Content-Based Image Retrieval Using Combined 2D Attribute Pattern Spectra

Tushabe, Florence; Wilkinson, Michael. H.F.

Published in:
ADVANCES IN MULTILINGUAL AND MULTIMODAL INFORMATION RETRIEVAL

Content-Based Image Retrieval Using Combined 2D Attribute Pattern Spectra

Florence Tushabe and Michael. H.F. Wilkinson

Institute of Mathematics and Computing Science, University of Groningen,
P. O. Box 407, 9700 AK Groningen, The Netherlands
m.h.f.wilkinson@rug.nl, florence@cs.rug.nl

Abstract. This work proposes a region-based shape signature that uses a combination of three different types of pattern spectra. The proposed method is inspired by the connected shape filter proposed by Urbach et al. We extract pattern spectra from the red, green and blue color bands of an image then incorporate machine learning techniques for application in photographic image retrieval. Our experiments show that the combined pattern spectrum gives an improvement of approximately 30% in terms of mean average precision and precision at 20 with respect to Urbach et al's method.

1 Introduction

The most popular content-based image retrieval descriptors follow the standard MPEG-7 visual tool-set [1]. They include descriptors based on color, texture, shape, motion and localisation. We test an alternative method of obtaining the image descriptor by the application of granulometric operations and machine learning techniques. Granulometric operations are applied to the image at different scales and levels of complexity to derive information about the distribution of its contents[2]. Attribute filtering is a relatively new and efficient way of implementing granulometry. Desired descriptors like size, spatial location or shape can be well represented with appropriate attributes like area [3], moments [4,5] or shape [6]. A size granulometry for example uses sieves of increasing sizes to obtain the size distribution of the image. Previous works like [7,8] use a structuring element approach for the granulometric operations. However, recent studies have found connected filtering to be faster and equal or sometimes better in performance than the SE approach [6]. In [6], a shape filter of a 2-D pattern spectrum consisting of an area and non-compactness spectrum is proposed. We extend the shape spectrum proposed in [6] and apply it to a photographic data set containing everyday vacation pictures [9]. This is because most shape-based image retrieval studies concentrate on artificial images or highly specialised domain-specific image data sets. The proposed shape spectrum consists of three rotation and scale invariant spectra: the area–non-compactness [10], area–compactness and area–entropy pattern spectra. They are weighted, combined and used for image retrieval within large-scale databases. The rest of the paper is organised

C. Peters et al. (Eds.): CLEF 2007, LNCS 5152, pp. 554–561, 2008.

as follows: Section 2 briefly describes the theory of the method employed, Section 3 contains the experimental set-up and Sections 4 and 5 give the results, discussions and concluding remarks.

2 Theory

Connected attribute filtering decomposes an image into sets of connected components. Each component adopts a single attribute value, r, and is considered for further processing only when r satisfies a given criterion. Attribute filtering is manifested through attribute openings or thinnings and is extensively discussed in [11]. Let C, D be connected components of set X and Ψ a binary image operator. Attribute openings are characterized by being increasing ($C \subseteq D \Rightarrow \Psi(C) \subseteq \Psi(D)$), idempotent ($\Psi\Psi(C) = \Psi(C)$) and anti-extensive ($\Psi(C) \subseteq C$). Example attributes include area, perimeter and moment-of-inertia. On the other hand, attribute thinnings are characterised by being idempotent, anti-extensive and non-increasing ($C \subseteq D \not\Rightarrow \Psi(C) \subseteq \Psi(D)$. Example attributes are length, compactness, non-compactness, circularity and entropy.

If 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:

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

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

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

$\forall r, s \in \Lambda$

Breen and Jones [11] show that attribute openings indeed provide size granulometries since equations (1),(2) and (3) define Γ_r as being anti-extensive, increasing and idempotent respectively. Similarly, Urbach and Wilkinson [10] show that a shape granulometry can be obtained from attribute thinnings. The shape granulometry, of X, is 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:

$$\Phi_r(X) \subseteq X \tag{4}$$

$$\Phi_r(tX) = t(\Phi_r(X)) \tag{5}$$

$$\Phi_s(\Phi_r(X)) = \Phi_{\max(r,s)}(X) \tag{6}$$

$\forall r, s \in \Lambda$ and $t > 0$

Equations (5),(6) and (7) define Φ_r as anti-extensive, scale invariant and idempotent respectively.

2.1 2-D Pattern Spectra

The results of the application of granulometry to an image can be stored in a pattern spectrum [3]. A 2D pattern spectrum represents the results of two granulometric operations in a single 2-dimensional histogram. The shape filter proposed in this work consists of a size-shape pattern spectrum.

