Development of an Augmented Reality System

Students should develop a software system aimed at creating an Augmented Reality (AR) video, as illustrated in the two images below:

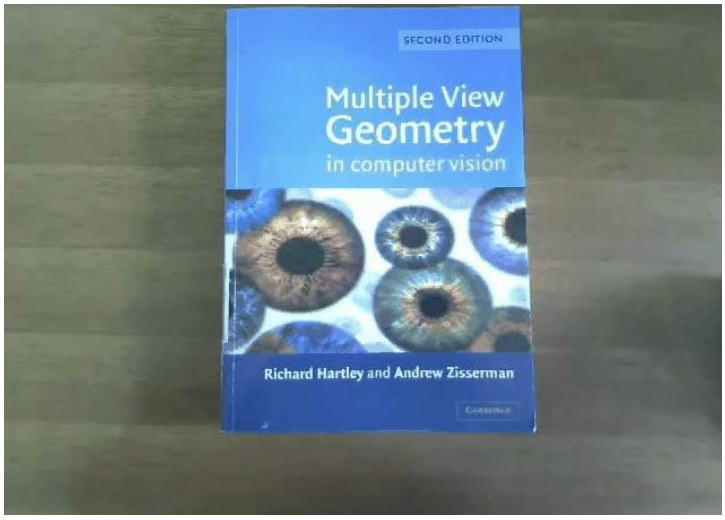

Figure 1: Original Video

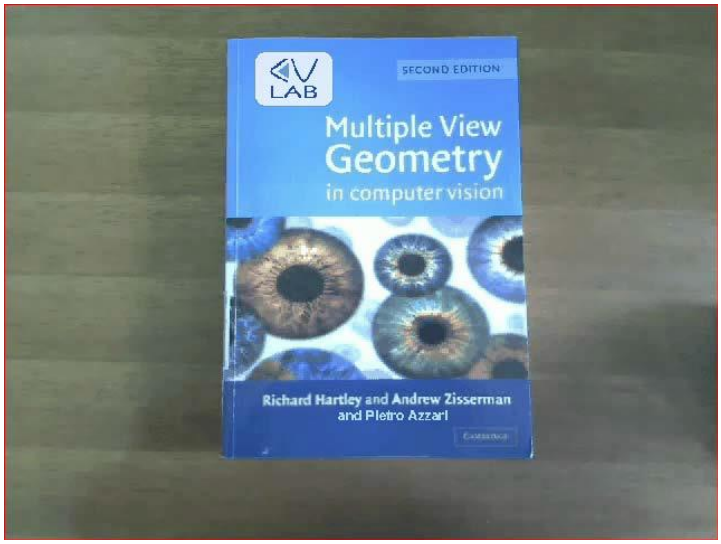

Figure 2: Augmented Video (CVLab logo and third author)

- **Characteristics of the video sequence**

The original video sequence, i.e. *Multiple View.avi*, depicts the well-known book "Multiple View Geometry in Computer Vision", by Richard Hartley and Andrew Zissermann, as seen by a moving camera. Initially, camera trajectory is a slow translation, followed then by rotations and slow as well as rapid brightness changes.

- **Functional specifications**

Given the input video, students should superimpose an augmented reality layer onto the book cover as realistically as possible, so that, ideally, when watching the augmented video one would perceive the overlaid graphics as real items present in the book cover. Purposely, the colour artefacts appearing when superimposing the graphics layer should be eliminated.

- **Input Data**

The augmented reality system should process the following input data:

- Input video sequence to be augmented: *Multiple View.avi*
- Reference frame (first frame of the input sequence): *ReferenceFrame.png*
- Binary mask that identifies the pixels belonging to the book in the reference frame: *ObjectMask.png*
- Image containing the augmented reality layer, made out of the CVLab logo and a different list of authors (i.e. Richard Hartley and Andrew Zissermann and Pietro Azzari) together with the associated binary mask: *AugmentedLayer.png, AugmentedLayerMask.png*

- **Output Data**

The developed software system should produce the following output data:

- A video sequence, referred to here as *Augmented Multiple View.avi,* containing realistic superimposition of the augmented reality layer onto the input video sequence.

SUGGESTIONS

Firstly, it is worth observing that the mapping between the augmented reality layer and the reference frame (first frame of the input video) is the identity. Therefore, if the transformation between the current and first frame is found, such a transformation can then be applied to project the augmented reality layer onto the current frame. The main issue to be dealt with consists in finding enough (i.e. ≥ 4) corresponding image points between pairs of video frames. Suitable approaches are feature detectors (e.g. Harris, KLT, SIFT available in OpenCV), template matching, segmentation and tracking of distinctive image regions, such as e,g. the irises.

The registration between frame pairs may occur either between successive frames (referred to here as Frame to Frame, F2F) or between the current and the reference frame (Frame to Reference, F2R). The F2F approach implies a simpler image correspondence task, as successive frames tend to look quite similar, but make it more difficult to determine the correct transformation required to overlay the augmented reality layer due to such a layer being defined with respect to the reference frame. On the other hand, the F2M approach allows for relying on the current transformation to project the augmented reality layer, although it turns out harder to establish correct image correspondences between the current and reference frame as they may look quite diverse.

It is suggested that students start the project according to the F2R method and then, perhaps, they also implement the F2F approach so to compare results.