Pattern Recognition 72 (2017) 236–253

Contents lists available at ScienceDirect

Pattern Recognition

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

Efficient planar affine canonicalization

Alberto Ruiz[a,*], Pedro E. Lopez-de-Teruel[b], Lorenzo Fernandez-Maimo[b]

[a] *Dept. Informática y Sistemas (DIS), University of Murcia, Spain*
[b] *Dept. Ingeniería y Tecnología de Computadores (DITEC), University of Murcia, Spain*

ARTICLE INFO

Article history:
Received 24 December 2016
Revised 26 May 2017
Accepted 17 July 2017
Available online 20 July 2017

Keywords:
Shape recognition
Affine canonicalization
ICA
Kurtosis

ABSTRACT

This paper presents a fast and accurate affine canonicalization method for planar shapes. This method improves on previous ones based on iterative optimization that produce multiple canonical versions. Canonicalization provides a common reference frame for shape comparison without the loss of discrimination ability often caused by invariant features. It also gives for free the alignment transformation between any pair of shapes. The proposed method is based on the properties of the joint angular distribution of marginal skewness and kurtosis, the so-called *SK signature*, which can be efficiently computed in closed form from the raw image moments. The experiments demonstrate that the method is robust to the non-affine distortions caused by natural perspective image conditions. Thus, it can be used as an automatic preprocessing step to add affine invariance in statistical pattern recognition applications.

1. Introduction

Visual recognition of planar shapes is a fundamental problem in pattern recognition and computer vision, with applications in many diverse fields including autonomous robot navigation, surveillance, document understanding, localization, and augmented reality. The proliferation of low-cost mobile devices equipped with high-quality cameras (e.g., smartphones and drones) increasingly demands simpler and more accurate shape recognition methods.

A common approach to solving this problem is based on standard classifiers using a suitable set of invariant features [1–6]. These methods are fast and do not require costly learning stages. However, simple techniques to achieve invariance may introduce additional perceptual aliasing, reducing discrimination ability. Aggregation methods based on point distributions [7] or shape contexts [8] have similar drawbacks.

Shape discrimination can be improved by alignment, obtaining an explicit model-target transformation [9–13]. This allows comparing registered shapes directly in the measurement domain by means of a simple Euclidean metric or the more powerful Hausdorff distance [14,15]. The inferred transformations are also useful to discard inconsistent matching hypotheses and provide pose estimates for self-localization and navigation applications [16]. Several ideas have been proposed for homography estimation from planar contours [17–21], including recursive probabilistic filters [22,23], statistical theory of shape [7,14,24], and differential geometry [25]. Another interesting approach to alignment is based on estimating a non-parametric probability model for the transformations of a set of training instances with respect to a "congealed" version determined by minimization of pixelwise entropy [26,27]. In general, alignment techniques are computationally expensive for multiclass shape recognition, especially when the parameters of the transformation cannot be expressed in closed form (due, for example, to the lack of explicit corresponding landmarks) and iterative approximations are needed for registration of all possible target-model pairs.

The extraordinary computing power of recent graphic processing units (GPU) has produced a considerable interest in machine learning techniques that use massive amounts of training data (natural or synthetic). In particular, deep convolutional neural networks have proved remarkably successful in many challenging vision applications [28], including image alignment [29,30]. In a promising step towards automatic canonicalization, generic spatial transformer neural modules allow the networks to learn how to transform feature maps to minimize the training error [31]. However, this kind of deep models have some disadvantages such as long learning times, ad-hoc selection of the network architecture, heuristic tuning of hyperparameters, and difficult interpretation of the learned models.

Hierarchical probabilistic generative models have also been recently proposed [32]. This approach admits a wide range of transformations, requires very few training samples, and supports transfer learning between categories. Moreover, the underlying perceptual model has some cognitive plausibility. These advantages, though, come at the cost of very long inference times.

* Corresponding author.
E-mail address: aruiz@um.es (A. Ruiz).

http://dx.doi.org/10.1016/j.patcog.2017.07.017

In contrast with the above general approaches, we are interested here in the specific problem of efficient planar shape recognition from natural images captured by ordinary cameras. Many computer vision methods assume weak perspective projection, modeled by simple affine transformations. This assumption usually holds in practice when object depth is small compared to the distance to the camera. In any case, it is not a severe limitation as full perspective shape recognition without explicit correspondences can be easily achieved by iterative refinement of a good affine initial solution [16,33]. Therefore, we will explore affine alignment methods that are robust to moderate departures from weak perspective caused by out-of plane rotation. We will focus on efficient one-shot learning (using a small number of training samples for each class, ideally just one) and closed-form algorithms for the whole data processing pipeline.

The rest of the paper is organized as follows. Section 2 reviews the canonicalization approach to registration. Section 3 introduces the so called *SK signature* and describes its applications to shape recognition and alignment. A closed-form, efficient canonicalization algorithm based on this signature is developed in Section 4. The stability and range of application of the proposed method is experimentally evaluated in Section 5. The paper concludes with a summary of contributions and future research directions.

