

Classifying Textile Designs using Bags of Shapes

Wei Jia, Stephen J. McKenna
 School of Computing, University of Dundee
 weijia,stephen@computing.dundee.ac.uk

Abstract

The use of region shape descriptors was investigated for categorisation of textile design images. Images were segmented using MRF pixel labelling and the shapes of regions obtained were described with generic Fourier descriptors. Each image was represented as a bag of shapes. A simple yet competitive classification scheme based on nearest neighbour class-based matching was used. Classification performance was compared to that obtained when using bags of SIFT features.

1. Introduction

The use of segmented regions for image recognition has seen some recent interest [2, 4, 5] but strategies based on low-level patch descriptors such as SIFT [9] are often preferred [8], at least in part due to concerns about unreliable segmentation. In this paper, the images of interest are mainly of textiles with printed or woven designs. We have previously shown that reasonable region segmentation results can be achieved on such images using Markov random field pixel-labelling [6]. This approach was partly motivated by knowledge of production which can involve the use of colour masks specified by designers. We investigate the use of bags of shapes computed from such segmentations, focusing on a classification task. This has application in image cataloguing. We are also interested in using such representations for retrieval and browsing applications [11].

Experiments reported in Section 4 compare the use of bags of shapes extracted from region segmentations with the use of bags of SIFT descriptors in terms of classification performance. This comparison was performed using a simple but effective classification scheme based on direct nearest neighbour searches in training data available for each class [3]. This scheme was adopted because it has obtained state-of-the-art performance on several well-known natural scene datasets using features such as SIFT [3] and because it involves



Figure 1. Two examples of each class.

neither descriptor quantization nor parameter-sensitive learning algorithms. Experiments were performed using images from a commercial archive (courtesy of Liberty Art Fabrics), categorised into seven classes based on the type of visual design (Figure 1). This choice of categories is not definitive and assigning these images to disjoint categories is not claimed to be an optimal approach to cataloguing them. Nevertheless, this classification problem is interesting and challenging because of large intra-class variations and because images from different classes often have much in common. Consider for example the potential for similar image parts in the floral and leaf classes, or in the geometric and check classes. Furthermore, the textile fragments span more than a century of design history, and are often damaged or degraded.

2. Bags of shapes

Images were segmented into regions using a pixel labelling method described in Jia *et al.* [6]. This method alternates between estimating Gaussian distributions in RGB space and minimising energy in a Markov random field model with the RGB distributions as likelihood functions. Such an energy function has the form

$$E(f) = \sum_p \sum_q \lambda \cdot (1 - \delta(f_p - f_q)) - \sum_p \ln P(x_p | f_p), \quad (1)$$

where f is an image labelling, f_p denotes the label assigned to a pixel p , and pixels denoted q are neighbours of pixel p . The first term rewards spatial coherence and the second term rewards a good fit for the colour distributions. The parameter $\lambda \geq 0$ specifies the penalty for assigning different labels to neighboring pixels. Optimization was performed using α -expansion steps with $\lambda = 4$. Further details can be found elsewhere [6].

Figure 2 shows an example of a pixel labelling. In this case, there are seven labels, each of which has multiple regions (connected components). In this paper, the labels are subsequently ignored and an image is represented as a single bag of regions with each region represented by a feature vector characterising its shape. Thus, an image is represented as a *bag of shapes*.

Shape was described using generic Fourier descriptors (GFD) [12]. Specifically, a 2D Fourier transform was applied to a polar representation $f(r, \theta)$ of each binary region image (see Figure 3):

$$F(\rho, \phi) = \sum_r \sum_\theta f(r, \theta) \exp[-j2\pi(r\rho + \theta\phi)], \quad (2)$$

where $0 \leq r < 1$ and $0 \leq \theta < 2\pi$. The generic Fourier descriptor (GFD) is:

$$\mathbf{d} = \left(\frac{|F(0,0)|}{\text{area}}, \frac{|F(0,1)|}{|F(0,0)|}, \dots, \frac{|F(m-1,n-1)|}{|F(0,0)|} \right) \quad (3)$$

Setting $m = 4$ and $n = 12$ gave a 48-dimensional feature vector. This representation is rotation, translation and scale invariant.