The size pattern spectrum, $s_\Gamma(X)$, obtained by applying the size granulometry, $\{\Gamma\tau\}$, to a binary image X is defined by [3] as:

$$(s_\Gamma(X))(u) = -\left.\frac{dA(\Gamma_r(X))}{dr}\right|_{r=u} \tag{7}$$

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

While the shape pattern spectrum, $s_\Phi(X)$, is obtained by applying the shape granulometry, $\{\Phi\tau\}$, to a binary image X and defined by [6] as:

$$(s_\Phi(X))(u) = -\left.\frac{dA(\Phi_r(X))}{dr}\right|_{r=u} \tag{8}$$

where the difference with (7) is in the use of the shape granulometry.

2.2 Computing the Pattern Spectra

The max-tree approach [12,13] was used to implement the attribute thinnings and openings. Let the peak components, P_h^k of an image represent the connected components of the threshold set at gray level h with k from some arbitrary index set. These peak components are arranged into a tree structure and filtered by removing nodes whose attribute values are less than a pre-defined threshold, T. Thus, the 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 [13]. 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. Note that two attributes are shown per node, the first of which is the size attribute which increases as the tree is descended. The second attribute, which is the shape attribute is not increasing.

The method of generating the 2D spectrum has been adopted from Urbach et al [6]. 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 [12].
- As the max-tree is built, compute the area $A(P_h^k)$, perimeter $P(P_h^k)$, histogram of the gray levels and moment of inertia $I(P_h^k)$ of each node.
- For each node C_h^k:

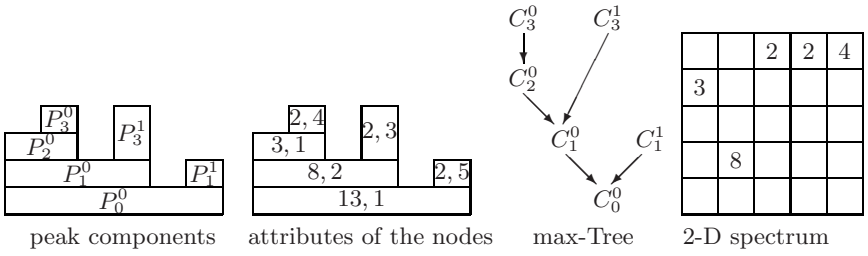

Fig. 1. Peak components(P_h^k), their attributes, corresponding (C_h^k) (the max-tree) and the resulting pattern spectrum (right)

- Compute the size class r from the area of P_h^k.
- Compute the shape class s from the shape attribute of 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 $S(r, s)$.

The shape attributes chosen are non-compactness, N, defined as

$$N = \frac{I(P_h^k)}{A^2(P_h^k)}, \tag{9}$$

compactness, C, defined as

$$C = \frac{P^2(P_h^k)}{A(P_h^k)}, \tag{10}$$

and finally, Shannon entropy

$$H = -\sum p(i) \log_2 p(i), \tag{11}$$

with $p(i)$ the probability with which gray level i occurs in P_h^k.

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 [14]. Our method uses the three query images that were provided per topic. The detailed methodology is as follows:

1. Separate the jpeg image into three different images, each representing its red, green and blue color bands. This is after initial analysis shows that RGB representation improves results unlike YUV and XYZ which performed worse than not separating the images.
2. Extract the desired pattern spectra from all the images including the query images. A 20 by 15 bin histogram that eventually translates into a 1×600 array representation was chosen. When concatenated, the spectra retrieved from the three color bands forms a 1×1800 vector per spectrum type. The three spectra that were tested are:

(a) Area and Non-Compactness (A-N) spectrum
- Area: Is a size filter that represents the number of pixels in the component. Initial experiments in [15] showed that the discriminative power lies more in the larger particles rather than the smaller ones. Therefore all particles less than 30% of the total image size were ignored.
- Non-Compactness: Thresholds of 1 - 53 were used for the non-compactness attribute since it gave the best MAP when compared with other thresholds ranging between $T = 1 : 100$.

(b) Area and Compactness (A-C) spectrum
- Area: Same thresholds as above.
- Compactness: The thresholds chosen for compactness are $T = 600$ since it registered the highest MAP when compared with other thresholds ranging between $T = 1 : 1000$.

(c) Area-Entropy (A-E) spectrum
- Area: Same thresholds as above.
- Entropy: A threshold of $T = 8$ was chosen because it is the maximum entropy that any component can achieve.

