Pattern Recognition Letters 125 (2019) 694–700

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Parameter optimization of a multiscale descriptor for shape analysis on healthcare image datasets

Allan C. Carneiro[a,b,*], José G.F. Lopes[a], Marcelo M.S. Souza[a], Jeová F. Rocha Neto[c], Flávio H.D. Araújo[d], Romuere R.V. Silva[d], Fátima N.S. Medeiros[a], Francisco N. Bezerra[b]

[a] Universidade Federal do Ceará, Ceará, Brazil
[b] Instituto Federal de Educação, Ciência e Tecnologia do Ceará, Ceará, Brazil
[c] Brown University, Providence, RI, USA
[d] Universidade Federal do Piauí, Piauí, Brazil

ARTICLE INFO

Article history:
Received 28 October 2018
Available online 19 June 2019

Keywords:
Shape analysis
Multiscale descriptor
Metaheuristic optimization
Clustering validation measures

ABSTRACT

Shape analysis is a key task in computer vision, and multiscale descriptors can significantly enhance shape characterization. However, these descriptors often rely on parameter adjustments to configure a meaningful set of scales that can enable shape analysis. Parameter adjustment in large image datasets is often done on a trial-and-error basis, and an alternative solution to mitigate such a limitation is the use of metaheuristic optimization. The main contribution of this paper is to provide a strategy that supports the automatic parameter adjustment of a multiscale descriptor within a metaheuristic optimization algorithm, where the choice of the cost function strongly influences and boosts the performance of the shape description, which is closely related to the problem domain, i.e. the image dataset. Our research considers synthetic data in a prior evaluation of the cost functions that optimize the scale parameters of the Normalized Multiscale Bending Energy (NMBE) descriptor through the Simulated Annealing (SA) metaheuristic. The cost functions that drive this metaheuristic are: Silhouette (SI), the Davies–Bouldin index (DB) and the Calinski-Harabasz index (CH). We conduct content-based image retrieval and classification experiments to assess the optimized descriptor using three healthcare image datasets: Amphetamine Type Stimulants (ATS) pills (Illicit Pills), pills from the National Library of Medicine (NLM Pills) and hand alphabet gestures (Hands). We also provide segmentation masks for Illicit Pills to guarantee reproducibility. We report the results of tests using a state-of-art method based on a deep neural network, Inception-ResNet-v2. The optimized NMBE with SI and DB achieved competitive and accurate values of above 94%, in terms of both the Mean Average Precision measure (MAP) and Accuracy (ACC) for Illicit Pills and NLM Pills. The precision recall curves demonstrate that it outperforms the Inception-ResNet-v2 for both of these datasets.

1. Introduction

Computer Vision is the science behind the understanding of how visual systems work and can be implemented in order to find computational solutions to tasks that have images as their primary input. Within this broad domain, shape modelling plays an important role, since it is known that most of the visual information of an object lies on its boundary [1]. On the other hand, finding shape descriptors that numerically represent generic objects and their geometry is, in general, challenging, although this is fundamental in recognition tasks.

Several shape descriptors can be found in the literature, and they are mostly parametric [2–4]. In general, the approaches used to handle the setup of the parameters are manual and involve exhaustive searches, leading to sub-optimal results [5]. This parameter setting becomes specific to a given task or dataset and cannot be transferred to other scenarios. An example of this remark is the Inner Distance-Shape Context (IDSC) descriptor [2], which has been used for leaf-based plant specimen identification [4] with parameter specification primarily calculated for the MPEG-7 [6] dataset.

Parameter setting is application-dependent, and the search for new parameters may be challenging for a particular assignment [5]. In order to tackle this problem, the authors of [7] proposed a framework that applies metaheuristic-based methods and a clustering validation measure called Silhouette [8] as the cost function for the parameter optimization of a multiscale shape

* Corresponding author.
E-mail addresses: allan.carneiro@ifce.edu.br, allanccarneiro@gmail.com (A.C. Carneiro).

https://doi.org/10.1016/j.patrec.2019.06.017
0167-8655/

descriptor. Inspired by [7], we investigate the role of different cost functions in this framework and apply our technique to retrieval and classification tasks using three healthcare image datasets. Our work is an effort to provide an optimized approach that improves the shape description and analysis of these healthcare images. Two of these datasets raise the problem of recognizing the shapes of pills, both legal and illegal.

Amphetamine Type Stimulants (ATS) represent a significant drug problem according to the World Drug Report 2011 published by the United Nations Office on Drugs and Crime [9]. Between 0.3% and 1.3% of the world population, with ages from 15 to 64 years, are affected by this problem [10]. Thus, the automatic identification of illicit pills based on pattern recognition algorithms is of great importance for forensic science, drug monitoring and the analysis of seized drugs around the world. The development of these computer systems for pill drug image recognition relies on color, imprint and shape features [10–13]. Shape descriptors can be employed as a preliminary filter to improve the performance of this task.

In [14], the authors exploit the importance of the identification of unknown medications in disaster and emergency situations as a major motivation to develop automated algorithms and software tools for drug recognition, and particularly prescription tablets and capsules. Pill misidentification may lead to catastrophic outcomes to patients and a huge financial burden on healthcare cost worldwide [14,15], and algorithms for pill recognition can benefit from computer vision techniques [11,14,15].

There are several studies that have aimed to develop gesture recognition systems to assist people with hearing limitations, some of which are based on computer vision techniques [16,17]. The performance of gestures as a sign language is the main form of communication of hearing impaired people, and the development of computer-assisted tools for automatic interpretation of these sign relies on hand gesture recognition.

