Content-based Image Retrieval Using Shape-Size Pattern Spectra

Florence Tushabe and Michael. H. F. Wilkinson

University of Groningen, The Netherlands

(m.h.f.wilkinson, florence)@cs.rug.nl

Abstract

This paper presents the results of using the shape, size and color features of an image for content-based image retrieval. The use of granulometries has been applied to model the size and shape of the connected components of the image. Granulometry are computed by successively sieving an image using filters of increasing size parameter so that information can be obtained about the components that filter through. A shape-size granulometry using sieves of increasing shape and sizeparameters represents the shape and size distributions of an image. The results of granulometry are stored in a 2-D pattern spectrum that is implemented using attribute thinnings and openings. Additionally, the pattern spectra is extracted from the red, green and blue color bands of the images, concatenated and compared using the L1 Norm and the Jensen-Shannon divergence (JSD). Our results show that incorporating color information improves the performance by 23% and L1 Norm outperforms JSD by 20%.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval; H.3.4 Systems and Software; H.3.7 Digital Libraries; I.5 [**Pattern Recognition**]: I.5.3 Clustering—*Similarity Measures*; G.2 [**Discrete Mathematics**]: G.2.2 Graph Theory

General Terms

Measurement, Performance, Experimentation

Keywords

Max-tree, pattern spectrum, shape-size pattern spectra, granulometry, attribute filtering, shape representation, IAPR TC-12

1 Introduction

The most recent advances in image retrieval show that text-based methods still outperform content-based ones by approximately 30% [3]. Yet the question: " What sentence can describe an image" is still being researched by linguists, computer scientists and other academicians. This is because different people will describe an image quite differently, depending on their interests and prior knowledge. Some users look at an image in terms of the shape of its components, others the size, color, texture, orientation or spatial location etcetera . In theory, using the features of an image as its signature presents the users with a precise and focussed way of describing an image according to their interests. User satisfaction is enhanced when one is presented with a variety of ways to define his need. This is because the retrieval engine is customised according to the selected choice / requirement. This study retrieves images based upon the shape of its components

by using mathematical morphology techniques. Attribute filtering is used to identify subsets of the image that satisfy the requirements of a particular property or attribute. Forexample, attributes like moments and length (or width / volume) are useful for spatial and size information respectively [10]. The selected components are then used for further analysis like the formation of a pattern spectrum. Pattern spectrum has been used to obtain information about the contents and distribution of an image including it's spatial location [10, 9, 5]. We present the results of using pattern spectra for content-based image retrieval in large-scale databases. It is an extension of the method proposed by Urbach et al [8] but combined with color information and applied within the image retrieval context. The objective was to identify the 1000 most similar images from the 20,000 IAPR TC-12 database by using a purely visual, fully automated modality [2].

Section 2 briefly describes the theory of our method, the experiments and results are contained in Section 3 and 4 and Section 5 has the concluding remarks and on-going work.

2   2-D Shape-Size Pattern Spectra

Pattern spectrum is a quantitative description of the contents of an image after the application of a granulometric operation [5]. Granulometry is when an image is successively sieved using filters of increasing diameters, r, so that information can be obtained about its components that satisfy the properties of r. A size granulometry obtains the size distribution of an image by using sieves of increasing sizes herein represented by area. Let X , Y represent an image then the size granulometry (Γ_r) is a set of filters $\{\Gamma_r\}$ with r from some totally ordered set Λ (usually $\Lambda \subset \mathbb{R}$ or $\mathbb{Z}$) satisfying the properties:

$$\Gamma_r(X) \subseteq X \tag{1}$$

$$X \subseteq Y \Rightarrow \Gamma_r(X) \subseteq \Gamma_r(Y) \tag{2}$$

$$\Gamma_s(\Gamma_r(X)) = \Gamma_{max(r,s)}(X), \tag{3}$$

for all $r, s \in \Lambda$. Equations (1),(2) and (3) define Γ_r as being anti-extensive, increasing and idempotent respectively. Attribute openings are size granulometries since they share the same properties [1, 10]. The pattern spectrum, $s_\Gamma(X)$, obtained by applying the size granulometry, $\{\Gamma\tau\}$, to a binary image X is defined by [5] as:

$$(s_\Gamma(X))(u) = - \left. \frac{dA(\Gamma_r(X))}{dr} \right|_{r=u} \tag{4}$$

where $A(X)$ is the area of X.

