Boosting Shape Classifiers Accuracy by Considering the Inverse Shape

Sébastien Piérard and **Antoine Lejeune** and **Marc Van Droogenbroeck**

{Sebastien.Pierard, Antoine.Lejeune, M.VanDroogenbroeck}@ulg.ac.be

INTELSIG Laboratory, Montefiore Institute, University of Liège, Belgium

Abstract

Many techniques exist for describing shapes. These techniques almost exclusively consider the contour or the inside of the shape; the major problem for describing the outside of a shape, or inverse shape, being that it has an infinite extension. In this paper, we show how to adapt two shape descriptors , one region based, the Cover By Rectangles, and one transform based, the Zernike moments, to be applicable to the inverse shape. We analyze their properties, and show how to deal with the infinite extension of the inverse shape. Then, we apply these descriptors to shape classification and compare representations that use the shape, its inverse, or both. Our experiments establish that, for shape classification, a representation integrating the inverse shape often outperforms a representation restricted to the shape. This opens the path for better techniques that could use, as a rule of thumb, both the representations of a shape and its inverse for the purpose of classification.

Keywords: Shape descriptor, Object classification, Inverse shape, Cover by rectangles, Zernike moments

1. Introduction

The shape of an object is an essential information for many tasks in computer vision including object classification, recognition, and information retrieval. Consequently, many shape representation and description techniques have been proposed and tried in the past. Surveys on shape description can be found in [18, 33, 35], and reviews on their use for applications in [2, 29, 31]. According to Zhang *et al.* [35], shape description techniques can be classified into two families: *contour-based* methods and *region-based* methods. Another, slightly different, taxonomy is proposed by da Fontoura *et al.* [8]. Their three families are based on the contours, regions, and transforms, respectively. They also subdivide region-based methods in three classes, which rely on a region decomposition, bounding regions, or internal features. Region-based methods include the well-know skeleton and medial axis [6] transform, the morphological shape decompositions [25], and other techniques that represent a shape as a combination of surfacic primitives. These components indirectly encode the topology of the object [5].

For the purpose of shape description, region-based approaches can use the object itself, or its complementary set (which includes the holes) named *inverse shape* hereafter. This is illustrated in Figure 1. Note that, for contour-based descriptors, there is no difference in describing a shape or its inverse, so that considering both descriptors does not increase the classification performance. Also, note that some alternative terminologies such as inner/outer, inside/outside, foreground/background exist to designate the shape and its inverse, but they might be ambiguous when the object contains holes or has an even more complex shape. The main contribution of this work is to highlight the boost of shape classifi-

Shape of an object X $\neq$ Inverse shape of X
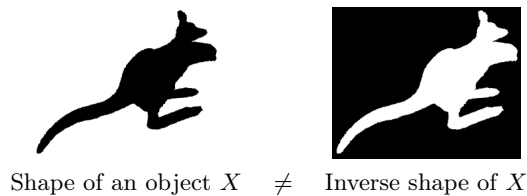
Figure 1: The main idea of this paper can be summarized as follows. Despite that the shape of an object and its inverse, both drawn in black, are dual types of information, they perform differently for the task of shape classification. A combination of both the shape and its inverse achieves the best classification results.

cation performance that can be obtained by considering the inverse shape. It seems unusual to consider the inverse shape in the common practice, and the resulting achievable boost is largely overlooked in the theoretical literature about shape description and classification. However, from our experience, it is always worth considering the inverse shape because the expected improvement is not related to a particular choice of region-based shape description technique, or to a particular choice of classification technique. In this paper, we illustrate the universality and the high potential of the technique with two shape descriptors and two classifiers. Despite the infinite extension of the inverse shape, we were able to adapt seamlessly the two descriptors to it.

While many papers compare shape descriptors, we are not aware of papers that evaluate the performance of a shape descriptor and its inverse descriptor simultaneously, except for the work of Kresch and Malah [16], which evaluates the performance from a coding perspective. In their conclusions, Kresch and Malah state that positive and negative skeletons, which are related to the shape and its inverse respectively, are not equivalent. One possible explanation for the lack of works that deal with the inverse might be the difficulty to deal with its infinite extension. While the natural way to handle this extension is to restrict it to the domain of definition or the image frame, this practical solution unfortunately introduces a bias towards the frame size and shape. Another explanation might be the intuitive, but wrong, impression that shape and its inverse carry the same type of information. This is discussed in this paper for shapes, but it has already been experienced that the outside of objects contributes to raise the recognition performance of classifiers in papers such as [1, 9, 27, 32].

The paper is organized as follows. Section 2 shows how we can adapt existing shape descriptors to deal with the inverse shape, and how to use them for classification tasks. In Section 3, we present several classification tasks and evaluate the performance of representation strategies for three case studies related to character recognition and binary silhouette classification. We establish that, although informations contained in the shape and its inverse are apparently equivalent, recognition performances based on the two representation strategies differ. For some families of shapes, a description of the shape or its inverse can outperform the dual descriptor significantly. In other cases, the combination of the two representations is the most promising strategy. Finally, Section 4 concludes this paper.

2. Adapting shape descriptors for leveraging the inverse shapes in classifiers

The main challenge when describing the inverse shape relates to its infinite extension. For example, we need shape descriptors that have an output size (*e.g.* the number of components for descriptions based on region decomposition, or values in the case of descriptions based on transformations) that is independent with respect to an increase of the size of the arbitrary
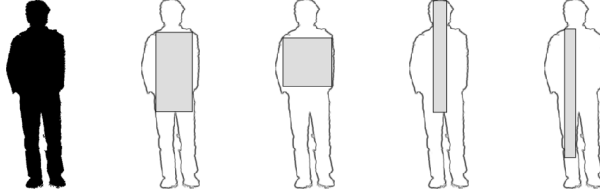
Figure 2: Some of the largest aligned rectangles included in an object (human silhouette); they form a subset of the cover by rectangles (adapted from [3]).

limit considered around the object (*e.g.* the bounding box, the image frame, the bounding circle, etc.). If not, the description of the inverse shape would be conditioned by the size of this limit, and its computation would become intractable. Also, it is important to consider how the descriptor is encoded. When dealing with shapes with an infinite extension, some of their characteristics, such as the centroids, can become ill-defined.

