

# Learning the Optimal Transformation of Salient Features for Image Classification

Jun Zhou<sup>1,2,3</sup>

Zhouyu Fu<sup>4</sup>

Antonio Robles-Kelly<sup>1,2,3</sup>

<sup>1</sup>Canberra Research Lab, NICTA\*

PO BOX 8001, Canberra, ACT 2601, Australia

<sup>2</sup>College of Engineering and Computer Science, The Australian National University  
Canberra, ACT 0200, Australia

<sup>3</sup>School of Engineering and Information Technology, UNSW@ADFA  
Canberra, ACT 2600, Australia

<sup>4</sup>Faculty of Information Technology, Monash University, VIC 3800, Australia

## Abstract

*In this paper, we address the problem of recovering an optimal salient image descriptor transformation for image classification. Our method involves two steps. Firstly, a binary salient map is generated to specify the regions of interest for subsequent image feature extraction. To this end, an optimal cut-off value is recovered by maximising Fisher's linear discriminant separability measure so as to separate the salient regions from the background of the scene. Next, image descriptors are extracted in the foreground region in order to be optimally transformed. The descriptor optimisation problem is cast in a regularised risk minimisation setting, in which the aim of computation is to recover the optimal transformation up to a cost function. The cost function is convex and can be solved using quadratic programming. The results on unsegmented Oxford Flowers database show that the proposed method can achieve classification performance that are comparable to those provided by alternatives elsewhere in the literature which employ pre-segmented images.*

## 1 Introduction

Recently, there has been considerable interest in the pattern recognition and computer vision communities on the image classification problem. An image classification system often summarises an image as a set of features or descriptors represented in the form of vectors. These vectors

---

\*NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program.

are combined into visual codebooks, which are then used to train image classifiers and assign testing images into their corresponding classes [5, 20, 15].

In practice, feature extraction is usually performed on the whole image. This draws in the use of scene matching as a means to image classification. An example of this is the work of Lazebnik *et al.* [10], where scene classification is effected by extending the bag-of-features approach to a spatial pyramid. Nilsback *et al.* [16] have used colour features to identify flowers based on the assumption that similar flowers are likely to grow in akin environments.

Despite effectiveness, in the methods above, the codebook depicts not only the objects of interest, but also the background in the scene. This is a result of using the whole image for feature extraction. Therefore, for those cases where the object of interest is not constrained by the nature of the scene, it is intuitive that “uncluttering” the image by removing the background would improve the performance of the classification task. Thus, extracting features from regions of interest (ROI) depicting the object under study rather than from the whole image can improve the classification results. This assertion is in accordance with the performance of those systems trained and tested on hand-segmented databases, which appear to provide a margin of advantage over those using unsegmented data [25, 26, 12, 27].

For hand-segmented data, ROIs are normally provided by the user as contours or object masks. This requires intensive human involvement [12, 16, 25]. To avoid these tedious and cumbersome labeling procedure, Bosch *et al.* [2] have presented an automatic ROI extraction approach in which rectangular regions are identified from the background by recovering areas with high similarity over the images belonging to the same class. Alternatively, ROIs

can be extracted from saliency maps. Itti *et al.* [7] proposed a bottom-up saliency-based visual attention model to analyse complex scenes. This model simulates the human search strategy by extracting simple feature maps in an image, which are then combined to generate small regions corresponding to the visual attention field. In [8], Kadir and Brady studied the relationship between saliency, image scales and image description. They developed an algorithm to select salient regions in an image using multi-scale analysis via entropy maximisation, and applied it to image retrieval problems. This work was extended by Lee *et al.* [11] using task-dependent processing and entropy derivative analysis so as to solve a scale adaptation problem.

Although saliency maps have been successfully used in object detection and eye tracking [22, 18, 19], its application to image classification is limited [15]. This is partly due to the lack of a mechanism to link the saliency map to the extraction of the image descriptors. Visual attention is normally focused on one or a set of points in the image, which further complicates deciding the appropriate region around the points for feature extraction. Another difficulty is more general to image classification problems and pertains how to optimally transform the extracted feature vectors so as to minimise the variance within the same class while maximising the classification performance. Such transformations may include feature construction, where new features are generated, or feature extraction, where only a subset of features are selected [13].

