Average Common Submatrix: A New Image Distance Measure

Alessia Amelio and Clara Pizzuti

National Research Council of Italy (CNR)
Institute for High Performance Computing and Networking (ICAR)
{amelio,pizzuti}@icar.cnr.it

Abstract. A new information-theoretic distance measure for images is proposed. The measure is based on the concept of average common sub-matrix by considering the pixel matrices associated with the images. An algorithm to compute such a value is described, and its computational complexity analyzed. Experimental results show that the measure is able to discriminate images by correctly reflecting human perception. Furthermore, comparison with state-of-the-art information-theoretic measures, points out that the new measure outperforms these measures in terms of retrieval precision.

Keywords: image retrieval, similarity measure, pattern matching.

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. Many different distance measures have been defined based on Euclidean distance among pixels [8], Hausdorff distance [4], cross correlation [5], and on the concept of entropy [1]. In particular, information-theoretic (dis)similarity measures rely 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 [6].

In this paper a new information-theoretic measure to compute the distance between two images I_A and I_B 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 respectively, and counts the number of square sub-matrices of matrix A that exactly occur in B, to quantify the distance between I_A and I_B. $ACSM$ is an extension in two dimensions of the average common substring (ACS) measure defined in [7] to measure the pairwise distances between sequences. Intuitively, if we have a matrix C on the same alphabet Σ, A is considered more similar to B than to C, if the average area of the sub-matrices of A that occur in B, is larger than the area of the sub-matrices of A occurring in C. In order to evaluate the $ACSM$ measure, two preliminary experimentations have been performed. The former computes distances among images containing similar objects, and shows that the $ACSM$ measure is able to reflect the concept of similar images as perceived by a human, i.e. it assigns smaller distance value to

A. Petrosino (Ed.): ICIAP 2013, Part I, LNCS 8156, pp. 170–180, 2013.

two images considered perceptually similar, and larger distance value to images deemed different. The second experimentation compares the $ACSM$ measure with other seven information-theoretic measures, and shows that $ACSM$ outperforms these measures in terms of retrieval precision.

The paper is organized as follows. In the next section the $ACSM$ measure is introduced, an algorithm to compute it described, and its complexity analyzed. In section 3 the experimental results are reported, showing that the new distance measure is very competitive with respect to the other information-theoretic measures. Section 4, finally, concludes the paper and gives some suggestion on future work.

2 Average Common Sub-matrix Measure

In this section we introduce a new (dis)similarity metric between matrices as extension in two dimensions of the average common substring (ACS) measure defined in [7], and we prove that it can be used to evaluate the distance among data in two dimensions, such as images. Intuitively, consider three matrices A, B and C defined on the same alphabet Σ. A can be considered more similar to B than to C if the average area of the sub-matrices in A that are also sub-matrices in B is larger than the same average area in C. This idea can be formalized as follows.

Let Σ be a finite alphabet, and A a square matrix over Σ of size $N \times N$.

Definition 1. *For any position (i, j) of A, let $A^n_{i,j}$ denote the set of all the square sub-matrices of size $n \times n$, for $1 \leq n \leq min\{i, j\}$, whose bottom right corner occurs at position (i, j). When $n = min\{i, j\}$, the sub-matrix P of A of size $n \times n$ is said maximal.*

Example 1. Figure 1 shows a 5×5 matrix A. Fixed position $(3, 4)$, being $min\{3, 4\} = 3$, the square sub-matrices starting at position $(3, 4)$ are $A^3_{3,4}$ of size 3×3, $A^2_{3,4}$ of size 2×2, and $A^1_{3,4}$ of size 1×1. Thus $A^n_{3,4} = \{A^3_{3,4}, A^2_{3,4}, A^1_{3,4}\}$, with $n = 1, 2, 3$. $A^3_{3,4}$ is the maximal sub-matrix.

