International Journal of Computer Applications (0975 - 8887)
Volume 122 - No. 7, July 2015

Visual 3D Model-based Tracking toward Autonomous Live Sports Broadcasting using a VTOL Unmanned Aerial Vehicle in GPS-impaired Environments

Inkyu Sa
Science and Engineering Faculty
Queensland University of Technology
2 George street, Brisbane, Queensland, Australia

Ho Seok Ahn
Department of Electrical and Computer Engineering, CARES
The University of Auckland,
368 Khyber Pass Roed, Newmarket, Auckland, New Zealand

ABSTRACT

This paper presents a novel approach for autonomous live sports broadcasting using visual 3D model-based tracking and a vertical take-off and landing (VTOL) unmanned aerial vehicle (UAV) such as a quadcopter or hexacopter in GPS-impaired environments. To achieve this level of autonomy, position estimation is essential and is a highly challenging problem using a monocular camera due to the scale ambiguity. In this paper, we track a tennis court, that is standard in dimension, using a moving edge-based tracker, and recover the scale with the prior knowledge of the fixed playing field. Experimental results are demonstrated in 3 different environments including static scenes, real broadcast video, and indoor flying. We also evaluate the proposed approach with the ground truth provided by a motion capture system and achieve a position estimation with less than $0.02\,\mathrm{m}$ standard deviation in the error.

General Terms:

Drones, Image Processing, Live Sports Broadcasting

Keywords

Vision, VTOL, UAVs, Model-Based Tracking, State Estimation, GPS-Impaired Environments

1. INTRODUCTION

Live sports broadcasting is getting a big market these days; for example, approximately 4.7 billion people, or roughly two-thirds of the world's population, watched the Beijing Olympics through TV [1]. Technological innovations in live sports broadcasting have been accelerated in the last decade. The first sports broadcasting only delivered textual data or symbols in the 1890s. Static cameras are traditionally used to ensure complete coverage of the playing field due to the large size or shape of the field. Recently, multiple moving colour cameras and image processing techniques allow viewers to see dynamic scenes from a variety of angles [2, 3]. As many cameras are used for providing better coverage of the field, especially for soccer and football, technology for choosing the optimal camera is important [4, 5]. Some systems detect multiple players and track them with PTZ camera automatically [6]. Some systems extract highlight scenes by analysing sports broadcasts using the intensity of acoustics, by using keywords such as "goal" spoken by an announcer, or using visual feature sequences and so on [7, 8, 9]. Many commercial products apply augmented reality to sports broadcasts, such as the BBC Piero system[1], Viz Arena[2], Tog Sports[3], and Spider Cam [4].

Fig. 1. The UAV captures the tennis court for autonomous live sports broadcasting. This figure is only for conceptual illustration of the proposed application.

All of these technologies provide improved quality of sports broadcasting, but the most basic task for applying these technologies is capturing dynamic scenes in a variety of angles. One good solution for this is using UAVs and drones with cameras. These days, there is a growing interest in UAVs and drones for various purposes, such as surveillance, reconnaissance operations, monitoring, and aerial photography. Many broadcasting companies use UAVs and drones with cameras to obtain dynamic scenes from

[1] `http://www.bbc.co.uk/rd/projects/piero`
[2] `http://www.vizrt.com/products/viz_arena`
[3] `http://rtsw.co.uk`
[4] `http://www.spidercam.org`

a wide view without complex equipment[5]. Amazon is planning to use UAVs for delivery in the USA and are now free to test its Prime Air service[6]. This means UAV technology is on the level for commercialisation and will become part of our daily life if we can verify the safety of UAVs. Someday Google and Facebook hope to use drones to assist in providing Internet access worldwide from the sky. UAVs and drones are used for ocean exploration [10], forest fire monitoring [11], wildlife tracking without human intervention that allows providing better protection and securing of animals [12], disaster recovery [13], infrastructure inspections [14], and crop, soil, and water status monitoring [15, 16].

In this paper, we present an approach for the application of sports live broadcasting by addressing the limitations of existing approaches with the proposed high-performance system. Especially, we focus on introducing a novel application that makes use of vision-based model tracking and an unmanned aerial vehicle for autonomous live sports broadcasting.

By recovering a scale using a predefined CAD model and a monocular camera, we demonstrate precise metric position estimation. We examine three case studies for tracking a tennis court: 1) desktop environment, 2) broadcasting video, and 3) using a real UAV system. We evaluate the feasibility and performance of our method through these intensive experiments using real broadcast video and ground truth provided by a motion capture system.

This paper is structured as follows: Section 2 introduces related work and background. Section 3 describes coordinate systems, 3D model-based tracking and dealing with unknown camera calibration parameters. We present our experimental results in Section 4 and conclusions in Section 5.

