

# Tackling Curse of Dimensionality for Efficient Content Based Image Retrieval

Minakshi Banerjee<sup>(✉)</sup> and Seikh Mazharul Islam

RCC Institute of Information Technology, Canal South Road,  
Beliaghata, Kolkata 700015, West Bengal, India  
mbanerjee23@gmail.com, mislam2911@gmail.com

**Abstract.** This paper proposes a content based image retrieval (CBIR) technique for tackling curse of dimensionality arising from high dimensional feature representation of database images and search space reduction by clustering. Kernel principal component analysis (KPCA) is taken on MPEG-7 Color Structure Descriptor (CSD) (64-bins) to get low-dimensional nonlinear-subspace. The reduced feature space is clustered using Partitioning Around Medoids (PAM) algorithm with number of clusters chosen from optimum average silhouette width. The clusters are refined to remove possible outliers to enhance retrieval accuracy. The training samples for a query are marked manually and fed to One-Class Support Vector Machine (OCSVM) to search the refined cluster containing the query image. Images are ranked and retrieved from the positively labeled outcome of the belonging cluster. The effectiveness of the proposed method is supported with comparative results obtained from (i) MPEG-7 CSD features directly (ii) other dimensionality reduction techniques.

**Keywords:** Kernel principal component analysis · Partitioning around medoids · Outliers detection · Support vector clustering · One-class support vector machine

## 1 Introduction

Developing efficient content-based image retrieval (CBIR) [3, 11] techniques have proven to be a challenging research area for facilitating tremendous demand of accessing digital image libraries in real life applications. CBIR is used in place of traditional image retrieval technique which employs manual keyword annotations. It combines different image processing, pattern recognition, computer vision techniques, human-computer interactions etc. to improve retrieval accuracy. Visual contents of an image such as color, shape, texture and spatial layout etc. are extracted automatically and represented in terms of multi-dimensional feature vectors [1, 11]. Similar images are returned based on measuring similarities between the feature vectors of the submitted query image and those stored in the database. Owing to the impact of multimedia retrieval, ISO/IEC has launched MPEG-7, which provides a collection of specific standard descriptors

[9] for evaluation of new image retrieval schemes. While there have been ample amount of publications in the area of CBIR, the performance of image retrieval systems is still not satisfactory due to semantic gap, which is created from the discrepancies between the computed low-level features (color, texture, shape, etc.) and users conception of an image. Iterative refinement of results through relevance feedback [12] has been popular for minimizing semantic gap in CBIR. An efficient image search method using small codes for large image databases is proposed in [13] which converts Gist descriptor to a compact binary code for representing images and their neighborhood structures.

Recently, automatic image annotation (AIA) is used as a distinctive method for bridging semantic gap. Bag-of-visual-words model which can represent images at object level and provides spatial information to build a visual dictionary between low level features and high-levels semantics for AIA is proposed in [8]. Scalable object recognition with vocabulary tree is also proposed in [6]. One important issue in AIA is high dimensional features which should be selected in right number and right features for efficient annotation. Performance of the classifiers may degrade when feature dimension is too high. Also large number of features may better represent the discriminative properties of visual contents. But this leads to dimensionality curse problem [5]. Proper dimensionality reduction methods must be employed to preserve the distinctive properties so that performance is not degraded. The popular dimensionality reduction algorithms include Principal Component Analysis (PCA), Kernel Principal Component Analysis (KPCA), Linear discriminant analysis (LDA), Factor analysis (FA), and Laplacian eigenmaps (LEM) etc. The kernel PCA [10] which is a nonlinear form of PCA, can efficiently compute principal components in high dimensional feature spaces. It captures more information than PCA as it is related to the input space by some nonlinear mapping. Image retrieval using Laplacian eigenmap [5] produces efficient results by preserving local neighborhood information of data points but suffers from the problem of new data point, that is if query image is absent in database. Although efficient feature reduction for a particular query is an important criteria to amend curse of dimensionality, to speed up the retrieval process data clustering is also an effective way. However selection of number of clusters for an unknown database is an important issue. A Fuzzy Clustering approach to CBIR is proposed in [7] but it makes no attempt in removing groups of images that are irrelevant to the search query, nor re-ranking the search results.