Faced with these related problems, we applied the proposed shape descriptor optimization methodology to public image datasets of ATS pills (Illicit Pills) [18], legal medicines (NLM Pills) [14,19] and hand alphabet gestures (Hands) [20] using a shape description technique called Normalized Multiscale Bending Energy (NMBE) [21]. We chose NMBE because it is invariant to scale, rotation, translation, and is robust to noise artifacts generated by image acquisition [21], and due to its simplicity and versatility in dealing with healthcare images.

We performed a quantitative evaluation of the optimized descriptor in terms of the Mean Average Precision (MAP) [22] measure and the precision recall curves [23,24] in content-based image retrieval (CBIR) [25] experiments. We also computed the classification Accuracy rate (ACC), in the supervised classification experiments. The main contributions of this work are as follows: a) the design and validation of a metaheuristic optimization solution for parameter adjustment of a shape descriptor to better characterize shapes in computer vision applications; b) a quantitative analysis of three clustering validation measures, i.e. Silhouette (SI) [8], the Davies–Bouldin index (DB) [26] and the Calinski-Harabasz index (CH) [27] using two synthetic databases (this can be extensible to other datasets); c) an investigation of how the cost function choice affects the output of our processing pipeline; d) understanding how the cost functions, SI, DB and CH, can handle different aspects of three healthcare image datasets in an optimization solution involving the parameter adjustment of a multiscale shape descriptor; e) a comparison with a deep feature extraction algorithm, i.e. the Inception-ResNet-v2 [28] applied to three healthcare image datasets for image retrieval and classification tasks; f) the release of segmentation masks for the Illicit Pills dataset for the benchmarking and reproducibility of our research.

The remainder of this paper is organized as follows. In the next section, we introduce and describe the materials and methods. Section 3 presents the results and discussion. Finally, Section 4 summarizes our findings and conclusions.

2. Materials and methods

This section introduces our optimization approach, which aims to provide the most informative scales of the NMBE shape descriptor using different cost functions (e.g. SI, DB or CH). By exploring each cost function, the proposed algorithm enables the optimized NMBE to better characterize shapes from the Illicit and NLM Pills, and Hands datasets. Fig. 1 illustrates the processing pipeline. It starts by computing NMBE [21] with random scale values and selects a single cost function to assess the cluster quality induced by the feature extraction output. Driven by the cost function, the optimization process then searches for new scales to tune the scale

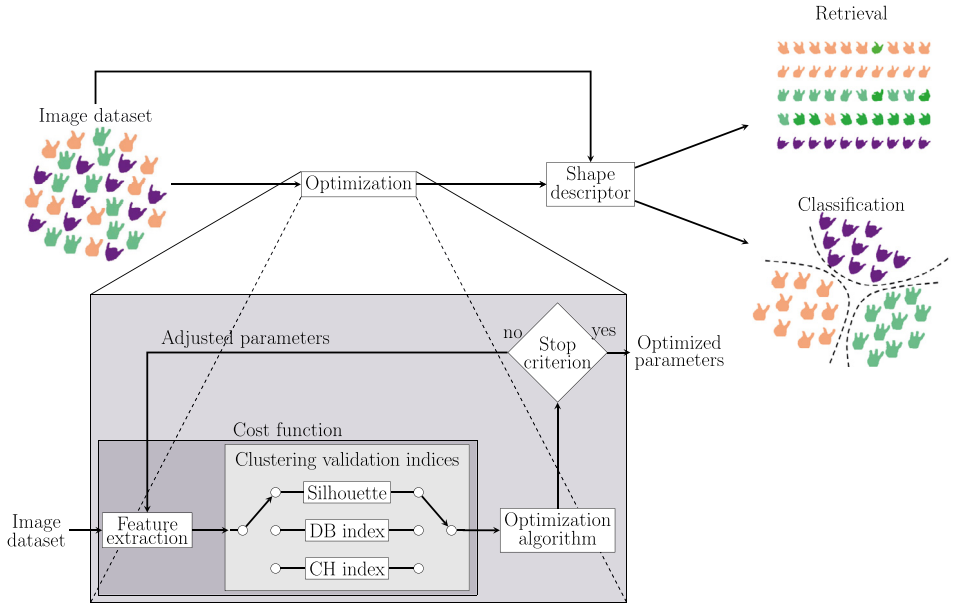
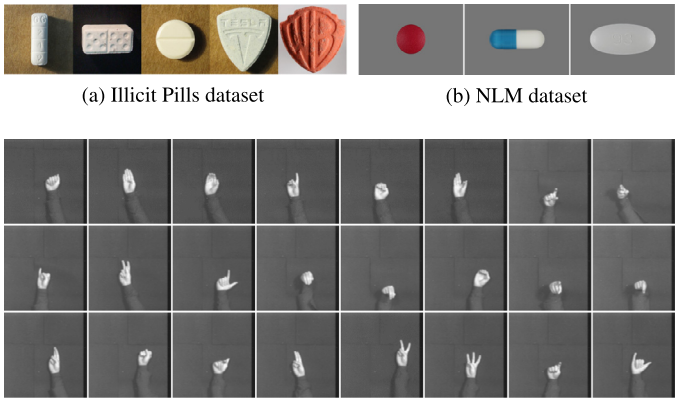


Fig. 1. Optimization methodology of a multiscale shape descriptor.

Table 1
Clustering validation measures.

