A tool for annotating homographies from hockey broadcast video

Pascale Walters
Mehrnaz Fani
David Clausi
Alex Wong
Email: pascale.walters@uwaterloo.ca
University of Waterloo

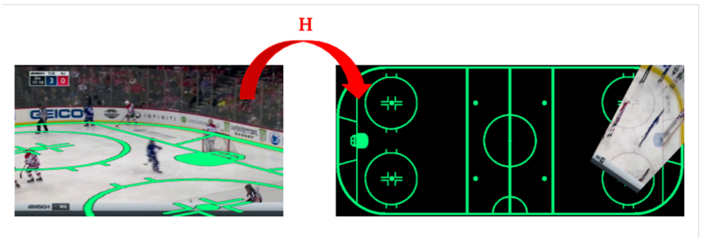


Fig. 1: Homography annotation involves determining the transform from the hockey broadcast video frame to the overhead view of the ice rink.

Dataset	Number of Images	Publicly Available?
World Cup Soccer [2]	500	
World Cup Soccer [1]	209 (train), 186 (test)	✓
Hockey [1, 5]	1.67M	
Basketball [6]	50,127	?
Volleyball [6]	12,987	?
MLS Soccer [6]	14,160	
Basketball [4]	1,232	
College Basketball [7]	526 (train), 114 (test)	

Table 1: Datasets for sports field localization. Datasets that have a ? in the Publicly Available column were reported in the paper as being made available, but do not seem to actually be so.

Abstract

In order to develop solutions for automatic ice rink localization from broadcast video, a dataset with ground truth homographies is required. Hockey broadcast video does not tend to provide camera parameters for each frame, which means that they must be gathered manually. A novel tool for collecting ground truth transforms through point correspondences between each frame and an overhead view of the ice rink is presented in this paper. Through collaboration with the users of the tool, we have added features to improve accuracy and efficiency, especially in frames with few lines on the playing surface visible. A dataset of 4,262 frames has been collected, which will be used for research into automatic camera calibration techniques.

1 Introduction

Automatic generation of hockey analytics is an interesting area of research, as insights about the players and the game can be provided to the teams and fans. These data can be used to influence management decisions, track player improvement, and increase fan interest in the game.

To generate these analytics, a critical first step is the localization of the playing surface. In broadcast footage, the camera pans, tilts, and zooms to follow the play. Registration of each frame is required in order to get the absolute locations of each of the players and the puck on the playing surface.

Several techniques have been described in the literature to perform this calibration with deep learning methods [1–7]. However, all of these methods do not release the datasets that they have used for training and testing, with the exception of the World Cup dataset for soccer games [1]. Therefore, a novel dataset is required in order to develop new sports field localization techniques.

In this paper, a new annotation tool for collecting homographies from frames of hockey broadcast videos. It relies on annotating corresponding points on each frame and a model of the overhead view of the ice surface. With this tool, we have collected a dataset of frames, each of which has a corresponding homography and time in the sequence.

2 Related Work

In the literature, each method for performing sports field localization tends to come with its own sports field localization dataset. These datasets, with the exception of the World Cup dataset [1], have not publicly been made available.

Several sports have been the focus of field localization techniques. Datasets described in the literature include volleyball [6], basketball [4], and hockey [5]. Table 1 compares the datasets used for sports field localization methods.

The only publicly available dataset for this problem, World Cup Soccer, has 395 annotated frames. Each frame has an associated homography, but no temporal information (i.e., the frames have been randomly collected from broadcast footage) [7].

All methods report that their datasets have been annotated with point correspondences, and the associated homography determined with the DLT algorithm. To our knowledge, no annotation tool for sports field localization has been released.

Tarashima describes an anotation method for their basketball dataset, whereby only pre-specified intersections of lines on the playing surface and corresponding points on overhead playing surface model are annotated [4]. Citraro *et al.*describe a semi-automated method that was used for their datasets [6]. The annotation tool automatically tracks keypoints and the user provides corrections as needed.