2. RELATED WORKS

2.1 Sports Broadcasting

Skycam is an instance of the current technological development in live sports broadcasting. A cable suspended camera hangs in open space over a playing ground and moves in horizontal and vertical directions driven by motors anchored at elevated corners of the stadium. This approach has been successfully demonstrated for popular sports leagues such as NFL, NBA, and the FIFA world cup. However, there are issues that must be considered. Firstly, using a cable mechanism can only generate tension force where the precision is subject to the flexibility of cables. Secondly, this system requires rigorous and costly pre-setup procedures including installation of reels, cameras, and a control tower. Using flying vehicles for autonomous sports broadcasting is a feasible alternative for resolving the limitations of Skycam. They can carry sufficient payloads including a stabilised camera system and can autonomously fly with GPS-based stabilised flight modes in outdoor environments. Off-the-shelf VTOL platforms are affordable.

However, it is challenging to use these platforms for fully autonomous broadcasting. There are many domed or covered stadiums and indoor arenas where GPS is not applicable. Moreover, GPS signals can experience interference from external disturbances in outdoor environments which in turn lead to inaccurate position estimation. Accurate position estimation of a VTOL platform is key for stable flight and high-quality broadcasting.

[5] http://www.aeronautics-sys.com/?CategoryID=259&ArticleID=191

[6] http://www.amazon.com/b?node=8037720011

To achieve such a level of accuracy with a VTOL platform, vision sensors are a favourable choice for small flying vehicles since they are lightweight, low-cost, and can provide information-rich data. However, there are also some challenges, including intensive computing requirements and scale recovery. A scale cannot be determined without the aid of other sensors (e.g., a laser range finder, IMU, or a sonar sensor) or prior knowledge such as CAD information. In this paper, we present an approach that makes use of known metric information (i.e., a tennis court) for state estimation. A fast and computationally cheap edge tracker with a CAD model can resolve both issues mentioned above. The proposed approach can be applied for any tennis court tracking since the dimensions of tennis courts are internationally uniform. Besides, it is easy to introduce another model for tracking sports such as soccer, baseball, and basketball fields.

2.2 UAVs for broadcasting

Recently, interest in unmanned aerial vehicles has increased, considering their broad range of applications in civil and military domains, such as aerial transport, geographical surveillance, and entertainment. Particularly, vertical take-off and landing (VTOL) micro-aerial vehicles (MAVs), e.g., a quadcopter or a hexacopter, offer a flexible and adaptable platform amenable to aerial research for such applications. They have advantages of small size, agile maneuverability, low-cost, and useful payloads [17, 18].

A large amount of impressive research has been presented using these platforms. One of the interesting and challenging problems is accurate state estimation that tracks internal states such as position, orientation, velocities, and biases. GPS-based systems have demonstrated stable performance for outdoor navigation, but their accuracy may be insufficient for autonomous flying and is not suitable for indoor systems. Using motion capture devices can provide accurate estimates within limited workspaces [10]. Stochastic filter frameworks, e.g., an Extended Kalman Filter, (EKF), or Particle Filter, can be utilised for state estimation using a vision sensor and an inertial measurement unit (IMU) [19].

In order to avoid this issue, we propose to use a visual 3D model-based tracking approach for providing position and orientation estimation. This approach has been widely developed for decades due to its simplicity and robustness under varying lighting conditions. Lepetit et al. [20] surveyed the state-of-the-art industrial and research grade 3D model-based tracking solutions. Teulire et al. [21, 22] demonstrated 3D model-based tracking and position control of UAVs. Translational velocity was estimated with inertial measurements and position estimation. This work is closely related to the method presented in this paper. However, we apply the technique to a different application (i.e., tennis court tracking) and quantitatively evaluate the performance of the system using accurate ground truth and real broadcasting video.

3. PRELIMINARIES

In this section, coordinate systems, 3D model-based tracking, manual model initialisation, and camera calibration are introduced. We utilise Visual Servoing Platform (ViSP) as a front-end feature tracking module as shown in Figure 2. This open-source software package includes useful computer vision algorithms, feature tracking, and visual servoing implementations. Below, we provide a brief overview of ViSP model-based tracking; more details can be found in [23].

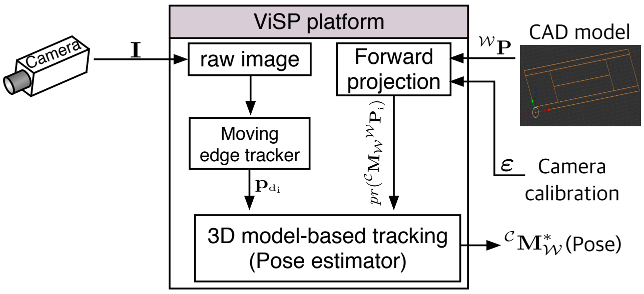


Fig. 2. 3D model-based tracking system diagram. **I** is a grey input image and $\mathbf{p}_{d_i}$ is a tracked set of points using a moving edge tracker. The CAD model in world coordinates, ${}^{\mathcal{W}}\mathbf{P}$, is projected onto an image plane with camera intrinsic and extrinsic parameters, $\boldsymbol{\varepsilon}$. ${}^{\mathcal{C}}\mathbf{M}^*_{\mathcal{W}}$ is an optimal estimator of the position and orientation of a camera.

