Int J Comput Vis (2013) 101:350–366
DOI 10.1007/s11263-012-0580-1

Shadow Casting Out Of Plane (SCOOP) Candidates for Human and Vehicle Detection in Aerial Imagery

Vladimir Reilly · Berkan Solmaz · Mubarak Shah

Received: 22 January 2012 / Accepted: 22 September 2012 / Published online: 12 October 2012

Abstract In this paper, we propose a method for detecting humans and vehicles in imagery taken from a UAV. This is a challenging problem due to a limited number of pixels on target, which makes it more difficult to distinguish objects from background clutter, and results in much larger search space. We propose a method for constraining the search based on a number of geometric constraints obtained from the metadata. Specifically, we obtain the orientation of ground plane normal, the orientation of shadows cast by out of plane objects in the scene, and the relationship between object heights and the size of their corresponding shadows. We use the aforementioned information in a geometry-based shadow, and ground-plane normal blob detector, which provides an initial estimation for locations of shadow casting out of plane (SCOOP) objects in the scene. These SCOOP candidate locations are then classified as either human or clutter using a combination of wavelet features and a Support Vector Machine. To detect vehicles, we similarly find potential vehicle candidates by combining SCOOP and inverted-SCOOP candidates and then classify them using wavelet features and SVM. Our method works on a single frame, and unlike motion detection based methods, it bypasses the entire pipeline of registration, motion detection, and tracking. This method allows for detection of stationary and slowly moving humans and vehicles while avoiding the search across the entire image, allowing accurate and fast localization. We show impressive results on sequences from VIVID and CLIF datasets and provide comparative analysis.

Keywords Human detection · Vehicle detection · Aerial surveillance · UAV · Shadow · Metadata

V. Reilly (✉) · B. Solmaz · M. Shah
University of Central Florida, Orlando, USA
e-mail: vsreilly@eecs.ucf.edu

B. Solmaz
e-mail: bsolmaz@eecs.ucf.edu

M. Shah
e-mail: shah@eecs.ucf.edu

1 Introduction

Every year Unmanned Aerial Vehicles, or UAVs, are becoming more widespread in both military and civilian applications, including surveillance, rescue, and reconnaissance (Xiao et al. 2008a, 2010; Quaritsch et al. 2010). In the course of these operations video data containing useful information is collected. This information may be useful during the mission itself or may become useful at a later date. The ever-increasing number of UAV missions equates to a backlog of data which becomes quite large; thus requiring too many man-hours to analyze manually. This calls for automated video analysis tools whose capabilities include registration (Yahyanejad et al. 2010), object detection (Kembhavi et al. 2011), tracking (Xiao et al. 2010; Yilmaz et al. 2006), classification (Xiao et al. 2008b), and scene and event analysis (Cheng et al. 2006; Kluckner et al. 2009). In this paper, we will focus on detecting pedestrians and vehicles in UAV imagery. This, however, is a challenge. Problems include smaller object sizes, varying orientations, motion blur, and camera motion.

One straightforward approach to this problem, is to apply a state-of-the-art static frame detection algorithm such as Dalal and Triggs (2005), Felzenszwalb et al. (2008), Leibe et al. (2005), Mikolajczyk (2004), Sabzmeydani and Mori (2007). This approach, however, runs into the problem of small object size, which may make it impossible to construct a meaningful model. Methods that perform parts detection