IMAGE SIMILARITY USING THE NORMALIZED COMPRESSION DISTANCE BASED ON FINITE CONTEXT MODELS

Armando J. Pinho and Paulo J. S. G. Ferreira

Signal Processing Lab, IEETA / DETI
University of Aveiro, 3810-193 Aveiro, Portugal
ap@ua.pt — pjf@ua.pt

ABSTRACT

A compression-based similarity measure assesses the similarity between two objects using the number of bits needed to describe one of them when a description of the other is available. Theoretically, compression-based similarity depends on the concept of Kolmogorov complexity but implementations require suitable (normal) compression algorithms. We argue that the approach is of interest for challenging image applications but we identify one obstacle: standard high-performance image compression methods are not normal, and normal methods such as Lempel-Ziv type algorithms might not perform well for images. To demonstrate the potential of compression-based similarity measures we propose an algorithm that is based on finite-context models and works directly on the intensity domain of the image. The proposed algorithm is compared with several other methods.

Index Terms— Image similarity, normalized compression distance, image compression

1. INTRODUCTION

Measuring similarity between images is an important open problem. The challenge is twofold: ideally one would like to assess the similarity without requiring the images to be perfectly aligned, illuminated etc. and on the other hand one would like the results to correlate well with our visual perception.

The simplest approach is to use a norm to calculate the difference between two images. Although it is often criticized for not being correlated with our perception, the L_2-norm is among those most often used. This approach works reasonably well when one image is a degraded or noisy version of the other. Under more general circumstances, and in particular when the images are not perfectly aligned, direct application of a norm becomes useless.

A good method would give a meaningful indication of how similar two images are, regardless of their geometry, orientation, scale, and other similar characteristics. A common approach is to extract a set of features from the images and then compare them. Popular choices include shape-based, color-based or texture-based features. A major difficulty associated with these methods is precisely how to select meaningful features.

Recently, there has been interest in image similarity measures based on compression methods [1, 2, 3, 4]. They rely on the notion of Kolmogorov complexity and opened a line of research that seems promising.

Given a string of bits A, its Kolmogorov complexity $K(A)$ is by definition the minimum size of a program that produces A and stops. A repetitive structure can be described by a small program ("print ab 1000 times") which scales with the logarithm of $|A|$, indicating low complexity. On the other hand, for a very complex pattern there might be no better program than "print A", the length of which scales with $|A|$, indicating high complexity.

A major drawback of the Kolmogorov complexity (also known as the algorithmic entropy) is that it is not computable. Therefore, one is forced to deal with approximations that provide upper bounds to the true complexity. Compression algorithms provide natural ways of approximating the Kolmogorov complexity, because, together with the appropriate decoder, the bitstream produced by a lossless compression algorithm allows the reconstruction of the original data. The number of bits required for representing these two components (decoder and bitstream) can be viewed as an estimate of the Kolmogorov complexity of the original bitstream. From this point of view, the search for better compression algorithms is directly related to the problem of how to improve the complexity bounds.

This theory and its relation with data compression has been known for some time but a systematic investigation of its possible interest for image analysis is still lacking. The assessment of compression-based similarity measures in the context of image analysis requires, first and foremost, a suitable image compression method. This does not appear to be a problem, since image compression has been a vibrant field of research for decades, but in reality there is a serious obstacle.

To measure similarity, a compression method needs to be *normal*, that is, it should generate essentially the same number of bits when compressing AA (the concatenation of A with A) and when compressing A alone. Lempel-Ziv based compressors are approximately normal and as such are frequently used in reference to compression-based similarity measurement. However, generally speaking, they do not perform well on images. On the other hand, most of the best performing image compression algorithms are not normal.

The goals of this paper are twofold. First and foremost, we want to emphasize the potential interest and applications of compression-based similarity measures in the field of image analysis. The present work is by no means closed and we hope that it may stimulate discussion and interest in the problem.

Since standard coding methods cannot be used for measuring similarity, and normal methods do not appear to lead to good performance, we need alternatives. Our second goal is to demonstrate the potential of compression-based similarity measures by using algorithms based on finite-context models that work directly on the intensity domain of the image and avoid the transform or predictive step that other encoders use.

978-1-4577-1302-6/11/$26.00 ©2011 IEEE

2. THE NORMALIZED COMPRESSION DISTANCE

The work of Solomonoff, Kolmogorov, Chaitin and others [5, 6, 7, 8, 9, 10] on how to measure complexity has been of paramount importance for several areas of knowledge. However, because it is not computable, the Kolmogorov complexity of A, $K(A)$, is usually approximated by some computable measure, such as Lempel-Ziv based complexity measures [11], linguistic complexity measures [12] or compression-based complexity measures [13].