2.1 Extending the Cover By Rectangles (CBR)

2.1.1 Shape representation

The first shape representation technique considered in this paper is the *Cover By Rectangles* (CBR) descriptor, introduced by Barnich *et al.* [4], which decomposes an object in its largest rectangular components. According to the comprehensive taxonomy of da Fontoura *et al.* [8], this is a region-based descriptor based on a region decomposition. In its original form, the CBR comprises all the rectangles, aligned with the axes, that fit within an object (see Figure 2 for an illustration). More formally, the CBR is defined as follows. Let us consider the family $\mathcal{S}$ of all $m \times n$ pixels large aligned rectangles on a digital grid. Let $\mathcal{R}(X)$ be the set of all the rectangles that fit within X: $\mathcal{R}(X) = \left\{ S_b^i \subseteq X \,|\, S^i \in \mathcal{S} \right\}$, where S_b^i is S^i translated by b. The cover of X by $\mathcal{S}$, denoted $\mathrm{CBR}\,(X)$, is then

$$\mathrm{CBR}\,(X) = \left\{ R \in \mathcal{R}(X) \,|\, \neg\exists R' \in \mathcal{R}(X) \ \text{such that } R \subset R' \right\}. \tag{1}$$

As shown in [30], the CBR is a fast to compute region-based descriptor; it is thus appropriate for real-time applications.

To derive a CBR descriptor usable for the inverse, we first consider the bounding box of X, denoted by $\mathrm{BB}\,(X)$, and expand it by one pixel in all directions (left, right, up, and down). Let us denote by $\mathrm{BB}^1\,(X)$ all the pixels included inside the bounding box of X expanded by one pixel. As the limits imposed by the box are arbitrary, we could as well extend it to infinity to obtain $\mathrm{BB}^\infty\,(X)$. This leads us to propose a variation of the CBR, adapted to the inverse shape and denoted $\mathrm{ICBR}\,(X)$. The cover by rectangles of the inverse of X is formally defined as

$$\mathrm{ICBR}\,(X) = \mathrm{CBR}\,(\mathrm{BB}^\infty\,(X) \setminus X), \tag{2}$$

where $\setminus$ denotes the set difference.

As the computation of $\mathrm{ICBR}\,(X)$ deals with the infinite extension of the complementary set, we elaborate on its cardinality:

1. First, note that if X is not empty, the CBR of $\mathrm{BB}^1\,(X) \setminus X$ and that of $\mathrm{BB}^2\,(X) \setminus X$ have the same amount of rectangles. The difference lays with some rectangles that are extended by one or two rows and by one or two columns. In other words, adding a one pixel large border to the bounding box only extends none, one, two or three borders of each rectangle of the CBR. This reasoning is recursively valid for any dimension
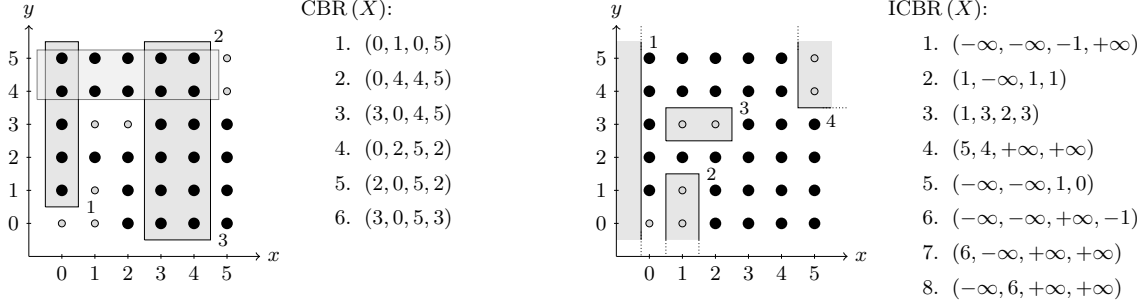
Figure 3: An example of a shape X (pixels are represented by disks: large solid ones for the shape and smaller hollow ones for its inverse) and its $\mathrm{CBR}(X)$ (left), and the inverse of X and $\mathrm{ICBR}(X)$, the new descriptor (right). For both $\mathrm{CBR}(X)$ and $\mathrm{ICBR}(X)$, a rectangle is represented by the coordinates of two of its corners, more precisely by $(\min(x), \min(y), \max(x), \max(y))$. Some of the rectangles of $\mathrm{CBR}(X)$ and $\mathrm{ICBR}(X)$ are drawn in their respective figure. The rectangles are listed only for illustrative purposes.

of the bounding box and, therefore, the cardinality of $\mathrm{CBR}\left(\mathrm{BB}^1(X)\setminus X\right)$ is equal to that of $\mathrm{ICBR}(X)$.

2. In addition, it has been proven in [30] that the cardinality of X is an upper bound of the cardinality of $\mathrm{CBR}(X)$.

Therefore, by combining the two properties, we obtain that the cardinality of $\mathrm{ICBR}(X)$ is upper bounded by the cardinality of $\mathrm{CBR}\left(\mathrm{BB}^1(X)\setminus X\right)$. This can be summarized as follows: the $\mathrm{ICBR}(X)$ is a finite representation, it does not depend on a bounding box, and it is uniquely defined.

