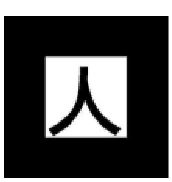
Matthew Walmer – mwalmer3

CIS II – Seminar Presentation – Critical Review

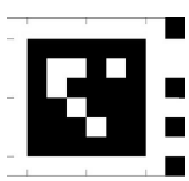
Publication:

Zhang, Xiang, Stephan Fronz, and Nassir Navab. "Visual marker detection and decoding in AR systems: A comparative study." Proceedings of the 1st International Symposium on Mixed and Augmented Reality. IEEE Computer Society, 2002.

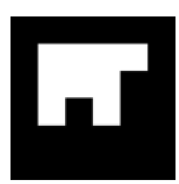
In “Visual Marker Detection and Decoding in AR Systems: A Comparative Study”, Zhang, Fronz and Navab compare four visual marker systems, namely ARToolKit (ATK), Hoffman marker system (HOM), Institut Graphische Datenverarbeitung marker system (IGD), and Siemens Corporate Research marker system (SCR). There is some initial risk of bias as this paper was published by the Augmented Reality Group at Siemens Corporate Research; however, their representation of the various marker systems does not seem to unfairly favor the SCR system. Instead, they try to highlight each system’s strengths and weaknesses, and in the end they find that no system was universally best in all of the metrics they defined. They do run into some difficulty running multi-marker tracking with IGD, however, they humbly take fault for this, saying “we believe the functionality was implemented and blame the failure to our own unfamiliarity to the IGD system.” While this humility is admirable, it still means this paper lacks certain data about IGD’s performance. At the end, the researchers provide an excellent qualitative analysis to guide users to the markers system which best suits their application.



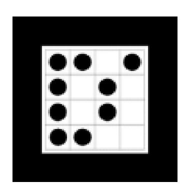
(ATK)



(HOM)



(IGD)



(SCR)

The researchers define four properties which they evaluate for each system: usability, efficiency, accuracy and reliability. The first property, usability, describes how easy it is to use the system, as well as the range of platforms the system can be used on. Of the four, this is the only metric which must be described qualitatively. There are some problems with this metric. For SCR they cannot provide an unbiased perspective of the "ease of use" of said system because they are already affiliated with Siemens Corporation and presumably would have prior experience with SCR. They state that of all the systems, ATK was the most accessible and the best documented, however, they do not make any mention of the documentation of the other systems. Another component of usability they address is each marker's encoding system. IGD, HOM and SCR use a systematic coding system, where each marker has a binary grid pattern that can be interpreted to give a marker a value. ATK allows for more flexibility. Rather than using a grid, the inside of an ATK marker can use any custom pattern for identification. While this may sound good, the researchers reasonably identify this as a drawback, arguing that ATK requires an extra step to "learn" the pattern, while the other systems can generate thousands of distinct grid patterns without requiring any extra effort to train. For this reason, they claim the other systems are more appropriate for large scale applications using hundreds or thousands of markers.

Their definition of efficiency is the time taken to recognize and decode markers. This was the most straight-forward of their metrics, and it was tested simply by timing the programs on the same hardware setup. ATK was the fastest for single and multi-marker detection and decoding. The researchers believe this is because ATK looks at a binary image, which is computationally faster to work with. If this study focused only on the speed of the programs, then it would give a very lacking view of these programs' performance. The researchers' later tests show that ATK's speed comes at a cost of both corner precision and marker identification. They also use these timing tests to highlight one of SCR's unique features. SCR has a "tracking" mode which uses a marker's position in the previous frame to help find it in the next frame. In this mode, SCR is much faster than any of the other systems, but it is limited to tracking one marker at a time.

Accuracy describes both the precision in identifying a marker's corner points as well as correctness in identifying markers. For their accuracy experiments, the researchers only measured the 2D pixel error in identifying the marker corner positions in each image. It would have been very useful to know the error in the 3D pose of the marker; however, the researchers could not devise a way to determine a ground truth for this value. To determine a 2D ground truth, the researchers used two different methods to identify the "correct" corner positions for each image. One method used OpenCV to detect corners (OCV), while the other used edge detection and line intersection (LIT). It is problematic that they do not pick one ground truth for their corner points, because the two sets give different results. SCR performed the best with respect to the LIT points, while IGD was best with the OCV points. The lack of clarity in establishing a single ground truth for the corner positions makes it difficult to draw conclusions about which system is most accurate. However, for either set, ATK was the least accurate. The researchers theorize that this is because of the binary image processing method ATK uses. Another component of accuracy testing was correct identification of markers in multi-marker images. They varied both the angle of view and the number of markers in each frame. They found that HOM performed the best at marker identification. A noteworthy discovery was that ATK has a tendency to confuse similar looking markers. This is because ATK uses a fairly simple template matching method to give fast identification. So even though ATK's custom marker system could be used to generate an unlimited number of unique markers, ATK is limited by how well it can identify similar looking markers.

The last metric, reliability, questioned how each tracker performed "under various unfriendly conditions". The researchers did a good job selecting a variety of challenges that the trackers would face in real-world applications, including low angle views, poorly focused images, and far-away markers. It would have been nice if they also performed tests with variable light levels and low contrast images. Multi-marker tests were also considered to be one of the "unfriendly conditions". These tests demonstrated that HOM's confidence value was very reliable and correlated reasonably with the difficulty of the test. ATK's confidence metric was very questionable, as it had higher confidence on the most unfocused images.

Overall, the researchers performed a good range of experiments to demonstrate the pros and cons of each marker tracking system. There are a few tests that are sadly missing, like variable lighting and accuracy of 3D poses. In their final qualitative discussion, the researchers summarize their results for each system. While their analysis for each system does touch on each of the four criteria they established, it does not adhere very rigidly to them. It would be nice if they had given a more concrete score, or a ranking of which system was best in each of the properties. Even though they don't give definitive scores, the qualitative discussion they provide for each marker system is still very useful for selecting a marker system.