Robustness Aspects of Optimized Centroids

Jan Kalina and Patrik Janáček

Abstract Centroids are often used for object localization tasks, supervised segmentation in medical image analysis, or classification in other specific tasks. This paper starts by contributing to the theory of centroids by evaluating the effect of modified illumination on the weighted correlation coefficient. Further, robustness of various centroid-based tools is investigated in experiments related to mouth localization in non-standardized facial images or classification of high-dimensional data in a matched pairs design. The most robust results are obtained if the sparse centroid-based method for supervised learning is accompanied with an intrinsic variable selection. Robustness, sparsity, and energy-efficient computation turn out not to contradict the requirement on the optimal performance of the centroids.

Keywords: image processing, optimized centroids, robustness, sparsity, low-energy replacements

1 Introduction

Methods based on centroids (templates, prototypes) are simple yet widely used for object localization or supervised segmentation in image analysis tasks and also within other supervised or unsupervised methods of machine learning. This is true e.g. in various biomedical imaging tasks [1], where researchers typically cannot afford a too large number of available images [3]. Biomedical applications also benefit from the interpretability (comprehensibility) of centroids [11].

This paper is focused on the question how are centroid-based methods influenced by data contamination. Section 2 recalls the main approaches to centroid-based object localization in images, as well as a recently proposed method of [6] for op-

Jan Kalina (✉) · Patrik Janáček
The Czech Academy of Sciences, Institute of Computer Science, Pod Vodárenskou věží 2, 182 07 Prague 8, Czech Republic, e-mail: kalina@cs.cas.cz;janacekpatrik@gmail.com

P. Brito et al. (eds.), *Classification and Data Science in the Digital Age*,
Studies in Classification, Data Analysis, and Knowledge Organization,
https://doi.org/10.1007/978-3-031-09034-9_22

timizing centroids and their weights. The performance of these methods to data contamination (non-standard conditions) has not been however sufficiently investigated. Particularly, we are interested in the performance of low-energy replacements of the optimal centroids and in the effect of posterior variable selection (pixel selection). Section 2.1 presents novel expressions for images with a changed illumination. Numerical experiments are presented in Section 3. These are devoted to mouth localization over raw facial images as well as over artificially modified images; other experiments are devoted to high-dimensional data in a matched pairs design. The optimized centroids of [6] and especially their modification proposed here turn out to have remarkable robustness properties. Section 4 brings conclusions.

2 Centroid-based Classification (Object Localization)

Commonly used centroid-based approaches to object localization (template matching) in images construct the centroid simply as the average of the positive examples and typically use Pearson product-moment correlation coefficient r as the most common measure of similarity between a centroid $\mathbf{c}$ and a candidate part of the image (say $\mathbf{x}$). While the centroid and candidate areas are matrices of size (say) $I \times J$ pixels, they are used in computations after being transformed to vectors of length $d := IJ$. This allows us to use the notation $\mathbf{c} = (c_1, \ldots, c_d)^T$ and $\mathbf{x} = (x_1, \ldots, x_d)^T$.

Assumptions $\mathcal{A}$: We assume the whole image to have size $N_R \times N_C$ pixels. We assume the centroid $\mathbf{c} = (c)_{i,j}$ with $i = 1, \ldots, I$ and $j = 1, \ldots, J$ to be a matrix of size $I \times J$ pixels. A candidate area $\mathbf{x}$ and nonnegative weights $\mathbf{w}$ with $\sum_i \sum_j w_{ij} = 1$ are assumed to be matrices of the same size as $\mathbf{c}$.

For a given image, E will denote the set of its rectangular candidate areas of size $I \times J$. The candidate area fulfilling

$$\arg\max_{\mathbf{x} \in \mathsf{E}} r(\mathbf{x}, \mathbf{c}) \tag{1}$$

or (less frequently)

$$\arg\min_{\mathbf{x} \in \mathsf{E}} ||\mathbf{x} - \mathbf{c}||_2 \tag{2}$$

are classified to correspond to the object (e.g. mouth).