In this paper, we introduce a method that optimises the generation and transformation of saliency features for image classification. Upon obtaining an image saliency map, the binarisation step is performed using an optimal cut-off value, which is recovered by maximising Fisher’s linear discriminant separability measure. Features are then extracted from the foreground region and transformed based upon the class conditional mean for descriptors under study. We recover an optimal feature transformation by minimising a cost function which can be extremised making use of either, quadratic or linear programming. This minimisation process aims at optimising the classification performance. We provide experimental results and show that the performance of the proposed method on an unsegmented image data set is comparable to those results reported on segmented data.

## 2 Saliency Map Extraction and Binarisation

The first step in the method presented here is to extract a saliency map. To this end, we follow the model by Itti *et al.* [7], which simulates the attention criteria of primate early vision. This model adopts a bottom-up strategy that decomposes visual input into component feature maps. For each feature map, saliency is extracted separately and then combined into a global map.

The procedure is implemented as follows. An input image is first smoothed using Gaussian filters so as to generate a scale pyramid. Simple features are then extracted at each scale to generate feature maps. Differences between fine and coarse scales for each feature are then computed by interpolating from coarse-to-fine in the scale domain and performing pixel-wise subtraction. This guarantees invariance over a pixel neighborhood for both the feature extraction task and the corresponding feature scale-differences. For the final recovery of the global saliency map, a two-level normalisation is effected. The first of these is at the individual feature level, where each feature map is normalised by the difference between the global maximum and the mean of the local maxima. The normalised feature maps at different scales are then combined into a “conspicuity” map, i.e. a map that indicates which regions on the image are visually salient. This conspicuity map corresponds to the global saliency map which we then binarise.

Here, we use three types of visual features. These are intensity, colour and orientation. The intensity feature is obtained by averaging the red, green and blue channel-values at each pixel in the input image. The color feature simulates the function of the cortex, which is represented by a set of color opponency between red green and blue channel values against the yellow basis, where each colour channel is computed as follows

$$\begin{aligned}
 R &= r - (g + b)/2 \\
 G &= g - (r + b)/2 \\
 B &= b - (r + g)/2 \\
 Y &= (r + g)/2 - |r - g|/2 - b
 \end{aligned}
 \tag{2.1}$$

where  $R$ ,  $G$  and  $B$  are the red, green and blue channel values at a pixel in the image and  $Y$  is the respective yellow basis.

The orientation information is obtained using Gabor filters [4]. Recall that the Gabor filters employ a Gaussian envelope modulated by a complex sinusoidal carrier. Here, the responses are computed at each scale and four orientations. For details, please refer to [7]. Upon obtaining the saliency map, the next step is to extract ROIs. It is worth mentioning that, here, in contrast with the sequential salient region recovery approach in [7], we attempt to simultaneously retrieve all salient regions in the image. This is done by optimally computing a cut-off value  $\tau$  so as to separate the ROIs from the image background. In this manner, the problem of ROI extraction can be reformulated into a binarisation task which operates on the normalised saliency map.

For purposes of recovering the cut-off value, we build on [17] and view  $\tau$  as the optimal split arising from a binomially distributed set of univariate random variables  $z_i \in \mathcal{Z}$  which correspond to the saliency values for every pixel in

the image. To do this, we maximise Fisher’s linear discriminant [6] separability measure. This measure is given by

$$\lambda = \frac{S_b^2}{S_w^2} \quad (2.2)$$

where  $S_b, S_w$  are between and within class variances given by

$$S_w^2 = \omega_1 S_1^2 + \omega_2 S_2^2 \quad (2.3)$$

$$S_b^2 = \omega_1 \omega_2 (\mu_1 - \mu_2)^2 \quad (2.4)$$

where  $\mu_i$  and  $S_i$  are the mean and variance of the class indexed  $i$  and  $\omega_1, \omega_2$  are real-valued class weights.

