Multimed Tools Appl
DOI 10.1007/s11042-015-2523-7

# Aligning shapes for symbol classification and retrieval

**Sebastiano Battiato · Giovanni Maria Farinella · Oliver Giudice · Giovanni Puglisi**

Received: 8 July 2014 / Revised: 25 December 2014 / Accepted: 18 February 2015


**Abstract** This paper proposes a method able to exploit peculiarities of both, local and global shape descriptors, to be employed for shape classification and retrieval. In the proposed framework, the silhouettes of symbols are firstly described through Bags of Shape Contexts. The shape signature is then used to solve the correspondence problem between points of two shapes. The obtained correspondences are employed to recover the geometric transformations between the shape to be classified/retrieved and the ones belonging to the training dataset. The alignment is based on a voting procedure in the parameter space of the model considered to recover the geometric transformation. The aligned shapes are finally described with the Blurred Shape Model descriptor for classification and retrieval purposes. Experimental results demonstrate the effectiveness of the proposed solution on two classic benchmark shape datasets, as well as on a large scale set of hand sketches composed by 20,000 examples distributed over 250 object categories.

**Keywords** Shape recognition · Shape retrieval · Symbol classification · Alignment

S. Battiato · G. M. Farinella · O. Giudice (✉)
Image Processing Laboratory, Department of Mathematics and Computer Science,
University of Catania, Catania, Italy
e-mail: giudice@dmi.unict.it

S. Battiato
e-mail: battiato@dmi.unict.it

G. M. Farinella
e-mail: gfarinella@dmi.unict.it

G. Puglisi
Dipartimento di Matematica e Informatica, University of Cagliari, Cagliari, Italy
e-mail: puglisi@unica.it

Published online: 13 March 2015

## 1 Introduction

The shape is a fundamental feature used by human visual system for object classification and recognition. The human visual system is one inspiration for investigating computer based shape classification algorithms, although researchers do not attempt to directly emulate it in approach, but merely in the outcome [14]. Motivations beyond that of pure scientific curiosity are provided by several important applications in which the shapes have been demonstrated to be a fundamental feature to solve the task under consideration (e.g., pedestrian detection, road symbols recognition, etc.) [1, 10, 11, 16]. Different works have been presented in literature for both, representation and classification of shapes. An overview of the most popular methods for shape description and classification is reported in [7] and [25]. Shape representation approaches can be grouped in two main classes [8]: local and global. Local methods describe shapes through a set of local properties (e.g., the context of each point of the shape [5]), whereas global techniques consider the shape as a whole entity by encoding general information of the shape (e.g., the statistical variance of a shape with respect to shape prototypes [21, 22]). Recent works combine both local and global features to obtain a more discriminative representation of the shape in order to improve the accuracy in shape classification and retrieval [2, 13, 15].

In this paper we propose a novel framework able to combine both, the local properties of Shape Context descriptor (SC) to solve the correspondence problem during the geometric alignment of shapes, as well as the global representation of the aligned shapes obtained through Blurred Shape Model (BSM). First, shapes are described making use of Bags of Shape Contexts to solve correspondence problem between points of two shapes. Starting from the computed correspondences, shapes are aligned by using a voting procedure in the parameter space of the considered model in order to filter out the geometric transformations [4, 18]. Finally, the aligned shapes are described with the Blurred Shape Model descriptor [10] and classified/retrieved by employing the K-nearest neighbor algorithm with $\chi^2$ test statistic. Experiments performed on the seventy-class MPEG-7 dataset, as well as on the seventeen-class symbols dataset used in [10], demonstrate that the proposed strategy outperforms both approached from which our solution originates: Shape Context (SC) [5] and Circular Blurred Shape Model (CBSM) [10]. The proposed approach is also tested on a large scale dataset of hand sketches composed by 20,000 examples distributed over 250 object categories [9]. The performances achieved by our method outperform the current state-of-the-art method [9] on the large hand sketches dataset.

The proof of concept of the proposed approach has been presented in [3]. Compared to [3], we have improved the algorithm in terms of robustness with respect to scale changes. Moreover, the performance of the proposed approach with respect to perturbations such as rotation, scale and shear has been tested. Finally, several tests and comparisons with state of the art methods, considering the retrieval scenario and a large dataset, have been performed.

The remainder of the paper is organized as follows: Section 2 presents related works. Section 3 describes the proposed framework. Section 4 details the experimental settings and reports the obtained results. Finally, conclusions and avenues for further research are given in Section 5.

## 2 Related works

In this paper we consider the problem of combination of local and global properties for classification purpose. The combination of local and global peculiarities can be obtained at level of the classifier [19] or at level of the features to be exploited in the classification problem to be solved [24]. We focus on the combination of local and global shape descriptors to address the problem of shape classification and retrieval.

Shape Context (SC) [5] is one of the most popular descriptor of points belonging to a shape. This descriptor encodes local properties of a point and it has been extensively and successfully exploited for shape matching and retrieval purposes due to its robustness in finding correspondences between point sets sampled from different shapes. Shape Context describes each point (sampled from the internal or external contour of an object) through a distribution of vectors originating from the considered point to all the other sample points of the shape. This description of a point, called Shape Context, is inherently insensitive to translation and small perturbations of parts of the shape. The Shape Context can be opportunely normalized to obtain invariance to rotation and scale. Moreover, it has been empirically demonstrated that the descriptor is robust to occlusions, noise, and outliers. Given sets of shape contexts obtained from different shapes, the matching problem is reduced to a weighted bipartite matching problem. In [5] the matching problem is solved by employing the Hungarian algorithm [12]. Once the correspondences between points on two shapes are obtained, a plane transformation (e.g., thin plate spline model) is used to map points from one shape to the other. For classification purposes, a K-nearest neighbor algorithm is used together with a shape distance computed as weighted sum of shape contexts $\chi^2$ distance, image appearance distance, and bending energy (see [5] for all related details).

On the other hand, the Blurred Shape Model (BSM) has been recently introduced in literature as global descriptor for shapes [10]. BSM is able to represent the spatial arrangement of shape parts through a correlogram structure from the center of the shape composed by regions defined with circles and sectors. The descriptor is computed by considering a set of points belonging to the contour of an object. It considers the distances from a point to the centroids of the corresponding correlogram region and from the closest ones. Those distances are normalized taking into account of the sum of total distances. The inverse of the normalized distances is then added to the corresponding positions of the distribution associated to the involved correlogram regions. This makes the descriptor more robust to geometric deformations. Rotating the correlogram by taking into account of the region with predominant density, the rotation invariance of the BSM descriptor is obtained. In this last case the descriptor is called Circular Blurred Shape Model (CBSM) [10]. Since the CBSM descriptor encodes a distribution, it is straightforward the exploitation of the $\chi^2$ test statistic as measure between two descriptors. The measure is coupled with the K-nearest neighbor algorithm for shape classification purposes.

In the last years other new shape descriptors have been presented with features of robustness to affine and non-linear transformation such as [20, 23] or others optimized for the recognition and retrieval task on hand-written sketches such as [6, 9]. The benefit of the combination of both local and global features to obtain a more discriminative representation of the shape has been demonstrated for shape classification and retrieval [2, 13, 15]. Here we propose to use local shape descriptors for the alignment step. Then, when geometrical variabilities are filtered out, a global descriptor is employed for the final representation of the shape to be recognized.