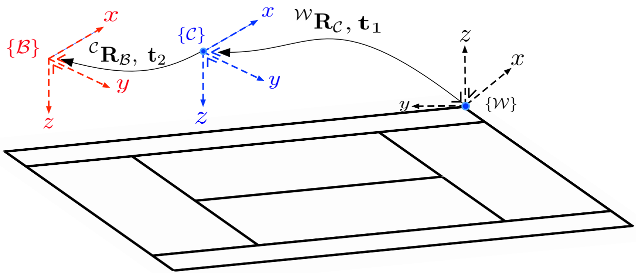


Fig. 3. Coordinate systems definitions. Transformation between $\{\mathcal{B}\}$ and $\{\mathcal{C}\}$ is constant whereas $\{\mathcal{W}\}$ and $\{\mathcal{B}\}$ vary as a camera moves. The world coordinate system is defined on the top-right corner of the tennis court.

3.1 Coordinate systems

We define 3 right-handed frames: world $\{\mathcal{W}\}$, camera $\{\mathcal{C}\}$ and body $\{\mathcal{B}\}$ as shown in Figure 3. $\{\mathcal{W}\}$ has its z-axis upward while $\{\mathcal{C}\}$ (camera optical axis) and $\{\mathcal{B}\}$ have their z-axis downward. We define the notation ${}^{a}\mathbf{R}_{b}$ which rotates a vector defined with respect to frame $\{b\}$ to a vector with respect to frame $\{a\}$. The rotation and translation between a camera and a body, ${}^{\mathcal{C}}\mathbf{R}_{\mathcal{B}}, \mathbf{t}_2$ are considered as constant in this paper. ${}^{\mathcal{W}}\mathbf{R}_{\mathcal{C}}, \mathbf{t}_1$ is estimated by the model-based tracking algorithm.

3.2 3D model-based tracking

3D model-based tracking is the trace of the projection of the known 3D model on an image [24]. Time-consuming global edge extraction is not required since this spatio-temporal edge tracker only needs positions of sample pixels and their intensities. For example, line tracking using moving edges is illustrated in Figure 4. Firstly, line, $\ell(\boldsymbol{r})^k$, in an image I^k is manually defined by the user with a step size, w. k and $\boldsymbol{r}$ are a time stamp and camera pose respectively. For each sample point denoted as blue in Figure 4(a), a 1D search is performed along the normal direction of each sample to determine the search interval, d. The maximum convolutional response with a pre-defined 3×3 mask is detected and an M-Estimator is utilised for outlier rejection. Given this tracked model, the camera pose can be defined as

$$ {}^{\mathcal{C}}\mathbf{M}^*_{\mathcal{W}} = \arg\min_{{}^{\mathcal{C}}\mathbf{R}_{\mathcal{W}}, \mathbf{t}_1} \sum_i (\mathbf{p}_{d_i} - pr({}^{\mathcal{C}}\mathbf{M}_{\mathcal{W}}{}^{\mathcal{W}}\mathbf{P}_i))^2 \qquad (1) $$

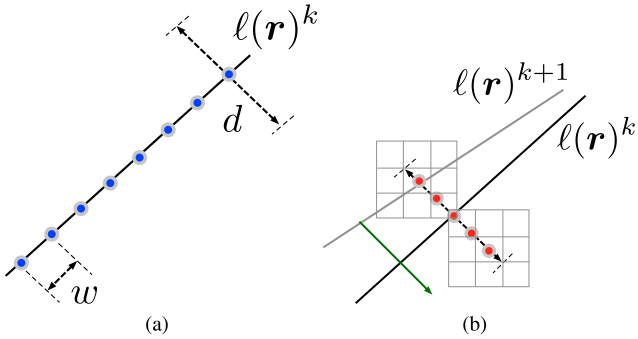


Fig. 4. (a) An example of a line $\ell(\boldsymbol{r})^k$ search at time stamp k. w is a constant step size in pixel units whereas d is a varying search interval. Blue dots denote sampling points. (b) At $k+1$, 1D convolutional searching seeks the maxima on $\ell(\boldsymbol{r})^{k+1}$.

where $\mathbf{p}_{d_i}$ are the matrix produced by the moving edge tracker and ${}^{\mathcal{W}}\mathbf{P}_i$ is 3D points of a CAD model in the world coordinate frame. ${}^{\mathcal{W}}\mathbf{M}_{\mathcal{C}}$ is a 4×4 homogeneous matrix containing camera pose, position, and orientation (usually called camera extrinsics). $pr(\cdot)$ is the projection that is a function of the camera pose. Therefore, Equation (1) estimates camera pose by minimising the sum of the errors between the edges in the image and those of the 3D model projected onto the image plane given the camera pose.

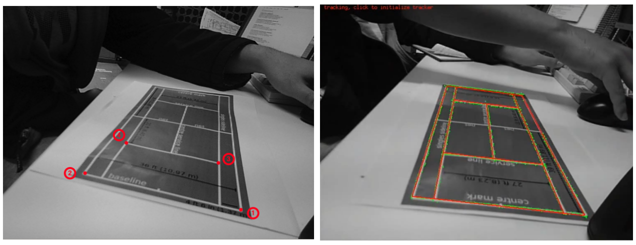
It is important to mention that this moving edge tracker has the advantages of fast tracking and providing mathematical representations of the tracked edges. However, the tracking speed and performance depends on the number of sample points determined by w and the search interval d. A user is still required to tune these parameters empirically with respect to the target application.