Definition 2. *Given two matrices A and B over Σ, of size $N \times N$ and $M \times M$ respectively, for any position (i, j) of A, let $P \in A^n_{i,j}$ be the sub-matrix of A of greatest size $r \times r$, for $1 \leq r \leq min\{i, j\}$, that exactly matches a sub-matrix $Q \in B^m_{k,l}$ starting at some position (k, l) of B. The size $r \times r$ of such sub-matrix is called* area *of P and it is denoted as $W(i, j)$.*

Example 2. Figure 2 shows a 5×5 matrix A and a 4×4 matrix B. For position $(3, 5)$ in A, $P = A^2_{3,5}$ is the sub-matrix with the greatest n value that exactly matches a sub-matrix $Q = B^2_{3,3}$ starting at position $(3, 3)$ in B. Consequently, the greatest n value $r = 2$. In fact, $A^3_{3,5}$, that is greater than P and of maximal size ($n = 3$) for position $(3, 5)$, does not match with any sub-matrices in B. The area $W(3, 5)$ of P is thus equal to 2×2.

The average of all these areas of the sub-matrices of A that match sub-matrices of B can be used to define a similarity measure between A and B. Note that m is dependent from n. In fact, if an exact match of size $n \times n$, $n = min\{i, j\}$, does not exists, we consider $A^{n-1}_{i,j}$ and search for an exact match in B of size $m \times m$, where $m = n - 1$, and so on until $n = 1$.

$$A_{[5,5]} = \begin{vmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & \mathbf{0} & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{vmatrix} \qquad A^3_{3,4} = \begin{vmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{vmatrix} \qquad A^2_{3,4} = \begin{vmatrix} 0 & 0 \\ 1 & 0 \end{vmatrix} \qquad A^1_{3,4} = |0|$$

Fig. 1. A matrix A of size 5×5, and the set of sub-matrices $A^n_{3,4}$, $n = 3, 2, 1$ whose bottom right corner occurs at $(3, 4)$

$$A_{[5,5]} = \begin{vmatrix} 1 & 0 & \mathbf{0} & \mathbf{1} & \mathbf{1} \\ 0 & 1 & \mathbf{0} & 0 & 1 \\ 1 & 1 & \mathbf{1} & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{vmatrix} \quad B_{[4,4]} = \begin{vmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{vmatrix} \quad A^3_{3,5} = \begin{vmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{vmatrix} \quad P = A^2_{3,5} = \begin{vmatrix} 0 & 1 \\ 0 & 0 \end{vmatrix} \quad A^1_{3,5} = |0| \quad Q = B^2_{3,3} = \begin{vmatrix} 0 & 1 \\ 0 & 0 \end{vmatrix}$$

Fig. 2. Two input matrices, A of size 5×5 and B of size 4×4. Fixed the position $(3, 5)$ in A, $P = A^2_{3,5}$ is the sub-matrix with the greatest n value that exactly matches a sub-matrix $Q = B^2_{3,3}$ starting at position $(3, 3)$ in B.

Definition 3. *Given two square matrices A and B, the* Average Common Sub-Matrix (ACSM) similarity *of A and B is defined as*

$$S(A, B) = \sum_{i=1}^{N} \sum_{j=1}^{N} W(i, j)/N^2 \tag{1}$$

Thus $S(A, B)$ computes the average of all the areas of the sub-matrices of A and B that match. Given another matrix C, if $S(A, B) > S(A, C)$, then A will be considered more similar to B than to C since the content of A is more embedded in B than in C.

In our basic measure, we need to identify, for any position (i, j) in A, the largest sub-matrix exactly matching some sub-matrix in B. Sometimes, this exact match is available only at a very small sub-matrix level. The problem is that, in some contexts, the similarity evaluation by using this thin granularity could be trivial due to the redundancy of very small patches in common between the two images. So, we introduce a parameterization in the size of the smallest sub-matrices considered in the similarity measure. A parameter α is introduced to fix a lower bound to the size of the sub-matrices.

More formally, the similarity measure is changed as,

$$S_\alpha(A, B) = \sum_{i=1}^{N} \sum_{j=1}^{N} W(i, j)/N^2 \ s.t.\ W(i, j) \geq \alpha \tag{2}$$

Now we derive a distance measure from this similarity measure. For a fixed α, since $S_\alpha(A, B)$ increases if B size increases, analogously to [7], we normalize $S_\alpha(A, B)$ with respect to the size of B by dividing it by $log(M^2)$. We take the inverse of the normalized similarity and then subtract a correction term in order to obtain zero if the two matrices are the same. The distance measure is then defined as,

$$d_\alpha(A, B) = \frac{log(M^2)}{S_\alpha(A, B)} - \frac{log(N^2)}{S_\alpha(A, A)} \tag{3}$$

