Over the last decade, the usage of unmanned systems such as Unmanned Aerial Vehicles (UAVs), Unmanned Surface Vessels (USVs) and Unmanned Ground Vehicles (UGVs) has increased drastically, and there is still a rapid growth. Today, unmanned systems are being deployed in many daily operations, e.g. for deliveries in remote areas, to increase efficiency of agriculture, and for environmental monitoring at sea. For safety reasons, unmanned systems are often the preferred choice for surveillance missions in hazardous environments, e.g. for detection of nuclear radiation, and in disaster areas after earthquakes, hurricanes, or during forest fires. For safe navigation of the unmanned systems during their missions, continuous and accurate global localization and attitude estimation is mandatory.

Over the years, many vision-based methods for position estimation have been developed, primarily for urban areas. In contrast, this thesis is mainly focused on vision-based methods for accurate position and attitude estimates in natural environments, i.e. beyond the urban areas. Vision-based methods possess several characteristics that make them appealing as global position and attitude sensors. First, vision sensors can be realized and tailored for most unmanned vehicle applications. Second, geo-referenced terrain models can be generated worldwide from satellite imagery and can be stored onboard the vehicles. In natural environments, where the availability of geo-referenced images in general is low, registration of image information with terrain models is the natural choice for position and attitude estimation. This is the problem area that I addressed in the contributions of this thesis.

The first contribution is a method for full 6DoF (degrees of freedom) pose estimation from aerial images. A dense local height map is computed using structure from motion. The global pose is inferred from the 3D similarity transform between the local height map and a digital elevation model. Aligning height information is assumed to be more robust to season variations than feature-based matching.

The second contribution is a method for accurate attitude (pitch and roll angle) estimation via horizon detection. It is one of only a few methods that use an omnidirectional (fisheye) camera for horizon detection in aerial images. The method is based on edge detection and a probabilistic Hough voting scheme. The method allows prior knowledge of the attitude angles to be exploited to make the initial attitude estimates more robust. The estimates are then refined through registration with the geometrically expected horizon line from a digital elevation model. To the best of our knowledge, it is the first method where the ray refraction in the atmosphere is taken into account, which enables the highly accurate attitude estimates.

The third contribution is a method for position estimation based on horizon detection in an omnidirectional panoramic image around a surface vessel. Two convolutional neural networks (CNNs) are designed and trained to estimate the camera orientation and to segment the horizon line in the image. The MOSSE correlation filter, normally used in visual object tracking, is adapted to horizon line registration with geometric data from a digital elevation model. Comprehensive field trials conducted in the archipelago demonstrate the GPS-level accuracy of the method, and that the method can be trained on images from one region and then applied to images from a previously unvisited test area.

The CNNs in the third contribution apply the typical scheme of convolutions, activations, and pooling. The fourth contribution focuses on the activations and suggests a new formulation to tune and optimize a piecewise linear activation function during training of CNNs. Improved classification results from experiments when tuning the activation function led to the introduction of a new activation function, the Shifted Exponential Linear Unit (ShELU).