From a practical point of view, it is important to consider how the rectangles of $\mathrm{ICBR}(X)$ are encoded. Barnich *et al.* [4] encoded a rectangle by the coordinates of its centroid, its width, and its height. That encoding is inappropriate here as the centroid is undefined for rectangles expanding to infinity on two opposite sides. Therefore, we encode a rectangle by the coordinates of its two opposite corners, as illustrated in Figure 3.

2.1.2 Classification technique

In order to classify shapes based on the CBR representation, we adopt the object classification method of Barnich *et al.* [4], for its simplicity. It has been proven to be efficient for the recognition of human silhouettes [4, 21, 22]. This classifier separates shapes or objects into several classes by means of machine learning methods. The original algorithm, modified to meet our needs, is organized in 3 steps:

1. An object is first represented by its shape or its inverse shape. This results in lists of rectangles such as the ones given in Figure 3. We use five attributes to describe each rectangle: the coordinates of two opposite corners (as shown in Figure 3), and a fifth attribute that indicates if the rectangle belongs to $\mathrm{CBR}(X)$ or $\mathrm{ICBR}(X)$, that is, to the shape or its inverse respectively. In addition, to ensure scale and translation invariance, we encode each rectangle with coordinates relative to the centroid of the object, and choose units along the axes that are normalized with respect to the object sizes (we use the horizontal and vertical standard deviations of an object to define its sizes). Note that this encoding differs from the convention taken to illustrate the representations in Figure 3, but ensures translation and scale invariance. In practice, to avoid difficulties for the machine learning techniques to deal with infinite values,

Figure 4: Noise on the object (a) can lead to five situations: (b, c) result from negative noise, (d, e, f) result from positive noise. When the objects to classify are separated by a connected components analysis, the (e) and (f) cases produce two blobs analyzed separately.

infinite values are encoded as a constant, much larger than all possible finite values produced by the descriptor.

2. For each rectangle of $\mathrm{CBR}(X)$ and $\mathrm{ICBR}(X)$, a machine learning algorithm (the *ExtRaTrees* [14] in our case, which is a forest of decision trees and therefore belongs to the category of classifiers most likely to be the best according to the comparison of Fernandez-Delgado *et al.* [12]), predicts the class of the rectangle. Consequently, each rectangle votes for one class. As proposed in [4] and [21], and for practical reasons, only a random subset of all the rectangles is used in practice.

3. The class assigned to the shape is obtained by thresholding the proportion of rectangles voting for that class (the class assignment is seen as a two classes problem).

In the following, we refer to this classification method as "Barnich's method". Analyzing the shapes by parts, as done in this method, decreases the influence of defects. This part-based shape classification method has the advantage that it does not require to give a semantic meaning to the parts (to the contrary, for example, of what was proposed earlier by Ramoser *et al.* [26]), and therefore can be used in a large variety of shape classification tasks.

2.1.3 Properties

The CBR descriptor is translation but not rotation invariant. For some shape classification tasks, it is an advantage. For example, in our experiments (see Section 3), we consider the task of recognizing human silhouettes. As most human silhouettes appear upright in images, the shape orientation is a relevant characteristic of the class of human silhouettes. Thus, in that case, it is advantageous for the CBR not to be rotation invariant.

Another important property relates to the influence of noise on the CBR as it highlights the benefits for considering the inverse shape. Shape descriptors should be both effective for shape analysis and classification, and robust to the noise occurring during the segmentation process. Kresch *et al.* [16] make the difference between positive noise (*e.g.* stains, which extends the object) and negative noise (*e.g.* holes, which reduces the object). The CBR of the shape is very sensitive to negative noise (as a single pixel wide hole can potentially disturb the whole set of rectangles) but nearly insensitive to positive noise (stains only add a few rectangles to the CBR, and the number of added rectangles is bounded by the area of the stains, as proven by Van Droogenbroeck *et al.* [30]). For the CBR of the inverse shape, it is the contrary. It follows that considering the inverse shape can help improving the robustness of classifiers when the shapes to classify originate from an image segmentation followed by a connected components analysis (in this case, stains are impossible, see Figure 4). Another strategy to increase the robustness consists in filling holes in a preprocessing step. The drawback is that there may be natural holes in the shapes to classify, and that these holes may carry useful information for the classification. It is difficult to give guidelines to decide if holes should be filled or not, since this depends on the probability for the noise to give rise

to holes. If the probability to observe holes due to noise is high, one should either fill holes or consider the inverse shape. The two strategies can be complementary, as our experiments (See Section 3, Figure 9) show that filling holes and considering both the shape and its inverse can lead to a better accuracy than the one achievable by applying only one of the two strategies. This being said, the experiments presented in this paper focus on noise-free images.

2.2 Extending the Zernike moments

2.2.1 Shape representation

The Zernike moments [34] belongs to the region-based method according to the taxonomy of Zhang *et al.* [35], and to the transform-based family according to the taxonomy of da Fontoura *et al.* [8]. It is considered as one of the best and most powerful shape descriptors [35]; it is described hereafter.

