Image Categorization Using Texture Features

Aya Soffer *
Computer Science and EE Department
University of Maryland Baltimore County
1000 Hilltop Circle, Baltimore, MD 21250 and
CESDIS, Goddard Space Flight Center
E-mail: soffer@cs.umbc.edu

Abstract

A method for finding all images from the same category as a given query image (termed categorization*) using texture features is presented. The hypothesis that two images that are similar in texture are likely to belong to the same category, is examined. A new texture feature termed $N \times M$-gram is presented. It is based on the N-gram technique that is commonly used for text similarity. The process of computing an image profile in terms of its $N \times M$-grams is described. Results of experiments on images from various categories are presented. The $N \times M$-gram method with three different similarity measures is compared to the results of categorization using other well known texture features and grey level distribution features. The results show that for our test images texture features are suitable for image categorization, and $N \times M$-gram based methods are the best overall choice of texture feature for this task.*

1 Introduction

With the proliferation of systems that store electronic images, there has recently been much interest in similarity based image retrieval. The goal is to retrieve all images that are similar to a given query image. The most common form of similarity is based on color [6, 9]. Another form of similarity is based on texture [5]. Shape based similarity is also used in some systems [2, 8].

Testing whether a particular method for image similarity is useful has proven to be an elusive task. One of the difficulties is that in many cases it is not clear even to the evaluator which images are most similar to a given query image. In our work, we assume that the images in our database can be divided into well defined categories (e.g., floorplans, music notes, satellite images), and that the goal of retrieval is to find other images from the same category as the query image. With this goal in mind, it is easier to assess how well the system performs.

*The support of USRA/CESDIS and NASA Goddard Space Flight Center is gratefully acknowledged.

In this paper we explore whether images can be categorized based on texture features. We hypothesize that images from similar categories have similar texture features. In particular, we present a method for categorizing images using a new texture feature termed $N \times M$-*gram* that is based on the N-gram technique commonly used for determining similarity of text documents [1]. We define the notion of $N \times M$-grams and show how to compute an image profile in terms of its $N \times M$-grams . We propose three similarity measures for comparing images based on their $N \times M$-gram profile. We present results of experiments that compare categorization using $N \times M$-grams to categorization using other texture features. Our results show that for our test images texture features are suitable for image categorization, and $N \times M$-gram based methods are the best overall choice of texture feature for this task.

2 Generating $N \times M$-gram vectors

An N-*gram* is a sequence of N consecutive characters in a document. The word "image", for example, consists of three tri-grams "ima", "mag", and "age". The N-*gram frequency profile* of a document consists of a frequency count of each possible N-gram in the language. For example, for the English case-insensitive alphabet, the N-gram frequency profile of a document is the frequency of each of the 26^3 possible tri-grams in the document. A similarity measure between the N-gram profile of two documents determines the similarity of these two documents.

An $N \times M$-*gram* is a sequence of $N \times M$ pixel values. For example, the image whose pixel values are shown in Figure 1 consists of four 3×3-grams, "001011110", "011110100", "011110100", and "110100000". The $N \times M$-*gram frequency profile* of an image consists of a frequency count of each possible $N \times M$-gram. For example, for binary images the 3×3-gram frequency profile is the frequency of each of the $2^{3 \times 3} = 512$ possible 3×3-grams in the image. A similarity measure between the $N \times M$-gram vector of two images determines the sim-

0	0	1	1
0	1	1	0
1	1	0	0
1	0	0	0

Figure 1: Example image pixel values: the corresponding 3×3-grams are "001011110", "011110100", "011110100", and "110100000".

ilarity of these two images.

We first transform the images into a binary representation. Currently we perform this binarization using a global threshold operation on the input images. The $N \times M$-grams are calculated for the binary image by sliding a window of size $N \times M$ over the entire image. The pixel values of the N rows of M pixels in this window are concatenated. The number of occurrences of each $N \times M$-gram in the image is counted. The $N \times M$-gram frequency is derived by dividing by the the total number of $N \times M$-grams in the image. We store both the absolute count and the frequency of each $N \times M$-gram for each image. In addition, we compute the average frequency and average count of each $N \times M$-gram for the entire corpus of images.