Let us consider here replacing r by the weighted correlation coefficient r_w

$$\arg\max_{\mathbf{x} \in \mathsf{E}} r_w(\mathbf{x}, \mathbf{c}; \mathbf{w}) \tag{3}$$

with given non-negative weights $\mathbf{w} = (w_1, \ldots, w_d)^T \in \mathcal{R}^p$ with $\sum_{i=1}^n w_i = 1$, where $\mathcal{R}$ denotes the set of all real numbers. Let us further use the notation $\bar{x}_w = \sum_{j=1}^d w_j x_j = \mathbf{w}^T \mathbf{x}$ and $\bar{c}_w = \mathbf{w}^T \mathbf{c}$. We may recall r_w between $\mathbf{x}$ and $\mathbf{c}$ to be defined as

$$r_W(\mathbf{x}, \mathbf{c}; \mathbf{w}) = \frac{\sum_{i=1}^d w_i (x_i - \bar{x}_w)(c_i - \bar{c}_w)}{\sqrt{\sum_{i=1}^d [w_i (x_i - \bar{x}_w)^2] \sum_{i=1}^d [w_i (c_i - \bar{c}_w)^2]}}. \tag{4}$$

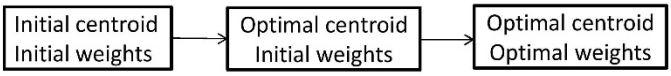

Fig. 1 The workflow of the optimization procedure of [6].

A detailed study of [2] investigated theoretical foundations of centroid-based classification, however for the rare situation when (1) is replaced by

The sophisticated centroid optimization method of [6], outlined in Figure 1, requires to minimize a nonlinear loss function corresponding to a regularized margin-like distance (exploiting r_w) evaluated for the worst pair from the worst image over the training database (i.e. the worst with respect to the loss function). Subsequently, optimization of the weights may be also performed, ensuring many pixels to obtain zero weights (i.e. yielding a sparse solution). The optimal centroid may be used as such, even without any weights at all; still, optimization of the weights leads to a further improvement of the classification performance. In the current paper, we always consider a linear (i.e. approximate) approach to centroid optimization, although a nonlinear optimization is also successful as revealed in the comparisons in [6].

2.1 Centroid-Based Object Localization: Asymmetric Modification of the Candidate Area

In the context of object localization as described above, our aim is to express $r_w(\mathbf{x}^*, \mathbf{c}; \mathbf{w})$ under modified candidate areas (say $\mathbf{x}^*$) of the image $\mathbf{x}$; we stress that the considered modification of the image does not allow to modify the centroid $\mathbf{c}$ and weights $\mathbf{w}$. These considerations are useful for centroid-based object localization, when asymmetric illumination is present in the whole image or its part. The weighted variance $S_w^2(\mathbf{x}; \mathbf{w})$ of $\mathbf{x}$ with weights $\mathbf{w}$ and the weighted covariance $S_w(\mathbf{x}, \mathbf{c})$ between $\mathbf{x}$ and $\mathbf{c}$ are denoted as

$$S_w^2(\mathbf{x}) = \sum_{i,j} w_{ij}(x_{ij} - \bar{x}_w)^2, \quad S_w(\mathbf{x}, \mathbf{c}) = \sum_{i,j} w_{ij}(x_{ij} - \bar{x}_w)(c_{ij} - \bar{c}_w). \tag{5}$$

Further, the notation $\mathbf{x} + a$ with $\mathbf{x} = (x_{ij})_{i,j}$ is used to denote the matrix $(x_{ij} + a)_{i,j}$ for a given $a \in \mathcal{R}$. We also use the following notation. The image $\mathbf{x}$ is divided to two parts $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)^T \in \mathcal{R}^d$, where $\sum_I$ or $\sum_{II}$ denote the sum over the pixels of the first or second part, respectively.

Theorem 1 *Under Assumptions $\mathcal{A}$, the following statements hold.*

1. *For $\mathbf{x}^* = \mathbf{x} + \varepsilon$, it holds $r_w(\mathbf{x}^*, \mathbf{c}) = r_w(\mathbf{x}, \mathbf{c})$ for $\varepsilon > 0$.*
2. *For $\mathbf{x}^* = k\mathbf{x}$ with $k > 0$, it holds $r_w(\mathbf{x}^*, \mathbf{c}) = r_w(\mathbf{x}, \mathbf{c})$.*
3. *For $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)^T$ and $\mathbf{x}^* = (\mathbf{x}_1, \mathbf{x}_2 + \varepsilon)^T$, it holds $r_w(\mathbf{x}^*, \mathbf{c}) =$*

$$= \frac{S_w(\mathbf{x},\mathbf{c}) + \varepsilon \sum_{II} w_{ij} c_{ij} - \varepsilon v_2 \bar{c}_w}{S_w(\mathbf{c}) \sqrt{S_w^2(\mathbf{x}) + v_2(1-v_2)\varepsilon^2 + 2\varepsilon(2v_2-1)(\sum_{II} w_{ij} x_{ij} - v_2 \bar{x}_w)}}, \quad (6)$$

where $v_2 = \sum_{II} w_{ij}$ *and* $\varepsilon \in \mathcal{R}$.