The shape X is represented by a weighted sum of functions (orthogonal polynomials V_{nm} with $n \geq 0$, $|m| \leq n$, $n - m$ even) and described by the set of weights (the Zernike moments; Z_{nm} denoting the m-th moment of order n):

$$X\left(x,y\right) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} Z_{nm} V_{nm}\left(x,y\right). \tag{3}$$

In practice, a good approximation of X can be reconstructed using only the lowest order moments (see Figure 5 for an example of reconstruction with an increasing number of moments). We have $\lim_{L \to \infty} \hat{X}_L = X$, with

$$\hat{X}_L\left(x,y\right) = \sum_{n=0}^{L} \sum_{m=-n}^{n} Z_{nm} V_{nm}\left(x,y\right). \tag{4}$$

Let us denote by (ρ, θ) the polar coordinates of the pixel (x, y). The Zernike polynomials are defined as

$$V_{nm}\left(\rho,\theta\right) = e^{im\theta} \sum_{k=0}^{\frac{n-m}{2}} \frac{(-1)^k (n-k)!}{k! \left(\frac{n+m}{2} - k\right)! \left(\frac{n-m}{2} - k\right)!} \rho^{n-2k}, \tag{5}$$

and the Zernike moments can be computed using [28]:

$$Z_{nm} = \frac{n+1}{\pi} \sum_{x=-1}^{1} \sum_{y=-1}^{1} X\left(x,y\right) V_{nm}^*\left(x,y\right). \tag{6}$$

As the polynomials are only valid inside the unit circle, the shape must be resized and translated to fit inside of it. Pixels outside the unit circle are discarded when computing the Zernike moments. The unit circle plays the same role as the frame considered around the shape in the CBR approach. However, to the contrary of the CBR, the Zernike moments are sensitive to the absolute size of the disk, and a larger set of moments (*e.g.* a higher L) has to be considered to reconstruct the shape when the bounding circle is enlarged. It follows that the description of the inverse shape is not as well defined as for the CBR, even if the difficulties can be circumvented in practice. In our implementation, the shape is normalized in such a way that its center of mass is located at the origin and that the farthest pixels with respect to the origin are located on the unit circle. Thanks to this normalization, the infinite extension of the inverse shape is not a problem for the Zernike moments and it is straightforward to extend these moments to the inverse.

$L = 1 \qquad L = 5 \qquad L = 10 \qquad L = 20 \qquad L = 30 \qquad L = \infty$
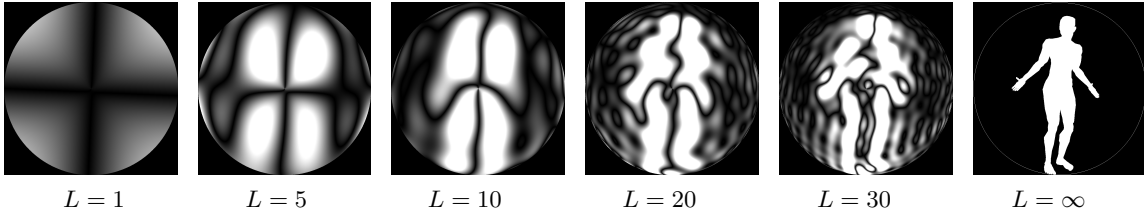
Figure 5: Reconstruction of a shape using an increasing number of Zernike moments. For $L = \infty$, the reconstruction is perfect and we obtain the original shape.

2.2.2 Classification technique

In our experiments with Zernike moments, shapes are classified with a *Support Vector Machine* with soft margin (C-SVM) classifier [7]. We chose a radial basis function kernel (γ being the inverse of the number of attributes) as described in [11]. The value of the parameter C, which controls the soft margin cost function, is optimized in the range $\log_2(C) \in [-5, 20]$ by cross-validation, for each experiment. We compute the moments up to the order $n = 5$, which results in a collection of 24 descriptors. The descriptors are stored by increasing sizes of n and $|m|$.

2.2.3 Properties

The Zernike moments are invariant to rotation [15] and are said to be more robust to noise than any other image moments. This robustness to noise has been underlined in several papers, but not specifically for the type of noise inducing holes in the shape. Nevertheless, our experiments (see Section 3) show that the performance of shape classification remains nearly the same when holes are filled in a preprocessing, indicating that the sensitivity of Zernike moments to holes is weak.

3. Evaluating the performance boost of the inverse shape description

In this section, we analyze the role of the shape and its inverse, and show that they are not equivalent in terms of their representative power, despite that they apparently provide the same amount of semantic information. Intuitively, for the CBR representation, we explain this difference as follows. When the shapes to classify and the elements used to cover the region (*i.e.* rectangles in our experiments) are connected and convex, each rectangle of the inverse descriptor ICBR (X) encodes a portion of the contour. This is similar to the observation of Loncaric [18], who pointed out that the representation by simple components is in fact related to the boundary of the described object. Therefore, for some objects, describing the inverse with a set of convex elements is equivalent to describing the contour by pieces, while the shape representation is closer to describing the inner surface of the object. But the existence of a link between the contour and interior of an object does not mean that contour-based or region-based representations of an object are equally effective for classification. As shown in this section, their roles are not equivalent when it comes to use them to classify objects.

In the following, we want to compare the shape description and the inverse shape description for the purpose of classification tasks. We develop three classification scenarios related to character recognition (scenarios 1 and 2) and silhouette classification (scenario 3). For each of them, both the learning set and the test set contain several thousands of samples, which all differ. However, we assume that the samples of the learning and test sets result from an identical probabilistic context. Because collecting a large amount of real and dis-

similar objects is a time-consuming task, it is common practice to fill learning sets with synthetic noiseless shapes (see the work of Piérard *et al.* [24] for example).