Kolmogorov theory also leads to an approach to the problem of measuring similarity. Li *et al.* proposed a similarity metric [14] based on an information distance [15], defined as the length of the shortest binary program that is needed to transform A and B into each other. This distance depends not only on the Kolmogorov complexity of A and B, $K(A)$ and $K(B)$, but also on conditional complexities, for example $K(A|B)$, that indicates how complex A is when B is known. Because this distance is based on the Kolmogorov complexity (not computable), they proposed a practical analog based on standard compressors, which they call the normalized compression distance [14],

$$\mathrm{NCD}(A, B) = \frac{C(AB) - \min\{C(A), C(B)\}}{\max\{C(A), C(B)\}}, \qquad (1)$$

where $C(A)$ gives the number of bits required by compressor C for compressing string A. Successful applications of these principles have been reported in areas such as genomics, virology, languages, literature, music, handwritten digits and astronomy [16]. However, applications of the normalized compressing distance to the imaging area are scarce. This seemingly surprising fact is due to the following reasons.

According to Li et al. [14], a compression method needs to be *normal* in order to be used in a normalized compression distance. One of the conditions for a compression method to be normal is that the compression of AA (the concatenation of A with A) should generate essentially the same number of bits as the compression of A alone [16]. This characteristic holds, for example, in Lempel-Ziv based compressors, making them a frequent choice in the applications of the complexity principles to image analysis [1, 2, 17]. However, generally speaking, Lempel-Ziv based compressors do not perform well on images. Moreover, most of the best performing image compression algorithms are not normal compressors, according to the definition of [16].

A normal compression algorithm accumulates knowledge of the data while compression is performed. It finds dependencies, collects statistics, i.e., it creates a model of the data. Most state-of-the-art image compressors start by decorrelating the data using a transformation (for example, the DCT or DWT as in JPEG or JPEG2000) or a predictive method (as in JPEG-LS). Therefore, they assume an a priori data model that remains essentially static during compression. Moreover, this first step destroys most of the data dependencies, leaving to the entropy coding stage the mere task of encoding symbols from an (assumed) independent source. This makes them unsuitable for measuring similarity and alternatives need to be sought.

3. A NEW ENCODER FOR MEASURING SIMILARITY

We implemented an image encoder for gray scale images based on finite-context modeling, similar in principle to that used by JBIG. However, it differs from JBIG in two major aspects. First, instead of addressing image coding in a bit-plane basis, as JBIG does, we designed an encoder that handles the whole pixel at once. The drawback of this approach is that since finite-context modeling usually requires memory resources that grow exponentially with the order of the model, it is unpractical to apply them to large alphabets. Although it is possible to alleviate this memory burden through the use of sophisticated data structures such as hashing, in this exploratory work we opted for reducing the size of the alphabet. Therefore, before being processed by the encoder, the images are quantized to just four levels, using a Lloyd-Max quantizer. At first thought this might sound too severe, but in fact the results that we have obtained show that even with only four levels our approach behaves globally better than the others to which we compared.

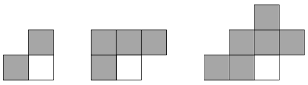

Fig. 1. Context templates used by the encoder, corresponding to model orders of 2, 4 and 6.

The second significant difference to JBIG is the use of simultaneous multiple contexts, combined using a mixture model that relies on the past performance of each of those models. In the remainder of this section we present some more details regarding this setup.

A finite-context model assigns probability estimates to the symbols of the alphabet $\mathcal{A} = \{s_1, s_2, \ldots, s_{|\mathcal{A}|}\}$, where $|\mathcal{A}|$ denotes the size of the alphabet, regarding the next outcome of the information source, according to a conditioning context computed over a finite and fixed number, $k > 0$, of the most recent past outcomes $c_{k,n} = x_{n-k+1} \ldots x_{n-1}x_n$ (order-k finite-context model) [18]. The number of conditioning states of the model is $|\mathcal{A}|^k$.