To take our analysis further, we note that the maximum of  $\lambda$  is given by  $\omega^* = \omega_1 = \omega_2$ , where  $\omega^*$  is the optimum value of the weights, which can be computed making use of the expression

$$\omega^* = \frac{\mu_1 - \mu_2}{S_1^2 + S_2^2} \quad (2.5)$$

Moreover, making use of  $\omega^*$ , the optimum cut-off value is given by

$$\tau = \{\eta | (\omega^*)^2 = \omega_\eta (1 - \omega_\eta)\} > 0 \quad (2.6)$$

where  $\omega_\eta$  is a real-valued function of the univariate random variables defined as follows

$$\omega_\eta = \frac{1}{|\Omega_\eta|} \sum_{z_i \in \Omega_\eta} z_i \quad (2.7)$$

and  $\Omega_\eta$  is the set of variables whose value is less than or equal to  $\eta$ , i.e.  $\Omega_\eta = \{x_i | z_i \leq \eta\}$ . Thus, in practice, we can recover  $\tau$  making use of linear search governed by the condition in Equation 2.6.

### 3 Descriptor Transformation

Once the ROIs are at hand, image descriptors can be recovered. As mentioned earlier, recovering an optimal descriptor transformation can potentially improve the performance of the image classification task. Descriptor transformation is however, a non-trivial task. In the proposed method, we view the descriptor transformation task as a regularised risk minimisation problem. In this manner, the optimal transformation of the image descriptors under study can be recovered via the optimisation of the risk function.

#### 3.1 Image Descriptors

Let the set of images in the database under study be  $\Gamma = \bigcup_{j=1}^K \gamma_j$ , where  $\gamma_j$  is the subset of images comprising the  $j^{th}$  image category. Similarly, consider the codeword  $\Phi_i$  corresponding to the  $i^{th}$  image in  $\Gamma$ . We note that each of

the codewords can be further expressed in terms of the image descriptors themselves. For the sake of simplicity, we view these codewords as a vector corresponding to the concatenation  $X_j$  of those descriptors corresponding to each of the images in the data set, i.e.  $\Phi_i = \{\phi_1, \phi_2, \dots, \phi_{|\Phi_i|}\}$ , where  $\phi_k$  is the  $k^{th}$  descriptor under study.

Moreover, we can view these descriptors as a field subject to a transformation  $\mathcal{T} : X_i \mapsto Y_i$ , where  $Y_i$  is the vector comprised by the optimisation of image descriptors for the image indexed  $i$ . In practice,  $\mathcal{T}$  is a family of matrices such that  $Y_i = \mathcal{T}X_i$ . Thus, with these ingredients, we can cast the descriptor combination problem in a risk minimisation setting, in which the aim of computation is the recovery of the optimal transformation matrix  $\mathcal{T}$ .

#### 3.2 Learning

In this section, we describe how the optimal transformation matrix for the image descriptors under study can be recovered using risk minimisation. Regularised risk minimisation [24] was first proposed so as to build large margin classifiers such as the support vector machine (SVM) [3]. In the classification setting, an optimal classifier is recovered through the minimisation of the following general purpose risk functional

$$\min g(f) = \|f\|_{\mathcal{H}}^2 + \sum_i L(\mathbf{x}_i, y_i, f) \quad (3.1)$$

In the above equation,  $f$  is a smooth discriminant function in the Hilbert space,  $\|\cdot\|_{\mathcal{H}}$  is the norm function, which is related to the complexity of the classifier, and  $L(\mathbf{x}_i, y_i, f)$  is the cost of applying the classifier  $f$  to the sample-label tuple  $(\mathbf{x}_i, y_i)$ .

Note that the first term on the right-hand-side of the above equation is the regularisation term, which favors classifiers with lower complexity. The second term is related to the classifier error over the training data set. Different choices of the norm and the loss function lead to different versions of the classification technique. For example, if the  $L_2$  norm and the Hinge loss are used, the above formulation defines the popular SVM classifier [3]. Alternatively, if the  $L_2$  squared loss is used, it is then equivalent to regularised least squares (RLS) [23]. We adopt the RLS version of risk minimisation framework due to both, its simplicity and its ability to handle multiple category settings without loss of generality.

