Motion-based object segmentation in video sequences

SUMMARY

This thesis addresses unsupervised video object segmentation based on motion information. As the motion information is implicit in a video sequence, we focus on acquiring, analyzing and utilizing motion data from a given video sequence. This research focuses on processing video and image data at different levels, using techniques and approaches from different fields. Chapter 2 and Chapter 3 give the introduction of related research fields, as well as the involved methodologies and techniques. Chapter 4 presents a general framework for addressing the research problem. In Chapter 5, we address the low-level image processing and discuss three types of motion data that can be extracted from a video sequence. The obtained motion data represent the 2D motions in the video frames. A motion segmentation algorithm is presented in Chapter 6, to segment the video frames based on the 2D motion data. In Chapter 7, we analyze the 3D motion consistency implied in the obtained 2D motions. Chapter 8 investigates how to segment non-moving objects based on information that was learned when the objects were moving. Chapter 8 also discusses the object identification based on motion segmentation.

In Chapter 4, we analyze the general pipeline of processing video sequences, for the purpose of video object segmentation. We investigate how to extract the implicit motion information hinted in the changes of video frames without using prior knowledge. A framework is presented for the system design that combines a bottom-up and a top-down scheme. At the bottom-up step, we investigate the extraction of motion data from the lower-level image features and the utilization of the obtained motion data for grouping image features into higher-level object descriptions. The top-down step then learns to segment non-moving objects by using the obtained object descriptions as the prior knowledge. Based on the scheme, we present three fundamental modules for building up the video object segmentation system: one for motion data extraction, one for motion based video objects segmentation, and one for non-moving object segmentation. We describe the explicit definitions of related terminologies in this system, and discussed the keypoints to be noticed for each module. We also introduced the datasets used in our experiments.

Chapter 5 focuses on the extraction of motion data based on the low level image features. We extract and track some salient image features in a video sequence, and represent the motion data as their position changes in the video frames. We investigate two ways for motion data extraction: one focuses on estimating the motion of elementary image pixels, another takes some local image patterns as the feature points to be tracked. Three approaches are described in this chapter. One of them focuses on the pixel-based motion estimation by using the optical flow technique. The other two approaches both extracts SIFT features from the images and tracks their movements, but differ in the methods used for feature extraction. One of the feature-based methods only focus on a sparse set of features in the images, while the other one adopts a dense field of features for tracking. As a result, three types of motion data are obtained from our data sets. As these data are used as the input for next step, i.e the motion segmentation, we can not conclude that which is

a better choice for our system in this chapter.

In Chapter 6, we investigate the segmentation of moving objects given the obtained motion data. An unsupervised 2D based motion segmentation algorithm is proposed. This algorithm focuses on segmenting the pixels or feature points into groups that undergo the unique 2D motion, based on the given motion data. Based on the assumption that points from the same object undergo the same motion, the segmented groups can be regarded as objects in the given videos. We model the points' motion between two successive image frames by a 2D affine transformation, and use a classification EM algorithm to segment points into groups of the same motion. Bayesian update is used to propagate the segmentation from one frame pair to the next. Thus the segmentation of the entire video sequence is the combination of segmentation of successive frames. In this way, the segmentation algorithm takes the motion information from the beginning to the end of a video sequence into account. The algorithm also automatically estimates the number of objects. The method has high flexibility in dealing with different types of motion data. We measure the reliabilities of the segmentation results and propose an optional step for compensating the camera motion, which leads to a different version of the proposed algorithm. The experimental results on the three types of motion data, which are described in Chapter 5, show that the performance of the proposed method is comparable with the state-of-the-art methods. The results also suggest that the proposed method prefers the motion data obtained by the dense feature tracking approach, on which the performance is more outstanding.

Chapter 7 complements Chapter 6, by analyzing the inherent 3D consistency in the obtained motion data, which is essentially the 2D movements in the successive images of a video sequence. As the 2D motions in images are a projection of the 3D motions in the real world carried out by a camera system, the assumption that the motion of points of an object can be described by an affine transformation, can be violated and leads to over-segmentation. Two theorems are presented in this chapter, which describes the properties of a given set of points from an object and their projected movements between two images. These theorems can effectively measure the 3D rigid motion consistency of a group of 2D projection points, and also estimate the original 3D body movement with one degree of freedom. Experiments shows that the segmentation obtained by our algorithm presented in Chapter 6 can be improved by using the proposed theorems. Unfortunately the recovering of 3D motion models only works well for the rotation about the z direction in the camera system. Nevertheless, the proposed theorems show promising potential, and further research is worth to take in the future.

Chapter 8 addresses the non-moving object segmentation, when motion segmentation is no longer possible. Without prior knowledge, we use the motion segmentation results to learn object descriptions from moving objects. These object descriptions are used to segment objects when they are no longer moving. We present a static object segmentation method based on the SIFT matching strategy, which obtains the segmentation by comparing the image data with the learned object descriptions. Additionally, we investigate object identification given the segmentation results based on a sparse coding algorithm. It addresses the problem of identifying the same objects without using the motion clue, and evaluates the quality of the segmented objects. The experimental results show that the non-moving object segmentation relies on the quantity of learning data that is obtained by motion segmentation. With more images in the motion segmentation results, the non-moving object segmentation performs better. The results also show that the segmented objects can effectively represent the ground-truth, as they can be identified even with errors in the segmentation results.

Finally in Chapter 9, we discuss the results with respect to the research questions. We also present several directions for further research.