## 3 Proposed framework

The proposed solution aims to combine the local peculiarities of shape context descriptor (SC) [5] together with the global representation obtained through Blurred Shape Model

(BSM) [10]. It is worth noting that a simple combination, in cascade, of Shape Context and Blurred Shape Model does not provide satisfactory results as pointed out by the experiments described in Section 4. Due to the considered model (thin plate spline), Shape Context is able to transform a shape to a very different one. Considering for example a template and a test image belonging to different classes, the alignment procedure employed by SC transforms the test image in a novel image very similar to the target image (see Fig. 1). Computing the BSM over the image provided by the SC does not provide hence any useful information. This problem can be mitigated by modifying the standard parameter setting proposed in [5] in order to limit the variability of the thin plate spline transformation. However, also in this case, the overall performance is not satisfactory (see Section 4).

Taking into account the considerations above, in this paper we propose to combine local (Shape Context) and global (Blurred Shape Model) descriptors by exploiting a robust registration algorithm [4] for shapes alignment. The alignment algorithm proposed in [5] is replaced by the proposed one which is based on affine transformations and is more appropriate to combine SC and BSM descriptors.

The workflow of our shape classification method is shown in Fig. 2. As first step, shapes are described as a Bags of Shape Contexts taking into account of a pre-computed vocabulary. The vocabulary is obtained by clustering the set of Shape Contexts [5] computed from points sampled on contours of training shapes. Specifically, the set of centroids obtained with the clustering procedure becomes the codebook used to represent training and testing shapes for

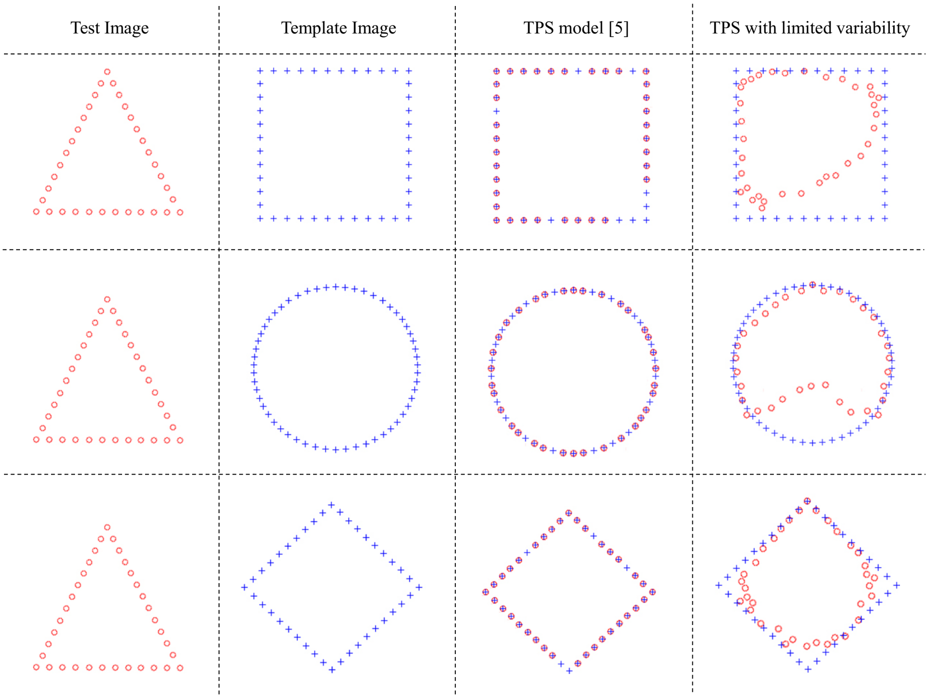



**Fig. 1** Simple combinations of SC [5] and BSM [10]. Shape Context with thin plate spline model transforms a shape in a very different one. Although belonging to different classes, the alignment procedure employed by SC transforms the test image in a novel image too similar to the target. Although the variability of the thin plane spline (TPS) model can be limited by modifying the standard parameter setting used in [5], the overall performance is not satisfactory (see Section 4)

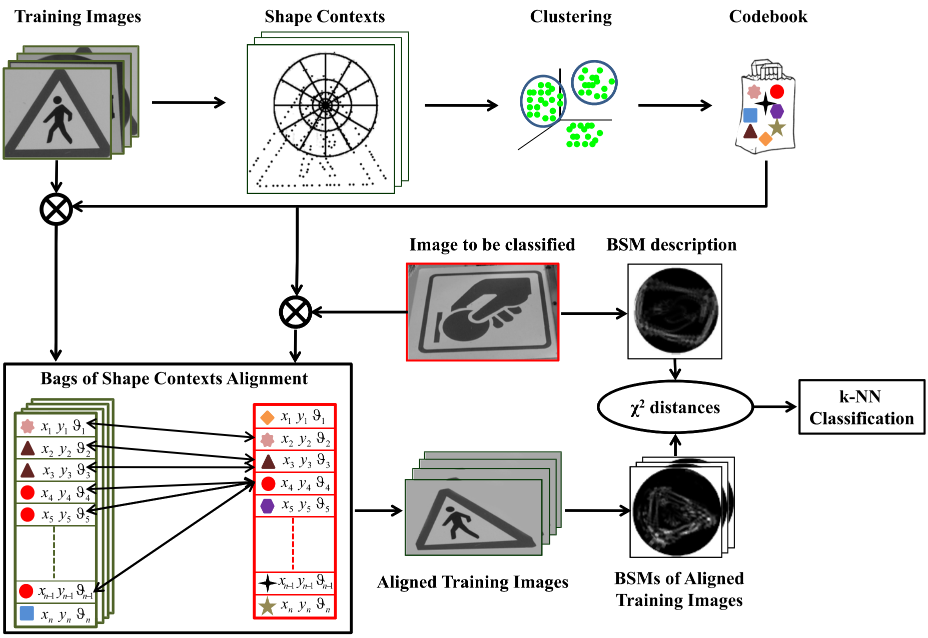



**Fig. 2** Overall schema of the proposed approach. A Bag of Shape Contexts is built and used in combination with BSM to properly classify the input image. A fundamental contribution is provided by the alignment process between Bags of Shape Contexts before exploiting the global BMS descriptor

the alignment step. From each training/test shape (extracted with Canny edge detector on the corresponding image), $n$ points ($n = 100$ in our tests) are uniformly sampled and described with the Shape Context descriptor. Then each point $p_i$ ($i = 1, \ldots, n$) of the considered shape, is associated to a tuple with four components: the $id$ label corresponding to the closest centroid belonging to the pre-computed codebook of shape contexts, the direction $\theta$ of the tangent at that point, and the point coordinates $(x, y)$. Hence, each training/test shape is represented with a signature $S = \{(id_i, x_i, y_i, \theta_i)|i = 1, \ldots, n\}$ which is exploited for alignment purposes.

