Simultaneous Feature Selection and Classification for Relevance Feedback in Image Retrieval

Reshma Prasanna
Dept. of Electrical Engineering
Indian Institute of Science
Bangalore, 560 012 INDIA
reshma@ee.iisc.ernet.in

K. R. Ramakrishnan
Dept. of Electrical Engineering
Indian Institute of Science
Bangalore, 560 012 INDIA
krr@ee.iisc.ernet.in

Chiranjib Bhattacharyya
Dept. of Computer Science and Automation
Indian Institute of Science
Bangalore, 560 012 INDIA
chiru@csa.iisc.ernet.in

Abstract— In image retrieval, relevance feedback uses information, obtained interactively from the user, to understand the user's perceptions of a query image and to improve retrieval accuracy. We propose simultaneous relevant feature selection and classification using the samples provided by the user to improve retrieval accuracy. The classifier is defined by a separating hyperplane, while the sparse weight vector characterizing the hyperplane defines a small set of relevant features. This set of relevant features is used for classification and can be used for analysis at a later stage. Mutually exclusive sets of images are shown to the user at each iteration to obtain maximum information from the user. Experimental results show that our algorithm performs better than feature weighting, feature selection and classification schemes.

1. Introduction

With the advent of the Internet, there is widespread access to virtually unlimited information and a very large part of it is images. Compression algorithms, high capacity storage devices, improved processor speeds, and high speed networks have all contributed to the creation of many large collections of images. An image retrieval system provides the user with a way to access, browse and retrieve efficiently, in real time from these databases.

In a content-based image retrieval system, automatically extracted features are used to represent the image content. Low-level image features such as color, texture and shape are used to represent images. But similarity measures based on these features do not necessarily match perceptual semantics. In addition, each of the visual image features represents only one of the many aspects of image similarity. If multiple features are to be used, the question arises as to how to combine all the different features. Finally, different users' perceptions of the same image can vary widely and the same user's perception of a single image may also vary with time. In order to overcome these problems, relevance feedback techniques were introduced [1], [2]. Relevance feedback is an interactive process where the user is asked to evaluate the retrieval results and based on the user's evaluation, the retrieval mechanism is adapted in order to improve subsequent retrieval efficiency.

We propose a new relevance feedback scheme that performs simultaneous feature selection and classification using a *sparse classifier* [3]. The sparse classifier is the hyperplane separating the relevant images from the non-relevant images such that the weight vector characterizing the hyperplane has very few non-zero elements and hence is sparse. The non-zero elements of the weight vector correspond to a set of relevant features. Thus, relevant feature identification and classification are performed simultaneously. Choosing a set of relevant features eliminates the effect of any extraneous features during the classification process and decreases the time taken for classification of the images. Section 2 presents an overview of selected relevance feedback techniques, Section 3 presents the construction of a sparse classifier and Section 4 explains our algorithm for relevance feedback. Section 5 reports the experimental results Section 6 presents the conclusions.

2. Relevance Feedback

In a typical relevance feedback scenario, the retrieval system shows the user a set of retrieval results and the user provides some form of judgment on the retrieved images. The system then learns from the feedback, updates its parameters and derives a new set of retrieved images and the process continues iteratively.

Most of the early techniques for relevance feedback were based on two principles: (i) query point movement which moves the query point in the feature space such that more relevant images are included in the neighborhood of the query (ii) feature relevance weighting which stretches the feature space along those directions in which relevant images are well separated from non-relevant images. The MARS system [1] uses feature relevance weighting where the weight for each component is the inverse of the standard deviation of this component across the relevant examples. The new query is updated as the linear combination of the previous query, the average of the relevant examples and the average of the non-relevant examples.

Various classifiers have been used for relevance feedback. A Support Vector Machine (SVM) classifier has been trained using the relevant and non-relevant examples to obtain an optimal hyperplane separating the two classes. The technique in [4] uses active learning in addition to SVMs for relevance feedback.

The feature selection scheme proposed in [5] selects those features that maximize a feature relevance measure (the ratio of inter-class scatter to intra-class scatter). A certain number of features is selected and the rest are discarded. The method we propose improves on [5] since the number of features to be selected is not a constant. It varies from iteration to iteration and from query to query depending on the query itself and the relevance feedback provided by the user. The use of a sparse classifier differs from the use of SVMs in that every feature contributes to the construction of the discriminating surface when using SVMs, whereas, only a subset of the features contributes to the structure of the sparse classifier. Both feature selection and classification are simultaneously performed in our algorithm.

