Electrical & Electronic Engineering | Research Article

Modified One-Class Support Vector Machine for Content-Based Image Retrieval with Relevance Feedback

Oluwole A. Adegbola, David O. Aborisade, Segun I. Popoola, Olatide A. Amole and Aderemi A. Atayero

Abstract: Image retrieval via traditional Content-Based Image Retrieval (CBIR) often incurs the semantic gap problem—non-correlation of image retrieval results with human semantic interpretation of images. In this paper, Relevance Feedback (RF) mechanism was incorporated into a traditional Query by Visual Example CBIR (QVER) system. The inherent curse of dimensionality associated with RF mechanism was catered for by performing feature selection using Principal Component Analysis (PCA). The amount of feature dimension retained was determined based on a not more than 5% loss constrain imposed on average precision of retrieval result. While the asymmetry and small sample size nature of the resultant image dataset informed the use of a modified One-Class Support Vector Machine (OC-SVM) classifier, three image databases (DB10, DB20 and DB100) were used to test the OC-SVM RF mechanism. Across DB10, DB20 and DB100, Average Indexing Time of 0.451, 0.3017, and 0.0904s were recorded, respectively. For a critical recall value of 0.3, precision values for QVER were 0.7881, 0.7200 and 0.9112, while OC-SVM RF yielded precision of 0.8908, 0.8409, and 0.9503, respectively. Also, the use of PCA yielded tolerable degradation of 3.54, 4.39 and 7.40% in precision on DB10, DB20, and DB100, respectively, with 80% reduction in feature dimension. The OC-SVM RF increased the precision and invariably the reliability of the CBIR system by ranking most of the relevant images higher. Also, the target class was identified faster than the conventional method, thereby reducing the image retrieval time of the OC-SVM RF.

ABOUT THE AUTHOR

Oluwole A. Adegbola is a lecturer in the Department of Electronic and Electrical Engineering, Ladoke Akintola University of Technology, Ogbomoso, Oyo State, Nigeria. He holds a PhD degree in Electronic, Computer and Control Engineering (2018). He obtained his MSc degree in Control Engineering from the University of Lagos in 2003. He holds a Bachelor's degree in Electronic and Electrical Engineering from Ladoke Akintola University of Technology, Ogbomoso, Nigeria (1997). His research interests are in application of machine learning techniques to multimedia information management, and deployment of Content-Based Image Retrieval (CBIR) in platforms with limited resources.

PUBLIC INTEREST STATEMENT

The need for efficient digital image retrieval has become highly imperative due to the exponential growth and diversity in the sources and deployment of digital images and image databases. Retrieval of relevant images from such image databases remains a problem. Although Content-Based Image Retrieval (CBIR) had been proposed to address the problem, image retrieval via traditional CBIR often incur the semantic gap problem—non-correlation of image retrieval results with human semantic interpretation of images. In this work, Relevance Feedback (RF) mechanism was incorporated into a traditional Query by Visual Example CBIR (QVER) system so as to improve the precision, reliability, and the image retrieval time of the system.
1. Introduction

Digital images now come from diverse sources (such as digital cameras, handheld mobile devices, scanners, and medical imaging devices) and for different applications (ranging from personal photograph, entertainment, crime prevention and control, satellite imaging, medicine, to research and training). Also, the volume of information stored in image format is increasing exponentially (Chung, 2007). Owing to the volume and the diversity of image databases, there is a need for effective and efficient means of retrieving specific images from the available dataset. The adoption of human experts to manually annotate images in large databases of generic image types can be very costly and laborious, highly subjective, and the interpretation of a particular image may be inconsistent under different scenarios. The efficiency of available text-based internet search engines is limited because sometimes the surrounding texts may not be directly related to the image or the textual description is grossly inadequate. Therefore, the visual content of digital images must be analyzed to ensure effective and efficient image retrieval.

CBIR was introduced to overcome the shortcomings in text-based retrieval systems (Kato, 1992). CBIR is generally concerned with the process of retrieving desired images from a large collection based on automatic extraction of content characteristics (features), such as color, texture, and shape, which are good representations of the images (Zhang & Chen, 2007). Several feature-extraction techniques have been proposed in the literature for indexing, annotation and retrieval purposes (Long, Zhang, & Feng, 2003). However, the application of low-level feature descriptors to a large image database has some limitations and thus they cannot completely express the user semantic concept, which is the way humans interpret images at an abstract level (Ye Lu, Hu, Zhu, Zhang, & Yang, 2000). This lack of coincidence between the information that can be extracted from the visual data and the interpretation that the same data have for a user in a given situation is commonly known as semantic gap in CBIR systems (Chen, Wang, & Krovetz, 2003; Smeluders, Worrin, Santini, Gupta, & Jain, 2000). Bridging the semantic gap remains a major challenge in CBIR systems.

Previous research efforts that achieved significantly higher semantic performance mostly focused on Nearest Neighbor (NN) search, in which retrieval performance is hinged majorly on the “quality” of features with which the image was indexed in the database (Eakins & Graham, 1999). An example of systems that employ this technique is the Query by Image Content (QBIC) system (Niblack et al., 1993). One major setback of this approach is that the feature vector space employed in similarity computation does not support semantic retrieval. Another concept is the dual modality query scheme, where a combination of text and image example is submitted as query. In (Yavlinsky, 2007), the author proposed to solve the semantic gap problem based on Automatic Image Annotation (AIA). A learning model that depends on the correlation between the low-level features and the training annotation words was developed and then machine learning techniques were employed to automatically apply annotations to new images using the learned model. However, the model designers are often limited in the choice of vocabulary that fully conveys the semantic concept in the user’s mind. Other limitations to this approach include language barrier, polysemy and synonymy.