3. The spectra were separated into two equal parts, A and B, referring to larger and smaller features in the images.
4. The baseline distance, $d_{x,j}$, of any two images x and j is given by:

$$d_{x,j} = w_a d_{A(x,j)} + w_b d_{B(x,j)} \tag{12}$$

where $w_{a,b}$ are the weights of parts A and B of the spectrum and $d_\alpha(x, j)$ the L1 norm distance of image x and j as computed from attribute α of the spectrum. The weights chosen for area - non-compactness is $w_a = 0.7$ and $w_b = 0.3$; area - compactness is $w_a = 0.7$ and $w_b = 0.3$; and area - entropy attributes $w_a = 0.5$ and $w_b = 0.5$. These weights were found by trial and error.
5. The 250 most significant features from the 1×1800 spectra are selected and used to train the query images using the naive bayesian classifier from [17,16]. The images are then classified by each of the spectra into classes consisting of the 60 topics. The distance, d_x, of an image from a particular topic is reduced by a given percentage, p if it has been classified within that topic. This is done because we wish to obtain a single distance, and the bayesian classifier from [17] works with a different distance measure than d_x. Parameter p is the classification weight and is 20% for for A-N and A-E and 70% for A-C feature sets respectively. These percentages were also determined empirically.
6. The distance, D_x of X from topic T is the minimum of its distances from the three topic images. The final distance of image x from topic image y is the weighted addition of its distances from the three spectra.

$$D_x = \min_{j \in T} \{0.75 d^N_{x,j} + 0.20 d^C_{x,j} + 0.05 d^H_{x,j}\} \tag{13}$$

where $d^N_{x,j}$, $d^C_{x,j}$, $d^H_{x,j}$ are the distances between x and j depending on their A-N, A-C and A-E spectra, respectively.

7. The similarity measure between images X and Y is then calculated using:

$$Sim(X,Y) = 1 - \frac{D_x}{D_{\max}} \tag{14}$$

where $D_{\max}$ is the maximum of D_x over the data set, which helps in normalising D_x.

4 Results

The experiments were implemented in C and matlab and run on an AMD Opteron-based machine. Feature extraction took approximately 3 seconds per image. The overall performance of this method has shown that combining the three spectra improves the MAP of the best performing single spectrum by over 28%. Table 1 gives the detailed results of the different combinations that were performed. They show that the A-N spectrum has the highest discriminative

Table 1. Performance of the spectra

Run	MAP	P20	Relevant	% improvement
A-N	0.0444	0.1258	830	-
A-C	0.0338	0.1100	819	-
A-E	0.0265	0.0767	622	-
A-N and A-C	0.0539	0.1508	932	21.4
A-N and A-E	0.0479	0.1358	846	7.9
A-N, A-C and A-E	0.0571	0.1608	926	28.6

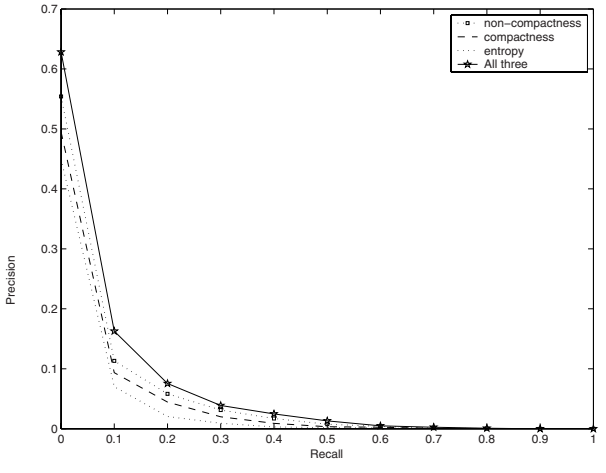

Fig. 2. Interpolated Precision - Recall Averages

power, followed by A-C and A-E respectively. Figure 2 illustrates the interpolated precision-recall average for the three separate and the combined spectra. As expected, at any given point, the precision of the combined spectrum is much higher than any of the individual ones. Initial results showed that Bayes classification out-performed k-nearest neighbour and decision tree. Bayes classification improves the MAP of the combined filter by 28% from 0.0444 to 0.0571 and precision at 20 from 0.1258 to 0.1333.

5 Discussion

Our experiments have shown that using only one technique, i.e, the 2D pattern spectra, produces very promising results for CBIR. There is no doubt that combining it with other visual descriptors like color or texture will further enhance performance for image retrieval. This work proposes a feature vector that combines three 2D pattern spectra: the area–non-compactness, area–compactness and area–entropy spectra. The combined spectrum translates in an improved performance in terms of both the mean average precision and precision at 20. Given the small training set used and simple retrieval scheme, the registered performance indicates that this feature set shows promise and should be developed further. Urbach et al [6] already showed that the area–non-compactness spectrum is very robust against noise in the application of diatom identification. The difference in performance between the different pattern spectra may be attributed to differences in robustness to noise. Compactness is probably less robust through the use of the perimeter parameter. The fact that the A-C spectrum required a classification weight of 70% compared to 20% for A-N and A-E respectively could indicate that the decision boundary with the simple nearest neighbor classifier is less reliable in the case of compactness. The relatively poor performance of entropy may mean that shape is relatively more important than variation in gray level. We believe that choosing features using more advanced relevance learning techniques [18,19] as well as using a larger training set will enhance the MAP scores registered here. Secondly, obtaining the spectra from specific objects (cartoon) as opposed to the whole image can also be tried out [20,21]. Further advancements should include relevance feedback by users.