As second step, the shapes of training set are aligned with respect to the shape to be classified/retrieved. To this purpose, we use the registration procedure detailed in [4]. The entries of the training signatures and the one related to the signature of the shape to be classified/retrieved are matched by considering their $id$ values. To cope with wrong matchings (i.e., outliers), a robust estimator has to be employed. A filtering is performed considering the differences between tangent orientations (i.e., $\theta$) of matched entries. After this preliminary elimination of outliers, the set of point pairs $((x_{test}; y_{test}); (x_{training}; y_{training}))$, corresponding to matched entries, is aligned by considering a similarity transformation model:

$$x_{training} = x_{test}\sigma \cos\alpha - y_{test}\sigma \sin\alpha + T_x \tag{1}$$

$$y_{training} = x_{test}\sigma \sin\alpha + y_{test}\sigma \cos\alpha + T_y \tag{2}$$

In order to recover the geometric transformations defined by the parameters ($\sigma$, $\alpha$, $T_x$, $T_y$), a robust estimator based on a voting procedure in the parameter space of the similarity transformation model [4] is used. Specifically, each matching votes for a line in the parameter space (Fig. 3). By analysing the densest region of this space, a first estimation of the transformation parameters can be obtained. All the matchings voting for the abovementioned solution can be then considered as inliers (see [4] for all mathematical details).

In this paper, a further filtering has been performed with respect to the registration strategy proposed in [4] where the outlier rejection (see Fig. 3) is performed in a reduced space $(\alpha, T_x, T_y)$. Actually, some matching pairs, close in this reduced space, could be related to very different scale parameter $\sigma$. A further filtering is then performed at the end of the algorithm by simple computing the median value of the already estimated scale parameters.

These matching pairs (inliers) are then used to estimate the parameters of the affine transformation:

$$x'_{training} = ax'_{test} + by'_{test} + T_x \quad (3)$$

$$y'_{training} = cx'_{test} + dy'_{test} + T_y \quad (4)$$

where $(x'_{test}, y'_{test})$ and $(x'_{test}, y'_{test})$ are training and test points considered as inliers and $a$, $b$, $c$, $d$, $T_x$, $T_y$ are the six real parameters of the affine transformation. The estimation of the affine parameters is performed by using the Least Squares algorithm. Outliers have been discarded in the previous steps and, considering only inliers, a good accuracy is achieved.

The obtained transformation parameters are used to align training and test shapes. The aligned shapes are then represented with the Blurred Shape Model [10]. Finally, K-nearest neighbor algorithm is used for classification and retrieval purposes (Fig. 2). To establish similarity between shapes we employ the $\chi^2$ measure.

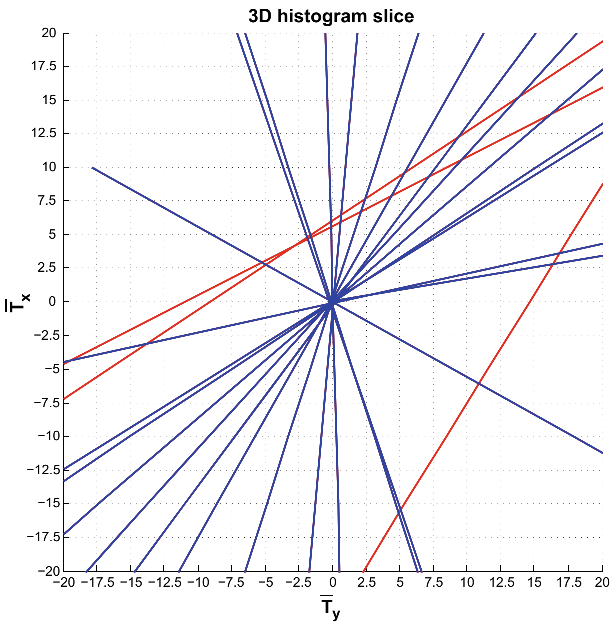



**Fig. 3** A slice of the 3D histogram representing the parameter space $(\alpha, T_x, T_y)$. Each pair of coordinates $(x_{training}, y_{training})$ and $(x_{test}, y_{test})$ votes for a line in the quantized 2D parameter space $(\overline{T_x}, \overline{T_y})$. Lines corresponding to inliers (*blue*) intersect in the bin $(\overline{T_x}, \overline{T_y}) = (0, 0)$, whereas the remaining lines (*red*) are related to outliers

## 4 Experimental settings and results

To properly test the effectiveness of the proposed solution several tests and comparisons with respect to state-of-the-art methods have been performed. In all the tests we considered two different datasets: the public available MPEG-7 CE Shape-1 Part-B dataset which includes 1400 shape samples of 70 different classes, and the 17-class dataset of greylevel symbols used in [10] for testing CSBM. Examples of the shapes contained in both datasets are shown in Fig. 4. The original code of SC provided by the authors has been used in our tests. CBSM approach has been reimplemented as described in [10]. The simple combination in cascade of SC and CBSM has been also considered in our comparisons. To limit the problems described in Section 3 (see Fig. 1) we have modified the standard parameter setting proposed in [5] in order to limit the variability of the thin plate spline transformation.

Tests on each dataset have been repeated five times (5-fold cross-validation procedure) considering 20 % samples of each class as test set and the remaining as training set. The 1-NN has been used for classification purposes. Confusion matrices were recorded at each run and the final classification results are obtained averaging on the results of all five runs. The percentage of the per-class correct classified shapes (diagonal of the final confusion matrix) obtained testing the different approaches on the MPEG-7 dataset is reported in Table 1, whereas the results obtained on the Symbol dataset are shown in Table 2. In both cases the proposed strategy outperforms the other state-of-the-art methods obtaining a good margin with respect to the average accuracy. It is worth nothing that, considering the MPEG-7 dataset, sometimes the proposed method is not the best one. This behaviour can be explained analysing the intraclass shape deformations that, sometimes, the considered affine model

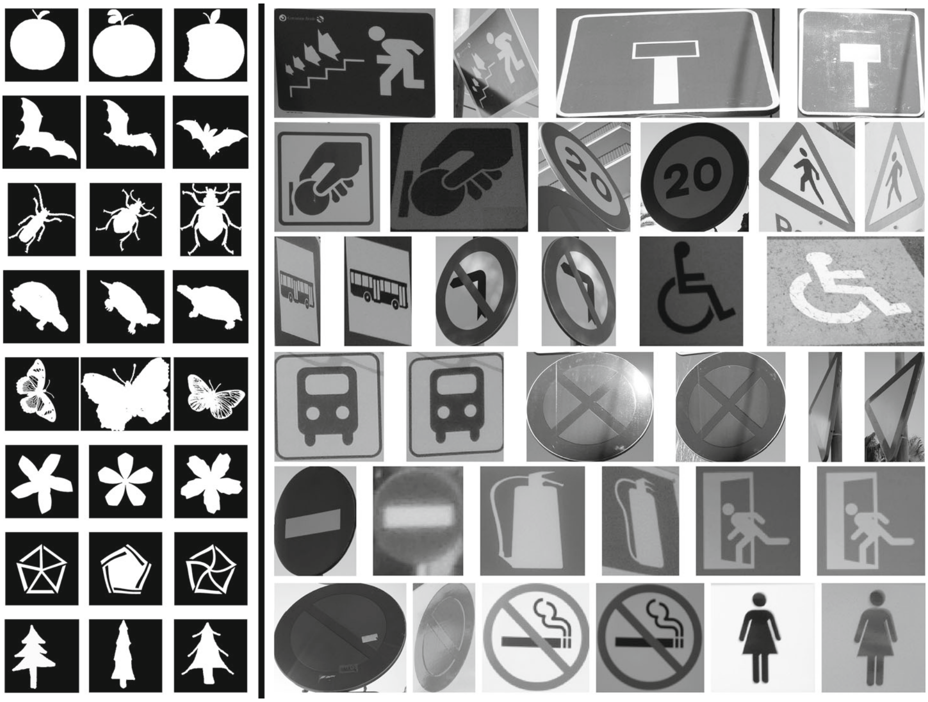



**Fig. 4** *Left*: examples of eight over seventy different classes of the MPEG-7 dataset. *Right*: examples of the seventeen different classes of the Symbol dataset