3 Comparing $N \times M$-gram vectors

Text documents are commonly compared using the normalized dot product of their N-gram vectors. We use a similar measure for images. Let T denote the highest $N \times M$-gram value (e.g., 511 for binary images and 3×3-grams), and let $f_{j,I}$ denote the frequency of $N \times M$-gram j in image I. The similarity between two images using dot product, S_{dp}, is given in Equation 1.

$$S_{dp}(A,B) = \frac{\sum_{j=0}^{T} f_{j,A} * f_{j,B}}{\sqrt{\sum_{j=0}^{T} f_{j,A}^2 * \sum_{j=0}^{T} f_{j,B}^2}} \quad (1)$$

In work on text documents [1], it was observed that the similarity can be improved by subtracting the average N-gram vector for a corpus of documents from the N-gram vector of the query and database documents. We have adopted this idea for images. Let af_j denote the average frequency of $N \times M$-gram j over all images in the database, and let $df_{j,A}$ denote $(f_{j,A} - af_J)$ (i.e., the distance from the average). The modified similarity score, S_{dpa}, is given in Equation 2.

$$S_{dpa}(A,B) = \frac{\sum_{j=0}^{T} df_{j,A} * df_{j,B}}{\sqrt{\sum_{j=0}^{T} df_{j,A}^2 * \sum_{j=0}^{T} df_{j,B}^2}} \quad (2)$$

The second similarity measure that we use takes the bin to bin difference of the $N \times M$-gram frequency vector. The similarity between two images A and B using bin to bin difference S_{bd}, is given by Equation 3.

$$S_{bd}(A,B) = \sum_{j=0}^{T} |f_{j,A} - f_{j,B}| \quad (3)$$

The weight of common $N \times M$-grams is reduced by multiplying the absolute difference between the bins by $\frac{1}{af_j}$ (Equation 4).

$$S_{bda}(A,B) = \sum_{j=0}^{T} \frac{|f(j,A) - f(j,B)|}{af(j)} \quad (4)$$

The third similarity measure that we use gauges the similarity between images only in terms of the number of common $N \times M$-grams. The similarity measure is the intersection between the histogram [9] of $N \times M$-gram values. Let $t_{j,I}$ denote the total number of occurrences of $N \times M$-gram j in image I. The similarity between two images using this measure, S_{int}, is given by Equation 5.

$$S_{int}(A,B) = \frac{2 * \sum_{j=0}^{T} min(t_{j,A}, t_{j,B})}{\sum_{j=0}^{T} (t_{j,A} + t_{j,B})} \quad (5)$$

The weight of common $N \times M$-grams is reduced by adjusting by the average (at_j). The modified similarity measure, S_{inta}, is given in Equation 6.

$$S_{inta}(A,B) = \frac{2 * \sum_{j=0}^{T} \frac{min(t_{j,A}, t_{j,B})}{at_j}}{\sum_{j=0}^{T} (\frac{t_{j,A}}{at_j} + \frac{t_{j,B}}{at_j})} \quad (6)$$

4 Other texture and grey level distribution features

4.1 Grey level features

In this method we categorize the images based on the grey level values of the input images using two different similarity measures. The first, S_{grlh}, compares the histograms of the images using histogram intersection [9].

The second similarity score uses statistics on the grey level distribution. We compute five global features that describe the grey level distribution throughout the image. These features are mean, standard deviation, RMS, skew, and kurtosis [1]. Each image is thus represented by a five dimensional feature vector. The similarity of two images using this method, S_{grls}, is the weighted Euclidean distance between their corresponding feature vectors.

[1] For a full description of these and other features see `http://www.cs.umbc.edu/~soffer/tr96-187.ps.gz`

4.2 Local standard deviation

Local standard deviation calculates the standard deviation in a local window. The output of this process is an image, in which each pixel p is replaced by the standard deviation in the window of size $dx \times dy$ with p in its center. We use a 3×3 window and the same two similarity measures that are used to compare grey scale levels (i.e., histogram intersection and statistical distribution) giving measures S_{sdvh} and S_{sdvs}, respectively .