Our evaluation methodology is the following. In the particular case of a two classes problem, it is common to draw the *Receiver Operating Characteristic* (ROC) and to interpret it. The ROC space represents the *True Positive Rates* ($TPR(\alpha)$) and *True Negative Rates* ($TNR(\alpha)$) pairs that can be obtained by a given classifier for a varying threshold α. In this paper, we evaluate the performance of classifiers by using the *maximum achievable balanced accuracy* metric, denoted in short by *accuracy* in the following, which is defined as the maximum average of $TPR(\alpha)$ and $TNR(\alpha)$, when α spans all the possible values:

$$\text{accuracy} = \max_{\alpha} \left(\frac{TPR(\alpha) + TNR(\alpha)}{2} \right). \tag{7}$$

3.1 First scenario: simple objects

The purpose of this first experiment is to understand the behavior of our descriptors for the classification task on a simple example. Therefore, our first scenario considers the shape and its inverse of simple objects (letters), but with a significant amount of intra-class variability.

Shape plays a central role in optical character recognition and many techniques have been proposed [10]. The aim in choosing our scenarios is not to elaborate a new technique for optical character recognition or to compete with well-established classification techniques, but rather to explain how describing the inverse shape can boost the performance.

For convenience, computer-rendered letters were used. The intra-class variability is obtained with 194 different fonts and 8 variations (lowercase, uppercase, normal, italic, bold, italic, and bold), and by rotating the glyphs randomly taken between $-30°$ and $30°$. This leads to $194 \times 8 = 1552$ samples in the learning set, and 1552 different samples (because of the random orientation chosen when they are generated) in the test set. Some samples of the dataset are displayed in Figure 6.

For simplicity, we only consider two classes classifiers. Furthermore, to avoid a prohibitive number of tests, we arbitrarily and intentionally restrict the tests to the comparison of vowels. Table 1 presents the classification results of Barnich's method and of the Zernike moments method, expressed in terms of accuracy, for classifiers that use either the shape or its inverse, but not both. These results show that, Barnich's method makes a better use of the inverse shape for all binary classification tasks to the exception of the (a/u) pair; $ICBR(X)$ is preferable to $CBR(X)$ for classification. However, we obtain opposite results for the Zernike moments method. The shape itself is preferable to its inverse for all but three pairs.

### 3.2 Second scenario: encircled letters

The first experiment indicates that classifiers based on the shape or its inverse have different performances, but it is impossible to know if the discrepancies result from the representation or the classifier itself. Therefore, we build another dataset. Instead of letters on a clean background, we invert the color of the letters of the previous experiment and put them inside an encircling black disk (see Figure 7 for some samples). Both the learning and the test sets contain 1552 samples, and the samples in the learning set differ from those in the test set.

Table 2 provides the classification results in terms of the accuracy. To the contrary of the first experiment, for all the cases except one, the CBR based classifier performs better on the shape than on its inverse and the Zernike moments classifier performs better on the inverse than on the shape itself. This is the inverse conclusion to that of our first experiment. This is not surprising as, by encircling the letters, we have, in broad terms, inverted the roles played by the shape and its inverse.

Table 1: Classification accuracy for different pairwise classification tasks of the first dataset. It appears that, for the CBR, it is more effective to encode the inverse shape (IS) than the shape itself (S), while for the Zernike moments, it is the opposite.

	Cover by rectangles			Zernike moments		
	S		IS	S		IS
a / e	96.97 %	<	**99.13** %	98.86 %	<	**98.98** %
a / i	98.03 %	<	**99.36** %	**98.19** %	>	97.64 %
a / o	92.75 %	<	**95.78** %	**97.81** %	>	96.91 %
a / u	**95.39** %	>	94.04 %	**97.94** %	>	96.69 %
a / y	97.52 %	<	**99.23** %	**98.47** %	>	96.55 %
e / i	98.20 %	<	**98.94** %	98.36 %	<	**98.42** %
e / o	92.33 %	<	**98.55** %	96.95 %	<	**97.26** %
e / u	95.30 %	<	**98.29** %	**97.50** %	>	96.34 %
e / y	98.29 %	<	**99.26** %	**99.01** %	>	98.38 %
i / o	99.48 %	<	**99.77** %	**98.94** %	>	97.86 %
i / u	98.55 %	<	**99.07** %	**98.49** %	>	97.95 %
i / y	97.33 %	<	**98.90** %	**98.25** %	>	97.58 %
o / u	91.59 %	<	**96.23** %	**97.37** %	>	96.68 %
o / y	98.49 %	<	**99.45** %	**99.01** %	>	96.97 %
u / y	96.17 %	<	**97.91** %	**98.04** %	>	97.02 %
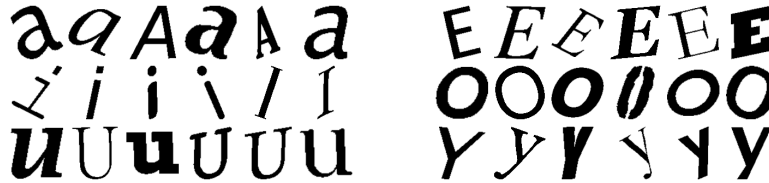
Figure 6: Samples of our first dataset.

From these two experiments, we can conclude that, for each binary classification problem, it is possible to find a dual problem such that describing the shape is preferable for one, and describing the inverse is better for the other one. Ultimately, this brings into question the systematic selection of the shape in a classification problem: in some cases, the dual classification problem, obtained by swapping the roles of the shape and its inverse, might be more appropriate.

3.3 Third experiment: classification of human silhouettes

In the previous sections, we have shown that either the shape or its inverse is preferable for classification, depending on the two classes to be separated. In a third experiment, we simultaneously consider the shape and its inverse to determine if it improves the classification performance.

Here, we focus on a practical application that consists in the recognition of binary human silhouettes, which is central to computer vision tasks involving humans [13, 17, 19]. The extraction of silhouettes of moving objects and persons in the scene is usually performed by a background subtraction algorithm. Silhouettes are then split into connected components, and each component is then classified as human or non-human. Both the CBR and Barnich's classifier were in fact initially designed for that purpose. However, to date, the CBR has exclusively been applied to the shape itself. Our results show that taking into account the inverse improves the results.

