Building Up Low-Level Centroids for Groups of Perceptually Similar Images[1]

Ilya MARKOV [a], Natalia VASSILIEVA [b]
[a] *Saint-Petersburg State University, ilya.markov@gmail.com*
[b] *HP Labs, Saint-Petersburg, Russia, nvassilieva@hp.com*

Abstract. Image retrieval by using content analysis is known as a difficult task. It is hard to work out common features and metrics that match all perceptually similar images well and distinguish non-similar ones. In our previous studies [8] and [7] mixed-metrics were proposed in order to combine color and texture metrics for image retrieval task. It was shown that it is always possible to mark out the best mixed-metrics for every group of similar images. Thus, usage of appropriate mixed-metrics for a given query improves retrieval effectiveness. To get the proper mixed-metrics a particular query-image should be classified to one of the predefined groups of perceptually similar images. In this study possible solutions for classification task are discussed, experimental results for one of the solution are given.

Keywords. Content-Based Image Retrieval, Mixed-Metrics, Image Classification

Introduction

Content-based image retrieval is a way of searching in image collections when no additional information about images is given. Every image in a collection is represented by one or several feature vectors. These *feature vectors* are numerical descriptors of image content. Feature vectors are built upon the information extracted from pixel values. Usually one feature vector represents one of the low-level image features (color, texture, shape). At query time feature vectors are extracted from a query-image and compared against existing vectors of images stored in an image collection.

Common approach is to match vectors describing different low-level features independently. All vectors of an image collection describing the same feature form multidimensional vector space (feature space). To measure image similarity one should define metrics for every feature space. The main assumption here is that two images are similar to each other if and only if corresponding vectors are similar too in accordance with the metrics specified.

Many researchers showed that it is necessary to combine various features for effective image retrieval. Merging results obtained for the same query by using different metrics in different feature spaces is a challenging task here.

Color and texture are the common features which are used for searching in natural images. In [8] we proposed a technique to combine color and texture metrics taking

[1] This work was partially supported by RFBR (grant 07-07-00268a).

into account a particular query-image. Weighed linear combination of color and texture metrics (mixed-metrics) is considered as a fusion function. Our approach is based on the hypothesis, proposed and proved in [8], that there are optimal weights to combine color and texture metrics for every query-image and these weights are unique for a given query. By using these optimal weights one can improve retrieval results. We showed that it is always possible to mark out the best mixed-metrics for every group of similar images (and thus for every query-image). In [7] it was shown that mixed-metrics based approach outperforms widely used CombMNZ data fusion method in some cases, and has close results in others.

It was stated that in order to perform a search over an image collection by using mixed-metrics one should prepare training set which represents the collection well. The training set should be somehow divided into groups of similar images and best mixed-metrics for every group should be calculated. To perform a search itself, one should classify query-image as belonging to one of the groups. After classification a search can be performed by using mixed-metrics of the group which the query-image is classified to.

The approach described above will improve retrieval efficiency only if classification task can be solved. The key problem here is that groups consist of images that are similar according to human visual perception, while the classification is performed based on low-level feature vectors. It is a well-known problem of "semantic gap" in content-based image retrieval, cased by the following reason: humans match images based on semantic content while image retrieval systems judge image similarity based on low-level features.

In this paper we present a simple solution for classification task in the described context. We use a notion of centroid taken from cluster analysis domain which denotes cluster center there. Every group of perceptually similar images is represented with a set of such "centers" or centroids. A query-image classification is performed based on distances between its feature vectors and groups' centroids. The better centroids represent groups the more precise classification and retrieval are.

Simple algorithm of building up one centroid per group is proposed and possible improvements of the algorithm are discussed. While classification results are not so good (precision is low for some groups), experiments show slight improvement of retrieval efficiency compared to CombMNZ method.

1. Related Works

Our goal is to improve image retrieval results by effective combination of color and texture features for a particular query-image. To achieve this goal we need a data fusion algorithm and a solution for query-images classification. Subsection 1.1 describes data fusion methods that can be used in image retrieval and subsection 1.2 describes different approaches which address image classification problem.

1.1. Data Fusion

Many researchers showed that it is necessary to combine various features for effective image retrieval. At the same time not enough attention is paid to the particular fusion methods. A number of works dedicated to similarity measures combining for image retrieval task is relatively small.