Note that if $A = B$ the left part of the formula coincides with the right part, thus $d_\alpha(A, A) = 0$. This distance measure is not symmetric, thus we compute,

$$d_s(A, B) = d_s(B, A) = \frac{d_\alpha(A, B) + d_\alpha(B, A)}{2} \tag{4}$$

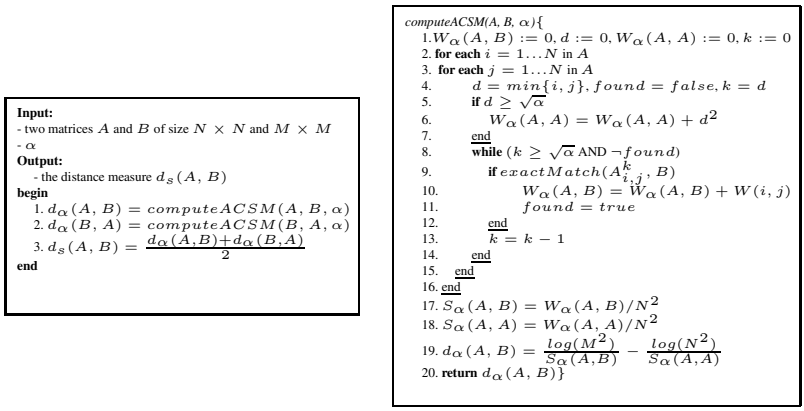


Fig. 3. The ACSM algorithm

that is the final distance measure in 2D. It is worth to note that the granularity is not lost in the definition of the symmetric version, as the same α is adopted for both $d_\alpha(A, B)$ and $d_\alpha(B, A)$.

2.1 The ACSM Algorithm

An algorithm to compute the distance between two matrices, by applying the concept of *Average Common Sub-Matrix*, introduced in the previous section, is shown in figure 3. The main procedure receives as input two matrices A and B, and the granularity parameter α. Then, it computes the $ACSM$ distance between A and B (step 1) and between B and A (step 2), given the α parameter. At the end of the procedure, the symmetric distance $d_s(A, B)$ is evaluated (step 3) by employing the computed $d_\alpha(A, B)$ and $d_\alpha(B, A)$, as in equation (4), and returned as output.

The $ACSM$ distance between two generic matrices A and B, given the α parameter, is calculated by the function *computeACSM(A,B,α)*. Step 1 of the function initializes some variables, including the cumulative area $W_\alpha(A, B)$ of the greatest common sub-matrices between A and B, and the cumulative area $W_\alpha(A, A)$ of the greatest common sub-matrices between A and itself. Observe that the similarity of the matrix A with itself $S_\alpha(A, A)$ can be computed simply by considering that, for each (i, j) in A, the greatest common sub-matrix matching inside A itself is exactly the sub-matrix of maximal extension at that position, whose size is $\geq \alpha$. Consequently, given a position (i, j) in A (steps 2-3), the function computes the area of the sub-matrix of maximal size $d \times d$, with $d = min\{i, j\}$ at that position and updates $W_\alpha(A, A)$ only if this area is larger than or equal to α, i.e. $d \geq \sqrt{\alpha}$ (steps 4-6). After that, the function finds the sub-matrix $A^k_{i,j}$ with the greatest size $k \times k$ that exactly matches a sub-matrix in B (steps 8-14). Searching starts from k equal to $d = min\{i, j\}$ and gradually decreases the k value (step 13). In any case, this value cannot be smaller than $\sqrt{\alpha}$, which is the lower bound for k (step 8). As soon as the greatest sub-matrix at position (i, j) in A matching inside B is detected, the cumulative area $W_\alpha(A, B)$ is augmented (step 10)

with the area $W(i,j)$ of the found greatest common sub-matrix, and the next position in A is considered. Note that if no common sub-matrix is found at position (i,j) within the α bound, the current position of A does not contribute to the computation of the distance measure. Finally, equations (2) and (3) are evaluated (step 17 and step 19), by computing also the similarity of the A matrix with itself $S_\alpha(A,A)$ (step 18), and the value of $d_\alpha(A,B)$ is returned as output of the function.

