Pattern Recognition 45 (2012) 3439–3450

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

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

Novelty detection in wildlife scenes through semantic context modelling

Suet-Peng Yong [a,b], Jeremiah D. Deng [a,*], Martin K. Purvis [a]

[a] Department of Information Science, University of Otago, Dunedin 9054, New Zealand
[b] Universiti Teknologi Petronas, Perak, Malaysia

ARTICLE INFO

Article history:
Received 10 June 2011
Received in revised form
16 January 2012
Accepted 27 February 2012
Available online 8 March 2012

Keywords:
Novelty detection
Co-occurrence matrices
Semantic context
Multiple one-class models

ABSTRACT

Novelty detection is an important functionality that has found many applications in information retrieval and processing. In this paper we propose a novel framework that deals with novelty detection in multiple-scene image sets. Working with wildlife image data, the framework starts with image segmentation, followed by feature extraction and classification of the image blocks extracted from image segments. The labelled image blocks are then scanned through to generate a co-occurrence matrix of object labels, representing the semantic context within the scene. The semantic co-occurrence matrices then undergo binarization and principal component analysis for dimension reduction, forming the basis for constructing one-class models on scene categories. An algorithm for outliers detection that employs multiple one-class models is proposed. An advantage of our approach is that it can be used for novelty detection and scene classification at the same time. Our experiments show that the proposed approach algorithm gives favourable performance for the task of detecting novel wildlife scenes, and binarization of the semantic co-occurrence matrices helps increase the robustness to variations of scene statistics.

1. Introduction

In our visual perception, something popping out of context is often considered novel or interesting to us. According to psychological studies, novelty seems intuitively tied to interest [1], but short-term novelty can also be referred to contrast with recent experience [2]. Indeed, in terms of contrast, novelty may relate to not only the presence of new features, but also the sudden lack of certain stimulus quality consistently observed in previous experience. The modelling of such kind of novelty detection mechanism, with the help of image analysis techniques, can be useful in a number of applications. For instance, on a social network website, it is desirable that images can be automatically analyzed, and abnormal or inappropriate contents can be detected and then either removed or hidden from certain users or groups. Similarly, when browsing a large photo collection, it is also desirable to use novelty detection to highlight or select scenes with novelty or 'interestingness', either of the content itself or the layout.

In general, there have been a number of statistical or machine learning approaches proposed for novelty or anomaly detection in various application domains [3]. To detect novel scenes, a straightforward idea would be to employ these methods and apply them to low-level visual features extracted from the images, such as colour and texture. This approach is however rather questionable, as content-based image retrieval research has long revealed the ambiguity of low-level features [4]. Different objects may produce quite similar visual features and the ambiguity cannot be resolved unless semantic analysis is conducted so that objects within a scene are recognized. Recent psychophysical studies have revealed the importance of high-level semantics in object recognition and scene interpretation. It has been shown that top-down facilitation in recognition is triggered by early information about an object, and also by contextual associations between the object and others with which it typically appears [5,6]. In light of these findings, it is reasonable to model the novelty of image scenes based on semantics within the scene.

While the 'interestingness' of an image can be a subjective concept, novelty as contrast in experience can be more tangibly modelled. Even so, novelty can also be treated differently depending on our individual experiences. For instance, a scene with penguins on a grass field may seem novel to a person, but may not be so to another who has seen that before. Hence, novelty is very much domain and experience dependent. In this paper, we consider the problem of novelty detection in wildlife scenes. The choice of the wildlife domain aligns with our love of animals in general—wildlife documentaries remain a favourite genre on TV and social media websites (such as YouTube), also reflecting the increasing awareness of natural environment preservation. On the other hand, even with a limited domain the visual variance

* Corresponding author. Tel.: +64 3 4798090; fax: +64 3 4798311.
E-mail addresses: jddeng@ieee.org, jeremiah.deng@otago.ac.nz (J.D. Deng).

0031-3203/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2012.02.036

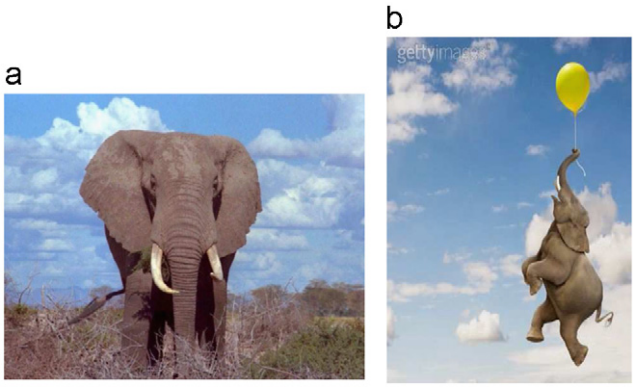


Fig. 1. Scenes with similar objects but different context can either be normal or novel: (a) 'normal' scene and (b) 'novel' scene.

generated by various animals and background objects still attribute to a non-trivial complexity for video analysis problems to be tackled. We are therefore particularly interested in using image analysis and machine learning techniques to detect interesting or unusual scenes in wildlife video. This capability can lead to useful applications such as content filtering and wildlife monitoring.

In this study, analogous to psychological findings, the concept of 'novelty' is not semantically pre-defined but rather decided in contrast to the visual 'experience'. This experience, as we have explained, is necessary to be defined by semantic context of the scene. To detect novelty, a computer system is not only required to have the abilities to analyze images and classify objects but also to model the image semantics. For wildlife images, our normal expectation is that the animals reside to their habitats or a normal natural environment. For instance, dolphins usually appear in water, and zebra pictures have grass or land in the background. When a zebra appears with a dolphin in water, a novelty or anomaly shall be indicated. On the other hand, an image with similar objects may be classified into the same scene category, however, the different spatial context of the objects may still create novelty. For example the two images in Fig. 1 have similar occurrence of object classes ('elephant' and 'sky'), but their spatial contexts are different. Fig. 1(b), even with the small balloon ignored, presents a novel scene because elephants normally walk under the sky rather than float in it.

We therefore propose a computational framework that models the semantic context, so that novelty detection can be conducted based on comparing the semantic context representations. In the remainder of the paper, we will first briefly review some relevant work in Section 2. Our computational framework and the novelty detection algorithm are introduced in Section 3, and the experiment results are presented in Section 4. This is followed by the conclusion, where possible applications of the proposed approach are identified, and the limitations of the current work are acknowledged and future directions outlined.

2. Related work

From a statistical point of view, novelty detection is equivalent to anomaly detection or outliers detection. There are three fundamental approaches to the problem of novelty detection [7]. Type 1 is to determine outliers without prior knowledge, an approach that is analogous to unsupervised clustering; Type 2 is to model both normality and abnormality with pre-trained samples, which is analogous to supervised classification; finally, Type 3 refers to semi-supervised recognition, where only normality is modelled but the algorithm learns to recognize abnormality. Recently there is a growing attention paid to novelty detection and there have been a few informative reviews of various statistical, machine learning, and neural network algorithms [7–9,3]. Generally, corresponding to different approaches for novelty detection, these include clustering algorithms for Type 1, two-class classifiers (such as support vector machines and multi-layer perceptrons) for Type 2, and one-class classifiers (such as hidden Markov models, support vector machines, and statistical tests) for Type 3.

