Neurocomputing 171 (2016) 362–378

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A patch-based measure for image dissimilarity

Alessia Amelio *, Clara Pizzuti *

National Research Council of Italy (CNR) Institute for High Performance Computing and Networking (ICAR), 87036 Rende, CS, Italy

ARTICLE INFO

Article history:
Received 17 October 2014
Received in revised form
22 April 2015
Accepted 23 June 2015
Communicated by Yongdong Zhang
Available online 6 July 2015

Keywords:
Image retrieval
Image similarity
Pattern matching
Pattern recognition

ABSTRACT

A measure for computing the dissimilarity for images is presented. The measure, based on information theory, considers the pixel matrices representing two images, and compares their greatest common sub-matrices. The algorithm to calculate the average area of square sub-matrices that exactly occur in both the images is described, together with its computational complexity, and an extension to accelerate its execution time is proposed. Experimental evaluation of the measure based on human perception of multiple subjects demonstrates that the measure is able to correctly discriminate (dis)similar images. Furthermore, an extensive quantitative evaluation on different kinds of image data sets shows the superiority of the measure with respect to other state-of-the-art measures in terms of retrieval precision.

1. Introduction

Distance computation between images is an important and challenging problem in computer vision, image recognition, image registration, and, more in general, pattern recognition. The determination of similarity between objects is of crucial importance in retrieval systems where a query object must be compared with all those contained in a database to compute the distance between their corresponding representations. Many different measures have been proposed and evaluated in various application domains [1–5], along with algorithms to compute them. Recently, the patch-based approaches are receiving a lot of attention because they allow a more abstract representation of an image [5]. These methods divide the image into small overlapping or non-overlapping rectangles of fixed size, called patches, and analyze the image at patch level, instead that at pixel level. As observed in [6], patches contain contextual information, and are more efficient in terms of computation. Given two images or image regions I and J, nonparametric patch approaches find, for every patch in I, the nearest neighbor patch in J under a patch distance metric. This computation is referred to as *Nearest Neighbor Field (NNF)* and both exact methods using kd-tree structures [7], and approximate methods [6,8,9] have been proposed. However, though approximate methods are more efficient, they could not explore the overall search space and thus produce bad matches. On the other hand, exact methods, when data dimensionality is high, could become unfeasible since the brute force complexity becomes $O(m^2 M^2)$, where m and M are the patch and image sizes, respectively.

In this paper, the information-theoretic measure to compute the (dis)similarity between two images I_A and I_B, presented in [10], is investigated, and an extension for its computation to accelerate execution time is proposed. The measure, named *Average Common Sub-Matrix* (*ACSM*), considers the pixel matrices A and B, defined on an alphabet Σ, associated with I_A and I_B, of size $N \times N$ and $M \times M$, respectively, and determines the area of square sub-matrices of matrix A that exactly occur in B, to quantify the dissimilarity between I_A and I_B. The measure is showed to have upper and lower bounds, moreover, for square matrices, it is demonstrated that it is a semi-metric. In fact, though *ACSM* satisfies nonnegativity, reflexivity, and symmetry properties, triangular inequality cannot be proved. An optimization of the algorithm to compute the measure, that exploits suffix tree data structure and reduces its complexity from $O(M^2N^2 \log(N))$ to $O(N^2+M^2)$, is also proposed. A thorough experimentation on real-world and handwritten digit images shows that the *ACSM* measure is able to reflect the concept of similar images as perceived by humans on the former data set, and to outperform well known information-theoretic measures. Moreover, *ACSM* obtains higher retrieval precision values than measures we compared with. The main advantages of the *ACSM* measure can be summarized as the following:

- *ACSM* is an exact measure that searches the overall search space, and not a limited spatial window, thus it avoids to find locally optimal solutions.
- ACSM does not need to fix the patch size, but only a parameter α representing a lower bound to the patch area. Current

* Corresponding authors.
E-mail addresses: amelio@icar.cnr.it (A. Amelio), pizzuti@icar.cnr.it (C. Pizzuti).

http://dx.doi.org/10.1016/j.neucom.2015.06.044
0925-2312/

patch-based approaches, by fixing the patch size, could miss some recurrent patterns with size bigger than that the fixed patch size.

- The computation of *ACSM* is faster than other image (dis) similarity measures. The computational complexity of accelerated *ACSM* is linear in the image size, and does not depend on the patch size.
- *ACSM* is not sensitive to noise, and thus it is able to correctly discriminate between similar and dissimilar images even when images are unclear.

The paper is organized as follows. In the next section an overview of state-of-the-art dis(similarity) measures is reported. Section 3 introduces the *ACSM* measure proposed in [10]. Section 4 studies its metric properties. Section 5 reviews the algorithm to compute it, and proposes the use of a suffix tree data structure to reduce its complexity. In Section 6 the experimental results are reported, showing that the measure is very competitive with respect to some other considered contestant measures. Section 7 discusses the advantages of *ACSM* with respect to the other state-of-the-art measures. Section 8, finally, concludes the paper and gives some suggestions for future work.

2. Related work

Many different measures have been defined to compute the (dis)similarity between images. An extensive overview and performance comparison can be found in [11]. A broad classification categorizes them into three main groups [3]: geometric measures, information theoretic measures, and statistical measures.