Measure	Definition	Bounds	Runtime [31]
SI (max)	$\frac{1}{N_C}\sum_i \left\{ \frac{1}{n_i} \sum_{x \in C_i} \frac{b(x)-a(x)}{max(b(x),a(x))} \right\}$	$(-1, 1)$	$\mathcal{O}(n^2)$
DB (min)	$\frac{1}{N_C}\sum_i \max_{j,j\neq i} \left\{ \frac{1}{d(c_i,c_j)} \sum_{k=\{i,j\}} \frac{1}{\eta_k} \sum_{x \in C_k} d(x,c_k) \right\}$	$(0, \infty)$	$\mathcal{O}(n)$
CH (max)	$\frac{\sum_i \eta_i d^2(c_i,c)/(N_C-1)}{\sum_i \sum_{x \in C_i} d^2(x,c_i)/(N_D-N_C)}$	$(0, \infty)$	$\mathcal{O}(n)$

N_D: number of objects in the dataset; c: centroid of the dataset; N_C: number of clusters; C_i: the ith cluster; n_i: number of objects in C_i; c_i: centroid of C_i; $d(x, y)$: distance between x and y; $a(x)$ is the average distance from x to the other elements in the same cluster as x; $b(x)$ is the minimum average distance from x to elements in a different cluster, minimized over clusters.



(a) Illicit Pills dataset (b) NLM dataset

(c) Hands gestures dataset

Fig. 2. Samples from each class in each of the investigated datasets.

parameters of the shape descriptor until the algorithm reaches the stop criterion. Here, the Simulated Annealing (SA) algorithm guides the search for the optimal set of scale parameters. We evaluated the extracted features using the optimized NMBE in CBIR and classification experiments using healthcare image datasets. As reported in [7], the SA and Differential Evolution (DE) algorithms performed more efficiently than Particle Swarm Optimization (PSO). Thus, we selected SA rather than DE because the former demands a lower number of cost function evaluations per round [7], which improves the computational efficiency.

2.1. Datasets

The Illicit Pills dataset[1] comprises 117 elements distributed in five classes which represent illicit pills with distinct shapes. Here, the NLM Pills dataset[2] is composed of 498 pill shapes corresponding to round tablets, capsules and tablets. We selected a subset of the original dataset and excluded the class "Others" because it contains shapes that did not match any other class. Furthermore, the random selection of 166 pill images out of the three remaining classes aimed at balancing this dataset. The Hands dataset[3] consists of 2038 elements that correspond to letters from the international hand alphabet, unevenly distributed in 24 classes. Fig. 2 a–c exhibit samples from these datasets.

[1] The masks of the Illicit Pills dataset [18] that were used here are available online at https://doi.org/10.17632/bmsz34n3dk.2. These masks correspond to a subset of elements from the large dataset available in www.ecstasydata.org.

[2] The NLM Pills dataset [14,19] is available online at www.webmd.com/pill-identification/default.htm.

[3] The Hands dataset [20] is available online at www-prima.inrialpes.fr/FGnet/data/12-MoeslundGesture/database.html.

2.2. Cost functions

The cost functions in the proposed optimization approach are designed to measure the quality of the solutions in terms of how well grouped the samples are. Clustering validation measures are therefore employed as cost functions, and the solutions are sets of shape descriptors parameters obtained by the SA metaheuristic. For the NMBE descriptor, these sets of optimized parameters are the shape descriptor scales.

Table 1 summarizes the cost functions under investigation, their bounds, their optimal values (i.e. maximum (max) or minimum (min)) and the computational complexity. We selected these measures because they are not closely related to the ones used to evaluate the performance of the shape descriptor in CBIR and classification experiments. Moreover, these functions use different methods to evaluate the quality of clusters, in terms of their compactness and separation. SI measures the degree of affinity between a given object and its closest class, assuming values between $(-1, 1)$. The higher value indicates the better assignment of objects into clusters. When an object is close to a boundary, the SI value is close to zero, and hence there is uncertainty about which class the object belongs to [5,8]. The mean Silhouette value of the objects labeled as being in the same class is indicative of the cluster quality. The DB index quantifies the cluster compactness and separation. This metric is based on the ratio of the sum of within-cluster scatter to between-cluster separation, taking into account cluster samples and centroids [29]. DB is a non-negative index in which minimal values indicate well-defined clusters. CH is a measure of cluster validity based on the average between-cluster and within-cluster sums of squares [27]. Likewise, both DB and CH use cluster centers to reckon the distances between samples. However, CH measures the sample separation in terms of the centroid of the dataset, rather than in terms of the cluster centroid [30].

Fig. 2 a–c show that each dataset poses its own challenges to the proposed optimization methodology for parameter adjustment in terms of the meaningful multiscale characteristics of the objects, cluster arrangements and similarity within and between clusters, to name a few. To understand how the cost functions handle these different aspects related to the datasets, we present two distinct synthetic datasets, namely A and B, as shown in Fig. 3a and b. Each dataset comprises five clusters and 100 elements. These samples are displayed into a two-dimensional space that follows a multivariate normal distribution with identity covariance matrices I_2 and mean vectors given by: dataset A – [10, 20], [20, 50], [30, 40], [40,30] and [50, 20]; dataset B – [70, 80], [22, 52], [20, 50], [32, 42], [30, 40]. Note that the respective vector position is in accordance with the class number, as shown in Fig. 3a and b. The clusters are assigned to different shades of gray and the starred points represent the centroids. The clusters in dataset A are compact and separated, whereas in dataset B there are two pairs of overlapping

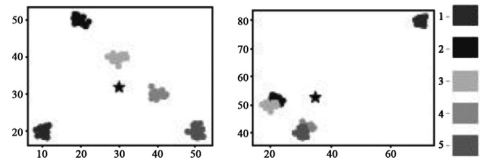


Fig. 3. Synthetic datasets A and B with: a) non-overlapping clusters and b) overlapping clusters, respectively.

clusters. Fig. 3 also presents the clustering evaluation results for both synthetic datasets. The results demonstrate that unlike with SI and DB, CH is strongly affected by the class five in dataset B, because CH relies on the distance between the global dataset centroid and cluster centroids. This analysis raises an interesting point about the clustering quality that can be misleading: CH indicates that dataset B presents the best cluster arrangement which departs from the information provided by SI and DB.