3. Simultaneous Relevant Feature Selection and Classification for Relevance Feedback

Simultaneous Feature Selection and Classification

A sparse classifier is used for simultaneous relevant feature identification and classification. The algorithm makes use of the property that minimization of the l_1 norm of a vector results in a sparse vector [6]. In this case, the vector corresponds to the weight vector of the hyperplane and hence minimizing its l_1 norm yields a hyperplane characterised by a sparse weight vector.

Consider the set $D = \{(\mathbf{x_i}, \mathbf{y_i}), \mathbf{i} = \mathbf{1}, \mathbf{2}, \ldots, \mathbf{n}\}$, where $\mathbf{x_i} \in \Re^\mathbf{k}$ are the k-dimensional training samples and y_i are the corresponding class labels. The samples are obtained by relevance feedback where the user marks a set of images shown to him/her as relevant or non-relevant to the query image. The two classes have labels +1 and -1 for relevant and non-relevant respectively, $y_i \in \{+1, -1\}$. Assuming that the samples of the two classes are linearly separable in $\Re^k$, we try to find a separating hyperplane, characterized by a weight vector $\mathbf{w} \in \Re^\mathbf{k}$ and a bias $b \in \Re$ such that

$$y_i[\mathbf{w}^\mathbf{T}\mathbf{x_i} + \mathbf{b}] \geq \mathbf{1}, \mid \mathbf{i} = \mathbf{1}, \mathbf{2}, \ldots, \mathbf{n} \tag{1}$$

A data vector $\mathbf{x}$, not part of the set of training samples, is classified by calculating $z = \mathbf{w}^\mathbf{T}\mathbf{x} + \mathbf{b}$. If $z \geq +1$, then $\mathbf{x}$ is assigned to the class with label +1. If $z \leq -1$, then $\mathbf{x}$ is assigned to the class with label -1.

We try to find a hyperplane that is characterized by a sparse weight vector $\mathbf{w}$ and that separates the relevant from non-relevant examples. Let $z = \mathbf{w}^\mathbf{T}\mathbf{x_i} + \mathbf{b}$, and let the j^{th} components of $\mathbf{w}$ and $\mathbf{x_i}$ be w_j and x_{ij} respectively. z can be written as $z = \sum_{j=1}^{k} w_j x_{ij} + b$. If the component $w_j = 0$, then this component of the weight vector does not contribute to the calculation of z and it can be ignored. Such components are then defined to be non-relevant whereas the components that actually do contribute to the calculation of z are defined to be relevant. Only the relevant components of $\mathbf{w}$ are then used during classification of all the images in the database. To find a sparse classifier, we try to maximize the number of zero elements in $\mathbf{w}$. Let the l_0 norm of $\mathbf{w}$, denoted as $||\mathbf{w}||_0$, be defined as $||\mathbf{w}||_0$ = number of $\{j \mid w_j \neq 0\}$ i.e. the number of non-zero elements of $\mathbf{w}$. The minimization of this norm, in general, requires a search through the set of all subsets of components of $\mathbf{w}$, looking for a particular subset that provides the sparse hyperplane required for classification. An alternate approach to this problem is to substitute the l_0 norm with the l_1 norm of the components of $\mathbf{w}$. This is a convex optimization problem and can be solved using linear programming methods such as the classical simplex method. The l_1 norm of a vector is the sum of the absolute values of the elements of the vector and it is defined by $||\mathbf{w}||_1 = \sum_{\mathbf{j}=1}^{\mathbf{k}} |\mathbf{w_j}|$. This leads to the optimization problem:

$$\begin{aligned} \min_{\mathbf{w},\mathbf{b}} \quad & ||\mathbf{w}||_1 \\ \text{such that} \quad & y_i[\mathbf{w}^\mathbf{T}\mathbf{x_i} + \mathbf{b}] \geq \mathbf{1}, \ \forall \mathbf{i} = \mathbf{1}, \mathbf{2}, \ldots, \mathbf{n} \end{aligned} \tag{2}$$

