GENERIC IMAGE SIMILARITY BASED ON KOLMOGOROV COMPLEXITY

Nima Nikvand   and   Zhou Wang

Dept. of Electrical & Computer Engineering, University of Waterloo, Waterloo, ON, Canada
Email: nnikvand@uwaterloo.ca,   zhouwang@ieee.org

ABSTRACT

Image similarity measurement is a fundamental and common issue in a broad range of problems in image processing, compression, communication, recognition and retrieval. Existing image similarity measures are limited to restricted application environments. The theory of Kolmogorov complexity and the related normalized information distance (NID) measure provide an attractive theoretic framework for generic image similarity that is applicable to any scenario. While this is appealing, the difficulty lies in the implementation due to the non-computable nature of Kolmogorov complexity. In this paper, we propose a practical framework to approximate NID, where the key is to find the shortest program within a set of potential transformations that convert one image to another and vice versa. As one of the initial attempts in this new and promising research direction, our preliminary experimental work demonstrates the wider applicability of the proposed approach than existing methods.

Index Terms— image similarity measurement, Kolmogorov complexity, normalized information distance, compression distance

1. INTRODUCTION

Measuring the similarity between two images is a fundamental issue in many problems throughout the entire field of image processing and machine vision. These include image restoration, denoising, coding, communication, interpolation, registration, fusion, classification and retrieval, as well as object detection, recognition and tracking. Many existing image similarity measures were proposed to work with very specific types of image distortions (e.g., JPEG compression) [1]. There are also methods such as the structural similarity (SSIM) index [2] that are applicable to a wider range of applications. However, even these "general-purpose" methods [1] are still limited in their application scopes. For example, SSIM does not apply or work properly when significant geometric changes (e.g., enlargement by a factor of 2, or rotation by 90 degrees) exist between the two images being compared.

The theory of Kolmogorov complexity provides solid groundwork to build a universal and generic information distance metric between any objects that minorizes all metrics in the class [3]. A practically more useful metric, namely the normalized information distance (NID), was introduced in [4]. To overcome the non-computable nature of Kolmogorov complexity and NID, an normalized compression distance (NCD) measure was proposed [4], which is an effective approximation of NID and has found many successful applications in the fields of bioinformatics, pattern recognition, and natural language processing.

Nevertheless, the application of NID for image similarity measurement is still in its early stage. Several authors have done pioneering work that applied the NID framework and the NCD algorithm to image clustering [5], image distinguishability [6], content-based image retrieval [7] and video classification problems [8], but most of them reported only moderate success. Moreover, due to their focuses on specific applications, the *generic* property of NID was not fully exploited.

In this paper, we attempt to develop a practical framework that can effectively approximate NID, where the most critical step is to find the shortest program within a list of available transformations that convert one image to another. Based on the framework, generic image similarity measures can be built that have much wider applicability than existing image similarity measures.

2. KOLMOGOROV COMPLEXITY AND NORMALIZED INFORMATION DISTANCE

The Kolmogorov complexity [3] of an object is defined to be the length of the shortest program that can produce that object on a universal Turing machine and halt:

$$K(x) = \min_{p:U(p)=x} l(p). \qquad (1)$$

In [4], the authors assume the existence of a general decompressor that can be used to decompress the presumably shortest program x^* to the desired object x. However, they note that due to the non-computability of this concept, a compressor that does the opposite does not have to exist.

The conditional Kolmogorov complexity of x relative to y is denoted by $K(x|y)$. An information distance between x and y can then be defined as $\max\{K(x|y), K(y|x)\}$, which is the maximum of the length of the shortest program that computes x from y and y from x. To convert it to a normalized

symmetric metric, a novel NID measure was introduced in [4]:

$$\text{NID}(x,y) = \frac{\max\{K(x|y^*), K(y|x^*)\}}{\max\{K(x), K(y)\}}. \quad (2)$$

It was proved that NID is a valid *distance metric* that satisfies the identity and symmetry axioms and the triangular inequality [4].