**Table 1** Per-class percentage of correct classified shapes obtained on the 70-class MPEG-7 dataset

| Method | apple | bat | beetle | bell | bird | bone | bottle | brick | butterfly | camel |
|---|---|---|---|---|---|---|---|---|---|---|
| CBSM [10] | 60.00 | *75.00* | 50.00 | 70.00 | 60.00 | 75.00 | 60.00 | 80.00 | *70.00* | *70.00* |
| SC [5] | 60.00 | 60.00 | 60.00 | *75.00* | *85.00* | 35.00 | 70.00 | 75.00 | *70.00* | 45.00 |
| SC [5] and CBSM [10] | 50.00 | 60.00 | 45.00 | 65.00 | 75.00 | 55.00 | 55.00 | 70.00 | 60.00 | 55.00 |
| Proposed approach | *75.00* | 65.00 | *80.00* | *75.00* | 65.00 | *85.00* | *90.00* | *85.00* | 50.00 | *70.00* |

| Method | car | carriage | cattle | cell phone | chicken | children | chopper | classic | comma | crown |
|---|---|---|---|---|---|---|---|---|---|---|
| CBSM [10] | *85.00* | *80.00* | 80.00 | *75.00* | 70.00 | *70.00* | 35.00 | 65.00 | *75.00* | 70.00 |
| SC [5] | 70.00 | 70.00 | 50.00 | 65.00 | *80.00* | 55.00 | 40.00 | *75.00* | 70.00 | *80.00* |
| SC [5] and CBSM [10] | 70.00 | 70.00 | 60.00 | 65.00 | 75.00 | 65.00 | 30.00 | 70.00 | 65.00 | 65.00 |
| Proposed approach | 75.00 | 60.00 | *85.00* | 65.00 | *80.00* | 60.00 | *80.00* | 60.00 | 70.00 | 75.00 |

| Method | cup | deer | device0 | device1 | device2 | device3 | device4 | device5 | device6 | device7 |
|---|---|---|---|---|---|---|---|---|---|---|
| CBSM [10] | 75.00 | 75.00 | 60.00 | 30.00 | 75.00 | 55.00 | 45.00 | 55.00 | 70.00 | 75.00 |
| SC [5] | 60.00 | 65.00 | 45.00 | 65.00 | *85.00* | 65.00 | *80.00* | 65.00 | *90.00* | 60.00 |
| SC [5] and CBSM [10] | 60.00 | 65.00 | 45.00 | 45.00 | 75.00 | 50.00 | 55.00 | 55.00 | 75.00 | 65.00 |
| Proposed approach | *80.00* | *85.00* | *70.00* | *80.00* | 80.00 | *85.00* | 65.00 | *100.00* | 80.00 | *85.00* |

| Method | device8 | device9 | dog | elephant | face | fish | flatfish | fly | fork | fountain |
|---|---|---|---|---|---|---|---|---|---|---|
| CBSM [10] | *85.00* | *85.00* | 75.00 | *85.00* | 60.00 | 45.00 | 60.00 | *85.00* | *75.00* | 65.00 |
| SC [5] | 50.00 | 55.00 | 60.00 | 55.00 | 75.00 | 55.00 | 65.00 | 70.00 | 70.00 | *70.00* |
| SC [5] and CBSM [10] | 55.00 | 60.00 | 60.00 | 65.00 | 60.00 | 40.00 | 60.00 | 75.00 | 65.00 | 60.00 |
| Proposed approach | 65.00 | 65.00 | *80.00* | *85.00* | *80.00* | *95.00* | *80.00* | 80.00 | 70.00 | *70.00* |

| Method | frog | glas | guitar | hammer | hat | hcircle | heart | horse | horseshoe | jar |
|---|---|---|---|---|---|---|---|---|---|---|
| CBSM [10] | 65.00 | 60.00 | 60.00 | 55.00 | *70.00* | 60.00 | 70.00 | 65.00 | 55.00 | *85.00* |
| SC [5] | 55.00 | 40.00 | 65.00 | 55.00 | 65.00 | *75.00* | 55.00 | 35.00 | 45.00 | 50.00 |
| SC [5] and CBSM [10] | 55.00 | 45.00 | 55.00 | 50.00 | 60.00 | 60.00 | 60.00 | 35.00 | 35.00 | 60.00 |
| Proposed approach | *80.00* | *75.00* | *85.00* | *70.00* | *70.00* | 70.00 | *75.00* | *90.00* | *85.00* | 70.00 |

| Method | key | lizzard | lmfish | misk | octopus | pencil | personal car | pocket | rat | ray |
|---|---|---|---|---|---|---|---|---|---|---|
| CBSM [10] | *80.00* | *85.00* | 60.00 | 55.00 | 55.00 | *85.00* | *75.00* | 70.00 | 60.00 | 75.00 |
| SC [5] | 70.00 | 75.00 | 50.00 | *75.00* | *70.00* | 55.00 | 50.00 | 60.00 | 70.00 | *90.00* |
| SC [5] and CBSM [10] | 65.00 | 75.00 | 45.00 | 55.00 | 55.00 | 60.00 | 55.00 | 65.00 | 60.00 | 75.00 |
| Proposed approach | 70.00 | *85.00* | *80.00* | 60.00 | *70.00* | *85.00* | *75.00* | *75.00* | *85.00* | 70.00 |

| Method | sea snake | shoe | spoon | spring | stef | teddy | tree | truck | turtle | watch |
|---|---|---|---|---|---|---|---|---|---|---|
| CBSM [10] | 75.00 | *75.00* | 65.00 | 50.00 | 80.00 | 50.00 | 60.00 | *70.00* | 70.00 | 70.00 |
| SC [5] | 70.00 | 80.00 | 60.00 | 55.00 | 80.00 | *65.00* | 70.00 | *70.00* | 60.00 | 55.00 |
| SC [5] and CBSM [10] | 70.00 | 70.00 | 55.00 | 45.00 | 80.00 | 50.00 | 60.00 | *70.00* | 60.00 | 55.00 |
| Proposed approach | *80.00* | 70.00 | *80.00* | *80.00* | *85.00* | *65.00* | *75.00* | *70.00* | *85.00* | *95.00* |

The average accuracy of CBSM, SC, the combination of CBSM and SC and the proposed approach are respectively 67.43 %, 63.79 %, 59.36 %, and 76.29 %. Emphasized numbers with italics show the approach with better results

