Computer Vision and Image Understanding 154 (2017) 182–191

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Water detection through spatio-temporal invariant descriptors

Pascal Mettes[a,b,*], Robby T. Tan[a,c], Remco C. Veltkamp[a]

[a] Department of Information and Computing Sciences, Utrecht University, Utrecht, the Netherlands
[b] Intelligent Systems Lab Amsterdam, University of Amsterdam, Science Park 904, Amsterdam, the Netherlands
[c] Multimedia Technology and Design Programme, SIM University, Singapore

ARTICLE INFO

Article history:
Received 29 October 2015
Revised 26 January 2016
Accepted 8 April 2016
Available online 12 April 2016

Keywords:
Water detection
Spatio-temporal descriptors
Fourier analysis
Invariants
Markov random fields

ABSTRACT

In this work, we aim to segment and detect water in videos. Water detection is beneficial for appllications such as video search, outdoor surveillance, and systems such as unmanned ground vehicles and unmanned aerial vehicles. The specific problem, however, is less discussed compared to general texture recognition. Here, we analyze several motion properties of water. First, we describe a video preprocessing step, to increase invariance against water reflections and water colours. Second, we investigate the temporal and spatial properties of water and derive corresponding local descriptors. The descriptors are used to locally classify the presence of water and a binary water detection mask is generated through spatio-temporal Markov Random Field regularization of the local classifications. Third, we introduce the Video Water Database, containing several hours of water and non-water videos, to validate our algorithm. Experimental evaluation on the Video Water Database and the DynTex database indicates the effectiveness of the proposed algorithm, outperforming multiple algorithms for dynamic texture recognition and material recognition.

1. Introduction

The goal of this work is water detection in both natural and man-made environments from videos. Spatio-temporal water detection finds applications in unmanned ground and aerial systems (e.g. self-driving cars, and UAV's (van Gemert et al., 2014)), outdoor surveillance, video search, and wildlife search. These applications are highlighted in Fig. 1. To the best of our knowledge, related work focuses on texture recognition in general, and thus does not specifically explore the motion properties of water.

We focus on investigating the spatio-temporal motion properties of water. In biological studies, the visual properties of water have been investigated in order to understand the visual attractiveness of water in human and animal vision. From the work of Schwind (1991), it is known that water insects are attracted to the horizontal polarization caused by the reflections of water surfaces. This observation has for example been used to explain why certain insects lay eggs on highways (Kriska et al., 1998). In videos however, polarization information is not captured. Human observers are still experts at water detection without polarization information, indicating that water contains valuable spatio-temporal motion properties that can be exploited. Here, we investigate which spatio-temporal motion properties make water distinctive.

Current methods for automatic water detection can be divided into two categories: in specialized systems or as part of a broader recognition framework. In the broader fields of material recognition (Hu et al., 2011; Mettes et al., 2014b; Sharan et al., 2013) and dynamic texture recognition (Chan and Vasconcelos, 2008; Doretto et al., 2003; Fazekas et al., 2009; Zhao and Pietikäinen, 2007) water is one of the target classes. In these works, the objective is to minimize the miss-classification rate over all classes and as a result, the distinctive properties of water specifically are not investigated. Furthermore, the focus is generally on classification or segmentation, but not on the joint problem as posed here. On the other hand, water detection in specialized settings, such as autonomous driving (Rankin and Matthies, 2006) and in maritime settings (Smith et al., 2003), either make non-generalizable restrictions on the movement and orientation of cameras (Rankin and Matthies, 2006) or use auxiliary data sources in their measurements (Rathinam et al., 2007; Scherer et al., 2012; Smith et al., 2003). To address the limitations of related work with respect to water detection specifically, this work provides an investigation into the temporal and spatial behaviour of water scenes.

This work reports three contributions. (1) We introduce a video pre-processing step to remove background reflections and inherent water colours. (2) We introduce a hybrid spatial and temporal descriptor for local water classification. For the temporal

* Corresponding author. Tel.: +31634724306.
E-mail addresses: P.S.M.Mettes@uva.nl (P. Mettes), robbytan@unisim.edu.sg (R.T. Tan), R.C.Veltkamp@uu.nl (R.C. Veltkamp).

http://dx.doi.org/10.1016/j.cviu.2016.04.003
1077-3142/

(a) Surveillance. (b) Water hazards [11]. (c) Aerial water maps [13]. (d) Wildlife search [1].

Fig. 1. Visual examples of practical applications that benefit from water detection.

descriptor, we analyze the periodicity and regularity of local water patches and derive a descriptor that captures these elements. For the spatial descriptor, we advocate Local Binary Patterns and we investigate what makes them suitable for local water detection. (3) We introduce a new dataset, the Video Water Database, for experimental evaluation and to encourage research into this topic. The Video Water Database, further discussed in Section 5, along with the code used in the experimentation will be made publicly available to encourage further research into this topic.