More recently, Relevance Feedback (RF), which was originally developed for text document retrieval in (Salton, 1989), was proposed for CBIR systems. This idea introduced human interaction with the system such that the system learns the user’s interest, which is the concept in the user’s mind. This paper presents a CBIR system that incorporate machine learning RF mechanism with capability to analyze user’s search interest and subsequently refine the retrieval result through
some feedback mechanism. It aims also to address key issues associated with machine learning RF mechanism namely: *curse of dimensionality*, small sample size and asymmetric nature of image dataset. It is expected that the inclusion of the user in the retrieval loop will significantly improve the performance of CBIR system.

2. Review of related works

Su, Li, and Zhang (2001) introduced Principal Component Analysis (PCA) into the RF framework to extract feature subspaces so that the subjective class implied in the positive examples can be captured. Similar to previous work (Cox, Miller, Minka, Papathomas, & Yianilos, 2000; Vasconcelos & Lippman, 2000), Bayesian learning was used to incorporate user’s feedback for updating the probability distribution of all the images and subsequently re-rank the images in the database. Although the method increased the retrieval speed and significantly reduced the amount of memory deployed, the retrieval accuracy in top 20 results in four feedback iterations was low (0.45 on a scale of 1). This may be attributed to the inability of Bayesian classifiers to estimate the class probability distribution with the few available image examples gathered over the feedback iterations. In addition, more feedback iterations are required to gather more samples that accurately approximate the probability distribution of the image examples; whereas, this is not always available in real-time retrieval systems (Yin, Bhanu, Chang, & Dong, 2005).

In order to resolve the computational complexity issue, a feature selection technique named Filtered and Supported Sequential Forward Search was proposed in the context of SVM (Liu & Zheng, 2006). The technique integrated the filter and wrapper parts into one scheme by utilizing their individual strengths. Experimental results on both synthetic and real data showed the effectiveness of their method in terms of classification accuracy. However, in spite of the fact that the volume (number of instances) and dimension of data used to evaluate this system were much smaller compared to what was obtained in CBIR system, the average run time recorded was 16.23 s. Such a lengthy run time is not acceptable for CBIR system with RF framework.

Chen et al. (2003) presented an unsupervised learning approach to tackle the semantic gap problem in CBIR. A novel image retrieval scheme, CLUster-based rEtrieval of images (CLUE), which retrieves images by unsupervised learning, was used in the work. Neighboring target images selection was done by NN search while weighted graph representation and spectral graph partitioning were done using the normalized cut (Ncut) algorithm for image clustering in a two-level display strategy. At the first level, the system showed a collection of representative images of all the clusters (one for each cluster). At the second level, the system displayed all target images within the cluster specified by a user. In addition, a linear organization of clusters (traversal ordering) was adopted in the work. The system was tested on a 60,000 images database from COREL and an overall average precision for 10 categories was 0.538. In terms of speed, CLUE took, on average, 0.8 s per query for similarity measure evaluation, sorting, and clustering.

Yijuan Lu, Cohen, Zhou, and Tian (2007) performed dimensionality reduction of a feature set by choosing a subset of the original features that contains most of the essential information. The method called Principal Feature Analysis (PFA), essentially a modification of PCA, was applied for choosing the principal features in face tracking and CBIR problems. The authors used Precision as the performance metric, and noted that the approach yielded a better performance as opposed to using the whole feature dimension for retrieval. However, no mention was made of the semantic gap problem in CBIR system; the dimensionality reduction technique was only applied to issue of selecting feature subspace that enhances retrieval performance based solely on similarity measure.

An interactive object-based image clustering and retrieval system, which incorporates two major modules: Preprocessing and Object-based Image Retrieval, was developed in (Zhang & Chen, 2007). The preprocessing module employed an unsupervised segmentation method (WavSeg) to segment images into meaningful semantic regions (image objects). Genetic Algorithm was used to cluster the image objects, hence, minimized the search space for...
object-based image retrieval. The learning and retrieval module adopted the Diverse Density algorithm to analyze the user’s interest. A RF framework based on OC-SVM, which provides progressive guidance to the learning process, was integrated. The performance of the system was tested by dividing the entire set of image regions in Corel image database consisting of 10,000 images from 100 categories into 40 to 150 clusters. The number of clusters was increased by 10 and it was found that the result is most reasonable in terms of the balance between accuracy and reduction of search space when the number of clusters \( k = 100 \). The result showed that the search space is reduced to 13.4\% of the original search space (10,000 images) on average. It was concluded from the work that the system has proved to be effective in better identifying the user’s real need and removing the noise data.

Marakakis, Galatsanos, Likas, and Stafylopatis (2009) proposed a feature selection technique to mitigate the curse of dimensionality problem in a CBIR System indexed with multidimensional feature vector. The authors proposed a RF framework for CBIR, which uses SVM in conjunction with methodology introduced for feature selection in (Guyon, Weston, Barnhill, & Vapnik, 2002). One advantage of their approach is that the learning model used for feature selection is the same as that adopted for the task of RF. This results in the benefit of selecting those features, which are important for the subsequent training of the SVM classifiers used for RF. Although the authors stated that the Recursive Features Elimination using SVMs (SVM-RFE) was based on linear-kernel SVMs, it is known that the real strength of SVMs to handle nonlinear separable data is in the use of kernel tricks.

