Textural lacunarity for semi-supervised detection in sonar imagery

J. D. B. Nelson and V. Krylov

Abstract

Wavelet energy-based lacunarity features, which measure deviations from translational statistical invariance over multiple scales, were recently proposed for object detection and classification in sonar imagery. We here extend the idea to incorporate further robustness to background type whilst retaining sensitivity to local changes in texture caused by the presence of man-made objects. The resulting textural-lacunarity features are constructed by estimating the joint distribution of local neighbourhoods with empirical distributions over an adaptive texton dictionary. Experiments on a synthetic aperture sonar imagery dataset suggest that the features offer significant improvements in the receiver operating curve.

I. INTRODUCTION

As autonomous underwater vehicles become ever more sophisticated there is growing interest from sectors such as the oil and gas industry, oceanography, and defence in computer aided approaches to object detection using sonar imagery. The complexity and variability of the seabed floor and of the potential objects of interest presents a major challenge to automatic target detection methodology. We focus on a defence application where the objects of interest, namely underwater mines, comprise various types and shapes and are located in rippled and non-rippled regions of the seabed.

One approach to detect objects of various shapes is to frame the problem as anomaly detection whereby the presence of a potential object of interest is detected by measuring its interference with the statistical homogeneity of the background. To this end a form of lacunarity [6], which measures deviations from translational statistical invariance, has been considered by Myers for mine hunting [7]. In its originally proposed form however, this feature only measures local variability in a sliding neighbourhood and depends as much on the statistics of the background as it does on the presence of an object. As such...
lacunarity is more commonly used in segmentation; examples include: remote sensing with SAR imagery [4], vision-based diagnosis skin cancer [1], and diagnosis of osteoporosis [12].

Instead, Nelson and Kingsbury proposed a measure of lacunarity which compared correlation of the wavelet energies between an inner and an outer region of a local neighbourhood [9]. Unlike the original lacunarity, this offered some robustness to background type.

Although multiscale and multi-directional, wavelet-based lacunarity was not intended as a complete description of statistical disruption. In particular, it is evident that the presence of a target also disrupts the texture of a region. Whilst much work in the texture classification literature has treated texture in terms of energy (of wavelet filters or otherwise), relatively recent work by Varma and Zisserman [11] has also shown the importance of the joint distribution of local patches. Persisting with the inter-regional aspect, this motivates the advancement of a complementary measure of lacunarity whereby the joint distributions of local patches in an inner region are compared to those from an outer region.

To validate these new features, we will adopt the same framework as [8], [9] where the oft-followed fully-supervised approach (c.f. [2], [3], [10] and Figure 1) is circumvented in favour of a semi-supervised scheme (c.f. Figure 2). Here, adaptive, wavelet-based sand ripple suppression together with a non-linear matched filter detector are used to prune the data. However, only false positives are used to train the classifier—in this case, a one-class support vector machine. This only requires examples of background data, and thus avoids the major difficulty of capturing examples of real threats in any significant quantity. It also potentially means that unanticipated, or even hitherto unknown, types of threats can be detected and will, in theory, allow the system to evolve according to the environmental conditions. As a result, the approach regards the background, or natural seabed, as normal whereas the unknown targets are treated as anomalies.

Sand ripples, which often excite mine detector matched filters, are suppressed by implementing a wavelet shrinkage preprocessing step from [9]. This performs a dual-tree complex wavelet transform, applies a soft shrinkage operator, and then performs an inverse dual-tree complex wavelet transform. The idea is that the shrinkage operator reduces the energy of any coefficients that contribute to the ripples. This energy reduction is proportionate to how much the coefficients deviate from the usual power law decay, observed in typical non-rippled conditions. It is important to note that the ripple suppression does not guarantee that all ripple energy is completely removed. Furthermore, there exist many other background types (vegetation etc) that the ripple suppression method is not designed to suppress. Hence, although the ripple suppression step greatly reduces the number of false detections, successful classifier features (be they lacunarity or otherwise) must still be robust to background type.
The proposed textural lacunarity features are introduced in Section II and experiments on a synthetic aperture sonar data set are discussed in Section III.

Fig. 1. Common detector/classification scheme. Detector finds (true and false) positives. Training phase: a binary classifier is trained on features extracted from both true and false positives. Testing phase: the detector is applied, features are extracted from positives, which are then classified by a learned decision function.

Fig. 2. Proposed detector/classification scheme. Pre-processing suppresses ripples. Detector finds (true and false) positives. Training phase: a unary classifier is trained on features extracted from false positives only. Testing phase: Pre-processing suppresses ripples. Detector is applied, features are extracted from positives, which are then classified by a learned decision function (true positives are anomalies).