References

1. Bober, M.: MPEG-7 Visual Descriptors. IEEE Transactions in Circuits and Systems for Video Technology 11(7), 703–715 (2001)
2. Matheron, G.: Random sets and integral Geometry. John Wiley and Sons, Chichester (1975)
3. Maragos, P.: Pattern Spectrum and Multiscale shape Representation. IEEE Transactions Pattern Analy. Mach. Intel. 11(7), 701–715 (1989)
4. Wilkinson, M.H.F.: Generalized Pattern Spectra sensitive to Spatial Information. In: Proceeding of the 16th International Conference on Pattern Recognition, Quebec City, vol. 1, pp. 701–715 (2002)

5. Hu, M.K.: Visual pattern recognition by moment invariants. IRE Transactions on Information Theory IT-8, 179–187 (1962)
6. Urbach., E.R., Roerdink., J.B.T.M., Wilkinson, M.H.F.: Connected Shape-Size Pattern Spectra for Rotation and Scale-Invariant Classification of Gray-Scale Images. Trans. Pattern Analy. Machine Intell. 29(2), 272–285 (2007)
7. Bagdanov., A., Worring, M.: Granulometric analysis of document images. In: Proceeding of the 16th International Conference on Pattern Recognition, vol. 1, pp. 468–471 (2002)
8. Fuertes Garcia., J.M., Lucena Lopez, M., Gomez, J.I., de la Blanca, N.P., Fdez-Valdivia, J.: Content Based Image Retrieval Using a 2D Shape Characterization. In: Fifth Iberoamerican Symposium On Pattern Recognition (SIARP), Lisbon, pp. 529–544 (2000)
9. Grubinger., M., Clough., P., Clement, L.: The IAPR TC-12 Benchmark for Visual Information Search. IAPR Newsletter 28(2), 10–12 (2006)
10. Urbach, E.R., Wilkinson, M.H.F.: Shape-Only Granulometries and Grey-scale shape filters. In: International Symposium on Mathematical Morphology, Sydney, Australia (2002)
11. Breen, E.J., Jones, R.: Attribute openings, thinnings and granulometries. Computer Vision Image Understanding 64(3), 377–389 (1996)
12. Salembier, P., Oliveras, A., Garrido, L.: Antiextensive connected operators for image and sequence processing. IEEE Transactions In Image Processing 7(4), 555–570 (1998)
13. Meijster, A., Wilkinson, M.H.F.: A comparison of Algorithms for Connected Set openings and Closings. IEEE Trans. Pattern Analy. Mach. Intell. 34(4), 484–494 (2002)
14. Nardi, A., Peters, C.: Working Notes of the 2007 CLEF Workshop, Budapest (2007)
15. Tushabe, F., Wilkinson, M.H.F.: Content-based Image Retrieval Using Shape-Size Pattern Spectra. In: Working notes for the CLEF 2007 Workshop, Hungary (2007)
16. Demsar, J., Zupan, B., Leban, G.: Orange: From Experimental Machine Learning to Interactive Data Mining. Faculty of Computer and Information Science, University of Ljubljana (2004)
17. Kira, K., Rendell, L.: A practical approach to feature selection. In: Proceedings of the 9th International Conference on Machine Learning, Aberdeen, pp. 249–256 (1992)
18. Hammer, B., Villmann, T.: Generalized relevance learning vector quantization. Neural Networks 15, 1059–1068 (2002)
19. Hammer, B., Strickert, M., Villmann, T.: On the generalization capability of GR-LVQ networks. Neural Processing Letters 21, 109–120 (2005)
20. Maragos, P., Evangelopoulos, G.: Levelling cartoons, texture energy markers and image decomposition. In: Proceeding of the 8th International Symposium on Mathematical Morphology, Rio de Janeiro, pp. 125–138 (2007)
21. Sofou, A., Evangelopoulos, G., Maragos, P.: Coupled Geometric and Texture PDE - based Segmentation. In: Proceeding of the International Conference on Image Processing, vol. II, pp. 650–653 (2005)