In [1] authors examine an application of a fuzzy logic approach to the fusion of image features. While it is a promising technique, there is no similarity measure proposed for observed fused feature and no experimental or other results are shown that can prove an efficiency of this approach.

Common solution is to fuse similarity measures calculated based on different features but not the features themselves. Linear combination of multiple similarity measures is usually treated as an aggregate measure (in [3] for instance).

Common data fusion algorithms like CombSUM, CombMNZ [2] and others ([6], [5]) are widely used in text retrieval. The same algorithms can be applied to image retrieval domain.

CombMNZ is considered to be one of the best data fusion algorithms. It performs as follows. Element in the result ranked-list gets rank equaled to the sum of all its ranks in fused lists divided by the number of lists in which this element exists with non-zero rank:

$$rank_{result}(obj) = \sum_{fused\ lists} rank_{list}(obj) \cdot nz(obj),\ \ \forall\ obj \in image\ collection,$$

where $nz(obj)$ is the number of fused lists in which the given object exists with non-zero rank. This algorithm is simple to use and outperforms other data fusion methods [5].

In [14] we proposed our own data fusion method "Weighted Total with Gravitation Function" (WTGF) and compared it to CombMNZ, applied to the image retrieval domain. WTGF function satisfies various criteria like symmetry, monotonicity and so called "cone rules". Experimental results showed that WTGF outperforms CombMNZ in case there are multiple inputs of non-equal reliability (we can trust to one input more than to others) and inputs do not overlap much. In case information about element ranks is not trusted (all inputs have the same reliability) and inputs overlap a lot, CombMNZ outperforms WTGF.

1.2. Image Classification

Grouping images into semantic classes by using low-level visual features is a challenging and important task. Many researchers claim that solving this problem can not only significantly improve image retrieval results ([12], [4], [10]), but could also allow using more convenient query forms instead of query-by-example ([15], [16]).

The first researchers following the idea of general image classification proposed usage of binary classifier to decide whether an image belongs to a certain semantic class or not. These studies include techniques that can decide whether an image is indoor/outdoor ([10], [11]), city/landscape ([13]). Vailaya et al. [13] stated that multiple two-class classifications may be more feasible than multi-class classifications. Usage of weighted k-NN classifier is typical for these binary classification tasks.

The same authors summarized binary classifiers in [12], proposing binary hierarchy of vacation images, where the first level of the hierarchy is indoor/outdoor images, the second is city/landscape, and the other levels are related to specific classes of natural scenes. A Bayesian classifier is used to assign an image to its best related group for each level of the binary hierarchy.

In [4] authors proposed multi-class classification of outdoor scenes by computing image similarity based on region matching. Experiments showed that this approach improves classifier performance: precision = 0.38, recall = 0.63 for global similarity, precision = 0.46, recall = 0.70 for the best case with region matching. k-NN classifier was used in this study for image classification.

2. Statement of Problem

In [8] we proposed a technique to combine image similarity measures which takes into account a particular query-image. We introduced mixed-metrics obtained from color and texture metrics (C and T respectively) by using their weighted linear combination $a \cdot C + (1-a) \cdot T$, where a is a varying coefficient which depends on a query-image. We stated and proved the hypothesis that optimal value of a is the same for similar query-images.

It was shown that in order to perform searches over an image collection by using mixed-metrics, training set which represents the collection well should be prepared. The training set should be somehow divided into groups of similar images and optimal mixed-metrics coefficient a should be calculated for every group by using algorithm proposed in [8].

To perform a search with a particular query-image, it should be classified to one of the groups of similar images defined on the previous step. While classification should be performed in real time during retrieval process it should be as fast as possible, therefore it should be simple enough and involve just a few computations. For this reason, many well-known classification algorithms cannot be used in our environment. Nearest Neighbor algorithm, for instance, assume as many distance evaluations as many images there are in a training set.

To reduce the number of distance calculations during classification we build centroids for every group for every feature (in our case color and texture features are combined, therefore two centroids should be built for every group: one representing ”central” color of the group and other representing ”central” texture). In this case the number of distance calculations is equaled to the number of groups multiplied by the number of features because query-image should be compared to the groups' centroids only.

As will be shown in section 4 it is better to use mixed-metrics instead of pure color or pure texture metrics in order to calculate distances between query-image and groups' centroids, but optimal mixed-metrics do not always perform well in this task. Therefore special mixed-metrics for classification should be defined for this purpose.

Thus, we should develop algorithms for building up centroids for groups of perceptually similar images and find appropriate mixed-metrics for every group to perform classification task.

