21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

Multiple Local Kernel Integrated Feature Selection for Image Classification

Yu Sun, Bir Bhanu
Center for Research in Intelligent Systems, University of California, Riverside, CA 92521.
ysun@vislab.ucr.edu, bhanu@cris.ucr.edu

Abstract

Feature redundancy and loss of local feature are central problems for image classification. Feature selection decreases the feature redundancy by choosing a subset of features and eliminating those with low prediction. The local feature representation is able to highlight objects in an image, thus, overcoming the drawbacks of global features. This paper presents a new method, called the local kernel for feature selection, which integrates a local kernel of the segmentation regions into feature selection to provide improved image classification, by means of the region-based image distance integrated into the kernel of the Bayesian classifier. The proposed method is tested on two standard image databases and the classification results are higher than the current feature selection and classification methods.

1. Introduction

Image classification [1-3] is used for applications such as object recognition [1], medical imaging [2] and image retrieval [3]. In image classification, an image is represented by a set of visual features, e.g. color and texture. The classification accuracy largely depends on the features used. Two challenging problems commonly exist for image classification approaches.

Problem (1): There exists a significant redundancy among a large number of low-level features which prohibits the achievement of ideal classification.

Problem (2): The global features are inadequate to represent the important and discriminative local object information present in an image.

The objective of feature selection is to address the *problem (1)* by identifying the feature subset that is most predictive, and minimizing feature redundancy and the classification error. Feature selection has been explored in computer vision and pattern recognition tasks [4-8]. A survey of state-of-the-art feature selection methods is provided in [9].

The essence of *problem (2)* in image classification is characterized by the loss of local image contents. Objects in an image cannot be resented by using only the global feature by the entire image. Solutions of *problem (2)* make use of the local image contents [10], or computing the region-based image distance [11, 12]. The local features help to improve the classification by means of highlighting important local objects.

In this paper, the local feature selection is realized by a novel approach, that integrates a local kernel into the Bayesian classifier. Also a multiple kernel learning is used to select an optimal classifier.

The outline of the paper is as follows. Section 2 gives the related work and the contributions of this paper. Section 3 describes the technical approach. Section 4 provides experimental results. Finally, conclusions are given in Section 5.

2. Related Work and Contributions

2.1 Related Work

The aim of feature selection is to solve the *problem (1)* by decreasing the information redundancy in the raw feature and improving the feature discrimination capability. The feature selection is usually described as a search problem in the feature space as follows [9]:

(1) *Feature space search*, a strategy with which the feature space is explored. Widely used search strategies are the sequential forward **s**earch **(**SFS**)** and the sequential forward floating search (SFFS) [6].

(2) *Evaluation measure* is a function by which the candidate features are evaluated. It uses Bayesian classifier [6] or SVM [7].

Local feature also improves the classification. They are divided into two categories: point-based features, e.g. the SIFT [13], and the region-based features [10-12], e.g., the earth mover distance (EMD) [11] and the integrated region matching (IRM) [12].

The combination of feature selection and local features, called the local feature selection, has not been thoroughly studied [20-22]. In this paper, local feature selection is realized by a novel way, in which a local kernel is integrated into the Beyesian classifier.

2.2 Contributions of this Paper

The paper makes the following contributions:

1) It provides a new way to combine the feature selection and local features, by integrating local kernel into the distance metric of a Bayesian classifier. Both *problem (1)* and *(2)* (Section 1) are solved.

978-4-9906441-1-6 ©2012 IAPR

2) A multiple kernel learning scheme is used to select the optimal inducer from multiple Bayesian classifiers.

3. Technical Approach

In the system diagram in Fig. 1, the region features are extracted, and they are further used to compute the region-based image distance in the Bayesian classifier. A local kernel learning selects the optimal classifier.

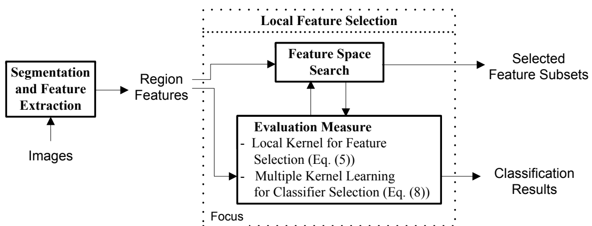

Fig 1. The overall system diagram for this paper.

3.1. Image Segmentation, Feature Extraction

3.1.1. Image Segmentation: As the first step, the images are segmented into regions. This paper uses 3 different segmentation approaches for the proposed method: region growing [14], watershed segmentation [18], and the normalized cut [19]. The three methods provides local regions for the proposed method, and their performances are compared in Table 1.