3.2.1 Manual model initialisation. Before tracking a 3D model, a model initialisation step is required to associate correspondences between the 3D CAD model and the 2D image projection of it. Given prior knowledge of a tennis court, we can link a 3D model to the 2D image by sequentially clicking points in the image as shown in Figure 5 (a). A minimum of three points that are not collinear are required to define a surface representing the tennis court. However, we use four symmetric points that are close to the camera for a more accurate initialisation. One point is redundant and can be removed. After the initialisation, the moving edge tracker is invoked as depicted in Figure 5 (b). Note that the number of points for the model initialisation vary depending on playing fields.

3.3 Camera calibration

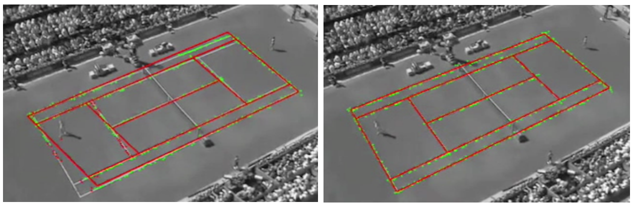
We faced an interesting and well-known problem: unknown camera calibration [25]. For instance, we demonstrate accurate 3D model fitting for the desktop experiment (see Figure 5-b), whereas the 3D model fitting is quite poor for broadcast video (see Figure 6-a). The right bottom of the 3D model is poorly aligned with the tennis court in the image. It was difficult to obtain intrinsic camera calibration parameters for the latter case (live broadcast video), whereas we were able to carry on camera calibration for the former case (desktop experiment).

There is a straightforward way to estimate unknown camera intrinsic parameters ξ such as principle points, focal length, and distortion parameters if we have information about an object. This



(a) (b)

Fig. 5. Guidance for manual model initialisation (a) and forward-projection of a CAD model (red) with the initialisation (b). Green dots denote tracked point features.



(a) (b)

Fig. 6. Tracking results without calibration (a) and with (b). Camera calibration yields longer tracking performance.

information determines how the object should be seen in an image. We can define a camera by ${}^{\mathcal{W}}\mathbf{M}_{\mathcal{C}}, \xi$ and an object ${}^{\mathcal{W}}\mathbf{P}$ with respect to the world coordinate frame. Then, the position of the object in $\{\mathcal{C}\}$ and its projection can be expressed as

$$\begin{aligned} {}^{\mathcal{C}}\mathbf{P} &= {}^{\mathcal{C}}\mathbf{M}_{\mathcal{W}} {}^{\mathcal{W}}\mathbf{P} \\ \mathbf{p}_{\text{prj}} &= pr_{\xi}({}^{\mathcal{C}}\mathbf{P}). \end{aligned}$$

The objective of camera calibration is to minimise the error $\mathbf{e} = \|\mathbf{p}_{\text{prj}} - \mathbf{p}_{\text{mea}}\|$ where $\mathbf{p}_{\text{mea}}$ denotes the measured visual feature.
The motion of measured features is a function of camera velocity $\boldsymbol{\nu}_{\text{c}}$ and the time differentiation of intrinsic parameters ξ.

$$\dot{\mathbf{p}}_{\text{prj}} = \mathbf{L}_{\mathbf{p}}\boldsymbol{\nu} \quad \text{where} \quad \boldsymbol{\nu} = \begin{bmatrix} \boldsymbol{\nu}_{\text{c}} \\ \dot{\xi} \end{bmatrix}, \mathbf{L}_{\mathbf{p}} = \begin{bmatrix} \frac{\partial \mathbf{p}_{\text{prj}}}{\partial \mathbf{r}}, \frac{\partial \mathbf{p}_{\text{prj}}}{\partial \xi} \end{bmatrix}$$

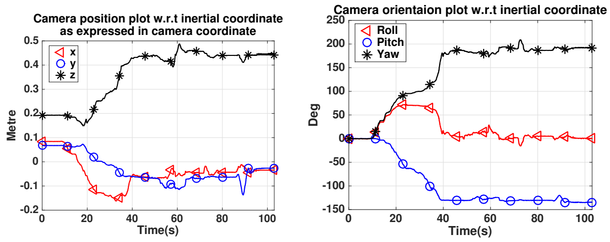
Finally, $\boldsymbol{\nu}$ can be expressed as:

$$\boldsymbol{\nu} = -\lambda \mathbf{L}_{\mathbf{p}}^{+}\mathbf{e}$$

where $\lambda \in \mathbb{R}$ is a positive gain, that determines the decay rate and $\mathbf{L}_{\mathbf{p}}^{+}$ is a pseudo inverse of $\mathbf{L}_{\mathbf{p}}$. Note that we introduce only general ideas of camera calibration using a 3D CAD model and image measurements. It is also recommended to read the related works for more detail and solid explanations [26, 27].