3 Annotation Tool

We have developed a tool for annotating point correspondences between frames in NHL broadcast footage and a standard model of the rink surface. The tool was developed in Python and released as an executable for use by analysts at Stathletes.

To use the tool, the user selects corresponding points, alternating between the frame and the rink model (Fig. 2).

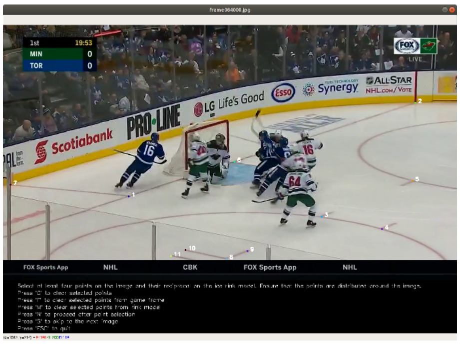
Users are instructed to select as many points as possible. The best points (i.e., highest precision) are located at the intersections and ends of lines on the playing surface. For example, the intersection of the goal line and boards, the intersection of the hash marks and faceoff circle, and the base of the goal post. The users are also able to zoom in and out to ensure that they are clicking on the correct position.

After the user has selected at least four corresponding points, the tool then displays the results of that homography warping. The frame is warped and overlayed on top of the rink model and the rink model is warped on top of the frame (Fig. 3).

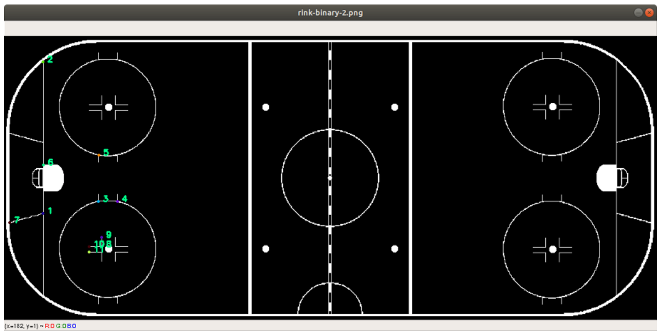
The homography is determined with the direct linear transform (DLT) algorithm with RANSAC for outlier rejection. We have tuned the RANSAC parameters for the highest visual accuracy, while requiring the fewest point annotations.

The user can then accept, reject, or edit their annotations after viewing the result of their annotations. The output is stored in JSON format, which is easily transferred once the annotations are complete. The output is further verified by the University of Waterloo team.

While Stathletes analysts have been using the annotation application, they have provided feedback and suggestions to the University of Waterloo developers. These suggestions have allowed for the annotations to be more accurate and the analysts to process more frames faster. For example, we added keyboard shortcuts to allow for the user to quickly delete pairs of annotations that add noise to the homography estimate. When the user reviews their annotation and choose to redo the selected points, rather then having to start the annotation from the beginning, they are able to alter the points they have already selected.

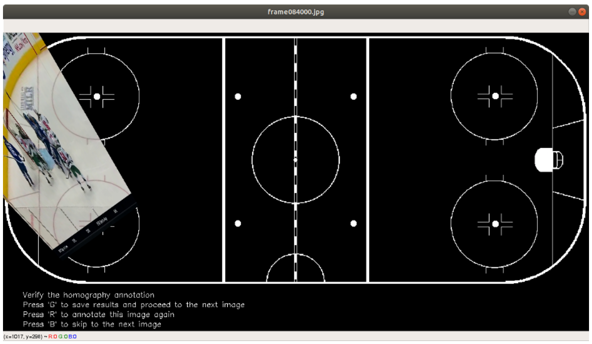


(a) Points annotated on a frame from a hockey broadcast video.

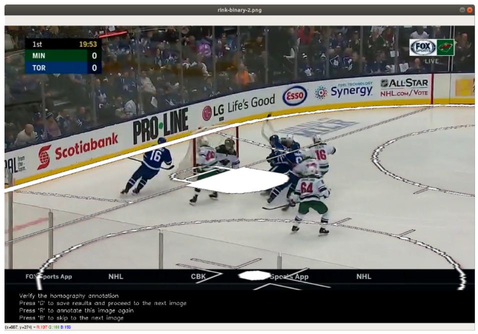


(b) Corresponding points annotated on the overhead model of an NHL rink.