3.1.2. Region Feature Extraction: 219 feature dimensions are extracted from segmented regions.
(1) RGB and HSV components, with region's mean and std. for each component: 12 dimensions. **(**2) The 8-dimension texture feature from the mean and std. of the filtered image/region by Gabor filters. **(**3) The 7-dimension shape feature derived from the first 7 central geometric moments of the image/region. **(**4) The 192 dimensional color histogram for RGB components.

3.2. Local Feature Selection

The realization of the feature selection is composed of the two steps below.

3.2.1. The Feature Space Search: SFFS [6] starts with an empty feature subset ($X' = \phi$). In each iteration, one feature is chosen among the remaining m-dimensional feature space $X = \{x_i\}, i = 1,...,m$ and is added into the subset. To determine which feature to add, it tests the performance of every addition of feature from the remaining features, by an evaluation measure $eva(\cdot)$, and select the one with the highest performance for the new subset. The above process, called the forward search, is shown as follows [9].

$$X' = X' \cup \{x_i \in X \mid \max(eva(X' \cup x_i)), i = 1,...,m\} \\ X = X - x_i \quad (1)$$

Also in each iteration, a backward search deletes a feature, after which the remaining set reaches the highest improvement of performance evaluation, compared to the subset before deletion,

$$X' = X' - \{x_i \in X \mid \max(eva(X' - x_i) - eva(X')), i = 1,..m\} \\ X = X + x_i \quad (2)$$

Equations (1) and (2) are run sequentially, until no additional features results in accuracy improvement.

3.2.2. Local Kernel in Bayesian Classifier: A local kernel is used in Bayesian classifier as the evaluation measure, realized by the integrated region matching (IRM) [12], a region-based similarity measure between two images. Images I_1 and I_2 have two region sets $\mathrm{I}_1 = \{r_1, r_2, ...,r_m\}$ and $\mathrm{I}_2 = \{r_1', r_2', ...,r_n'\}$. The IRM distance between two images is the summation of all the region distance,

$$dist_{IRM}(I_1, I_2) = \sum_{i,j} s_{i,j} d_{i,j} \quad (3)$$

where $d_{i,j}$ is the Euclidean distance between regions r_i and r_j' of two images, and $s_{i,j}$ is the distance weights. The larger weights indicate the importance of the two regions in similarity measure. The weights are proportional to the region importance values p_i and p_j', for which a larger region has a higher region importance value. The equations are,

$$\sum_{j=1}^{n} s_{i,j} = p_i, i = 1,....,m \\ \sum_{i=1}^{m} s_{i,j} = p_j', j = 1,....,n \quad (4)$$

The sum of the distance weights is equal to its importance value, and the distance of a region can be measured to multiple regions of the other image. So important regions contribute largely in computing the distance. Examples of important regions with their values (p) are shown in Fig. 2. The region-based distance (Eq. (3)) is integrated into the Bayesian classifier as a local kernel, by the following equation,

$$f(x) = \frac{1}{(2\pi)^{k/2} |\Sigma|^{1/2}} \cdot \exp\left(-\frac{1}{2} \cdot V\_dist_{IRM}{}^T(x, \mu) \cdot \Sigma^{-1} \cdot V\_dist_{IRM}(x, \mu)\right) \quad (5)$$

where the IRM distance is used instead of the global image distance. The term $V\_dis_{IRM}(x, \mu)$ is an IRM distance vector for k dimensional feature space,

$$V\_dist_{IRM}(x, \mu) = \{dist_{IRM}(x_1, \mu_1), ...,dist_{IRM}(x_k, \mu_k)\} \quad (6)$$

where x_i, μ_i , i=1,…..,k, are components of k-dimensional region vectors of testing image x and the cluster μ of the trained Gaussian model, respectively. The IRM distance is computed for each feature dimension of the data, and for a specific dimension i, the distance $dist_{IRM}(x_i, \mu_i)$ is computed by Equation (3). The testing image x is composed of segmented region feature vectors to compute IRM distance.

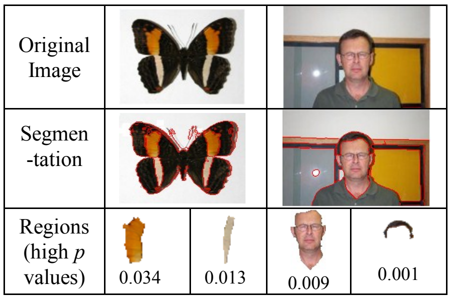

Fig. 2. Images segmentation, and regions with high p values (butterfly and Caltech-101 database).