**Table 2** Per-class percentage of correct classified shapes obtained on the 17-class Symbol dataset

| Method | class 1 | class 2 | class 3 | class 4 | class 5 | class 6 | class 7 | class 8 | class 9 |
|---|---|---|---|---|---|---|---|---|---|
| CBSM [10] | 51.00 | 56.00 | 68.00 | 65.00 | 60.00 | 76.00 | *79.00* | 65.33 | 76.00 |
| SC [5] | 56.00 | 56.00 | 64.00 | 74.00 | 72.00 | 70.00 | 59.00 | 77.33 | 64.00 |
| SC [5] and CBSM [10] | 63.00 | 64.00 | 60.00 | 65.00 | 60.00 | 71.00 | 63.00 | *80.67* | 68.00 |
| Proposed approach | *87.00* | *84.00* | *84.00* | *87.00* | *80.00* | *83.00* | 75.00 | 65.33 | *80.00* |

| Method | class 10 | class 11 | class 12 | class 13 | class 14 | class 15 | class 16 | class 17 | Average |
|---|---|---|---|---|---|---|---|---|---|
| CBSM [10] | 80.00 | 28.00 | 16.22 | 42.27 | 90.00 | *100.00* | *96.67* | *100.00* | 67.62 |
| SC [5] | 80.00 | 64.00 | 18.44 | 51.17 | 90.00 | *100.00* | *96.67* | 98.33 | 68.29 |
| SC [5] and CBSM [10] | 80.00 | 52.00 | 14.00 | 50.33 | 75.00 | *100.00* | *96.67* | *100.00* | 66.42 |
| Proposed approach | *100.00* | *68.00* | *26.67* | *54.27* | *95.00* | *100.00* | *96.67* | *100.00* | *79.12* |

The average accuracy of CBSM, SC, the combination of CBSM and SC and the proposed approach are respectively 67.62 %, 68.29 %, 66.42 % and 79.12 %

employed in our technique is not able to properly take into account. The 17-class Symbol dataset, considering images of plane objects taken from different points of view can be better modeled by the affine transformation and the proposed approach obtains the best performance in almost all cases (15 over 17). This demonstrates the effectiveness of the proposed framework which exploits the peculiarities of both approaches from which it has origin.

Further tests have been performed to analyze the behaviour of the proposed solution at varying of the parameters of the classifier. Specifically, the performances related to different values of the $k$ parameter of the k-NN classifier have been tested through cross validation procedure. The MPEG-7 and Symbol datasets have been divided in two parts. The first half samples of each dataset have been used to set the parameter $k$ involved in the classification procedure. To this aim a 5-fold cross validation procedure has been employed, and the average results over five runs of each considered description technique have been used to fix the final $k$ parameter to be used in the classification procedure. The results obtained with this procedure are reported in Fig. 5. The second half samples of the datasets have been used for testing purposes after setting $k$ in correspondence to the best results obtained with cross validation procedure. The final comparison on the test sets is reported in Table 3. The proposed solution outperforms the other methods in all tests obtaining an average accuracy of 96.30 % and 98.04 % for MPEG-7 and Symbol datasets respectively when $k = 5$.

To properly study the robustness of the considered methods at varying of transformation parameters, several tests have been carried out. An image has been selected from each class and several transformation have been artificially performed (scale, rotation, shear). The average over all classes has been hence reported. The shearing transformation is defined as follows:

$$x_r = x_s + k y_s \tag{5}$$

$$y_r = y_s \tag{6}$$

where $(x_s, y_s)$ and $(x_r, y_r)$ are points in the source image $I_s$ and transformed image $I_r$ respectively, and $k$ is the shear parameter.

As shown in Figs. 6, 7 and 8 the proposed solution obtains satisfactory performances at varying of rotation, scale and shear factor and outperforms the other techniques in all the considered transformations.

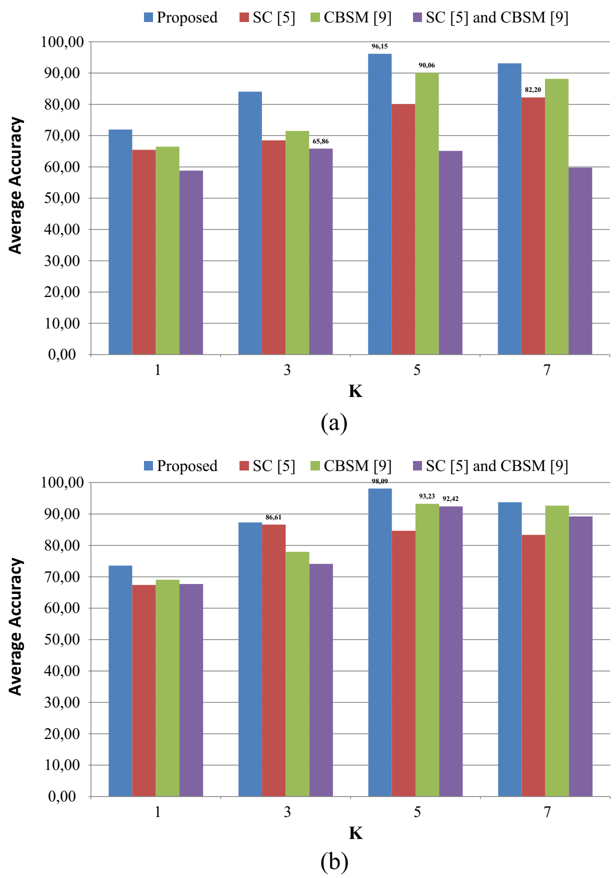
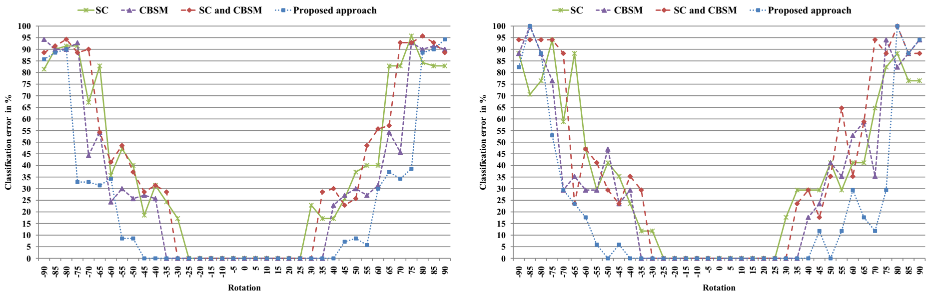



**Fig. 5** Performance evaluations at varying of the $k$ parameter of k-NN classifier for MPEG-7 (**a**) and Symbol (**b**) datasets

To further demonstrate the effectiveness of the proposed framework, retrieval tests have been performed. Considering a specific dataset, for each class an image has been selected as training and the remaining ones as test images to be used to perform the query. This is repeated such that each image belonging to a class is used once in the training set. For example, considering MPEG-7 CE Shape-1 Part-B dataset, we selected 20 pairs of training and test sets made up of 70 and 1330 images respectively.