It has been shown [6] that the solution to this optimization problem is quite frequently sparse. In this manner, we get a sparse weight vector $\mathbf{w}$ for the separating hyperplane. In order to eliminate the $|\cdot|$ in the objective function, we replace w_j by two non-negative variables w_j^+, $w_j^- \in \Re$ so that $|w_j| = w_j^+ + w_j^-$. Similarly, the bias b is decomposed into b^+ and b^- so that $b = b^+ - b^-$. Let $\mathbf{w}^+$ and $\mathbf{w}^-$ be the vectors whose components are respectively w_j^+ and w_j^-, $j = 1, 2 \ldots, k$. Mis-classifications can occur if the training samples in the set D are not linearly separable, resulting in wrong labels y_i and we will not be able to find a separating hyperplane . In such a case, we modify the problem slightly to minimize the errors in addition to the minimization of the l_1 norm of the weight vector $\mathbf{w}$. The modified problem now includes the non-negative slack variables ξ_i. The final optimization problem is now:

$$\begin{aligned} & \min_{\mathbf{w}^+,\mathbf{w}^-,\mathbf{b}^+,\mathbf{b}^-} \sum_{j=1}^{k}(w_j^+ + w_j^-) + C\sum_{i=1}^{n}\xi_i \quad \text{such that} \\ & y_i[(\mathbf{w}^+ - \mathbf{w}^-)^\mathbf{T}\mathbf{x_i} + (\mathbf{b}^+ - \mathbf{b}^-)] \geq \mathbf{1} - \xi_\mathbf{i}, \\ & b^+ \geq 0,\ b^- \geq 0,\ \xi_i \geq 0,\ \forall i = 1, 2, \ldots, n \\ & w_j^+ \geq 0,\ w_j^- \geq 0,\ \forall j = 1, 2, \ldots, k \end{aligned} \tag{3}$$

C is an adjustable parameter that weighs the contribution of misclassified samples.

4. Algorithm For Relevance Feedback

The simultaneous relevant feature identification and classification are applied to the relevance feedback scenario as follows: The user provides a query image to the content-based retrieval system. The system returns to the user, a set of p images classified as relevant (p is the number of images displayed to the user). The images marked relevant or non-relevant by the user are used to train a sparse classifier which is used to classify the images in the database. The images classified as relevant by the sparse classifier are sorted in descending order of their distance to the separating hyperplane. The images farthest from the hyperplane are the most relevant. The top p images in this sorted list that have not been seen by the user in previous iterations are then displayed to the user for feedback. The feedback obtained is combined with the feedback from previous iterations to build a new sparse classifier. In this way, the process of building a classifier, classifying images in the database, showing the relevant images to the user and obtaining feedback is repeatedly performed. In the final iteration, the user is shown the p relevant images farthest from the hyperplane over all the previous iterations.

Analysis of Relevant Features Selected

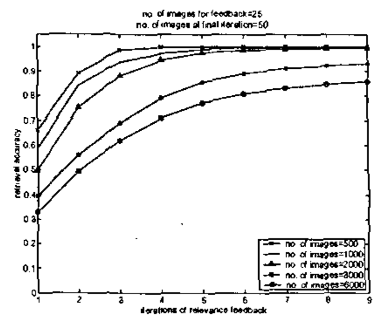
The identification of the relevant features is performed during the formation of the sparse classifier. Consider a set of images belonging to the same semantic category. Each of these images is used as a query image and a few iterations of relevance feedback are performed for each query. The set of selected feature components is stored for each query image. A histogram of the selected feature components for all the images in the set is then constructed and it is used to analyze which feature components and which features were selected the most number of times. Feature selection thus enables the association of a set of components of visual features with a semantic category. This association may aid future experiments to find and verify the features most pertinent to an image depending on its semantic content.

5. Experimental Results

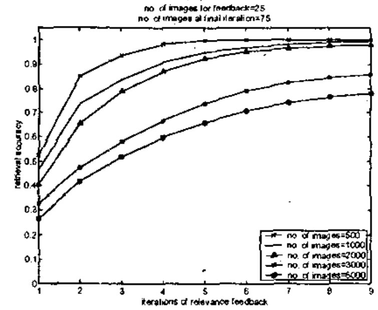
The datasets used contain natural images obtained from the Corel Image collection. The images are extremely diverse and heterogeneous and include cars, buses, horses, birds, flags, planes, etc. The datasets have been categorized a priori and we have restricted the selection of categories such that an image belongs to only one category. We define the *ground-truth* for a query image as the set of all the images in the same category as the query image. We use five different datasets of varying size:(1) 500 images, 5 categ. (2) 1000 images, 10 categ. (3) 2000 images, 15 categ. (4) 3000 images, 23 categ. (5) 6000 images, 45 categ.

