University Defence Research Collaboration in Signal Processing

LSSC Consortium White Paper

Automatic Target Detection in Videos Using Incongruence

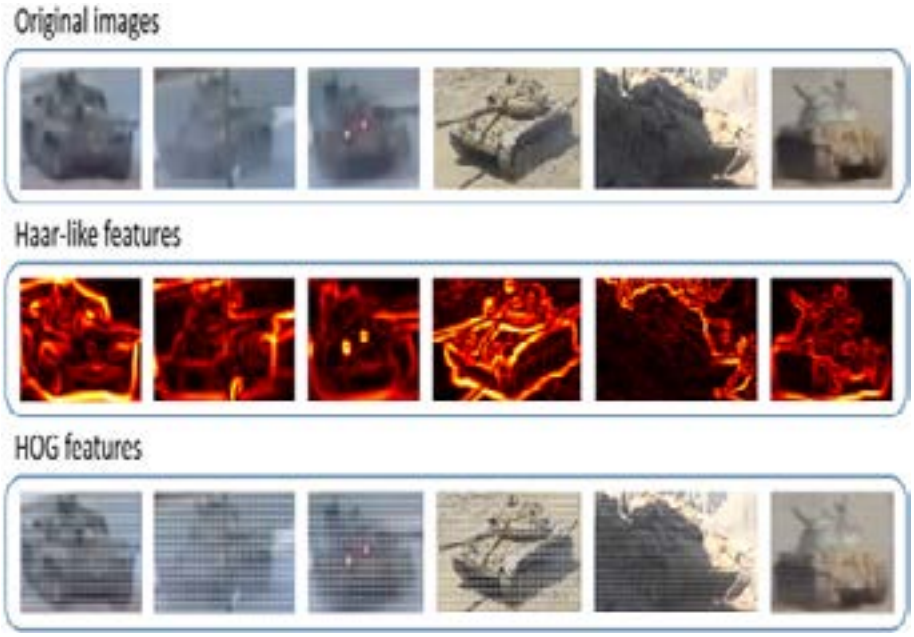
Introduction

Automatic target detection is a prominent challenge in defence applications. It involves the development of automatic techniques which enable the detection and recognition of objects of interest from sensory input. A wide variety of sensors can be utilised for target detection, ranging from cameras, synthetic aperture radars (SARs) and lidars to acoustic sensors. This work focuses on target detection using visual sensors. There are still, however, many open research challenges such as efficient target acquisition from moving vehicles (e.g. UAVs), detection of camouflaged objects and accurate target recognition and tracking in the highly cluttered battlefield environment, where visibility is often limited due to occlusions.

In this collaborative work between the Universities of Surrey and Cardiff, the problem of efficiently detecting the location of the target is tackled by employing an ensemble of object detectors via utilizing an incongruence measure. The aim is to show an increase on the performance of a given non-contextual classifier by firstly aiding it with contextual information to create a contextual classifier. The incongruence between the contextual and non-contextual classifiers is then used to control the decision making process.

Methodology

Two tank detectors have been developed, based on established feature descriptors. The first of those utilises the histogram of oriented gradients (HOG) descriptor, which represents an object of interest as a set of local features, extracted by counting occurrences of gradient orientation in localised parts of images used for training. The second tank detector is based on Haar-like features which describe an object as a group of adjacent rectangular regions. Both tank detectors have been built with very few training samples.

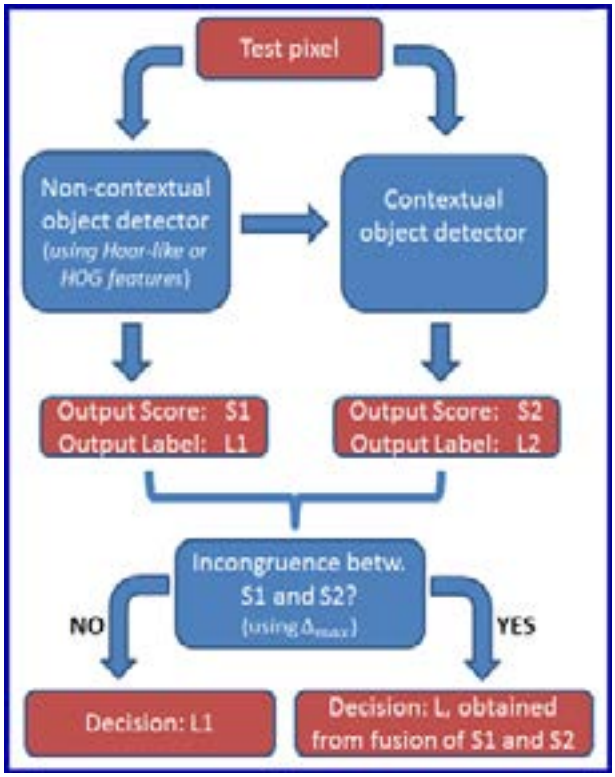


EPSRC Engineering and Physical Science Research Council

[dstl]

UDRC

At the test stage, given a sample image, a trained non-contextual object detector provides information about each pixel's score of belonging to an object class of interest. Then, in order to aid the decision making process, a contextual classifier is obtained by combining the ioutputs of a given non-contextual classifier over a number of neighbouring frames. In particular, while deciding the label of a pixel at the frame of interest, the incongruence between the non-contextual and contextual classifiers is measured using the proposed incongruence measure, Δ_max^ [1]. When deemed congruent, the decision of the non-contextual classifier is accepted. On the other hand, at an instance of incongruence, intelligent fusion of scores is carried out leading to a final decision.



The incongruence measure used in this study, Δ_max^ , gauges the consistency of classifier outputs by concentrating on the dominant hypotheses. Δ_max^ boosts the incongruence value when two experts disagree on the favoured hypotheses compared to when they agree.

Future work will investigate the use of motion-based target detectors in conjunction with the appearance-based detectors used in the study so far. Moreover, novel methodologies on the intelligent fusion of scores aided by incongruence detection are to be explored to further increase the classification performance.

References:

[1] J. Kittler and C. Zor: A Measure of Surprise for Incongruence Detection, 2nd IET International Conference on Intelligent Signal Processing, 2015, London (accepted).