POSTER: A novel Content-based Image Retrieval system based on Bayesian Logistic Regression

J.P. Arias-Nicolás
University of Extremadura, Spain
jparias@unex.es

F. Calle-Alonso
University of Extremadura, Spain
fcalonso@unex.es

ABSTRACT

In this work, a novel content-based image retrieval (CBIR) method is presented. It has been implemented and run on "Qatris IManager" [14], a system belonging to SICUBO S.L. (spin-off from University of Extremadura, Spain). The system offers some innovative visual content search tools for image retrieval from databases. It searches, manages and classifies images using four kinds of features: colour, texture, shape and user description.

In a typical CBIR system, query results are a set of images sorted by feature similarities with respect to the query. However, images with high feature similarities to the query may be very different from the query in terms of semantics. This discrepancy between low-level features and high-level concepts is known as the semantic gap.

The search method presented here, is a novel supervised image retrieval method, based in Bayesian Logistic Regression, which uses the information from the characteristics extracted from the images and from the user's opinion who sets up the search. The procedure of search and learning is based on a statistical method of aggregation of preferences given by Arias-Nicolás et al. [1] and is useful in problems with both a large number of characteristics and few images.

The method could be specially helpful for those professionals who have to make a decision based in images, such as doctors to determine the diagnosis of patients, meteorologists, traffic police to detect license plate, etc.

Keywords: Computer Vision, CBIR, Logistic Regression, Similarity, Pairwise Comparisons.

1 INTRODUCTION

Research in Content-Based Image Retrieval (CBIR) is today a very active discipline, concentrating on depth issues, such as learning or management access to information content in images. Two fundamental problems remain largely unsolved: how to best learn from users' query concepts, and how to measure perceptual similarity. A very popular framework in the 1970s was to first annotate the images by text, and then to use text-based Database Management Systems (DBMS) to perform image retrieval. The essential difficulty in this method results from rich content in the images and subjectivity in human perception: the same image content may be perceived differently by other users, [16].

Classification techniques are usually applied in CBIR systems. Image categorization contributes to perform more effective searches. In the repertoire of images under consideration there is a gradual distinction between narrow and broad domains. A broad domain has an unlimited and unpredictable variability in its appearance even for the same semantic meaning [15]. The good performance of classifiers has been proved when the image domain is specific, i.e, it is a narrow domain which has a limited and predictable variability in relevant aspects for the specific purpose, [6, 7].

Actually, there are lots of works directly related to CBIR, and QBIC system is one of the first [8]. It was developed by IBM Corporation and is commercially available. This kind of CBIR systems focus on work with images over a broad domain. Another system is PicHunter [5] which uses Bayesian learning based on a probabilistic model of user's behaviour. This system works with features from images and introduce relevance feedback from the users' opinions. Most recently Liu et al. [12] focused on a powerful feature selection method always addressed to cover the semantic gap; however, these authors do not develop a good method to analyze similarity, but do so for metric distance.

When classification methods are applied to general-purpose image collections the results are not positive, even less if we hope that the performance of the classifier match with the classification developed by non-expert humans. We find some examples in [17, 18].

We thus aim to develop a pairwise comparison method based on the Binary Logistic Regression in order to determine the images that can match one lacking some information. The proposed framework is based on [1], which focused on how to aggregate personal preferences to arrive at an optimal group decision. We are interested in searching for similar images with respect to several features, when we only know the similarity between some pairs of the images. The method for CBIR proposed, combining information of both computational features and user's knowledge.

2 THE METHOD

2.1 Feature Extraction

Each image is represented by a feature vector of features. We have considered three kinds of features: color, texture and shape features.