Theorem 1. *The ACSM algorithm takes $O(M^2N^3)$ time. It can be reduced to $O(M^2N^2log(N))$ time by performing a binary search on d.*

Proof. The cost of the ACSM algorithm is mainly dependent on the $exactMatch$ procedure. Given the current pattern $A^k_{i,j}$ and the input matrix B, it searches $A^k_{i,j}$ in B, by verifying if the pattern exactly occurs into the input matrix. Consider the worst case where $\alpha = 1$ and the size $k \times k$ of the greatest common sub-matrix is equal to 1 for each position (i,j) in A. This means that, for each (i,j) in A, the algorithm will exactly match all the patterns with k varying from $d = min\{i,j\}$ to 1 with the input matrix B, and that the correspondence will be found between the pattern of size $k \times k = 1$ and B. The number of comparisons is:

$$nc = \sum_{i,j}\sum_{k=1}^{d} MC = \sum_{i,j}\sum_{k=1}^{d} M^2 = M^2 \sum_{i,j}\sum_{k=1}^{d} 1 \tag{5}$$

where MC is the execution time of $exactMatch$. A pattern matching procedure for searching a two dimensional pattern P inside the matrix B can take $O(M^2)$ time [2], independently from the size $k \times k$ of the pattern P, with k that varies from 1 to $d = min\{i,j\}$. Because the number of positions (i,j) is N^2 and d is at most equal to N, the overall cost is $O(M^2N^3)$.

However, for a given position (i,j) in A, the cost for searching the largest pattern P exactly matching a sub-matrix of B can be improved by employing a binary search strategy. In particular, starting from the pattern P of size $k \times k$ with $k = \frac{d}{2}$, the matching of P inside B is verified. If a correspondence is found, P occurs in B and eventually also a larger pattern containing P could be there. Consequently, the larger patterns of size $k \times k$, with $d \leq k \leq \frac{d}{2} + 1$ will be checked as possible expansion of P to match with B. Otherwise, looking at the patterns which are larger than P and that contain P is useless. In fact, if P doesn't match inside B, none of the larger patterns containing P can match inside B. Consequently, the smaller patterns of size $k \times k$, with $\frac{d}{2} - 1 \leq k \leq 1$ will be considered as possible reductions of P to match with B. In both cases, the process will start from the new middle points of the two intervals and it will continue, until the greatest common sub-matrix is found at that position. By performing this binary search along d, the number of patterns of size $k \times k$ to match with B for each position (i,j) in A, is reduced to $log(N)$. Consequently, nc is:

$$nc = \sum_{i,j} M^2 log(N) \tag{6}$$

Because the number of positions (i,j) is N^2 in A, the overall cost can be reduced to $O(M^2N^2log(N))$.

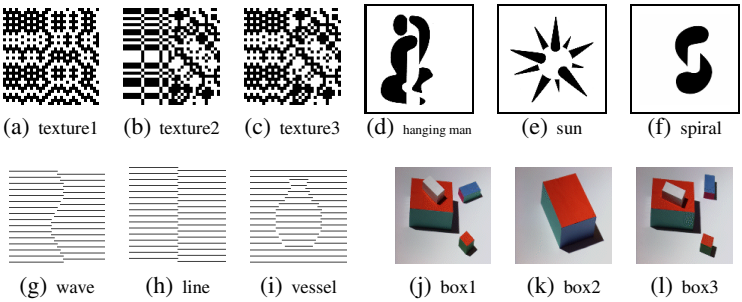


Fig. 4. The test images: the distance is computed for the triples (a-c), (d-f), (g-i) and (j-l)

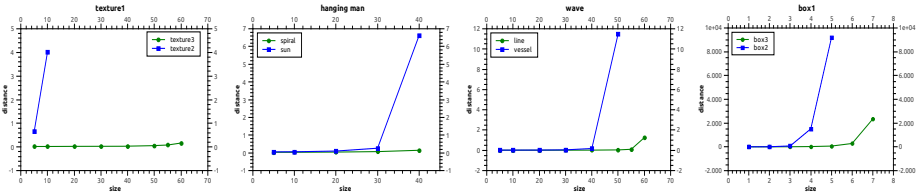
2.2 Accelerating the ACSM Procedure

Recall that, for each position (i, j) in A, the ACSM algorithm finds the largest sub-matrix exactly matching some sub-matrix in B. In order to find, for each position (i, j) in A, the largest sub-matrix matching inside B, a generalized suffix tree in two dimensions can be constructed by employing the *Lsuffix tree* for a matrix [3], generalized for a set of matrices $\{A^1...A^s\}$, each of size $n_i \times n_i$, $1 \leq i \leq s$. In this case, the set of matrices is composed of A and B. The generalized Lsuffix tree is a compacted trie representing the set of all the square sub-matrices of both A and B matrices in a linearized form. Visiting properly this trie, the sub-matrices of maximal extension in A that are also sub-matrices in B, for each position (i, j) in A, can be discovered. This is mainly because a path from the root to a leaf node in the tree represents a sub-matrix starting at a given position inside A or B, and all the positions inside the two matrices are considered. This procedure would reduce the execution time from $O(M^2N^2log(N))$ to $O(N^2 + M^2)$, which is linear in the size of the input images, i.e. the area of the matrices.