3.3. Multiple Kernel Learning for Classifier Selection

A multiple kernel learning is performed to learn the parameters of multiple Bayesian classifiers, each of which is individually trained in Section 3.2. For a testing image x, we employed a so-called kernel function $y(x)$ (see Equation (8)), which intuitively selects an optimal classifier for a specific testing data.

$$y(x) = \underset{c=1,\dots,C}{argmax}\ [\beta_c \cdot (f_c(x) \cdot a_c + b_c)] \quad (7)$$

where C is the number of candidate Bayesian classifiers, and β_c, a_c and b_c are parameters with specific values learned for each classifier. The parameters satisfy the constraints $\sum_{c=1}^{C} \beta_c = 1$, $\sum_{c=1}^{C} a_c = 1$, $\sum_{c=1}^{C} b_c = 1$, $f_c(x)$ is the individual probability computed by the Bayesian classifier as shown in Equation (5). The weighted linear function of the posterior probability enables to put different weights on classifiers with different performance during the training session. The kernel function aims to select an optimal weighted linear posterior probability from one of the multiple classifiers. The three parameters are learned using the multi-objective optimization by NSGA-II [23].

4. Experimental Results

We compare our approach with various approaches applied into diagram of Fig. 1. Three segmentation approaches compared are shown in Section 3.1.1. The three feature selection methods compared are:

(1) Sequential forward floating search (SFFS) evaluated by Bayesian classifier [6],
(2) Feature selection evaluated by max-dependency, max-relevance and min-redundancy (mRMR) [5],
(3) Feature selection evaluated by entropy [15].

4.1. Datasets

The proposed method is applied to the following two publicly available image databases,
(1) Caltech-101 image database [16]: composed of images of objects belonging to 101 categories, for 40 to 800 images per category.
(2) Butterfly database (http://janzen.sas.upenn.edu/): It contains 30 classes with a total of 7600 images of butterflies. Example images are shown in Fig. 2.

4.2. Different Segmentation and Feature Selection Methods

Comparison among different methods (using the same system in Fig. 1) is shown in Table 1. The region growing and watershed segmentation outperform the normalized cut. Also among feature selection, the SFFS-Bayesian has the highest performance, with more compact dimensions. Table 1 also indicates the advantages of local features over the global features.

Table 1. Comparison of the final classification results, with corresponding feature subset dimensions and contents.

Approaches			Caltech-101		Butterfly	
Feature form	Feature selection	Segmentation	Classif-ication	Reduced dimension	Classif-ication	Reduced dimension
Local feature with multiple kernels	SFFS-Bayesian	Region growing	86.7%	8	89.5%	8
		Watershed	86.4%	8	89.6%	10
		N-cut	84.9%	8	88.5%	11
	mRMR	Region growing	73.6%	8	60.4%	8
		Watershed	73.1%	9	58.6%	9
		N-cut	72.2%	7	58.8%	9
	Entropy	Region growing	70.9%	11	55.2%	8
		Watershed	71.3%	11	54.6%	8
		N-cut	70.3%	11	54.8%	6
Global feature + multiple kernels	SFFS-Bayesian	NA	85.4%	11	88.4%	12
	mRMR	NA	72.5%	9	58.0%	8
	Entropy	NA	70.3%	11	54.1%	8

4.3. Local Kernel for Feature Selection: Compare with Local Feature Selection

In this section, the proposed method using local kernel for feature selection, is compared to another local feature selection method [22], which uses region-based image representation for feature selection. The method in [22] is integrated into the 'evaluation measure' block in Fig. 1, instead of the local kernel. Fig. 3 shows the comparison between the two methods in different feature selection iterations. Our method has the higher performance compared to that of [22].

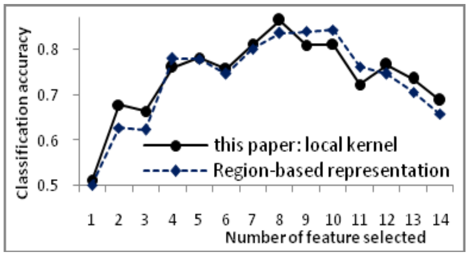

Fig. 3. Proposed method compared to [22], for Caltech-101 database, with SFFS feature selection.

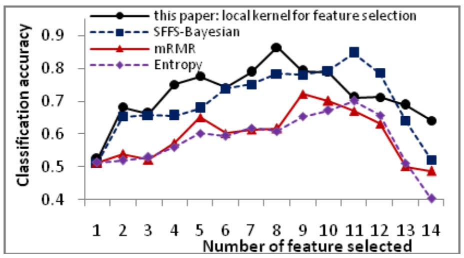

Fig 4. Comparison of classification results with feature selection iterations (Caltech-101).

4.4. Local Kernel for Feature Selection: Compare with Global Feature Selection

The benefits of the proposed method against the other three global feature selection methods are shown in Fig. 4 for the Caltech-101 datasets, where the classification results are displayed for incremental feature selection iterations. The peak accuracy is regarded as the final image classification results, with corresponding optimal feature subset. It can be seen in Fig. 4 that our method outperforms other global feature selection methods in most feature selection iterations, and reaches the highest peak classification accuracy.