In an attempt to improve retrieval accuracy with relevance feedback, Bose, Pal, Mallick, and Kumar proposed a hybrid approach to CBIR by combining a CBIR system based on color features. The approach used is based on segmentation similarity using k-means algorithm in which \( k \) is taken to be equal to 8 while an instance-based cluster density with Euclidean distance was employed for feature reweighting. The proposed approach was tested with two databases namely; DB2000 and DB2020, and its performance was finally evaluated based on retrieval efficiency and false discovery as the performance metrics. The result of the proposed method indicated an improved performance with retrieval efficiency and false discovery of 94.69\% and 41.89\%, respectively.

In this paper, a framework is proposed to reduce the feature dimension based on PCA. This will mitigate the curse of dimensionality associated with machine learning RF in CBIR system. In order to determine the amount of feature dimension to be retained, a more objective approach is adopted. This method employs the area under the precision-recall curve, as opposed to using a percentage value of the sum of eigenvalues. Also, a modified OC-SVM is developed with nonlinear kernel to address the asymmetric and small sample size challenges.

The remaining part of this paper is organized as follows: Section 3 presents the methodology adopted for this study while Section 4 discusses the results and major findings. Finally, Section 5 concludes the paper.

3. Materials and method

The entire CBIR system, being a complex system comprising of many modules, was viewed as two broad processes namely: image indexing process and image retrieval process. The image database indexing, which is an offline process, includes copying images to specific database folder, image preprocessing, feature extraction, and feature representation, while the image retrieval includes image query submission, feature extraction of query image, similarity measurement, ranking, and display of top \( n \) retrieved images. Meanwhile, the image retrieval process was performed online. Typically, \( n \) was chosen to be 20 for quick browsing and relevance judgment. The two processes formed the Query by Visual Example Retrieval (QVER) and it was implemented using the block diagram shown in Figure 1.

Furthermore, the machine learning-based RF framework, shown in Figure 2, was implemented. First, the QVER was performed and the retrieved results were labeled as either “relevant” or “irrelevant”. Afterward, the images labeled as “relevant” were used as training samples. The
system learned based on the information that was fed back into it and it consequently updated the results. This process was repeated until the retrieval result is considered satisfactory.

3.1. Image feature database indexing

Standard online image databases (DB10 and DB100) of various image categories and themes were employed for experimentations. A generic domain image database (DB20) consisting of African images was created in addition to the standard databases acquired online. Detailed information about the image databases are presented in Table 1. The image databases were transformed into feature databases. The transformation process includes: image preprocessing, feature extraction, feature representation, and post processing. The dimensions of the feature vectors were reduced using PCA.

The images in the databases were preprocessed to address the challenges of variations in image format and size. Image resizing, image partitioning, and color space transformation were performed. The images were resized to enable the partitioning of each image into $3 \times 3$ equal tiles to facilitate feature extraction and comparison at a finer resolution. In addition, the images were transformed to Hue-Saturation-Value (HSV) color space to aid the extraction of non-device-dependent color features.

The image databases were transformed into feature databases by extracting appropriate low-level features. Color and texture features were chosen as the low-level features. Two color descriptors (color moments and HSV color histogram) and two texture descriptors (Gabor filters and wavelet moments) were used to encode the low-level information based on the results obtained in previous work (Adegbola, Aborisade, Popoola, & Atayero, 2018). Also, the proposed CBIR system is expected to perform retrieval at image level and not at object level.

The color feature descriptors represented the images at global and partitioned levels. The partitioning was done with a view of discovering whether retrieval performance can be enhanced by taking into account the spatial information of color properties. At the global level, for each color channel, three statistical moments (the mean, the standard deviation, and the skewness) were
computed using equations (1) to (3), respectively. This produced a nine-dimensional feature vector which was denoted by CM9.

To extract color moments at the partitioned level, the images were partitioned into $3 \times 3$ equal tiles. For each partition, the first two statistical moments were computed resulting into a $6 \times 3 \times 3$ dimensional feature vector named CM54. For the HSV color histogram, a color quantization was first performed to reduce the number of colors. Although this reduced the quality of information retained in the image, it allowed the histograms to be computed and used in practical situation as obtained in CBIR. Afterward, the color histogram, which represents the estimation of the statistical distribution of colors in the image, was computed using the model of equation (4). This generated a feature vector of dimension $\mathbb{R}^d$, where $d$ is dependent on the number of bins used for quantizing each of the color channels.
Table 1. Selected reference image databases

| Name of Database           | Number of classes | Number of images | Properties                                                                 |
|----------------------------|-------------------|------------------|---------------------------------------------------------------------------|
| DB10 (Corel)               | 10                | 100              | Strong within class low-level visual diversity; clearly defined high-level semantics of each class. Most appropriate for RF evaluation. |
| DB100 (Columbia color database) | 100               | 72               | Sufficient visual variability within each class; unique class identity not subject to interpretation. Contains sufficient number of image for RF. |
| DB20 (various sources)     | 20                | 100              | Reasonable level of overlapping high-level semantics and high variability in terms of source and image size. |

\[
\mu_i = \frac{1}{N} \sum_{j=1}^{N} f_j
\]  
(1)

\[
\sigma_i = \left( \frac{1}{N} \sum_{j=1}^{N} (f_j - \mu_i)^2 \right)^{\frac{1}{2}}
\]  
(2)

\[
S_i = \left( \frac{1}{N} \sum_{j=1}^{N} (f_j - \mu_i)^3 \right)^{\frac{1}{2}}
\]  
(3)

\[
h(c) = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \delta(I(i,j) - c)
\]  
(4)

\(\mu_i, \sigma_i, S_i, N, \) and \(f_j\) represent the mean, standard deviation, skewness, total number of pixels, and the pixel value, respectively.