Table 2: Classification accuracy for different pairwise classification tasks of the second dataset. For this scenario, it appears that, for the CBR, it is more effective to encode the shape itself (S), and that, for the Zernike moments, the inverse (IS). The results are opposite to what we observed in Table 1 (note that accuracies given in Table 1 differ because of the presence of the encircling disks).

|	Cover by rectangles			Zernike moments		
	S		IS	S		IS
ⓐ / ⓔ	**92.33 %**	>	90.30 %	**97.90 %**	>	97.59 %
ⓐ / ⓘ	**97.00 %**	>	96.49 %	97.12 %	<	**97.44 %**
ⓐ / ⓞ	**93.33 %**	>	91.11 %	96.50 %	<	**96.96 %**
ⓐ / ⓤ	**92.24 %**	>	91.43 %	96.16 %	<	**96.60 %**
ⓐ / ⓨ	**98.16 %**	>	94.43 %	97.06 %	<	**97.46 %**
ⓔ / ⓘ	**97.71 %**	>	96.36 %	97.04 %	<	**97.23 %**
ⓔ / ⓞ	**90.59 %**	>	85.79 %	95.75 %	<	**97.69 %**
ⓔ / ⓤ	94.72 %	<	**96.10 %**	95.60 %	<	**96.79 %**
ⓔ / ⓨ	**97.49 %**	>	93.56 %	96.84 %	<	**97.40 %**
ⓘ / ⓞ	**98.68 %**	>	97.42 %	97.07 %	<	**97.54 %**
ⓘ / ⓤ	**98.13 %**	>	98.07 %	97.29 %	<	**97.69 %**
ⓘ / ⓨ	**95.72 %**	>	95.39 %	97.34 %	<	**97.92 %**
ⓞ / ⓤ	**93.85 %**	>	91.46 %	95.81 %	<	**97.14 %**
ⓞ / ⓨ	**98.84 %**	>	97.42 %	97.13 %	<	**97.28 %**
ⓤ / ⓨ	**96.59 %**	>	94.68 %	96.79 %	<	**97.23 %**
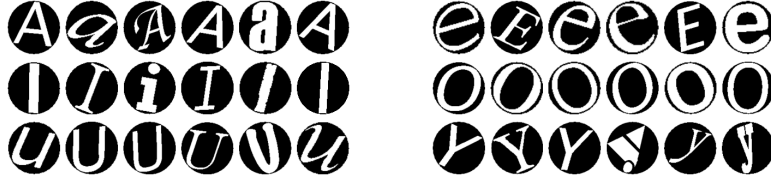
Figure 7: Samples of the second dataset.

The recognition of human silhouettes is a challenging problem due to the wide variety of poses taken by humans, and the difficulty (or perhaps the impossibility) to model non-human silhouettes in a comprehensive way. Also, the frontier between human and non-human silhouettes is sometimes hard to determine, which in classification terms means that the two classes overlap. This kind of difficulty did not show up in our previous experiments.

To assess the classification performance, we need samples of both human and non-human silhouettes. Our database is filled with synthetic silhouettes generated by the method described in [21] (see Figure 8 for some samples). As the results of person detection systems are known to be sensitive to the set of human poses [21, 23], two different sets of human silhouettes are considered: one set with an important pose variability and another one with silhouettes closer to the one of a walker; hereafter, these sets are named "weakly" and "strongly" constrained poses, respectively. In each of our experiments, the learning set and test set contain both 5000 human silhouettes and 5000 non-human silhouettes. The selection of silhouettes in the learning set and in the test set is randomized; in addition, we ensure that the silhouettes in the test set are different from those contained in the learning set.

In our experiments, we compare the results obtained with Barnich's classifier and the Zernike moments classifier for a variable relative importance of the descriptors derived from the shape and its inverse. For Barnich's classifier, this means that we randomly resam-
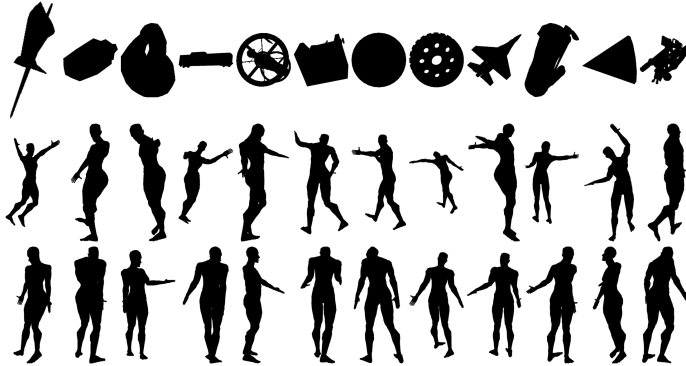
Figure 8: Classification of non-human silhouettes (top row), weakly constrained human silhouettes (middle row), and strongly constrained poses (lower row).

ple the rectangles sets of $\mathrm{CBR}\,(X)$ and $\mathrm{ICBR}\,(X)$ between the first and second steps (see Section 2.1.2) to obtain the adequate proportion. For the Zernike moments, we select the moments from the shape and its inverse by starting from the lowest order. We also compare these results with those obtained when all the holes of the silhouettes are filled during a preprocessing step (as discussed in Section 2.1.3).

According to the graphs of Figures 9 and 10, we can observe for both classification methods that, for noise-free silhouettes:

1. a representation only based on the shape itself is preferable to a representation only based on the inverse.

2. the best results are obtained when we mix descriptors from the shape and its inverse. The optimal proportion is about $70\,\%$, regardless of whether the holes are filled or not. So, the inverse shape carries some information.

