IET Computer Vision

Research Article

IET Journals
The Institution of Engineering and Technology

Curve description by histograms of tangent directions

ISSN 1751-9632
Received on 24th September 2018
Revised 18th March 2019
Accepted on 28th March 2019
E-First on 28th June 2019
doi: 10.1049/iet-cvi.2018.5613
www.ietdl.org

Ali Köksal[1], *Mustafa Özuysal*[1] ✉

[1]*Department of Computer Engineering, İzmir Institute of Technology, İzmir, Turkey*
✉ *E-mail: mustafaozuysal@iyte.edu.tr*

Abstract: The authors propose a novel approach for the description of objects based on contours in their images using real-valued feature vectors. The approach is particularly suitable when objects of interest have high contrast and texture-free images or when the texture variations are high so textural cues are nuisance factors for classification. The proposed descriptor is suitable for nearest neighbour classification still popular in embedded vision applications when the power considerations outweigh the performance requirements. They describe object outlines purely based on the histograms of contour tangent directions mimicking many of the design heuristics of texture-based descriptors such as scale-invariant feature transform (SIFT). However, unlike SIFT and its variants, the proposed approach is directly designed to work with contour data and it is robust to variations inside and outside the object outline as well as the sampling of the contour itself. They show that relying on tangent direction estimation as opposed to gradient computation yields a more robust description and higher nearest neighbour classification rates in a variety of classification problems.

1 Introduction

Description of locally defined image features such as interest points and edges is a long standing problem. For object classification purposes, description of edges and contours has a particular appeal since a variation of object outlines within a single class is generally less than textural differences between instances of the same class. Moreover, high contrast images captured under low light and infrared images contain considerably less texture. Unfortunately, robust description of image contours is a challenging problem since such a description has to rely on mostly ambiguous low-level intensity information.

In this study, we propose a simple and effective algorithm for the description of image contours that does not depend on textural information such as image gradients. The proposed description is purely based on the histograms of contour tangent directions. We show that avoiding gradient information inside, along, and outside the extracted contour is essential to increased performance on tasks such as character and shape classification.

Our approach belongs to the group of heuristically designed descriptors such as shape context (SC) [1]. Such approaches require a relatively small number of operations. Therefore, they are suitable when the power considerations outweigh performance requirements such as in embedded vision applications. Moreover, the descriptors are usually classified using the nearest neighbour (NN) approach, which requires only Euclidean distance computations. Therefore, classification can be very efficiently implemented and ported to existing and novel hardware. Efficient algorithms for approximate NNs [2] exist to scale these approaches to very large descriptor data sets.

The dominant approach for shape description – the SC descriptor and its variations [3–5] – is based on a sampling of the contours and representing the relative positions of the samples with respect to each other by histograms. This is a powerful approach that is invariant to variations in object texture and robust to geometric deformations; however, including relative positions of all points with respect to all others in the description results in a global representation. This means that if a certain part of the contour is significantly altered or occluded, the description at every point is affected by this change. This is in contrast to descriptors such as scale-invariant feature transform (SIFT) [6], where the histograms only store local gradient orientation information. If an image patch is partially changed due to lighting or occlusion, its SIFT description is only partially affected. Since SIFT is designed to work with textured image patches, it is not directly suitable for the description of image contours and it does not perform well for shape classification when intra-class texture information varies a lot.

There are many approaches that are tailored directly for curve description. Some approaches such as [3, 5] build on SC to improve its characteristics, some others exploit shape skeletons [4, 7], while others partition the shape contour and describe the resulting sub-contours [8, 9]. However, these do not take advantage of the heuristics exploited by SIFT-like texture descriptors that yield impressive performance for the key point matching task. The key design heuristics that drive that performance are the discriminative power of direction information and the robustness of histograms computed on a spatial grid.

Based on this insight, we have designed a descriptor that gathers only local tangent distributions along the curve to be described. Fig. 1 illustrates the steps of the proposed descriptor computation. We call the resulting descriptor as histograms of tangent directions (HOTD). In the experiments section, we show that such a local approach yields improved robustness compared to global descriptors. Moreover, we demonstrate that, at least for shape description applications, HOTD outperforms histograms of gradient directions even when gradients are only taken at the shape boundary.

2 Related work

Various approaches have been developed to describe shapes using numerical vectors. One alternative is to simply exploit intensity/texture-based descriptors such as the SIFT descriptor for this purpose. However, when the shape outline is more discriminative and the texture of the objects of interest varies a lot, a contour-based approach is arguably more suitable. Therefore, specific approaches targeting curves have been developed, SC [1] being one of the most prominent ones. The main advantage of both lines of work is their implicit simplicity. They are fast to compute and effective even when used in combination with a NN classifier. In this study, we show that HOTD have all the same benefits and they are an effective description for several applications from letter outlines to shape boundaries.