With regard to application areas, novelty detection has gained popularity in textual information retrieval research [10,11]. Novelty detection in video sequence has also been studied with the growing demand in surveillance applications [12,13]. These methods dealing with video frames or motion trajectories are usually data-driven and do not consider object extraction or semantic modelling. Low-level features, such as edge histogram [14], can be extracted from webcam video frames which are then processed by nearest neighbour classifiers to decide whether a scene is abnormal or rare. A cognitively motivated approach is also proposed [15], employing low-level image features based on Itti's visual attention theory [16], but semantic analysis is avoided so as to keep computational cost to the minimum. Our approach differs from these in that we do not consider motion and trajectory information acquirable from video data, but rather concentrate on conducting semantic analysis of individual scenes and then employ statistical analysis to detect novelty.

For wildlife surveillance, it seems that image analysis techniques have found little utilization yet [17]. Other application areas include mechanic damage detection [18], TCP traffic monitoring [19] and intrusion detection [20]. One interesting approach is to transform network traffic into image sequences and employs discrete cosine transform to detect sudden changes related to anomaly [21].

Anomaly detection in images has been explored in specific domains such as biomedicine and geography where unconventional imaging methods are involved, e.g., high-content screening assays for pharmaceutical research [22] in which a set of descriptors that determines the types of detectable anomalies are previously defined; and the algorithms to search the image for pixels whose spectral content is significantly different from that of background in hyper-spectral imagery [23]. These novelty detection methods are however not applicable to wildlife images. This is because the settings of anomalies in natural images are sparse and can hardly be pre-defined. Besides, we are interested not in pixel or region level novelty, but in the context of the overall scene. In this paper, we therefore adopt the Type-3 approach of novelty detection, only modelling the normality of wildlife scenes that are commonly seen. The reason is straightforward—there are numerous possible variations in the abnormality data which will be troublesome either to collect or to model. Since we adopt the Type-3 approach and model the semantic context of normal scenes only, there is no need to generate artificial abnormality data [24].

In image processing, co-occurrence matrices were proposed by Haralick as a texture feature representation [25]. It has been mainly used in low-level contexts, e.g., colour texture analysis [26], or in combination with colour information for object recognition to deal with intensity variance [27]. Co-occurrence matrices have also found application in other areas, e.g., TCP traffic monitoring [19].

Closely related to novelty detection in images is that of interestingness discrimination. Even though human studies on interestingness discrimination have been carried out [28], identifying relevant properties of images that are helpful for computational modelling of interestingness remains a challenge as the matter relates not only to computer vision but also to psychology

and aesthetics. On the other hand, novelty stands out from the background of normal scenes and it is psychologically plausible that scene categorization occurs at the same time.

Semantic representations of images have been explored for scene classification [29,30]. Overall, these approaches are to identify semantics as the set of objects that appear in the image, and the scene is described by the statistical modelling of the semantic objects, e.g., using probabilistic latent semantic analysis [31,32]. A recent work introduced a hierarchical generative model that makes use of text information together with visual features to perform three tasks: image level classification, individual object annotation as well as pixel level segmentation [33]. Other state-of-art techniques include the Gist descriptor which has been found effective in modelling outdoor scenes [34,35], and the bag-of-visual-words approach which is inspired by text retrieval research but builds a visual vocabulary from visual words extracted from local descriptors such as SIFT in order to model the scene types [36,37]. These models however are aimed mainly at scene classification or understanding, and do not deal with novelty detection.

This work differs from the previous work in several aspects. We consider semantic modelling as the basis for novelty detection. The co-occurrence matrix is constructed on the semantics level, modelling the co-occurrence of image block labels. We then use within-class distance modelling and thresholding in deciding outliers [18] and construct a system of multiple one-class classifiers. In short, the statistical modelling in our approach is conducted on the co-occurrence of image semantics.

3. Computational framework

Here we propose a framework as shown in Fig. 2. The input images are first segmented into homogeneous segments, then cut into equal-size image blocks, after which colour and texture features are extracted. These blocks can then be classified using classifiers trained on the visual features. To achieve the goal of novelty detection, we construct co-occurrence matrices of block labels and train a number of one-class classifiers, each modelled on a scene category. If a scene is rejected by all these one-class classifiers, it is then considered as novel or interesting. Hence, by going through the same process, both scene classification and novelty detection tasks can be accomplished.

Next, we present detailed description of the data processing procedure step by step.

3.1. Image segmentation

A visual scene is usually complicated in the sense that it contains multiple objects, which are typically of different visual characteristics. A single object, e.g., a bird, may also have several regions of different visual appearance such as the beak, eyes and feathers. By segmenting an image into homogeneous regions, it facilitates detection or classification of these objects. We choose the JSEG algorithm [38] for segmentation because of its efficiency and effectiveness.

In JSEG, colours in the image are first quantized to some representing classes that are used to separate the regions in the image. Image pixel colours are then replaced by their corresponding colour class labels to form a class-map of the image. Two parameters are involved in the algorithm: a colour quantization threshold and a merger threshold. The smaller the colour quantization threshold is, the more number of quantized colours will result from the image. On the other hand, the merger threshold controls the merging of identical colours in neighbouring regions; the smaller the value, the more regions to be separated. We have experimented with the default setting of the JSEG package, with the colour quantization threshold set as 250, and the merger threshold as 0.4, and it is found that good segmentation results can be obtained from most of the images that we deal with (Fig. 3 gives an example). Small segments, often of background clutter, will be ignored if they are less than 10% of the image size. The remaining segments are then stored as individual segment images, and when used for training will be labelled manually as the groundtruth.

The parameter setting allows us to over-segment an object so as to deal with the visual variance within the segments. However, even with over-segmentation sometimes the visual homogeneity of the segments cannot be always satisfactory. This means that it will be challenging to build classifiers to recognize these image

Fig. 3. Segmentation of a 'giraffe' scene.

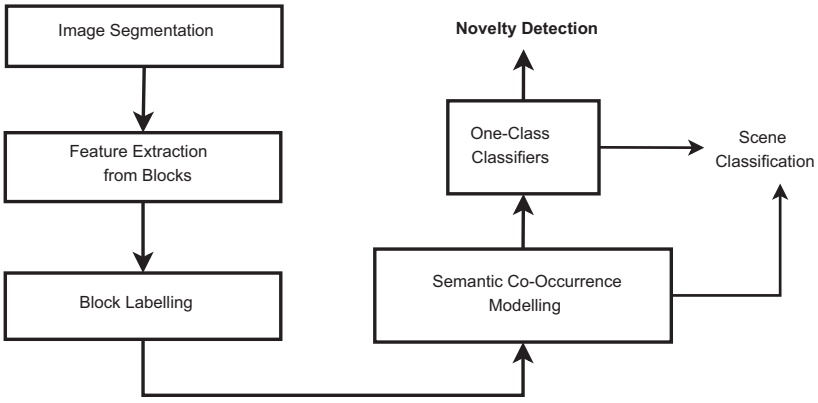


Fig. 2. The computational framework for image novelty detection and scene classification.

segments. To cope with this drawback, each segment image is further tiled into $b \times b$-pixel blocks where $b \in N$. We want to ensure that for small segments they contribute at least one image block. The smallest segment image we have in the training data is of the size 31×67 pixels. Hence we set $b=25$ in the further experiments, as smaller b may cause inefficiency in generating texture features. Image blocks that fall out of the segment boundaries will be ignored. There are two options for us to build classifiers, either on the image segments, or on the image blocks. Results will be presented later in the following subsection.

