

Abstract

In Computer Vision, finding simple features often entails filtering a 2D image to find basic patterns. Often these are used for tracking or co-ordination (e.g. in stereo vision, or the motion vectors enabling MPEG compression), and requires a method for measuring tracking performance, usually through repeatability of simple points of interest.

Interest points are used regularly in Computer Vision in various applications. However, those applications, along with their requirements, are becoming increasingly complex and demanding. 2D interest point detectors (such as the Harris detector) are now regularly applied to 3D environments with varying degrees of success. With forays into automated optimisation of 2D feature classifiers via GP, there is a demand for better evaluation approaches that are more relevant to 3D, but the disconnect between 2D and 3D means they are still tightly coupled to their respective domains. 2D interest points are, by their design, poorly optimised for 3D, or even unoptimisable, due to their inability to work with, or awareness of, scene depth. Currently, evaluation of 2D-based interest points has innovated little in recent years, with research still highly dependent on image datasets and approximations of 3D, and little, or no reliable ground truth.

Interestingly some approaches prove effective even though not specifically designed to find interest points, notably the Fast and Harris detectors. Since these are effective in 2D images of 3D scenes, it seems they must be capturing some kind of 3D information, but this raises the question of how optimal they are. Testing 2D-based detectors in a real-world environment is not normally possible though due to their inability to be properly assessed, but a virtual scene can help bridge this gap. The concept of virtual spaces are seeing greater usage as they can address problems where a real world environment can't be used. Through the use of a virtualised ground truth, it is possible to probe aspects of the real world to solve problems or gain insight, where in normal situations, it would not be possible.

This thesis seeks to utilise virtual 3D spaces to bridge the gap between 2D detectors and 3D scenes, by emulating performance evaluation of 2D interest points with virtual spaces. A virtual ground truth can evaluate features detected based on 2D interest points in a simulated 3D space, for realistic evaluation of repeatability performance. By doing so, a virtual scene is able to utilise more

sophisticated evaluation strategies like ROC, and informedness, for correct feature classification. This enables 2D feature detectors to be properly evaluated, and potentially optimised, when tested for real world applications in virtual 3D spaces.

To test these new approaches, a virtual space is used to emulate well known repeatability evaluation via 2D and Euclidean space to find closest points. This is tested using conventional 2D detectors, as well as GP-based optimisation of 2D classifiers. Both images sets, and 3D-scanned model datasets are used in testing, as well as comparing the performance of 2D, 3D, and color optimisation approaches with GP. Additionally, we demonstrate that the use of virtual spaces enables other types of evaluation approaches like “informedness” which can similarly evaluate performance based on ϵ thresholds. Informedness demonstrates that it can incorporate more information about classified features identified in 2D, and can more effectively evaluate classifier performance due to the virtual ground truth. Our tests also empirically support that optimisation with depth data not only optimised 2D classifiers in virtual spaces comparatively better than without it, but also supports the argument that certain well accepted conventions regarding interest point repeatability and the Moore neighbourhood ($\epsilon = 1.5$) are not necessarily the best best performance tradeoff.