The motivation behind the proposed method lies in addressing the task of dimensionality reduction, search space reduction by efficient clustering and proper searching and retrieving from reduced search space by labeling them efficiently as positive (relevant) and negative (non-relevant) images so that the overall performance is improved. To pursue these motivations the proposed method uses kernel principal component analysis for dimensionality reduction, Partitioning Around Medoids (PAM) [4] algorithm for clustering. The clusters are further processed using Support Vector Clustering (SVC) [15] to remove possible outliers from the cluster containing the query. One-class support vector machine [2] is proposed for classification as this classifier is biased to the learned

concept of a particular category. Training samples are generated from displayed results obtained using KPCA-reduced CSD feature and  $L_1$  similarity distance. The proposed method is compared with others dimensionality reduction techniques, namely, Principal Component Analysis (PCA), Factor analysis (FA), and Laplacian eigenmaps (LEM). The remaining sections are organized as follows. Section 2 gives the mathematical background of Kernel PCA and one-class SVM. Proposed method is highlighted in Sect. 3. Experimental discussion and results are reported in the Sect. 4. Section 5 concludes.

## 2 Mathematical Preliminaries

### 2.1 Kernel Principal Component Analysis (KPCA)

Suppose  $x_i \in R^l, i = 1, 2, \dots, n$  are  $n$  observations. The basic idea of KPCA [10], is as follows. First, the samples are mapped into some potentially high-dimensional feature space  $F$  by following.

$$\mathcal{O} : R^l \rightarrow F, x_i \rightarrow \mathcal{O}(x_i), (i = 1, 2, \dots, n) \quad (1)$$

where  $\mathcal{O}$  is a nonlinear function. Now these mapped samples in  $F$  using a kernel function  $k$  occur in terms of dot product are given by following equation.

$$k(x_i, x_j) = (\mathcal{O}(x_i), \mathcal{O}(x_j)) \quad (2)$$

For simple notation assume  $\sum_{i=1}^n \mathcal{O}(x_i) = 0$ . To find eigenvalues  $\lambda > 0$  and associated eigenvectors  $v \in F \setminus \{0\}$  which are computed by:

$$v = \sum_{i=1}^n \alpha_i \mathcal{O}(x_i) \text{ where } \alpha_i (i = 1, \dots, n) \text{ are coefficients.} \quad (3)$$

Then the eigenvalue equation is  $n\lambda\alpha = K\alpha$ , where  $\alpha$  denotes a column vector with entries  $\alpha_1, \dots, \alpha_n$  and  $K$  is called kernel matrix. Kernel principal components of a test point  $t$  are computed by projecting  $\mathcal{O}(t)$  onto the  $k^{th}$  eigenvectors  $v^k$  is:

$$(v^k \cdot \mathcal{O}(t)) = \sum_{i=1}^n \alpha_i^k (\mathcal{O}(x_i), \mathcal{O}(t)) = \sum_{i=1}^n \alpha_i^k k(x_i, t), \text{ where } \lambda_k (\alpha^k \cdot \alpha^k) = 1 \quad (4)$$

Kernel function for our experiment is:

$$k(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / 2\sigma^2), \text{ where } \sigma > 0 \quad (5)$$

### 2.2 One Class Support Vector Machine (OCSVM)

The OCSVM [2] maps input data into a high dimensional feature space using a kernel and iteratively finds the maximal margin hyperplane which best separates

the training data from the origin. The OCSVM may be viewed as a regular two-class SVM where all the training data lies in the first class, and origin is taken as only member of the second class. Thus, the hyperplane (or linear decision boundary) corresponds to the classification rule is:

$$f(x) = \langle w, x \rangle + b \quad (6)$$

where  $w$  is the normal vector and  $b$  is a bias term. The OCSVM solves an optimization problem to find the rule with maximal geometric margin. To assign a label to a test sample  $x$  if  $f(x) < 0$  then the sample is non-relevant otherwise relevant.