3 Experimental Results

In this section preliminary tests to evaluate the proposed distance measure are presented. In particular, two kinds of experimentations have been performed. The former aims at assessing the correspondence between the distance values computed and the visual perception of a human. The latter quantitatively compares the capability of the $ACSM$ measure, with respect to state-of the art similarity measures, in finding images belonging to the same class of a given query image.

The test images used are extracted from the online database of the Computer Vision Group, University of Granada, freely available at http://decsai.ugr.es/cvg/dbimagenes/. This database contains gray level and color images of various size. For our experimentation, we used gray level illusory and color miscellaneous images of size 128×128. Without loss of generality, the size of the selected images is always the same. Gray level illusory images represent synthetic objects with some recurrent patterns inside, and relevant shapes useful for testing the effectiveness of the distance measure. Color miscellaneous images consist of real world objects under different poses, helpful for



(a) (b) (c) (d)

Fig. 5. $ACSM$ distance for different values of $\sqrt{\alpha}$ between (a) $texture1$ with $texture2$ and $texture3$, (b) $hanging\ man$ with sun and $spiral$, (c) $wave$ with $line$ and $vessel$, and (d) $box1$ with $box2$ and $box3$

evaluating the robustness of the distance measure. The used images have been manually grouped into five classes: *textures*, *symbols*, *lines*, *boxes* and *carafes*.

3.1 Human Perception Evaluation

An important characteristics of a distance measure between images is that it should reflect human perception, i.e. images deemed more similar than others by a human observer should have smaller distance among them. In order to evaluate the ability of the $ACSM$ measure to discriminate according to the visual similarity, we considered four out of the five different classes of objects, namely *textures*, *symbols*, *lines*, and *boxes* with three images each. Then we chose one target image out of the three images and computed the distance between the target image and the other two. Figure 4 shows the four image classes, the first of each is selected as the target image. A visual inspection of the figure clearly points out that *texture1* (Figure 4(a)) is more similar to *texture3* than to *texture2*, *hanging man* (Figure 4(d)) is more similar to *spiral* than to *sun*, *wave* (Figure 4(g)) is more similar to *line* than to *vessel*, and *box1* (Figure 4(j)) is more similar to *box3*, than to *box2*. This insight is confirmed by the computation of the $ACSM$ distance at different granularity values, as depicted in Figure 5. In all cases we can observe that the distance measure increases as the minimum size of the patches grows, until it goes to infinity. The motivation is that the distance between two images depends on the exact match of increasingly large patches. If such an exact match does not exists, then the value of formula (2) becomes zero, and consequently, the value of distance as computed in (3) is infinity. Figure 5(a) shows that the distance between $texture1$ and $texture3$, with α varying from 5×5 to 60×60, is lower than that between $texture1$ and $texture2$, and it slightly increases for larger values of α. This is due to the presence of recurrent regular and compact patterns shared by the first and the third image, even for sub-matrices of larger size. On the contrary, even if $texture1$ and $texture2$ have common small patches, if α is above 10×10, there is no overlap between the two objects, thus their similarity is zero, and, consequently, their distance grows to infinity. As regards the *hanging man* figure, the distance with sun and $spiral$ becomes to be distinguished for sub-matrices of size larger than 20×20, and clearly returns a higher similarity between *hanging man* and $spiral$ for $\alpha \geq 30 \times 30$. The necessity of higher values of α comes from the presence in both sun and $spiral$ of many small pure black and white parts that overlap and that should not contribute to the similarity evaluation.

