ATHLETE TRACKING IN MONOCULAR SPORTS VIDEO

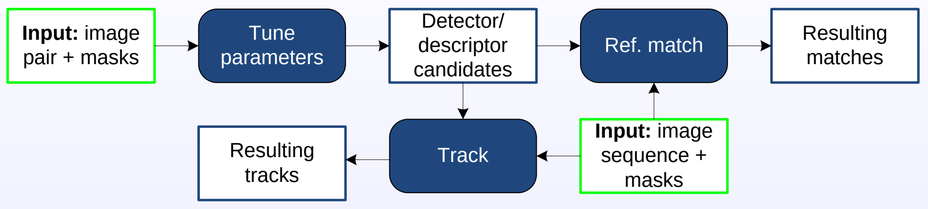
Fastovets M., Guillemaut J.-Y., Hilton A. - University of Surrey
{mykyta.fastovets, j.guillemaut, a.hilton}@surrey.ac.uk

Abstract

In this work we consider the task of tracking athletes in monocular sports sequences. We evaluate sparse feature tracking approaches and propose an evaluation framework to test the suitability of existing sparse feature detectors and descriptors for this difficult problem. The framework consists of two stages: parameter tuning and tracking evaluation. Preliminary results indicate that readily available detector-descriptor pairs are able to provide meaningful information in this scenario.

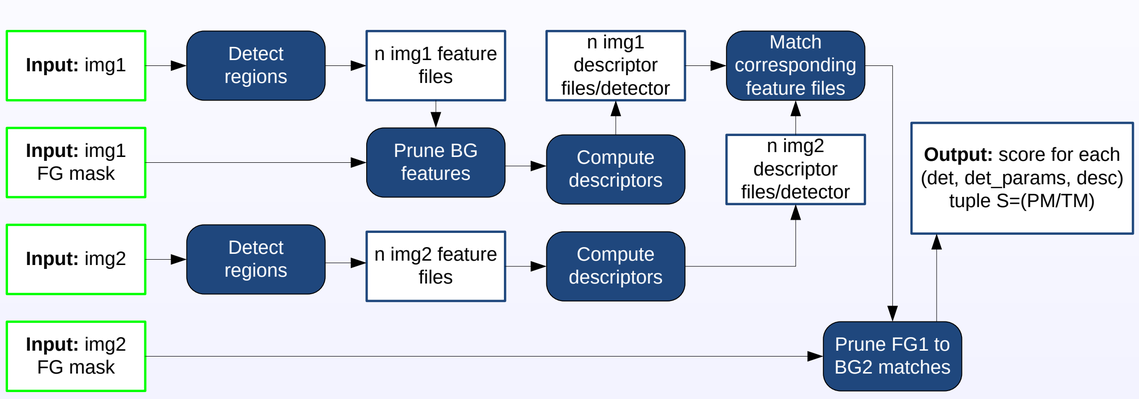
Framework Layout

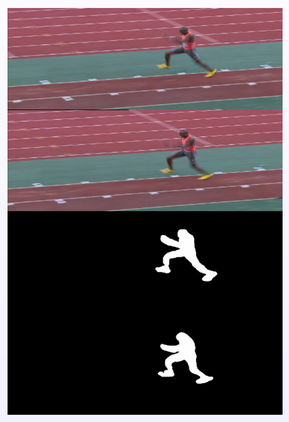
The objective is identifying good candidate detector/descriptor pairs for a given image sequence. We first handle the parameter tuning for each detector/descriptor pair, before selecting the best few to process our input data with.