3. Naive-Bayes Nearest-Neighbor

A query image, Q , can be classified by assigning it to the class, \hat{C} , with largest posterior probability. This is equivalent to maximising the class likelihood if a uniform prior is adopted. Given a representation of Q in terms of a set of descriptors $\mathbf{d}_1, \dots, \mathbf{d}_n$ assumed to be conditionally independent, the classification rule becomes

$$\hat{C} = \arg \max_C \sum_{i=1}^n \log(p(\mathbf{d}_i | C)). \quad (4)$$

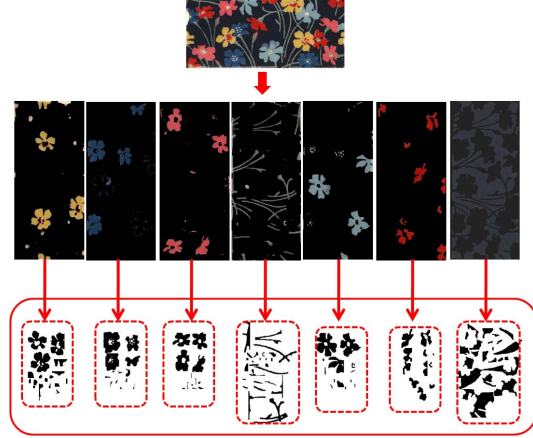


Figure 2. Top: original image. Middle: pixel labelling. Bottom: regions.



Figure 3. A region in polar coordinates

Since the number of descriptors in a training set will be very large, Parzen density estimation can be used to estimate $p(\mathbf{d}|C)$:

$$\hat{p}(\mathbf{d}|C) = \frac{1}{M} \sum_{j=1}^M K(\mathbf{d} - \mathbf{d}_j^C), \quad (5)$$

where M is the number of descriptors obtained from all of the training images of class C and K is a Parzen kernel function. Computing the sum over M terms is expensive. However, this sum is well approximated using a sum only over the nearest neighbours because of the long-tail properties of the descriptor distribution. Boiman *et al.* [3] provided empirical evidence using several popular image classification data sets that even using a single nearest neighbour can be effective and indeed highly competitive with the state-of-the-art in image classification. They observed very little effect on the discriminative ability of local feature descriptors. In the case of a Gaussian kernel,

$$K_g(\mathbf{d} - \mathbf{d}_j^C) = \frac{1}{(2\pi)^{d/2} \sigma^d} \exp\left(-\frac{\|\mathbf{d} - \mathbf{d}_j^C\|^2}{2\sigma^2}\right), \quad (6)$$

this results in the following simple algorithm [3]: Calculate descriptors $\mathbf{d}_1, \dots, \mathbf{d}_n$ for a query image Q and for each such descriptor \mathbf{d}_i , find its nearest neighbor $NN_C(\mathbf{d}_i)$ for each class C . Assign the query to the class $\hat{C} = \arg \min_C \sum_{i=1}^n \|\mathbf{d}_i - NN_C(\mathbf{d}_i)\|^2$. In order to

speed up nearest neighbour search, KD-trees were used. The time complexity is $O(N \log N)$, where N is the number of items to search¹.

We also experimented with an exponential kernel instead of a Gaussian:

$$K_e(\mathbf{d} - \mathbf{d}_j^C) = a \exp(-a \|\mathbf{d} - \mathbf{d}_j^C\|) \quad (7)$$

resulting in the classification rule $\hat{C} = \arg \min_C \sum_{i=1}^n \|\mathbf{d}_i - NN_C(\mathbf{d}_i)\|$. This kernel has heavier tails. Additionally, we experimented with the L_1 norm in place of the L_2 norm since it is known to perform better for some high-dimensional matching problems (see e.g. [1]).

4. Experiments

A set of 490 images, 70 per class, was used. For each trial, the image set for each class was divided at random into n_{train} training images and n_{test} test images with $n_{train} + n_{test} = 70$. The classification methods described in Section 3 were compared when feature descriptors were region shape descriptors (MRF-GFD) and when they were SIFT local feature descriptors [9]. Two methods of computing bags of SIFT features were used:

EP-SIFT: Local patches were horizontally and vertically separated by 20 pixels.