2. Canonicalization

Optimal registration is computationally expensive for classification applications, requiring a separate optimization process for each model. A faster alternative is provided by *canonicalization*, which allows precomputation of a good approximation to all possible alignment transformations and evaluation of shape similarity in a common reference frame. A traditional alignment transformation works with two input images, while canonicalization needs just one.

Invariance to a group of transformations can be achieved without any additional loss of class separability by using canonical representatives. Different shapes correspond to the classes of equivalence induced on the set of planar regions by the transformations in the group. Each shape is represented by a particular canonical element in the class, characterized by certain conventional geometric properties. The alignment transformation T_{ab} between any two elements a and b can be immediately obtained as $T_{ab} = C_a^{-1} C_b$ from the respective canonicalization transformations C_a and C_b. This process is much faster than computing a different transformation from scratch for every model-target pair[1]. Furthermore, shape similarity can be directly evaluated in the canonical frame, making on-line classification very efficient as only one 'warping' transformation of the target shape is required. This approach is still suboptimal because the abstract canonical frame does not have any physically meaningful metrics, and the canonicalization transformations are not optimized to reduce registration residuals. However, as demonstrated in the experimental section, it provides excellent approximations for most practical purposes.

Canonicalization for the planar affine group (6 d.o.f.) is generally thought to be a simple task: we first apply a whitening transformation[2] and then fix one single remaining rotational degree of freedom [34,35]. In other words, we must find a characteristic, or "intrinsic" orientation of the (whitened) shape. In principle this can be easily done by considering geometric properties like big concavities, bitangents, most distant points, or Fourier phases, among many other ideas [36]. Unfortunately, most of these proposals only work well for special sets of shapes. Moreover, they do not always provide a unique orientation even for clearly asymmetric figures, and are unstable under noise or small non-affine distortions.

A popular method to detect a dominant orientation is based on the mode of gradient directions. This method is commonly used by keypoint detectors like SIFT to normalize salient image patches in order to compute invariant feature descriptors [37]. This has proved very successful for highly textured images, but in flat or binary regions typically arising in shape recognition the aggregated gradient is a purely local property of the boundary. This feature disregards the relative location and structure of the internal points and it is sensitive to noise. Rounded shapes do not have strongly dominant gradient orientations, and those with straight edges may have multiple histogram maxima even though their structure is rich enough to induce a single distinguished orientation.

A promising canonicalization approach using global image information is based on Independent Component Analysis (ICA) [38]. In contrast with the maximum variance projections of a data set obtained by the principal components, ICA looks for a linear transformation such that the new variables are as much statistically independent as possible. This new representation may provide a useful reinterpretation of the data set in terms of meaningful components. For example, a common application is blind separation of mixed signals. Computational ICA techniques typically start from whitened data and then iteratively optimize an orthogonal transformation to get new variables as different from Gaussian distributions as possible. The key idea is that any linear combination of (non Gaussian) random variables has more entropy, and therefore is "more Gaussian", than the original variables. Practical optimization costs are marginal kurtosis and relative entropy. Efficient implementations include FastICA [39] and RobustICA [40].

In the context of shape recognition, ICA has been applied in order to compute affine invariant descriptors and alignment homographies [41,42]. These methods eventually work with Fourier or Zernike rotation invariant features, which partially defeat the advantages of canonicalization. Other ICA methods [43,44] work with the contour coordinates as separate 1D signals instead of the whole set of 2D points (the joint distribution) in a general figure, which may include separate fragments and holes.

The above proposals use standard ICA implementations and produce unnecessary multiple orientations. While general multidimensional ICA require expensive iterative local optimization, the 2D case arising from the shape orientation problem is the simplest one, with just one degree of freedom. In offline applications it could even be solved by exhaustive search of all rotation angles. In this paper we present a fast and simple closed-form solution for affine canonicalization, based on the concepts developed in the next section.

3. The skewness-kurtosis signature

We will informally use the term 'shape' to refer to a planar binary region, although most of the following results are equally valid for gray level images.

A region R is defined by its indicator function $I_R(x, y) = 1$ if $(x, y) \in R$ and $I_R(x, y) = 0$ otherwise. The sum of function f over R is

$$E_R\{f\} = \iint_{\mathbb{R}^2} I_R(x, y) f(x, y) dx dy, \tag{1}$$

and the moments of R are

$$m_{pq} = E_R\{x^p y^q\}. \tag{2}$$

Let λ_1^2 and λ_2^2 be the eigenvalues of the covariance matrix of R. Degenerate regions ($\lambda_2/\lambda_1 \ll 1$) cannot be whitened in a numerically

[1] Certain highly symmetric shapes can be taken to the canonical version via different and equally valid alternative transformations; this ambiguity should be managed in an application-dependent fashion.

[2] This denotes an affine transformation that produces uncorrelated variables with zero mean and unit variance. This preprocessing transformation is widely used in data analysis and can be easily computed in closed-form from the first and second-order moments (see Appendix A).