For an image \(I(x, y)\) with \((p \times q)\) pixels, the Gabor wavelet is defined by equation (5). To extract texture information, a set of Gabor filter banks was applied to cover the entire frequency spectrum of the image. This created a multichannel representation of the image, where each original pixel was described by a vector of the filter's response values at the pixel's location. Useful frequency information was obtained from images by applying a set of Gabor filters ranging across four scales and six orientations. Each filter output, that is, image channel, was encoded by its mean and standard deviation as given by equations (6)-(7) to yield a 48-dimensional real-valued feature vector that represents the image based on equation (8).

\[
W_{mn}(x, y) = \sum_{x_1} \sum_{y_1} I(x_1, y_1) \psi_{mn}^\text{r}(x - x_1, y - y_1)
\]  
(5)

\[
\mu_{mn} = \frac{1}{pq} \sum_x \sum_y |W_{mn}|
\]  
(6)

\[
\sigma_{mn} = \left[ \frac{1}{pq} \left( \sum_x \sum_y (|W_{mn}| - \mu_{mn})^2 \right) \right]^{\frac{1}{2}}
\]  
(7)

\[
f = [\mu_{00}, \sigma_{00}, \ldots, \mu_{mn}, \sigma_{mn}]
\]  
(8)
where * indicates the complex conjugate. The mean $\mu_{mn}$ and the standard deviation $\sigma_{mn}$ of the magnitude of $W_{mn}(x, y)$, represent the texture feature of homogeneous texture region.

### 3.2. Similarity measurement

Similarity or dissimilarity (distance) measurement is a fundamental database operation essential to solve many pattern recognition problems such as clustering, classification, and information retrieval problems. Similarity measures how close to each other two instances are, the larger the similarity value the “closer” the instances are to each other. In CBIR systems, instead of exact matching, distance measure actually computes the level of “closeness” of a query image to the images in the database. Thus, the retrieval result is a list of images ranked in order of their similarities with the query image. In this paper, Minkowski-form distance measure was employed. Equation (9) gives the general $L_p$ norm, while equations (10)-(11) represent the $L_1$ and $L_2$ norm explored in this work.

$$d_{LP}(x, y) = \left[ \sum_{i=1}^{n} |x_i - y_i|^p \right]^{1/p}$$  \hspace{1cm} (9)

$$d_{L1}(x, y) = \sum_{i=1}^{n} |x_i - y_i|$$ \hspace{1cm} (10)

$$d_{L2}(x, y) = \left[ \sum_{i=1}^{n} |x_i - y_i|^2 \right]^{1/2}$$ \hspace{1cm} (11)

### 3.3. Feature selection model

In a generic system, it is extremely difficult to know the particular feature model(s) to be used to uniquely identify certain groups of images. Therefore, a combination of several image feature models is usually employed with the assumption that at least one will have the ability to capture the unique identity of the targeted images. However, this approach poses several challenges. First, such arrangement may increase the chances of “diluting” the feature component that uniquely identifies the targeted image group because the image features are cascaded as a flat vector. This may also lead to what is known as “curse of dimensionality” in approaches that employ machine learning techniques to CBIR system. Another issue is the cost of feature extraction algorithm, which may become prohibitive as the number of feature models increases. In view of this, including too many features is obviously not feasible for application involving human interaction because such system is expected to be fast enough for smooth interaction. Therefore, the selection of most appropriate features (i.e. dimensionality reduction) becomes imperative and to achieve this, a technique that employs PCA was used.

For a binary classification problem, given a set of label training data $\{(X_i, y_i) | i = 1, N \ y_i \neq 0\}$ where sample $X_i \in \mathbb{R}^d$ and $y_i \in \{-1, 1\}$ . Let

$$F = \{f_1, f_2, \ldots, f_d\}$$ \hspace{1cm} (12)

be the set of all features under examination, and let

$$S = \{(X_i, y_i) | i = 1, 2, \ldots, N\} = \left\{ [x_{i1}^1, x_{i2}^1, \ldots, x_{iD}^1]^T, y_i | i = 1, 2, \ldots, N \right\}$$ \hspace{1cm} (13)

denote the training set containing $N$ training pairs, where $x_{id}^i$ is the numerical value of feature $f_d$ for the $i$th training sample. The goal of dimensionality reduction is to find a minimal set of features $F_s = \{f_{s1}, f_{s2}, \ldots, f_{sk}\}$ to represent the input vector $X$ in a lower dimensional space as:

$$X_s = [x_{s1}, x_{s2}, \ldots, x_{sk}]$$ \hspace{1cm} (14)

where $k < d$ , while the classification obtained in the low-dimensional space still yielded the desired accuracy.

PCA is a statistical tool for data analysis, which de-correlates second-order moments corresponding to low-frequency property, and identifies directions of principal variation of data (Diamantaras & Kung,
Considering an ensemble of \( n \)-dimensional vectors \( \{ x = [x_1, \ldots, x_n]^T \} \) whose distribution is centered at the origin, \( E(x) = 0 \), the covariance between each pair of variables is given by equation (15):

\[
\sigma_{ij} = E\{ (x_i - E(x_i))(x_j - E(x_j)) \} = E\{ x_ix_j \}
\]

where \( E \) is the expectation operator. The parameters, \( \sigma_{ij} \), can be arranged to form the \( n \times n \) covariance matrix.

\[
\Sigma = E\{ (x - E(x))(x - E(x))^T \} = E\{ xx^T \}
\]

Assuming \( \text{det}(\Sigma) \neq 0 \) and applying eigenvector decomposition, \( \Sigma \) can be decomposed into the product of three matrices:

\[
\Sigma = W\Lambda W^{-1}
\]

where \( \Lambda = \text{diag}\{\lambda_1, \ldots, \lambda_n\} \) are the eigenvalues and \( W = [w_1, \ldots, w_n]^T \) are the corresponding eigenvectors. \( W \) is orthonormal because \( W^TW = I \). The columns of \( W \) form a set of orthonormal basis vectors, which spans a linear space.