RP-SIFT: Local patches were positioned uniformly at random.

In both cases, patch diameters were sampled at random in the range 10 to 30 pixels. The use of SIFT descriptors extracted at interest points has been reported to be inferior to these methods in a scene classification task [8]. Each patch was described using a 128-dimensional SIFT feature vector².

When using MRF-GFD, the number of regions obtained varied from image to image, averaging around 500. Instead of using all regions, only the largest regions in an image were used to represent it. As well as reducing computational expense, this reflected the intuition that larger regions are more likely to be informative for classification. Specifically, in each experiment an upper bound was set on the number of regions per image to be used. If the number of regions in an image was below this bound, all its regions were used. Figure 4 shows how classification accuracy and average matching time varied with this bound on the number of descriptors per image, and in the case of SIFT

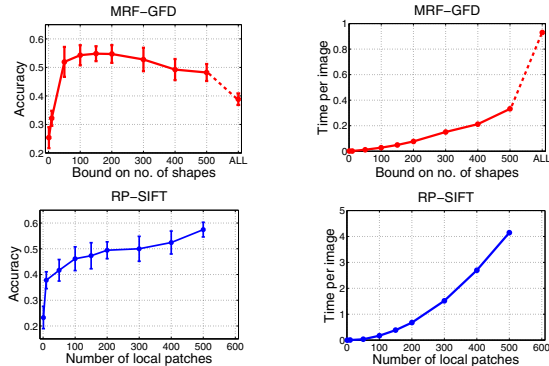


Figure 4. Classification accuracy and average matching time per image.

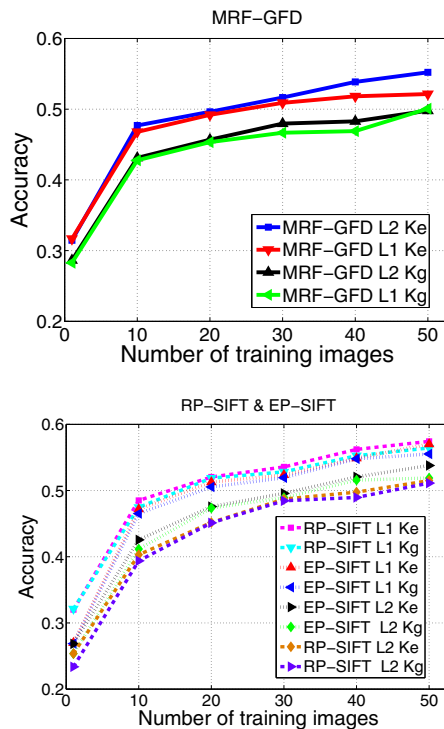


Figure 5. Classification accuracy.

features (RP-SIFT) with the number of descriptors per image. The number of training examples per class was $n_{train} = 50$. Other experiments described in this Section used 500 descriptors in the case of RP-SIFT, and an upper bound of 150 descriptors in the case of MRF-GFD.

Figure 5 illustrates the effect of using different kernels (Gaussian, K_g or exponential, K_e) and different norms (L_1 or L_2) for classification. The results are plotted for training set sizes $n_{train} \in$

¹The implementation available at www.cs.umd.edu/~mount/ANN/ was used [10]

²www.robots.ox.ac.uk/~vgg/research/affine/descriptors.html

	fl	ps	st	lf	gm	sp	ck
floral	70%	5%	0	25%	0	0	0
paisley	15%	50%	5%	20%	0	0	0
strip	0	0	75%	0	15%	5%	0
leaf	5%	15%	5%	65%	5%	5%	5%
geometric	5%	5%	5%	0	45%	5%	35%
spot	15%	20%	0	5%	5%	50%	5%
check	10%	5%	5%	0	20%	0	60%

Figure 6. An example confusion matrix using bags of shapes for classification

{1, 10, 20, 30, 40, 50}. Each point represents an average of 10 trials with randomly selected training examples. For both descriptor types, the exponential kernel gave better accuracy than the Gaussian kernel, irrespective of the norm. MRF-GFD had slightly better results using L_2 than L_1 . In contrast, L_1 gave better results than L_2 with the (higher dimensional) SIFT features.