The real-world application of NID is difficult because Kolmogorov complexity is a non-computable quantity [3]. By using the fact that $K(xy) = K(y|x^*) + K(x) = K(x|y^*) + K(y)$ (subject to a logarithmic term), and by approximating Kolmogorov complexity K using a practical data compressor C, a normalized compression distance (NCD) was proposed in [4] as

$$\text{NCD}(x,y) = \frac{C(xy) - \min\{C(x), C(y)\}}{\max\{C(x), C(y)\}}. \quad (3)$$

NCD has been proved to be an effective approximation of NID and achieves superior performance in bioinformatics applications such as the construction of phylogeny trees using DNA sequences [4].

3. THE PROPOSED FRAMEWORK

When NCD was used to quantify image similarities, it did not achieve the same level of success as in other application fields. For example, it was reported in [6] that NCD works well when parts are added or subtracted from an image, but struggles when image variations involve form, material and structure. We believe that this is mainly due to the poor approximation of $K(xy)$ using $C(xy)$, which is often implemented by applying a regular image compressor to the concatenation of two images. For example, when an image is a ninety-degree rotated copy of another, concatenating two images would not facilitate any efficient compression. To avoid this problem, we propose to approximate the conditional Kolmogorov complexity in Eq. (2) directly by designing a conditional image compressor denoted by C_T, so that

$$K(y|x) \approx C_T(y|x) \quad \text{and} \quad K(x|y) \approx C_T(x|y). \quad (4)$$

This leads to a normalized conditional compression distance (NCCD) measure given by

$$\text{NCCD}(x,y) = \frac{\max\{C_T(x|y), C_T(y|x)\}}{\max\{C(x), C(y)\}}. \quad (5)$$

It remains to define the conditional compressor C_T. Here we propose a practical solution by making use of a set of transformations that convert one image to another. Let $\{T_i | i = 1, \cdots, N\}$ be the set of transformations, let $T_i(x)$ represent the transformed image when applying the i-th transform to image x, and let $p(T_i, x)$ denote the parameters used in the transformation. Each type of transformation is also associated with a parameter compressor, and C_i^p denotes the parameter compressor of the i-th transformation. We can then define our conditional compressor as

$$C_T(y|x) = \min_i \{C[y - T_i(x)] + C_i^p[p(T_i, x)] + \log_2(N)\}, \quad (6)$$

where C remains to be a practical image compressor which encodes the difference between y and the transformed image $T_i(x)$, and the $\log_2(N)$ term computes the number of bits required to encode the selection of one out of N potential transformations.

The idea of finding the simplest transformation between two images is sensible from the viewpoint of human visual perception, for which it has long been hypothesis that the biological visual system is an efficient coder of the visual world [9]. For example, given two images that are rotated copy of each other, our visual system would not interpret the difference between them by directly differencing their intensity values (which requires a large number of bits to encode the residual), but by estimating the amount of rotation (which can be coded very efficiently).

4. IMPLEMENTATION

An advantage of NCCD (as opposed to NCD) is that it provides a more flexible framework so that different types of transformations can be included. The list of transformations can also be incremental, in the sense that new transformations, when available, can be easily added into the existing list, and expanding the list always improves the approximation of NCCD to NID. Of course, exhausting all possible transformations is practically impossible. However, by going through a handful of transformations, it may be sufficient to appropriately cover most image distortions encountered in real-world applications.

Our current implementation of NCCD are as follows. First, we adopt the content adaptive lossless image compression algorithm (CALIC) [10] as the base image compressor, which achieves superior performance when compared with state-of-the-art algorithms. CALIC is employed in computing the denominator of Eq. (5) as well as the first term in Eq. (6). Since $y - T_i(x)$ in Eq. (6) can generate negative values and CALIC applies to grayscale images with positive intensity values only, the mean intensity value of $y - T_i(x)$ is shifted to mid-gray level before the application of CALIC. Second, the types of transformations involved in the computation of C_T include

