Robotics and Computer-Integrated Manufacturing 37 (2016) 90–102

Contents lists available at ScienceDirect

Robotics and Computer-Integrated Manufacturing

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

An approach for identifying classifiable regions of an image captured by autonomous robots in structural environments

Andrew Wing Keung To*, Gavin Paul, Dikai Liu

Centre for Autonomous Systems, University of Technology, Sydney, Australia

ARTICLE INFO

Article history:
Received 31 March 2014
Received in revised form
17 July 2015
Accepted 23 July 2015

Keywords:
Surface-type classification
RGB-D
Likelihood
Texture features

ABSTRACT

When an autonomous robot is deployed in a structural environment to visually inspect surfaces, the capture conditions of images (e.g. camera's viewing distance and angle to surfaces) may vary due to unideal robot poses selected to position the camera in a collision-free manner. Given that surface inspection is conducted by using a classifier trained with surface samples captured with limited changes to the viewing distance and angle, the inspection performance can be affected if the capture conditions are changed. This paper presents an approach to calculate a value that represents the likelihood of a pixel being classifiable by a classifier trained with a limited dataset. The likelihood value is calculated for each pixel in an image to form a likelihood map that can be used to identify classifiable regions of the image. The information necessary for calculating the likelihood values is obtained by collecting additional depth data that maps to each pixel in an image (collectively referred to as a RGB-D image). Experiments to test the approach are conducted in a laboratory environment using a RGB-D sensor package mounted onto the end-effector of a robot manipulator. A naive Bayes classifier trained with texture features extracted from Gray Level Co-occurrence Matrices is used to demonstrate the effect of image capture conditions on surface classification accuracy. Experimental results show that the classifiable regions identified using a likelihood map are up to 99.0% accurate, and the identified region has up to 19.9% higher classification accuracy when compared against the overall accuracy of the same image.

1. Introduction

The typical manufacturing factory setting provides a well-structured and predictable environment that is suitable for implementing pre-planned routines on a robot to perform repetitive tasks. However, pre-planned routines are impractical for a field robot due to the environment changing over time and/or the robot being freely moved within the environment. One such scenario is steel bridge maintenance conducted by autonomous robots [1,2], as shown in Fig. 1.

In steel bridge maintenance, a mobile robot is moved to various sections of a bridge to conduct grit-blasting for the removal of rust and old paint from steel surfaces in preparation for repainting. In order for the robot to grit-blast autonomously in each position, an up-to-date geometric map of the surrounding environment is provided to the robot such that a plan for grit-stream trajectory and robot movements can be newly generated. At present, there are well-developed approaches for a robot to explore and build an update geometric map of an environment using a depth sensor mounted on the robot's end-effector [3,4]. Provided with a geometric map of the environment, a robot can only autonomously grit-blast all the surfaces without the capability to target specific surface areas based on surface-type/conditions (e.g. mildly rusted and heavily rusted).

For a robot to be capable of selectively grit-blasting specific surface areas, it must also explore and inspect the surface's condition. One possible approach to this is to mount a vision camera to the robot's end-effector and capture images during (1) pre-grit-blasting for identifying specific surface areas to grit-blast based on rust grading, and (2) post-grit-blasting for assessing whether the required steel cleanliness has been achieved or re-blasting is necessary. A robot can inspect the surfaces in the captured images by using a classifier trained with surface samples from a visual inspection standard such as the rust grading and steel cleanliness visual metrics provided in BS EN ISO 8501-1 [5]. In this way, information about the surface's condition can be produced that will enable a robot to intelligently (re)grit-blast specific surface areas on a bridge.

A review of vision-based classification approaches shows that colour and/or texture features can be extracted to accurately distinguish between various surface-types (surface appearance of

* Corresponding author.
E-mail addresses: Andrew.To-1@uts.edu.au (A.W.K. To), Gavin.Paul-1@uts.edu.au (G. Paul), Dikai.Liu@uts.edu.au (D. Liu).

http://dx.doi.org/10.1016/j.rcim.2015.07.003
0736-5845/

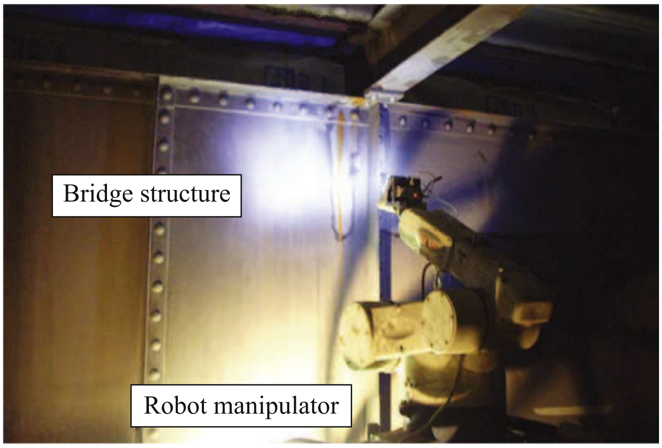

Fig. 1. An autonomous robot for grit-blasting steel bridge structures.

materials) [6], and steel surface conditions [7,8]. Colour features have been shown to provide high resolution (per pixel) classification for rust. In the simplest implementation, RGB intensity values can be directly used as the features to classify rust from a single colour background surface [9]. However, RGB intensity values are shown to be affected by illumination changes and thus further investigation into improving classification accuracy for non-uniformly illuminated surfaces have been performed in [10] and [11]. The investigation examined 14 colour-spaces and found that the a*b* colour component from the CIELab colour-space can provide high classification accuracy for rust in non-uniformly illuminated images.