They are no more overlapping only if areas of larger size are considered. In such a case the value computed for the distance allows to correctly discriminate the shape of the inner objects and, consequently, the more similar object. An analogous behavior can be observed in Figure 5(c). In this case $wave$, $line$, and $vessel$ are rather similar among them, thus a granularity of 40×40 is necessary in order to obtain a higher distance between $wave$ and $vessel$. It's important to observe that, if small sub-matrices are also included, the difference in distance computed between $wave$ and $vessel$ and between $wave$ and $line$ is low. In fact, the background of all the three images is almost the same. However, visually *wave* and *line* contain a curved and straight line splitting the image area in two vertical parts, while in *vessel* two lines "draw" something similar to a vessel and split the image in two concentric regions. Finally, Figure 5(d) shows the distance values for $boxes$. These images are color images having an alphabet of very large size. The high variability in the pixel values drastically reduces the probability to have an exact match between two sub-matrices if their size is large. So, differently from the previous tests, we chose a range of small values of the α parameter for the computation of ACSM distance. In particular, because $\sqrt{\alpha}$ is fixed between 1 and 7, α is between 1 and 49. *box1* and *box3* images contain some details that are absent in *box2*, that consists of a single box. Furthermore, in *box1* and *box3* the same internal objects are placed in a different position, and some of them are rotated or located under a different perspective. Although the objects around the boxes are only small details, their presence influences the computation of the distance. In fact, *box1* appears as more distant from *box2* than from *box3*, for α values between 3×3 and 5×5.

3.2 Comparative Evaluation

In this section we compare the $ACSM$ measure with other seven standard similarity measures well known in Information Theory: Joint entropy, Conditional entropy, Mutual information, Normalized mutual information, Kullback-Leibler divergence, Arithmetic geometric mean divergence, Jensen divergence [6].

The performance index adopted to compare $ACSM$ and the above measures is the *retrieval precision* used in *content-based image retrieval*, and employed by Tourassi and Harrawood [6] in the medical context.

(a) (b) (c) (d) (e) (f)

Fig. 6. Images representing the queries. Images (a) and (b) belong to *symbols* class, image (c) to *textures* class, image (d) to *lines* class, image (e) to *boxes* class and image (f) to *carafes* class.

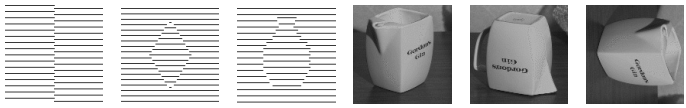
As pointed out in [6], relevance in image retrieval can be of two types: $visual$ and $semantic$. A retrieved image I_R is considered relevant if it belongs to the same class of the query image I_Q. Since the precision depends on the query, precision results are averaged across multiple queries. In our case, the concept of relevant retrieved image that belongs to the correct class is interpreted as follows. We consider six query images belonging to the five different classes. The queries and the classes are reported in Figure

(a) (b) (c) (d) (e) (f) (g)

Fig. 7. Images belonging to the *symbols* class. They represent abstract objects, although some of them are similar to objects of real life, such as sun (d-g), spirals (e), columns (b).

(a) (b) (c) (d) (e) (f) (g) (h) (i) (j)

Fig. 8. Images belonging to the *textures* class (a-e). They represent synthetic textures. Images belonging to the *boxes* class (f-j).



(a) (b) (c) (d) (e) (f)

Fig. 9. Images belonging to the *lines* class (a-c). Two different figures in images (b) and (c) are depicted by the same lines background. Images belonging to the *carafes* class (d-f).

6. In particular, two query images (Figures 6(a-b)) belong to the class $symbols$, and there is one query image for each of the classes $textures$, $lines$, $boxes$, and $carafes$ (Figures 6(c-f)). For each class we examine a set of particularly significative images. For the class $symbols$ we have seven images (Figure 7), for $textures$ five images (Figure 8 (a-e)), for the classes $lines$ and $carafes$ three images each are considered (Figure 9), and for the class $boxes$ five images (Figure 8 (f-j)). As regards the total number of retrieved images, this number depends on the number of images contained in the query class. In order to detect the number of relevant images, for each query image I_Q, the similarity between I_Q and each image I of the five classes is computed. After that, the top K most similar images are selected. This procedure is performed for each similarity measure and all the similarity measures are evaluated by counting how many of the top K images belong to the query class.