2.3. Multiscale shape descriptor

The bending energy for a specific scale σ is defined as E_σ [21]. It relies on the contour curvature, a differential geometry concept with interesting properties to describe shapes [32]. The derivative nature of the discrete curvature makes its calculation highly sensitive to noise. Therefore, its direct application to digital image contours is unfeasible, due to the noisy nature of these contours. Thus, in [33], the concept of curvature scale space was introduced as a robust method for curvature-based shape-invariant description. It represents shapes at various levels of details and employs a Gaussian kernel in the spatial frequency or spatial domain to filter out the parametric coordinates of the shape contour prior to its curvature calculation. Thus, the standard deviation, σ, of the Gaussian kernel represents the scale that controls the details for shape representation. Here, we perform this operation in the frequency domain. The Normalized Multiscale Bending Energy is given by [7]:

$$\mathrm{NMBE} = (\log E_{\sigma_1},\ \log E_{\sigma_2},\ \ldots,\ \log E_{\sigma_M}), \tag{1}$$

where

$$E_\sigma = \frac{L^2}{N}\sum_{n=0}^{N-1} |\kappa(\sigma, n)|^2, \tag{2}$$

L represents the contour perimeter, N is the number of points on the contour and $\kappa(\sigma, n)$ is the N-sampled value of the multiscale curvature at scale σ.

2.4. Performance assessment

Quantitative evaluation

We perform quantitative evaluation of the retrieval experiments in terms of the MAP metric [34] and the precision recall curves [23,24]. For the classification tests, we applied the Monte Carlo cross-validation [35], and each set of shape descriptors was split into two random subsets (70% and 30%) for training and testing, respectively. Then, we repeated this process for 100 rounds and we computed the mean Accuracy of the kNN classifier ($k = 3$).

The MAP [22] is derived from the average precision $AP(Q)$ that is computed for each query image Q in the ranking list, i.e. a list of objects in the decreasing order of similarity with query. $AP(Q)$ is given by:

$$AP(Q) = \frac{1}{S}\sum_{i=1}^{N_D} P(i) \times rel(i), \tag{3}$$

where $P(i)$ is the precision up the position i of the ranking list. Precision is the ratio of the retrieved objects that belong to the same class as the query image Q and the number of retrieved objects i, $rel(i)$ is a binary function, that is equal to 1 if the object i belongs to the same class as the query image Q, and 0 otherwise. N_D is the number of total objects in the dataset, and S is the number of objects of the same class obtained by the query object. The MAP is calculated by averaging the AP score over all query images [34].

To compute the precision recall curves, we plot the average values of precision versus recall over a threshold modulating the number of retrieved images i. Note that the average values of the precision measure are different from the average precision $AP(Q)$ in Eq. (3), because the precision $P(i)$ is not multiplied by $rel(i)$. Recall is the ratio of the retrieved objects that belong to the same class as the query image Q and the number of objects that belong to the same class as the query in the whole dataset. An ideal precision recall curve has precision of one over all recall levels [23,24].

The Accuracy (ACC) measure is defined in terms of the ratio of correctly predicted classes to the number of total predicted classes. The MAP and Accuracy measures are within the range (0,1] and higher values indicate better performance.

Qualitative evaluation

We applied the Kohonen clustering algorithm [36] represented by the U-matrix [37] to assess the quality of the clustering arrangements. The U-matrix displays the shapes in their corresponding topology using a particular color for each class. The darker the cluster borders in the U-matrix, the greater the neighbor distance in the topology. Moreover, a well-defined cluster is surrounded by a dark border and presents all elements close to each other with a light neighborhood.

2.5. Design of experiments

We carried out experiments by applying our methodology to the NMBE descriptor using each cost function separately. Our experiments on three datasets considered five scales, corresponding to the σ parameter values to be optimized by SA in the range [0.4,100]. We tuned SA with the following parameters: initial temperature ($T_0 = 100$), epoch length ($L_e = 10$), cooling ratio ($\alpha = 0.9$) of the exponential cooling schedule as reported in [38], and number of rounds ($N_r = 100$) the stop criterion.

The computational cost of the whole optimization process depends on the number of times that the chosen cost function is evaluated and its computational complexity. In our SA implementation, the cost function runs $L_e \times N_r = 1000$ times. Table 1 summarizes the time complexity of each clustering validation measure, and we note that DB and SI demand less linear $O(n)$ compared to the complexity of SI, i.e. $O(n^2)$. The optimization algorithm runs 30 times for each cost function to optimize the NMBE descriptor. Thus, we calculate the mean and variance of the evaluation measures for retrieval and classification experiments.

All CBIR and classification experiments and the Kohonen clustering algorithm use a Euclidean distance to measure the shape similarity of features from the healthcare images.

3. Results and discussion

In this paper, the cost functions are used to assess the cluster quality and to adjust the scales of NMBE according to the shape details of each healthcare dataset. In fact, these datasets posed different challenges to our approach, which were revealed by the cost functions and the descriptor scales. Thus, the differences among the three datasets imply different scales for the same descriptor in order to reveal subtle shape details.

Table 2
Mean average precision (MAP) and Accuracy (ACC) values for the retrieval and classification results.

