Discriminative Data Transform for Image Feature Extraction and Classification

Yang Song[1], Weidong Cai[1], Seungil Huh[2], Mei Chen[3], Takeo Kanade[2], Yun Zhou[4], and Dagan Feng[1]

[1] BMIT Research Group, School of IT, University of Sydney, Australia
[2] Robotics Institute, Carnegie Mellon University, United States
[3] Intel Science and Technology Center on Embedded Computing, United States
[4] Johns Hopkins University School of Medicine, United States

Abstract. Good feature design is important to achieve effective image classification. This paper presents a novel feature design with two main contributions. First, prior to computing the feature descriptors, we propose to transform the images with learning-based filters to obtain more representative feature descriptors. Second, we propose to transform the computed descriptors with another set of learning-based filters to further improve the classification accuracy. In this way, while generic feature descriptors are used, data-adaptive information is integrated into the feature extraction process based on the optimization objective to enhance the discriminative power of feature descriptors. The feature design is applicable to different application domains, and is evaluated on both lung tissue classification in high-resolution computed tomography (HRCT) images and apoptosis detection in time-lapse phase contrast microscopy image sequences. Both experiments show promising performance improvements over the state-of-the-art.

1 Introduction

Image classification in medical imaging has many important applications such as differentiation between different disease types. Other types of medical imaging problems can also be approached with image pixel- or patch-wise classification as the basis, such as object detection and segmentation. Image classification is normally performed with two main components: feature extraction and classification. While many different classifiers have been proposed, such as k-nearest neighbor (kNN) and support vector machine (SVM), it is critical that the feature set is descriptive and discriminative, so that the classifier would be able to differentiate between different classes.

Feature design has been an active research topic in computer vision and medical imaging. The various features that are often used include the relatively traditional ones such as intensities [5], wavelets [4] and shapes [6]. The more recently proposed generic descriptors, such as the local binary patterns (LBP) and scale-invariant feature transform (SIFT), have also been incorporated to represent the texture and gradient information [11,12]. In addition, some studies focus

on designing customized feature set and have demonstrated higher classification performance [2,10].

In this work, we propose a new feature design with two main contributions. First, based on two popular texture descriptors, i.e. LBP and binary robust independent elementary features (BRIEF) [1], we design a more discriminative way of feature extraction with learning-based image transform. Second, we design another learning-based method to further transform the extracted feature descriptor for more accurate image classification. Our method is inspired by the recent studies on face recognition [7,8], in which the images are transformed with learned filters to obtain more discriminative features. We extend the LBP-based filter learning [8] to the BRIEF feature and descriptor transform.

To evaluate the proposed method, our feature design is integrated into two different applications: classification of lung tissues with interstitial lung disease (ILD) on HRCT images, and apoptosis detection in time-lapse phase contrast microscopy image sequences. In both applications, classification is required at image patch-level, hence the features are computed at patch-level. Note that our feature design can be applied to other levels, such as regions and images.

2   Methods

2.1   Patch-Based Feature Vectors

We choose to use a feature set including: LBP, BRIEF, histogram of oriented gradients (HOG) [3], and grayscale histogram. These four types of features are chosen since they provide complementary information, and can be generally effective for different imaging domains. These feature descriptors also have smaller dimensions, compared to features such as SIFT. This property is especially useful for medical imaging typically with a small size of training data.

To enhance the discriminative power of LBP and BRIEF features, we propose a data-adaptive image transform method with learning-based image filtering. A common characteristic of LBP and BRIEF is that both benefit from initial image filtering. Specifically, it has been reported that LBP becomes more descriptive when combined with Gabor filters [10]; and the design of BRIEF requires image preprocessing with Gaussian filters [1]. The choice of Gabor or Gaussian filters and their parameters is important to achieve good classification, but such choices are generally made based on empirical studies. Our proposed method addresses this limitation of filter selection by replacing the filters designed manually with filters learned from data, resulting in more discriminative features. The method details are described in Section 2.2.