4.3 Grey level cooccurrence

The grey level cooccurrence is the two-dimensional spatial dependency of the grey levels for a fixed distance and/or angular spatial relationship. We compute the cooccurrence matrix in the horizontal and vertical directions with distance 1. We then compute eight features from the cooccurrence matrix (GLCM) of image I [3]. The similarity score between two images using grey level cooccurrence, S_{glcm}, is the weighted Euclidean distance between the 8-dimensional GLCM feature vector of the two images.

4.4 Laws texture features

This method extracts texture features from an image by convolving the original image with a spatial filter. The spatial filters are comprised of 5×5 kernels derived from center-weighted vectors defined by the LAWS texture measure [4]. We convolve the images with the L5E5, L5S5, L5R5, E5L5, E5E5, S5L5, W5W5, and R5L5 kernels, a mixture of horizontal, vertical, and diagonal sensitive kernels. We then compute the mean, variance, RMS, skew, and kurtosis, over the grey level values of the convolved images and the similarity between two images using Laws texture features, S_{laws}, is the weighted Euclidean distance between these 5-dimensional feature vectors.

5 Implementation and Experimentation

At this point we have tested our method on 200 images from 22 different categories (see Table 1).

We computed the 3×3-gram profile for each image. In addition, we computed the other texture features for each image, as described in Section 4. For each image I in our database we ran the system with I as the query image, in effect ranking all of the other images with respect to I. We repeated this process using 3×3-grams and the similarity measures described in Section 3 as well as the texture features along with the corresponding similarity measures as described in Section 4.

5.1 Evaluation procedure

Retrieval accuracy is evaluated using the *normalized recall* metric [7] (NR), which is applicable when the result of retrieval is a list of items sorted by similarity. NR measures how close to the top of the list of retrieved items the relevant items appear compared to the ideal retrieval in which all N relevant items appear in the top N positions. NR is computed using the following two measures:

- $IAVRR$ - ideal average rank of relevant items. Ideally all C_n images from the same category C should be ranked at the top. Thus, $IAVRR(C) = \frac{\sum_{n=1}^{C_n} n}{C_n}$ (since the ideal rankings are 1 – C_n).
- $AVRR$ - average rank of relevant items for a particular retrieval seeking images from the same category as a query image Q. $AVRR(Q) = \frac{\sum_{I \in cat(Q)} R(I)}{cat(Q)_n}$, where $R(I)$ denotes the ranking of image I in the result, *cat(Q)* is the category of Q, and $cat(Q)_n$ is the number of images in the category of Q.

The ratio of AVRR to IAVRR is a measure of the deviation from the ideal ranking. $\frac{AVRR}{IAVRR} = 1$ for a perfect retrieval, $\frac{AVRR}{IAVRR} > 1$ for all other cases. For each category C we compute the average normalized recall for C by taking the average of the results of $\frac{AVRR(Q)}{IAVRR(C)}$ for each image Q in category C. In addition, we compute the average normalized recall value over the whole database as the average of the NR of all categories, and the total normalized recall as $\frac{AVRR(Q)}{IAVRR(cat(Q))}$ for each image Q in the database.

5.2 Results

	dp	dpa	int	inta	bd	bda
fingerprints	1.28	1.79	1.66	1.73	2.15	2.17
floorplans	4.41	2.54	2.19	1.16	2.97	3.47
comics	3.44	4.51	2.47	1.69	3.31	3.67
music	3.51	2.30	3.31	1.07	3.40	3.05
tables	1.63	1.42	1.30	1.26	1.32	1.26
mug shots	4.14	1.18	2.38	1.88	1.84	1.84
houses	9.51	11.9	1.80	2.02	11.3	11.9
aircraft	9.65	10.5	11.7	9.53	10.5	11.6
animals	1.07	1.52	1.90	2.30	1.10	1.07
topo maps	2.01	2.68	1.54	1.92	3.35	2.99
road maps	6.11	9.62	7.17	5.85	8.65	8.94
satellite	3.24	4.03	4.40	3.60	6.03	5.51
aerial	5.23	3.88	4.28	5.12	11.4	10.9
Arabic	8.30	8.65	3.68	2.64	7.41	5.76
Chinese	1.00	1.00	4.55	1.07	1.00	1.00
Eng. print	7.14	5.30	6.27	4.50	8.35	8.86
Eng. lyrics	1.00	1.00	3.05	1.00	1.00	1.00
Eng. hand	4.48	3.27	2.76	1.80	4.05	3.90
Heb. script	1.33	2.50	1.17	1.11	1.22	1.11
Heb. hand	13.6	10.7	7.40	4.42	10.9	12.5
Heb. print1	1.00	1.00	1.39	1.00	1.00	1.00
Heb. print2	3.19	4.13	2.68	2.15	5.44	5.20
total	3.68	3.47	2.83	2.14	3.83	3.86
average	4.23	4.34	3.48	2.60	4.73	4.95