For Barnich's classifier, the preprocessing step of filling holes always decreases the performance. Holes thus contain discriminative information about human silhouettes. This is an observation similar to that of Mori and Malik [20], who also noticed that the internal contours are essential for estimating configurations of self-occluding bodies.

With the shape descriptors based on Zernike moments, we obtain very similar results whether holes are filled or not. The Zernike moments are invariant to inner holes within the shapes. When such holes represent a distinctive part of the shape that we want to recognize, this can be problematic.

4. Conclusions

This paper proposes a novel approach for shape description which uses the inverse of the shape instead of the shape itself, as almost exclusively seen in scientific literature.

First, we showed that it is possible to adapt region-based shape descriptors to the inverse shapes, despite their infinite extension. We presented the adaptation for two different descriptors, one based on a bag of words representation (the Cover By Rectangles), and the other one providing potentially an infinite number of descriptors (the Zernike moments).

Then, we challenged the potential of the inverse shape for the task of classification for three scenarios. Our first observation is that the descriptions of a shape and its inverse are not equivalent for the task of classification, even if they are dual types of information. Depending on the scenario, the inverse shape might perform significantly better than the
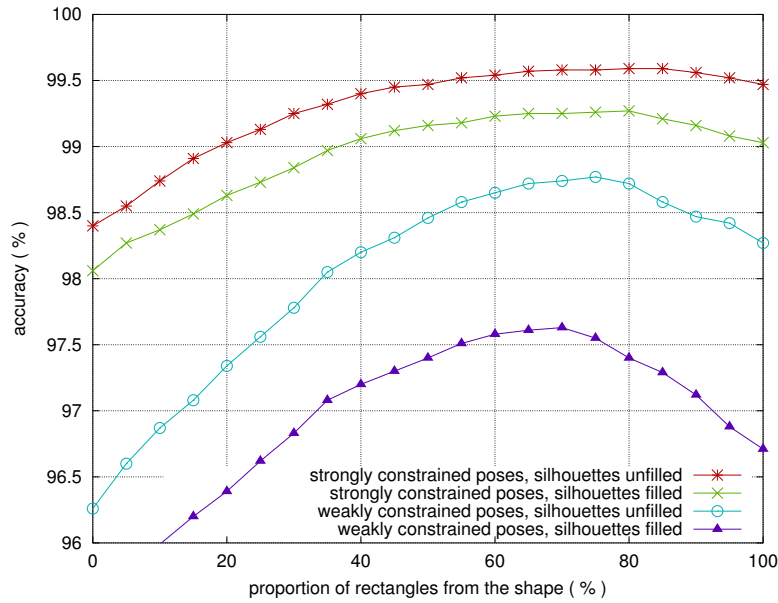
Figure 9: Classification accuracy obtained for human silhouette recognition based on Barnich' method [4], with respect to the proportion of rectangles taken from the shape when mixing rectangles from both the shape and its inverse, for several situations.
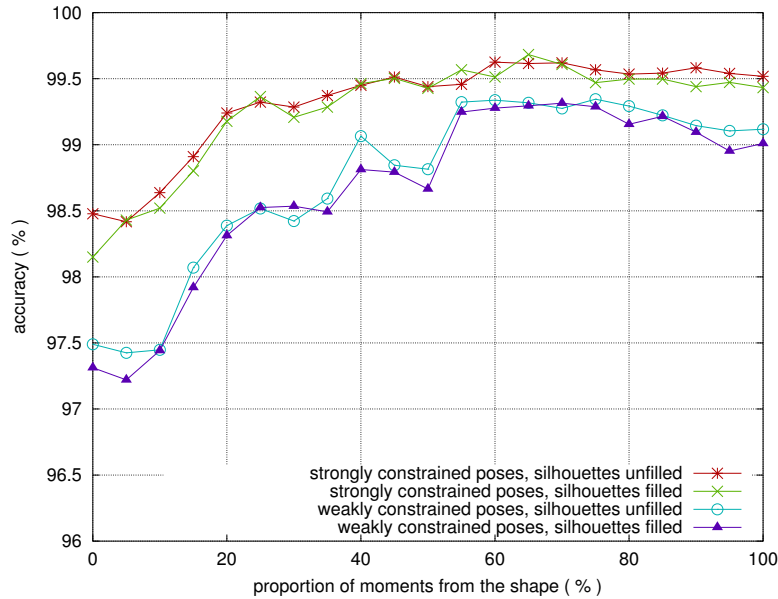
Figure 10: Classification accuracy obtained for human silhouette recognition based on Zernike moments [11], with respect to the proportion of moments from the shape when mixing rectangles from both the shape and its inverse, for several situations.

shape for classification. Whether to choose the shape or its inverse not only depends on the shapes themselves, but also on the selected shape descriptor. Our experiments also show that, for the classification of binary human versus non-human silhouettes, the optimal result is obtained by combining the shape and inverse shape descriptors. While the optimal relative

importance of the shape versus its inverse is problem specific, we conclude that representing the inverse shape is helpful for the task of classification, and also that it is advised to mix the representations of the shape and its inverse for object classification.

Acknowledgments

References

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face recognition with local binary patterns. In *European Conference on Computer Vision (ECCV)*, volume 3021 of *Lecture Notes in Computer Science*, pages 469–481. Springer, 2004.

[2] A. Amanatiadis, V. Kaburlasos, A. Gasteratos, and S. Papadakis. Evaluation of shape descriptors for shape-based image retrieval. *IET Image Processing*, 5(5):493–499, 2011.

[3] O. Barnich. *Motion detection and human recognition in video sequences*. PhD thesis, University of Liège, Belgium, September 2010.