3. Centroid Building Algorithms

3.1. Best Image of the Group

Color/texture group's centroid is an image, belonging to the group, having the highest precision when performing a search by using color/texture metrics with this image as a query. Building algorithm is the following: for every image of the group a search by using color/texture metrics should be performed and retrieval precision should be calculated. Images from the same group as the query-image are treated as relevant, while others are not. Image with the highest precision is treated as a color/texture centroid of the group.

3.2. Improvements

”Best image of the group” is very simple and computationally cheap strategy, but it has one major drawback. When one group of semantically similar images consists of several subgroups where images are similar according to their low-level content, the feature vector of the ”best” image in the biggest subgroup is taken as a centroid. It becomes hard to classify images belonging to the same group semantically but similar to other subgroups by their low-level content. The following solution can solve this problem.

For every group of similar images clusterisation should be performed based on inter-image distances in particular feature space. Then centroids for every cluster are to be built by using the same ”best image” strategy. In this case each group of similar images will be represented by several centroids: two per every cluster (for color and texture features). The disatvantage of this approach in increasing number of computations.

The solution described above can be also improved by running clusterisation with regional matches for color (spatial information is encoded into color feature which is used in this study) and particular filter matches for texture (we use ICA filters to build texture feature vectors).

4. Experiments and Result Analisys

4.1. Experiment Preparations

We use experimental image database from [7] which consists of 650 images from Corel Photo Set collection. It is divided into 9 groups based on images content by 2 experts. Result groups are: City, Clouds, Coasts, Contemporary buildings, Fields, Lakes, People, Rocks and Trees. This set of images can be considered as a training set for some larger collection.

For every image in the database color and texture features are extracted and for every pair of images color and texture distances are computed. Metrics values are normalized according to the following rule:

$$metrics_{result}(image) = (metrics(image) - Average)/Deviation.$$

Therefore color and texture metrics values distributions have the same $Average$ and $Deviation$.

We implement ”best image of the group” centroid building algorithm because it is the simplest one and still has acceptable performance as shown below. Color and texture centroids are calculated for every group.

4.2. Classification

Optimal mixed-metrics for search and optimal mixed-metrics for classification are shown in Table 1. It can be seen that they are different (while still very close in most cases). That is because these two classes of mixed-metrics serve for different tasks and their efficiency estimated in different ways.

Group’s optimal mixed-metrics for search task is the mixed-metrics which has the highest precision at N when searching with query-images from that group for all or most

Table 1. Mixed-metrics and classification results

Group	Optimal mixed-metrics for search task	Optimal mixed-metrics for classification task	Recall (%)	Precision (%)
City	$0.2 \cdot C + 0.8 \cdot T$	$0.3 \cdot C + 0.7 \cdot T$	6	19
Clouds	$0.5 \cdot C + 0.5 \cdot T$	$0.4 \cdot C + 0.6 \cdot T$	77	86
Coasts	$0.8 \cdot C + 0.2 \cdot T$	$1 \cdot C + 0 \cdot T$	70	44
Buildings	$0.2 \cdot C + 0.8 \cdot T$	$0.5 \cdot C + 0.5 \cdot T$	30	71
Fields	$1 \cdot C + 0 \cdot T$	$0.9 \cdot C + 0.1 \cdot T$	62	66
Lakes	$0.8 \cdot C + 0.2 \cdot T$	$1 \cdot C + 0 \cdot T$	55	56
People	$0.7 \cdot C + 0.3 \cdot T$	$0.6 \cdot C + 0.4 \cdot T$	82	35
Rocks	$1 \cdot C + 0 \cdot T$	$1 \cdot C + 0 \cdot T$	95	28
Trees	$0.2 \cdot C + 0.8 \cdot T$	$0.9 \cdot C + 0.1 \cdot T$	51	78

of N values. Precision at N is a rate of relevant objects among the first N retrieved. It means that when a search by using such mixed-metrics with a query-image from that group is performed, there is maximum number of relevant images among the first N retrieved (N varies from 1 to the size of the observed group) comparing to the searches by using other mixed-metrics.

Optimal mixed-metrics for classification task have different goal: maximum number of images from the training set should be classified to their real groups. Since the number of predefined groups is relatively small and the number of different mixed-metrics is small too, optimal mixed-metrics for classification can be found by using the following iterative algorithm. Optimal mixed-metrics for search task can be taken as a starting point.