Similarly, shape granulometry filters image contents using sieves of different shapes [8, 9]. Urbach and Wilkinson [9] defined shape granulometry, of X, as a family of filters, $\{\Phi_r\}$, with shape parameter, r, from some totally ordered set Λ (usually $\Lambda \subset \mathbb{R}$ or $\mathbb{Z}$) with the following properties:

$$\Phi_r(X) \subseteq X \tag{5}$$

$$\Phi_r(tX) = t(\Phi_r(X)), \tag{6}$$

in which tC stands for a scaling of set C by a factor t, and,

$$\Phi_s(\Phi_r(X)) = \Phi_{max(r,s)}(X), \tag{7}$$

for all $r, s \in \Lambda$ and $t > 0$. Equations (5),(6) and (7) define Φ_r as anti-extensive, scale invariant and idempotent respectively. Therefore, attribute thinnings are shape granulometries since they share the same properties including the non-increasing criterion[8]. The shape pattern spectrum, $s_\Phi(X)$, is obtained by applying the shape granulometry, $\{\Phi\tau\}$, to a binary image X and defined as [8]:

$$(s_\Phi(X))(u) = - \left. \frac{dA(\Phi_r(X))}{dr} \right|_{r=u} \tag{8}$$
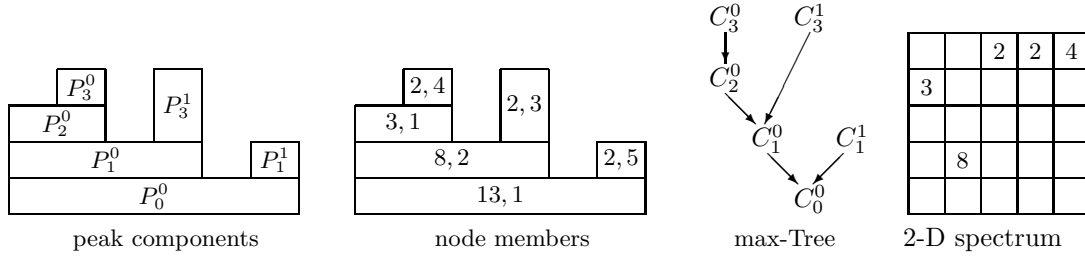
Figure 1: Peak components(P_h^k), corresponding (C_h^k), the resultant max-tree and the corresponding pattern spectrum (right)

as above. In our case we use a shape granulometry based on the ratio $I(C)/A^2(C)$, with $I(C)$ the moment of inertia of C. This shape parameter is equivalent to Hu's first moment invariant [8]. The final feature vector resulting from the combination of size and shape granulometries formed the 2-D shape-size pattern spectrum as a shape descriptor that we used. Further details about this are found in [8].

2.0.1 Computing the shape-size pattern spectra

The same method as was used in [8] has been adopted for application within CBIR. The max-tree approach [6, 7] was used to implement the attribute thinnings and openings in grey-scale. The subsets (connected components) of the image are arranged into a tree structure and filtered by removing the nodes which do not satisfy a given criterion. A max tree is a rooted tree in which each of its nodes, C_h^k, at gray-level h corresponds to a peak component, P_h^k [6]. An example is shown in Figure 1 which illustrates the peak components, P_h^k, of a 1-D signal, the corresponding C_h^k at levels $h = 0, 1, 2, 3$, the resultant max-tree and corresponding spectrum.

Let $\{\Gamma_r\}$ be a size distribution with r from some finite set Λ_r and $\{\Phi_s\}$ a shape distribution with s from some index set Λ_s. If S is the 2-D array that stores the final 2-D spectrum, then each cell, $S(r, s)$, contains the sum of gray levels of C_h^k that falls within size class $r-$ and r and shape class $s-$ and s. The 2-D pattern spectrum is then computed from the max-tree as follows:

- Set all elements of the array S to zero.

- Compute the max-tree according to the algorithm in [7].

- As the max-tree is built, compute the area $A(P_h^k)$ and moment of inertia $I(P_h^k)$ of each node.

- For each node C_h^k:

    - Compute the size class r from the area $A(P_h^k)$;
    - Compute the shape class s from $I(P_h^k)$ / $A^2(P_h^k)$;
    - Compute the gray level difference δ_h, between the current node and its parent;
    - Add the product of δ_h and $A(P_h^k)$ to $S(r, s)$.