II. LOCALISED APPROXIMATE JOINT DISTRIBUTIONS FOR ANOMALY DETECTION

For notational convenience, we introduce a labelling \( I \) of the lattice \( \mathcal{T} \) over which the random field is defined, namely \( t: I \rightarrow \mathcal{T} \subset \mathbb{Z}^2 \). We write the real-valued random variable of the field \( x: \mathcal{T} \rightarrow \mathbb{R} \) at a site or location, labelled by the index \( i \in I \), as \( x_i := x(t_i) \in \mathbb{R} \) and its neighbours as \( (x_j)_{j \in \partial i} := (x(t_j))_{j \in \partial i} \in \mathbb{R}^{\partial i} \) over some neighbourhood structure indexed by \( \partial i \) such that \( \partial i \cap \{i\} = \emptyset \). We also use the shorthand convention that \( x_{-i} = x_{I \setminus \{i\}} \in \mathbb{R}^{I \setminus \{i\}} \).

Our texture descriptor is loosely based on that of Varma and Zisserman [11] whose approach lies amongst the state-of-the-art for the texture classification problem. Here, we adapt their idea for anomaly detection. This results in what might be regarded as a ‘textural lacunarity’ measure. Instead of comparing the energy (i.e. 2nd moment) of \( x \) over the inner and outer regions, as in [9], we attempt to compare
representations of the joint distributions of the inner and outer regions. This task is more realistic when
Markovianity is assumed/invoked, namely that, given the neighbours the value of a random field at a
site is conditionally independent from all others, i.e. \( x_i \perp x_{\neq i(\partial i)} | x_{\partial i} \), or \( p(x_i | x_{\neq i}) = p(x_i | x_{\partial i}) \). In this
sense, that the neighbourhoods fully describe a Markov random field motivates us to estimate the joint
local neighbourhood \( p(x_{\partial i}) \). In particular, if we denote the set of labels in the inner region by \( \mathbb{I}_0 \) and the
outer region by \( \mathbb{I}_* \), where \( \mathbb{I} \supset \mathbb{I}' = \mathbb{I}_0 \cup \mathbb{I}_* \), we are interested in estimating the distance
\[
d (p(x_{\partial i}; i \in \mathbb{I}_0), p(x_{\partial i}; i \in \mathbb{I}_*)) \ ,
\]
over some suitable metric \( d(\cdot, \cdot) \). In practice, the joint local distribution \( p(x_{\partial i}) \) can be estimated as
follows by first clustering all the neighbourhoods in the entire region, via k-means or otherwise. This
attributes a cluster label, 1 to \( m \) say, to each neighbourhood: \( C : \mathbb{R}^{\partial |} \mapsto \{1, \ldots , m\} \). In this way, the
joint local neighbourhood is represented instead by the empirical mass function, or normalised histogram
\( (p(C(x_{\partial i}) = j))_{j=1}^m \) over a ‘dictionary’ (the cluster centres) of local neighbourhood clusters. The distance
(1) can then be approximated by
\[
d (p(C(x_{\partial i}) ; i \in \mathbb{I}_0), p(C(x_{\partial i}) ; i \in \mathbb{I}_*)) \ .
\]
This process is summarised by Algorithm 1. To reiterate: all the neighbourhoods in the entire region are
clustered via k-means or otherwise. Each cluster is assigned a label. The centres of each cluster is a
dictionary element. The texton feature is simply the distribution of neighbourhoods over the cluster labels.
The algorithm is performed in both training and testing and forms part of the ‘feature extraction’ step which
takes place after the initial ‘detector’ is applied (cf. Figure 2). It is important to note that the dictionary is
rebuilt for each positive (recall the pseudo code: the clustering is done about a region ‘for each positive
from detector’). In this sense, the dictionary adapts to the data. A more explicit representation, also

**Algorithm 1** Textural lacunarity algorithm

```
for each positive from detector do
    use k-means to cluster the neighbourhoods over region \( \mathbb{I}' \) centred about positive
    for inner \( \mathbb{I}_0 \) and outer \( \mathbb{I}_* \) region do
        form normalised histogram \( p(C(x_{\partial i})) \) over cluster labels
    end for
    compute distance \( d (p(C(x_{\partial i}) ; i \in \mathbb{I}_0), p(C(x_{\partial i}) ; i \in \mathbb{I}_*)) \) between inner and outer histograms
end for
```
attempted by Varma and Zisserman, is to include the marginal information in the local joint neighbourhood. The texture is thus characterised by \( p(x_{\partial i \cup \{i\}}) = p(x_i, x_{\partial i}) \). In practice, the neighbourhoods are clustered and histogrammed as before. The marginal distribution is approximated by a separate histogram—the values taken by the centre pixels are discretised and binned so that each \( x_i \) can be attributed a bin number: \( C_0: \mathbb{R} \mapsto \{1, \ldots, m_0\} \). Together, the distribution of the centres and neighbourhoods are represented by the joint mass function \( p(C_0(x_i), C(x_{\partial i})) \). The distance measure then becomes

