1 Diploma Thesis: "Adaptive Motion-Based Thresholding of a Natural Feature Tracking Application with Hybrid Tracking Techniques"

Supervisors:

- Prof. Gudrun Klinker (Technical University Munich)
- Prof. Mark Billinghurst (Hitlab New Zealand)
- Martin Wagner (Technical University Munich)

Introduction

A major problem in augmented reality systems is the calculation of the users viewpoint. The viewpoint is needed to display virtual information into the user´s view. Several image processing methods and toolkits have been introduced. This work deals with the problem that these image processing technologies, like for example the ARToolkit oder Natural Feature Tracking, are very error-prone to movements. The approach of this thesis is to use additional orientation information to configure an augmented reality application using a natural feature tracking algorithm during runtime to improve the performance.

Brief overview of the used technologies

- ARToolkit
  The ARToolkit is a well-know software for building Augmented Reality systems. Black and white markers are printed out and placed into the environment. Based on marker recognition the viewpoint of the user is calculated and a virtual 3D model is displayed on it. This technology is used in many application domains like collaborative work, pose estimations, maintenance, etc ...

  - **Example Application: AR Volcano**
    AR Volcano is an application in the education domain and gives an introduction to the development of volcanoes. A physical-book has ARToolkit markers on each page. A user equipped with a hand hold device can explore the volcano scenes and turn the page for the next scene.

- Natural Feature Tracking (NFT)
  The basic idea of NFT is not to use markers for the registration of the virtual models. Any 2D plane or even 3D objects can be used to calculate the viewpoint of the camera and to display the 3D model. Advantage is that there is no need to configure the environment with markers. A newer version of the ARToolkit uses NFT for 2D planar scenes. All the research done in this thesis is based on this version by H. Kato and M. Billinghurst

([1]). Images with certain properties can be used now instead of markers. Only a small marker in the image is needed for the initialization of the scene (see figure 1).

Figure 1: A 2D planar image configured for NFT

The NFT algorithm initializes itself with the common ARToolkit marker detection. After that it chooses feature points in the image for the calculation. These feature points, matching certain requirements, must be extracted of the image in a preprocessing step.

- **Example Application: Magic Book**
  With a handhold device a user is able to navigate through a fairy tale landscape. This device consists of a camera and a head mounted display (HMD) (see figure 2) For the navigation the user can move into the scene or rotate to observe certain objects in the landscape.

- Inertial Tracking
  Inertial trackers are used to measure the object position and orientation. While accelerometers are used to track the position, gyros measure the orientation of the tracked object. In this thesis only a 3 degrees of freedom (3 DOF) tracker is used (Intertrax2 by Intersense). This rotation around the perpendicular axis are also called pitch, yaw and roll. This means that the tracker only gives information about the object´s orientation (by using a gyro). But there are trackers available providing full 6 DOF.

- Magnetic Tracking
  Magnetic Tracker use a magnetic field to measure position and orientation of an object (6 DOF). These fields can either be a low

Figure 2: A user holding a hand held device

frequency AC field or a pulsed DC field. In this work a Ascension Flock of Birds system will be used.

Template Matching Algorithm of the ARToolkit

NFT uses a Template Matching algorithm that estimates the similarity between two pixels. This is necessary to find a feature point in the next frame again. However every "Magic Book" user moves differently through the scenes. But there are constraints to movements caused by the template matching algorithm. Under the assumption that changes between two single video frames are very small the feature point is searched in the next frame within a certain search window. This window must be very small to keep computation time within reasonable limits. This is constrictive to fast movements at all. Therefore the usability of NFT applications using the toolkit is restricted.

The idea now is to adjust the search window size due to the movements of the handhold:

- Low movement
  If the system can predict a low movement, a small search window size is best suited. This will speed up the computational expensive template matching algorithm.

- High movement
  If we would keep the search window size constant, the movement might be too fast to detect the feature point in the next frame. If we are able to enlarge the search window the point might be in the larger window. If the system loses the orientation it has to be reinitialized through looking at the initial marker again. This should be avoided.

Hybrid Tracking Approach: Combining Inertial and Vision Based Tracking

The combination of inertial and vision based tracking is quite an important topic in the research area of Hybrid Tracking. If the vision based tracker loses orientation the inertial information is used until the vision based tracker delivers reliable information again. This is also used for a predictive tracking approach. Through a prediction- and correction loop the error-prone inertial and vision based trackers are refined. This leads to a better tracking performance.