- **Color Features.** The color features used in this work are based on HLS model (Hue, Saturation, Luminosity), since the human perception is quite similar to this model. On the other hand, local color features are used in order to achieve information about the spatial distribution [4].
- **Texture Features.** They have been obtained applying two well known methods. The first one works on a global processing of images, it is based on the Gray Level Co-ocurrence Matrix proposed by Haralick [10]. The second method is focused on detecting only linear texture primitives. It is based on features obtained from Run Length Matrix proposed by [9].
- **Shape Features.** The images are processed using Active Contours as segmentation method and then some shape features are obtained from these contours. Shape features are based on Hu's moments(first and second moments), centroid (center of gravity), angle of minimum inertia, area, perimeter, ratio of area and perimeter, and major and minor axis of fitted ellipse. The methods to obtain these features are explained in [3].

2.2 Clasification

Once features vectors are obtained, we apply our proposed search method. An efficient images supervised classification method based on bayesian logistic regression has been implemented, which stands out for his high probability of wise move and his facility of incorporating the user's opinion in the learning phase.

The method needs a training stage and a posterior testing one. In the first phase, we require a set of previously classified images to determine the pattern that will be used in the phase of test for classification of new images. One of the important advantages of using the bayesan logistic regression, with regard to others methods, is to obtain a progressive process of the reliability of classifier incorporating the user's opinions and corrections in the successive phases of learning.

The multinomial logistic regression model is a direct generalization of the binary logistic regression for K classes. So, we can classify a new element $\mathbf{x}$ in a class $k \in \{1,2,\ldots,K\}$, for what we can assign to it a K-dimensional vector with values $0-1$, $\mathbf{y} = (y_1, y_2, \ldots, y_K)^t$ where $y_k = 1$ and the others ones 0. Multinomial logistic regression is a model of conditional probability of the form:

$$p(y_k = 1|\mathbf{x}, \mathbf{B}) = \frac{e^{(\mathbf{B_k^t x})}}{\sum_{i=1}^{K} e^{(\mathbf{B_i^t x})}}, \quad (1)$$

and standardize by the matrix $\mathbf{B} = (\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_K)$. Every column of $\mathbf{B}$ is a vector of parameters corresponding to each of the classes: $\mathbf{B_k} = (\beta_{k1}, \beta_{k2}, \ldots, \beta_{km})^t$.

The most widely used bayesian approach to the model of logistic regression is to impose a gaussian distribution with mean 0 and variance σ_{kj}^2, for every parameter β_{kj} ([2, 13]):

$$p(\beta_{kj}|\sigma_{kj}) = \frac{1}{\sqrt{2\pi}\sigma_{kj}} \exp\left(\frac{-\beta_{kj}^2}{2\sigma_{kj}^2}\right). \quad (2)$$

The classification of a new image is based on the vector of conditional probabilities estimated by the model. For this, simply the image is assigned to the class with the highest estimated probability. The maximum likelihood estimation of the parameters $\mathbf{B}$ is equivalent to maximize:

$$L(\mathbf{B}) = l(\mathbf{B}|\mathbf{X}) + \ln p(\mathbf{B}), \quad (3)$$

being

$$l(\mathbf{B}|\mathbf{X}) = -\sum_{i=1}^{n}\left[\sum_{k=1}^{m} y_{ik}\mathbf{B}_k^t\mathbf{x}_i - \ln\sum_{k=1}^{m} e^{(\mathbf{B_k^t x_i})}\right] \quad (4)$$

and $p(\mathbf{B})$ is the joint distribution of vector $\mathbf{B}$.

$L(\mathbf{B})$ is optimized by iteratively maximizing a surrogate function Q, thus (see e.g. [11]):

$$\hat{\mathbf{B}}^{(t+1)} = arg\max_{\mathbf{B}} Q(\mathbf{B}|\hat{\mathbf{B}}^{(t)}). \quad (5)$$

Once the classifier has been trained, it will be applied to new images whose classification is unknown. The resultant model is applied on a new vector of characteristics to obtain a vector of K probabilities, where K is the number of classes. The $k-$element of the vector represents the probability of the new image belongs to class k. Therefore, the resultant value to apply the classifier will be the class with major probability of belonging.