The probability estimates, $P(X_{n+1} = s|c_{k,n}), \forall_{s\in\mathcal{A}}$, are calculated using symbol counts that are accumulated while the image is processed, making them dependent not only of the past k symbols, but also of n. We use the estimator

$$P(X_{n+1} = s|c_{k,n}) = \frac{n_s^{c_{k,n}} + \alpha}{n^{c_{k,n}} + \alpha|\mathcal{A}|}, \qquad (2)$$

where $n_s^{c_{k,n}}$ represents the number of times that, in the past, the information source generated symbol s having $c_{k,n}$ as the conditioning context and where

$$n^{c_{k,n}} = \sum_{a\in\mathcal{A}} n_a^{c_{k,n}} \qquad (3)$$

is the total number of events that has occurred so far in association with context $c_{k,n}$. Parameter α allows balancing between the maximum likelihood estimator and an uniform distribution. Note that when the total number of events, n, is large, the estimator behaves as a maximum likelihood estimator. For $\alpha = 1$, (2) is the well-known Laplace estimator.

The per symbol information content average provided by the finite-context model of order-k, after having processed n symbols, is given by

$$H_{k,n} = -\frac{1}{n}\sum_{i=0}^{n-1} \log_2 P(X_{i+1} = x_{i+1}|c_{k,n}) \quad \text{bps}, \qquad (4)$$

where “bps” stands for bits per symbol.

The encoder was implemented using three finite-context models, of orders $k = 2, 4, 6$, that operate simultaneously. Figure 1 displays the configuration of the three contexts. For each symbol, the probability estimate given by each of the three models is combined using averaging, according to

$$P(X_{n+1} = s) = \sum_{k=2,4,6} P(X_{n+1} = s|c_{k,n})\, w_{k,n}, \tag{5}$$

where

$$w_{k,n} \propto w_{k,n-1}^{\gamma} P(X_n = x_n|c_{k,n-1}) \tag{6}$$

and

$$\sum_{k=2,4,6} w_{k,n} = 1. \tag{7}$$

Although not explained in detail here, for lack of space, it can be shown that these weights favor the models that have provided better performance in the recent past of the sequence of symbols. Parameter γ is usually very close to one (we used $\gamma = 0.99$).

4. EXPERIMENTAL RESULTS

For performing the experiments reported in this paper, we used the ORL face database, which contains 92×112, 8 bits per pixel gray level image faces of 40 distinct subjects (10 from each one), taken between April 1992 and April 1994 at the Olivetti Research Laboratory in Cambridge, UK [19]. Figure 2 shows some examples taken from this image face database.

For comparison, we calculated the normalized compression distance using four general purpose compressors and three image coding standards, namely lossless JPEG2000 [1], JPEG-LS [2], and JBIG [3]. The set of general purpose compressors was composed by the Linux implementations of GZIP version 1.3.12 (based on LZ77 Lempel-Ziv coding), BZIP2 version 0.9.0b (based on Burrows-Wheeler block-sorting and Huffman coding), LZMA version 4.32.0beta3, SDK 4.43 (Lempel-Ziv-Markov chain-Algorithm), and PPMd (based on prediction by partial matching).

We divided the set of 400 face images in two subsets. The first subset (which we call the reference subset) contains the first image of each of the 40 subjects. The second subset (the test subset) contains the 360 remaining images. For each image in the reference subset, we calculated the normalized compression distance to all images in the test subset. Then we picked the nine images having the shortest distances and we determined how many of them corresponded to the reference subject. These are the numbers presented in Table 1, where nine means that all face images have been correctly associated to that subject.

5. DISCUSSION

Measuring image similarity using the concept of Kolmogorov complexity might open new ways for dealing with problems such as content-based image retrieval. The normalized compression distance as been applied with success to several types of uni-dimensional data, usually using general purpose data compressors. Being an approximation of the normalized information distance, the normalized compression distance relies on a good approximation of $K(A)$ by $C(A)$ and, therefore, requires good compressors for the type of data addressed.