4. *For* $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)^T$ *and* $\mathbf{x}^* = (\mathbf{x}_1, k\mathbf{x}_2)^T$ *with* $k > 0$, *it holds*

$$r_w(\mathbf{x}^*, \mathbf{c}) = r_w(\mathbf{x}, \mathbf{c}) \frac{S_w(\mathbf{x})}{S_w^*(\mathbf{x})} + \frac{(k-1)\sum_{II} w_{ij} x_{ij} (c_{ij} - \bar{c}_w)}{S_w(\mathbf{c}) S_w^*(\mathbf{x})}, \quad (7)$$

where

$$\left(S_w^*(\mathbf{x})\right)^2 = S_w^2(\mathbf{x}) + (k^2-1) \sum_{II} w_{ij} x_{ij}^2 - \frac{k^2-1}{n} \left(\sum_{II} w_{ij} x_{ij} \right)^2 -$$

$$-\frac{2}{n}(k-1) \left(\sum_{I} w_{ij} x_{ij} \right) \left(\sum_{II} w_{ij} x_{ij} \right). \quad (8)$$

The proofs of the formulas are technical but straightforward exploiting known properties of r_w. The theorem reveals r_w to be vulnerable to the modified illumination, i.e. all the methods based on centroids of Section 2 may be too influenced by the data modification.

3 Experiments

3.1 Data

Three datasets are considered in the experiments. In the first dataset, the task is to localize the mouth in the database containing 212 grey-scale 2D facial images of faces of healthy individuals of size 192×256 pixels. The database previously analyzed in [6] was acquired at the Institute of Human Genetics, University of Duisburg-Essen, within research of genetic syndrome diagnostics based on facial images [1] under the projects BO 1955/2-1 and WU 314/2-1 of the German Research Council (DFG). We consider the training dataset to consist of the first 124 images, while the remaining 88 images represent an independent test set acquired later but still under the same standardized conditions fulfilling assumptions of unbiased evaluation. The centroid described below is used with $I = 26$ and $J = 56$.

Using always raw training images, the methods are applied not only to the raw test set, but also to the test set after being artificially modified using models inspired by Section 2.1. On the whole, five different versions of the test database are considered; the modifications required that we first manually localized the mouths in the test images:

1. Raw images.

2. Illumination. If we consider a pixel $[i, j]$ with intensity x_{ij} in an image (say) f, then the grey-scale intensity f_{ij} will be

$$f^*_{ij} = f_{ij} + \lambda |j - j_0|, \quad i = 1, \ldots, I, \quad j = 1, \ldots, J, \tag{9}$$

where $[i_0, j_0]$ are the coordinates of the mouth and $\lambda = 0.002$.
3. A more severe version of the modification (ii) with $\lambda = 0.004$.
4. Asymmetry. In every test image, each true mouth $\mathbf{x}$ of size 26×56 pixels with intensities x_{ij} is replaced by

$$x^*_{ij} = \begin{cases} x_{ij} + 0.2, & i = 1, \ldots, 26, \ j = 1, \ldots, 15, \\ x_{ij}, & i = 1, \ldots, 26, \ j = 16, \ldots, 41, \\ x_{ij} + 0.1, & i = 1, \ldots, 26, \ j = 42, \ldots, 56. \end{cases} \tag{10}$$

5. Rotation. Such candidate area is classified as the mouth in the given image, which maximizes the loss (1) or (3) over the three versions of the image, namely after rotations by +5, 0, and −5 degrees.
6. Image denoising (for raw images). The LWS-filter [5], replacing each grey value by the least weighted squares estimate [7] computed from a circular neighborhood with radius 4 pixels, was applied to each test image.

The optimized centroids were explained in [6] to be applicable also to classification tasks for other data than images, if they follow a matched pairs design. We use two datasets from [6] in the experiments and their classification accuracies are reported in a 10-fold cross validation.