Figure 6 shows a confusion matrix obtained using the bag-of-shapes (MRF-GFD) method with $n_{train} = 50$. It reveals the ambiguities that occur between the three classes leaf, floral and paisley, and between the two classes geometric and check.

Pixel labelling [6] implemented in C took approximately 20s (3 iterations) to process a 500×500 pixel image on a 2.40GHz PC. However, it should be noted that faster optimization methods now exist for such functions [7]. Using 50 training images per category it took on average 4.14s to compare one image with one class using 500 SIFT descriptors per image. Bag-of-shapes took 0.04s with up to 150 regions. When up to 500 regions were used it took 0.35s. Bag-of-shapes obtained similar accuracy to SIFT (Figure 5). In the above experiment, matching was two orders of magnitude faster using shapes than when SIFT features were used. Shape features were of lower dimensionality than SIFT and there were fewer per image.

5. Conclusion and future work

This paper introduced the problem of representing images of textile designs for classification. Motivated in part by knowledge of production, a method for computing a representation in terms of regions was used. Each region’s shape was characterised using generic Fourier descriptors. Each image was thus represented as a bag of shapes. A simple yet competitive classification scheme based on nearest neighbour matching was used.

The method was compared to the use of bags of SIFT features and accuracy was comparable. Once features had been computed, matching using bags of shapes was considerably faster. A potential advantage of the pixel labelling method that was not exploited in this paper is the availability of the labels (see the bottom row of Figure 2). This additional structure should enable further improvements to be obtained. Future work should also explore sensitivity to model order, i.e. the number of distinct labels used to represent an image.

Acknowledgments: The authors are grateful to Annette Ward (Univ. of Dundee) and Anna Buruma (Liberty Art Fabrics) for helpful discussions, and to the anonymous reviewers. This research was supported by the UK Technology Strategy Board *FABRIC* project, a collaboration with Liberty Art Fabrics, System Simulation, the Victoria & Albert Museum, and Calico Jack.

References

- [1] C. C. Aggarwal, A. Hinneburg, and D. A. Keim. On the surprising behavior of distance metrics in high dimensional space. *Proc. of ICDT*, pages 420–434, 2001.
- [2] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. From contours to regions: An empirical evaluation. In *CVPR*, 2009.
- [3] O. Boiman, E. Shechtman, and M. Irani. In defense of nearest-neighbor based image classification. In *CVPR*, 2008.
- [4] C. H. Gu, J. J. Lim, P. Arbelaez, and J. Malik. Recognition using regions. In *CVPR*, 2009.
- [5] Z. Harchaoui. Image classification with segmentation graph kernels. In *CVPR*, 2007.
- [6] W. Jia, S. J. McKenna, and A. A. Ward. Extracting printed designs and woven patterns from textile images. In *Int. Conf. Computer Vision Theory and Applications*, 2009.
- [7] N. Komodakis, G. Tziritas, and N. Paragios. Performance vs computational efficiency for optimizing single and dynamic MRFs: Setting the state of the art with primal dual strategies. *Computer Vision and Image Understanding*, 112:14–29, 2008.
- [8] F. F. Li and P. Perona. A Bayesian hierarchical model for learning natural scene categories. In *CVPR*, 2005.
- [9] D. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60 (2):91–110, 2004.
- [10] D. Mount and S. Arya. ANN: A library for approximate nearest neighbor searching. In *CGC 2nd Annual Workshop on Comp. Geometry*, 1997.
- [11] A. A. Ward, S. J. McKenna, A. Buruma, P. Taylor, and J. Han. Merging technology and users: Applying image browsing to the fashion industry for design inspiration. In *Content-Based Multimedia Indexing (CBMI)*, pages 288–295, June 2008.
- [12] D. S. Zhang and G. J. Lu. Shape-based image retrieval using generic Fourier descriptors. *Sig. Proc.: Image Communication*, 17:825–848, 2002.