For dimensionality reduction, only the set of orthonormal bases vectors corresponding to the \( k \)-largest eigenvalues were retained. This produced a significant reduction in the feature dimension. In conventional approach, the \( k \)-largest eigenvalues that constitute 95% of the sum of all eigenvalues are retained for dimensionality reduction. But in this work, the amount of feature dimension to be retained was determined using the precision-recall graph. This was more objective as the resulting lower dimensional feature vectors were used for distance similarity measures in image retrieval and RF. Consequently, the number of feature dimension retained was based on a not more than 5% loss constraint imposed on the precision-recall graph.

**3.4. Modified OC-SVM classifier model**

OC-SVMs were first introduced in Schölkopf, Williamson, Smola, Shawe-Taylor, & Platt, 2000. They are essentially binary functions that capture regions in the input space where the probability density lies (i.e. supports). OC-SVMs have been applied to perform Automatic Image Annotation (AIA) in CBIR system (Setia, Teynor, Halawani, & Burkhardt, 2008). In this study, the OC-SVM was slightly modified for learning and classification. The OC-SVM was trained using the positive examples that were obtained from the images retrieved based on the RF framework. In order to keep the complexity of the overall system in check, it was assumed that image classes are not interdependent (i.e. image classes do not overlap). This relaxed the need for class conditional probability, given the presence of other classes. The smallest possible hypersphere was obtained in \( \mathbb{R}^n \) which contains most of the training data. This can be written in primal form as given in equation (18):

\[
\min_{R \in \mathbb{R}, \xi \in \mathbb{R}} R^2 + \frac{1}{\nu} \sum \xi
\]

subject to \( \|\phi(x_i) - c\|^2 \leq R^2 + \xi_i \quad \xi_i \geq 0, 1, \ldots, l \)

\( \nu \in [0, 1] \) is a tunable parameter that sets the tradeoff between \( R \) (i.e. the radius of the hypersphere) and the number of outliers. Small \( R \) allowed the hyperplane to grow so that more training examples can be accommodated while large \( \nu \) kept the hyperplane compact allowing that a fraction of the training examples lies outside; \( \phi \) is the mapping function, \( c \) and \( R \) are the center and the radius of the hypersphere, respectively. Vapnik (2013) showed that it is possible to work in the transformed space without computing the map \( \phi(x_i) \) explicitly if the kernel trick is used. A kernel function is defined by equation (19):

\[
K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle
\]
Furthermore, the dot products between the vectors were computed as parameters for the algorithm. Using Lagrange multipliers, equation (18) becomes equation (20) with $\alpha_i \geq 0$.

$$L(R, \xi, c, \alpha) = R^2 + \frac{1}{\eta} \sum_{i=1}^{n} \xi_i + \sum_{i=1}^{n} \alpha_i \left( \|\phi(x_i) - c\|^2 - R^2 - \xi_i \right)$$  \hspace{1cm} (20)

Equation (20) was minimized based on the Karush–Kuhn–Tucker (KKT) condition to yield the dual form as given in equation (21):

$$\min_{\alpha} \sum_{i,j} \alpha_i \alpha_j K(x_i, x_j) - \sum_i \alpha_i K(x_i, x_j)$$ \hspace{1cm} (21)

subject to $0 \leq \alpha_i \leq \frac{1}{\eta}, \sum_i \alpha_i = 1$

The optimal $\alpha$’s were computed with the help of optimization algorithm to yield the decision function of the form:

$$f(x) = \text{sign} \left[ R^2 - \sum_{i,j} \alpha_i \alpha_j K(x_i, x_j) + 2 \sum_i \alpha_i K(x_i, x) - K(x,x) \right]$$ \hspace{1cm} (22)

In the original formulation, the decision function of equation (22) returns positive for points enclosed by the hypersphere and negative for points outside. However, for CBIR system, a ranked result of images to be retrieved is desirable, so that the result can be sorted on the basis of their positiveness. To this end, OC-SVM decision function of equation (22) was modified to yield the form of equation (23), which is essentially, the normalized distance from the boundary in the transformed space.

$$q(x) = 1 - \frac{1}{R^2} \left[ \sum_{i,j} \alpha_i \alpha_j K(x_i, x_j) + 2 \sum_i \alpha_i K(x_i, x) - K(x,x) \right]$$ \hspace{1cm} (23)

3.5. Simulation setup

The simulations in this study were carried out using MATLAB® Integrated Development Environment (IDE) (Release, 2013). A Graphical User Interface (GUI) was developed as shown in Figure 3. In the case of searching with RF, a new window pops up for the user to select a number of relevant images from the list of retrieved images as shown in Figure 4. Afterward, another RF search, in which the system uses the selected relevant images to form a new model for the modified OC-SVM, was performed. The process was repeated at every click of the Perform RF until the maximum number of clicks is reached.