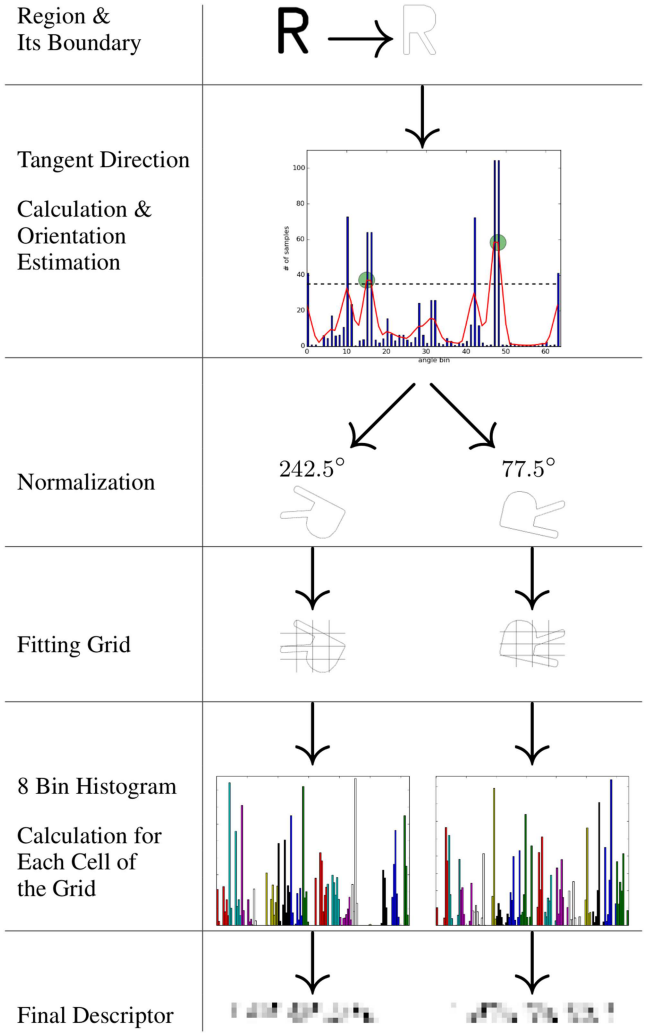

Fig. 1 *Summary of the descriptor computation. We describe shapes by extracting their outer contours. The tangent directions for the contour are computed and the peaks of their orientation histogram determine the major orientations for the shape. A grid is fitted to a rotation normalised contour to fix the scale of description. The parts of the contour that fall in a specific grid cell contribute to a tangent orientation histogram within that cell. The concatenation of the histograms from each cell yields the histograms of oriented directions (HOTD), which mimics the key design heuristics of texture descriptors but it is tuned to the contour shape and robust to variations inside and outside the object contour*

The simplest way to describe image patches is to encode the intensity information and compare it using a metric such as cross correlation [10] or variations [11]. By gaining invariance to local deformations, it is possible to compute moments [12], normalised histograms [13], or colour histograms [14] of the intensity values. The disadvantage is the variance due to the viewing direction and the inability to effectively describe binary object masks or high contrast images.

Apart from SIFT, several descriptors exploit gradients and their directions to characterise an image neighbourhood. Gradient location orientation histogram (GLOH) [10] bins gradient orientations on a circular grid. Steerable filters [15] and differential invariants [16, 17] extend the filters used to directional and higher-order ones. Schiele and Crowley [18] compute rotation invariant gradient histograms at multiple scales. Principal component analysis (PCA)–SIFT [19] applies dimensionality reduction on top of gradient histograms. We exploit the design of these texture descriptors, but we replace gradients with robust tangent computation along the curve. We show that this is very effective to encode only relevant information for a given object outline.

There are many approaches developed specifically for the description of object contours. SC [1] encodes the positions of points on the curve with respect to a reference in log-polar space. Inner distance SC [3] is a variant of SC that changes how the distance to the reference is computed. Geometric blur [20] and curvature scale space [21] filter the curve with Gaussians and record various statistics of the resulting shapes. Similar to SC, contour points distribution histogram [5] relies on point position histograms on a circular grid. Revollo *et al*. [22] exploit bilateral and radial symmetries in the object contour to better describe the shape of objects.

Instead of the contour boundary, skeletal context [4] and shock edit [7] describe curves based on the skeleton of the region they enclose. Distance set [23] computes a distance set for each point and its neighbours and selects significant ones for description. Our approach bears similarity with these since we also restrict computations to the contour boundary. However, instead of point locations, we compute and record tangent orientations.

Some contour-based approaches require pairwise matching and/or normalisation of curves. Curve edit [24] matches high curvature points of aligned curves before description. Tu and Yuille [25] measure the similarity of two curves based on the transformation required to compute one from the other. These approaches certainly increase the matching performance, but they require pairwise processing. As a result, they are not immediately scalable to larger databases, unlike description based on a numerical vector in a Euclidean space, which can potentially employ approximate NN approaches [2].

Several recent approaches describe shapes by first partitioning the contour into sub-parts. Wang *et al*. [9] split the contour after fitting a parametric curve and finding critical points of this representation. The fragments are then encoded with their SC descriptors and these are vector quantised and histograms of the resulting codebooks represent the overall shape of the curve. Laiche *et al*. [8] fit a polynomial to the contours and compute sub-curves based on high curvature points. The sub-curves are normalised and further described by cubic polynomials. Such approaches work well for certain data sets, however splitting the curves into sub-curves is risky, since curvature and other metrics used in the splits are not affine invariant.

3 Description of curves

Our approach computes a real-valued descriptor vector for a curve given as a discretely closed chain of points. Depending on the application, this curve might be computed using a segmentation mask corresponding to the object boundary or it might be extracted by a region detector such as one computed by maximally stable extremal regions (MSERs) [26]. In the latter case, if multiple regions are detected on the object, we take the outermost one as input to our algorithm, ignoring inner holes and structures. During the computation of the descriptor, points on this input curve are traversed in the counter-clockwise direction.

3.1 HOTD descriptor

Since our descriptor is based on tangent directions, a tangent direction is estimated for each point on the input curve. For this purpose, we employ the median filtered differencing algorithm [27], which we found to be quite robust to variations of the object shape and point sampling. To compute the tangent direction at a given point p on the curve, m vectors to the preceding points, $\boldsymbol{v}_{\mathrm{p}}^{-}$, and m vectors to the following points, $\boldsymbol{v}_{\mathrm{p}}^{+}$, are taken into consideration each one starting at point p. The vectors $\boldsymbol{v}_{\mathrm{p}}^{-}$ and $\boldsymbol{v}_{\mathrm{p}}^{+}$ are converted to their polar representations and the tangent direction is taken to be the median of their angles. As m increases, the calculated tangent direction becomes more robust to discretisation errors along the curve, but the accuracy is reduced due to smoothing. In practice, we take m to be three, which we found to present a good balance between robustness and accuracy. However, the value of m should be considered as a hyper-parameter that depends on the contour resolution and noise levels. As a result, it should be optimised based on the input characteristics.