Content-based retrieval for databases with heterogeneous images depends heavily on the features used for retrieval. In our experiments, we have tried to include features that address different aspects of visual similarity. These include color histogram in the HSV space, color moments, color coherence vector, color correlogram, Gabor filter responses, energy of wavelet coefficients and edge histogram.

The retrieval accuracy is defined as the ratio of the number of relevant images shown to the user to the number of images shown to the user. We have compared our algorithm with three other schemes: (i) MARS system[1]. (ii) feature selection scheme [5]. (iii) classification scheme using SVMs and active learning [4]. We compare with these schemes since each of them performs only one of the two tasks that our method performs simultaneously i.e. feature selection and classification. These schemes and our algorithm are denoted by MARS, FS, SVMAL and SFSC resp. in the graphs shown.



(a)



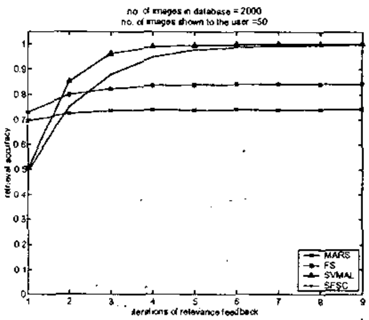
(b)

Figure 1. Retrieval Accuracy (a) No. of images shown = 50 (b) No. of images shown = 75

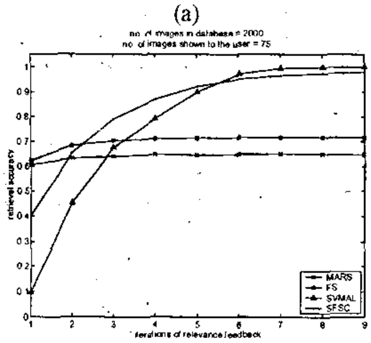
Results

From Fig. 1(a), we can see that for dataset 1 (500 images), with 50 images shown to the user, the retrieval accuracy increases from 0.66 after 1 iteration, to 1.0 after 5 iterations of relevance feedback. As the number of images in the dataset increases, the retrieval accuracy decreases. This is expected since in a larger dataset, there are more images that are similar but non-relevant to the query image. Even for the largest dataset (6000 images), the retrieval accuracy improves from 0.33 after 1 iteration, to 0.85 after 9 iterations of relevance feedback. For an average size dataset of 2000 images (dataset 3), when showing the user 50 images, the retrieval accuracy increases from 0.5, after 1 iteration, to 1.0 after 8 iterations of relevance feedback. For the same dataset, when showing the user 75 images in the final iteration (Fig. 1(b)), the retrieval accuracy improves from 0.4, after 1 iteration, to 0.98 after 9 iterations of relevance feedback. Thus, even when a large number of images are shown to the user, up to 98% percent retrieval accuracy is achieved.

Fig. 2 compares the performances of the four algorithms for an average size dataset (dataset 3, 2000 images). The retrieval accuracy of both the algorithms that build classifiers (SVMAL and SFSC) is seen to be initially worse than the other two algorithms. But after the initial improvement, the retrieval accuracy stays almost constant for MARS and FS. By contrast, SVMAL and SFSC improve retrieval accuracy with each iteration. Fig. 2 also shows that our algorithm per-



(a)



(b)

Figure 2. Retrieval Accuracy for dataset 3 (2000 images) for 4 different algorithms. (a) No. of images shown = 50 (b) No. of images shown = 75

forms slightly worse than the SVMAL when the number of images shown to the user is 50. But as we increase this number to 75, our algorithm outperforms SVMAL for the first 5 to 6 iterations and after that, its performance is comparable.

Our algorithm scores over SVMAL in the retrieval time. The sparse classifier is constructed by solving a linear programming optimization problem whereas SVMAL solves a quadratic programming optimization problem, thus making our algorithm faster. Another factor in saving time is that our algorithm performs feature selection, using a small subset of the original features to classify the images in the database, resulting in time saving. In contrast, every single feature is used for classification in SVMAL. The experiments were conducted on a 1-GHz Pentium III processor powered PC running Linux operating system. The total retrieval time t_{tot} is split into *training time* t_{tr} and *classification time* t_{cl}. t_{tr} is the time taken to construct the classifier using the set of training samples. t_{cl} is the time taken to classify all the images in the database using the classifier. From Table 1, it can be seen that the training time for our algorithm for dataset 1 is 3 to 6 times smaller than that for SVMAL and our algorithm takes half the time taken by SVMAL for classification.