2.3 Similarity measure

Having a prior knowledge about the similarity between images, our objective is to find the most similar images to the one given (with respect to the obtained features).

First, we sample r pairs of images from the image database. The idea is to determine a unique discrepancy distance (d_g) for all images, that models the similarity between images.

Then for every pair (a,b) of images, we compute:

- The independent variables, i.e., the distance between their features:

  $$\mathbf{x}_{ab} = (d_1(a,b), d_2(a,b), ..., d_n(a,b)), \quad (6)$$

  where d_i is a distance function that models the similarity with respect to the feature i ($i = 1, \ldots, n$). Note that, since d_i are distances, these variables are non negative.

- The answer variable or dependent variable, Y_{ab}, is a boolean set of variables, with a value of 0 if the images (a and b) belong to the same class, or 1 if not.

This is the training stage of the system.

We will apply modified Logistic Regression to

$$(x_{j1}, x_{j2}, \ldots, x_{jn}, y_j),\ j = 1, \ldots, r.$$

Since our objective is to determine a measurement of discrepancy among images, and as the independent variables are non-negative, the linear predictor will be non-negative. Thus we consider a link function that transforms π into a quantity that takes values in the interval $[0, +\infty)$. Then we consider the link function:

$$g(\pi) = \log\left(\frac{1+\pi}{1-\pi}\right). \tag{7}$$

In this case,

$$\pi = \frac{e^{\mathbf{x}^t\beta} - 1}{e^{\mathbf{x}^t\beta} + 1} \text{ and } 1 - \pi = \frac{2}{e^{\mathbf{x}^t\beta} + 1}. \tag{8}$$

We estimate that parameters $\beta = (\beta_1, \beta_2, \ldots, \beta_n)$, which maximize the likelihood function:

$$L(\beta) = \sum_{i=1}^{r} y_i \log\left(0.5(e^{\mathbf{x}_i^t\beta} - 1)\right) - \log\left(0.5(e^{\mathbf{x}_i^t\beta} + 1)\right),$$

being $\beta \geq \mathbf{0}$.

Unfortunately there is no analytical solution for $\hat{\beta}$ (estimated β), but we may resort to a Newton-Raphson iterative procedure. Each cycle in this procedure provides an updating formula given by:

$$\hat{\beta}^{(\mathbf{k+1})} = \hat{\beta}^{(\mathbf{k})} + (\mathbf{X^tWX})^{-1}\mathbf{X^t\hat{Z}}(\mathbf{Y} - \hat{\mathbf{Y}}), \tag{9}$$

where $\mathbf{Y}$ denotes the vector of response values, $\mathbf{X}$ denotes a matrix with each row by $\mathbf{x}_i^t$, $\hat{\mathbf{Y}}$ the vector of estimated values at that iteration and $\mathbf{W}$, $\mathbf{Z}$ denotes the diagonal matrix with elements:

$$\hat{z}_i = \frac{e^{\mathbf{x}_i^t\hat{\beta}^{(\mathbf{k})}}}{e^{\mathbf{x}_i^t\hat{\beta}^{(\mathbf{k})}} - 1}, \text{ and}$$

$$\hat{w}_i = y_i \frac{e^{\mathbf{x}_i^t\hat{\beta}^{(\mathbf{k})}}}{(e^{\mathbf{x}_i^t\hat{\beta}^{(\mathbf{k})}} - 1)^2} + \frac{e^{\mathbf{x}_i^t\hat{\beta}^{(\mathbf{k})}}}{(e^{\mathbf{x}_i^t\hat{\beta}^{(\mathbf{k})}} + 1)^2}, \tag{10}$$

respectively. This formula is used until the estimates converge.