Once tangent directions are computed at each point, the proposed HOTD descriptor is computed in two steps, orientation estimation followed by spatial and angular binning. To estimate an

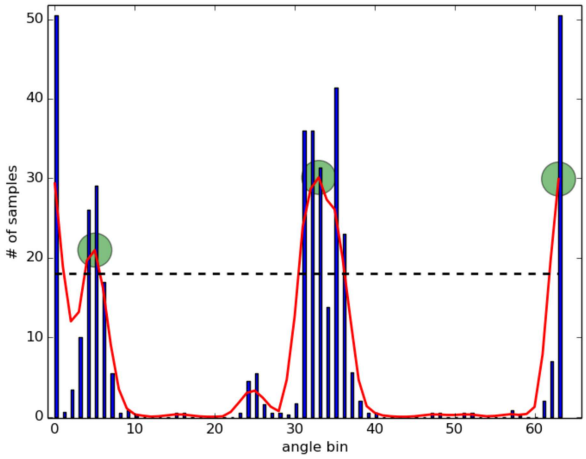

Fig. 2 *To estimate the shape orientation, the histogram of the tangent directions of the shape contour is computed. The peaks are found after smoothing the histogram with a box filter multiple times. The peaks of the smoothed histogram marked by the red line are given by the green circles. The highest peak determines the primary orientation and the threshold for secondary orientations given by the dotted black line. For this sample shape, the primary orientation is at* 167.5°*, and the secondary orientations are at* 317.5° *and* 27.5°

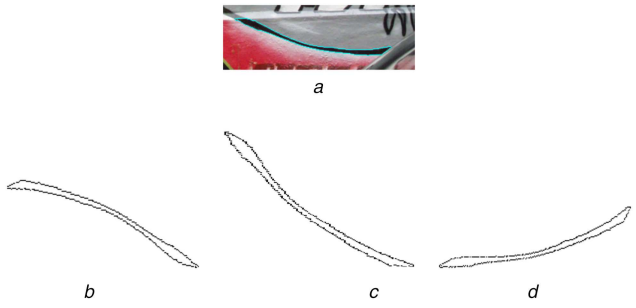

Fig. 3 *Orientation normalisation*

(a) Original curve, a region in the image detected by the MSER detector whose contour is drawn in light blue, ***(b)–(d)*** Rotated curves according to the computed primary and secondary orientations, respectively, at 167.5°, 317.5°, and 27.5°

Fig. 4 *Rotated shapes are scale normalised by fitting a rectangular grid to their bounding boxes. This creates 16 spatial grid cells and a separate eight bin histogram is computed in each one. The final descriptor is visualised under each shape by a grey-scale map that encodes the weight of each histogram bin. There are four rows in the map and each row contains the four histograms corresponding to a single row of the grid. As a result, there are 32 columns in each row. The top-left eight boxes in the first row of the visualisation correspond to the eight histogram bins of the histogram at the top-left grid cell. Grid cells that do not intersect any parts of the shape contour are left empty*

orientation for the curve, the histogram of the tangent directions is computed using 64 bins that approximately yield $\pm 5.6°$ angular resolution. We use linear interpolation to compute contributions to several bins corresponding to the calculated tangent orientation. The histogram is smoothed to remove small local minima by twice convolving with a box filter of size three. We have observed that the amount of smoothing affects the descriptor performance although not by a large percentage, so it should be considered as a hyper-parameter of the approach that might potentially be tuned based on input characteristics.

The maxima of the histogram are taken to be the primary curve orientation. Since multiple local maxima are possible, we also store a list of secondary orientations containing angles for all local maxima that is higher than 60% of the primary peak. This scheme is illustrated by Fig. 2 and it replicates the one used by a gradient-based detector and descriptors for point features. Estimating potentially multiple dominant orientations and extracting more than one descriptor is not an optimal solution. Nevertheless, this is a common solution to rotation invariance employed for key point matching and our experiments show that it works reasonably well for shape recognition. Since the computation of the proposed descriptor is independent of the way orientation is estimated, if a more suitable approach exists for the problem at hand, the steps outlined above could be replaced by it.

For each computed orientation (be it primary or secondary), we extract a separate descriptor by performing the following three operations: shape orientation normalisation, shape scale normalisation by fitting a spatial grid, and shape descriptor computation.

In the first step, the shape is normalised by rotating according to the estimated orientation to achieve robustness to planar rotations of the object. This corresponds to multiplying points along the curve with a rotation matrix. Fig. 3 illustrates an example shape and its normalised versions corresponding to each estimated orientation.

Once orientation is normalised, in the second step, we also scale normalise the shape and rescale the spatial grid that will be used for descriptor computation according to the estimated scale. The scale is estimated by computing the bounding box for the orientation normalised shape. Since we use a 4×4 spatial grid for descriptor computation each cell is of width and height set to one-fourth of the bounding box width and height, respectively.

Once the cells of the spatial grid are determined, each point along the curve is assigned to one or more spatial cells according to their spatial location. We use bilinear interpolation to compute multiple cell weights for a single curve point to improve robustness to small deformations and localisation errors.

In the third and last step, we recomputed the tangent directions as before on the normalised curve with m set to two. For each grid cell, we compute an orientation histogram with eight bins again by linearly interpolating the tangent orientation to determine the contribution to each histogram bin. The histograms are concatenated and the resulting vector is normalised to unit Euclidean length. The resulting HOTD descriptor is a robust description of the curve shape.

Fig. 4 visualises the computed HOTD descriptors for each normalised shape shown in Fig. 3. As is evident from the visualisations, the HOTD descriptor encodes both the distribution of the curve point over the spatial grid cells as well as the dominant tangent directions within each cell. As the experiments demonstrate this leads to a robust and discriminative representation of the curve shape. All the steps in the descriptor computation are summarised in Fig. 1.

3.2 Comparing gradient and tangent orientations

In order to clearly separate the description power of tangent orientations from the particular way the HOTD descriptor is computed, we present an alternative formulation based on gradient orientations that are similar to feature descriptors in the existing literature. The only difference between the proposed HOTD descriptor and the description below is the use of gradient information as opposed to tangent directions.

To replace tangent orientations with gradient information, we simply calculate the image gradient using forward differences at each point along the curve. We do so both before orientation estimation and also during descriptor computation for the orientation and scale normalised shapes. While this is similar to the HOTD computation, in contrast to feature descriptors such as SIFT, the gradient information is confined to the curve boundary and the gradients inside and outside the curve do not affect the descriptor computation. In Section 4, we contrast the performance of HOTD