### Kernels:

The optimization problem of OCSVM is obtained by solving the dual quadratic programming problem, is given by:

$$\min_{\alpha} \frac{1}{2} \sum_{ij} \alpha_i \alpha_j K(x_i, x_j), \text{ s.t. } 0 \leq \alpha_i \leq \frac{1}{vl} \text{ and } \sum_i \alpha_i = 1 \quad (7)$$

where  $\alpha_i$  is a lagrange multiplier (or weight on example  $i$  such that vectors associated with non-zero weights are called support vectors and solely determine the optimal hyperplane),  $v$  is a parameter that controls the trade-off between maximizing the distance of the hyperplane from the origin and the number of data points contained by the hyper-plane,  $l$  is the number of points in the training dataset, and  $K(x_i, x_j)$  is the kernel function. By using the kernel function to project input vectors into a feature space, we allow for nonlinear decision boundaries. Given a feature map:

$$\mathcal{O} : X \rightarrow R^N \quad (8)$$

where  $\mathcal{O}$  maps training vectors from input space  $X$  to a high-dimensional feature space, we can define the kernel function as:

$$K(x, y) = \langle \mathcal{O}(x), \mathcal{O}(y) \rangle \quad (9)$$

The commonly used kernels are linear, radial basis function, and sigmoid.

## 3 Proposed Method

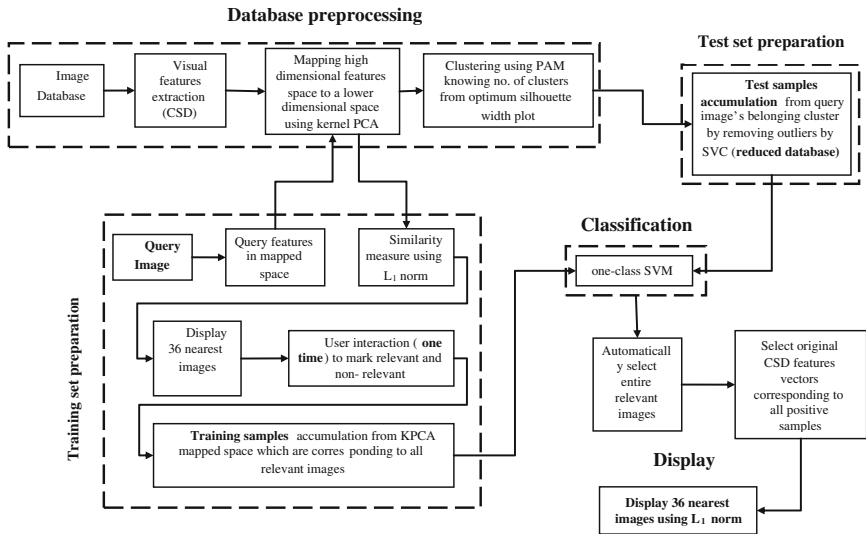
The proposed method is explained in Fig.1 and the algorithmic steps are also shown below.

### 3.1 Feature Extraction

The MPEG-7 Color Structure Descriptor (CSD) (64-bins) is extracted for each image. Number of times a particular color is contained within the structuring element is counted using  $c_0, c_1, c_2, \dots, c_{M-1}$  quantized colors. A color structure histogram can then be denoted by  $h(m)$ ,  $m = 0, 1, 2, \dots, M - 1$  where the value in each bin represents the number of structuring elements in the image containing one or more pixels with color  $c_m$ . The hue-min-max-difference (HMMD) color space is used in this descriptor [9].

**Algorithm 1.** Steps in the proposed method

- Input:** A query Image  
**Output:** Similar Images  
**Step1:** Extract the MPEG-7 CSD (64-bins) features of all the images in database.  
**Step2:** Apply KPCA to map feature vectors in lower-dimensional space.  
**Step3:** Cluster the reduced dataset obtained in **Step2** using PAM algorithm. Select the cluster containing the query by feature matching.  
**Step4: Training samples collection** - Display 36 nearest images with respect to query image using  $L_1$ -norm in KPCA - reduced space and mark relevant images to get training set.  
**Step5: Test samples collection** - Apply SVC to the cluster containing the query obtained in **Step3** to remove possible outliers.  
**Step6:** Train OCSVM using samples obtained in **Step4**.  
**Step7:** Input test samples obtained in **Step5** in OCSVM for prediction (prediction label 1 is for positive sample and 0 for negative sample).  
**Step8:** Consider all positive samples obtained in **Step7** automatically using the labels of prediction.  
**Step9:** Select original CSD (64-bins) feature vectors corresponding to all positive samples and calculate the similarity value by  $L_1$  norm. Sort these similarity values in ascending order.  
**Step10:** Display the ranked images obtained in **Step9**.