We obtain:

$$\hat{\pi}_{ab} = \hat{P}[Y_{ab} = 1] = \frac{e^{\mathbf{x}_{ab}^t\hat{\beta}} - 1}{e^{\mathbf{x}_{ab}^t\hat{\beta}} + 1}, \tag{11}$$

the probability of a and b are different.

Note that for all $a \in \mathscr{A}$, it holds that $d_i(a,a) = 0$ ($i = 1, \ldots, n$). Therefore, $\hat{\pi}^{aa} = 0$. Given $a, b \in \mathscr{A}$, we say that a is similar to b if $\hat{\pi}^{ab}$ is close to 0. However if two images a and b are very different with respect to some feature ($d_i(a,b)$ tends to infinite) and $\beta_i > 0$, then the probability that a and b be different, $\hat{\pi}^{ab}$ tends to 1.

We aim to determine the most similar images to the image under study by applying this method. For a new image c, we can compute $\hat{\pi}_{ca_j}$, that can be interpreted as the discrepancy degree between c and a_j. Note that if c and a_j are very similar, the probability of discrepancy is near 0. Then, we must look for the images a_j so that the probability of this one being different from c is near 0. Observe that $\hat{\pi}^{ab}$ near 0 is equivalent to $\mathbf{x}_{ab}^t\hat{\beta}$ near 0. Then, we could consider:

$$d_g(\cdot) = \sum_{i=1}^{n} \hat{\beta}_i d_i(\cdot), \tag{12}$$

as the measure of discrepancy between images.

2.4 Relevance feedback to searching

The method may improve by providing user's opinion about retrieved images at each step. The β parameter is then updated by applying the Bayesian Logistic Regression method, and by showing new images to the user. A measure is thus obtained for each query image and user. This step corresponds to the learning process in the methodology, see Figure 1.

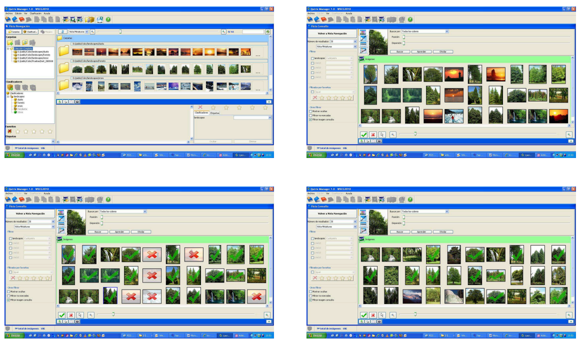

Figure 1: Searching and relevance feedback.

The system displays the most similar images on database to the Query Image (QI). The user can interact with the system, indicating which of the retrieval images (RI) correspond to their search conditions (RI-YES) and which do not (RI-NO). Therefore, the opinion of the user in every query allows to add new pairs by means of a process of feedback, see Table 1.

The system can learn from the answers provided by the user and update vector β with the above mentioned information (feedback). Once optimized, the system offers new similar images. The process repeats itself until the user demonstrates an agreement with the result.

Image	QI	RI-YES	RI-NO
QI	...	$Y=0$	$Y=1$
RI-YES	$Y=0$	$Y=0$	$Y=1$
RI-NO	$Y=1$	$Y=1$	...

Table 1: Learning process (QI: Query Image; RI: Retrieval Image; RI-YES: RI corresponds to the QI; RI-NO: RI does not correspond to the QI).

3 CONCLUSIONS

In this paper we have tested a novel pairwise comparison method to find most similar images and a classification method based on Bayesian logistic regression. The method has sorted query results by similarity, so, we can say that it works like a similarity measure. The technique is based on the Logistic Regression method, and it is useful to get information for Decision Making. In addition the method is integrated in a commercial image database system. It is also particularly useful to solve ranking issues involving a large number of features and few images. Finally, the method is easily applied in practice. By extracting color, texture and shape features from the digital images, acceptable results are obtained for some real CBIR problems. This method proposes an approach to the semantic gap.