- AMI. The gene expressions of 4000 genes over 92 individuals in two versions (raw or contaminated by outliers). The aim is to learn a classification rule allowing to assign a new individual to one of the two given groups (controls or patients with acute myocardial infarction (AMI)).
- Simulated data. The design mimicks a 1:1 matched case-control study with 2500 variables over 60 individuals in two versions (raw or contaminated by outliers) and the aim is again to classify between two given groups (patients and controls).

Fig. 2 The average centroid used as the initial choice for the centroid optimization.

3.2 Methods

The following methods are compared in the experiments; standard methods are computed using R software and we use our own C++ implementation of centroid-based methods. The average centroid is obtained as the average of all mouths of the training set, or the average across all patients. The centroid optimization starts with the average centroid as the initial one, and the optimization of weights starts with equal weights as the initial ones:

A. Centroid-based method (2).
B. Centroid-based method (1) with average centroid (Figure 2) and equal weights.
C. Centroid-based method (1) with average centroid, replacing r_w by cosine similarity defined for $\mathbf{x} \in \mathcal{R}^d$ and $\mathbf{y} \in \mathcal{R}^d$ as

$$\cos\,\theta = \frac{\mathbf{x}^T\mathbf{y}}{||\mathbf{x}||_2||\mathbf{y}||_2} = \frac{\sum_{i=1}^d x_i y_i}{\left(\sum_{i=1}^d x_i^2\right)^{1/2}(\sum_{j=1}^d y_j^2)^{1/2}}. \tag{11}$$

D. Centroid-based method (1) with optimal centroid and equal weights [6].
E. Centroid-based method (1) with optimal centroid and optimal weights as in [6] (optimizing the centroid and only after that the weights), i.e. with posterior variable selection (pixel selection).
F. Centroid-based method (1) as in [6], where however the weights are optimized first, and then the centroid is optimized.
G. Centroid-based method (1) as in [6], where however each step of centroid optimization is immediately followed by optimization of the weights; this method performs (in contrary to [6]) intrinsic variable selection.
H. Centroid-based method (1) as in [6], where however each optimization step proceeds over 10 worst images (instead of the very worst image).
I. Centroid-based method (1) with average centroid, where r_w is used as r_{LWS} [7] with weight function

$$\psi_1(t) = \exp\left\{-\frac{t^2}{2\tau^2}\right\} 1\left[t < \frac{3}{4}\right], \quad t \in [0,1], \tag{12}$$

corresponding to a (trimmed) density of the Gaussian $\mathsf{N}(0,1)$ distribution; 1 denotes an indicator function. To explain, the computation of $r_{\mathsf{LWS}}(x,y)$ starts by fitting the LWS estimator in the linear regression of y as the response of x, and r_w is used with the weights determined by the LWS estimator.
J. The method (I) with the weight function $\psi_2(t) = 1\left[t < \frac{3}{4}\right]$ for $t \in [0,1]$.
K. The approach of [12] that is meaningful however only for the mouth localization dataset.

Table 1 Classification accuracy for three datasets. For the mouth localization data, modifications of the test images are described in Section 3: (i) None (raw images); (ii) Illumination; (iii) Asymmetry; (iv) Rotation; (v) Image denoising. A detailed description of the methods is given in Section 3.2.

	Dataset									
	Mouth localization						AMI		Simul.	
Method	(i)	(ii)	(iii)	(iv)	(v)	(vi)	Raw	Cont.	Raw	Cont.
A	0.90	0.86	0.81	0.88	0.81	0.93	0.73	0.66	0.71	0.67
B	0.93	0.90	0.86	0.92	0.86	0.95	0.76	0.70	0.77	0.70
C	0.89	0.84	0.74	0.89	0.84	0.93	0.72	0.61	0.70	0.64
D	1.00	0.98	0.95	0.99	0.93	0.98	0.85	0.83	0.80	0.77
E	1.00	1.00	0.98	1.00	0.95	0.98	0.87	0.85	0.83	0.80
F	1.00	0.98	0.96	1.00	0.89	0.97	0.86	0.82	0.79	0.73
G	1.00	0.96	0.95	1.00	0.93	0.99	0.88	0.85	0.86	0.82
H	1.00	1.00	0.98	1.00	0.92	0.96	0.86	0.83	0.84	0.79
I	0.96	0.96	0.93	0.99	0.94	0.96	0.77	0.72	0.75	0.71
J	0.94	0.93	0.89	0.95	0.89	0.93	0.74	0.69	0.72	0.66
K	1.00	1.00	0.97	0.95	0.97	0.96	Not meaningful			