Geometric measures represent an image $I=\{x_1,\dots,x_n\}$ as a vector, where each x_i is an image pixel mapped to a feature in the feature space. The feature space is obtained through a feature extraction process, and includes information such as color, position, shape, or texture. The distance between two images I and J can then be computed by using known measures, such as the cosine distance or the *Minkowski* distance family $d(I,J)=(\sum_{i=1}^{n}|x_i-y_i|^p)^{1/p}$, where $p=1$ corresponds to *city block* norm, if $p=2$ the *Euclidean* distance among pixels, while $p=\infty$ the *Chebyshev* distance. These measures, however, often are inadequate because of the semantic gap between the information extracted from the image and the semantic assigned by a user to the image [12]. In order to overcome this problem, recently, distance metric learning techniques have been proposed [13–17]. The concept of distance metric learning comes from the machine learning and data mining community [18,13]. It is based on the idea of exploiting side information, like class labels and similarity/dissimilarity couples between two objects. In particular, given a set of images, each represented as a feature vector, and a set of pairwise constraints (I,J) indicating if I and J are deemed either similar or dissimilar by domain experts, the goal of distance metric learning is to learn a matrix M such that the Mahalanobis distance $d_M(I,J)=\sqrt{(I-J)^T M(I-J)}$ is optimized. To this end, Davis et al. [13] presented an information theoretic approach, *Information Theoretic Metric Learning* (*ITML*), that formalizes the problem as that of minimizing the relative entropy between two multivariate Gaussians. Chechik et al. [14] proposed an algorithm, *Online Algorithm for Scalable Image Similarity* (*OASIS*), that learns a bilinear similarity measure over sparse representations. Hoi et al. [15] formulated the learning problem as a convex optimization task and proposed a semi-supervised metric learning technique, along with two algorithms to perform it. Li et al. [16] presented a low rank distance metric learning *LRML* that uses visual and textual information instead of similarity/dissimilarity constraints. The metric is formalized as a convex optimization problem, and an algorithm based on gradient method is given. Gao et al. [17] introduced a *Sparse Online Metric Learning* (*SOML*) scheme to learn distance functions from sparse large-scale high-dimensional data, and two algorithms to solve the optimization problem.

Distance metric learning requires a form of supervision, since a training data set must be available in order to allow the method to find out an optimal Mahalanobis matrix. If the training set is small, the solution obtained could generate overfitting, thus reducing the generalization capability to unknown images. Moreover, existing techniques are sensitive to noise and their performances degrade with noisy and small training data, which, often, are common situations in a real-world setting.

Information theoretic measures are derived from the concept of entropy H defined by Shannon [19], $H=-\sum_x p(x)\log_2[p(x)]$, where $p(x)$ is the probability that an image pixel will have the intensity value x, estimated by the image histogram. They are based on pixel intensity distributions and use the histograms of two images, i.e. the number of times each gray value occurs in an image, to determine the similarity between the images to be matched. Several information-theoretic measures have been defined and successfully applied in different contexts, such as medical imaging [2]. Table 1 describes some of them that have been compared with our measure in the experimental results section. Information theoretic measures need the joint probability distribution of two images, computed from their joint histogram. If the images contain noise, the dispersion in the joint probability distribution increases, thus causing mismatch between images.

Statistical measures, such as Pearson correlation coefficient, χ^2 statistics, cross correlation [20], compare probability distributions of image pixels.

More recently, the idea to analyze images at patch level, rather than at pixel level, has been flourishing [6,8,9]. A patch is a small pixel area, typically 3×3 or 4×4, extracted from an image. Given two images I and J, the idea underlying the patch-based methods

Table 1
Information-theoretic (dis)similarity measures.

Measure	Formula	Description
Joint entropy	$JE(X,Y)=H(X,Y)=-\sum_x\sum_y p_{XY}(x,y)\log[p_{XY}(x,y)]$	Entropy of the joint histogram of two images X and Y
Conditional entropy	$CE(X,Y)=H(X\mid Y)=H(X,Y)-H(Y)$	Entropy of the image X given the truth regarding image Y
Mutual information	$MI(X,Y)=H(X)+H(Y)-H(X,Y)$	Uncertainty reduction about one image given the information about the second one
Normalized mutual information	$NMI(X,Y)=\frac{H(X)+H(Y)}{H(X,Y)}$	Normalized version of *MI*
Kullback–Leibler divergence	$KL(q\parallel p)=\sum_x q(x)\log\frac{q(x)}{p(x)}$	Average inefficiency to use the histogram of one image to code another one, where p,q probability distributions of the two images
Arithmetic–geometric mean divergence	$AGM(p,q)=\sum_x\frac{p(x)+q(x)}{2}\log\frac{p(x)+q(x)}{2\sqrt{p(x)q(x)}}$	KL divergence between the arithmetic and geometric means of $p(x)$ and $q(x)$
Jensen divergence	$JD(p,q)=\sum_x\left(q(x)\log\frac{2q(x)}{p(x)+q(x)}+p(x)\log\frac{2p(x)}{p(x)+q(x)}\right)$	Symmetric and more robust modification of KL