3.6. System performance evaluation

To demonstrate the effectiveness of the developed CBIR system, an easy-to-use GUI was developed in MATLAB environment for experimental and performance evaluation purposes. In order to test the performance of the system, experiments were performed using some standard reference image databases. This decision was guided by the QVER method and a more involved method namely, RF. Since system performance evaluation relies only on a priori categorization, one obvious advantage is that the evaluation was objective. The performance was evaluated using the Average Indexing Time (AIT), Recall and Precision metrics. The AIT was used to evaluate the ability of the developed system to meet real-time requirement. The AIT for a particular feature model where $n$ is the number of images in the database and $t$ is time in seconds is given by equation (24):

$$\text{Average Indexing Time} = \frac{\sum_{i=1}^{n} t_i}{n}$$  \hspace{1cm} (24)

Recall (R) and Precision (P) are the most common performance metrics for comparing the performances of different feature descriptors and evaluation of CBIR systems (Feng & Chua, 2003; Ferecatu, 2005). The computation of recall and precision was tied to the a priori categorization information of the image database (i.e. the query image is always considered as belonging to a class of images judged visually similar based on some selection criteria that defines the class).
Given an image database $D$, a set of images in a class belonging to the given category and $q \in G$ a query image. The set of images in $G$ is termed relevant to the given query. Assume now that user submits the query image $q$ to the system, a set of retrieved images $I \subset D$ shall be displayed to the user, where the size of $I$ denoted by $|I|$ is between 1 and $|D|$ the size of the
database. For a given query, the recall is thus the ratio of the retrieved relevant images to the total number of relevant images as given by equation (25):

$$R = \frac{|I \cap G|}{|G|}$$  \hspace{1cm} (25)

R is a number between 0 and 1: the bigger the recall value, the better the retrieval result. It is, however, not sufficient in itself to measure system performance, as one could push the recall value to 1 by setting $|I| = |D|$. Thus, we further defined precision as given by equation (26):

$$P = \frac{|I \cap G|}{|I|}$$  \hspace{1cm} (26)

P is essentially the ratio of the retrieved relevant images to the total number of images retrieved by the system. It is also a number between 0 and 1 with higher values corresponding to better results.

Precision and recall are not independent measures, in fact for a given set $I$:

$$P = \frac{|G|}{|I|}R$$  \hspace{1cm} (27)

Equation (27) revealed that increasing the size of the retrieved set $I$ will make the precision value to reduce for any query $q$. The P and R values were plotted against each other for different values of $|I|$ to generate the precision-recall graph.

4. Results and discussion

In this section, the mean precision-clicks graphs are presented. At the first click, the CBIR system displayed the first 30 images closest to the query image. At the next click, relevant images were selected and the database was queried again. The images selected after each retrieval session were used to create an OC-SVM model classifier for the next query. This process was repeated until a maximum number of clicks of 10 were reached. The images were selected from each class in the database and the mean precision from all the classes was calculated. Figure 5 shows that the Manhattan distance (L1) measure is better suited for image retrieval compared to the Euclidean distance (L2) measure. Figure 6 shows the mean precision performance of the OC-SVM RF on the
The mean precision increased from 0.5820 to 0.6940 for 54-dimensional Color Moment (CM54); from 0.4340 to 0.6260 for 48-dimensional Gabor Wavelet (GW48); from 0.3520 to 0.5460 for 40-dimensional Wavelet Moment (WM40); and from 0.7100 to 0.8460 for Histogram (Hist).

The mean precision increased from 0.2540 to 0.6540 for CM54; from 0.3800 to 0.4540 for GW48; from 0.3520 to 0.4340 for WM40; and from 0.3320 to 0.5460 for Hist. The mean precision performance of the OC-SVM RF on the DB100 database using the L1 measure is in Figure 8. The mean precision increased from 0.3660 to 0.7660 for CM54; from 0.5680 to 0.6600 for GW48; from 0.3860 to 0.7920 for WM40; and from 0.3340 to 0.7420 for Hist. These results showed that there is an appreciable improvement on the mean precision with the application of the OC-SVM RF.

**Figure 6.** Mean precision performance of the OC-SVM relevance feedback on the DB10 database using L1 similarity measure.

**Figure 7.** Mean precision performance of the OC-SVM relevance feedback on the DB20 database using L1 Similarity measure.

**Figure 8.** Mean precision performance of the OC-SVM RF on the DB100 database using the L1 measure.
The performance of the developed OC-SVM RF was also investigated with combination of visual descriptors. The mean precision performance of the OC-SVM RF with combinations of selected descriptors on the databases using the L1 measure is shown in Figure 9. The mean precision increased from 0.6467 to 0.7573 for GW48+ WM40; from 0.6973 to 0.8040 for Hist+ WM40; and from 0.6513 to 0.7313 for Hist+ GW48. Figure 10 shows the mean precision performance of the OC-SVM RF with combinations of all descriptors on the databases using the L1 measure. The mean precision increased from 0.7467 to 0.8067 for GW48+ WM40+ Hist and from 0.7380 to 0.8287 for CM+ GW48+ WM40+ Hist. These results revealed that the use of the OC-SVM RF improved the mean precision performance of two combined descriptors. The comparison between the L1 and L2 measure with the OC-SVM RF using the feature model CM54+ GW48+ WM40+ Hist on each of
the databases is illustrated in Figure 11. The mean precision on DB10 increased from 0.8140 to 0.9400 for L1 and from 0.5800 to 0.6880 for L2. On DB20, the mean precision increased from 0.6340 to 0.7600 for L1 and from 0.3880 to 0.5880 for L2. The mean precision on DB100 increased from 0.7660 to 0.7860 for L1 and from 0.1460 to 0.4480 for L2. These results show that the mean precision is relatively higher with the L1 measure than with the L2 measure.

