppe. 2018. FingertipCubes: An Inexpensive D.I.Y Wearable for 6-DoF per Fingertip Pose Estimation using a Single RGB Camera. In *Proceedings of SA '18 Posters*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3283289.3283349

1 ABSTRACT

It is natural to use our hands for interacting with a virtual world, but this remains to be widely available. The Leap Motion controller has brought 3D hand tracking to consumers, but its high cost prohibits its mass adoption, especially for users in developing countries. To facilitate mass adoption, we present a do-it-yourself wearable that has a material cost of only 1 US Dollar, and which coupled with a webcam, can provide 6-DoF(degrees of freedom) per fingertip tracking in real-time. We also propose a novel solution to the pose ambiguity problem of a single square planar fiducial marker in monocular view.

SA '18 Posters, December 04-07, 2018, Tokyo, Japan

ACM ISBN 978-1-4503-6063-0/18/12.
https://doi.org/10.1145/3283289.3283349

2 INTRODUCTION

3D motion tracking of hands has been of great interest in the visual computing community due to its applications in 3D user interfaces, Virtual Reality, etc. Over the last decade, depth sensing based approaches have become the de facto standard for this purpose, and monocular markerless 3D hand tracking in the absence of depth information is considered impractical. Our goal in this work is to perform 3D tracking of fingertips, using only equipment which is often readily available or easy to acquire.

[Huang et al. 2015] have shown that fingertips can be tracked by using custom designed markers and a magnetic tracking system. Although their results are impressive with robustness to self-occlusions, the system remains highly inaccessible for an end user due to the equipment cost and lack of availability.

Recently, fiducial markers have been used for motion tracking of physical objects. [Wu et al. 2017] tracked a passive stylus for writing and drawing in mixed reality using a single camera, whereas [Zoss et al. 2018] captured jaw motion using multiple cameras. Both of these use 3D-printed polyhedrons which have binary fiducial markers glued onto them. Model based 2D-3D pose estimation techniques are then used to track the polyhedrons. Inspired by these approaches, we present our system - FingertipCubes, which provides real-time 6DoF per fingertip tracking using a single RGB camera and a wearable consisting of D.I.Y. cardboard cubes textured with binary fiducial markers, as described in the next section. See figures 1b to 1e for some tracking results.

3 OUR METHOD

In this section we discuss the design of our wearable and then describe the algorithm for 6-DoF pose estimation of each worn cube, and hence the enclosed fingertip. Figure 1a describes our setup.

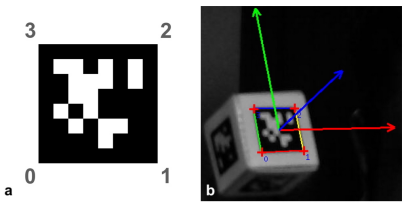

Figure 2: (a) An Apriltag with numbered corners (b) Same tag deployed on a FingertipCube, with pose displayed

3.1 Our wearable

Our wearable consists of multiple cubes textured with Apriltag[Wang and Olson 2016] binary fiducial markers. These cubes can be worn on the fingertips by either directly inserting the fingers into them or by sticking the cubes on gloves and then wearing them. We created them in a D.I.Y. fashion by printing the unfloded 3D model of each cube on regular paper, sticking it onto cardboard, then cutting, folding and gluing it into a cube. We chose cubes as our wearables as Apriltags need to be planar and we wanted to maximize the number of visible markers, without making the wearable cumbersome to wear on the fingertips. An alternate construction of the wearable could be 3D-printing the cubes and gluing the markers, as in [Wu et al. 2017][Zoss et al. 2018].

3.2 Our proposed algorithm

To find the 6-DoF pose for each cube, we run the Apriltags detector on each frame of a webcam feed, which gives the unique ID number and pose of each visible marker, see figure 2b. Now for each cube, if more than one marker is visible, we use the pose of the marker with maximum projected area onto the image plane. The pose is adjusted to the center of the cube by the relative transformation between the particular marker and the cube's local coordinate system.

As the rotation part of the pose estimated from 4 coplanar points is known to be ambiguous [Jin et al. 2017], and is highly unstable in our case of small markers, we rectify it as described in section 3.3. We do not further refine this new pose by minimizing the norm of the reprojection error by a nonlinear approach such as Gauss Newton or Levenberg Marquardt technique [Marchand et al. 2016], as it is not guaranteed to converge to the correct local minimum, and can reintroduce the pose ambiguity. Unlike [Wu et al. 2017][Zoss et al. 2018], we don't use model based 2D-3D pose estimation techniques when more than one markers are visible, as the D.I.Y. cubes are not precise in construction.