4. EXPERIMENTAL RESULTS

In this section, the experimental hardware and software setup and results in three different environments: indoor desktop, outdoor broadcasting, and indoor flying are presented. The image instances for each test are shown in Figure 7. Each result consists of a metric position, orientation, and 3D camera pose plots with respect to the world coordinate frame. We also present ground truth



(a) Position estimation (b) Orientation estimation

Fig. 8. Indoor desktop experimental results. The CAD model is rotated 180° during the experiment whereas the camera is stationary. Both position and orientation estimates show qualitatively reasonable results but it is challenging to evaluate the performance without ground truth.

comparison provided by a motion capture system that can measure sub-millimetre accuracy at >100 Hz for the indoor flying test in order to quantitatively evaluate the proposed system. It is worth mentioning that the flying experiments are conducted with safety nets and at low-altitude under the QUT legislation and guidance regulations. Video demonstration is available from the link below[7].

4.1 Experimental hardware and software setup

In this paper, we utilise a different camera for each test. Firstly, a low-cost, $8 AUD, high-speed, up to 100 Hz at 320×240, Playstation EyeToy camera is used for the desktop test. This cost effective CMOS camera has a rolling shutter that introduces problems for moving platforms. Secondly, a global shutter industrial low-end grayscale Bluefox camera from MatrixVision is used for the flying test. 752×480 image sequences are recorded at 30 Hz with manual piloting. Third, an unknown commercial zoom camera provides footage for a real tennis game. Tracking is performed by post-processing with a computer (Intel i5 3.2 GHz CPU and 16 GB RAM). Ubuntu Linux, 14.04 trusty and Visual Servoing Platform (ViSP, 2.10.0) on Robot Operating System (ROS indigo) software packages are utilised. The broadcasting dataset is acquired from a previous tennis match [8].

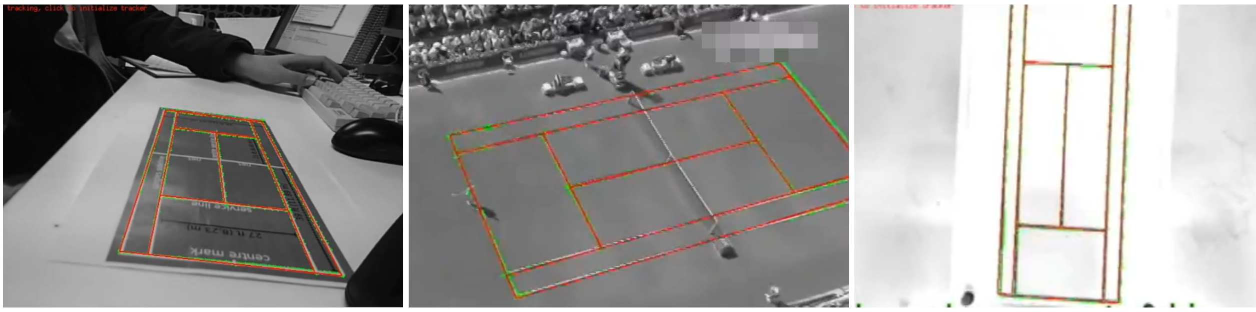
4.2 Desktop tracking results

The first experiment is performed in a controlled, static environments. An accurately calibrated EyeToy static camera is placed about 20 cm away from a scaled tennis court. After initialisation, a human rotates the tennis court 180° and introduces occlusion and shape variations. For most desktop testing, the tracker is able to track the court well as shown in Figure 8. The object starts to move at about 20-60 s and the camera moves from 60-90 s. There are two peaks at 60 s and 90 s. The former is caused by object shape variation and the latter is due to the camera shaking just before moving. Even though we do not have access to ground truth for this experiment, the performance is qualitatively demonstrated in the accompanying demonstration video clip.

4.3 Broadcasting video results

It is interesting to apply the proposed approach to the real broadcast image sequences. There are fundamental challenges for this test

[7] http://youtu.be/5eH0U3iMDDw
[8] The authors do not hold the copyrights of this testing video.



(a) Indoor desktop (b) Outdoor broadcasting video (c) Indoor UAV flying

Fig. 7. Three experimental environments.

in comparison to the previous test. Firstly, we do not have access to intrinsic camera parameters which yields poor model fitting and lost tracking. Secondly, image scenes are more complex and dynamic with moving players and camera zooming. Lastly, the tracker often loses the tennis court when only small portions of the tennis court are visible.

To address the first issue, we employ an unknown camera calibration technique. The moving edge tracker performs reasonably well with illumination changes and small occlusions, but it still requires fine parameter tuning. We cannot resolve the last issue in this paper, but propose some feasible solutions in section 4.5.

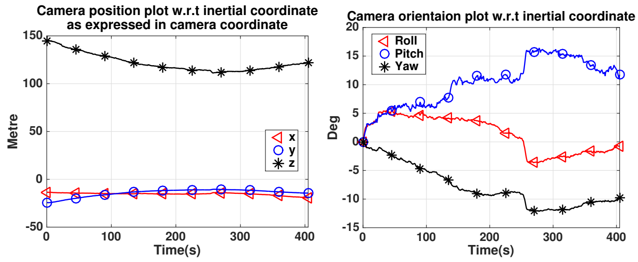
Figure 9 shows position and orientation estimation with 3D camera pose plots. Like the previous test, we do not have ground truth for this experiment; however, the performance is qualitatively demonstrated in the accompanying video clip. The camera is initially located about 145 m away from the tennis court and approaches 30 m from the tennis court with small linear and angular velocities and decelerates at 115 m.