- *Global contrast and luminance change*. This is computed by a pointwise intensity transformation defined as $s = \alpha(r - \bar{r}) + \bar{r} + \beta$, where r and s are the intensity values before and after the transformation, respectively, $\bar{r}$ is the average value of r, and α and β are the parameters that determine the degrees of contrast and mean luminance changes, respectively. In a special case when

NCCD = 0.8121, SSIM = 0.4649, MSE = 73.4   NCCD = 0.2189, SSIM = 0.7059, MSE = 49.8   NCCD = 0.3901, SSIM=0.8425, MSE = 78.4   NCCD =0.0091, SSIM = 0.9094, MSE = 105.5

Fig. 1. Comparison of MSE, SSIM and NCCD measures using images distorted by JPEG compression, blur, JPEG2000 compression and contrast reduction.

$\alpha = 1$ and $\beta = 0$, it reduces to an identity transform, i.e., $T(x) = x$.

- *Global Fourier power spectrum scaling*. This transformation attempts to match two images by scaling the power spectrum of one image in the Fourier transform domain. Let $X(\omega)$ and $Y(\omega)$ be the Fourier transforms of x and y, respectively. We first find the best linear transform parameters p_1 and p_2, such that $\||Y(\omega)| - (p_1|X(\omega)| + p_2)\|^2$ is minimized. We then define the transform $T(x)$ as the inverse Fourier transform of $p_1 X(\omega) + p_2$.

- *Global affine transform*. This transformation tries to match one image by applying a global affine transform to another. The transformation can be encoded using six parameters and covers a variety of image changes including translation, scaling (zooming in or zooming out), rotation, and shearing.

- *Local registration transformation*. This is implemented by aligning two images using the coherent point drift registration algorithm [11] that allows for both rigid and affine non-rigid transformations.

Given a pair of images x and y for comparison, we attempt all the above transformations from both x to y and y to x (multiple transformations are also allowed). This is important because the values of $K(x|y)$ and $K(y|x)$ can be drastically different (and so do the values of $C_T(x|y)$ and $C_T(y|x)$). For example, converting the "Lena" image x to a blank image y is easy (as y can be created by a very short program), but the opposite is not.

5. EXPERIMENT

The goal of our preliminary experimental work is to test the applicability of the proposed NCCD implementation for various distortion types and compare it with existing measures such as the mean squared error (MSE) and SSIM. Figure 1 shows four original images distorted with JPEG compression, blur, JPEG2000 compression and contrast reduction. MSE appears to be a poor measure in this test, because the best quality image (contrast reduced) that does not exhibit any structural distortion, results in the worst MSE value. Both SSIM and NCCD works reasonably well, which give the best quality values (highest SSIM and lowest NCCD) to the contrast reduced image.