For purposes of learning optimal descriptor transformation, we used the same formulation as in Equation 3.1. Let  $\mu_w^* = E(\cup_{i \in w} \Phi_i | f(\Phi_i) = w \forall i \in w)$  be the class conditional mean of the descriptor variables from the class  $w$ , i.e. the mean for the descriptors that have been correctly classified by the current classifier, where  $\Phi_i$  is the concatenated feature vector for all the descriptors as defined in the

previous section, and  $f(\Phi_i)$  denotes the predicted label values for  $\Phi_i$ . Note that the purpose of the transformation task is to achieve a more compact representation of the descriptors by mapping descriptor vectors from the same class to the class representative given by the class conditional mean such that the within-class scatter is minimised. Hence, we can view  $\Phi_i$  as the input and its corresponding  $\mu_w^*$  as the target variable. The input and the target variables are then related by a linear transformation  $\mathcal{T}\Phi_i = \mu_w^*$ . This leads to the following formulation for learning the feature transformation based on a slight modification of the regularised risk minimisation framework in Equation 3.1

$$\min g(\mathcal{T}) = \lambda \|\mathcal{T}\|^2 + \sum_{w \in \Omega} \sum_{i \in w} \|\mathcal{T}\Phi_i - \mu_w^*\|^2 \quad (3.2)$$

where  $\Omega$  is the set of class categories in the dataset. Again, the first term on the right-hand-side is the regularisation term that minimises the  $L_2$  norm of the transformation matrix. The second term is the data term that enforces the similarity between the transformed descriptor and the class conditional mean feature vector.

Note that the cost function defined in Equation 3.2 can be solved directly. The closed form solution of  $\mathcal{T}$  is given by

$$\mathcal{T} = \mathbf{Y}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T + \lambda\mathbf{I})^{-1} \quad (3.3)$$

where  $\mathbf{X}$  is a  $N \times d$  matrix with input descriptor feature vectors stacked in columns,  $\mathbf{Y}$  is a matrix of the same dimension with  $\mu_{l(\phi_i)}^*$  stacked in columns corresponding to each input descriptor and  $\mathbf{I}$  is an  $N$ -dimensional identity matrix.

It is worth noting that, we can use the  $L_1$  loss function as an alternative to the  $L_2$  norm in Equation 3.2. Using the  $L_1$  norm for the regularisation term encourages the sparseness of the recovered transformation matrices and minimises the number of nonzero entries. Further, the use of an  $L_1$  loss function for the data term is believed to be more robust to the outliers in the input data [9]. By pursuing the use of the  $L_1$  norm, the cost function can be transformed to the following linear program

$$\begin{aligned} \min g_{L_1}(\mathcal{T}, \Delta, \epsilon) &= \lambda \sum_{i=1}^N \sum_{j=1}^N \delta_{i,j} + \sum_{i=1}^N \sum_{j=1}^d \epsilon_{i,j} \quad (3.4) \\ \text{s.t. } & -\delta_{i,j} \leq \mathcal{T}_{i,j} \leq \delta_{i,j} \quad \text{and} \quad \delta_{i,j} \geq 0 \\ & -\epsilon_{i,j} \leq \sum_r \mathcal{T}_{i,r} \phi_{r,j} - \mu_{l(\phi_i)}^* \leq \epsilon_{i,j} \quad \text{and} \quad \epsilon_{i,j} \geq 0 \end{aligned}$$

where  $\delta_{i,j}^{(l)}$ 's and  $\epsilon_{i,j}^{(l)}$ 's are slack variables defining the upper bound of the regulariser  $\|\cdot\|_{\mathcal{H}}^2$  and the cost  $L(\mathbf{x}_i, y_i, f)$  in Equation 3.1, respectively.

## 4 Experimental Results

In this section, we present the experimental results for our algorithm. For purposes of image categorisation, we have used the Oxford Flower database [16]. This database contains flower images for 17 species. Images in the dataset are divided into a training set of 680 images, a validation set of 340 images, and a testing set of 340 images. Samples of images from the database is displayed in Figure 1.

Here, we compare the performance of the proposed method with those yielded by the alternatives in [16] and [25], which employ the same set of image descriptors as we do. Note that, in our experiments, we have used unsegmented images and saliency as an alternative to the segmented flower database used in [16] and [25]. In [16], feature transformation was treated as finding the optimal weights for individual features when they are combined into the codebook, where the weights are treated equally across all classes and recovered through an intensive search on the validation set. The method of Varma and Ray [25] also treats the feature transformation as a feature combination problem. Their method recovers domain specific descriptors by learning the trade-off between invariance and discriminative power of the classification system. Additionally, we also provide comparison with the performance yielded by the Block 1-norm method (MKL-Block) [1, 25], which employs a multiple kernel learning approach based upon regularisation.