3.2. Feature extraction and block labelling

3.2.1. LUV histograms

Visual features are then extracted from image segment blocks. First we employ a colour histogram in the LUV colour space to encode the colour information of image blocks. Colour histograms are found to be robust to resolution and rotation changes. The LUV colour space is adopted because it models human's perception of colour similarity very well, and is also machine independent [39]. As the segment blocks are quite homogeneous in colour, we do not use a joint 3-D histogram (which would be too sparse) but concatenate bins of three separate colour bands into a 1-D histogram. The bins are quantized in a fine granularity. The LUV channels have different ranges: L (0−100), U (−134 to 220) and V (−140 to 122). Each channel is quantized with the same interval, thus giving 20 bins to the L channel, 70 bins to the U channel, and 52 bins to the V channel. Apart from these, the standard deviation of all the bin values is also included. The LUV histogram feature code thus has 143 dimensions.

3.2.2. Texture features

Texture features extracted from the image blocks are also included as local features. Image blocks are first converted to gray scale. We consider Edge Histogram Descriptor and Gabor filtering features [40] with the Haralick features [25], the latter performed the best in our test as revealed from the results in Table 1 and are therefore adopted in further experiments. The Haralick texture features consist of a few statistical properties based on the gray scale co-occurrence matrix. The co-occurrence matrix is a two-dimensional histogram composed from the pairwise statistics of gray scale co-occurrence among adjacent pixels. Four orientations are considered, each giving a co-occurrence matrix for an image block. A total of 13 statistical measures can be calculated for each co-occurrence matrix. The mean and deviation values of each of these 13 measures over the four orientations form a feature vector of 26 dimensions.

3.2.3. Block labelling

Classification was done on the segments and also on the image blocks with similar feature extractions by using 10-fold cross validation test on the training data, the classifier employed is nearest neighbour (1-NN). The results are as shown in Table 1. It can be seen that for segment classification the highest classification accuracy achieved is only as low as 74%. The existence of heterogeneity in segments probably contribute to the low performance for segment texture features. It is even detrimental to combine colour and texture features for segment images. On the other hand, these combinations are usually beneficial for block classification. From the results, it can be seen that in terms of texture representations only, Gabor features outperform Haralick and EHD textures. The best performance is from the combination of LUV colour and Haralick texture features that can classify 90.05% of the image blocks correctly. It seems that LUV colour histograms and Haralick texture features can complement each other to boost up the classification performance in our case.

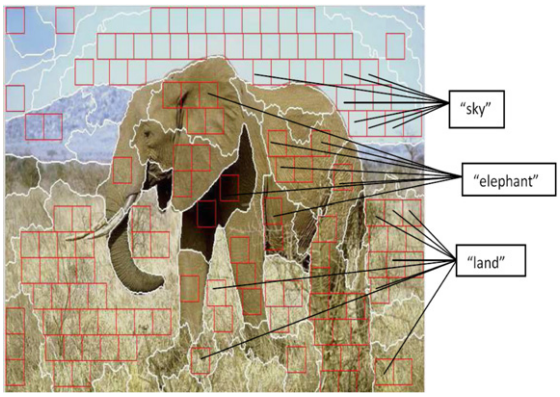
Having evaluated the features with the above experiment, we adopt the LUV and Haralick features and concatenate them together, giving a feature vector of 169 dimensions to represent an image block. Through manual labelling of image segments, semantic ground truth is assigned to the training images. The image blocks inherit semantic labels from their corresponding segments. Their feature codes along with the relevant class labels are used to train object classifiers.

Table 1
Validation of feature schemes used for block labelling: 10-fold cross validation is conducted over training image blocks and average accuracy values (in %) are listed.

Feature scheme	Segments	**Image blocks**
LUV	74.16	87.29
EHD	40.71	47.67
Gabor	33.28	87.47
Haralick	37.5	65.55
LUV+EHD	62.5	81.07
LUV+Gabor	59.12	87.47
LUV+Haralick	59.97	**90.05**

3.3. Semantic context modelling

To model the semantic context within a scene, we further generate a *block label co-occurrence matrix* (BLCM). After the labels of all image blocks are obtained the image are scanned from left to right and top to bottom, and the co-occurrence of labels for blocks within a distance threshold R is collected. The co-occurrence statistics is gathered across the entire image and normalized by the total number of image blocks. Obviously the variation on the object sizes will affect the matrix values of BLCM. To reduce this effect, one option is to binarize the values of BLCM elements, with non-zero elements all set to '1'. Fig. 4 shows an 'elephant' image as an example, with its image block labels and the relevant binary BLCM displayed under the image. To demonstrate the idea, there are eight object labels in this example: 'coast', 'land', 'sky', 'water', 'dolphin', 'elephant', 'penguin', and 'zebra', thus giving an 8×8 BLCM.



class	coast	land	sky	water	dolphin	elephant	penguin	zebra
coast	0	0	0	0	0	0	0	0
land	0	1	0	0	0	1	0	0
sky	0	1	1	0	0	1	0	0
water	0	0	0	0	0	0	0	0
dolphin	0	0	0	0	0	0	0	0
elephant	0	1	1	0	0	1	0	0
penguin	0	0	0	0	0	0	0	0
zebra	0	0	0	0	0	0	0	0

Fig. 4. Image blocks in an 'elephant' image and its corresponding co-occurrence matrix. The matrix is read row-column, e.g., a 'sky'→'land' entry of '1' indicates there is 'sky' block above (or to the left of) 'land' block.

The dimension of the BLCM will depend on the number of object classes in the knowledge or database. Hence, with C classes of objects to model from the given scenes, it gives a BLCM with $C \times C$ dimensions. Since the scanning of image blocks is directional, the BLCM is consequently asymmetric. This is however a desirable feature as we intend to keep the spatial location information in the semantic context representation. For example, scanning top-down, a scene context of 'sky'→'land' is common, while a 'land'→'sky' BLCM entry may suggest some novelty. This is however not explicitly defined but rather depends on the actual context statistics built across images. On the other hand, images with similar characteristics, i.e., with similar types of foreground and background settings, will produce quite similar BLCM features, which is helpful for scene classification.

Note that for images with salient foreground objects, normally the BLCM is quite sparse. We concatenate the BLCM matrix rows into a 1-D vector for further processing. It is possible to use principle component analysis (PCA) for dimension reduction. The dimension-reduced BLCM vectors are then ready to be further processed by classifiers.

3.4. Building scene classifiers

The scene classification task is two-fold. Given an image, if it is of an ordinary scene, the classifiers need to classify it into the right scene category; otherwise, novelty should be reported. This is however not a typical multi-class classification scenario, since there may be neither similarity nor restriction in the appearance of 'novelty', and it is usually hard to find sufficient training data especially for the 'novelty' class. Hence we resort to one-class classification. To build a one-class classifier for each of the scene types, the classifiers need only normally labelled images for training. We assume, for each scene type, in the BLCM feature space there will be a dense cluster related to 'normal' images, while 'novel' images will be sparsely distributed around these clusters. Testing images are assessed by calculating their BLCM feature distance to the trained one-class centres. If the distance is smaller than a defined threshold for that class, it is accepted by that one-class classifier; otherwise it is rejected from that class. A data instance rejected by all one-class classifiers is then reported as an outlier. With this approach, scene classification and novelty detection can be carried out at the same time.

We call our method 'Multiple One-class Classification with Distance Thresholding' (MOC-DT). Its training and testing algorithms are summarized as follows:

Training of MOC-DT.