3 Experiments

The objective of our experiments was: given a sample image, find as many relevant images as possible from the IAPR TC-12 photographic collection [3]. The procedure that was undertaken is as follows:
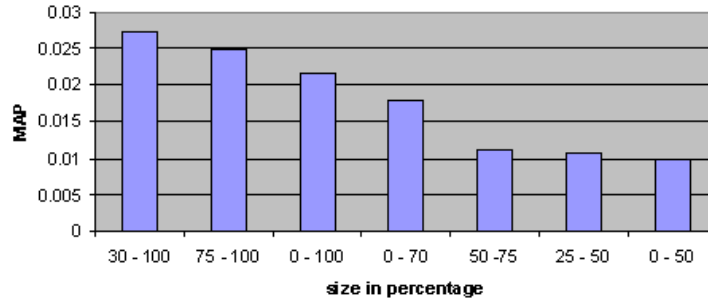
Figure 2: Performance of component sizes

1. Images are first pre-processed by filtering off all components whose area is less than 30% of the total image size. This was after comparisons with other size ranges indicated that the discriminative power lies in the large particles. Figure 2 illustrates detailed MAP results after using different class sizes. It is observed that the smaller particles performed poorer than the larger ones and that using all particles without any filtering performs worse than using particles over 30% or 75% of total area.

2. The shape-size pattern spectrum is then extracted from all images including the query images. The size and shape ranges were between $50,000 - 172,800$ and $1 - 53$ respectively. A 20 by 15 bin histogram which translates into a 1×600 array representation per image was chosen. Additionally, conversion of the images into XYZ and YUV color spaces as well as grayscale (PGM) was performed. This was to test whether additional information is captured from the color matrices. Visual analysis showed that unlike YUV and XYZ which performed worse than the grayscale images, RGB representation improved the results.

3. One of the three topic images was selected as the query image according to what the authors thought provided the most meaningful retrievals. The list of the chosen topic images is provided in the appendix.

4. The similarity measure used is the reciprocal of the L1 Norm.

$$Sim(X,Y) = \frac{1}{\sum (x_i - y_i)} \tag{9}$$

3.1   Submitted Runs

Four runs were submitted:

1. PGM2007 - the shape-size pattern spectrum of the grayscale images are compared.

2. Max2007 - the maximum of red, green and blue bands of the images are compared.

3. Concatenate2007 - the concatenated red, green and blue bands of the images are compared (a 1×1800 array).

4. JSD2007 - same as (3) above but comparison with the Jensen-Shannon divergence [4].

Table 1: performance of submitted runs

Run	MAP	P20	Relevant	% improvement
Concatenate	0.0337	0.1050	724	23
Max	0.0288	0.0933	699	5.5
JSD	0.0281	0.0925	701	2.9
PGM	0.0273	0.0925	693	-

4 Results

We observed that most queries returned images based on their shapes. For example an image of a hill returns images of hills / mountains. So is the case with images of people, beaches, skies, buildings etcetera. An image with more buildings than people returns more results of buildings (some of them without people) than people. A group of people standing in front of a mountain landscape returns a group of people in front of buildings, mountains or in a city square or anywhere else. But if the mountains cover a larger percentage of the picture, then the returned results are more of mountains than people. Similarly, a duck swimming in water returns many images of birds flying in the sky and hens eating on a road. This is because the shapes of ducks, birds and hens are similar. The texture of the image also affects the performance of this method. For example, the image of a bird flying in a clear sky returns many images of birds or aeroplanes flying in the sky. But the one with a bird flying in a cloudy sky returns pictures of beds, clouds without birds, mountains covered in snow and some beach pictures. One of the poorest performances was the meat dishes which returned buildings, landscapes, beds, and people with no food and without food.

The overall performance of the submitted runs is shown in Table 1. The RGB concatenated version using L1 Norm produces the best results and improves performance by 23%. As expected, all methods that incorporate color information perform better than the one without (PGM). This shows that more information has been obtained from the color matrices and so images with similar colors are grouped closer together. For example, images taken at night return more black pictures just as the ones taken in deserts return more brown ones. Interestingly, the images with two dominant colors seemed to outperform those with only one dominant color or many colors. Table 1 also gives surprising results that show how the maximum of the RGB bands using L1 norm outperformed the concatenated version using JSD. This could be due to the loss of information during normalization. This clearly shows that the use of an appropriate similarity measure significantly contributes to the performance of our system.