Subject	bzip2	gzip	lzma	ppmd	jbig	jp2k	jpls	fcm
s01	5	3	0	3	2	1	3	7
s02	9	9	6	5	2	0	0	9
s03	3	4	1	4	2	1	1	4
s04	4	4	0	2	1	2	2	4
s05	4	4	2	4	2	2	1	4
s06	6	3	0	1	2	0	0	5
s07	7	6	0	0	1	2	0	6
s08	7	4	1	3	5	1	4	7
s09	2	2	2	2	1	3	0	3
s10	6	7	2	3	3	0	4	9
s11	7	9	1	6	4	0	0	9
s12	3	5	0	2	2	0	5	4
s13	3	4	0	3	2	0	0	4
s14	8	7	1	4	4	0	1	9
s15	4	4	0	2	3	0	1	4
s16	0	0	1	0	0	0	0	6
s17	6	6	2	3	2	3	0	8
s18	6	8	0	0	0	1	3	6
s19	9	9	3	9	3	2	0	9
s20	3	5	0	0	1	0	0	3
s21	4	6	0	1	0	0	0	2
s22	7	9	2	8	1	0	0	8
s23	4	4	4	6	0	2	0	7
s24	5	9	0	3	3	0	0	8
s25	6	6	2	4	3	3	0	7
s26	4	5	1	3	1	0	2	7
s27	9	9	1	7	4	0	0	9
s28	4	4	1	1	1	0	3	4
s29	4	4	3	3	3	1	3	2
s30	8	5	2	6	3	0	1	7
s31	2	4	2	4	0	0	0	2
s32	7	9	1	4	6	1	6	9
s33	5	2	0	2	2	0	0	6
s34	9	7	2	3	2	0	2	7
s35	5	6	0	3	1	1	0	3
s36	1	1	0	1	1	0	0	1
s37	4	4	0	0	2	0	0	4
s38	4	4	3	3	2	2	1	2
s39	2	1	0	2	1	0	0	4
s40	3	4	1	0	0	0	0	4
Sum	199	206	47	120	78	28	43	223

Table 1. Comparison of the compression methods. Columns 2–5 correspond to general purpose compressors. Columns 6–8 correspond, respectively, to the JBIG, JPEG2000 and JPEG-LS standards. The last column shows the results of the method proposed. The values indicate the number of correct images within the nine shortest normalized compression distances of each reference image.

It is known for long that dedicated algorithms for image compression behave (much) better than general purpose data compressors. The problem is that the most popular image compression techniques, and particularly the image compression standards, do not comply with the requirements of a *normal* compressor. This is confirmed by the results presented in Table 1, where we can see the poor performance of the three standards, JBIG, JPEG2000 and JPEG-LS. Even JBIG, which is based on global context modeling, although on a bit-plane basis, performs poorly, but, even so, better that the other two image coding standards.

Regarding the four general purpose compressors used, both BZIP2 and GZIP behaved surprisingly well, with a small advantage to the one based on LZ77. The other two, LZMA and PPMd performed much worse.

[1] JPEG2000 codec from `http://jj2000.epfl.ch`

[2] The original website of this codec, `http://spmg.ece.ubc.ca`, is currently unavailable, but it can be obtained from `ftp://www.ieeta.pt/~ap/codecs/jpeg_ls_v2.2.tar.gz`.

[3] JBIG codec from `http://www.cl.cam.ac.uk/~mgk25/jbigkit/`.

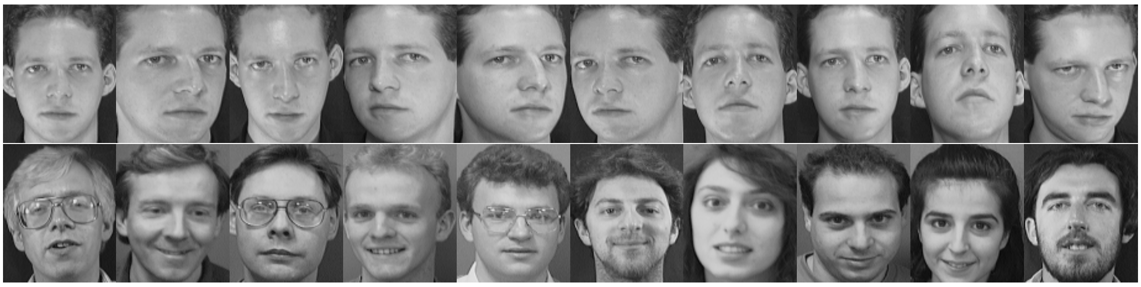

Fig. 2. Examples from the ORL face database. In the first row, ten face images corresponding to the same subject (s01). In the second row, the first face image of subjects s02 to s11.

The simple encoder that we have developed for the purpose of beginning studying this new paradigm of image similarity provided the best results in this experimental setup. Due to the memory requirements imposed by the finite-context models on large alphabets, we decided to quantize the images to just four levels before compression. This was done only for this encoder. All the other encoders have been provided with the original images. We also tested them on the quantized images, but the results have been much worse than those attained with the original images.