**Table 3** Performance evaluations considering the best $k$ for each description method

| Method | MPEG | symbols |
|---|---|---|
| SC [5] | 82.90 | 86.76 |
| CBSM [10] | 89.42 | 94.67 |
| SC [5] and CBSM [10] | 66.19 | 90.06 |
| Proposed approach | *96.30* | *98.04* |




(a) (b)

**Fig. 6** Performance evaluation with respect to rotational transformation. The proposed approach outperforms the other methods both on the MPEG-7 CE Shape-1 Part-B dataset (**a**) and 17-class dataset of greylevel symbols (**b**)

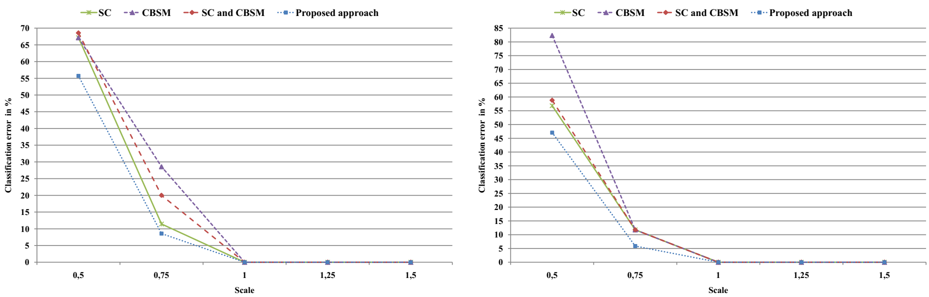
Each query image has been then associated to a list of training images taking into account the different methods to be compared. The retrieval performance has been evaluated with the probability of the successful retrieval $P(n)$ in a number of test queries:

$$P(n) = \frac{Q_n}{Q} \tag{7}$$

where $Q_n$ is the number of successful queries according to $top-n$ criterion, i.e., the number of correct queries among the first $n$ retrieved images, and $Q$ is the total number of queries.

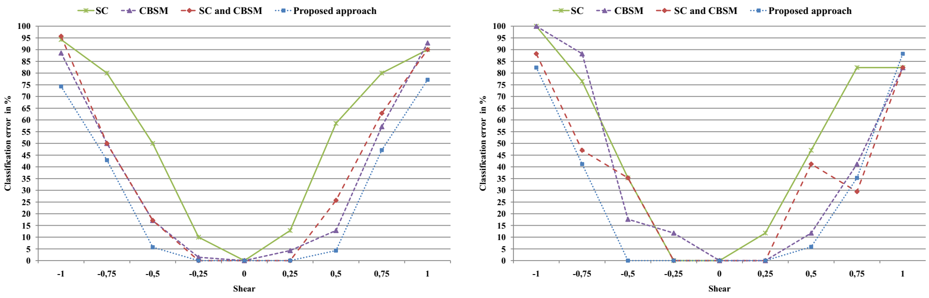
The average of $P(n)$ values with respect to the training and test set pairs is reported in Fig. 9. Results show that the proposed solution achieves a good margin of performances with respect to the other techniques on both the MPEG-7 CE Shape-1 Part-B dataset and the 17-class dataset of greylevel symbols. We also show the precision/recall values at $top - n = 1$ in Table 4. Note that the precision and recall for $top - n = 1$ are equivalent because there is only one correct match for each query.

Finally the retrieval results are evaluated through the mean average precision (mAP) measure, i.e., the area under the precision-recall curve (see [17] for further details). The results obtained on the considered datasets are reported in Table 5 Also in this case, the results are in favor to the proposed method.



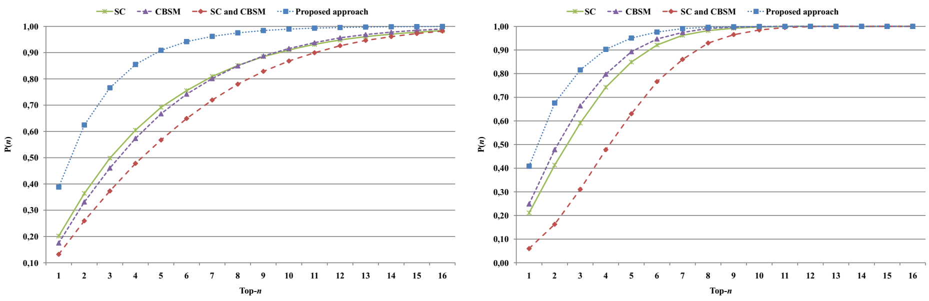

(a) (b)

**Fig. 7** Performance evaluation with respect to scale transformation. The proposed approach outperforms the other methods both on the MPEG-7 CE Shape-1 Part-B dataset (**a**) and 17-class dataset of greylevel symbols (**b**)




(a) (b)

**Fig. 8** Performance evaluation with respect to shear transformation. The proposed approach outperforms the other methods both on the MPEG-7 CE Shape-1 Part-B dataset (**a**) and 17-class dataset of greylevel symbols (**b**)




(a) (b)

**Fig. 9** Retrieval performance evaluation. The proposed approach achieves the best performances both on the MPEG-7 CE Shape-1 Part-B dataset (**a**) and the 17-class dataset of greylevel symbols (**b**)

**Table 4** Precision/recall values on the considered datasets at $top - n = 1$

| Method | Precision/recall MPEG | Precision/recall symbols |
|---|---|---|
| SC [5] | 0.2017 | 0.2110 |
| CBSM [10] | 0.1752 | 0.2486 |
| SC [5] and CBSM [10] | 0.1322 | 0.0601 |
| Proposed approach | *0.3882* | *0.4088* |

**Table 5** Mean Average Precision (mAP) values on the considered datasets

| Method | mAP MPEG | mAP symbols |
|---|---|---|
| SC [5] | 0,408 | 0,453 |
| CBSM [10] | 0,384 | 0,494 |
| SC [5] and CBSM [10] | 0,329 | 0,285 |
| Proposed approach | *0,599* | *0,628* |