ACKNOWLEDGEMENTS

This research was partially supported by project TIN2008-06796-C04-03 from *MEC* and project PDT09A009 from *Junta de Extremadura*, Spain.

REFERENCES

[1] Arias-Nicolás, J.P., Martín, J., Pérez, C.: A logistic regression-based pairwise comparison method to aggregate preferences. Group Decision and Negotiation, **17**(3), 237–247 (2008).

[2] Berger, J.O.: Statistical Decision Theory and Bayesian Analysis. Springer (1985).

[3] Caro, A., Rodríguez, P.G., Antequera, T., Palacios, R.: Feasible application of shape-based classification. Pattern Recognition and Image Analysis. LNCS Vol.4477, pp.588-595 IbPRIA (2007).

[4] Cinque L., Levialdi S., Pellicano A., Olsen K.A.: Color-Based Image Retrieval Using Spatial-Chromatic Histograms. IEEE International Conference on Multimedia Computing and Systems (ICMCS'99), Vol.2, pp. 969-973, (1999).

[5] Cox, I.J., Miller, M.L., Minka, T., Papathornas, T., Yianilos, P.: The Bayesian Image Retrival System, PicHunter: Theory, Implementation and Psychophysical experiments. IEEE Transactions on Image Processing, **9**, 20–37 (2000).

[6] Dumais, S., Platt, J., Heckerman, D., Sahami, M.: Inductive learning algorithms and representations for text categorization. Proc. of 7th International conference on Information and Knowledge Management. (1998).

[7] El-Naqa, I., Yang, Y., Galatsanos, N.P., Nishikawa, R.M., Wernick, M.N.: A similarity learning approach to content-based image retrieval: application to digital mammography. IEEE Transactions on Medical Imaging **23**(10), 1233–1244 (2004).

[8] Flickner, M.D., Sawhney, H.S., Niblack, C.W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., Yanker, P.C.: Query by Image and Video Content: The QBIC System. IEEE Computer **28**(9), 23–32 (1995).

[9] Galloway M. M.: Texture Analysis using Grey Level Run Lengths. Computer Graphics and Imag. Processing, vol. 4, pp. 172-179 (1975).

[10] Haralick, R.M., Shapiro, L.G.: Computer and Robot Vision. Vol.I Addison-Wesley (1993).

[11] Krishnapuram, B., Carin, L., Figueiredo, M., Hartemink, A.J.: Sparse Multinomial Logistic Regression: Fast Algorithms and Generalization Bounds. IEEE Transactions on Pattern Analysis and Machine Intelligence, **27**, 6, 957–968 (2005).

[12] Liu, W., Peng, F., Feng, S., You, J., Chen, Z., Wu, J., Yuan, K., Ye, D.: Semantic Feature Extraction for Brain CT Image Clustering Using Nonnegative Matrix Factorization. Lecture Notes in Computer Science **4901**: Medical Biometrics, 41–48 (2008).

[13] McCullagh, P., Nelder, J.A.: Generalized Linear Models, second ed.. Chapman & Hall (1989).

[14] Qatris Imanager, SICUBO S.L. (www.sicubo.com), Cáceres, Spain.

[15] Smeulders, A., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-Based Image Retrieval at the End of the Early Years. IEEE Transactions On Pattern Analisys and Machine Intelligence **22**(12), 1349–1380 (2000).

[16] Tversky, A.: Features of similarity. Phychological Review **84**(4), 327–352 (1977).

[17] Vailaya, A., Figueiredo, M., Jain, A., Hong Jiang, Z.: Content-based hierarchical classification of vacation images. Multimedia Computing and Systems, IEEE International Conference. Vol. 1, pp. 518-523 (1999).

[18] Li, J., Wang, J.Z.: ALIP: the Automatic Linguistic Indexing of Pictures system. Computer Vision and Pattern Recognition. IEEE Computer Society Conference. Vol 2, pp. 1208-1209 CVPR (2005).