This work extends an earlier investigation into this topic (Mettes et al., 2014a) in multiple aspects. An improvement is proposed in the pre-processing stage to deal with areas on the border of multiple objects of reflection, by modeling the density of pixel values over time. Also, further analysis is performed to investigate whether the hybrid descriptor is able to capture the spatial and temporal behaviour water ripples. In the experiments, we evaluate whether our method can generalize to water conditions and water types not seen during training. Lastly, another fusion of the temporal and spatial descriptor is evaluated.

The layout of the rest of this paper is as follows. In Section 2, an overview of water detection in related work is provided. Section 3 introduces the pre-processing step of the videos and the analysis of the local behaviour of water. This is followed by the discussion on local probabilistic classification and spatio-temporal regularization in Section 4. Finally, Section 5 provides the experimental evaluation of the algorithm and the paper is concluded in Section 6.

2. Related work

Given the lack of specific attention given to water detection, an overview is provided with respect to two broader recognition tasks: material recognition and dynamic texture recognition. Also, an overview of water localization in specialized systems is provided.

2.1. Water in material recognition

The classification of materials and static textures in images has a long history of investigation (Everts et al., 2012; Ojala et al., 2002; Sharan et al., 2013; Varma and Zisserman, 2009). Works on this topic are in line with Biederman (1987), who conjectured that materials are recognized in human vision by their surface characteristics such as texture and colour. Well-known approaches include filter bank distributions (Varma and Zisserman, 2005), Local Binary Patterns (Ojala et al., 2002), and image patch exemplars (Varma and Zisserman, 2009). In recent works, a shift has been made from laboratory settings (Ojala et al., 2002; Varma and Zisserman, 2005; 2009) to real-world image databases (Hu et al., 2011; Mettes et al., 2014b; Sharan et al., 2013). In these works, a range of surface characteristics, e.g. texture, colour, and reflectance, to find out what characteristics are best for classification. The results of these works indicate that spatial information is informative for distinguishing different materials. For water detection in videos however, there are two limiting aspects. First, only the spatial characteristics are investigated, excluding valuable temporal information. Second, research into material recognition has focused mostly on solving the classification problem or the segmentation problem, but not their joint problem.

2.2. Water in dynamic texture recognition

Dynamic textures are part of a class of motions with either structural or statistical similarity in both space and time (Nelson and Polana, 1992). Exemplary dynamic textures include fire, water, flags, and weather patterns. One of the dominant approaches in dynamic texture recognition is based on optical flow statistics (Chen et al., 2013; Fazekas et al., 2009; Fazekas and Chetverikov, 2007; Vidal and Ravichandran, 2005). In these approaches, either a global description is generated using invariant flow statistics such as characteristic direction and magnitude of flow vectors (Fazekas and Chetverikov, 2007), or flow vectors are binned into Histograms of Optical Flow (HOOF) (Chen et al., 2013). The use of optical flow is intuitively interesting for water detection, as the spatio-temporal movement of water seems statistically different to related textures. The use of conventional optical flow is however problematic for water detection in videos, as water meets none of the requirements for a proper flow estimation: Lambertian surface reflectance, pure translational motion parallel to the image plane, and uniform illumination (Beauchemin and Barron, 1995). A representation using optical flow will therefore be heavily influenced by the noise of the flow estimation, which makes optical flow not desirable for water detection.

Another popular research direction focuses on modeling dynamic textures as Linear Dynamical Systems (LDS) (Chan and Vasconcelos, 2008; Doretto et al., 2003; Ravichandran et al., 2013; Saisan et al., 2001). In dynamic texture recognition, the use of LDS has been made popular by Saison et al. (Saisan et al., 2001) and Doretto et al. (Doretto et al., 2003), mostly due to the proposed efficient sub-optimal learning procedure. As the original formulation of LDS requires a modeling of whole videos, it is unfit for local detection purposes. In order to deal with multiple textures within a video, several extensions have been provided. These include mixtures of dynamic textures (Chan and Vasconcelos, 2008), hierarchical EM clustering (Mumtaz et al., 2013), and Bags of Dynamical Systems (Ravichandran et al., 2013). These algorithms can potentially handle multiple textures in a video, but they have so far not been applied to detection problems. A noteworthy exception is the work of Ravichandran et al. (2011), where the joint segmentation and classification problem of dynamic textures is tackled by dividing a video into parts using Dynamic Appearance Images computed from LDS, after which the parts are represented by a bag-of-words representation with SIFT features. The representations are however more general and not tailored to water detection. Also, it is explicitly assumed that the texture class of a pixel does not change over time, restricting potential applications. Rather than performing a holistic modeling as with LDS, this work attempts to detect water from a local scale. The local scale is essential, as water is not bound to specific shapes in a scene.