	NMBE+SI		NMBE+DB		NMBE+CH		Non-optimized NMBE [21]		Inception-ResNet-v2 [28]	
Dataset	MAP(%)	ACC(%)	MAP(%)	ACC(%)	MAP(%)	ACC(%)	MAP(%)	ACC(%)	MAP(%)	ACC(%)
Illicit Pills	95.83 ± 0.31	97.94 ± 0.28	94.32 ± 3.47	96.77 ± 2.38	67.23 ± 2.33	78.72 ± 3.48	72.58	90.11	83.87	93.64
NLM Pills	97.45 ± 0.09	99.24 ± 0.13	97.44 ± 0.07	99.24 ± 0.09	97.56 ± 0.03	99.34 ± 0.06	90.38	99.36	85.33	99.46
Hands	51.66 ± 1.89	82.44 ± 2.81	48.36 ± 2.32	76.89 ± 4.44	42.92 ± 2.86	64.43 ± 5.69	43.12	79.48	56.18	95.48

Table 2 presents the retrieval and classification results achieved with NMBE optimized with SI, DB and CH (NMBE+SI, NMBE+DB and NMBE+CH, respectively), the non-optimized NMBE introduced in [21], and Inception-ResNet-v2 [28]. We employed the Inception-ResNet-v2 available in the open-source Python tool pyCBIR [34] for feature extraction which delivers 1536 features per image for each dataset.

In order to perform a fair comparison, Inception-ResNet-v2 was also applied to the segmented shapes of the healthcare datasets, as for the other methods. This deep learning-based method requires thousands of images for training, mainly to avoid overfitting. Our datasets contain a reduced number of images, and we therefore used the Inception-ResNet-v2 which had already been trained on the ImageNet dataset [39] consisting of around 1 million elements divided into 1000 classes. The main goal of this training process is to allow the convolutional layers to learn and extract different attributes such as: texture, shape and color, which are used to characterize the input images in order to minimize the classification error. The attributes are then obtained by removing the last fully connected layer, known as the classification layer [34].

Experiments performed with the optimized NMBE demonstrated that it is capable of improving the retrieval results for the Illicit Pills and Hands datasets. We can also observe from Table 2 that the Illicit Pills and Hands datasets posed challenges to NMBE+CH. In fact, the scales optimized with CH were not able to capture global and local shape details of these challenging datasets. This is due to the overlapping clusters in the Illicit Pills and Hands datasets, which can be seen in their respective U-matrices (Figs. 5c and 7 c) as mixed and unbounded groups. The presence of overlapping clusters hinders the efficiency of CH in adequately revealing the subtle details of the shapes in these datasets, as already mentioned in Section 2.2. However, the CH cost function achieved similar results to the others for the NLM Pills dataset, as shown in Fig. 6c, which does not provide evidence of overlapping between pairs of clusters.

Conversely, NMBE+SI and NMBE+DB outperformed the non-optimized descriptor. Although the Inception-ResNet-v2 [28] performed better than the optimized NMBE for the Hands dataset, NMBE+SI achieved closer results. The presence of images of hand gestures in the ImageNet dataset may explain the good performance of the Inception-ResNet-v2. For the Illicit Pills dataset, NMBE+SI and NMBE+DB performed better on both retrieval and classification tasks than the Inception-ResNet-v2. This confirms the reliance of shape recognition on tailored shape description algorithms for particular datasets.

The classification results with the NLM Pills showed that the optimization approach using SI, DB and CH performed similarly to the non-optimized method and Inception-ResNet-v2. However, the retrieval results significantly benefited from our optimization approach. In fact, the three classes of pills exhibit small differences within classes and are well separated, which may explain the similarity in the results. This means that a specific dataset requires the selection of an appropriate cost function to take advantage of the multiscale nature of the NMBE descriptor and its ability to represent fine to coarse shape details [7].

It is worth noting that there is a trade-off between the proposed optimization methodology and the cost function complexity (see Table 1). Overall, the MAP and Accuracy values in Table 2 led us to conclude that SI and DB are more likely to improve the performance of the optimized multiscale descriptor than CH, for the Illicit Pills and Hands. However, CH and DB are computationally equivalent, while SI presents a high computational cost, as shown in Table 1. Table 2 shows that CH performed worse than DB and the non-optimized method for the Illicit Pills and Hands. The poor performance of the Inception-ResNet-v2 in retrieving the NLM pills may be due to the its reliance on a large number of samples during the training process. The ImageNet dataset is, indeed, poorly correlated to our image datasets, especially the datasets of pill images. It worth noting that the Inception-ResNet-v2 requires thousands of images for training, and describes an image with a vector of 1536 features, whereas NMBE does it with only five features.

We computed the precision recall curves to evaluate our retrieval results. Fig. 4 shows the average values of precision versus recall. For the Illicit Pills, Fig. 4a indicates NMBE+SI and NMBE+DB outperformed the other methods, while NMBE+CH and the non-optimized NMBE achieved the poorest results and the Inception-ResNet-v2 gave intermediate outcomes. For the NLM Pills dataset in Fig. 4b, the optimized NMBE with SI, DB and CH performed