We also propose learning-based descriptor transform to further enhance the quality of feature descriptors. Note that all four types of feature descriptors can be considered as histograms. The quantizations would result in very different features between similar image patches due to small deformation. Recently diffusion distance [9] has been proposed to reduce the quantization effect for similarity measure. Its idea of Gaussian filtering to support cross-bin distance computation is inspiring, but the choice of using a Gaussian filter and its parameters
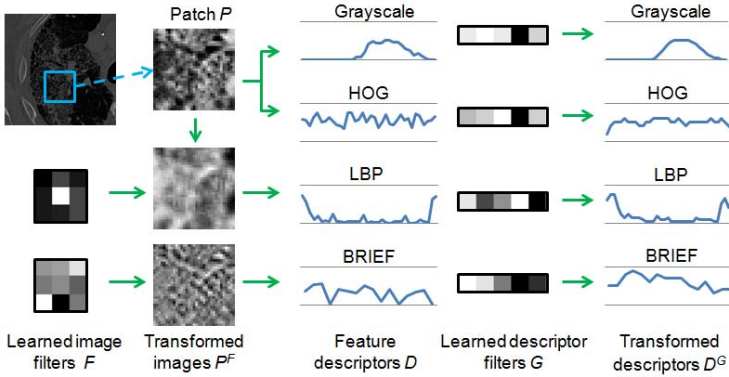
Fig. 1. Illustration of the proposed design. Grayscale histogram and HOG are computed based on the original image patch. To compute LBP and BRIEF features, the image patch is first transformed with learned image filters. The feature descriptors are then transformed with learned descriptor filters to produce the final feature vector.

might not be suitable for all types of data. Our approach incorporates cross-bin information in a data-adaptive manner with learning-based descriptor filtering, achieving better classification. The method details are described in Section 2.3. The overall feature extraction process is illustrated in Fig. 1.

2.2 Image Transform for Feature Extraction

Let us abstract the two types of feature descriptors – Gabor filter with LBP and Gaussian filter with BRIEF – as one feature extraction technique in two steps: (1) image patch transform and (2) feature descriptor computation. Rather than predefining the image filters, we propose a learning-based image transform method. Formally, for an image patch $P = \{p_i : i = 1, ..., N\}$ with N pixels, assume an $X \times X$ pixel image filter F is given ($X^2 < N$). The transformed image patch can be derived as $P^F = P * F$, based on which the feature descriptor $D(P^F)$ can be computed. The goal is to learn an image filter F that enhances the discriminative power of $D(P^F)$.

To do this, we formulate the image filter as a vector $F \in \mathbb{R}^{X^2 \times 1}$, and the transformed value at each pixel p_i as $p_i^F = F^T \mathcal{B}(p_i)$, where $\mathcal{B}(p_i) \in \mathbb{R}^{X^2 \times 1}$ indicates the pixel block centered at p_i. Given an H-dimensional pixel-level feature descriptor $d(p_i^F) \in \mathbb{R}^{1 \times H}$ computed for p_i^F, we expect $d(p_i^F)$ to exhibit small within-class and large between-class variations. In this way, the patch-level feature descriptor $D(P^F)$ obtained based on the pixel-level descriptors would be highly discriminative. This establishes the underlying objective for deriving F.

Specifically, with a set of training patches $\{P\}$, assume the training patches are transformed with F, and Q pixels $\{p_q^F\}_{q=1}^Q$ are sampled from the transformed training patches. The within- and between-class variations are defined as V_w^F and V_b^F, respectively:

$$V_w^F = \sum_l \sum_{q:L(q)=l} (d(p_q^F) - |d(p_q^F)|_l)(d(p_q^F) - |d(p_q^F)|_l)^T \tag{1}$$

$$V_b^F = \sum_l Q_l(|d(p_q^F)|_l - |d(p_q^F)|)(|d(p_q^F)|_l - |d(p_q^F)|)^T \tag{2}$$

where l indexes the different classes, $L(q)$ indicates the labeling of pixel p_q, and Q_l denotes the number of samples of class l. Here $|\cdot|$ represents the element-wise mean: $|d(p_q^F)|_l = Q_l^{-1} \sum_{q:L(q)=l} d(p_q^F)$, and $|d(p_q^F)| = Q^{-1} \sum_q d(p_q^F)$.

Next, with $p_q^F = F^T \mathcal{B}(p_q)$ and our design of $d(p_i)$ based on the standard LBP and BRIEF (described later in this section), we obtain $d(p_q^F) = F^T \mathcal{D}(p_q)$. $\mathcal{D}(p_q) = \{d(p_i) : p_i \in \mathcal{B}(p_q)\} \in \mathbb{R}^{X^2 \times H}$ represents the matrix by concatenating the pixel-level descriptors $d(p_i) \in \mathbb{R}^{1 \times H}$ surrounding p_q. By substituting $d(p_i)$, the formulation of $d(p_q^F) = F^T \mathcal{D}(p_q)$ can be easily verified. The variations are then rewritten as:

$$V_w^F = F^T(\sum_l \sum_{q:L(q)=l} (\mathcal{D}(p_q) - |\mathcal{D}(p_q)|_l)(\mathcal{D}(p_q) - |\mathcal{D}(p_q)|_l)^T)F \tag{3}$$

$$V_b^F = F^T(\sum_l Q_l(|\mathcal{D}(p_q)|_l - |\mathcal{D}(p_q)|)(|\mathcal{D}(p_q)|_l - |\mathcal{D}(p_q)|)^T)F \tag{4}$$