The experiments that we have performed confirm that current image coding standards are not suited for calculating the normalized compression distance. We believe that this is due, in a large part, to the violation of the condition $C(AA) = C(A)$ that is required by the complexity measure. For example, the $\mathrm{NCD}(A, A)$ values given by GZIP, BZIP2, JBIG, JPEG-LS and JPEG2000, where A was a random image, were 0.0143, 0.2200, 0.8165, 0.9543 and 0.9796, respectively. The encoder that we have developed gave 0.6595, which is still a high value, but lower than that provided by the other image encoders. When compared to the general purpose compressors, we believe that the good performance attained by the encoder based on finite-context models is related to the fact that 2D contexts are being used. Therefore, it is more suited for handling 2D data, despite using images with only four gray levels and not being able to attain values of $\mathrm{NCD}(A, A)$ closer to zero. Much work still needs to be done, but there it seems to be also plenty of room for improvement.

6. ACKNOWLEDGMENT

This work was supported in part by the FCT, Fundação para a Ciência e Tecnologia.

7. REFERENCES

[1] I. Gondra and D. R. Heisterkamp, "Content-based image retrieval with the normalized information distance," *Computer Vision and Image Understanding*, vol. 111, pp. 219–228, 2008.

[2] N. Tran, "The normalized compression distance and image distinguishability," in *Human Vision and Electronic Imaging XII — Proc. of the SPIE*, Jan. 2007, p. 64921D.

[3] J. Perkiö and A. Hyvärinen, "Modelling image complexity by independent component analysis, with application to content-based image retrieval," in *Proc. of the Int. Conf. on Artificial Neural Networks, ICANN 2009*, Limassol, Cyprus, 2009.

[4] J. Mortensen, J. J. Wu, J. Furst, J. Rogers, and D. Raicu, "Effect of image linearization on normalized compression distance," in *Signal Processing, Image Processing and Pattern Recognition*, D. Slezak, S. K. Pal, B.-H. Kang, J. Gu, H. Kuroda, and T.-H. Kim, Eds. 2009, vol. 61 of *Communications in Computer and Information Science*, pp. 106–116, Springer Berlin Heidelberg.

[5] R. J. Solomonoff, "A formal theory of inductive inference. Part I," *Information and Control*, vol. 7, no. 1, pp. 1–22, Mar. 1964.

[6] R. J. Solomonoff, "A formal theory of inductive inference. Part II," *Information and Control*, vol. 7, no. 2, pp. 224–254, June 1964.

[7] A. N. Kolmogorov, "Three approaches to the quantitative definition of information," *Problems of Information Transmission*, vol. 1, no. 1, pp. 1–7, 1965.

[8] G. J. Chaitin, "On the length of programs for computing finite binary sequences," *Journal of the ACM*, vol. 13, pp. 547–569, 1966.

[9] C. S. Wallace and D. M. Boulton, "An information measure for classification," *The Computer Journal*, vol. 11, no. 2, pp. 185–194, Aug. 1968.

[10] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.

[11] A. Lempel and J. Ziv, "On the complexity of finite sequences," *IEEE Trans. on Information Theory*, vol. 22, no. 1, pp. 75–81, Jan. 1976.

[12] G. Gordon, "Multi-dimensional linguistic complexity," *Journal of Biomolecular Structure & Dynamics*, vol. 20, no. 6, pp. 747–750, 2003.

[13] T. I. Dix, D. R. Powell, L. Allison, J. Bernal, S. Jaeger, and L. Stern, "Comparative analysis of long DNA sequences by per element information content using different contexts," *BMC Bioinformatics*, vol. 8, no. Suppl. 2, pp. S10, 2007.

[14] M. Li, X. Chen, X. Li, B. Ma, and P. M. B. Vitányi, "The similarity metric," *IEEE Trans. on Information Theory*, vol. 50, no. 12, pp. 3250–3264, Dec. 2004.

[15] C. H. Bennett, P. Gács, M. Li P. M. B. Vitányi, and W. H. Zurek, "Information distance," *IEEE Trans. on Information Theory*, vol. 44, no. 4, pp. 1407–1423, July 1998.

[16] R. Cilibrasi and P. M. B. Vitányi, "Clustering by compression," *IEEE Trans. on Information Theory*, vol. 51, no. 4, pp. 1523–1545, Apr. 2005.

[17] A. Mallet, L. Gueguen, and M. Datcu, "Complexity based image artifact detection," in *Proc. of the Data Compression Conf., DCC-2008*, Snowbird, Utah, 2008, p. 534.

[18] T. C. Bell, J. G. Cleary, and I. H. Witten, *Text compression*, Prentice Hall, 1990.

[19] F. Samaria and A. Harter, "Parameterisation of a stochastic model for human face identification," in *2nd IEEE Workshop on Applications of Computer Vision*, Sarasota, Florida, Dec. 1994, pp. 138–142.