1. For images of each image scene type, extract their BLCM code and calculate the centroid, denoted as μ_i, of the image group. Here $i = 1, 2, \ldots, N$, N is the number of scene types.
2. Calculate the distance towards μ_i for all instances of the same scene and obtain the distance mean m_i, and the standard deviation, σ_i.
3. Set the outlier distance threshold as $T_i = m_i + k\sigma_i$ for the scene type, where k is a constant.

Testing of MOC-DT.

1. Given a query image, calculate its BLCM code and its distance towards each image group as d_i.
2. If $d_i > T_i$, $\forall i = 1, \ldots, N$, label the image as 'novel'; otherwise, assign the scene label c to the image, $c = \arg_i \min(d_i)$, if $d_i < T_i$.

The value of parameter k remains to be decided. If k is too low, it may cause the 'normal' images misclassified as 'novel'; if it is too high, it may become insensitive to 'novel' images. If we assume a Gaussian distribution of the distance values, a $k=1$ setting would mean that about 84% of 'normal' cases will be counted as true 'normal' ones. While this seems unfair to other less normal cases, it gives the classifier a good sensitivity to detect novelty. We therefore adopt $k=1$ in our experiments.

Both the Manhattan distance as well as the Mahalanobis distance can be used on the BLCM codes. In the next session we present our experiment results, verifying our assumptions based on which the MOC-DT algorithm is derived, and comparing the performance of the proposed algorithm with that of other two one-class classifiers.

4. Experiments and results

4.1. Experiment settings

To validate the proposed approach, we conduct a few experiments using wildlife images and have the performance on scene classification and novelty detection measured.

The normal image set consists of 335 normal wildlife scenes and are taken from Google image Search and Microsoft Research (the 'cow' image set). These wildlife images usually feature one type of animal in the scene with a few other object classes in background to form the semantic context. There are six scene types (each followed by the number of instances in each type): 'cow' (51), 'dolphin' (57), 'elephant' (65), 'giraffe' (50), 'penguin' (55) and 'zebra' (57). The distribution of number of images across different scene types is roughly even. The background objects belong to eight object classes, namely: 'grass', 'land', 'mount', 'rock', 'sky', 'snow', 'trees' and 'water'. Therefore, counting both animal and background objects, there are 14 object classes in total. For these normal images, their sizes vary from 197×240 to 1111×444 pixels. Some sample 'normal' images are shown in Fig. 5.

We use 43 'novel' images taken from Google Image Search for testing. Compared with the normal images, these 'novel' images basically carry quite different semantic context. For instance, we regard scenes with zebra and dolphin together, penguin and grass, zebra and snow etc., as 'novel', because the same context is not present in the training set. Sample 'novel' images are shown as in Fig. 6.

For block labelling we have therefore 14 class labels, covering both the background and animals in the foreground. We include images of both 'normal' and 'novel' categories so that all object classes get enough image blocks for the training of the block classifiers. After segmentation, there are 65,000 image blocks obtained in total, and these are used to train the classifiers using feature schemes presented in Section 3. We set aside 30 'normal' images together with 43 'novel' images to make up a testing image set of 73 images to be used in the novelty detection evaluation. For the classification of image blocks, an average accuracy of 84.6% is achieved in a 10-fold cross validation using a 1-NN classifier. This we believe gives a sound basis for constructing the BLCM semantic context based on the labelled blocks.

To ensure that block classifiers are not too dependent on the training data but rather generalize quite well we adopt another 10-fold cross validation process when dealing with the testing images. Specifically, a random sampling of 90% of all image blocks are set aside to be used to train a classifier each time, which then labels the image blocks of the testing images. This process is repeated 10 times. Each time the BLCM codes of the testing images are generated and further computation is involved to either generate a scene class prediction, or report novelty.

Fig. 5. Sample 'normal' images from six scene classes: 'dolphin', 'elephant', 'giraffe', 'penguin', 'zebra' and 'cow'.

4.2. Results

4.2.1. Pre-processing

The semantic context of image scenes are calculated, undergoing binarization and PCA. The BLCM data has a higher dimensionality (196) than the number of images in each scene type (50–65), making it impossible to use the Mahalanobis distance. Therefore PCA is conducted to reduce the dimension of BLCM. The numbers of PCs are decided empirically in a test run of scene classification and novelty detection using the MOC-DT algorithm. As can be seen from Fig. 7, for scene classification the performance over different number of PCs is rather stable with four PCs or more included, while the performance on novelty detection hits a plateau when the number is between 6 and 12. Taking both into account, we choose 10 to be the number of PCs and extracted the first 10 PCs of BLCM codes in our further experiments when PCA is involved.

To visualize the generated BLCM codes of different scene types, a 2-D projection is produced using the first two PCs of the data, as shown in Fig. 8. Note that the projected data display a clustered structure with good separability for all the scene types despite minor overlaps. The 'cow', 'elephant' and 'zebra' data points are quite close to each other, probably due to the fact that the relevant images have very similar background objects.

4.2.2. BLCM for scene classification

Next, we assess the performance of BLCM for scene classification. The objective of this assessment is to check if the high-level BLCM feature is representative enough for scene types. If so, then the novelty detection based on the BLCM features can proceed. We conduct 10-fold cross validations on the training dataset. Five feature schemes are employed for a comparison study: global scene descriptor (GSD), local binary pattern (LBP) [41], Gist [34], BLCM, BLCM with PCA (BLCM/PCA), binary BLCM (B-BLCM), and binary BLCM with PCA (B-BLCM/PCA). The GSD uses the same low-level features as used to construct the BLCM: LUV colour histograms and Haralick texture features, but these are extracted globally from the entire image. As a frequently used texture descriptor, LBP is a global histogram on local binary patterns assessed across the entire image. It is a powerful texture feature that has been used for scene classification [35]. We use a basic version of LBP, with a dimensionality of 256. Gist as a state-of-art scene classification feature scheme has been found to perform very well on outdoor scenes [34]. The features are computed from

Fig. 6. Sample 'novel' images used for testing.

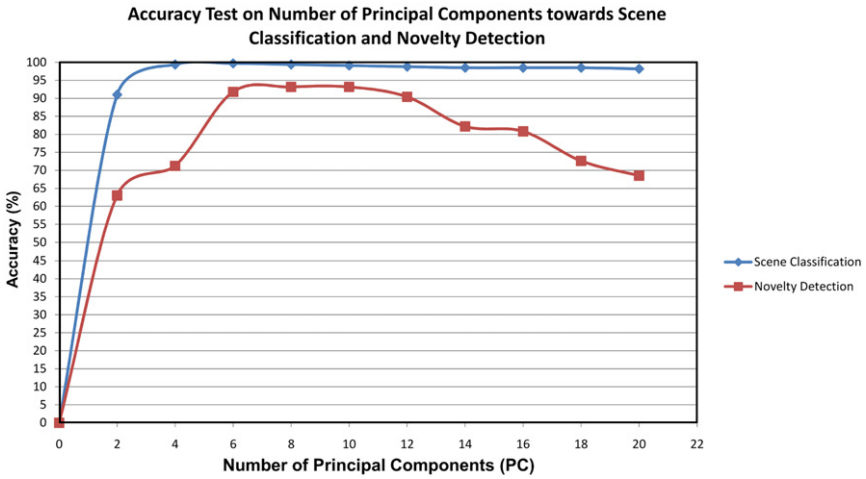


Fig. 7. Accuracy results for classifying 'normal' images using MOC-DT based on different number of principal components.

Gabor transforms by multiscale oriented filters. We use filters of eight orientations and four scales, and the dimensionality of the Gist feature is 512.