Parameter Tuning

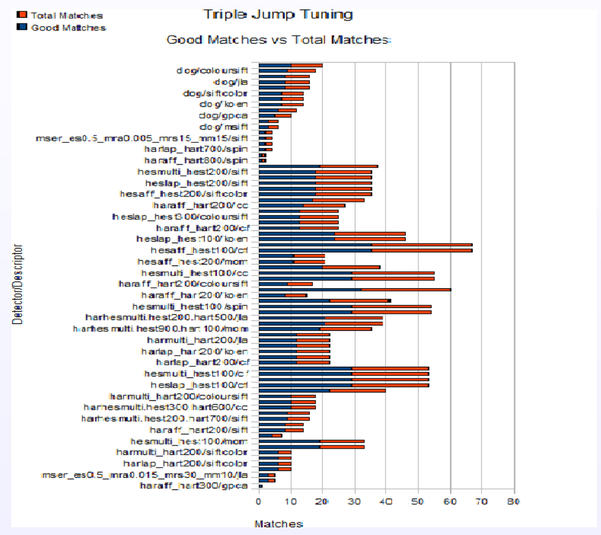
The first step is to tune the parameters of each detector/descriptor pair to be optimal for a particular dataset. The score is a ratio of "good" feature matches to the total number of matches. Good matches are ones which fall within the foreground of the image. Below is a an example of a sample input and the flowchart for parameter tuning.





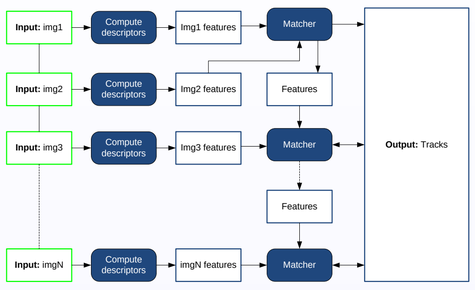
Tuning Results

In this particular case Hessian affine and Hessian-Laplace are the best detectors, as shown below.

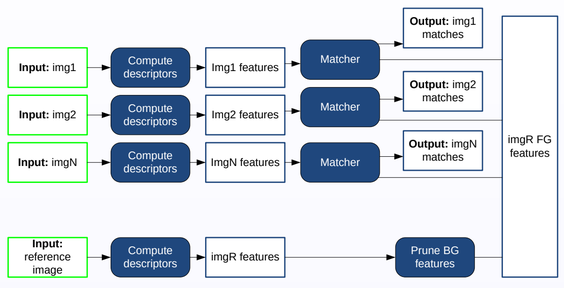


Tracking

We explore two ways of tracking. One that builds a model of the feature as the algorithm analyses the sequence, and the other a simple match to a static frame. The flow charts below demonstrate both principles.

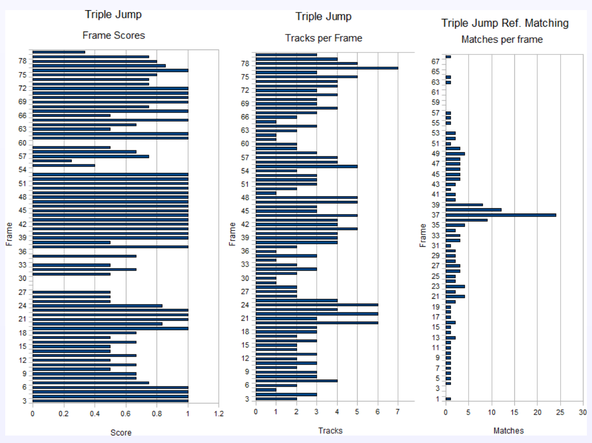


TRACKING



REFERENCE FRAME MATCHING

Once tracking has been performed we evaluate the results using the same method of scoring by ratio of true positives to the total number of features. Below is a sample of the results.



The procedure has been tested on a variety of sports sequences including gymnastics, triple jump, long and short runs. Standard and high definition footage was experimented with. Overall much better results are obtained using the tracking algorithm.

Conclusions

The framework produces an objective score for each detector/descriptor pair for a particular dataset. Experiments have shown that it is possible to obtain reasonable trajectories for some parts of the athlete's body using readily available methods. Qualitatively, feet, heads and knees produce strong feature trajectories, but their robustness is yet to be tested. Reference frame matching only produces good matches within several frames of the reference Additionally, there is no one best detector/descriptor pair as our results indicate, as each pair's success depends on the data. Future work will focus on exploring how, and if, some of the strong tracks (feet/heads, knees) can be used to track the athlete's motion.

The authors would like to thank the BBC for funding this research.

References

[1] K. Mikolajczyck, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *International Journal of Computer Vision*, 2005

[2] K. Mikolajczyck, C. Schmid. A performance evaluation of local descriptors. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2005