Every image in the training set should be classified to one of the groups' centroids by using mixed-metrics for classification. Minimal distance is taken as a criterion. Rate of images classified to a group, which they really belongs to, is calculated for every group. On the next steps mixed-metrics for classification should be changed one by one in order to increase rate for every group (one mixed-metrics change can influence several groups' rates) and overall rate among all groups. After optimal mixed-metrics for classification task are found, they can be used in conjunction with groups' centroids for classification task.

Classification results by using mixed-metrics obtained on the previous step are shown in Table 1. While classification recall is quite large for most of the groups, it is achieved at the expense of low precision for some of them ("Coasts", "People", "Rocks"). There are two reasons for this situation. First is a "semantical gap" mentioned before and the second one is that centroids are obtained with the simple and rough "best image of the group" algorithm. Therefore they not necessarily represent their groups well as was mentioned in section 3. We plan to implement improvements of the "best image of the group" algorithm in order to increase precision.

Distribution of images among groups is shown in Table 2. The highest values are on the diagonal in most cases. At least half of images in the training set are classified right. The highest deviations can be seen for group "City". Most of its images are classified to "Coasts" and "People" groups due to low-level features similarity. Moreover it is not clear should "People" and "City" groups be semantically separated one from the other or should not.

Table 2. Distribution of images among groups: groups which images belong to are in rows, groups which images are classified to are in columns

	City	Clouds	Coasts	Buildings	Fields	Lakes	People	Rocks	Trees
City	3	0	23	5	3	5	14	0	1
Clouds	3	57	0	0	0	0	0	14	0
Coasts	2	4	57	3	3	3	1	8	0
Buildings	1	0	35	29	2	7	19	3	2
Fields	1	3	7	2	69	15	8	4	3
Lakes	1	2	9	1	13	60	0	19	4
People	3	0	0	1	1	0	23	0	0
Rocks	1	0	0	0	0	0	0	20	0
Trees	1	0	0	0	13	18	0	3	36

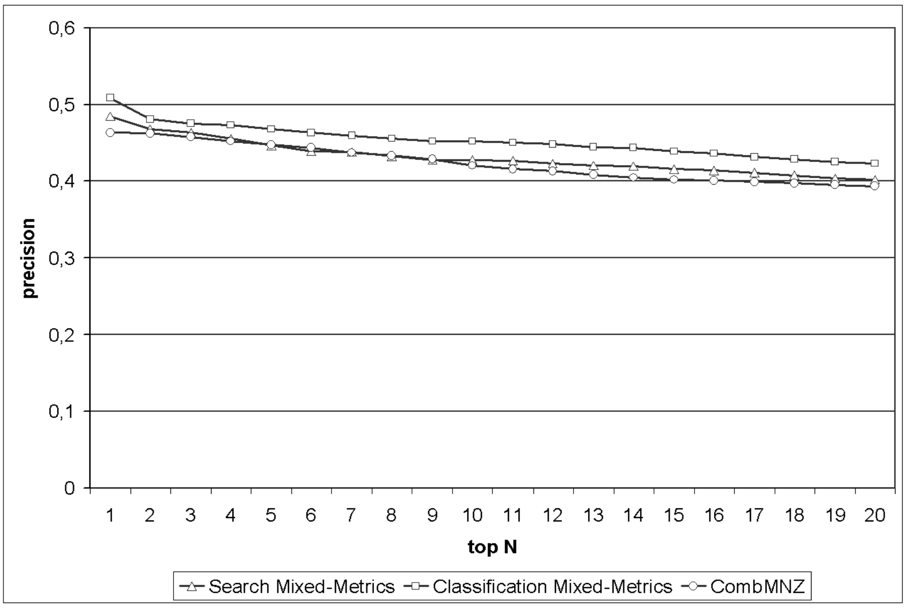

Figure 1. Precision/Top N dependencies for mixed-metrics-based searches and for CombMNZ algorithm

Similar situation is observed for groups ”Fields”, ”Lakes” and ”Trees”. They intersect a lot in their semantics and low-level features (containing a lot of green and blue colors, grass and leaves texture). Therefore images are usually misclassified between these groups.

4.3. Search

Our final goal is to perform a search with an arbitrary query-image by using appropriate mixed-metrics. The following experiment is held to evaluate precision of the search with proposed mixed-metrics approach.