Apart from the average recognition accuracy for the scene types, we also report the area under the ROC curve (AUC) [42], which is often used to assess the performance of machine learning algorithms. The results obtained by the nearest neighbour algorithm (1-NN) are presented in Table 2. It can be seen that among the low-level features Gist outperforms GSD and LBP, but the BLCM schemes achieve much better performance in scene classification compared with using low-level visual features directly. Also feature schemes with binarization (B-BLCM and B-BLCM/PCA) compare favourably with their non-binary counterparts. The PCA operation seems to weaken the features slightly in general, but their performance remains high. This suggests that the proposed BLCM can facilitate scene representation comparing with low-level visual features. To save space we only list the performance of the nearest neighbour (NN) algorithm, but other algorithms such as C4.5, random forest, and support vector machine have been experimented and report similar results.

4.2.3. Novelty detection

For novelty detection, the testing images are processed by the proposed MOC-DT algorithm. We use the Mahalanobis distance to calculate the distance between the BLCM code to its relevant class centre. This gives us better performance than using Manhatten distance which was reported in Ref. [43].

Fig. 9 demonstrates the one-class modelling of the 'elephant' class. The distance data calculated fit well into a normal distribution, as shown by the normal quantile plot in Fig. 9(a). The p-value calculated for the χ^2-test is 0.085, meaning that the null hypothesis on a normal distribution is accepted on a 0.05 significance level. Fig. 9(b) plots out the fitted Gaussian curve (normalized), the distance data points for all normal and novel images encountered by the one-class model, where a distance threshold is also derived based on the statistics of normal data.

On novelty detection performance we compare the MOC-DT algorithm with the probability density based one-class classier (PDOC) [44] and Support Vector Machine one-class classifier (SVM-OC) [45]. To train the one-class classifiers, for each class model the corresponding class is treated as target class while the

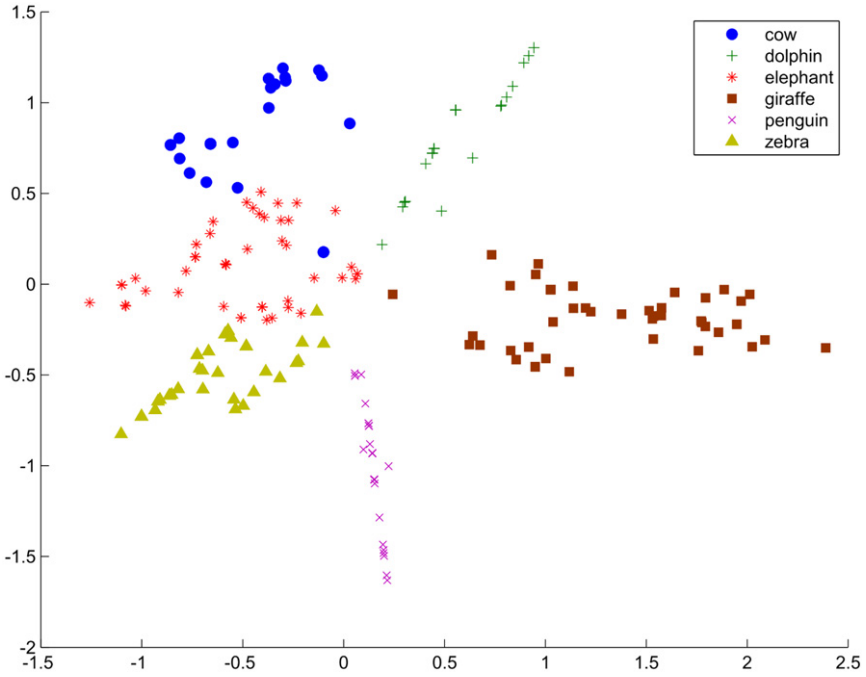


Fig. 8. The 2-D PCA projection of BLCM data for 'normal' images shown with their respective classes.

Table 2
Ten-fold cross validation of feature schemes used for scene classification: average accuracies (in %) and average AUC values are listed.

Feature	AUC	Accuracy
GSD	0.763	62.4
LBP	0.747	58.0
Gist	0.808	67.3
BLCM	0.977	95.8
BLCM/PCA	0.934	89.0
B-BLCM	1.000	99.4
B-BLCM/PCA	0.997	99.1

rest are considered outliers. Therefore PDOC, SVM-OC and MOC-DT are compared under the same treatment on the BLCM data, but the difference is only on the one-class classification algorithm used.

For performance evaluation, we denote cases when novel images classified as 'novel' as True Positive (TP), normal images classified as 'novel' as False Positive (FP), and novel images classified as 'normal' as False Negative (FN). The precision, recall, and F-measure can be calculated as follows:

$$\text{Precision } (P) = \frac{\text{TP}}{\text{TP}+\text{FP}} \tag{1}$$

$$\text{Recall } (R) = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{2}$$

$$\text{F-measure} = \frac{2PR}{P+R} \tag{3}$$

We experiment with two BLCM codes: BLCM/PCA and B-BLCM/PCA. The image block label data are used in a 10-fold manner to train a block labeller, which then generates the BLCM code from the training and testing image sets. One-class classifiers are then trained and then tested 10 times, each time with potentially different BLCM data. The average precision, recall, F-measure, and AUC value for each of the three one-class classifiers are calculated and given in Table 3.

As the results clearly indicate, MOC-DT has the highest score on all of the performance indicators as compared with the other two classifiers. Unlike other algorithms, MOC-DT seems to have less difference between the precision and recall values, which is desirable. Note however that we see no obvious gain in conducting binarization on the BLCM code as the F-measure is basically the same in two cases and the AUC value only differs slightly.

4.2.4. Scene classification

During testing, scene images either are rejected by all one-classifiers and labelled as 'novel', or, they will be classified into a certain scene type, according to the testing algorithm given in Section 3.4. Table 4 reports the per-class performance of MOC-DT implementations on BLCM/PCA and B-BLCM/PCA. The latter seems to improve the performance over all scene types except 'dolphin' and 'elephant', where lower recalls are reported. Overall, B-BLCM/PCA gives higher precision but lower recall than BLCM/PCA while the average F-measure is slightly improved.

For both feature schemes, it seems that MOC-DT outperforms PDOC and SVM-OC in scene classification by evaluating the average F-measure, as shown in Tables 5 and 6. The difference between precision and recall values is much larger for PDOC and SVM-OC over most of the scene types, while for MOC-DT this difference is mostly rather small. It is noted that the binarization of BLCM seems to boost the performance of these classifiers.

The benefit of using binarized BLCM code is however demonstrated in a few cases where the abnormal size contrast of objects seems to mislead the one-class classifiers working on the BLCM-PCA code. Fig. 10 shows two of such kind of 'novel' images incorrectly classified into normal scene types. However, when using the B-BLCM-PCA code, both of them can be correctly classified as 'novel'.

5. Conclusion

Novel scene detection can be better achieved by working with semantic context modelling in images. In this paper we have proposed a simple but effective computational framework that conducts semantic context modelling and novelty detection.