4.4 UAV experiments

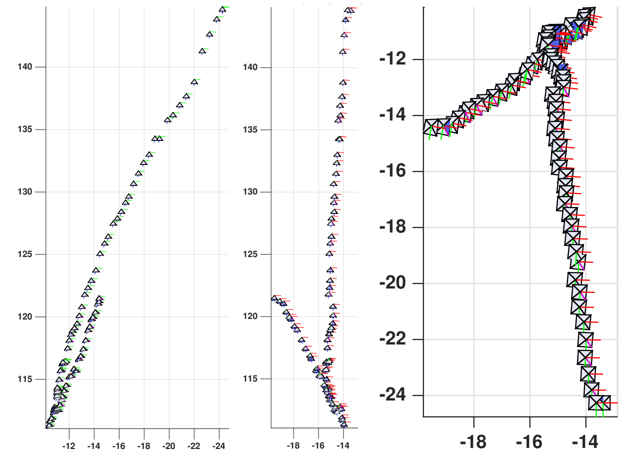
The last test is conducted with a down-facing camera on a manually piloted flying vehicle. A scaled-down tennis court is placed on the ground and image sequences are recorded for post processing; as shown in Figure 10.

Figures 11 and 12 show position and orientation estimation results, respectively. The results are plotted only for a successful tracking period: 42–67 s. For position estimation, the tracker does not suffer from drift, which is often a challenging issue for such problems due to small error accumulation over a long period. In our position estimation problem, error accumulation is not included, but rather is minimised the error between the model projection and measurements. We compute the standard deviation of each error (i.e., the differences between ground truth and estimates) in the period between 37–67 s and achieve $\sigma_x = 0.017\,\mathrm{m}$, $\sigma_y = 0.015\,\mathrm{m}$, and $\sigma_z = 0.010\,\mathrm{m}$. This performance is impressive and promising for future work such as closed-loop control and autonomous navigation.

For orientation estimation, pitch and roll, the rotation along the camera x and y-axes respectively, suffer from large noise. It is found that those angle changes introduce errors in image features that, in turn, produce errors for the moving edge tracker. It is necessary for VTOL platforms to tilt into the direction they want to move, which implies that the translation and rotation



(a) Position estimation (b) Orientation estimation



(c) Left view (d) Rear view (e) Top view

Fig. 9. Outdoor broadcasting video experimental results. (a) and (b) are position and orientation estimation results and (c-e) are camera trajectory plots in different views. The skycam is initially located around 145 m away from the tennis court and slowly flies to the target. Similar to desktop experiments, both position and orientation estimates show qualitatively reasonable results, but it is difficult to evaluate the performance without ground truth.

behaviours are coupled. Thus, it may be difficult to avoid this issue if we are exclusively using a vision sensor. However, we can alternatively utilise a low-end onboard IMU sensor for providing drift-fee reliable pitch and roll angle estimation. For the yaw angle, this model-based tracking approach can also provide drift-free

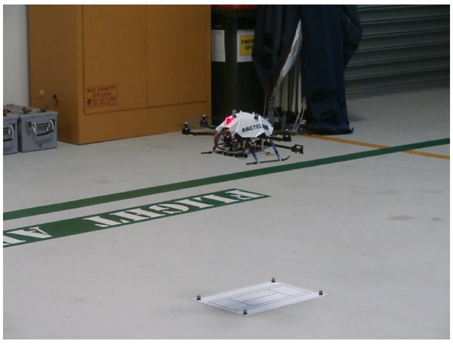


Fig. 10. Indoor UAV flying experiment with 62.6 times scaled-down tennis court. A motion capture system measures ground truth positions of reflective markers (grey dots) on the vehicle and the tennis court. For safety reasons, we can only demonstrate low-altitude flight tests.

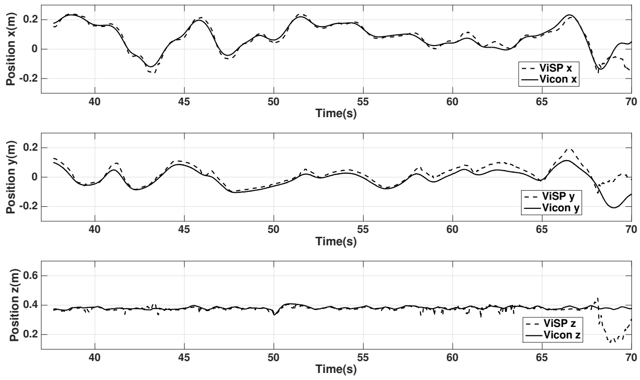


Fig. 11. Position estimation results with the ground truth for indoor UAV flying. The dotted line is our estimation and the solid line is ground truth provided from motion capture devices. Note that the tracker is able to track the tennis court for only 30 s and we discuss this issue in section 4.5.