Table 1 reports the average retrieval precision obtained by averaging on multiple query images for the top K retrieved images. Observe that the number of images in each class is different. Consequently, we computed the retrieval precision by averaging on all the queries, for $K = 1, 2, 3$, because some query classes don't contain more than 3 images. Then, for $K = 4, 5$, we considered queries (a), (b), (c) and (e) in Fig. 6, because their classes contain at least 5 images. Finally, for $K = 6, 7$, we averaged on queries (a) and (b) in Fig. 6, because they are the only queries whose number of images in the corresponding class is at least 7. The table points out that the $ACSM$ measure obtains the same precision of *Arithmetic-geometric mean divergence*, *Jensen divergence*, and *Kullback-Leibler divergence* for the first top 1 and 2 most similar images. In all the other cases $ACSM$ outperforms the other measures for increasing values of K. These results show that the new measure, based on the concept of average common sub-matrix is able to better discriminate (dis)similar images with respect to state-of-the-art measures.

Table 1. Average retrieval precision achieved by multiple similarity measures: Joint entropy (JE), Conditional entropy (CE), Mutual information (MI), Normalized mutual information (NMI), Arithmetic-geometric mean divergence (AGM), Jensen divergence (JD), KL divergence (KL), Average common submatrix distance (ACSM)

k	JE	CE	MI	NMI	AGM	JD	KL	ACSM
1	0.34	0.50	0.50	0.67	**1.00**	**1.00**	**1.00**	**1.00**
2	0.34	0.58	0.58	0.84	**1.00**	**1.00**	**1.00**	**1.00**
3	0.29	0.50	0.50	0.72	0.89	0.89	0.89	**0.94**
4	0.19	0.50	0.50	0.75	0.69	0.69	0.75	**0.94**
5	0.25	0.45	0.45	0.65	0.60	0.60	0.65	**0.80**
6	0.50	0.00	0.00	0.34	0.42	0.42	0.50	**0.67**
7	0.43	0.00	0.00	0.00	0.50	0.43	0.43	**0.58**

4 Conclusions and Future Work

A new information-theoretic distance measure for two dimensional matrices A and B of symbols has been proposed. The measure is based on the concept of average common sub-matrix, and considers the number of square sub-matrices of matrix A that exactly occur in B, to quantify the distance between the two matrices. The $ACSM$ measure has been applied to compute the similarity between images. Preliminary experimental results showed that $ACSM$ outperforms other information-theoretic similarity measures well known in the literature, by obtaining higher precision values in finding images belonging to the same class of query images. The $ACSM$ distance requires as input parameter the granularity level α. Experimentations have pointed out that, if images to compare are visually very different, small values of α are necessary to better discriminate similar images. Future work will extend the ACSM approach to rectangular matrices. Furthermore it will realize a more efficient implementation of the algorithm, by taking in account also an approximate matching between sub-matrices, and investigating more deeply the invariance of the distance measure to object rotation and scaling.

Acknowledgements. This work has been partially realized while the first author was visiting Georgia Institute of Technology, under the supervision of Prof. Alberto Apostolico. The work has been supported by the project *MERIT : MEdical Research in Italy*, funded by MIUR.

References

1. Cover, T.M., Thomas, J.A.: Elements of Information Theory. Wiley, New York (1991)
2. Crochemore, M., Gasieniec, L., Rytter, W., Plandowski, W.: Two-dimensional pattern matching in linear time and small space. In: Mayr, E.W., Puech, C. (eds.) STACS 1995. LNCS, vol. 900, pp. 181–192. Springer, Heidelberg (1995)
3. Giancarlo, R.: A generalization of the suffix tree to square matrices, with applications. SIAM Journal on Computing 24(3), 520–562 (1995)
4. Huttenlocher, D., Klanderman, G.A., Rucklidge, W.J.: Comparing images using the hausdorff distance. IEEE Trans. on Pattern Analysis and Machine Intelligence 15(9), 850–863 (1993)
5. Pratt, W.K.: Digital Image Processing. Wiley, New York (1991)

6. Tourassi, G.D., Harrawood, B.: Evaluation of information-theoretic similarity measures for content-based retrieval and detection of masses in mammogragrams. Medical Physics 34(1), 140–150 (2007)
7. Ulitsky, I., Burstein, D., Tuller, T., Chor, B.: The average common substring approach to phylogenomic reconstruction. Journal of Computational Biology 13(2), 336–350 (2006)
8. Wang, L., Zhang, Y., Feng, J.: On the euclidean distance of images. IEEE Trans. on Pattern Analysis and Machine Intelligence 27(8), 1334–1339 (2005)