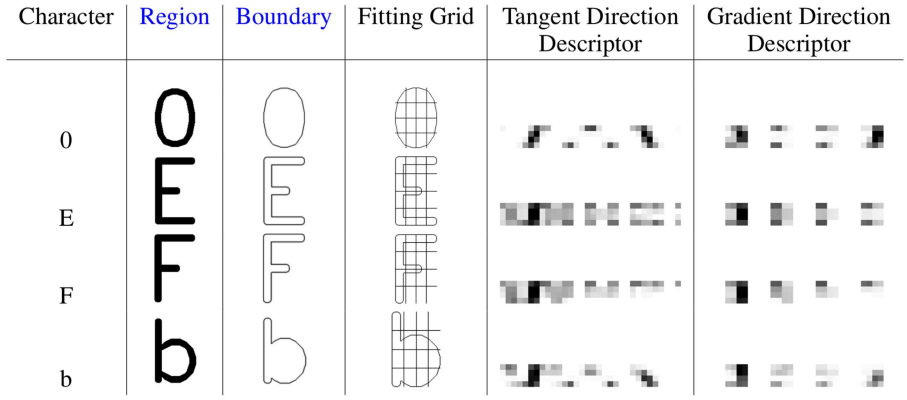

Fig. 5 *Comparison of the description using the computed tangent and gradient directions. The descriptors for four characters demonstrate the differences between the gradient and tangent direction histograms. We compute the descriptors in the figure for upright characters so that it is easier to visually match the descriptor bins to the image of the letters. In all the experiments, the characters are rotated such that their dominant orientations point upwards before the descriptor is computed to handle arbitrary rotations in the image plane. While the tangent and gradient information is very similar in each case, the tangent estimation relies on a robust metric that takes into account the discrete nature of the curve, which is represented as a sequence of point samples. As a result, the descriptors based on estimated tangent orientations yields a richer representation of the underlying curve than those based on calculated image gradients*

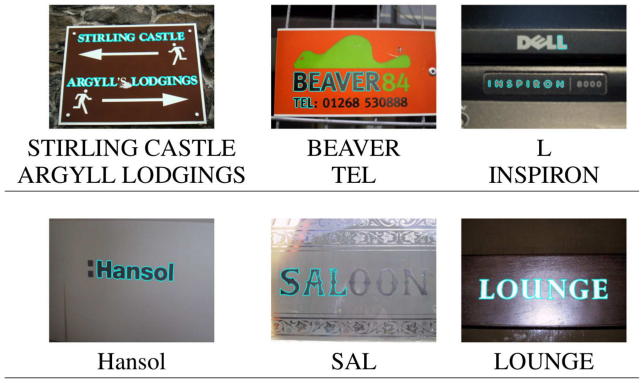

Fig. 6 *Samples from the modified ICDAR 2013 data set. The scene text images from the original ICDAR 2013 data set are processed by the MSER region detector. We only keep the characters that are successfully detected by the MSER detector whose outlines are depicted in light blue. These curves are placed in the modified ICDAR 2013 data set with the ground truth character codes provided in the ICDAR 2013 data set. This process mimics an actual character recognition pipeline based on the MSER detector*

with that of the gradient-based variant and show that tangent directions yield higher performance in several classification tasks.

To visualise the differences between these two alternative approaches, we visualise the corresponding descriptors for various letter shapes in Fig. 5. For simplicity, the letter images are generated synthetically and we only visualise the descriptors for the upright orientation. Note that the descriptors for 'E' and 'F' are very similar except the missing segment in 'F'. In this idealised high contrast situation, the gradient and tangent information are nearly identical. However, as can be observed from the visualisations, the tangent direction computation is more robust to discretisation artefacts and yields a richer description of the curve.

4 Experiments

The description of curves is usually associated with classification problems that involve the outline of objects of interest. For different kinds of classification tasks, the type of input image and how the curve is specified varies. To account for such differences, we test the proposed approach on three application domains: character classifications from outlines, shape recognition from segmentation masks and object recognition from boundary curves.

In each case, we describe the input images and how the curves to be described are extracted from them. Depending on the data set, sometimes image intensity values are also available and sometimes only a binary segmentation mask is provided. The object boundaries might be provided or a region detector or an automatic segmentation algorithm might be required to extract the image curves.

Once a curve is computed for each image, we compute the proposed descriptor as detailed in Section 3 and possibly the variant based on image gradients. Usually, the classification is performed by finding one or more NNs in the Euclidean descriptor space. However, different data sets have different performance evaluation criteria as described separately for each data set.

4.1 Character classification from outlines

To test the suitability of the HOTD descriptor for character classification, we have used the standard *Chars74k* data set and created a data set based on the popular ICDAR 2013 scene text recognition images that we call the *modified ICDAR 2013*. Since the ICDAR 2013 challenge aims end-to-end text detection and classification, the available ground truth is limited to bounding boxes for the letters. To prepare the ground truth for letter recognition based on curve description, we have run the MSER detector on the ICDAR 2013 images and extracted the MSER curves corresponding to letters in the scene text images. This yields extremal regions corresponding to 3141 of the 4419 existing characters. These characters mostly correspond to upper and lower case English letters with an uneven distribution. Fig. 6 depicts selected images with the detected characters outlined.

We could have used all the character bounding boxes in ICDAR 2013 and segmented curves corresponding to the character outlines in a preprocessing stage. The reason we restrict the experiment to MSER detector output is to separate the descriptor performance from the preprocessing stages. This lets us estimate the value of the HOTD descriptor for scene text recognition pipelines that also exploit the MSER detector for character detection [28]. On the *Chars74k* data set, we also test the classification performance coupled with segmentation algorithms, which is the standard protocol for the natural images subset of *Chars74k*.

We use the *modified ICDAR 2013* data set to compare the proposed descriptor to existing descriptors of similar complexity. For each detected character, we take the extended boundary and the ellipse representation of the detected region. The executables kindly provided by Mikolajczyk *et al.* [10] are used to compute state-of-the-art texture and curve descriptors such as scalable invariant feature transform (SIFT) [6], GLOH [10], SC [1], PCA-SIFT [19], spin images (SPIN) [13], steerable filters (JLA) [15], differential invariants (KOEN) [16], complex filters (CF) [29], moment invariants (MOM) [12], and cross-correlation (CC) [10].

We set aside 80% of the descriptor data as the training set and the remaining are placed in the test set. The NN classifier is used to retrieve the closest descriptor for each test shape. The matching score is computed as the correctly classified test samples divided by the total number of test cases. Since the train/test split is random, we repeat the same process 20 times and report both the

mean matching score together with the standard deviation in Table 1.

The texture-based descriptors such as SIFT and GLOH require a scale of extraction as their input. Since different descriptors compute the scale differently, for each descriptor we have experimented with a range of scale settings to be fair in our comparison. The reported results correspond to the scale that achieves the highest matching score for each texture descriptor. For curve description, the sampling and locations of the contour points determine the scale implicitly, so such a scale sweep is not required. Overall, the proposed HOTD descriptor based on tangent directions achieves the best matching score of 88.75% followed by GLOH with a score of 81.53% and closely followed by SIFT. The gradient-based variant of the proposed descriptor achieves 75.04%, which shows that tangent directions are better for description than simply taking the gradients along the curve.