Fig. 2: Annotation of points on the frame and ice rink model.



(a) The warped frame overlayed on the ice rink model.



(b) The warped ice rink model overlayed on the frame.

Fig. 3: During the verification stage, the homography calculated from the point correspondences is visualized by warping the frame and ice rink model.

During annotation, the area of the ice that was the most difficult to annotate was when the area around either blue line filled most of the frame. The lack of high fidelity keypoints made it difficult to obtain an accurate annotation. A feature was added to allow the users to draw guidelines in order to obtain accurate annotations in areas of open ice.

The average time to annotate each frame is 90 seconds.

4 Description of Dataset

The hockey homography dataset collected with the annotation tool has 4,262 annotated frames. During the annotation process, 178 frames were rejected due to the inability to get a satisfactory annotation.

The frames are collected from 15 separate game broadcasts from the 2018-19 NHL season. 2 to 3 shots, or sequences of dynamic game play, were extracted from each game, each lasting approximately 30 seconds, for a total of 42 shots. The videos were sampled at 1.5 fps. The frames come from several different broadcasters, which means that the overlayed graphics are not the same across the dataset. A variety of home teams also means that there are different designs (e.g., team logo) embedded in the ice.

5 Future Work

Due to the nature of homography annotation with point correspondences, there is no true way to get the homography transform for the rink. In the proposed solution, the output from point correspondences is verified visually, which means that there is potential for human error. Furthermore, while rejecting or accepting annotations is mostly clear, there are still many frames that are borderline on the quality. These tend to be frames where the blue line is mostly visible in the frame. There were 937 frames that were marginally accepted. More work could be done in this area to improve annotations in areas with fewer keypoints.

Future work with this homography annotation tool would be to expand the size of the dataset. While this dataset is relatively large, compared to other sports field localization datasets, a larger dataset would potentially allow for better solutions.

Another area of research would be to augment the tool with line segmentation. This would potentially allow for more precise ground truth annotations.

6 Conclusion

We have developed a homography annotation tool for frames from hockey broadcast video to an overhead rink model. The tool determines the ground truth transform through point correspondences of landmarks on the ice. We also have collected a dataset of 4,262 hockey broadcast homographies. Further work can be done to augment the dataset and collect more precise annotations.

Acknowledgments

This work was supported by Stathletes through the Mitacs Accelerate program.

References

[1] N. Homayounfar, S. Fidler, and R. Urtasun, "Sports field localization via deep structured models," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5212–5220.

[2] R. A. Sharma, B. Bhat, V. Gandhi, and C. Jawahar, "Automated top view registration of broadcast football videos," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2018, pp. 305–313.

[3] J. Chen and J. J. Little, "Sports camera calibration via synthetic data," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0.

[4] S. Tarashima, "Sflnet: Direct sports field localization via cnn-based regression," in *Asian Conference on Pattern Recognition*. Springer, 2019, pp. 677–690.

[5] W. Jiang, J. C. G. Higuera, B. Angles, W. Sun, M. Javan, and K. M. Yi, "Optimizing through learned errors for accurate sports field registration," in *The IEEE Winter Conference on Applications of Computer Vision*, 2020, pp. 201–210.

[6] L. Citraro, P. Márquez-Neila, S. Savarè, V. Jayaram, C. Dubout, F. Renaut, A. Hasfura, H. B. Shitrit, and P. Fua, "Real-time camera pose estimation for sports fields," *Machine Vision and Applications*, vol. 31, no. 3, pp. 1–13, 2020.

[7] L. Sha, J. Hobbs, P. Felsen, X. Wei, P. Lucey, and S. Ganguly, "End-to-end camera calibration for broadcast videos," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 13 627–13 636.