Then the image filter F can be derived by minimizing V_w^F and maximizing V_b^F. In particular, we formulate $V_w^F = F^T U_w^F F$ and $V_b^F = F^T U_b^F F$, and solve a generalized eigenvalue problem $U_b^F F = \lambda U_w^F F$ to obtain F [8]. F thus becomes adaptive to training images, and is not restricted to a certain filter type such as Gabor or Gaussian. The only parameter is the filter size X. By using multiple settings of X, a multi-scale feature descriptor $D(P^F)$ can be generated. Note that while the filter learning is based on Fisher criterion, our method is entirely different from the fisher discriminant analysis (FDA) in that it does not involve feature dimension reduction and is coupled with pixel-level descriptors of features to enhance the discriminative power of the features.

Finally, we describe how the pixel-level descriptor $d(p_i)$ is computed. In LBP, to compute the patch-level feature descriptor, each pixel is compared with its surrounding neighbors, and the difference relationships of all pixels in a patch are combined into a feature histogram. At pixel-level, we define $d(p_i)$ as a row vector: $d(p_i) = \{\mathcal{N}(p_i, 0) - p_i, \cdots, \mathcal{N}(p_i, M-1) - p_i\}$, where $\mathcal{N}(p_i, m)$ indicates the mth neighboring pixel of p_i and M is the number of neighbors. The sampling of p_q follows a uniform distribution with inter-sample spacing of X.

In BRIEF, a pixel is compared with a corresponding pixel to obtain the binary bit 0 or 1; and a number of such pairwise comparisons are sampled in P and combined as patch-wise feature. Our sampling of p_q follows the default scheme used in BRIEF. The descriptor is thus defined as a scalar in two ways: for a sampled pixel p_q, $d(p_q) = p_q - p_{q'}$ with $p_{q'}$ denoting the corresponding pixel of p_q in P; for pixels $p_i \in \mathcal{B}(p_q)$ and $i \neq q$, $d(p_i) = p_i - p_{i'}$ with $p_{i'} \in \mathcal{B}(p_{q'})$ and at the corresponding location as p_i.

Note that one filter F is learned for each feature type and scale. In other words, assume we compute a three-scale LBP and a three-scale BRIEF. Three F filters would be learned for LBP and another three for BRIEF. Each F is then convoluted with the image patch P to produce a transformed image P^F, based on which, a feature descriptor $D(P^F)$ is derived. For HOG and grayscale histogram, the feature descriptors $D(P)$ are computed directly on P.

2.3   Descriptor Transform for Feature Classification

Next, we propose a learning-based method to further enhance the discriminative power of the feature descriptors, using a similar approach as the image transform. Such descriptor transforms are performed on the feature descriptors D derived as described in the previous section (i.e. $D(P^F)$ for LBP or BRIEF and $D(P)$ for HOG or grayscale histogram), and one filter G is learned for each feature type and scale. Formally, denote a feature descriptor of a certain type (i.e. LBP, BRIEF, HOG or grayscale histogram) at a certain scale as $D = \{h_k : k = 1, ..., K\}$ with K as the feature dimension (K varies for different feature types). The goal is to find a filter G so that the transformed feature descriptor D^G achieves small within-class and large between-class variations. The final feature vector of an image patch P is then the concatenation of all D^G of all feature types and scales, and P is classified with this feature vector.

We define the filter as a Y-dimensional vector: $G \in \mathbb{R}^{Y \times 1}$ with $Y < K$. The descriptor transform is computed as $D^G = D * G$, and each transformed feature element can be represented as $h_k^G = G^T \mathcal{H}_k$, with $\mathcal{H}_k \in \mathbb{R}^{Y \times 1}$ indicating the Y elements centered at h_k. Given a set of training feature descriptors $\{D\}$ and J feature elements $\{h_j\}_{j=1}^{J}$, with J_l elements of class l, the within- and between-class variations are defined as V_w^G and V_b^G:

$$V_w^G = \sum_l \sum_{j:L(j)=l} (h_j^G - |h_j^G|_l)(h_j^G - |h_j^G|_l)^T \tag{5}$$

$$V_b^G = \sum_l J_l(|h_j^G|_l - |h_j^G|)(|h_j^G|_l - |h_j^G|)^T \tag{6}$$

By substituting $h_j^G = G^T \mathcal{H}_j$, we obtain the following:

$$V_w^G = G^T(\sum_l \sum_{j:L(j)=l} (\mathcal{H}_j - |\mathcal{H}_j|_l)(\mathcal{H}_j - |\mathcal{H}_j|_l)^T)G \tag{7}$$

$$V_b^G = G^T(\sum_l J_l(|\mathcal{H}_j|_l - |\mathcal{H}_j|)(|\mathcal{H}_j|_l - |\mathcal{H}_j|)^T)G \tag{8}$$

Similar to the image transform, a generalized eigenvalue problem is formulated to solve for G. The derived G is thus data-adaptive without requiring a predefined filter type, and the filter size Y is the only parameter involved.