\[
d(p(C_0(x_i), C(x_{\partial i}); i \in I_0), p(C_0(x_i), C(x_{\partial i}); i \in I_*))
\]

and the process is summarised by Algorithm 2. In the experiments conducted here, we follow Varma and

\begin{algorithm}
\textbf{Joint textural lacunarity algorithm}
\begin{algorithmic}
\For {each positive from detector}
\State use \( k \)-means to cluster the neighbourhoods over region \( I' \) centred about positive
\State uniformly discretise range of centre pixels
\For {inner \( I_0 \) and outer \( I_* \) region}
\State form normalised histogram \( p(C(x_{\partial i})) \) over cluster labels
\State form normalised histogram \( p(C_0(x_i)) \) of centre pixels
\EndFor
\State compute distance
\State \[
d(p(C_0(x_i), C(x_{\partial i}); i \in I_0), p(C_0(x_i), C(x_{\partial i}); i \in I_*))
\]
\State between inner and outer histograms
\EndFor
\end{algorithmic}
\end{algorithm}

Zisserman and use the \( \chi^2 \) metric

\[
d(p, q) = \frac{1}{2} \sum_j \frac{|p_j - q_j|^2}{p_j + q_j},
\]

to measure the difference between two histograms \( \{p_j\} \) and \( \{q_j\} \). For Algorithm 1, the bin index \( j \) runs over the cluster labels \( 1, \ldots, m \); for Algorithm 2, it runs over both the cluster labels and the marginal bin numbers \( 1, \ldots, m_0 \). In the second case \( j = (j_0, j_*) \), say, where \( j_0 = 1, \ldots, m_0 \) and \( j_* = 1, \ldots, m \). The neighbourhood structure \( \partial \) was chosen to be the 8-neighbourhood with a radius of 7 pixels and the number of clusters/bins used was \( m = m_0 = 12 \). For our experiments, we simply chose these by a short process of trial and error over a small subset of images. Ideally, these parameters should be optimised by more formal means such as cross-validation or otherwise.

III. Experiments

Experiments were conducted on MUSCLE data, provided by the NATO Centre for Maritime Research and Experimentation. Examples of the utility of the textural-lacunarity features are shown in Figures 3, 4, and 5.

A. Non-rippled region

A typical true and false positive in a non-rippled region is shown in Figure 3. These are taken directly from the output of the matched filter detector; no manual adjustments have been made. The size of the region is chosen to be the same size as the matched filter which will ideally be adapted to the sensor and environment geometry—i.e. the length is increased with respect to range in a piecewise manner to accommodate longer shadow regions further away from the sensor.

The wavelet based lacunarity features introduced in [9] are constructed by computing the root sum of squares of the inner and outer regions and then using a normalised dot product to compare them. The idea is that the presence of a genuine target will disrupt the wavelet energy at a certain scale level. This will cause a decorrelation between the energies of the inner and outer regions and will thus manifest an abnormally small dot product. Indeed, the wavelet energies in Figure 3 (c) show that the wavelet energies of the false positive region are more correlated than those in the true positive region.

Likewise, the new textural-lacunarity features also reflect such a difference. In Figure 3 (d), the empirical distributions of the inner and outer regions are plotted with respect to the respective texton dictionaries ((e) and (f)). It can be seen that the distributions in the false positive region are similar but those in the true positive region are markedly different.

B. Rippled region

Figure 4 shows an example of a true and false positive in a ripple field. Again, both lacunarity features corroborate a change when a target is present. Figure 5 shows an example where the wavelet energies are not sufficiently decorrelated in the presence of a genuine target but where the texture-based features do expose a notable difference in the distributions of the two regions.

C. ROC curves

The textural-lacunarity features are added to the wavelet-lacunarity features. As in [9], a one-class support vector machine is used to learn the decision boundary (mine/not mine) by training on the false positives only.
Fig. 3. Comparison of the wavelet-based (c) and textural-based (d) features of a true and false positive that lie in a non-rippled region. Wavelet energy decorrelation between the inner and outer regions and/or a difference in the distributions over the texton dictionary reflects the presence of an anomaly.

Consistent with the experimental setup of Hill et al [5] and Nelson and Kingsbury [9], half of the 180 sonar images were used to train the data and the other half held back for testing. In our experiment, a matched filter detector phase generates the false positives required for training whereas Hill et al. [5] generate 1000 randomly sampled non-target points from each 2000-by-7000 image and then used a two-class SVM trained on both false and true positives. They also reported the results of two human
(a) False positive in the ripple field

(b) True positive in the ripple field

(c) Energy at 4th & 5th finest level (left: false, right: true)

(d) Empirical mass functions (left: false, right: true).