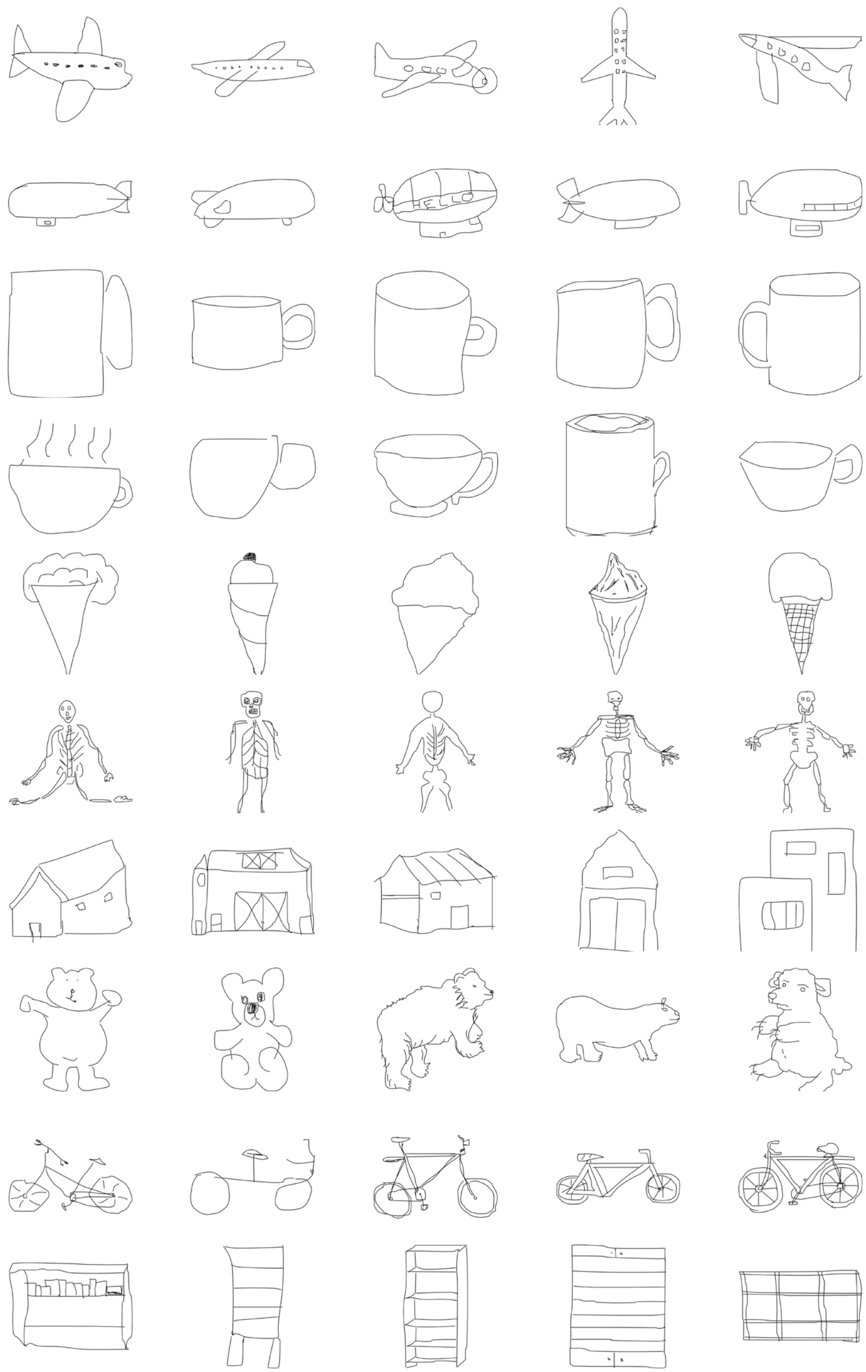


**Fig. 10** Examples of shapes of the hand sketch dataset [9]. Rows correspond to the following ten categories: *airplane*, *blimp*, *mug*, *cup*, *ice-cream-cone*, *human-skeleton*, *barn*, *bear*, *bicycle*, *bookshelf*. The dataset presents high within variability (see 8th row) as well as contains classes with low between vaiability (see 2nd and 3rd rows)

### 4.1 Classification performances on a large sketches dataset

The experiments performed so far demonstrated the performances and the robustness of the proposed method with respect to geometric variability on two classic datasets. To further assess the proposed technique we have performed further tests on a large hand sketches dataset composed by 20,000 examples distributed over 250 object categories [9]. Some examples of sketch sampled at random from the dataset are shown in Fig. 10. At the best of our knowledge, this is the larger labeled dataset of shapes on which the proposed method can be tested. The dataset covers most objects of everyday life. Each category is recognizable from its shape alone and does not require a context for recognition and each class is specific enough for testing purpose (i.e., it does not contain general classes composed by subcategories such as “Animal” or “musical instrument”). Humans are able to recognize correctly on average 73.1 % percent of all 20,000 sketches.

In [9] a method based on bag of feature paradigm (BoW) has been presented. Local histograms of orientations have been considered as main local feature. Both variants hard BoW and soft BoW have been tested considering k-NN and SVM as classifiers. To test their approach the authors used a 3-fold cross-validation method. The final classification accuracy have been obtained averaging over the 3 runs. We compared the proposed approach with respect to the one proposed in [9]. We used a k-NN classifier and SVM in one against all modality to run our tests after representing sketches with our method. The k-NN and SVM parameters have been fixed as in [9] (see original paper for the details). As shown in Fig. 11 the proposed approach outperforms [9] in both k-NN and SVM versions.

### 4.2 Execution time

In order to characterize the proposed solution also in terms of execution times, the average ones related to the main steps of the proposed solution are reported in Table 6. The experiments have been performed on a notebook equipped with a CPU Pentium Core Duo 2.4 GHz and the reported values refer to a single comparison between shapes. To obtain the

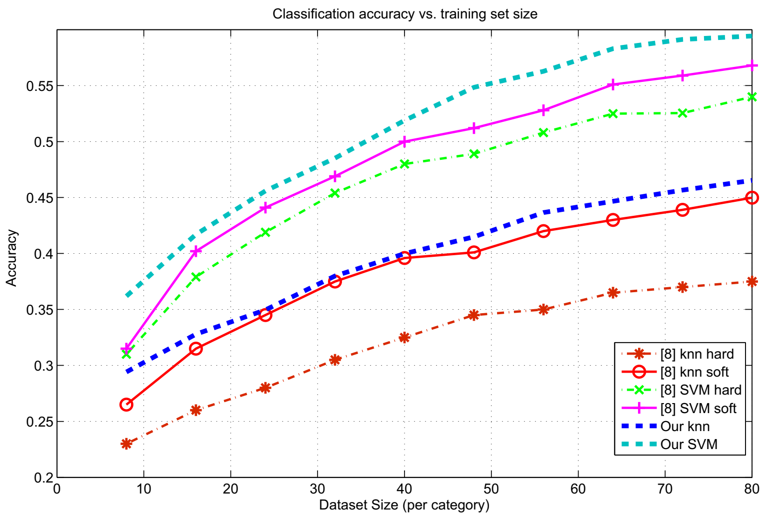



**Fig. 11** Classification accuracy comparison between the proposed approach (Our) Vs. [9] in both k-NN and SVM on growing data-set size

**Table 6** Execution time related to the main steps of the proposed approach

| | Average execution time (ms) |
|---|---|
| BoW representation based on SC | 20 |
| Shape alignment | 158 |
| BSM computation | 18 |
| $\chi^2$ distance computation | 9 |

execution time related to a generic query the reported times have to be multiplied by the number of images in the considered dataset. Note that *BoW representation based on SC*, *BSM computation* and $\chi^2$ *distance computation* steps have been implemented in C whereas *Shape alignment* in Matlab.

## 5 Conclusion and future works

This paper addresses the problem of shape classification and retrieval and proposes a framework to effectively combine local and global shape descriptors. The described technique exploits the peculiarities of the Shape Context in order to locally describe points of a shape. The local descriptors are used to solve the correspondence problem through a bag of feature approach. The shapes are hence aligned with a voting based algorithm. The Blurred Shape Model is then used to represent the aligned shapes in a global way. The assessment of the proposed technique has been performed by considering two representative datasets. The results show the effectiveness of the proposed method which obtains a valuable margin of performance with respect to other state-of-the-art techniques. Moreover tests on a large dataset of hand Sketch composed by 20,000 samples confirmed the superiority of the proposed approach with respect to the current state-of-the-art [9].