Other than the *modified ICDAR 2013* data set we have prepared, we have also tested the proposed approach on the publicly available *Chars74k* character data set. We use the English characters and digits for evaluation yielding a total of 62 classes. *Chars74k* has three types of data in three subsets: 62,992 images of synthetically generated characters from fonts, binary images corresponding to handwritten characters, and finally 7705 cropped image patches of scene text. For the first two subsets, we already have the binary image masks and use the contour extraction code from OpenCV [30] to compute the image curves. To extract the curves in the last subset, we first segment the images into foreground and background masks using the approach proposed in [31]. Once we obtain the binary mask, the curves are extracted in the same manner with the other subsets of *Chars74k*. Fig. 7 shows examples from each subset.

For *Chars74k*, we use the same standard experimental protocol proposed by de Campos *et al*. [32]. For each one of the three subsets, 15 randomly selected samples for each class are set aside as test data. For each test sample, the NN descriptor is found within the remaining training examples. The matching score is calculated as in the experiments of the *modified ICDAR 2013* data set.

The proposed method is compared to the stated results that use the NN classifier in Table 2.

Each column gives the results for a different subset. The HOTD descriptor achieves matching rates of 87.83, 77.18, and 60.40%, which is the best in all the subsets of *Chars47k*. The gradient-based variant of the HOTD descriptor yields 85.78, 75.55, and 56.45% showing that the tangent directions are again the better choice. Fig. 8 shows the confusion matrix for the natural images subset of the *Chars74k* data set.

In the last group, the natural character subset, we would like to note that the results depend on the segmentation approach employed. While the texture-based descriptors in [33] do not require such a step, contour descriptors expect a curve as their input. As a result, the matching scores of contour descriptors are functions of both segmentation and descriptor quality. Our initial experiments using a simple segmentation approach based on the intensity histogram yielded lower classification accuracy. A more suitable segmentation approach [31] improves the quality of the contours hence the descriptor classification accuracy is higher.

We believe that one of the biggest advantages of our approach for character classification is the restriction of the description to the letter outline. Texture descriptors such as SIFT risk including too much information from the surrounding context such as nearby letters and unrelated image features. The contour descriptors such as SC do not have this disadvantage, but they rely on different heuristics than SIFT that makes use of orientation histograms. Our approach combines the advantages of both, it does not take into account anything beyond the letter outline and it exploits the powerful heuristics from the SIFT design. Coupled with robust tangent estimation, the proposed descriptor is really suitable for classification of character images from a computed character outline.

4.2 Shape classification from segmentation masks

The main challenge in character classification is the number of classes and the similarities between characters. The shape and viewpoint variations are relatively smaller compared to the more general shape recognition task. The outlines of various shapes might contain an increased number of protrusions and gaps. The objects might also be arbitrarily rotated.

We evaluate the shape classification performance of the HOTD descriptor on the *Kimia-99* data set. It has 11 examples for nine classes and a total of 99 images each containing object silhouettes as shown in Fig. 9. Similar to previous cases where the object mask is available, we extract a contour curve using OpenCV.

The evaluation protocol is the same as in [3], which is a leave-one-out variant of k-NNs for k from one to ten. Table 3 lists the matching scores obtained by our approach as well as several existing approaches. HOTD descriptor achieves the best results except for $k = 6$, so it has the highest overall performance. Its score of 981 means that only nine images out of 990 test cases were misclassified. The gradient- based variant makes ∼200 more mistakes.

These results indicate that the histogram-based approach that we propose is able to capture the finer details of the classes, such as the fingers in the hand class and the feet of the cat class, at least for these low-resolution binary images. Moreover, the orientation estimation and peak determination work well enough to capture the different possible main orientations of a wide range of shape classes.

We also evaluate the effect of changing the value of m and input resolution. The results in Table 4 show that the optimal value of m is quite robust to changes in resolution down to a certain level beyond which the value of m needs to be adjusted.

4.3 Object classification from boundary curves

The test data for the previous experiments were mostly binary segmentation masks except for the natural subset of the *Chars74k* data set. When only a binary image is provided contour-based approaches such as ours has a clear advantage since the intensity gradients are not informative. To better evaluate the proposed HOTD in the context of natural images, we test its performance on the *ETH-80* object classification data set.

ETH-80 contains 41 images per each of the 80 objects from ten categories. Fig. 10 shows the images of the objects, which contain a simple background, and the boundary curves of the objects are also provided in the training and test data. The images have higher resolution and contain colour data, which the algorithms may or may not exploit. The experimental protocol is similar to that of *Kimia-99*. For each category, we compute a separate recognition rate and also report the average recognition rate over all categories.

Table 5 shows recognition rates achieved by our approach and others. The recent shape normalisation approach of Laiche *et al*. [8] outperforms others including ours. Comparing these to the results on *Kimia-99* data set where our approach outperforms [8], we can confirm that the type of shape present and the existence of intensity information affect the descriptor performance. This is further evident from the confusion table given in Table 6. The mistakes of our approach are mostly a result of the confusion between *tomato* and *apple* classes which have similar outlines but differ in texture and colour distribution.

Moreover, the three categories that represent animals, *horse*, *cow*, and *dog*, are confused with each other albeit to a lesser extent. One possible explanation is that our descriptor is only able to describe the coarser shape here and does not capture the smaller variations in the heads and the tails for these animals, which contain the most discriminating information. Therefore, smaller grid cells and a larger number of histograms might need to be used for higher performance. However, this would also increase the number of dimensions and as a result the classification time and power cost.