3.3 Results

The results as ratios of correctly classified cases are presented in Table 1. For the mouth localization, the optimized centroids of methods D, F, and H turn out to outperform simple centroids (A, B, and C); the novel modifications E and G performing intrinsic variable selection yield the best results. Simple standard centroids (A, B, and C) are non-robust to data contamination; this follows from Section 2.1 and from analogous considerations for other types of contaminating the images. On the other hand, the robustness of optimized centroids is achieved by their optimization (but not by using r_w as such). Methods E and G are even able to overcome methods I and J based on r_{LWS}. We recall that r_{LWS} is globally robust in terms of the breakdown point [4]), is computationally very demanding, and does not seem to allow any feasible optimization. Other results reported previously in [6] revealed that also numerous standard machine learning methods are too vulnerable (non-robust) with respect to data contamination, if measuring the similarity by r or r_w.

For the AMI dataset, methods E and G with variable selection perform the best results for raw as well as contaminated datasets. For the simulated data, the method G yields the best results and the method E stays only slightly behind as the second best method.

4 Conclusions

Understanding the robustness of centroids represents a crucial question in image processing with applications for convolutional neural networks (CNNs), because centroids are very versatile tools that may be based on deep features learned by deep

learning. We focus on small datasets, for which CNNs cannot be used [10]. This paper is interested in performance of centroid-based object localization over small databases with non-standardized images, which commonly appear e.g. in medical image analysis.

The requirements on robustness with respect to modifications of the images turn out not to contradict the requirements on optimality of the centroids. The method G applying an intrinsic variable selection on the optimal centroid and weights [6] can be interpreted within a broader framework of robust dimensionality reduction (see [8] for an overview) or low-energy approximate computation. Additional results not presented here reveal the method based on optimized centroids to be robust also to small shift. Neither the theoretical part of this paper nor the experiments exploit any specific properties of faces. The presented robust method has potential also for various other applications, e.g. for deep fake detection by centroids, robust template matching by CNNs [9], or applying filters in convolutional layers of CNNs.

Acknowledgements The research was supported by the grant 22-02067S of the Czech Science Foundation.

References

1. Böhringer, S., de Jong, M. A.: Quantification of facial traits. Frontiers in Genetics **10**, 397 (2019)
2. Delaigle, A., Hall, P.: Achieving near perfect classification for functional data. Journal of the Royal Statistical Society **74**, 267–286 (2012)
3. Gao, B., Spratling, M. W.: Robust template matching via hierarchical convolutional features from a shape biased CNN. ArXiv:2007.15817 (2021)
4. Jurečková, J., Picek, J., Schindler, M.: Robust statistical methods with R. 2nd edn. CRC Press, Boca Raton (2019)
5. Kalina, J.: A robust pre-processing of BeadChip microarray images. Biocybernetics and Biomedical Engineering **38**, 556–563 (2018)
6. Kalina, J., Matonoha, C.: A sparse pair-preserving centroid-based supervised learning method for high-dimensional biomedical data or images. Biocybernetics and Biomedical Engineering **40**, 774–786 (2020)
7. Kalina, J., Schlenker, A.: A robust supervised variable selection for noisy high-dimensional data. BioMed Research International **2015**, 320385 (2015)
8. Rousseeuw, P. J., Hubert, M.: Anomaly detection by robust statistics. WIREs Data Mining and Knowledge Discovery **8**, e1236 (2018)
9. Sun, L., Sun, H., Wang, J., Wu, S., Zhao, Y., Xu, Y.: Breast mass detection in mammography based on image template matching and CNN. Sensors **2021**, 2855 (2021)
10. Sze, V., Chen, Y. H., Yang, T. J., Emer, J. S.: Efficient processing of deep neural networks. Morgan & Claypool Publishers, San Rafael (2020)
11. Watanuki, S.: Watershed brain regions for characterizing brand equity-related mental processes. Brain Sciences **11**, 1619 (2021)
12. Zhu, X., Ramanan, D.: Face detection, pose estimation, and landmark localization in the wild. IEEE Conference on Computer Vision and Pattern Recognition 2012. IEEE, New York, pp. 2879–2886 (2012)