For every image in the training set three different searches are performed: by using mixed-metrics of the group which the query-image belongs to, by using mixed-metrics of the group which the query-image is classified to and by using CombMNZ data fusion algorithm. Precision at N for every search is calculated (N = 1..20). Images from the same group as the query-image are treated as relevant, while others are not. Average precision at N is calculated for every group of similar images.

Result charts are shown on Figure 1. Retrieval precision for mixed-metrics of the group which the query-image is classified to lies between CombMNZ precision (lower bound) and retrieval precision for appropriate mixed-metrics (upper bound). For position 6 CombMNZ algorithm outperforms mixed-metrics approach. And for near position their results are similar. It can be seen that even for rough centroids some benefits can be achieved comparing to CombMNZ data fusion algorithm.

5. Conclusions and Future Work

In this study we addressed image classification problem applied to the mixed-metrics approach developed in [8]. We proposed the usage of centroids as representative feature vectors for every group of perceptually similar images. The advantage of this approach is that centroids can be calculated offline. During query execution process query-image should be compared to centroids only in order to perform a classification. It significantly reduces computational cost of query execution compared to the usage of common classification techniques for this task. A simple algorithm for building up centroids for groups of similar images is proposed. Possible improvements are also discussed.

Experiments show that even for simple and rough algorithm at least half of images in the training set is classified right and retrieval precision for mixed-metrics of the group which an image is classified to is higher than for CombMNZ data fusion algorithm in most cases. In future works we plan to implement proposed improvements and evaluate their performance.

References

[1] P.J. Deer, P.W. Eklund. On the Fusion of Image Features. citeseer.ist.psu.edu/162546.html.

[2] E.A. Fox, J.A. Shaw. Combination of multiple searches. In *TREC 2* (1994), 243–249.

[3] A. Guerin-Dugue, S. Ayache, C. Berrut. Image retrieval: a first step for a human centered approach. In *Proceedings of the Joint Conference of ICI, CSP and PRCM* **1** (2003), 21 – 25.

[4] O. van Kaick, G. Mori. Automatic Classification of Outdoor Images by Region Matching. In *Proceedings of the 3rd Canadian Conference on Computer and Robot Vision* (2006).

[5] J.H. Lee. Analyses of Multiple Evidence Combination. In *ACM-SIGIR*, USA (1997), 267–276.

[6] D. Lilis, F. Toolan, A. Mur, L. Peng, R. Colier, J. Dunnion. Probability-Based Fusion of Information Retrieval Result Sets. In *Proceedings of AICS Conference* (2005).

[7] I. Markov, N. Vassilieva. Image retrieval. Color and texture combining based on query-image. Submitted to the *ICISP conference*, France (2008).

[8] I. Markov, N. Vassilieva, A. Yaremchuk. Image retrieval. Optimal weights for color and texture features combining based on query object. In *Proceedings of RCDL*, Russia (2007), 195–200.

[9] Y. Rui, T.S. Huang, S.-F. Chang. Image Retrieval: Past, Present and Future. In *Proceedings of Int. Symposium on Multimedia Information Processing* (1997).

[10] N. Serrano, A. Savakis, J. Luo. A computationally efficient approach to indoor/outdoor scene classification. In *Proceedings of International Conference on Pattern Recognition* **4** (2002), 146–149.

[11] M. Szummer, R. W. Picard. Indoor-outdoor image classification. In *Proceedings of IEEE Int. Workshop on Content-Based Access of Image and Video Database* (1998), 42–51.

[12] A. Vailaya, M. A. T. Figueiredo, A. K. Jain, H. J Zhang. Image classification for content-based indexing. In *IEEE Transactions on Image Processing* **10**(1) (2001).

[13] A. Vailaya and A. Jain, H. Zhang. On Image Classification: City Images vs. Landscapes. In *Pattern Recognition* **31**(12) (1998) 1921–1935.

[14] N. Vassilieva, A. Dolnik, I. Markov. Image retrieval. Fusion of the result sets retrieved by using different image characteristics. In *"Internet-mathematics" Collection* (2007), 46–55.

[15] N. Vassilieva, B. Novikov. A Similarity Retrieval Algorithm for Natural Images. In *Proceedings of the Baltic DB&IS*, Latvia (2004).

[16] N. Vassilieva, B. Novikov. Establishing a correspondence between low-level features and semantics of fixed images (In Russian). In *Proceedings of RCDL*, Russia (2005), 236–240.