5. Conclusions

In this paper, we presented a new approach for the local feature selection with multiple kernel learning. The proposed approach combines feature selection and local feature information by integration of region-based image similarity metric into the Gaussian kernel of a Bayesian classifier. We performed experiments to indicate the benefits of the proposed method. Our future work will focus on integrating the proposed approach with image retrieval.

Acknowledgements: This research was supported in part by US National Science Foundation grants 0641076 and 0727129.

References

[1] M. Toews, T. Arbel. Detection, Localization, and Sex Classification of Faces from Arbitrary Viewpoints and Under Occlusion. *PAMI*, vol. 31, no. 9, pp. 1567-1581, Sept. 2009.

[2] L. I. Kuncheva, J. J. Rodriguez, C. Plumpton, D. E. J. Linden, S. J. Johnston. Random Subspace Ensembles for fMRI Classification. *TMI*, vol. 29, no. 2, pp. 531-542, Feb. 2010.

[3] L. Jing, N. Allinson, T. Dacheng, L. Xuelong. Multitraining Support Vector Machine for Image Retrieval. *IEEE Trans. on Image Processing*, vol. 15, no. 11, Oct. 2006.

[4] H. Wei, S. Billings. Feature Subset Selection and Ranking for Data Dimensionality Reduction. *PAMI*, vol. 29, no. 1, 2007.

[5] H. Peng, F. Long, C. Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE TPAMI*, Vol. 27, No. 8, Aug 2005.

[6] D. Ververidis, C. Kotropoulos. Fast and Accurate Sequential Floating Forward Feature Selection with the Bayes Classifier Applied to Speech Emotion Recognition. *Elsevier Signal Processing*, Vol. 88, No. 12, 2008.

[7] J. Brank, M. Grobelnik, N. Milic-Frayling, D. Mladenic. Feature Selection Using Linear Support Vector Machines. *Microsoft Research Technical Report*, June 2002.

[8] P. Dollar, Z. Tu, H. Tao, S. Belongie. Feature Mining for Image Classification. *CVPR*, 2007.

[9] L. C. Molina, L. Belanche. Feature selection algorithms: a survey and experimental evaluation. *ICDM*, 2002.

[10] B. Bhanu, R Li, John Heraty, E. Murray. Automated Classification of Skippers based on Parts Representation. Entomology Society of America, 2007.

[11] C. Tomasi, L. J. Guibas. The Earth Mover's Distance as a Metric for Image Retrieval. *Intl. Journal of Computer Vision*, vol. 40, no. 2, pp. 99–121, Nov. 2000

[12] J. Li , J. Z. Wang, G. Wiederhold. IRM: Integrated Region Matching for Image Retrieval. *ACM Multimedia*, 2000.

[13] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *IJCV*, vol. 60, no. 2, pp. Jan. 2004.

[14] R. Adams, L. Bischof. Seeded Region Growing. *IEEE Trans on Image Processing*, Vol. 16, No. 6, June 1994.

[15] N. Abe, M. Kudo. Entropy Criterion for Classifier-Independent Feature Selection. *Intl. Conf. on Knowledge - Based and Intelligent Information & Engin. Systems*, 2005.

[16] L. Fei-Fei, R. Fergus, and P. Perona. One-Shot Learning of Object Categories. *PAMI*, vol. 28, no. 4, pp. 594-611, 2006.

[17] D. Kim, S. Yu. A new region filtering and region weighting approach to relevance feedback in content-based image retrieval. *J. of Systems and Software*, vol. 81, no. 9, Sept. 2008.

[18] S. Beucher. The Watershed Transformation Applied to Image Segmentation. *Pfefferkorn Conf. on Signal and Image Processing in Microscopy and Microanalysis*, 1991.

[19] J. Shi, J. Malik. Normalized Cuts and Image Segmentation. *IEEE TPAMI*, vol. 22, no. 8, pp. 888-905, Aug. 2000.

[20] D. Mohamad. Multi Local feature Selection Using genetic Algorithm For Face Identification. *Intl. Journal of Image Processing*, vol. 1, no. 2, pp. 1-28, July 2007.

[21] F. Li, Q. Dai, W. Xu. Improved Similarity-Based Online Feature Selection in Region-Based Image Retrieval. *ICME*, 2006.

[22] W. Jiang, G. Er, Q. Dai, L. Gu. Similarity-based Online Feature Selection in Content-based Image Retrieval. *IEEE Trans. on Image Processing*, vol. 15, no. 3, Mar. 2006.

[23] S. P. Kodali, R. Kudikala and K. Deb, "Multi-Objective Optimization of Surface Grinding Process Using NSGA-II," *Proc. Intl. Conference on Emerging Trends in Engineering and Technology*, Nagpur, India, 2008, pp. 763-767.