5 Conclusion

This paper uses the shape-size pattern spectrum as the fundamental image signature for content-based image retrieval. The pattern spectrum is obtained from the red, green and blue color bands of the image and compared using the L1-Norm. Our experiments have shown that this results in very logical and interesting performance although the MAP scores remain low. We believe that one of the ways of improving the MAP scores is to represent a single topic by the combination of the three provided sample images. This can be achieved by using relevance learning techniques already in place. Secondly, the results have shown that textural information, especially within the background, affects the performance of this system. We intend to incorporate both textural and spatial information into the pattern spectrum. Other works in progress include identifying the most appropriate shape parameters, attributes and similarity measures.

References

[1] E.J. Breens and R. Jones. Attribute openings, thinnings and granulometries. *Computer Vision Image Understanding*, 64(3):377–389, 1996.

[2] M. Grubinger, P. Clough, and L. Clement. The iapr tc-12 benchmark for visual information search. *IAPR Newsletter*, 28(2):10–12, 2006.

[3] M. Grubinger, C. Paul, A. Hanbury, and M. Henning. Overview of the ImageCLEF 2007 photographic retrieval task. In *Working Notes of the 2007 CLEF Workshop*, Budapest, Hungary, Sep 2007.

[4] J Lin. Divergence measures based on the shannon entropy. *IEEE Transactions on Information Theory*, 37(1):145–151, 1991.

[5] P Maragos. Pattern spectrum and multiscale shape representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11(7):701–715, 1989.

[6] A Meijster and M. H. F. Wilkinson. A comparison of algorithms for connected set openings and closings. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(4):484–494, 2002.

[7] P Salembier, A Oliveras, and L Garrido. Antiextensive connected operators for image and sequence processing. *IEEE Trans. Image Proc.*, 7(4):555–570, 1998.

[8] E R. Urbach, J. B. T. M. Roerdink, and M. H. F. Wilkinson. Connected shape-size pattern spectra for rotation and scale-invariant classification of gray-scale images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(2):272–285, 2007.

[9] E. R. Urbach and M. H. F. Wilkinson. Shape-only granulometries and grey-scale shape filters. In *Proc. Int. Symp. Math. Morphology (ISMM) 2002*, Sydney, Australia, April 2002.

[10] M.H.F. Wilkinson. Generalized pattern spectra sensitive to spatial information. In *Proceeding of the 16th International Conference on Pattern Recognition*, volume 1, pages 701–715, Quebec City, Canada, August 2002.

Appendix - Selected topic images and corresponding MAP

Topic	Image	MAP	Topics	Image	MAP
1	topics/01/3793	0.0019	31	topics/31/3616	0.001
2	topics/02/16432	0	32	topics/32/14973	0.0016
3	topics/03/31	0.028	33	topics/33/32317	0.0703
4	topics/04/23221	0.0012	34	topics/34/11601	0.0189
5	topics/05/30823	0.1956	35	topics/35/25004	0.1554
6	topics/06/37754	0.0655	36	topics/36/13033	0.0192
7	topics/07/15515	0.0455	37	topics/37/16336	0.0027
8	topics/08/1541	0.0134	38	topics/38/18059	0.0288
9	topics/09/1082	0	39	topics/39/14560	0.0002
10	topics/10/4072	0.0089	40	topics/40/11645	0.0037
11	topics/11/37367	0.0094	41	topics/41/2721	0.0224
12	topics/12/31646	0.069	42	topics/42/39273	0.0033
13	topics/13/26719	0.0421	43	topics/43/23909	0.0364
14	topics/14/31609	0.0665	44	topics/44/31871	0.0055
15	topics/15/2125	0.2106	45	topics/45/12477	0.0001
16	topics/16/38220	0.0185	46	topics/46/31312	0.0085
17	topics/17/33458	0.0113	47	topics/47/35974	0.001
18	topics/18/38195	0.0184	48	topics/48/35705	0.0491
19	topics/19/35791	0.009	49	topics/49/40214	0.0594
20	topics/20/38894	0.0022	50	topics/50/7301	0.003
21	topics/21/26437	0.0003	51	topics/51/23792	0.0058
22	topics/22/31715	0.0694	52	topics/52/32350	0.0029
23	topics/23/37505	0.0128	53	topics/53/2174	0.0601
24	topics/24/34159	0.0061	54	topics/54/35879	0.1016
25	topics/25/37874	0.0535	55	topics/55/14097	0.2525
26	topics/26/17051	0.0019	56	topics/56/4614	0.0362
27	topics/27/32195	0.0599	57	topics/57/32629	0.0036
28	topics/28/16432	0	58	topics/58/31189	0.003
29	topics/29/32759	0.0065	59	topics/59/17327	0.0133
30	topics/30/17057	0.0016	60	topics/60/27037	0.0007