In Fig. 2, the test images underwent certain geometric distortions, for which the MSE and SSIM measures do not apply or cannot work appropriately. By contrast, NCCD produces meaningful similarity evaluations. In particular, it works when the two images are of different size and shape; it works if parts of an image are missing; it also works when
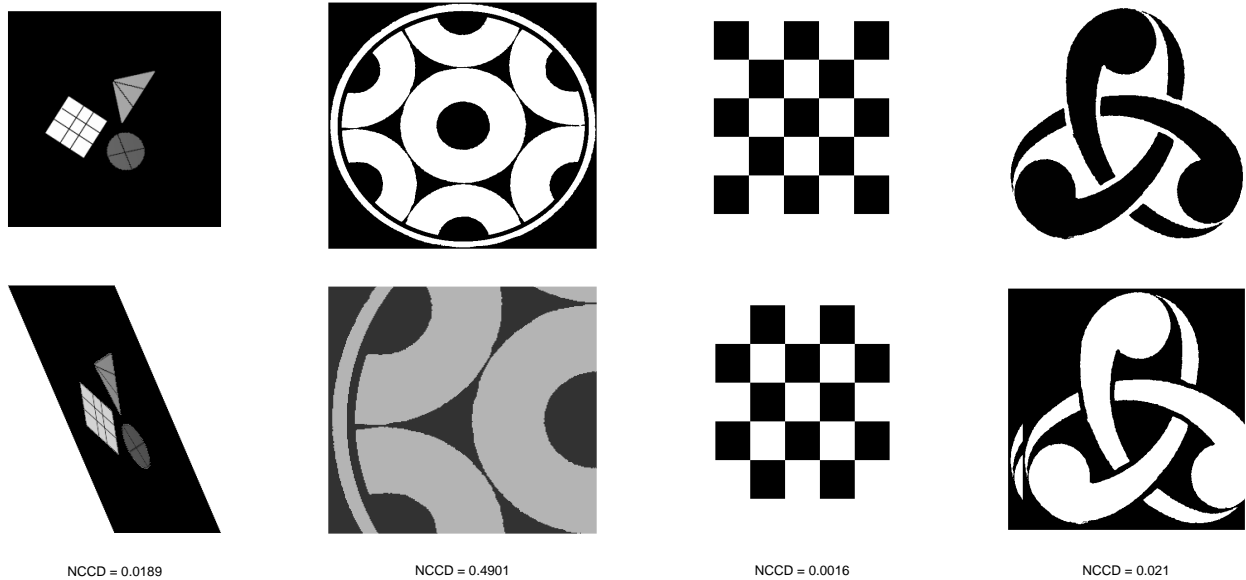
NCCD = 0.0189   NCCD = 0.4901   NCCD = 0.0016   NCCD = 0.021

Fig. 2. Tests using images with geometric and compound distortions.

images are shifted and/or the black/white pixels are reverted. All of these demonstrate the wider applicability of NCCD.

6. CONCLUSION

This work aims to develop a generic image similarity measure based upon the theoretic groundwork of Kolmogorov complexity and the NID metric. The most important contribution of this paper is to propose a practical framework of NCCD for the approximation of NID. The framework is flexible and expandable to include any image transformations that may help find the shortest description that converts one image to another and vice versa. Although the current implementation and experimental work is only preliminary, the resulting similarity measure works properly in a wide variety of scenarios. To the best of our knowledge, no existing image similarity measure has achieved the same level of wide applicability. In the future, the implementation of NCCD can be improved by incorporating more transformations. To further refine the NCCD framework, it is also useful to take into account the degrees of perceptual relevance of different transformations.

7. ACKNOWLEDGMENT

8. REFERENCES

[1] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*. Morgan & Claypool Publishers, Mar. 2006.

[2] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[3] M. Li and P. Vitányi, *An Introduction to Kolmogorov Complexity and Its Applications*. Berlin, 2nd edition: Springer, 1997.

[4] M. Li, X. Chen, X. Li, B. Ma, and P. M. B. Vitányi, "The similarity metric," *IEEE Trans. Info. Theory*, vol. 50, pp. 3250–3264, Dec. 2004.

[5] B. Hescott and D. Koulomzin, "On clustering images using compression," CS Department, Boston University, Tech. Rep., 2007.

[6] N. Tran, "The normalized compression distance and image distinguishability," in *The 19th IS&T/SPIE Symposium on Electronic Imaging Science and Technology*, San Jose, Jan. 2007.

[7] D. Cerra, A. Mallet, L. Gueguen, and M. Datcu, "Complexity based analysis of earth observation imagery: an assessment," in *ESA-EUSC*, Mar. 2008.

[8] K. Kaabneh, A. Abdullah, and Z. Al-Halalemah, "Video classification using normalized information distance," in *Geometric Modeling and Imaging: New Trends*, Aug. 2006, pp. 34–40.

[9] H. Barlow, *Possible principles underlying the transformation of sensory messages*. Cambridge, MA: MIT Press, 1961.

[10] X. Wu and N. Memon, "Context-based, adaptive, lossless image codec," *IEEE Trans. Comm*, vol. 45, no. 4, pp. 437–444, Apr. 1997.

[11] A. Myronenko and X. Song, "Point-set registration: Coherent point drift," 2009. [Online]. Available: http://www.citebase.org/abstract?id=oai:arXiv.org:0905.2635