Table 1: Normalized recall using 3×3-grams (normalized recall= 1 for perfect retrieval, > 1 otherwise).

	grls	grlh	sdvs	sdvh	glcm	laws
fingerprints	3.35	1.14	3.46	1.08	1.55	3.00
floorplans	3.14	4.70	4.11	3.03	2.84	4.48
comics	4.57	2.68	2.01	1.85	3.08	1.75
music	15.5	16.3	15.8	12.5	17.5	15.9
tables	8.31	1.80	4.34	1.40	5.53	2.80
mug shots	9.39	9.60	4.41	2.67	7.02	5.17
houses	7.97	16.74	3.44	5.52	7.45	5.76
aircraft	9.44	5.53	4.11	3.45	3.70	4.58
animals	2.23	1.00	1.43	1.00	1.05	1.02
topo maps	3.10	1.23	4.88	2.21	3.87	5.67
road maps	3.27	9.83	2.50	7.81	7.89	3.43
satellite	4.92	11.88	4.84	4.08	5.04	6.77
aerial	5.71	20.00	4.77	4.66	8.66	8.60
Arabic	11.1	6.6	15.5	8.1	12.4	15.5
Chinese	6.00	5.96	4.63	3.87	2.45	3.23
Eng. print	4.31	4.86	3.20	4.37	3.40	4.57
Eng. lyrics	8.82	4.95	2.20	1.85	2.15	11.7
Eng. hand	3.15	8.03	4.34	7.56	7.14	4.23
Heb. script	1.11	1.00	1.17	1.00	1.00	1.94
Heb. hand	5.22	7.62	2.25	5.83	2.98	3.58
Heb. print1	1.44	1.00	1.00	1.00	1.00	1.00
Heb. print2	7.47	3.64	6.77	2.77	2.17	5.65
total	4.65	5.25	4.06	3.46	4.48	4.40
average	5.70	6.40	4.44	3.85	4.82	5.28

Table 2: Normalized recall using grey level features, local standard deviation, cooccurrence matrix, and Laws features (normalized recall$= 1$ for perfect retrieval, > 1 otherwise) .

Tables 1 and 2 summarize the normalized recall (NR) results for each image category using various similarity criteria. The total and average normalized recall over the whole database are also presented. Figures 2 and 3 show the first 16 ranked images for the aircraft and musical notes categories. Each figure has three parts (a), (b), and (c) reporting the results using 3×3-grams with *inta*, *sdvh*, and *grls*, respectively. The first image is the query image. The names of all images that are from the same category as the query image are displayed in reverse video.

From these results it is apparent that overall the $N \times M$-gram based methods perform better than the other methods. In addition, between the similarity measures for $N \times M$-gram vectors, *inta* is best with an average NR of 2.6 and a total NR of 2.14. Among the other methods *sdvh* did best (average NR of 3.85 and total NR of 3.46), while *grlh* did worst (average NR of 6.4 and total NR of 5.25). Note, however, that this performance was not observed for all categories. An example of the superiority of *sdvh* in the aircraft category can be seen in Figure 2 where 7 out of the 8 aircraft images in our database are among the top 16 using *sdvh*, as opposed to 3 for *grls* and *inta*. Figure 3 demonstrates the superiority of *inta* in the music category. The *inta* method ranked the 9 musical scores in our database perfectly, while *sdvh* found 6, and *grls* only found 3.