[4] O. Barnich, S. Jodogne, and M. Van Droogenbroeck. Robust analysis of silhouettes by morphological size distributions. In *Advanced Concepts for Intelligent Vision Systems (ACIVS)*, volume 4179 of *Lecture Notes in Computer Science*, pages 734–745. Springer, September 2006.

[5] S. Biasotti, S. Marini, M. Mortara, and G. Patane. An overview on properties and efficacy of topological skeletons in shape modeling. In *Shape Modeling International*, pages 245–254, Seoul, South Korea, May 2003.

[6] H. Blum. A transformation for extracting new descriptors of shape. In W. Wathen-Dunn, editor, *Proceedings of the Symposium on Models for the Perception of Speech and Visual Form*, pages 362–380. MIT Press, 1967.

[7] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, September 1995.

[8] L. da Fontoura and R. Marcondes. *Shape classification and analysis: theory and practice*. CRC Press, 2009.

[9] P. Dollár, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: An evaluation of the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(4):743–761, April 2012.

[10] O. Due Trier, A. Jain, and T. Taxt. Feature extraction methods for character recognition – A survey. *Pattern Recognition*, 29(4):641–662, 1996.

[11] L. Fedorovici and F. Dragan. A comparison between a neural network and a SVM and Zernike moments based blob recognition modules. In *IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI)*, pages 253–258, May 2011.

[12] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim. Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research*, 15:3133–3181, 2014.

[13] D. Geronimo, A. Lopez, A. Sappa, and T. Graf. Survey of pedestrian detection for advanced driver assistance systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(7):1239–1258, July 2010.

[14] P. Geurts, D. Ernst, and L. Wehenkel. Extremely randomized trees. *Machine Learning*, 63(1):3–42, April 2006.

[15] A. Khotanzad and Y. Hong. Invariant image recognition by Zernike moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5):489–497, May 1990.

[16] R. Kresch and D. Malah. Two-sided skeleton: a representation composed of both positive and negative morphological elements. In *Mathematical Morphology and its Applications to Signal Processing*, pages 133–138, Barcelona, Spain, May 1993.

[17] Y. Kuno, T. Watanabe, Y. Shimosakoda, and S. Nakagawa. Automated detection of human for visual surveillance system. In *IEEE International Conference on Pattern Recognition (ICPR)*, volume 3, pages 865–869, Vienna, August 1996.

[18] S. Loncaric. A survey of shape analysis techniques. *Pattern Recognition*, 31(8):983–1001, August 1998.

[19] T. Moeslund, A. Hilton, and V. Krüger. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 104:90–126, November 2006.

[20] G. Mori and J. Malik. Recovering 3D human body configurations using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(7):1052–1062, 2006.

[21] S. Piérard, A. Lejeune, and M. Van Droogenbroeck. 3D information is valuable for the detection of humans in video streams. In *Proceedings of 3D Stereo MEDIA*, pages 1–4, Liège, Belgium, December 2010.

[22] S. Piérard, A. Lejeune, and M. Van Droogenbroeck. A probabilistic pixel-based approach to detect humans in video streams. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 921–924, Prague, Czech Republic, May 2011.

[23] S. Piérard, A. Marcos Alvarez, A. Lejeune, and M. Van Droogenbroeck. On-the-fly domain adaptation of binary classifiers. In *Belgian-Dutch Conference on Machine Learning (BENE-LEARN)*, pages 20–28, Brussels, Belgium, June 2014.

[24] S. Piérard and M. Van Droogenbroeck. A technique for building databases of annotated and realistic human silhouettes based on an avatar. In *Workshop on Circuits, Systems and Signal Processing (ProRISC)*, pages 243–246, Veldhoven, The Netherlands, November 2009.

[25] I. Pitas and A. Venetsanopoulos. Morphological shape decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):38–45, January 1990.

[26] H. Ramoser, T. Schlögl, C. Beleznai, M. Winter, and H. Bischof. Shape-based detection of humans for video surveillance applications. In *IEEE International Conference on Image Processing (ICIP)*, volume 3, pages 1013–1016, Barcelona, Spain, September 2003.

[27] P. Sabzmeydani and G. Mori. Detecting pedestrians by learning shapelet features. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, Minneapolis, Minnesota, USA, June 2007.

[28] C. Singh and E. Walia. Algorithms for fast computation of Zernike moments and their numerical stability. *Image and Vision Computing*, 29(4):251–259, March 2011.

[29] O. Terrades, S. Tabbone, and E. Valveny. A review of shape descriptors for document analysis. In *International Conference on Document Analysis and Recognition (ICDAR)*, volume 1, pages 227–231, September 2007.

[30] M. Van Droogenbroeck and S. Piérard. Object descriptors based on a list of rectangles: method and algorithm. In *International Symposium on Mathematical Morhology (ISMM)*, volume 6671 of *Lecture Notes in Computer Science*, pages 155–165, Ispra, Italy, 2011. Springer.

[31] O. van Kaick, H. Zhang, G. Hamarneh, and D. Cohen-Or. A survey on shape correspondence. *Computer Graphics Forum*, 30(6):1681–1707, September 2011.

[32] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 511–518, Kauai, USA, December 2001.

[33] M. Yang, K. Kpalma, and J. Rosin. A survey of shape feature extraction techniques. In *Pattern Recognition Techniques*, pages 43–90. InTech, 2008.

[34] F. Zernike. Beugungstheorie des schneidenver-fahrens und seiner verbesserten form, der phasenkontrastmethode. *Physica*, 1(7):689–704, 1934.

[35] D. Zhang and G. Lu. Review of shape representation and description techniques. *Pattern Recognition*, 37(1):1–19, January 2004.