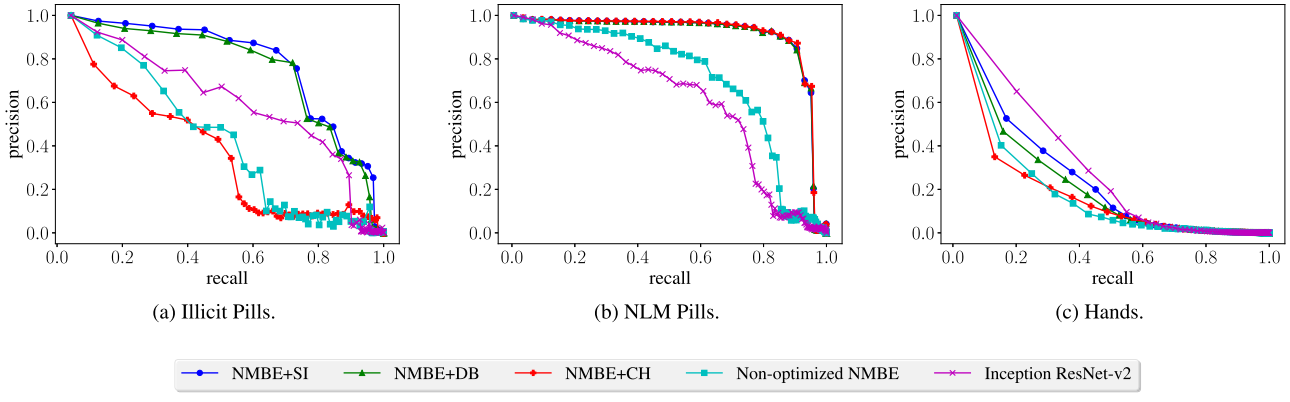
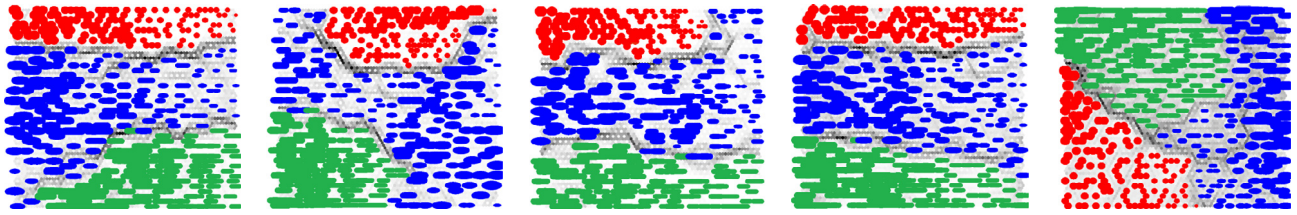


Fig. 4. Precision recall curves for: a) Illicit Pills, b) NLM Pills and c) Hands, respectively.

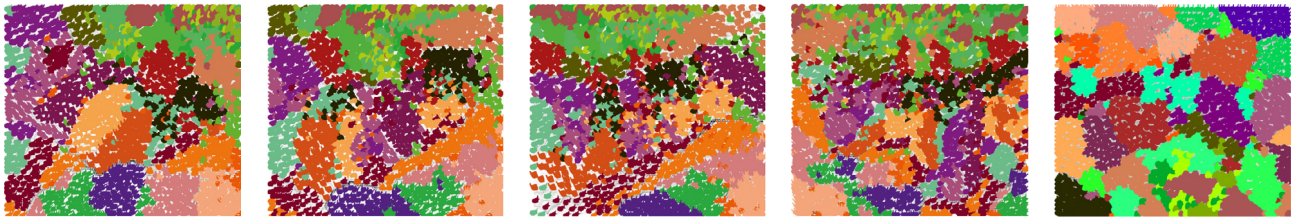
(a) NMBE+SI. (b) NMBE+DB. (c) NMBE+CH. (d) Non-optimized [20]. (e) Inception-ResNet-v2[27].

Fig. 5. U-matrices for shapes from the Illicit Pills using the optimized and non-optimized NMBE and Inception-ResNet-v2 [28].



(a) NMBE+SI. (b) NMBE+DB. (c) NMBE+CH. (d) Non-optimized [20]. (e) Inception-ResNet-v2[27].

Fig. 6. U-matrices for shapes from the NLM Pills using the optimized and non-optimized NMBE and Inception-ResNet-v2 [28].



(a) NMBE+SI. (b) NMBE+DB. (c) NMBE+CH. (d) Non-optimized [20]. (e) Inception-ResNet-v2[27].

Fig. 7. U-matrices for shapes from the Hands using the optimized and non-optimized NMBE and Inception-ResNet-v2 [28].

similarly, and these outperformed the other methods. These curves confirmed that the NLM Pills dataset benefited significantly from our methodology. The curves for the challenging Hands dataset in Fig. 4c show that the Inception-ResNet-v2 achieved the best result, but its advantage against the other methods decreases markedly when the number of retrieved images increases.

Fig. 5 shows the U-matrices for the Illicit Pills obtained with the optimized and non-optimized NMBE and Inception-ResNet-v2. The optimization with SI and DB achieved similar results and outperformed the others. Furthermore, the optimization with CH underperformed the non-optimized NMBE, indicating that it is not a suitable cost function for use with Illicit Pills. The U-matrix of the Inception-ResNet-v2 in Fig. 5e merged two rectangular-shaped pills, and performed worse than NMBE+SI and NMBE+DB. However, it gives results that are better than NMBE+CH and the non-optimized version.

Fig. 6 displays the U-matrices for the NLM Pills dataset and indicate that the optimized and non-optimized descriptors present similar results. This is due to the geometry of the shapes and the high level of within-class similarity in this dataset. In this case, our methodology achieved fairly similar results regardless the cost function. The results of the Inception-ResNet-v2 show that there are classes separated in subsets, and the dark frontiers represent a large distance between the samples.