## References

1. Azzaro G, Caccamo M, Ferguson J, Battiato S, Farinella GM, Guarnera G, Puglisi G, Petriglieri R, Licitra G (2011) Objective estimation of body condition score by modeling cow body shape from digital images. J Dairy Sci 94(4):2126–2137
2. Bai X, Liu W, Tu Z (2009) Integrating contour and skeleton for shape classification. In: IEEE international conference on computer vision workshops, pp 360–367
3. Battiato S, Farinella GM, Giudice O, Puglisi G (2012) Aligning bags of shape contexts for blurred shape model based symbol classification. In: Proceedings of international conference on pattern recognition (ICPR), pp 1598–1601
4. Battiato S, Farinella GM, Messina E, Puglisi G (2012) Robust image alignment for tampering detection. IEEE Trans Inf Forensics Secur 7(4):1105–1117
5. Belongie S, Malik J, Puzicha J (2002) Shape matching and object recognition using shape contexts. IEEE Trans Pattern Anal Mach Intel 24(4):509–522
6. Caglar T, Berrin Y, Metin TS (2012) Sketched symbol recognition with auto-completion. Pattern Recog 45:3926–3937
7. da Fontoura Costa L, Cesar RM Jr (2009) Shape classification and analysis: theory and practice, 2nd edn. CRC Press, Inc., Boca Raton
8. Daliri MR, Torre V (2008) Robust symbolic representation for shape recognition and retrieval. Pattern Recog 41(5):1782–1798
9. Eitz M, Hays J, Alexa M (2012) How do humans sketch objects? ACM Trans Graph (SIGGRAPH) 31(4):44:1–44:10
10. Escalera S, Fornés A, Pujol O, Lladós J, Radeva P (2011) Circular blurred shape model for multiclass symbol recognition. IEEE Trans Syst Man Cybern 41(2):497–506

11. Farinella GM, Impoco G, Gallo G, Spoto S, Catanuto G, Nava MB (2006) Objective outcome evaluation of breast surgery. In: Medical image computing and computer-assisted intervention—MICCAI 2006. Lecture Notes in Computer Science, vol 4190. Springer, Berlin Heidelberg, pp 776–783
12. Kuhn HW (1955) The hungarian method for the assignment problem. Nav Res Logist Q 2:83–97
13. Lim KL, Galoogahi H (2010) Shape classification using local and global features. In: Pacific-rim symposium on image and video technology, pp 115–120
14. Marr D (1982) Vision: a computational investigation into the human representation and processing of visual information. Henry Holt and Co., Inc., New York
15. McNeill G, Vijayakumar S (2006) Hierarchical procrustes matching for shape retrieval. In: IEEE computer society conference on computer vision and pattern recognition, pp 885–894
16. Munder S, Schnörr C, Gavrila DM (2008) Pedestrian detection and tracking using a mixture of view-based shape-texture models. IEEE Trans Intell Transp Syst 9(2):333–343
17. Philbin J, Chum O, Isard M, Sivic J, Zisserman A (2007) Object retrieval with large vocabularies and fast spatial matching. In: Conference on computer vision and pattern recognition
18. Puglisi G, Battiato S (2011) A robust image alignment algorithm for video stabilization purposes. IEEE Trans Circ Syst Video Technol 21(10):1390–1400
19. Qi GJ, Tian Q, Huang T (2011) Locality-sensitive support vector machine by exploring local correlation and global regularization. In: IEEE international conference on computer vision and pattern recognition, pp 841–848
20. Qi J, Xin F, Zhongxuan L, Yu L, He G (2014) A new geometric descriptor for symbols with affine deformations. Pattern Recogn Lett 40:128–135
21. Velasco-Forero S, Angulo J (2010) Statistical shape modeling using morphological representations. In: International conference on pattern recognition, pp 3537–3540
22. Wang B, Shen W, Liu WY, You XG, Bai X (2010) Shape classification using tree-unions. In: International conference on pattern recognition, pp 983–986
23. Wang J, Bai X, You X, Liu W, Latecki L (2011) Shape matching and classification using height functions. Pattern Recogn Lett 33:134–143
24. Yan Y, Shen H, Liu G, Ma Z, Gao C, Sebe N (2014) GLocal tells you more: coupling GLocal structural for feature selection with sparsity for image and video classification. Comp Vision Image Underst 124:99–109. Large Scale Multimedia Semantic Indexing
25. Zhang D, Lu G (2004) Review of shape representation and description techniques. Pattern Recog 37(1):1–19

**Sebastiano Battiato** received his degree in computer science (summa cum laude) in 1995 from University of Catania and his Ph.D. in computer science and applied mathematics from University of Naples in 1999. From 1999 to 2003 he was the leader of the “Imaging” team at STMicroelectronics in Catania. He joined the Department of Mathematics and Computer Science at the University of Catania as assistant professor in 2004 and became associate professor in the same department in 2011. His research interests include image enhancement and processing, image coding, camera imaging technology and multimedia forensics. He has edited 4 books? and co-authored more than 150 papers in international journals, conference proceedings and book chapters. He is a co-inventor of about 15 international patents, reviewer for several international journals, and he has been regularly a member of numerous international conference committees. Prof. Battiato has participated in many international and national research projects. Chair of several international events (IWCV2012, ECCV2012, VISAPP 2012-2013-2014, ICIAP 2011, ACM MiFor 2010–2011, SPIE EI Digital Photography 2011-2012-2013, etc.). He is an associate editor of the?IEEE Transactions on Circuits and System for Video Technology?and of the?SPIE Journal of Electronic Imaging. Guest editor of the following special issues: “Emerging Methods for Color Image and Video Quality Enhancement” published on EURASIP Journal on Image and Video Processing (2010) and “Multimedia in Forensics, Security and Intelligence” published on IEEE Multimedia Magazine (2012). He is the recipient of the 2011 Best Associate Editor Award of the IEEE Transactions on Circuits and Systems for Video Technology. He is director (and co-founder) of the International Computer Vision Summer School (ICVSS), Sicily, Italy. He is a senior member of the IEEE.

**Giovanni Maria Farinella** received the M.S. degree in computer science (egregia cum laude) from the University of Catania, Italy, in 2004, and the Ph.D. degree in computer science in 2008. He joined the Image Processing Laboratory (IPLAB) at the Department of Mathematics and Computer Science, University of Catania, in 2008, as a researcher. He is Adjunct Professor of Computer Science at the University of Catania (since 2008) and a Contract Professor of Computer Vision at the Academy of Arts of Catania (since 2004). His research interests lie in the fields of computer vision, pattern recognition and machine learning. He has edited four volumes and coauthored more than 70 papers in international journals, conference proceedings and book chapters. He is a co-inventor of four international patents. He serves as a reviewer and on the programme committee for major international journals and international conferences. He founded (in 2006) and currently directs the International Computer Vision Summer School (www.dmi.unict.it/icvss).

**Oliver Giudice** was born in Augusta, Italy in 1986. He received his master degree in Computer Science Engineering (summa cum laude) from University of Catania in 2011. He is currently Ph.D. student at the Image Processing Laboratory (IPLAB) at the Department of Mathematics and Computer Science, University of Catania. His interests lie in the fields of computer vision, pattern recognition and mobile development.

**Giovanni Puglisi** was born in Acireale, Italy, in 1980. He received the M.S. degree in computer science engineering (summa cum laude) from Catania University, Catania, Italy, in 2005, and the Ph.D. degree in computer science in 2009. He is currently a contract Researcher at the Department of Mathematics and Computer Science, Catania University. His research interests include image/video enhancement and processing, camera imaging technology and multimedia forensics. He edited one book, coauthored more than 40 papers in international journals, conference proceedings and book chapters. He is a co-inventor of several patents and serves as reviewer different international journals and conferences.