Following the sections above, we commence by applying the saliency map extraction algorithm to each of the images in the data set. Once the saliency maps are at hand, we compute the optimal cut-off  $\tau$  for each image. The ROIs used for classification are given by the binarisation of the saliency maps. Examples of saliency maps and their corresponding ROIs are displayed in Figure 2.

A set of features are extracted from the ROIs. The descriptors used in our experiments are the SIFT [14], colour and the texture features from the MR-filter bank in [21]. Following Nilsback and Zisserman [16], we have recovered visual words by clustering the descriptor-set using the K-means algorithm. The visual word for each image is given by the frequency histogram for the descriptors. This is done such that the bin-centres for the histogram correspond to the cluster-centres recovered by the K-means algorithm. Finally, the extracted visual words are concatenated so as to be used as input to an SVM classifier. We have tuned the SVM parameters via cross validation for optimum performance of the classifiers.

Here, we have trained the SVM classifier on the training set. The feature transformation process commences by computing the classification results on the training set using the SVM classifiers. The results yield an expected feature vector from correctly classified images for each image



Figure 1. Sample images for the Oxford flowers database.



Figure 2. Saliency map and ROI.

class, i.e. a  $K$ -dimensional vector specifying the class conditional mean for the images in the training data. These expected feature vectors and the input image descriptors are then used to recover the transformation matrix. The transformation matrix is multiplied to all original feature vectors, which are then used in the testing phase.

It is worth noting that we can either use the  $L_2$  regulariser in Equation 3.2 or the  $L_1$  norm in Equation 3.4 for the purpose of learning optimal transformation matrix. In practice, we found  $L_1$  regulariser outperforms  $L_2$  marginally. This is a reasonable result as  $L_1$  norm is less sensitive to outliers than  $L_2$  norm. With regard to the computational complexity, the  $L_2$  regulariser is much more computationally efficient than the  $L_1$  counterpart.

From Figure 2, it can be observed that the saliency detection algorithm is very effective in extracting the area where intensity, color and orientation changes. The resulting ROIs preserve the objects of interest. For instance, in the case of the flowers, they have been separated from the grassy background. Moreover, this is evidenced by the classification rates yielded by both, with and without saliency map extraction steps. In Table 1, we compare the image classification performance given saliency extraction steps, direct descriptor computation and their combinations into a concatenated

visual word. For all individual descriptors, as well as their combinations, the classification performance is greatly improved through the application of saliency map extraction and binarisation.

In terms of classification performance, we present results in Table 2. We implemented both the  $L_1$  and the  $L_2$  regularisers. Our method achieved correct classification rates of up to 78.5%. Nonetheless we use an unsegmented dataset, our method provides a margin of improvement over the MKL-Block method in [1, 25] using the  $L_1$  norm and is comparable to the other alternatives. These results suggest that, with our method, the human hand-labeling process can be omitted. Moreover, by providing comparable results to those yielded by the alternatives, we can conclude that the method can achieve classification rates similar to those obtained through intensive search on the validation set. This greatly reduces the amount of human intervention and permits optimally transforming image descriptors for image classification.

## 5 Conclusion

In this paper, we have proposed a method which hinges in the use of saliency feature extraction and risk minimi-

	C	S	T	C-S-T
Results using ROIs	51.5%	61.5%	40.3%	71.4%
Results on raw imagery (no ROIs used)	45.2%	58.2%	30.6%	62.6%

**Table 1. Performance on classification with and without saliency map extraction and ROI recovery. In the table, we show classification rates for color (C), SIFT (S), and texture (T) descriptors alone and their combination as a concatenated vector.**

Classification Method	Flower Dataset	Classification Rate
proposed method $L_1$	unsegmented	78.5%
proposed method $L_2$	unsegmented	77.1%
Nilsback [16]	segmented	81.3%
Bach [1, 25]	segmented	77.8%
Varma [25]	segmented	82.6%

**Table 2. Performance comparison.**

sation so as to optimally transform salient descriptors for image classification. The contributions are, hence, twofold. Firstly, we have built upon the saliency mapping method from Itti *et al.* [7] to recover ROIs suitable for image classification tasks. Secondly, we have proposed a means to optimally recover a transformation matrix so as to maximise the classification performance. The optimisation approach presented here is quite general in nature and can be applied to similar applications using different classifier configurations. We have presented results on a real world data set and compared our method against alternatives elsewhere in the literature.