(e) Texton dictionary for false positive

(f) Texton dictionary

Fig. 4. Comparison of the wavelet-based (c) and textural-based (d) features of a true and false positive that lie inside a ripple field. Wavelet energy decorrelation between the inner and outer regions and/or a difference in the distributions over the texton dictionary reflects the presence of an anomaly.

experts.

Table I and Figure 6 summarise the classification results for the lacunarity-based one-class SVM result with/without the addition of the textural features (‘textural + wavelet’) and the two-class SVM result of Hill et al (with results of two expert human operators) [5]. In addition, Figure 6 also shows the ROC curves of the one-class SVM result when the textural features are used without the wavelet features.
(a) True positive in the ripple field

(b) Empirical mass function.

(c) Energy at 4th & 5th finest levels.

(d) Texton dictionary

Fig. 5. Comparison of the wavelet-based (b) and textural-based (c) features of a true positive in a ripple field. In this example the wavelet energies of the inner and outer regions remain reasonably correlated and thus fail to provide an indication of an anomaly whereas, on the other hand, the difference in the distributions over the texton dictionary does successfully reflect the presence of an anomaly.

IV. CONCLUSIONS

Both the table and ROC curve show that the new textural features add discrimination power to the original wavelet-lacunarity features proposed by Nelson and Kingsbury [9]. For example, to recover 95% of true positives, the textural features reduce the number of false positives by around 28%. To recover 98% of true positives, the new features incur less than half the number of false positives of the previous method. Likewise, the curve shows that the textural features benefit from the addition of the wavelet features. Together, both sets of features deliver better results. There does not appear to be a clear overall winner between the textural and joint textural methods. Each one holds an advantage over different regions of the ROC curve. The two-class approach of Hill et al [5] appears to be superior over much of the lower part of the ROC curve (below 95% true positive rate). However, we again note that the one-class approach described here does not require any target exemplars for training.

Amongst some of the fixed parameter settings, the inner and outer region sizes are worthy of some note. Both feature types are accumulative in nature—the wavelet features computes mean energy, the
TABLE I

NUMBER OF FALSE POSITIVES INCURRED TO RECOVER 80%, 85%, 90%, 95%, 98%, AND 99% OF THE TOTAL NUMBER OF TRUE POSITIVES.

<table>
<thead>
<tr>
<th>Number of true positives</th>
<th>Number of false positives</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nelson &amp; Kingsbury [9]</td>
</tr>
<tr>
<td>(80%) 147</td>
<td>95</td>
</tr>
<tr>
<td>(85%) 156</td>
<td>130</td>
</tr>
<tr>
<td>(90%) 166</td>
<td>194</td>
</tr>
<tr>
<td>(95%) 175</td>
<td>337</td>
</tr>
<tr>
<td>(98%) 180</td>
<td>1114</td>
</tr>
<tr>
<td>(99%) 182</td>
<td>3508</td>
</tr>
<tr>
<td></td>
<td>Hill et al</td>
</tr>
<tr>
<td>textural lacunarity</td>
<td>74</td>
</tr>
<tr>
<td>45</td>
<td></td>
</tr>
<tr>
<td>78</td>
<td></td>
</tr>
<tr>
<td>105</td>
<td></td>
</tr>
<tr>
<td>310</td>
<td></td>
</tr>
<tr>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>joint textural lacunarity</td>
<td>81</td>
</tr>
<tr>
<td>99</td>
<td></td>
</tr>
<tr>
<td>131</td>
<td></td>
</tr>
<tr>
<td>225</td>
<td></td>
</tr>
<tr>
<td>535</td>
<td></td>
</tr>
<tr>
<td>668</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 6. ROC curves of: one-class SVM result without textural features; the two-class SVM result of Hill et al [5] (which requires training on true positives); and the one-class SVM results with textural and joint-textural distributional features. Also shown are the results from two human expert operators (cf. [5]).

texton features are simply estimates of distributions. This naturally affords some robustness to region size. In practice, the ultimate system would adapt the region size according to the altitude and internal parameters of the sensor. Since such parameters were reasonably constant over the data set, no attempt was made to implement this adaptation in our experiments.

Aside from applications to other areas in remote sensing and beyond, further work may include exploring the possibility of constructing a textural dictionary over multiple scale levels and directional subbands in the wavelet domain. This would effectively encode the intuitive observation that coarse scale information carries Markov-type dependencies over larger spatial neighbourhoods than finer scale levels.

V. ACKNOWLEDGEMENTS

This work was supported by Dstl. All data used here was collected and provided by the NATO Centre for Maritime Research and Experimentation. The authors would like to thank Duncan Williams (DSTL) for discussions on the data and mine hunting literature; and Dave Bull and Paul Hill (University of Bristol) for providing ground truth of the data and other useful information.

REFERENCES