2.4   Datasets and Implementation

We use two datasets of different applications for the experiments. First, 93 HRCT image sets [4] are used to classify lung tissues with five categories of lung tissues:

normal, emphysema, ground glass, fibrosis, and micronodules. We use the same sparse representation classifier, HOG descriptor and grayscale histogram as [10]. Changes in feature extraction include: (1) three-scale LBP descriptors with three F filters of size $X = 3$, $X = 5$ and $X = 7$, and 8 immediate neighbors; (2) three-scale 128-bit BRIEF descriptors with three F filters of size $X = 3$, $X = 5$ and $X = 7$. Furthermore, the descriptor of each feature type and scale is transformed with filter G of size $Y = 5$, and a total of eight G filters are used. The dataset is divided into four parts, within each part, leave-one-subject-out method is applied for filter learning and classifier training.

Second, three time-lapse microscopy image sequences of C2C12 myoblastic stem cells [5] are used for apoptosis detection. Each image sequence comprises 540 frames, and a total of 1154 apoptosis cases are annotated. In this work, we follow the three-step design in [5] but replace the feature extraction step. In particular, for each image patch containing an apoptosis candidate, four types of features are computed: (1) two-scale LBP descriptors with two F filters of size $X = 3$ and $X = 5$ and 8 immediate neighbors; (2) two-scale 128-bit BRIEF descriptors with two F filters of size $X = 3$ and $X = 5$; (3) HOG descriptor of 4 cells and 9 bins [10]; and (4) grayscale histogram between adjacent frames [5]. The descriptor of each feature type and scale is further transformed with filter G of size $Y = 5$. The final concatenated feature vector is classified using a linear-kernel SVM [5]. Altogether four F filters and six G filters are learned using cross-validation, and the same training set is also used to train classifiers.

3   Results

For lung tissue classification, we evaluate the following feature extraction methods besides the proposed approach: (L1) three-scale rotation-invariant Gabor filter with LBP, HOG and grayscale histogram [10]; (L2) three-scale rotation-invariant Gabor filter with LBP, HOG, grayscale histogram and three-scale Gaussian filter with BRIEF; (L3) proposed three-scale image filter with LBP, HOG and grayscale histogram, and proposed three-scale image filter with BRIEF. None of the above methods involves the descriptor transform as proposed.

Fig. 2 clearly shows that our proposed method enhances the discriminative power of features. The usefulness of a multi-scale Gabor filter with LBP, HOG and grayscale histogram is already demonstrated in [10], and comparison between (L2) and (L1) shows the benefit of adding BRIEF features. While (L2) is computed with customized three-scale and rotation-invariant Gabor and Gaussian filters, (L3) is based on learning-based image filters. The advantage of (L3) over (L2) suggests that the learning-based data-adaptive filters can be more effective than the manually-customized filter design. Further performance improvement is then observed by adding the descriptor transforms, as indicated by comparing our results with (L3).

For apoptosis detection, we compare the proposed approach with the following feature extraction methods: (C1) LBP and grayscale histogram [5]; (C2) LBP, grayscale histogram, BRIEF and HOG; (C3) two-scale Gabor filter with LBP,
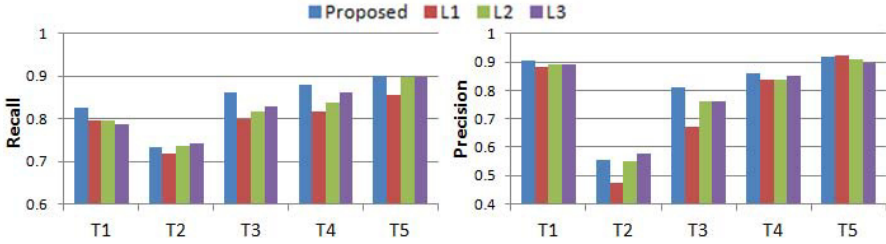
Fig. 2. Classification recall and precision of lung image patches. T1: normal, T2: emphysema, T3: ground glass, T4: fibrosis, and T5: micronodules.

grayscale histogram, two-scale Gaussian filter with BRIEF, and HOG; (C4) proposed two-scale image filter with LBP, grayscale histogram, proposed two-scale image filter with BRIEF, and HOG.