## References

- [1] F. R. Bach, R. Thibaux, and M. I. Jordan. Computing regularization paths for learning multiple kernels. In *Neural Information Processing Systems Conference (NIPS)*, 2004.
- [2] A. Bosch, Z. A. and X. Muoz. Image classification using random forests and ferns. In *ICCV*, pages 1–8, 2007.
- [3] B. Boser, I. Guyon, and V. Vapnik. A training algorithm for optimal margin classifiers. In *ACM Conference on Computational Learning Theory*, 1992.
- [4] J. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two dimensional visual cortical filters. *Journal of the Optical Society of America*, 2(7):1160–1169, 1985.
- [5] L. Fei-Fei and P. Perona. A bayesian hierarchical model for learning natural scene categories. In *IEEE Conference on Computer Vision and Pattern Recognition (IEEE Conference on Computer Vision and Pattern Recognition (CVPR))*, pages II:524–531, 2005.
- [6] R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7:179–188, 1936.
- [7] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE TPAMI*, 20(11):1254–1259, 1998.
- [8] T. Kadir and M. Brady. Saliency, scale and image description. *International Journal of Computer Vision*, 45(2):83–105, 2001.
- [9] Q. Ke and T. Kanade. Robust l1 norm factorization in the presence of outliers and missing data by alternative convex programming. In *CVPR*, pages 739–746, 2005.
- [10] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2169–2178, 2006.
- [11] S. Lee, J. Moon, and M. Lee. A region of interest based image segmentation method using a biologically motivated selective attention model. In *International Joint Conference on Neural Networks*, pages 1413–1420, 2006.
- [12] B. Leibe and B. Schiele. Analyzing appearance and contour based methods for object categorization. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 409–415, 2003.
- [13] H. Liu and H. Motoda. Feature transformation and subset selection. *IEEE Transactions on Intelligent Systems and Their Applications*, 13(2):26–28, 1998.
- [14] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [15] F. Moosmann, B. Triggs, and F. Jurie. Fast discriminative visual codebooks using randomized clustering forests. In *Neural Information Processing Systems Conference (NIPS)*, pages 985–992, 2006.
- [16] M. E. Nilsback and A. Zisserman. A visual vocabulary for flower classification. In *IEEE Conference on Computer Vision and Pattern Recognition (IEEE Conference on Computer Vision and Pattern Recognition (CVPR))*, pages II:1447–1454, 2006.

- [17] N. Otsu. A thresholding selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.
- [18] C. Papageorgiou and T. Poggio. A trainable system for object detection. *International Journal of Computer Vision*, 38(1):15–33, 2004.
- [19] C. Privitera and L. Stark. Algorithms for defining visual regions-of-interest: Comparison with eye fixations. *IEEE TPAMI*, 22(9):970–982, 2000.
- [20] P. Quelhas, F. Monay, J. Odobez, D. Gatica-Perez, T. Tuytelaars, and L. V. Gool. Modelling scenes with local descriptors and latent aspects. In *IEEE International Conference on Computer Vision*, pages I:883–890, 2005.
- [21] J. Sivic, B. Russell, A. Efros, A. Zisserman, and W. Freeman. Discovering objects and their location in images. In *ICCV*, pages I:370–377, 2005.
- [22] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE TPAMI*, 22(12):1349–1380, 2000.
- [23] J. Suykens and J. Vandewalle. Least squares support vector machine classifiers. *Neural Processing Letters*, 9:293–300, 1999.
- [24] V. Vapnik. *Statistical Learning Theory*. John Wiley and Sons, 1998.
- [25] M. Varma and D. Ray. Learning the discriminative power-invariance trade-off. In *ICCV*, pages 1–8.
- [26] J. Zhou and A. Robles-Kelly. A quasi-random sampling approach to image retrieval. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2008.
- [27] M. Zhu and A. Martinez. Subclass discriminant analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(8):1274–1286, 2006.