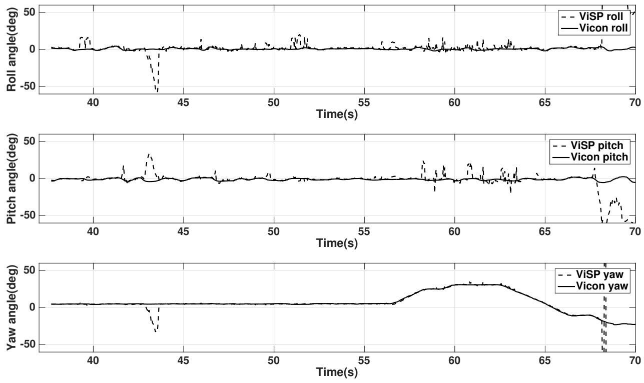
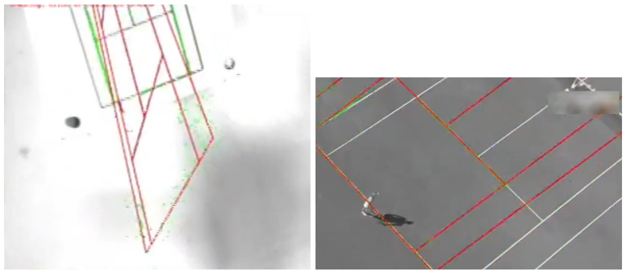


Fig. 12. Orientation estimation results with the ground truth for indoor UAV flying. The dotted line is our estimation and the solid is ground truth provided from motion capture devices. Abundant errors in `pitch` and `roll` estimation reflect a fact that a VTOL platform has to rotate first into the direction of the traverse and this motion introduces errors in an image and angle estimation.



(a) (b)

Fig. 13. Model tracking failure cases. The model tracker often loses tracking when observations are insufficient due to occlusions or out of FOVs.

estimates similar to the position estimation. This will be useful for future closed-loop development. We calculate standard deviations for the same period and achieve $\sigma_{\text{pitch}} = 6.71°$ $\sigma_{\text{roll}} = 5.26°$, and $\sigma_{\text{yaw}} = 3.84°$.

4.5 Discussions and limitations

As we presented in the previous section, we can see promising results for tracking. The pose estimator is also simple and drift-free. However, there are still many challenges that can be divided into two major issues: safety and continuous tracking.

Regarding the first issue, we only demonstrate low-altitude flying experiments due to the university's regulations. UAVs must be flown more than 30 m away from people, vehicles, and buildings and it is prohibited to fly over any populated areas. The maximum height allowance is 120 m in good weather conditions with visual-line-of-sight. This is a sensitive issue in operating UAVs and must be resolved before flight tests. One feasible approach is hiring a certified CASA operator.

Regarding the second issue, the tracker often loses tennis court tracking and is required to be manually re-initialised as shown in Figure 13. This stems from insufficient edge measurements of a 3D model, e.g., large occlusions or being beyond the field of views (FOVs). In order to address this challenge, we are planning to exploit supervised machine learning techniques that are able to provide a bounding box for the model. Within this image region, a Particle Filter based approach can estimate the optimal pose of a camera. More specifically, a particle is a forward-projection of a 3D model given a random camera pose and we are then able to compute a fitting score that is the residual of edge measurements and the projection. This local window search approach will be able to perform re-initialisation with low computation demands.

5. CONCLUSIONS AND FUTURE WORKS

In this paper, we address a novel approach for live sports broadcasting applications using a VTOL platform and 3D model-based tracking in GPS-denied environments. State estimation of metric position and orientation is demonstrated by using a moving edge-based 3D model tracker and a monocular camera in different experimental environments. Ground truth from a motion capture system is utilised for quantitative evaluation. Developing a precise and accurate tracking system is an essential first step in developing an automated UAV-based sports broadcasting system before tackling closed-loop control system issues.

Future work includes using an IMU for feature de-rotation (motion compensation) and feature prediction, auto initialisation, lost tracking recovery, and closed-loop control.

Acknowledgement

The authors would like to thank Aaron Mcfadyen from Queensland University of Technology for assistance while using the VICON system and Alex Bewley for his valuable comments and feedback.

6. REFERENCES

[1] C. Chen, O. Wang, S. Heinzle, P. Carr, A. Smolic, and M. Gross. Computational sports broadcasting: Automated director assistance for live sports. In *Multimedia and Expo (ICME), 2013 IEEE International Conference on*, pages 1–6, July 2013.

[2] Lawrence L Cone. Skycam-An aerial robotic camera system. *Byte*, 10(10):122, 1985.

[3] Christopher Hanson. Television Sport Broadcasting and Technology. *Producing Television, Spectator*, 2008.

[4] Sergio Ilarri, Eduardo Mena, Arantza Illarramendi, Roberto Yus, Maider Laka, and Gorka Marcos. A friendly location-aware system to facilitate the work of technical directors when broadcasting sport events. *Mobile Information Systems*, 8(1):17–43, 2012.