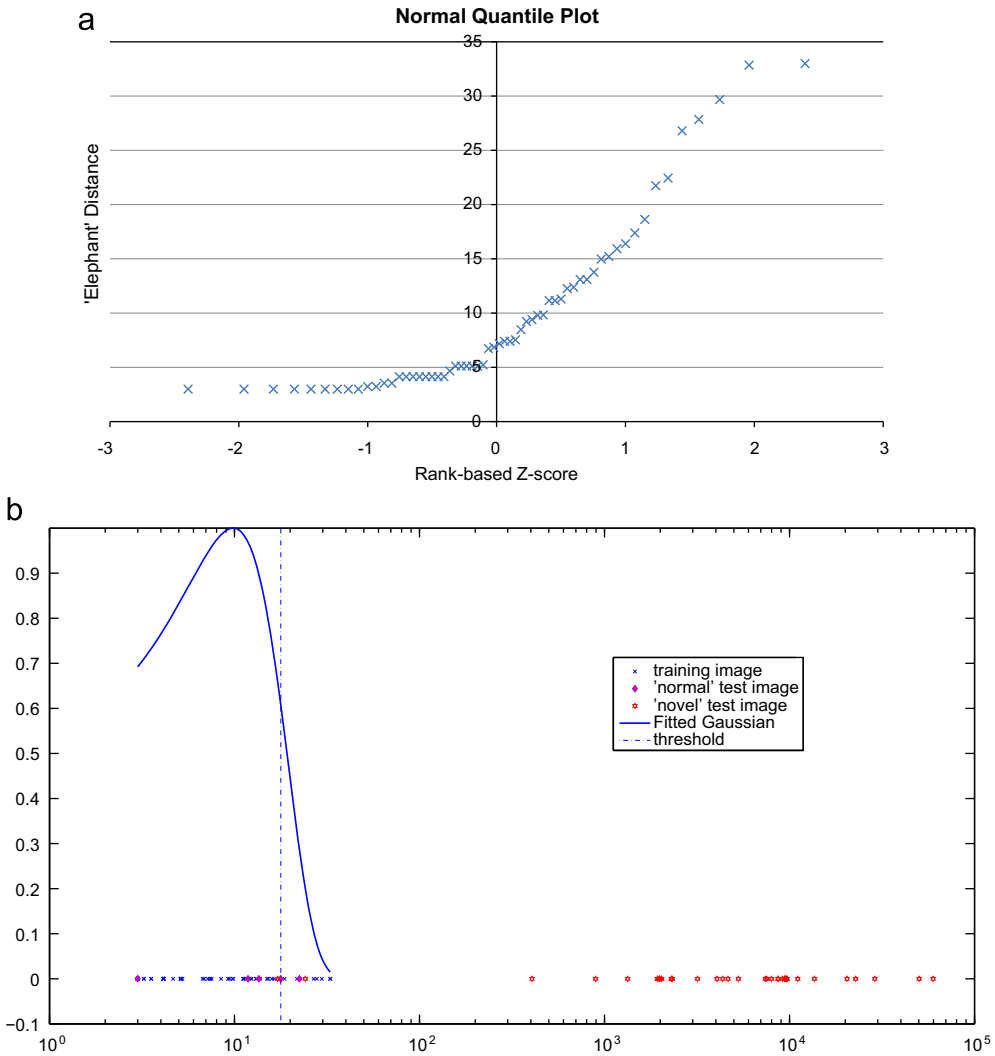


Fig. 9. Gaussian modelling of the semantic distances of the 'elephant' set. (a) The distance data fit well into a Gaussian despite a few outliers. (b) The fitted Gaussian curve is displayed along with the data points and the threshold line. (a) Normality test. (b) Distance thresholding.

Table 3
Performance on novelty detection. The F-measure and AUC values of the best classifiers are shown in bold.

Feature	Perf.	Classifier		
		MOC-DT	PDOC	SVM-OC
BLCM/PCA	Pre.	0.90	0.75	0.62
	Rec.	0.93	0.53	0.46
	F	**0.91**	0.62	0.53
	AUC	**0.85**	0.70	0.58
B-BLCM/PCA	Pre.	0.86	0.62	0.79
	Rec.	0.98	0.39	0.95
	F	**0.91**	0.48	0.84
	AUC	**0.89**	0.57	0.81

Co-occurrence matrices, which are usually used as a texture feature, are extended to model high-level semantic context. The proposed BLCM code then undergoes principal component analysis for dimension reduction. One-class classifiers are built around the BLCM code for each scene type and a simple distance threshold method is employed for classification. One feature of the proposed framework is that apart from detecting novelty, these one-class classifiers can also classify a normal image into its corresponding scene category.

We have found that by employing the Mahalanobis distance better performance can be achieved on the BLCM code, but because of the relative high dimensionality of the BLCM code we have to conduct PCA so that the dimensionality is low enough to enable the computation. Nevertheless, this scheme seems to work well. Another finding is that binarized BLCM with PCA can better cope with object size variations when encoding the semantic context. Our experiments on a set of images of six scene classes have given some promising results, and the proposed MOC-DT algorithm performs favourably compared with other one-class classification algorithms.

Although the scale of our experiments is limited, and the performance of image block labelling can be further enhanced, our framework seems flexible for further expansion and improvement. In the future we will take more object/scene categories into consideration and conduct more experiments using large-scale datasets, e.g., [35]. Also, instead of using crisp classification of the labels using the BLCM code, probability-based or fuzzy prediction can be considered in order to mitigate the side-effect caused by potential misclassification of image blocks. We also intend to conduct more comparison with other state-of-art scene classification methods, including bag-of-visual-words.

The idea of exploring semantic clues so as to automatically identify novelty out of a context can easily find practical

Table 4
MOC-DT performance (precision, recall and F-measure) on scene classification per scene class: C, cow; D, dolphin; E, elephant; G, giraffe; P, penguin; Z, zebra. Average performance across all classes is shown on the last column. Better F-measures are in bold.

Feature	Perf.	Classes						Avg.
		C	*D*	*E*	*G*	*P*	*Z*	
BLCM/PCA	Pre.	0.823	1.000	0.805	0.773	1.000	0.847	0.875
	Rec.	0.940	0.920	0.770	0.760	0.560	0.960	0.818
	F	0.876	0.956	**0.786**	**0.763**	0.710	**0.896**	0.831
B-BLCM/PCA	Pre.	1.000	1.000	0.777	1.000	1.000	1.000	0.963
	Rec.	0.923	0.920	0.720	0.380	0.900	0.760	0.767
	F	**0.955**	0.956	0.743	0.548	**0.942**	0.861	**0.834**

Table 5
PDOC performance (precision, recall and F-measure) on scene classification per scene class: C, cow; D, dolphin; E, elephant; G, giraffe; P, penguin; Z, zebra. Average performance across all classes is shown on the last column. Better F-measures are in bold.

Feature	Perf.	Classes						Avg.
		C	*D*	*E*	*G*	*P*	*Z*	
BLCM/PCA	Pre.	0.199	1.000	0.850	0.165	0.080	0.167	0.410
	Rec.	0.400	0.380	0.180	0.420	0.380	0.400	0.360
	F	0.266	**0.551**	0.297	**0.237**	0.132	0.236	0.287
B-BLCM/PCA	Pre.	1.000	0.167	0.400	0.066	0.583	0.333	0.333
	Rec.	0.600	0.500	0.400	0.267	0.800	0.300	0.478
	F	**0.750**	0.250	**0.400**	0.106	**0.674**	**0.316**	**0.416**

Table 6
SVM-OC performance (precision, recall and F-measure) on scene classification per scene class: C, cow; D, dolphin; E, elephant; G, giraffe; P, penguin; Z, zebra. Average performance across all classes is shown on the last column. Better F-measures are in bold.

