Int J Comput Vis
DOI 10.1007/s11263-013-0648-6

Generative Methods for Long-Term Place Recognition in Dynamic Scenes

Edward Johns · Guang-Zhong Yang

Received: 11 October 2012 / Accepted: 25 July 2013

Abstract This paper proposes a new framework for visual place recognition that incrementally learns models of each place and offers adaptability to dynamic elements in the scene. Traditional Bag-Of-Words (BOW) image-retrieval approaches to place recognition typically treat images in a holistic manner and are not capable of dealing with sub-scene dynamics, such as structural changes to a building façade or seasonal effects on foliage. However, by treating local features as observations of real-world landmarks in a scene that is observed repeatedly over a period of time, such dynamics can be modelled at a local level, and the spatio-temporal properties of each landmark can be independently updated incrementally. The method proposed models each place as a set of such landmarks and their geometric relationships. A new BOW filtering stage and geometric verification scheme are introduced to compute a similarity score between a query image and each scene model. As further training images are acquired for each place, the landmark properties are updated over time and in the long term, the model can adapt to dynamic behaviour in the scene. Results on an outdoor dataset of images captured along a 7 km path, over a period of 5 months, show an improvement in recognition performance when compared to state-of-the-art image retrieval approaches to place recognition.

Keywords Scene recognition · Appearance-based localization · Topological localization · Image retrieval · Simultaneous localization and mapping

E. Johns (✉) · G.-Z. Yang
Imperial College London, London, UK
e-mail: ej09@imperial.ac.uk

G.-Z. Yang
e-mail: g.z.yang@imperial.ac.uk

Published online: 13 August 2013

1 Introduction

The recognition of a place instance depicted in an image has seen a wide range of applications including object retrieval (Arandjelovic and Zisserman 2012), loop closure, topological localisation and appearance-based Simultaneous Localisation and Mapping (SLAM) (Johns and Yang 2013a; Cummins and Newman 2009), 3D reconstruction (Agarwal et al. 2009) and building recognition for tourists (Zheng et al. 2009; Johns and Yang 2011a). Typically, the approach to large-scale tasks (Schindler et al. 2007; Nister and Stewenius 2006) is based on adaptation of image retrieval methods, whereby a query image is compared to all images in a database, each representing a distinct location, to find the closest match. In recent years, efficient matching has been inspired by the Bag-Of-Words (BOW) model (Sivic and Zisserman 2003) where comparisons of histograms of quantised features select candidate images for stronger geometric verification. In this paper, we present a new framework for place recognition that improves both the BOW filtering and geometric verification components of traditional approaches.

Databases for image retrieval often have significant redundancy due to dynamic behaviour influencing an image. In this paper, we define two types of dynamics: feature dynamics and scene dynamics. Feature dynamics arise due to the instability of a keypoint when the same real-world point is viewed under different viewpoints or illumination conditions. Scene dynamics arise due to long-term structural changes in a scene, such as renovations of building façades or seasonal effects on foliage, and short-term dynamic bodies such as pedestrians or cars. As a result of both these types of dynamic behaviour, many features exist in the database that are never matched to by features in a query image, causing significant redundancy and inefficiency in both memory and computational time. Figures 1 and 2 demonstrate two examples of dynamic