An alternative might be to increase the number of training samples. This might especially be helpful since the three-dimensional nature of the animal classes creates an additional layer of complexity. More samples from varying viewpoints might help

Table 1 NN matching scores of the common texture and contour descriptors as computed on the *modified ICDAR 2013* data set

Algorithm	$k = 1$, %	$k = 3$, %	$k = 5$, %
CC [10]	79.44 ± 1.21	77.07 ± 1.28	80.06 ± 1.56
CF [29]	76.05 ± 2.24	77.07 ± 2.03	74.98 ± 1.72
GLOH [10]	81.53 ± 1.51	82.75 ± 1.53	81.13 ± 1.70
JLA [15]	74.28 ± 2.70	76.44 ± 3.22	74.52 ± 2.86
KOEN [16]	66.82 ± 2.44	68.75 ± 1.91	67.16 ± 2.04
MOM [12]	76.26 ± 2.22	77.05 ± 2.25	75.14 ± 2.38
PCA-SIFT [19]	77.30 ± 2.23	79.99 ± 2.56	78.38 ± 2.15
SC [1]	78.81 ± 2.23	79.73 ± 2.16	78.20 ± 1.96
SIFT [6]	81.16 ± 2.57	82.50 ± 2.57	80.24 ± 2.53
SPIN [13]	56.15 ± 3.23	57.88 ± 3.05	55.41 ± 3.00
ours gradient	75.04 ± 1.43	74.10 ± 1.39	74.25 ± 1.59
ours tangent	88.75 ± 1.21	87.03 ± 1.11	86.55 ± 1.26

The standard deviations correspond to 20 repetitions on the data set based on random sampling of the test and training subsets.

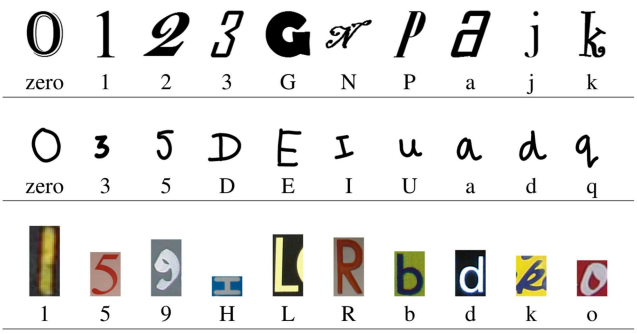

Fig. 7 *Examples from the Chars74k data set. The first row depicts the samples from the synthetically generated subset corresponding to font data. The second row shows the samples from the handwritten character subset. These two subsets have ground truth masks, which we use in the experiments. The last row illustrates patches for natural images containing characters which we segment using the approach of [31]*

to better distinguish these classes. Since approximate NN methods work well with descriptors like ours, the number of training samples per class in the *ETH-80* data set is much lower than what can actually be used in practice.

5 Discussion and conclusion

The description of curves has been a longstanding problem in computer vision. We have proposed a curve descriptor that carries the same spirit and complexity of the classical SC and the SIFT descriptors. Our descriptor is simple to compute and it captures the essential contour information in a robust and discriminative way. It is well suited for applications that require low computational power and small memory print. In the following, we briefly discuss limitations and possible extensions of the presented descriptor.

Since we only consider the outermost boundary of objects, objects with different internal structures might be confused with each other. It should be possible to extend our approach to multiple possibly overlapping curves, either by accumulating tangent directions in the same grid cells or by concatenating descriptors for the individual curves.

Due to the scale and orientation normalisation, characters with similar lower and upper case forms are confused with each other. Usually, such ambiguities might be dealt with in a post-processing stage that takes into account the surrounding context.

Our approach disregards the grey scale and colour intensity information. As such, it is more suitable in applications where high contrast/binary images are provided or the categories have high variance in their intensity/texture information. This result is evidenced by the variance in the performance of our approach on the *Kimia-99* and *ETH-80* data sets.

Table 2 NN matching scores of the descriptor-based approaches as computed on the *Chars74k* subsets

Algorithm	Fonts, %	Handwritten, %	Natural images, %
geometric blur (GB) [20]	69.71 ± 0.64	65.40 ± 0.58	47.09
maximum response 8 (MR8) [17]	30.71 ± 0.67	25.33 ± 0.63	10.43
patches [11]	44.93 ± 0.65	69.41 ± 0.72	21.40
SC [1]	64.83 ± 0.60	67.57 ± 1.40	34.41
SIFT [6]	46.94 ± 0.71	44.16 ± 0.79	20.75
SPIN [13]	28.75 ± 0.76	26.32 ± 0.42	11.83
histogram of oriented gradients (HOG) [33]	—	—	57.5
NATIVE + FERNS [34]	—	—	54.0
rank-1 tensor decomposition (R1-TD) [35]	73.0	—	56.0
ours gradient	85.78 ± 0.73	75.55 ± 1.08	56.45
ours tangent	87.83 ± 1.37	77.18 ± 0.50	61.40

The proposed descriptor achieves the best reported *NN classification* results in all the subsets.

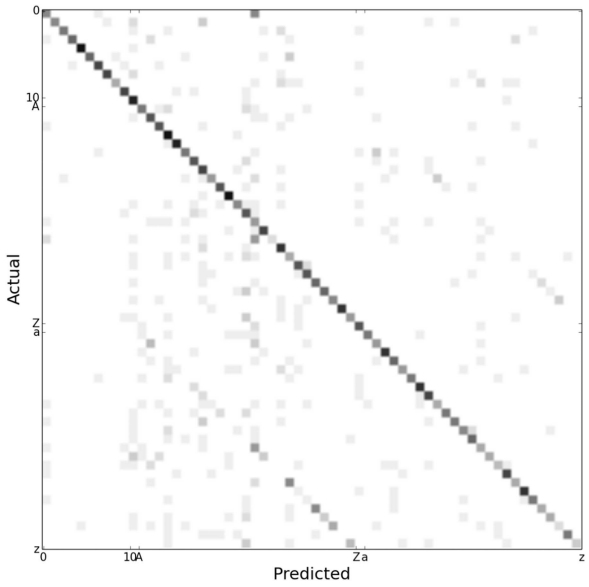

Fig. 8 *Confusion matrix for natural image subset of the Chars74k data set. The heavily weighted main diagonal corresponds to correctly classified characters. Some characters are confused due to the similarity between their scale normalised uppercase and lowercase versions or similar looking numbers*

Fig. 9 *Samples from the Kimia-99 data set. The top row shows examples from different shape classes. The bottom row illustrates the variations of samples from a single class*

Since the histograms that we exploit only allow for robustness to local deformations, they only perform well under two-dimensional transformations and a limited range of three-dimensional view changes. To truly achieve three-dimensional shape and object classification, descriptors from multiple orientations would need to be recorded and used during the classification phase.