Given an image of a bus as a query, Fig. 3 shows the 1^{st} and 3^{rd} iterations of relevance feedback, and the final set of images after 5 iterations of feedback. It can be seen that our retrieval system is able to retrieve 48 images of a bus with 5

Table 1. Training, classification and total retrieval time (in seconds) for dataset 1 (500 images) for SVMAL and SFSC algorithms

Itn.	SVMAL			SFSC		
No.	t_{tr}	t_{cl}	t_{tot}	t_{tr}	t_{cl}	t_{tot}
1	0.1	2.4	2.5	0.8	0.8	1.6
2	1.0	2.9	3.9	2.3	0.6	2.9
3	3.3	3.1	6.4	3.6	0.8	4.4
4	8.5	3.4	11.9	5.4	0.6	6.0
5	15.1	3.5	18.6	8.6	0.7	9.3
6	24.7	3.4	28.1	12.4	0.8	13.2
7	31.4	3.6	35.0	15.3	0.6	15.9
8	40.5	3.7	44.2	18.3	0.8	19.1

iterations of feedback. Fig. 4 shows similar results for a query image of an eagle. It can also be seen that the sets of images at each iteration are mutually exclusive, and all the relevant images are consolidated in the final results.

6. Conclusions

We have proposed a scheme for simultaneous feature selection and classification for relevance feedback using a sparse classifier. The weight vector characterizing the sparse hyperplane has non-zero elements that correspond to the relevant features. Experiments have shown that retrieval accuracy improves at every iteration. A comparative evaluation of our relevance feedback algorithm and three other existing algorithms shows that our algorithm performs better than the feature relevance weighting and feature selection schemes and comparably with the classification scheme using SVMs in terms of retrieval accuracy, and it has the advantage of being faster than the classification scheme using SVMs.

References

[1] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, "Relevance feedback: A power tool in interactive content-based image retrieval," *IEEE Transactions on Circuits and Systems for Video Technology*, **8**, 644–655, 1998.

[2] Y. Ishikawa, R. Subramanya, and C. Faloutsos, "Mindreader: Querying databases through multiple examples," in *Int. Conf. on Very Large Databases*, Aug 1998.

[3] L. R. Grate, C. Bhattacharyya, M. I. Jordan, and I. S. Mian, "Simultaneous relevant feature identification and classification in high-dimensional spaces," in *Workshop on Algorithms in Bio-informatics*, Sept 2002.

[4] S. Tong and E. Chang, "Support vector machine active learning for image retrieval," in *ACM Int. Conf. on Multimedia*, Oct 2001.

[5] C. V. Jawahar, P. J. Narayanan, and S. Rakshit, "A flexible scheme of representation, matching and retrieval of images," in *Indian Conference on Computer Vision, Graphics and Image Processing*, Dec 2000.

[6] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Transactions on Information Theory*, **47**, 2845–2862, Nov 2001.