Fig. 7 a shows the U-matrices for the Hands dataset. The U-matrix of the Inception-ResNet-v2 presents compact clusters, although there are several sub-clusters within the same class. The clusters in Fig. 7a, b and d are better organized than in Fig. 7c. The U-matrices using SI and DB achieved similar results and the cluster arrangements provided by NMBE+CH and the non-optimized NMBE are worse than the others. Fig. 7c exhibits more regions with merged elements of different classes than Fig. 7a and b, confirming that CH is an inappropriate cost function to deal with the Hands dataset. We concluded that CH cannot optimize the scale parameters of NMBE in order to provide a more reliable shape representation for this dataset. One major finding from our results is that the cost function selection and the dataset are closely related and hence, this selection impacts directly on the performance of the optimization algorithm. Note that the U-matrices in Figs. 5–7 confirmed the results in Table 2.

4. Concluding remarks

This paper investigates three clustering validation measures, as cost functions (SI, DB and CH) within a metaheuristic optimization solution for parameter adjustment of a multiscale shape descriptor. We search for the set of descriptor parameters that best describes the shapes of three healthcare image datasets. Our experiments confirmed the assumption that different datasets may require different sets of parameters such that the NMBE descriptor extracts the best shape features. Our solution for parameter adjustment improves the performance of the multiscale shape descriptor and, furthermore, it is a feasible alternative to the manual setting. The benefits of the optimization methodology for the NLM Pills are confirmed by the precision recall curves and U-matrices, which show significant improvement over the non-optimized NMBE. The optimization approach with CH was unable to output multiscale parameters that produced a more reliable shape representation for the Illicit Pills and Hands datasets. The calculation of CH relies on the distance between global dataset centroid and cluster centroids and is therefore strongly sensitive to the presence of distinct classes with a high intergroup dissimilarity. Overall, NMBE

optimized with SI and DB achieved similar results to, or outperformed, the Inception-ResNet-v2. Thus, we state that the role of the cost function in the parameter optimization is essential to achieve the best performance of a multiscale descriptor in retrieval and classification experiments for a specific dataset. It is worth mentioning that the proposed processing pipeline can be extended to other shape descriptors and other datasets. The main limitation of our method is the absence of specific criteria to guide the selection of the cost function that best matches the problem domain. Another challenge is that parameter optimization may require much computational effort, especially for large datasets, since it evaluates the cost function and performs feature extraction several times. Thus, the selection of the function should not consider only its adequacy for the problem domain but also its computational complexity.

Acknowledgments

This work was developed with partial financial support from CNPq (#444784/2014 – 4, #401442/2014 – 4 and #306600/2016 – 1) and CAPES-Brazil (Finance Code 001). We thank Dr. Deborah Maria Vieira Magalhães for the enlightening suggestions. We are also grateful to the reviewers for their valuable comments, which greatly improved the quality of the manuscript.

Supplementary material

All necessary code files to optimize the Normalized Multiscale Bending Energy descriptor (NMBE) by Simulated Annealing (SA) with the cost functions: Silhoutte, Davies-Bouldin index or Calinski-Harabasz index are available online at https://github.com/AllanCCarneiro/Descriptor_Optimization.git and the Illicit Pills dataset is available online at https://doi.org/10.17632/bmsz34n3dk.2.