**Fig. 1.** Proposed method

### 3.2 Number of Clusters Selection Using PAM

Searching similar images in few nearest clusters are more effective than sequential search in large database. The PAM [4] clustering algorithm is usually less sensitive to outliers than k-means. It groups similar objects by minimizing sum of the dissimilarities of all the objects to their nearest medoids. The aim is to find the clusters  $C_1, C_2, \dots, C_k$  that minimize following target function:

$$\sum_{i=1}^k \sum_{r \in C_i} d(r, m_i) \text{ where for each } i \text{ the medoid } m_i \text{ minimizes } \sum_{r \in C_i} d(r, m_i) \quad (10)$$

where  $d(r, m_i)$  indicates dissimilarity between  $r$  and  $m_i$ . The numbers of clusters are determined from optimum average silhouette width plot. For every point  $i$ , the silhouette width  $silw(i)$  is calculated as follows: Let  $p(i)$  be the average dissimilarity between  $i$  and other all points of the partition to which  $i$  lies. If  $i$  is only object in its belonging partition then  $silw(i) = 0$  barring further computation. For all remaining partitions  $C$  we get  $d(i, C) =$  average dissimilarity of  $i$  to all objects of  $C$ . The smallest of these  $d(i, C)$  is  $q(i) = \min_{\forall C} (d(i, C))$  indicates the dissimilarity between  $i$  and its nearest partition obtained from minimum leads to Eq. 11 and  $\max(\text{avg}(silw_C), \forall C)$  determines the number of clusters. For example, if number of clusters is  $k$  (where  $k = 2, 3, 4, \dots, 25$ ) then silhouette width for every point is computed and average is found. Finally, the cluster number which gives the maximum average silhouette width plot is selected. In Fig. 2 it is obvious that the number of clusters is 3 by taking upto 25 clusters.

$$silw(i) = (q(i) - p(i)) / \max(p(i), q(i)) \quad (11)$$

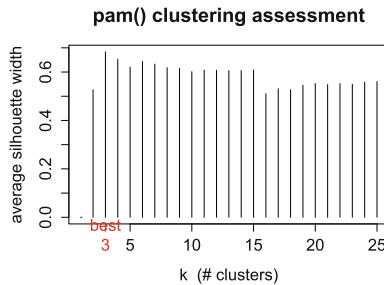


Fig. 2. Computation of number of clusters from optimum average silhouette width

### 3.3 Outliers Detection Criteria of Support Vector Clustering (SVC)

Let  $\{x_i\}$  be a dataset with dimensionality  $d$  and  $x$  be a data point. SVC [15] computes a sphere of radius  $R$  and center  $a$  containing all these data. The computation of such a smallest sphere is obtained from solving minimization problem