The precision-recall for the OC-SVM RF was compared with that of conventional QVER on the databases with the L1 measure using combination of all the descriptors (CM54+ GW48+ WM40+ Hist) in Figure 12. Taking average of the recalls, on DB10, the average precision for the OC-SVM RF and Without RF were 0.8581 and 0.7247, respectively. On DB20, the average precision for the OC-SVM RF and Without

Figure 11. Comparison between L2 and the L1 measures with the OC-SVM RF using combination of all the descriptors (CM54+ GW48+ WM40+ Hist).
RF were 0.8093 and 0.6643, respectively. The average precision on DB100 for the OC-SVM RF and Without RF were 0.8937 and 0.8102, respectively. The results showed that the OC-SVM RF gives precision performance improvement of about 15.55% on DB10, 17.91% on DB20, and 9.34% on DB100. The mean precision performance of OC-SVM RF for combined database effect is shown in Figure 13. Taking average of the recalls, the average mean precision for the OC-SVM RF was 0.8537, while the average mean precision for the Without RF case was 0.7331. This showed that, on the average, the OC-SVM RF gives about 14.13% retrieval performance improvement over the conventional QVER (or Without RF).

Combination of two or more visual descriptors increased the dimension of the final feature vector. For example, the final feature model, which is the concatenation of individual feature

Figure 12. Comparison between the performance of the conventional QVER and the OC-SVM RF-based QVER on the databases with L1 similarity measure.

Figure 13. Mean precision performance comparison between conventional QVER and the OC-SVM RF-based QVER with the L1 measure.
vectors, could have very high dimensions and thus increase the latency of RF even on medium-size image databases. Hence, in order to minimize the complexity of the RF-based CBIR system, reducing the dimensions of feature vectors was considered necessary. PCA was introduced to the developed OC-SVM RF for the purpose of reducing the dimension of the feature vector. A criterion of not more than 5% degradation in mean precision performance was used to determine the dimension of feature to keep.

The effect of the reduction in the dimensions of feature vector is shown in Figure 15. When the feature dimension was reduced by 80%, the maximum mean precision obtained on DB10, DB20, and DB100 was 0.9067, 0.7266, and 0.7275, respectively, while a reduction of feature dimension by 83% resulted into mean precision values of 0.6933, 0.5093, and 0.3657 for DB10, DB20, and DB100, respectively. Figure 14 shows the comparison between the OC-SVM RF that used the whole 174 dimensional feature (labeled STD) and the OC-SVM RF with PCA that used 35 dimensional features (labeled PCA). The maximum mean precision achieved with STD on the DB10, DB20, and DB100 was 0.9400, 0.7600, and 0.7860, respectively. The maximum mean precision achieved with PCA on the DB10, DB20, and DB100 was 0.9067, 0.7266, and 0.7275, respectively. Thus, with 80% reduction in feature dimension, tolerable degradation of 3.54%, 4.39%, and 7.4% in precision performance was achieved on DB10, DB20, and DB100, respectively.

5. Conclusion

This paper developed a method for CBIR system based on RF with modified OC-SVM classifier to learn the user’s semantic concept for the purpose of achieving faster and acceptable relevant contents retrieval. Dimensionality reduction of features for medium to large image databases was achieved by employing a linear PCA. The developed OC-SVM RF technique is an advanced interactive search CBIR system, which helps to reduce the semantic gap. The OC-SVM RF increases the precision and, invariably, the reliability of the CBIR system by ranking most of the relevant images higher. Also, by identifying the target class faster than the conventional method, the image retrieval time of the OC-SVM RF was reduced. The use of PCA helped to cope with large databases; however, great care is required in selecting the percentage of dimensionality reduction. The percentage of dimensionality reduction of the features should be chosen such that reasonable percentage of signal energy is kept for proper querying.
In future work, the color-level features may be combined with gray-level features. Also, weighted visual descriptors may improve retrieval performance. In addition, the retrieval performance of other classifiers, apart from SVM, may be investigated. Finally, other similarity measures or distance functions can be checked for performance improvement.

Acknowledgments
The authors wish to appreciate the Center for Research, Innovation, and Discovery (CUCRID) of Covenant University, Ota, Nigeria for the partial funding of this publication.

Funding
This work was supported by the Covenant University, Ota, Nigeria [CUCRID-SMARTCU-000445].

Author details
Oluwole A. Adegbola
E-mail: oaadegbola@lautech.edu.ng
David O. Aborisade
E-mail: daoaborisade@lautech.edu.ng
Segun I. Popoola
E-mail: segun.popoola@stu.cu.edu.ng
ORCID ID: http://orcid.org/0000-0002-3941-5903
Olatide A. Amole
E-mail: olatide.amole@bellsuniversity.edu.ng
Aderemi A. Atayero
E-mail: atayero@covenantuniversity.edu.ng

1 Department of Electronic and Electrical Engineering, Ladoke Akintola University of Technology, P.M.B. 4000, Ogbomoso, Nigeria.
2 Department of Electrical and Information Engineering, Covenant University, P.M.B. 1023, Ota, Nigeria.
3 Department of Electrical, Electronic and Computer Engineering, Bells University of Technology, P.M.B. 1015, Ota, Nigeria.