Feature	Perf.	Classes						Avg.
		C	*D*	*E*	*G*	*P*	*Z*	
BLCM/PCA	Pre.	0.387	0.395	0.310	0.344	0.326	0.496	0.376
	Rec.	0.400	0.600	0.400	0.600	0.800	0.580	0.563
	F	0.393	0.476	0.350	0.437	0.463	0.535	0.442
B-BLCM/PCA	Pre.	1.000	1.000	0.667	1.000	1.000	0.800	0.911
	Rec.	0.600	0.500	0.400	0.600	1.000	0.800	0.650
	F	**0.750**	**0.667**	**0.500**	**0.750**	**1.000**	**0.800**	**0.744**

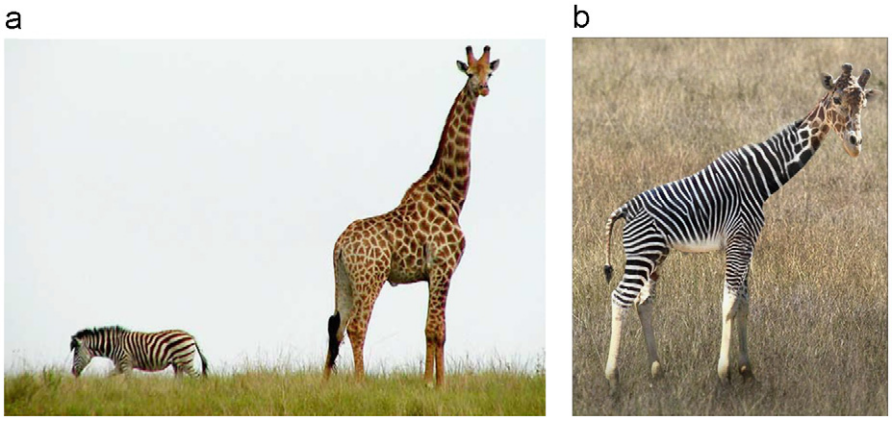


Fig. 10. Two 'novel' scenes that are misclassified as 'giraffe' and 'zebra' respectively by using BLCM/PCA, but correctly detected by using B-BLCM/PCA.

applications. Our approach may be introduced into a picture collection management system that automatically detects pictures of novelty. The idea can be extended into other areas of data processing. For instance, our approach may provide an improved solution to video data management and retrieval, such as by incorporating semantic clues for better key frame extraction and video summarization [46], or building an online wildlife monitoring system.

The novelty or 'interestingness' of a scene is tightly coupled with visual experience and therefore can be dynamic and also subjective to different viewers. It is our intention to investigate using some incremental algorithms to model the semantic

context and its classification. On another note, we have taken basically a statistical approach in modelling novelty, without taking any further semantic or aesthetic aspects of novelty or interestingness into consideration. To be able to tell whether a picture is 'natural', 'surreal', or 'stunning', while a statistical modelling approach may remain relevant, some other visual features that are inherently semantic or computational aesthetics oriented [47] need to be extracted from the images so that the system can learn these abstract concepts.