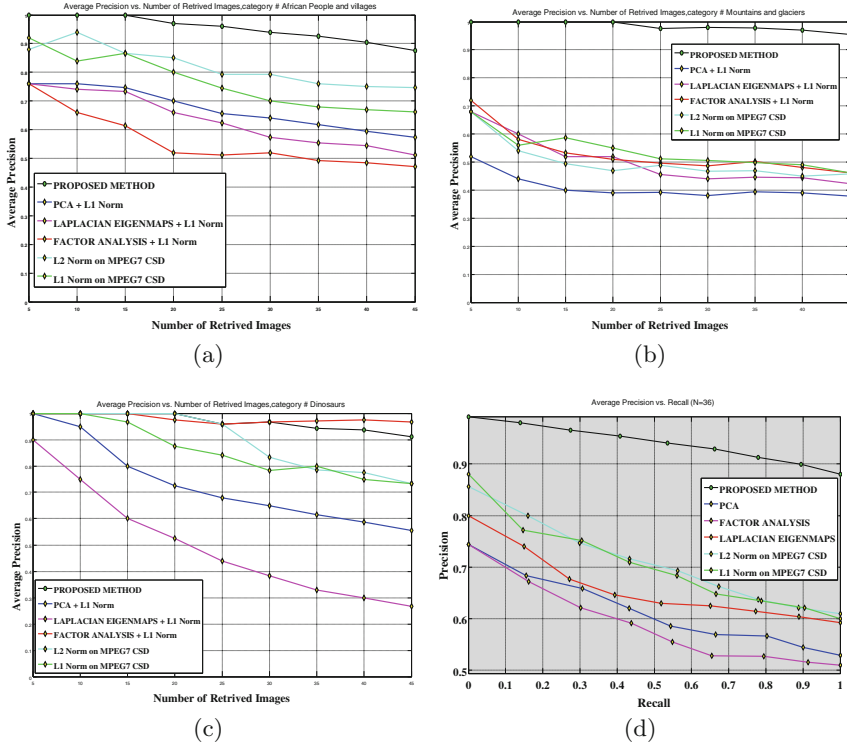


**Fig. 3.** (a) and (c) represent MPEG-7 CSD(64-bins) with  $L_1$  norm, relevant/scope ratios are 16/36 and 20/36 respectively, (b) and (d) represent proposed method, relevant/scope ratios are 35/36 and 35/36 respectively, with KPCA projected dimension=1, OCSVM based prediction from refined cluster containing the query. The top-left most image is the query.

considering Lagrangian formulation which produces following expression.

$$\|x - a\|^2 = (x \cdot x) - 2 \sum_{i=1}^N \alpha_i (x \cdot x_i) + \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j (x_i \cdot x_j) \leq R^2, \text{ where } \alpha_i \text{ is the Lagrangian multipliers. To test a data } x \text{ for outlier, the necessary condition is } \alpha_i \geq 0.$$

**Computational Issues of the Proposed Method.** The proposed method, at first, applies KPCA to reduce original data set in  $O(n^3)$  times, where  $n$  is the number of data points, then the clustering algorithm PAM operates on this reduced dataset at least in quadratic times. The complexity concerned with SVC is  $O(n^2d)$  if the number of support vectors is  $O(1)$  (where  $d$  is the number of dimensions of a data point). OCSVM algorithm uses sequential minimal optimization to solve the quadratic programming problem, and therefore takes time  $O(dL^3)$ , where  $L$  is the number of entities in the training dataset. The total complexity is involved mainly in SVC and OCSVM which are applied on reduced features and search space.



**Fig. 4.** (a), (b) and (c) represent Average precision vs. Number of retrieved images. Category in (a) African people and villages (b) Mountains and glaciers (c) Dinosaurs and (d) represents Average precision vs. Average recall on overall database.

## 4 Experiments and Results

The experimental results have been demonstrated from Figs. 3 to 4 using 1000 SIMPLIcity database [14] (<http://wang.ist.psu.edu>) which is divided into 10 Semantic groups including African people and villages, Buses, Dinosaurs, Elephants etc. Each group comprises of 100 sample images. Experiments are carried out extensively on each group. The class label of each image in the database are not predefined and label of training samples for classification by OCSVM is based on visual judgement of similarity which is obtained by marking relevant and non-relevant images. Therefore the performance of retrieval is evaluated on the finally displayed images based on the visual judgement of similarity instead of classification scores by predefined labels. The performances metrics used are Recall rate and Precision rate. Let  $n_1$  be the number of images retrieved in top  $n$  positions that are close to a query image. Let  $n_2$  be the number of images in the database similar to the query. Evaluation standards, Recall rate ( $R$ ) is given by  $n_1/n_2 \times 100\%$  and Precision rate ( $P$ ) are given by  $n_1/n \times 100\%$ .