References
Adegbola, O. A., Aborisade, D. O., Popoola, S. I., & Atayero, A. A. (2018). Performance evaluation of visual descriptors for image indexing in content based image retrieval systems. Paper presented at the International Conference on Computational Science and Its Applications.

Cite this article as: Modified one-class support vector machine for content-based image retrieval with relevance feedback, Oluwole A. Adegbola, David O. Aborisade, Segun I. Popoola, Olatide A. Amole & Aderemi A. Atayero, Cogent Engineering (2018), 5: 1541702.

Citation information
Adegbola, O. A., Aborisade, D. O., Popoola, S. I., & Atayero, A. A. (2018). Performance evaluation of visual descriptors for image indexing in content based image retrieval systems. Paper presented at the International Conference on Computational Science and Its Applications. Proceeding of Seventh International Symposium (vol. 1, pp. 197-200). Paris. France. doi:10.1109/ISSPA.2003.1224674

Chung, K.-P. (2007). Intelligent content-based image retrieval framework based on semi-automated learning and historic profiles. Western Australia: Murdoch University.

Cox, T. J., Miller, M. L., Minka, T. P., Papathomas, T. V., & Yianilos, P. N. (2000). The Bayesian image retrieval system, PicHunter: Theory, implementation, and psychophysical experiments. IEEE Transactions on Image Processing, 9(1), 20-37. doi:10.1109/83.817596
Diamantaras, K. I., & Kung, S. Y. (1996). Principal component neural networks: Theory and applications. New Jersey, NJ: John Wiley & Sons, Inc.

Eskins, J., & Graham, M. (1999). Content-based image retrieval. United Kingdom: University of Leeds.

Feng, H., & Chou, T.-S. (2003). A bootstrapping approach to annotating large image collection. Paper presented at the Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval. Berkeley, California.

Ferencz, M. (2005). Image retrieval with active relevance feedback using both visual and keyword-based descriptors. University of Versailles Saint-Quentin-en-Yvelines, INRIA.

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. Machine Learning, 46(1), 389–422. doi:10.1023/A:1012487302797

Kato, T. (1992). Database architecture for content-based image retrieval. Image Storage and Retrieval Systems, 1662, 112–123.

Liu, Y., & Zheng, Y. F. (2006). FS_SFS: A novel feature selection method for support vector machines. Pattern Recognition, 39(7), 1333–1343. doi:10.1016/j.patcog.2005.10.006

Long, F., Zhang, H., & Feng, D. D. (2003). Fundamentals of content-based image retrieval. In Multimedia information retrieval and management (pp. 1–26). Berlin: Springer.

Lu, Y., Cohen, I., Zhou, X. S., & Tian, Q. (2007). Feature selection using principal feature analysis. Paper presented at the Proceedings of the 15th ACM international conference on Multimedia. Augsburg, Germany.

Lu, Y., Hu, C., Zhu, X., Zhang, H., & Yang, Q. (2008). A unified framework for semantics and feature based relevance feedback in image retrieval systems. Paper presented at the Proceedings of the eighth ACM international conference on Multimedia. California, USA.

Marakakis, A., Galatsanos, N., Likas, A., & Stafylopatis, A. (2009). Combining Gaussian mixture models and support vector machines for relevance feedback in content-based image retrieval. In Iliadis, Maglogiann, Tsoumakakis, Vlahavas, Bramer (Eds.), Artificial intelligence applications and innovations III (pp. 249–258). Boston, MA: Springer.

Niblack, C. W., Barber, R., Equitz, W., Flickner, M. D., Glasman, E. H., Petkovic, D., … Taubin, G. (1993). QBIC project: Querying images by content, using color, texture, and shape. Paper presented at the IS&T/SPIE’s Symposium on Electronic Imaging: Science and Technology. San Jose, CA, United States.

Release, M. (2013). The MathWorks. Inc., Natick, Massachusetts, 488.

Salton, G. (1989). Automatic text processing (pp. 6). Reading, MA: Addison Wesley.

Schölkopf, B., Williamson, R. C., Smola, A. J., Shawe-Taylor, J., & Platt, J. C. (2000). Support vector method for novelty detection. Paper presented at the Advances in neural information processing systems. Denver, CO.

Setia, L., Teynor, A., Halawani, A., & Burkhart, H. (2008). Grayscale medical image annotation using local relational features. Pattern Recognition Letters, 29(15), 2039–2045. doi:10.1016/j.patrec.2008.05.018

Smeulders, A. W., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(12), 1349–1380. doi:10.1109/34.895972

Su, Z., Li, S., & Zhang, H. (2001). Extraction of feature subspaces for content-based retrieval using relevance feedback. Paper presented at the Proceedings of the ninth ACM international conference on Multimedia (pp. 98–106). Ottawa, Canada.

Vapnik, V. (2013). The nature of statistical learning theory. Springer science & business media. Berlin, Heidelberg: Springer-Verlag.

Vasconcelos, N., & Lippman, A. (2000). Learning from user feedback in image retrieval systems. Paper presented at the Advances in neural information processing systems. Denver, CO.

Yovelinsky, A. (2007). Image indexing and retrieval using automated annotation. United kingdom: Department of Computing, Imperial College London.

Yin, P.-Y., Bhanu, B., Chang, K.-C., & Dong, A. (2005). Integrating relevance feedback techniques for image retrieval using reinforcement learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(10), 1536–1551. doi:10.1109/TPAMI.2005.201

Zhang, C., & Chen, X. (2007). OCRS: An interactive object-based image clustering and retrieval system. Multimedia Tools and Applications, 35(1), 71–89. doi:10.1007/s11042-007-0116-9