The idea for the NFT configuration is not that complicated:

To estimate the movement of the user a inertial sensor could be attached to the handhold. The information given by the inertial is evaluated and used for the adaptive configuration of the search window size. Therefore an interpretation of the rotation data given by the inertial is needed.

Interpretation of the Inertial Data with a User Study

A first approach would be to study how the user actually uses the Magic Book. This would give information about the common movements and rotations. This data could be used to figure out a mapping between inertial data and search window size:

$$InertialData \Rightarrow Movement \Rightarrow SearchWindowSize$$

To perform the user study a magnetic tracking system is used. This tracking system allows a 6 DOF tracking, though it is also possible to get translation information of the movements. It might be possible to observe a relationship between translation and rotation for the Magic Book.

The data of the user study is now used to configure the interpretation of the inertial data. It might also be interesting to work with user profiles caused by the fact that user moves very differently through the scene.

A first idea is too look at the absolute angels and the angular velocities of the movements. Figure 3 shows an example record gained by attaching a inertial tracker to the hand hold device. The chart only shows the pitch-rotations. At a first glimpse the behavior of the angular velocity seems to be interesting. Although not shown in the chart (for simplification) the dependencies of pitch, yaw and roll might be interesting.

In a first step different measurements can be compared to get a feeling about the movements occurring. Fast movements will result in a higher angular velocity peak. Some users might rotate the device around two or more axes at a moment other might rotate only around a single axis. Then movements can be categorized (low -moderate -fast movements). After that experiments have to be made to find the mapping between movements and search window size.

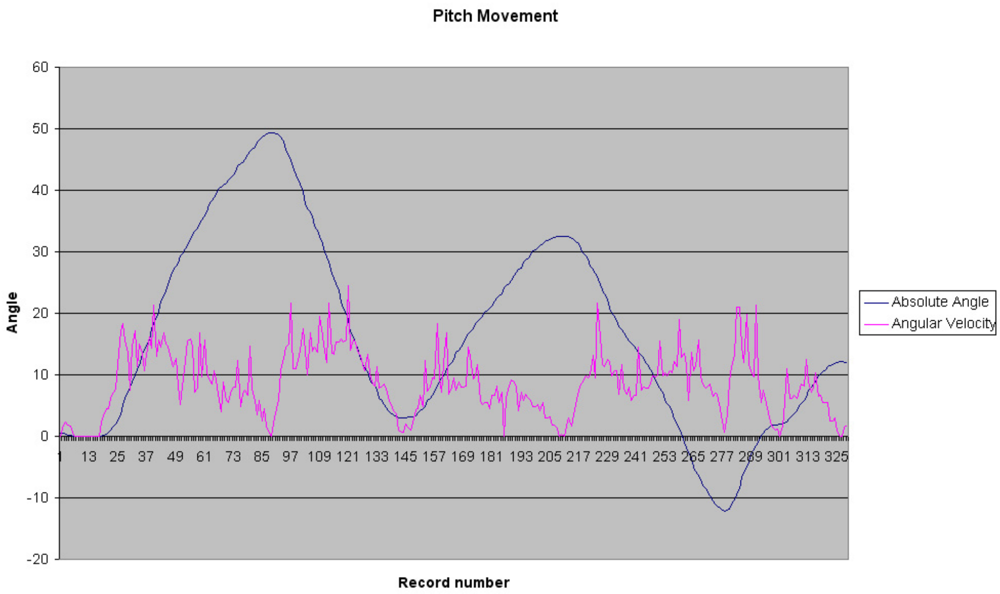

Figure 3: Example Pitch Measurement of Hand Hold Rotation

Evaluation of the dynamic configuration

Hopefully the adaptive thresholding of the search window size leads to a better performance of the Magic Book. So the performance of a fixed and an adaptive search window size has to be compared and conclusions have to be drawn.

Conclusions

Outcome of this work should be an answer to the question if the performance and the robustness of applications that adopt their behavior to motions can be improved with the approach described above.

Other ideas might be evaluated concerning the improvement of the NFT-ARToolkit. The inertial approach considers two main aspects on the performance of the system: robustness and computation speed. A trade-off between these two factors always has to be made. Other approaches to speed up the template matching algorithm, like numerical approaches can be discussed here.

Also other interesting topics that came up during the research phase will be discussed.

References

[1] H. Kato, K. Tachibana, M. Billinghurst, and M. Grafe. A Registration Method based on Texture Tracking using ARToolkit. In *IEEE Intl. Augmented Reality Toolkit Workshop*, 2003.