3.3 Unambiguous pose from a single marker

Perspective-n-point(PnP) pose estimation from 4 coplanar points is known to have an ambiguous rotation, under measurement noise and variance in corner detection, which is significant in our case due to small markers. [Jin et al. 2017] proposed using sensor fusion with a depth sensor for this issue. We instead stick to monocular RGB and exploit the parallel lines in the marker to disambiguate the pose by estimating the rotation from the vanishing points.

Figure 2a shows an Apriltag with corners marked 0 to 3, which are given by the Apriltags algorithm. The line joining 0 to 1 and the one joining 3 to 2 are intersected to get the X-vanishing point(V_x). Similarly the Y-vanishing(V_y) point is obtained from lines joining 0 to 3 and 1 to 2. In order to compute these vanishing points, we perform computations on the Gaussian sphere as in [Bazin and Pollefeys 2012] instead of on the image space. V_z is given by $V_x \times V_y$. These 3 vanishing points on the Gaussian sphere give the columns of the rotation matrix we seek.

A drawback of this approach is that measurement noise in the marker corner detection still affects the rotation estimate, and it increases as the marker moves away from the camera or appears skewed due to large angular deviation. To deal with this we perform temporal smoothing by using a moving average of the new rotation and the rotations of the last 5 frames.

4 RESULTS AND DISCUSSION

We implemented our system using ViSP[Marchand et al. 2005] which has an Apriltags module inbuilt, for pose estimation, and Unity for pose visualization. Under good lighting conditions, our system achieves real-time performance at 25 frames per second, has a range of upto 80 centimeters but is limited to 60 degrees field of view of the webcam. It is able to handle fast hand movements by lowering the exposure of the webcam, without which markers are not detected due to high motion blur. Our system is inexpensive as the wearable has a material cost of only 60 INR(1 USD), and a 20 USD 720p webcam is used. Considering the alternate use value of the webcam, our system is significantly cheaper than an 80 USD Leap Motion controller.

We show one application of our system, in-air writing, see figure 1f. Simple extensions to our system in Unity can enable other applications such as 3D object creation using two hands like Leap Motion Orion or providing input to games: controlling a plane in a flight simulator, or controlling a gun in a first person shooter game.

5 FUTURE WORK

In future work we envision to extend our system to perform monocular 26-DoF full hand tracking, where the big challenge is to deal with self occlusions. Due to the lightweight nature of our system, it is also possible to extend it to Google Cardboard based mobile Virtual Reality, where self occlusions in egocentric view will again be a challenge.

REFERENCES

Jean-Charles Bazin and Marc Pollefeys. 3-line ransac for orthogonal vanishing point detection. In *IROS*, pages 4282–4287. IEEE, 2012.

Jiawei Huang, Tsuyoshi Mori, Kazuki Takashima, Shuichiro Hashi, and Yoshifumi Kitamura. Im6d: magnetic tracking system with 6-dof passive markers for dexterous 3d interaction and motion. *ACM Transactions on Graphics (TOG)*, 34(6):217, 2015.

P. Jin, P. Matikainen, and S. S. Srinivasa. Sensor fusion for fiducial tags: Highly robust pose estimation from single frame rgbd. In *IROS*, pages 5770–5776. IEEE, Sept 2017.

E. Marchand, F. Spindler, and F. Chaumette. Visp for visual servoing: a generic software platform with a wide class of robot control skills. *IEEE Robotics and Automation Magazine*, 12(4):40–52, December 2005.

Eric Marchand, Hideaki Uchiyama, and Fabien Spindler. Pose estimation for augmented reality: a hands-on survey. *IEEE transactions on visualization and computer graphics*, 22(12):2633–2651, 2016.

John Wang and Edwin Olson. Apriltag 2: Efficient and robust fiducial detection. In *IROS*, pages 4193–4198. IEEE, 2016.

Po-Chen Wu, Robert Wang, Kenrick Kin, Christopher Twigg, Shangchen Han, Ming-Hsuan Yang, and Shao-Yi Chien. Dodecapen: Accurate 6dof tracking of a passive stylus. In *UIST*, pages 365–374. ACM, 2017.

Gaspard Zoss, Derek Bradley, Pascal Bérard, and Thabo Beeler. An empirical rig for jaw animation. *ACM Trans. Graph.*, 37(4):59:1–59:12, July 2018.