References

[1] P.J. Silvia, Exploring the Psychology of Interest, Oxford University Press, 2006.
[2] D. Berlyne, Stimulus Selection and Conflict, McGraw-Hill Book Company, 1960.
[3] V. Chandola, A. Banerjee, V. Kumar, Anomaly detection: a survey, ACM Computing Surveys 41 (2009) 1–58.
[4] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, Content-based image retrieval at the end of the early years, IEEE Transaction on Pattern Analysis and Machine Intelligence 22 (12) (2000) 1349–1380. doi:10.1109/34.895972.
[5] M.J. Fenske, E. Aminoff, N. Gronau, M. Bar, Top-down facilitation of visual object recognition: object-based and context-based contributions, Progress in Brain Research 155 (2006) 3–21.
[6] J.A. Stirk, G. Underwood, Low-level visual saliency does not predict change detection in natural scenes, Journal of Vision 7 (10) (2007) 1–10.
[7] V.J. Hodge, J. Austin, A survey of outlier detection methodologies, Artificial Intelligence Review 22 (2004) 85–126. doi:10.1007/s10462-004-4304-y.
[8] M. Markou, S. Singh, Novelty detection: a review. Part 1. Statistical approaches, Signal Processing 83 (12) (2003) 2481–2497. doi:10.1016/j.sigpro.2003.07.018.
[9] M. Markou, S. Singh, Novelty detection: a review. Part 2. Neural network based approaches, Signal Processing 83 (12) (2003) 2499–2521. doi:10.1016/j.sigpro.2003.07.019.
[10] E. Gabrilovich, S. Dumais, E. Horvitz, NewsJunkie: providing personalized newsfeeds via analysis of information novelty, in: WWW'04: Proceedings of the 13th International Conference on World Wide Web, ACM, New York, NY, USA, 2004, pp. 482–490. doi:10.1145/988672.988738.
[11] X. Li, W.B. Croft, An information-pattern-based approach to novelty detection, Information Processing and Management 44 (3) (2008) 1159–1188. doi:10.1016/j.ipm.2007.09.013.
[12] D. Pokrajac, A. Lazarevic, L.J. Latecki, Incremental local outlier detection for data streams, in: Proceedings of IEEE Symposium on Computational Intelligence and Data Mining, 2007, pp. 504–515.
[13] S. Khalid, Motion-based behaviour learning, profiling and classification in the presence of anomalies, Pattern Recognition 43 (1) (2010) 173–186. doi:10.1016/j.patcog.2009.04.025.
[14] M. Breitenstein, H. Grabner, L. Van Gool, Hunting nessie—real-time abnormality detection from webcams, in: 2009 IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops), 2009, pp. 1243–1250. doi:10.1109/ICCVW.2009.5457468.
[15] J. Kang, M. Aurangzeb Ahmad, A. Teredesai, R. Gaborski, Cognitively motivated novelty detection in video data streams, in: V.A. Petrushin, L. Khan (Eds.), Multimedia Data Mining and Knowledge Discovery, Springer, London, 2007, pp. 209–233.
[16] L. Itti, C. Koch, Computational modeling of visual attention, Nature Neuroscience Review 2 (3) (2001) 194–203.
[17] C. Connolly, Wildlife-spotting robots, Sensor Reviews 27 (2007) 282–287.
[18] G. Manson, G. Pierce, K. Worden, On the long-term stability of normal condition for damage detection in a composite panel, Key Engineering Materials 204–205 (2001) 359–370.
[19] C. Callegari, S. Giordano, M. Pagano, On the use of co-occurrence matrices for network anomaly detection, in: IWCMC '09: Proceedings of the 2009 International Conference on Wireless Communications and Mobile Computing, ACM, New York, NY, USA, 2009, pp. 96–100. doi:10.1145/1582379.1582401.
[20] A. Patcha, J.-M. Park, An overview of anomaly detection techniques: existing solutions and latest technological trends, Computer Networks 51 (12) (2007) 3448–3470. doi:10.1016/j.comnet.2007.02.001.
[21] S.S. Kim, A. Reddy, Image-based anomaly detection technique: algorithm, implementation and effectiveness, IEEE Journal on Selected Areas in Communications 24 (10) (2006) 1942–1954. doi:10.1109/JSAC.2006.877215.
[22] A. Goode, R. Sukthankar, L. Mummert, M. Chen, J. Saltzman, D. Ross, S. Szymanski, A. Tarachandani, M. Satyanarayanan, Distributed online anomaly detection in high-content screening, in: 5th IEEE International Symposium on Biomedical Imaging, ISBI 2008: From Nano to Macro, 2008, pp. 249–252. doi:10.1109/ISBI.2008.4540979.
[23] D. Stein, S. Beaven, L. Hoff, E. Winter, A. Schaum, A. Stocker, Anomaly detection from hyperspectral imagery, IEEE Signal Processing Magazine 19 (1) (2002) 58–69. doi:10.1109/79.974730.
[24] M. Markou, S. Singh, A neural network-based novelty detector for image sequence analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (10) (2006) 1664–1677. doi:10.1109/TPAMI.2006.196.
[25] R.M. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification, IEEE Transactions on Systems, Man, and Cybernetics 3 (1973) 610–621.
[26] C. Palm, Color texture classification by integrative co-occurrence matrices, Pattern Recognition 37 (5) (2004) 965–976. doi:10.1016/j.patcog.2003.09.010.
[27] D. Muselet, L. Macaire, Combining color and spatial information for object recognition across illumination changes, Pattern Recognition Letters 28 (10) (2007) 1176–1185. doi:10.1016/j.patrec.2007.02.001.
[28] H. Katti, K.Y. Bin, T.S. Chua, M. Kankanhalli, Pre-attentive discrimination of interestingness in images, in: IEEE International Conference on Multimedia and Expo, 2008, pp. 1433–1436. doi:10.1109/ICME.2008.4607714.
[29] L. Fei-Fei, P. Perona, A Bayesian hierarchical model for learning natural scene categories, in: IEEE Conference on Computer Vision and Pattern Recognition, 2005, pp. 524–531.
[30] Y.-G. Jing, J. Yang, C.-W. Ngo, A.G. Hauptmann, Representations of keypoint-based semantic concept detection: a comprehensive study, IEEE Transactions on Multimedia 12 (2010) 42–53.
[31] R. Fergus, L. Fei-Fei, P. Perona, A. Zisserman, Learning object categories from Google's image search, in: International Conference on Computer Vision, vol. 2, 2005, pp. 1816–1823.
[32] A. Bosch, A. Zisserman, X. Munoz, Scene classification via pLSA, in: Proceedings of the European Conference on Computer Vision, 2006, pp. 517–530.
[33] L.-J. Li, R. Socher, L. Fei-Fei, Towards total scene understanding: classification, annotation and segmentation in an automatic framework, in: Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR), 2009, pp. 2036–2043.
[34] A. Oliva, A. Torralba, Building the Gist of a scene: the role of global image features in recognition, in: S. Martinez-Conde, S. Macknik, L. Martinez, J.-M. Alonso, P. Tse (Eds.), Visual Perception Fundamentals of Awareness: Multi-Sensory Integration and High-Order Perception, Part B of Progress in Brain Research, vol. 155, Elsevier, 2006, pp. 23–36. doi:10.1016/S0079-6123(06)55002-2.
[35] J. Xiao, J. Hays, K.A. Ehinger, A. Oliva, A. Torralba, SUN database: large-scale scene recognition from abbey to zoo, in: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2010, pp. 3485–3492. doi:10.1109/CVPR.2010.5539970.
[36] S. Lazebnik, C. Schmid, J. Ponce, Beyond bags of features: spatial pyramid matching for recognizing natural scene categories, in: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR'06, vol. 2, IEEE Computer Society, Washington, DC, USA, 2006, pp. 2169–2178. doi:http://dx.doi.org/10.1109/CVPR.2006.68.
[37] J. Yang, Y.-G. Jiang, A.G. Hauptmann, C.-W. Ngo, Evaluating bag-of-visual-words representations in scene classification, in: Proceedings of the International Workshop on Multimedia Information Retrieval, MIR'07, ACM, New York, NY, USA, 2007, pp. 197–206. doi:http://doi.acm.org/10.1145/1290082.1290111.
[38] Y. Deng, B. Manjunath, Unsupervised segmentation of color-texture regions in images and video, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (8) (2001) 800–810.
[39] Y.-F. Ma, H.-J. Zhang, Contrast-based image attention analysis by using fuzzy growing, in: Multimedia'03: Proceedings of the 11th ACM International Conference on Multimedia, ACM, New York, NY, USA, 2003, pp. 374–381.
[40] B.S. Manjunath, J.R. Ohm, V.V. Vasudevan, A. Yamada, Color and texture descriptors, IEEE Transactions on Circuits and Systems for Video Technology 11 (6) (2001) 703–715. doi:10.1109/76.927424.
[41] T. Ojala, M. Pietikinen, D. Harwood, A comparative study of texture measures with classification based on featured distributions, Pattern Recognition 29 (1) (1996) 51–59. doi:10.1016/0031-3203(95)00067-4.
[42] D.J. Hand, R.J. Till, A simple generalisation of the area under the ROC curve for multiple class classification problems, Machine Learning 45 (2001) 171–186. doi:10.1023/A:1010920819831.
[43] S.-P. Yong, J.D. Deng, M.K. Purvis, Modelling semantic context for novelty detection in wildlife scenes, in: 2010 IEEE International Conference on Multimedia and Expo (ICME), 2010, pp. 1254–1259. doi:10.1109/ICME.2010.5583899.
[44] K. Hempstalk, E. Frank, I.H. Witten, One-class classification by combining density and class probability estimation, in: ECML PKDD'08: Proceedings of the 2008 European Conference on Machine Learning and Knowledge Discovery in Databases—Part I, Springer-Verlag, Berlin, Heidelberg, 2008, pp. 505–519.
[45] C.-C. Chang, C.-J. Lin, LIBSVM: A Library for Support Vector Machines, Software, Available at 〈http://www.csie.ntu.edu.tw/~cjlin/libsvm〉, 2001.
[46] S.-P. Yong, J. Deng, M. Purvis, Wildlife video key-frame extraction based on novelty detection in semantic context, Multimedia Tools and Applications 1–18, doi:10.1007/s11042-011-0902-2 (Published online: 4 November 2011).
[47] R. Datta, D. Joshi, J. Li, J. Wang, Studying aesthetics in photographic images using a computational approach, in: Computer Vision—ECCV 2006, Lecture Notes in Computer Science, vol. 3953, Springer, Berlin, Heidelberg, 2006, pp. 288–301. doi:10.1145/1291233.1291364.

Suet-Peng Yong obtained her B.IT (Hons.) in 1998, and M.IT in 2001, both from the National University of Malaysia. She has been a Lecturer in Universiti Teknologi Petronas, Malaysia before joining the Department of Information Science, University of Otago, as a Ph.D. candidate in 2007. Her research interests include computer vision, machine learning and multimedia computing.

Jeremiah D. Deng obtained the B.Eng degree from the University of Electronic Science and Technology of China in 1989, and the M.Eng and Ph.D. from the South China University of Technology (SCUT) in 1992 and 1995 respectively. From 1993 to 1995 he was a research assistant with the Department of Computer Science, University of Hong Kong. He joined SCUT as a lecturer in 1995, and then University of Otago, New Zealand in 1999 as a Research Fellow. He is now a Senior Lecturer in the Department of Information Science, University of Otago. Dr. Deng's research interests include pattern recognition, machine learning and neural networks.

Martin K. Purvis obtained his B.Sc in Physics from Yale University in 1967, and the M.Sc and Ph.D. from University of Massachusetts in 1971 and 1974 respectively. He is now a Professor of Information Science, University of Otago. His research interests include multiagent systems, software engineering, multimedia computing and wireless communications.