Augmented Reality - VU WS 2022/23

Camera Pose Estimation

Fuzail Palnak
fuzail.palnak@student.tugraz.at
Matriculation Number - 12129320

November 29, 2022

Graz

1 Description

The application comprises of four steps mainly,

- Feature Detection and Description
- Feature Matching
- Pose Estimation
- Rendering

In **Feature Detection and Description**, features from reference frame, along with its probable match on the input frame are detected, which is then followed by a **feature matching** step, as this features are plenty and also have false positives, a matching step is carried out to reduce the number of false positive. The next step is **pose estimate**, here we assume the matched features are not free from outliers, therefore, RANSAC is used to find a good estimate, which is not influenced by outliers. Finally, a 3D object is **rendered** using the estimated pose from the previous step.

2 Motivation for selecting the algorithms

ORB[1] feature detector and descriptor along with FLANN[2] feature mature did not yield satisfactory results (1), while SIFT provided better results (2) in contrast to ORB, hence, SIFT detector with FLANN matcher was chosen.

2.1 Other reasons for choosing SIFT

SIFT is rotation and scale invariant and the features are robust to occlusion and clutter. The descriptor is robust to typical variations in viewing conditions. All these properties allow for better image matching.

Figure 1: ORB with FLANN

[1] https://docs.opencv.org/3.4/dc/dc3/tutorial$_p y_m atcher.html$
[2] https://docs.opencv.org/3.4/dc/dc3/tutorial$_p y_m atcher.html$

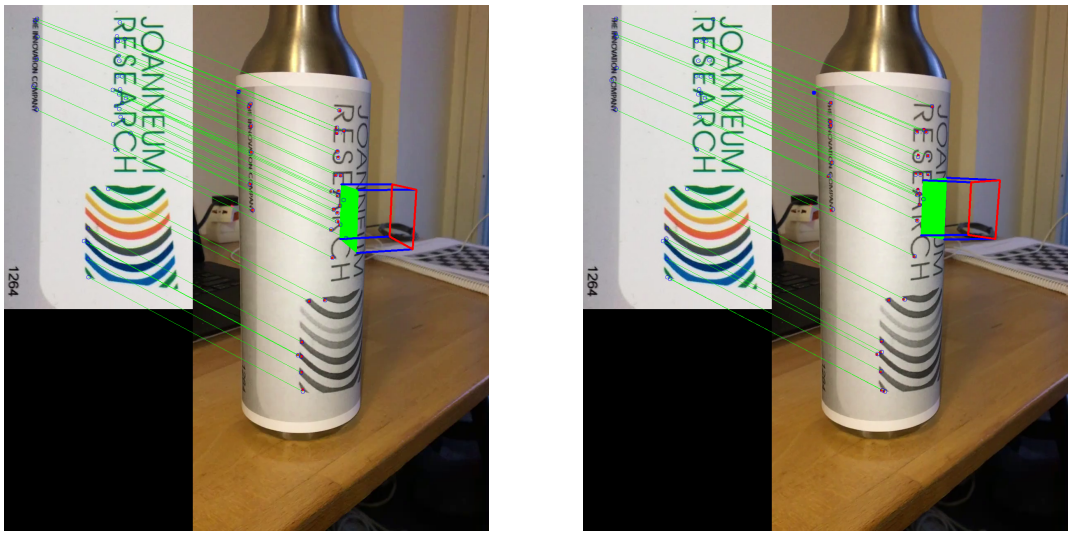

Figure 2: SIFT with FLANN

3 Target

3.1 Why did you chose the particular target?

The chosen target (3) has natural features all around it, meaning, you are able to find minimum set of features to estimate the pose from all different perspective. The features in the target are not isolated in just one region.

Figure 3: Target Image

3.2 Which kinds of targets would perform differently (better/worse), and why?

A Target must posses following properties to be considered for tracking.

1. **Repeatability** - Features should be stationary and should not change over different frames
2. **Distinctiveness** - Features should be informative
3. **Locality** - Same local features can be found even with view point changes
4. **Quantity** - Sufficiently large number of features are present on the target

Targets where the results can be worse.

1. Targets with features isolated in just one region, are bad, as it would get challenging to detect these features as the camera starts to move.
2. Targets which have similar surface and characteristics throughout, with no edge changes, color changes are a bad choice.
3. Target which are not stationary and keep moving
4. Target which change properties that could affect the detection pipeline are also a bad choice

4 FPS

Rendering a video shot at 720p with SIFT resulted in a processing time of 2 frames per second. The processing time was significantly better for lower resolution videos.