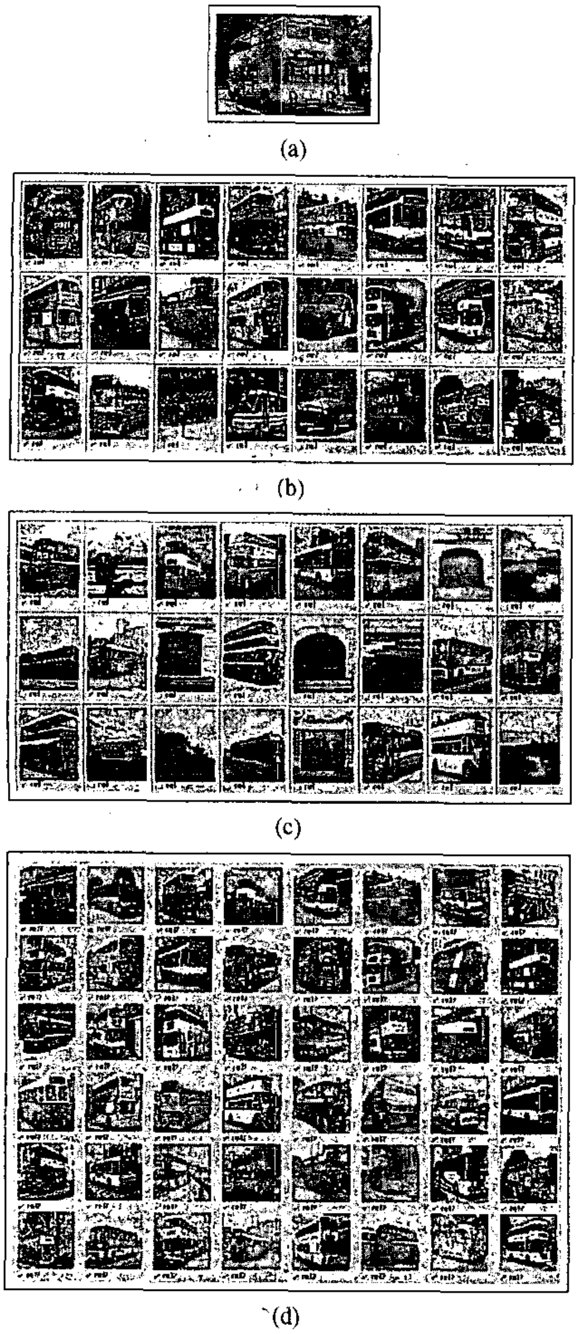


Figure 3. Query:'Bus' (a) query image (b) 1^{st} iteration of relevance feedback (c) 3^{rd} iteration of relevance feedback (d) Final set of relevant images after 5 iterations

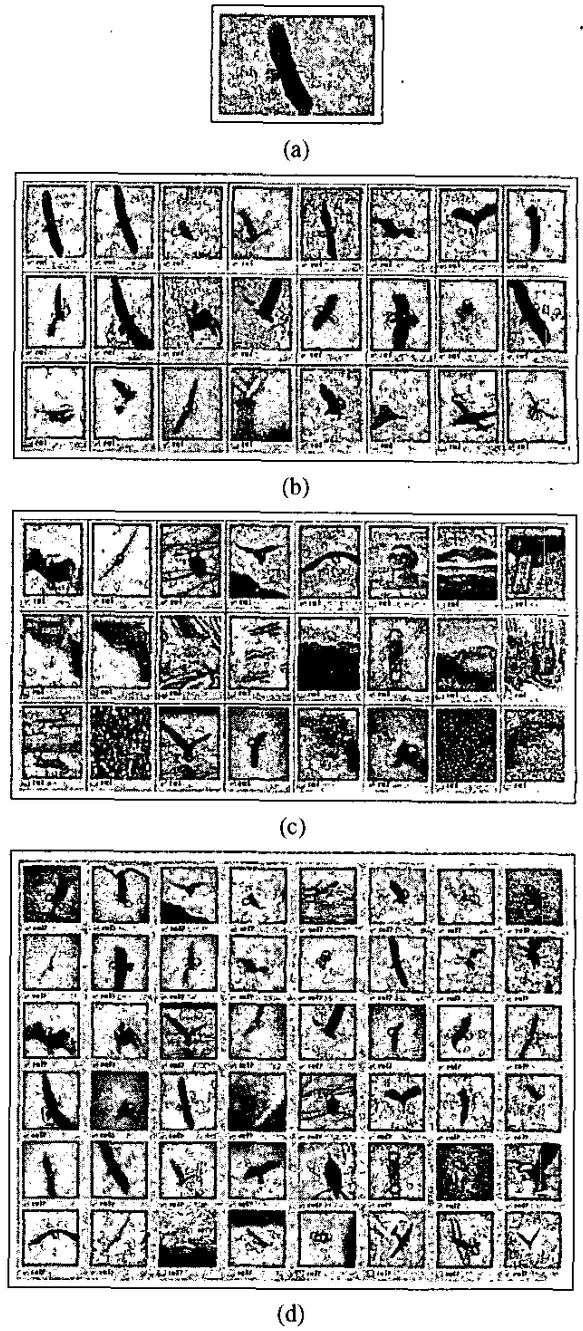


Figure 4. Query:'Eagle' (a) query image (b) 1^{st} iteration of relevance feedback (c) 3^{rd} iteration of relevance feedback (d) Final set of relevant images after 5 iterations