KPCA is used to reduce the CSD dataset taking Gaussian function with  $\sigma = 0.2$  as a parameter of kernel function using kernlab R language Package. In the classification stage by OCSVM the training of classifier is done with radial basis function using LIBSVM package with  $c = 0.01$ ,  $\gamma = 0.00001$ . Finally the predicted relevant images of OCSVM have been ranked using  $L_1$ -norm considering original CSD features of predicted set. The result of Fig. 3(a) and (b) indicate a query from category elephant having diverse background. Figure 3(a) shows result with MPEG-7 CSD (64-bins) features using  $L_1$ -norm as a similarity measure. Figure 3(b) represents proposed method with KPCA where the projected dimension is equal to one. The query image is in top-left corner. The comparative results suggest that the precision obtained in case of Fig. 3(a) within the scope of 36 images is 44.44%. For the same query the proposed method gives 97.22% precision where similar images have attained higher rank compared to Fig. 3(a), in the reduced search space. The effectiveness of the proposed method is also observed in Fig. 3(c) for the category of Bus. Here color structure descriptors have proven to be good features for retrieving buses not only for similar shapes but also for similar colors. Precision and ranking are improved in Fig. 3(d) compared to Fig. 3(c). Intuitively, nonlinear KPCA based feature mapping captures discriminating information of the CSD (64-bins) into a lower dimensional subspace. The results are compared with other unsupervised dimensionality reduction methods namely, PCA with  $L_1$ -norm, Laplacian eigenmaps with  $L_1$ -norm, Factor analysis with  $L_1$ -norms, and with MPEG-7 CSD (64-bins) dataset with  $L_2$  and  $L_1$  norms. Average precision is chosen as a comparative measure because it gives a good measure within the displayed scope and it is not always possible to know the number of relevant image in advance for a particular query in an unknown database. The proposed method gives 95% average precision within the scope of 30 images for the category of African people. This is shown in Fig. 4(a) which performs far well than other techniques. In case of the category of mountain as shown in Fig. 4(b) the proposed method gives 98% average precision within the scope of 30 images and performs far well than other methods. For the category of dinosaur as shown in Fig. 4(c), Factor analysis gives marginally better result. Particularly for this category the images are covered by white background and lacking prominent color structure information which may be a possible reason for this performance. Considering images from all categories the average precision vs. recall graph on the overall database is plotted in Fig. 4(d) which is giving more than 95% precision before recall reaches at 0.5. This proves the effectiveness of the proposed method.

## 5 Conclusion

In this paper we have proposed an efficient content-based image retrieval system to address the issues of dimensionality reduction and search space reduction followed by possible outlier removal, without degrading the retrieval performance. With the proposed method the MPEG-7 Color Structure Descriptor (CSD) in reduced dimensions is performing very well for many categories compared to

other retrieval systems which combine several descriptors together. As a further scope of research we intend to make the system more generic considering several newer datasets for different retrieval applications. We would also like to combine others MPEG-7 features and extend the proposed method for object retrieval by unifying textual cues.

## References

1. Agarwal, S., Verma, A., Dixit, N.: Content based image retrieval using color edge detection and discrete wavelet transform. In: 2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), pp. 368–372 (2014)
2. Chen, Y., Zhou, X.S., Huang, T.: One-class SVM for learning in image retrieval. *Int. Conf. Image Process.* **2001**, 34–37 (2001)
3. Jaworska, T.: Application of fuzzy rule-based classifier to CBIR in comparison with other classifiers. In: 11th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), pp. 119–124 (2014)
4. Kaufman, L., Rousseeuw, P.J.: *Partitioning Around Medoids (Program PAM)*. Wiley, New York (2008)
5. Lu, K., He, X., Zeng, J.: Image retrieval using dimensionality reduction. In: Zhang, J., He, J.-H., Fu, Y. (eds.) *CIS 2004*. LNCS, vol. 3314, pp. 775–781. Springer, Heidelberg (2004)
6. Nistér, D., Stewénus, H.: Scalable recognition with a vocabulary tree. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2161–2168 (2006)
7. Ooi, W., Lim, C.: A fuzzy clustering approach to content-based image retrieval. In: *Workshop on Advances in Intelligent Computing*, pp. 11–16 (2009)
8. Philbin, J., Chum, O., Isard, M., Sivic, J., Zisserman, A.: Object retrieval with large vocabularies and fast spatial matching. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2007)* (2007)
9. Salembier, P., Sikora, T.: *Introduction to MPEG-7: Multimedia Content Description Interface*. Wiley, New York (2002)
10. Schölkopf, B., Smola, A., Müller, K.R.: Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.* **10**(5), 1299–1319 (1998)
11. Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(12), 1349–1380 (2000)
12. Su, J.H., Huang, W.J., Yu, P., Tseng, V.: Efficient relevance feedback for content-based image retrieval by mining user navigation patterns. *IEEE Trans. Knowl. Data Eng.* **23**(3), 360–372 (2011)
13. Torralba, A., Fergus, R., Weiss, Y.: Small codes and large image databases for recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8. *IEEE* (2008)
14. Wang, J., Li, J., Wiederhold, G.: Simplicity: semantics-sensitive integrated matching for picture libraries. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(9), 947–963 (2001)
15. Yang, J., Estivill-Castro, V., Chalup, S.: Support vector clustering through proximity graph modelling. In: *Proceedings of the 9th International Conference on Neural Information Processing, ICONIP 2002*, pp. 898–903 (2002)