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.patrec.2019.06.017

References

[1] L.F. Costa, R.M. Cesar Jr, Shape Classification and Analysis: Theory and Practice, Second ed., CRC Press Taylor & Francis Group, Boca Raton, Florida, FL, USA, 2009.
[2] H. Ling, D.W. Jacobs, Shape classification using the inner-distance, IEEE Trans. Pattern Anal. Mach.Intell. 29 (2) (2007) 286–299.
[3] J. Wang, X. Bai, X. You, W. Liu, L.J. Latecki, Shape matching and classification using height functions, Pattern Recognit. Lett. 33 (2) (2012) 134–143.
[4] B. Wang, D. Brown, Y. Gao, J. La Salle, March: multiscale-arch-height description for mobile retrieval of leaf images, Inf. Sci. 302 (2015) 132–148.
[5] A.C. Carneiro, J.G.F. Lopes, J.F.S.R. Neto, M.M.S. Souza, F.N.S. Medeiros, F.N. Bezerra, On the evaluation of cost functions for parameter optimization of a multiscale shape descriptor, in: 2017 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), 2017, pp. 045–050.
[6] L.J. Latecki, R. Lakamper, T. Eckhardt, Shape descriptors for non-rigid shapes with a single closed contour, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 1, IEEE, 2000, pp. 424–429.
[7] M.M.S. Souza, F.N.S. Medeiros, G.L.B. Ramalho, I.C. de Paula Jr, I.N.S. Oliveira, Evolutionary optimization of a multiscale descriptor for leaf shape analysis, Expert Syst. Appl. 63 (2016) 375–385.
[8] P.J. Rousseeuw, Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, J. Comput. Appl. Math. 20 (1987) 53–65.
[9] UNODC, United Nations Office on Drugs and Crime: World Drug Report 2010, Technical Report, United Nations Publication, Sales No. E.11.XI.10, 2011.
[10] J. Camargo, P. Esseiva, F. Gonzálezre, J. Wist, L. Patiny, Monitoring of illicit pill distribution networks using an image collection exploration framework, Forensic Sci. Int. 223 (1) (2012) 298–305.
[11] Y.-B. Lee, U. Park, A.K. Jain, S.-W. Lee, Pill-id: Matching and retrieval of drug pill images, Pattern Recognit. Lett. 33 (7) (2012) 904–910. Special Issue on Awards from ICPR 2010
[12] M. Lopatka, W. van Houten, Automated shape annotation for illicit tablet preparations: a contour angle based classification from digital images, Sci. Justice 53 (1) (2013) 60–66.
[13] K.T. Maddala, R.H. Moss, W.V. Stoecker, J.R. Hagerty, J.G. Cole, N.K. Mishra, R.J. Stanley, Adaptable ring for vision-based measurements and shape analysis, IEEE Trans. Instrum.Meas. 66 (4) (2017) 746–756.
[14] D. Ushizima, A. Carneiro, M. Souza, F. Medeiros, Investigating pill recognition methods for a new national library of medicine image dataset, in: G. Bebis, R. Boyle, B. Parvin, D. Koracin, I. Pavlidis, R. Feris, T. McGraw, M. Elendt, R. Kopper, E. Ragan, Z. Ye, G. Weber (Eds.), Advances in Visual Computing, Lecture Notes in Computer Science, 9475, Springer International Publishing, Cham, 2015, pp. 410–419.
[15] Y.F. Wong, H.T. Ng, K.Y. Leung, K.Y. Chan, S.Y. Chan, C.C. Loy, Development of fine-grained pill identification algorithm using deep convolutional network, J. Biomed. Inform. 74 (2017) 130–136.
[16] M.M. Islam, S. Siddiqua, J. Afnan, Real time hand gesture recognition using different algorithms based on american sign language, in: 2017 IEEE International Conference on Imaging, Vision Pattern Recognition (ICIVPR), 2017, pp. 1–6.
[17] V. Sombandith, A. Walairacht, S. Walairacht, Hand gesture recognition for lao alphabet sign language using hog and correlation, in: 2017 14th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2017, pp. 649–651.
[18] A.C. Carneiro, J.G.F. Lopes, F.N.S. Medeiros, Illicit Pills, 2019. Mendeley Data, V2, https://doi.org/10.17632/bmsz34n3dk.2.
[19] Z. Yaniv, J. Faruque, S. Howe, K. Dunn, D. Sharlip, A. Bond, P. Perillan, O. Bodenreider, M.J. Ackerman, T.S. Yoo, The national library of medicine pill image recognition challenge: an initial report, in: 2016 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), IEEE, 2016, pp. 1–9.
[20] H. Birk, T.B. Moeslund, C.B. Madsen, Real-time recognition of hand alphabet gestures using principal component analysis, in: In 10th Scandinavian Conference on Image Analysis, 1, 1997, pp. 261–268.
[21] R.M. Cesar Jr, L.F. Costa, Application and assessment of multiscale bending energy for morphometric characterization of neural cells, Rev. Sci. Instrum. 68 (5) (1997) 2177–2186.
[22] Y. Hu, et al., Multiple-instance ranking: learning to rank images for image retrieval, in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2008, pp. 1–8.
[23] N.V. Shirahatti, K. Barnard, Evaluating image retrieval, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 1, 2005, pp. 955–961vol. 1, doi:10.1109/CVPR.2005.147.
[24] O. Chum, J. Philbin, J. Sivic, M. Isard, A. Zisserman, Total recall: automatic query expansion with a generative feature model for object retrieval, in: 2007 IEEE 11th International Conference on Computer Vision, 2007, pp. 1–8, doi:10.1109/ICCV.2007.4408891.
[25] J.Z. Wang, Integrated Region-Based Image Retrieval, 11, Springer Science & Business Media, 2001.
[26] D.L. Davies, D.W. Bouldin, A cluster separation measure, IEEE Trans. Pattern Anal. Mach.Intell. PAMI-1 (2) (1979) 224–227.
[27] T. Caliński, J. Harabasz, A dendrite method for cluster analysis, Commun. Stat.-Theory Methods 3 (1) (1974) 1–27.
[28] C. Szegedy, S. Ioffe, V. Vanhoucke, Inception-v4, inception-resnet and the impact of residual connections on learning, Comput. Res. Repos. (2016) 1–12. abs/1602.0726
[29] C.-H. Chou, M.-C. Su, E. Lai, A new cluster validity measure and its application to image compression, Pattern Anal. Appl. 7 (2) (2004) 205–220.
[30] J. Baarsch, M.E. Celebi, Investigation of internal validity measures for k-means clustering, in: Proceedings of the International MultiConference of Engineers and Computer Scientists, 1, 2012, pp. 14–16.
[31] S. Saitta, B. Raphael, I.F. Smith, A comprehensive validity index for clustering, Intell. Data Anal. 12 (6) (2008) 529–548.
[32] V.V. Kindratenko, On using functions to describe the shape, J. Math. Imaging Vis. 18 (3) (2003) 225–245.
[33] F. Mokhtarian, A. Mackworth, A theory of multiscale, curvature-based shape representation for planar curves, IEEE Trans. Pattern Anal. Mach.Intell. 14 (8) (1992) 789–805.
[34] F.H.D. Araujo, R.R.V. Silva, F.N.S. Medeiros, D.D. Parkinson, A. Hexemer, C.M. Carneiro, D.M. Ushizima, Reverse image search for scientific data within and beyond the visible spectrum, Expert Syst. Appl. 109 (2018) 35–48.
[35] W. Dubitzky, M. Granzow, D.P. Berrar, Fundamentals of Data Mining in Genomics and Proteomics, Springer Science & Business Media, 2007.
[36] , Self-Organizing Maps, T. Kohonen, M.R. Schroeder, T.S. Huang (Eds.), Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2001.
[37] A. Ultsch, H.P. Siemon, Kohonen's self organizing feature maps for exploratory data analysis, in: Proceedings of International Neural Networks Conference, 1, Kluwer Academic Press, Paris, 1990, pp. 305–308.
[38] M.-W. Park, Y.-D. Kim, A systematic procedure for setting parameters in simulated annealing algorithms, Comput. Oper. Res. 25 (3) (1998) 207–217.
[39] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei, Imagenet large scale visual recognition challenge, Int. J. Comput. Vis. 115 (3) (2015) 211–252.