6 Concluding Remarks

As can be seen from the results, $N \times M$-grams work very well on simple documents such as floorplans, music notes, and comics. The $N \times M$-gram method was however weak on richer images such as aircraft, and aerial photos. Numerous modifications are anticipated. The $N \times M$-grams are computed on the images following a threshold operation on the original greyscale image. Better results may be achieved by computing these on the gradient image or on the images following edge detection. Furthermore, the choice of 3×3-grams may be too small. Using a larger window may yield better results, but will increase the storage requirements. We intend to use multiple window sizes by applying the same size window to various resolutions of the image using the wavelet transform.

References

[1] M. Damashek. Gauging similarity with n-Grams: Language-independent categorization of text. *Science*, 267:843–848, February 1995.

[2] W. I. Grosky and R. Mehrotra. Index-based object recognition in pictorial data management. *Computer Vision, Graphics, and Image Processing*, 52(3):416–436, December 1990.

[3] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Trans. on Systems, Man and Cybernetics*, 610–621, 1973.

[4] K. I. Laws. Rapid texture identification. *Proc. of the SPIE*, volume 238, pages 376–380, 1980.

[5] B. S. Manjunath and W. Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(8):837–842, August 1996.

[6] W. Niblack, et. al. The QBIC project: Querying images by content using color, texture, and shape. In *Storage and Retrieval of Image and Video Databases*, volume 1908, pages 173–187, San Jose, 1993.

[7] Gerard Salton and Michael J. McGill. *Introduction to Modern Information Retrieval.* McGraw-Hill, New York, 1st edition, 1983.

[8] H. Samet and A. Soffer. MARCO: MAp Retrieval by COntent. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(8):783–798, August 1996.

[9] M. Swain. Interactive indexing into image databases. In *Storage and Retrieval for Image and Video Databases*, volume 1908, pages 95–103, San Jose, 1993.

Figure 2: Results for image from the aircraft category (a) 3×3-gram with vector intersection score normalized by averages (*inta*). (b) local standard deviation with histogram intersection (*sdvh*). (c) grey levels with distribution statistics features (*grls*).

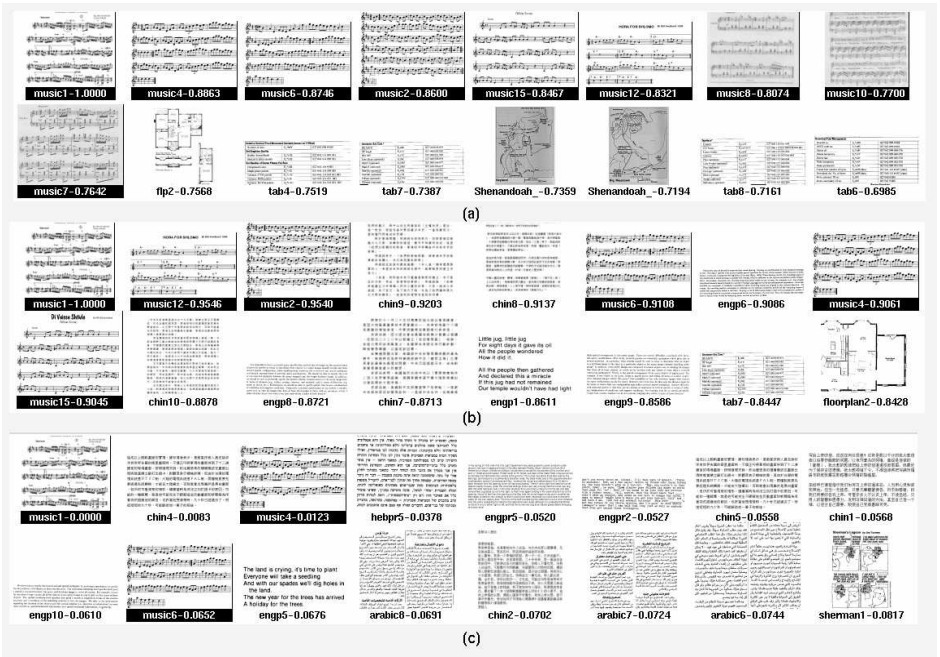

Figure 3: Results for image from the music category (a) 3×3-gram with vector intersection score normalized by averages (*inta*). (b) local standard deviation with histogram intersection (*sdvh*). (c) grey levels with distribution statistics features (*grls*).