Table 1 demonstrates the advantage of our proposed method. The results using (C2) show that simply adding more features onto (C1) does not bring any measurable benefit. The usefulness of adding Gabor and Gaussian filters as in (C3) is also negligible. Incorporating the learning-based image transforms, however, does improve the performance, as shown by (C4). This suggests that the data-adaptive approach can however be applied in the same way for both lung and apoptosis datasets with similar benefits. Comparison between the proposed method and (C4) shows the advantage of descriptor transforms. The results also show that our feature design is generally effective with different classifiers, e.g. sparse representation for lung images and SVM for apoptosis images.

Table 1. The average recall, precision and F-score of apoptosis detection

	Proposed	C1	C2	C3	C4
Recall (%)	91.3	89.4	89.3	89.5	90.6
Precision (%)	94.2	92.4	92.5	92.4	92.8
F-score (%)	92.7	90.8	90.9	90.9	91.7

Note that although there seems to be a large number of F and G filters involved, filter learning is fast due to the simple operations for computing $U_w^{F/G}$ and $U_b^{F/G}$ and solving the generalized eigenvalue problem. The image/descriptor transforms are also low-cost with convolution. The overall computational complexity is thus comparable to the basic feature extraction approaches.

4   Conclusion

In this work, we proposed a new feature design to improve image classification performance. A data-adaptive approach is designed to learn discriminative image

and descriptor filters, which are then used to transform the images prior to computing feature descriptors, and to transform the derived feature descriptors as the final input to classifiers. The feature design is integrated into two different applications for performance evaluation, and achieved encouraging performance improvements compared to the state-of-the-art.

References

1. Calonder, M., Lepetit, V., Strecha, C., Fua, P.: BRIEF: Binary Robust Independent Elementary Features. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part IV. LNCS, vol. 6314, pp. 778–792. Springer, Heidelberg (2010)
2. Cheng, L., Ye, N., Yu, W., Cheah, A.: Discriminative Segmentation of Microscopic Cellular Images. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part I. LNCS, vol. 6891, pp. 637–644. Springer, Heidelberg (2011)
3. Dalal, N., Triggs, B.: Histograms of Oriented Gradients for Human Detection. In: CVPR, pp. 886–893 (2005)
4. Depeursinge, A., de Ville, D.V., Platon, A., Geissbuhler, A., Poletti, P.A., Muller, H.: Near-Affine-Invariant Texture Learning for Lung Tissue Analysis Using Isotropic Wavelet Frames. IEEE Trans. Inf. Technol. Biomed. 16(4), 665–675 (2012)
5. Huh, S., Ker, D.F.E., Su, H., Kanade, T.: Apoptosis Detection for Adherent Cell Populations in Time-Lapse Phase-Contrast Microscopy Images. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 331–339. Springer, Heidelberg (2012)
6. Jacobs, C., Sánchez, C.I., Saur, S.C., Twellmann, T., de Jong, P.A., van Ginneken, B.: Computer-Aided Detection of Ground Glass Nodules in Thoracic CT Images Using Shape, Intensity and Context features. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 207–214. Springer, Heidelberg (2011)
7. Kumar, R., Banerjee, A., Vemuri, B.C.: Voterrafaces: Discriminant Analysis using Volterra Kernels. In: CVPR, pp. 150–155 (2009)
8. Lei, Z., Yi, D., Li, S.Z.: Discriminant Image Filter Learning for Face Recognition with Local Binary Pattern Like Representation. In: CVPR, pp. 2512–2517 (2012)
9. Ling, H., Okada, K.: Diffusion Distance for Histogram Comparison. In: CVPR, pp. 246–253 (2006)
10. Song, Y., Cai, W., Zhou, Y., Feng, D.D.: Feature-based Image Patch Approximation for Lung Tissue Classification. IEEE Trans. Med. Imag., 1–12 (2013)
11. Sorensen, L., Shaker, S.B., de Bruijne, M.: Quantitative Analysis of Pulmonary Emphysema Using Local Binary Patterns. IEEE Trans. Med. Imag. 29(2), 559–569 (2010)
12. Xu, Y., Zhang, J., Chang, E.I.-C., Lai, M., Tu, Z.: Context-Constrained Multiple Instance Learning for Histopathology Image Segmentation. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part III. LNCS, vol. 7512, pp. 623–630. Springer, Heidelberg (2012)