[5] Jinjun Wang, Changsheng Xu, Engsiong Chng, Hanqing Lu, and Qi Tian. Automatic composition of broadcast sports video. *Multimedia Systems*, 14(4):179–193, 2008.

[6] Guangyu Zhu, Changsheng Xu, Qingming Huang, and Wen Gao. Automatic multi-player detection and tracking in broadcast sports video using support vector machine and particle filter. In *Multimedia and Expo, 2006 IEEE International Conference on*, pages 1629–1632. IEEE, 2006.

[7] Chung-Yuan Chao, Huang-Chia Shih, and Chung-Lin Huang. Semantics-based highlight extraction of soccer program using DBN. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*, volume 2, pages ii–1057. IEEE, 2005.

[8] Yoshiaki Itoh, Shigenobu Sakaki, Kazunori Kojima, and Masaaki Ishigame. Highlight scene extraction of sports broadcasts using sports news programs. In *Multimedia Signal Processing, 2008 IEEE 10th Workshop on*, pages 646–649. IEEE, 2008.

[9] Jinjun Wang, Changsheng Xu, Engsiong Chng, and Qi Tian. Sports highlight detection from keyword sequences using HMM. In *Multimedia and Expo, 2004. ICME'04. 2004 IEEE International Conference on*, volume 1, pages 599–602. IEEE, 2004.

[10] PB Sujit, Joao Sousa, and F Lobo Pereira. UAV and AUVs coordination for ocean exploration. In *Oceans 2009-Europe*, pages 1–7. IEEE, 2009.

[11] David W Casbeer, RW Beard, TW McLain, Sai-Ming Li, and Raman K Mehra. Forest fire monitoring with multiple small UAVs. In *American Control Conference, 2005. Proceedings of the 2005*, pages 3530–3535. IEEE, 2005.

[12] St D Ilcev and Ivan Skoryk. Automatic landing of animal tracking Lighter-Than-Air Systems (LTAS) on the recharging platform. In *Microwave & Telecommunication Technology (CriMiCo), 2014 24th International Crimean Conference*, pages 275–277. IEEE, 2014.

[13] Gurkan Tuna, Tarik Veli Mumcu, and Kayhan Gulez. Design strategies of unmanned aerial vehicle-aided communication for disaster recovery. In *High Capacity Optical Networks and Emerging/Enabling Technologies*. 2012.

[14] Inkyu Sa, Stefan Hrabar, and Peter Corke. Inspection of Pole-Like Structures Using a Vision-Controlled VTOL UAV and Shared Autonomy. In *IEEE International Conference on Intelligent Robots and Systems*, pages 4819–4826, Sep. 2014.

[15] J Berni, Pablo J Zarco-Tejada, Lola Suárez, and Elias Fereres. Thermal and narrowband multispectral remote sensing for vegetation monitoring from an unmanned aerial vehicle. *Geoscience and Remote Sensing, IEEE Transactions on*, 47(3):722–738, 2009.

[16] S Nebiker, A Annen, M Scherrer, and D Oesch. A light-weight multispectral sensor for micro UAV—Opportunities for very high resolution airborne remote sensing. *The international archives of the photogrammetry, remote sensing and spatial information sciences*, 37:1193–1200, 2008.

[17] Aaron Mcfadyen, Peter Corke, and Luis Mejias. Rotorcraft Collision Avoidance using Spherical Image-based Visual Servoing and Single Point Features. In *Proc. of the International Conference on Intelligent Robot Systems (IROS)*, Oct. 2012.

[18] Inkyu Sa and Peter Corke. System Identification, Estimation and Control for a Cost Effective Open-Source Quadcopter. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2012.

[19] Stephan M. Weiss. *Vision based navigation for micro helicopters*. PhD thesis, ETHZ, 2012.

[20] Vincent Lepetit and Pascal Fua. Monocular model-based 3D tracking of rigid objects: A survey. *Foundations and trends in computer graphics and vision*, 1(CVLAB-ARTICLE-2005-002):1–89, 2005.

[21] Céline Teuliere, Eric Marchand, and Laurent Eck. 3D Model-Based Tracking for UAV Indoor Localization. *IEEE Trans. on Cybernetics*, 2014.

[22] C. Teulière, L. Eck, E. Marchand, and N. Guénard. 3D model-based tracking for UAV position control. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 1084–1089, Oct 2010.

[23] Éric Marchand, Fabien Spindler, and François Chaumette. ViSP for visual servoing: a generic software platform with a wide class of robot control skills. *Robotics & Automation Magazine, IEEE*, 12(4):40–52, 2005.

[24] Patrick Bouthemy. A maximum likelihood framework for determining moving edges. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11(5):499–511, 1989.

[25] Zhengyou Zhang. A flexible new technique for camera calibration. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(11):1330–1334, 2000.

[26] S. Hutchinson, G.D. Hager, and P.I. Corke. A tutorial on visual servo control. *Robotics and Automation, IEEE Transactions on*, 12(5):651 –670, oct 1996.

[27] Lagadic project. Visual Servoing Platform, Computer vision algorithms, 2012.