We have shown that the experiments convincingly show that the proposed approach is quite robust to the sampling of the described curves. However, in some applications, a much higher scale ratio might exist between image samples to be classified. In such cases,

Table 3 NN matching scores of the existing approaches as computed on images of the *Kimia-99* data set [3]

Algorithm	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	Total
contour points distribution histogram (CPDH) + Earth mover's distance (EMD) [5]	96	94	94	87	88	82	80	70	62	55	808
generative models [25]	99	97	99	98	96	96	94	83	75	48	885
inner distance shape context (IDSC) + dynamic programming (DP) [3]	99	99	99	98	98	97	97	98	94	79	958
multidimensional scaling (MDS) + SC + DP [3]	99	98	98	98	97	99	97	96	97	85	964
SC [1]	97	91	88	85	84	77	75	66	56	37	756
shock edit [7]	99	99	99	98	98	97	96	95	93	82	956
shape normalisation [8]	97	86	87	75	76	70	55	59	46	44	695
ours gradient	91	91	83	78	75	75	73	75	74	72	787
ours tangent	99	99	99	98	98	98	97	98	97	98	981

We use the experimental protocol of [3] to be able to compare our results to those reported in the literature.

Table 4 Effect of image resolution and parameter m on the recognition performance on the *Kimia-99* data set

Resolution	m	Total
100%	1	658
	3	981
	5	977
	7	975
50%	1	695
	3	949
	5	940
	7	922
25%	1	576
	3	870
	5	865
	7	861
10%	1	583
	3	557
	5	676
	7	675

For a wide range resolution, the setting $m = 3$ works well, however, for very low resolution images, different settings outperform the default setting.

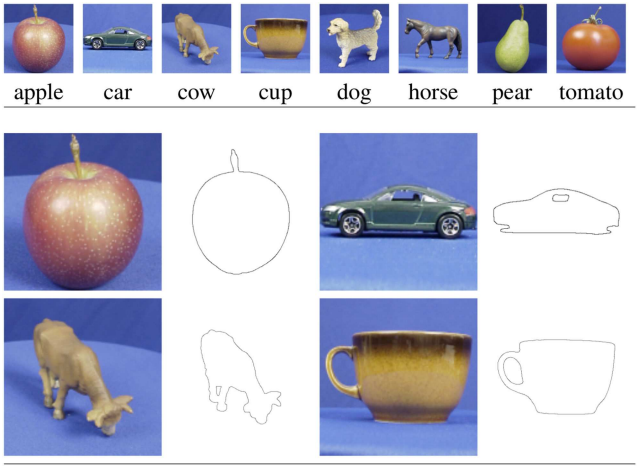

Fig. 10 *Examples from the ETH-80 data set. The top row shows the sample colour images of each class in the data set. The rest of the figure shows four sample images with the corresponding boundary curve provided in the data set. Our approach only relies on the boundary information, disregarding any intensity and colour information provided in the training and test samples*

Table 5 NN matching scores of the descriptors as computed on images of the *ETH-80* data set. Our approach yields fair results largely due to a confusion between the apple and tomato classes

Algorithm	Recognition Rate (%)								
Categories	Apple	Car	Cow	Cup	Dog	Horse	Pear	Tomato	Avg
colour histogram [14]	57.6	62.9	86.6	79.8	34.6	32.7	66.1	98.5	64.9
D_xD_y [18]	85.4	98.3	82.7	66.1	62.4	58.8	90.0	94.6	79.8
IDSC + DP [3]	—	—	—	—	—	—	—	—	88.1
mag – Lap [18]	80.2	77.6	94.4	77.8	74.4	71.0	85.4	97.1	82.2
MDS + SC + DP [3]	—	—	—	—	—	—	—	—	86.8
PCA grey [36]	88.3	97.1	62.4	96.1	66.3	77.3	99.8	76.6	83.0
PCA masks [36]	78.8	100.0	75.1	96.1	72.2	77.8	99.5	67.8	83.4
SC greedy [1]	77.1	99.5	86.8	99.8	82.0	84.6	90.7	70.7	86.4
SC + DP [1]	76.3	100.0	86.3	99.0	82.9	84.6	91.7	70.2	86.4
shape normalisation [8]	97.5	99.37	91.87	100.0	80.93	89.37	100	95	94.2
kernel edit [37]	—	—	—	—	—	—	—	—	91.3
bag of contour fragments (BoCF) [9]	—	—	—	—	—	—	—	—	91.5
ours gradient	78.5	99.0	73.9	90.5	43.7	59.8	93.2	90.2	78.6
ours tangent	76.1	99.8	88.5	99.3	76.3	86.3	90.2	64.1	85.1

The best results are achieved by Laiche *et al*. [8], which we outperform on *Kimia-99*. This shows that the relative descriptor performance depends on the particular variations in the test data.

it might be prudent to interpolate all curves using an approach such as splines and resample these to yield a fixed number of samples per each curve before tangent direction estimation. The sampling might be adaptive so that the tangent variations are adequately captured by the contour points.

We believe further work is necessary to design a handcrafted simple curve descriptor that works well on a variety of data sets, both with and without strong intensity/colour variations. Such a descriptor would simplify the machine vision pipelines of contour-based recognition and classification applications. We have shown that the tangent direction information can be a valuable component of such a descriptor especially for the recognition of character images.

Table 6 Confusion matrix for images of the ETH-80 data set. Descriptors are computed by the proposed approach when their directions are the tangent direction

		Predicted							
		Apple	Car	Cow	Cup	Dog	Horse	Pear	Tomato
Actual	apple	**312**	0	0	3	0	0	6	89
	car	0	**409**	0	1	0	0	0	0
	cow	0	11	**363**	5	14	15	1	1
	cup	1	0	0	**407**	0	0	0	2
	dog	0	3	35	0	**313**	59	0	0
	horse	0	1	24	0	31	**354**	0	0
	pear	14	0	0	12	0	0	**370**	14
	tomato	131	2	0	6	2	2	4	**263**

In the table, the *y*-axis shows the actual values of the samples and the *x*-axis shows the predicted values for the samples. So samples that are shown in the diagonal axis of the matrix are classified correctly.