Texture features can also be extracted to accurately classify non-uniformly illuminated surfaces [12] and are typically preferred in applications where it is difficult to classify based on colour features (i.e. surface-types that have similar colour appearance). The main disadvantage of using texture features when compared against colour features is that an image region ($n \times n$ pixels) is required to extract the features; thus reducing the resolution of the classification results. However, due to the inter-pixel information provided from an image region, it is possible to extract richer features to distinguish a wider range of surface-types. Approaches for extracting texture features can be categorised into four main groups including statistical, geometrical, model-based and signal processing [13–15]. The work in [16] demonstrates the use of these approaches to extract texture features that accurately classifies for rust grading on steel surfaces. Specifically, texture feature extraction approaches such as Gray Level Co-occurrence Matrix (GLCM), Fast Fourier Transform (FFT) and Wavelet Transform (WT) have been used towards visual inspection of bridges given the high performance in space–frequency decomposition at different scales [17]. To summarise, a texture-based classification approach is generally more suitable for inspecting surface-types and surface conditions given that texture features can distinguish a wider range of surfaces.

A review of visual inspection systems implemented in manufacturing environments shows that a fixed camera position is preferred to ensure consistent image capture conditions (viewing distance and viewing angle) for high inspection performance. For example, a camera mounted in a fixed position on a conveyor system has been used to achieve inspection accuracy of greater than 90%; for marble slab grading [18] where captured images are processed to categorise marble slabs into different aesthetic groups, and for steel manufacturing [7,19,20] where captured images are processed to detect surface defects on metal sheets/strips. Essentially, by fixing the camera's position the appearance of surface(s) in each image in terms of focus quality, spatial resolution and perspective distortion will remain consistent and a classifier can be trained using pre-captured samples to accurately classify subsequently captured samples.

However, in field applications consistent image capture conditions cannot be assumed. Examples of field applications are a camera being mounted onto a robot end-effector for autonomous bridge inspection [1,2], welding inspection [21,22] and parts assembly [23]. As a result, when a classifier is trained with samples captured at a limited range of viewing distances and angles (i.e. surface samples provided in a visual standard guide, BS EN ISO 8501-1), the inspection performance can be affected when the image capture conditions are varied due to un-ideal robot poses selected to position the camera in a collision-free manner. Presently an autonomous bridge surface inspection robot [24] is capable of inspecting for rust on steel surfaces by using colour features which are not affected by surface appearance changes (e.g. focus quality, spatial resolution and perspective distortion). However, in order to extend the robot to inspect for different rust grading and cleanliness on steel surfaces, texture-based classification will be required and consequently an approach for identifying regions in an image that are classifiable. A classifier's prediction certainty cannot be used to identify the classifiable regions because of the potential surface appearance changes between training and test samples. For example, an out-of-focus (blurry) test sample of a blasted metal surface may extract texture features that resemble the timber surface. As a result, the classifier will misclassify this sample as timber and "falsely" have a high degree of prediction certainty.

This paper presents an approach to calculate a value that represents the likelihood of a pixel being classifiable by a classifier trained with a limited dataset. The likelihood is estimated based on the assumption that identical capture conditions used for capturing the training dataset and the test images will produce accurate results and deteriorate if the capture conditions are varied. This likelihood value can be calculated for each pixel in an image to form a likelihood map and used to identify regions in the image that are classifiable. The capture conditions necessary for calculating the likelihood value can be obtained by capturing additional depth data that maps to an image (collectively referred to as a RGB-D image). The subsequent sections of this paper are organised as follows: Section 2 details the overview of the approach, the process for calculating the image capture conditions of viewing distance and viewing angle, the process for selecting threshold values based on surface appearance factors and the process for calculating the likelihood value using the identified threshold ranges. Section 3 presents three experiments conducted to test the proposed approach and the corresponding experimental results, and Section 4 provides a discussion and the conclusion

2. The approach

2.1. Overview

Fig. 2 shows the overview of the approach for identifying classifiable regions of an image by generating a likelihood map. From a RGB-D image captured to inspect a surface(s), the depth image is used to calculate the capture conditions including the viewing distance d_c, and the viewing angle θ_c, for each pixel (detailed in Section 2.2). Provided with the viewing distance and viewing angle for each pixel, and a set of calibrated threshold values $\{n_1, f_1, n_2, f_2, \tau_\theta\}$ (detailed in Section 2.3), the likelihood value for each pixel is calculated to produce a likelihood map (detailed in Section 2.4). Finally, the classifiable image regions in the map can be identified.