6 References

[1] Belongie, S., Malik, J., Puzicha, J.: 'Shape matching and object recognition using shape contexts', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, **24**, (4), pp. 509–522

[2] Muja, M., Lowe, D.G.: 'Scalable nearest neighbor algorithms for high dimensional data', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2014, **36**, pp. 2227–2240

[3] Ling, H., Jacobs, D.W.: 'Shape classification using the inner-distance', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2007, **29**, (2), pp. 286–299

[4] Xie, J., Heng, P.A., Shah, M.: 'Shape matching and modeling using skeletal context', *Pattern Recognit.*, 2008, **41**, (5), pp. 1756–1767

[5] Shu, X., Wu, X.J.: 'A novel contour descriptor for 2D shape matching and its application to image retrieval', *Image Vis. Comput.*, 2011, **29**, (4), pp. 286–294

[6] Lowe, D.G.: 'Distinctive image features from scale-invariant keypoints', *Int. J. Comput. Vis.*, 2004, **60**, (2), pp. 91–110

[7] Sebastian, T.B., Klein, P.N., Kimia, B.B.: 'Recognition of shapes by editing their shock graphs', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2004, **26**, (5), pp. 550–571

[8] Laiche, N., Larabi, S., Ladraa, F., *et al.*: 'Curve normalization for shape retrieval', *Signal Process. Image Commun.*, 2014, **29**, (4), pp. 556–571

[9] Wang, X., Feng, B., Bai, X., *et al.*: 'Bag of contour fragments for robust shape classification', *Pattern Recognit.*, 2014, **47**, (6), pp. 2116–2125

[10] Mikolajczyk, K., Schmid, C.: 'A performance evaluation of local descriptors', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2005, **27**, (10), pp. 1615–1630

[11] Varma, M., Zisserman, A.: 'Texture classification: are filter banks necessary?'. Proc. 2003 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 2003, vol. 2, pp. II–691

[12] Van Gool, L., Moons, T., Ungureanu, D.: 'Affine/photometric invariants for planar intensity patterns'. European Conf. on Computer Vision, 1996, pp. 642–651

[13] Lazebnik, S., Schmid, C., Ponce, J.: 'A sparse texture representation using affine-invariant regions'. Proc. 2003 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 2003, vol. 2, pp. II–319

[14] Swain, M.J., Ballard, D.H.: 'Color indexing', *Int. J. Comput. Vis.*, 1991, **7**, (1), pp. 11–32

[15] Freeman, W.T., Adelson, E.H.: 'The design and use of steerable filters', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1991, **13**, (9), pp. 891–906

[16] Koenderink, J.J., van Doorn, A.J.: 'Representation of local geometry in the visual system', *Biol. Cybern.*, 1987, **55**, (6), pp. 367–375

[17] Varma, M., Zisserman, A.: 'Classifying images of materials: achieving viewpoint and illumination independence'. European Conf. on Computer Vision, 2002, pp. 255–271

[18] Schiele, B., Crowley, J.L.: 'Recognition without correspondence using multidimensional receptive field histograms', *Int. J. Comput. Vis.*, 2000, **36**, (1), pp. 31–50

[19] Ke, Y., Sukthankar, R.: 'PCA-SIFT: a more distinctive representation for local image descriptors'. Proc. 2004 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'04), Washington, DC, USA, 2004, pp. 506–513, Available at http://dl.acm.org/citation.cfm?id=1896300.1896374

[20] Berg, A.C., Berg, T.L., Malik, J.: 'Shape matching and object recognition using low distortion correspondences'. Proc. 2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'05). 2005, vol. 1, pp. 26–33

[21] Mokhtarian, F., Abbasi, S., Kittler, J.: '*Efficient and robust retrieval by shape content through curvature scale space*' (World Scientific, Singapore, 1996), pp. 35–42

[22] Revollo, N.V., Delrieux, C.A., González José, R.: 'Set of bilateral and radial symmetry shape descriptor based on contour information', *IET Comput. Vis.*, 2017, **11**, (10), pp. 226–236

[23] Grigorescu, C., Petkov, N.: 'Distance sets for shape filters and shape recognition', *IEEE Trans. Image Process.*, 2003, **12**, (10), pp. 1274–1286

[24] Sebastian, T.B., Klein, P.N., Kimia, B.B.: 'On aligning curves', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2003, **25**, (1), pp. 116–125

[25] Tu, Z., Yuille, A.L.: 'Shape matching and recognition–using generative models and informative features'. European Conf. on Computer Vision, 2004, pp. 195–209

[26] Matas, J., Chum, O., Martin, U., *et al.*: 'Robust wide baseline stereo from maximally stable extremal regions'. British Machine Vision Conf., 2002, pp. 384–393

[27] Matas, J., Shao, Z., Kitter, J.: 'Estimation of curvature and tangent direction by median filtered differencing'. 8th Int. Conf. on Image Analysis and Processing, San Remo, 1995

[28] Liu, Z., Li, Y., Qi, X., *et al.*: 'Method for unconstrained text detection in natural scene image', *IET Comput. Vis.*, 2017, **11**, pp. 596–604(8)

[29] Schaffalitzky, F., Zisserman, A.: 'Multi-view matching for unordered image sets, or 'how do I organize my holiday snaps?'. European Conf. on Computer Vision, 2002, pp. 414–431

[30] Bradski, G.: 'The OpenCV library', Dr Dobb's Journal of Software Tools, 2000

[31] Köksal, A., Işık, Z.: 'Character segmentation on natural images with graph-based representation'. IEEE 26th Signal Processing and Communications Applications Conf. (SIU), 2018, pp. 1–4

[32] de Campos, T.E., Babu, B.R., Varma, M.: 'Character recognition in natural images'. VISAPP(2), 2009, pp. 273–280

[33] Wang, K., Belongie, S.: 'Word spotting in the wild'. European Conf. on Computer Vision, 2010, pp. 591–604

[34] Wang, K., Babenko, B., Belongie, S.: 'End-to-end scene text recognition'. 2011 IEEE Int. Conf. on Computer Vision (ICCV), 2011, pp. 1457–1464

[35] Ali, M., Foroosh, H.: 'Character recognition in natural scene images using rank-1 tensor decomposition'. 2016 IEEE Int. Conf. on Image Processing (ICIP), 2016, pp. 2891–2895

[36] Leonardis, A., Bischof, H., Maver, J.: 'Multiple eigenspaces', *Pattern Recognit.*, 2002, **35**, (11), pp. 2613–2627

[37] Daliri, M.R., Torre, V.: 'Shape recognition based on kernel-edit distance', *Comput. Vis. Image Underst.*, 2010, **114**, (10), pp. 1097–1103