SEMANTIC DISCRIMINANT MAPPING FOR CLASSIFICATION AND BROWSING OF REMOTE SENSING TEXTURES AND OBJECTS

Julien Fauqueur, Nick Kingsbury and Ryan Anderson

Signal Processing Group, Department of Engineering, University of Cambridge, UK

ABSTRACT

We present a new approach based on Discriminant Analysis to map a high dimensional image feature space onto a subspace which has the following advantages: 1. each dimension corresponds to a semantic likelihood, 2. an efficient and simple multiclass classifier is proposed and 3. it is low dimensional. This mapping is learnt from a given set of labeled images with a class groundtruth. In the new space a classifier is naturally derived which performs as well as a linear SVM. We will show that projecting images in this new space provides a database browsing tool which is meaningful to the user. Results are presented on a remote sensing database with eight classes, made available online. The output semantic space is a low dimensional feature space which opens perspectives for other recognition tasks.

1. INTRODUCTION

Recognising objects and textures based on their visual appearance is a challenging task because the correspondence between the visual features of an image and its associated meaning is often ambiguous.

In the context of remote sensing imagery huge amounts of data are produced for various purposes including mosaicing, terrain classification, or the detection of changes, anomalies and manmade structures. We focus on the problem of recognition of various types of manmade entities as well as different types of vegetation. This problem presents various challenges due to the high heterogeneity both within and across classes. The within class heterogeneity is due to the difference of acquisition process, orientation, and intrinsic appearance. On the other hand some classes, as we will see, can be very similar (such as grass and fields) while some others are of different nature: the vegetation classes tend to relate to textures while the manmade ones relate to objects. In this context, choosing an adequate set of features to discriminate between these classes is difficult. Indeed our intuition would suggest that different types of features may

be useful in different cases: for example, structure or color features to discriminate manmade versus vegetation, color for grass versus river, texture for grass versus trees. This relates to the feature selection problem. Based on discriminant analysis, the presented method handles the problem of finding the optimal weighting of the feature components to discriminate each class against the others.

Let us consider a set of images manually assigned to C classes of interest. We introduce a mapping, referred to as **Semantic Discriminant Mapping** (SDM), from the image feature space into a C-dimensional subspace referred to as the semantic space. In the semantic space, each of the C dimensions discriminates between a given class and the others. Along each dimension a one-class predictor for the corresponding class will be built and the combination of the predictors will result in a multi-class classifier. The component values on those semantic dimensions give a classification score and can be used to display an image database content with the semantic directions of interest.

In section 2 we give an overview of all the image features used. The SDM is presented in section 3. In section 4 we will introduce the multiclass classifier which operates in the semantic space. Results are presented in section 5, and conclusions in section 6.

2. COLOR, TEXTURE AND STRUCTURE FEATURES

Since the nature of the images we want to classify is very heterogeneous, various types of features must be integrated to achieve a good recognition rate. We therefore build a big feature vector to characterise the image content which consists of the concatenation of different color and texture features. Since our method can use any type of feature vector, we give only an overview of the features involved. Our texture and structure descriptors are based on the Dual Tree Complex Wavelet Transform (DTCWT) [1] computed on four levels. They involve central moments, histograms and a new feature, the Inter-Level Product (ILP) described in [2]. The entire feature vector description is comprised as follows with each subfeature dimension:

- *color*: mean HSV value (3), mean RGB value (3), RGB-pixel variance (3), $6 \times 6 \times 6$ uniformly quantised RGB color histogram (216), entropy of the RGB histogram (1)

- *texture*: DTCWT magnitude mean, variance, skew, kurtosis, and directional statistics (36 for all 4 levels)

- *structure*: 4-bin ILP magnitude histogram (16 for all 4 levels), mean, variance, skew and kurtosis of the 4-bin ILP magnitude histogram (16 for all 4 levels), 4-bin ILP histogram of generic structures (edges, ridges, ...) (16 for all 4 levels)

The total dimension of the feature vector is $p = 310$. This original feature space is denoted as $\mathcal{F}$. Each dimension is normalised so that the mean over the data set is zero and the standard deviation is one.

3. SEMANTIC DISCRIMINANT MAPPING

The feature space $\mathcal{F}$ contains various subfeatures which we thought relevant for our remote sensing recognition problem. In this context, Discriminant Analysis [3] is a simple and powerful technique to project data into a reduced dimensional space in which the data are optimally separated given a set of labeled images. The implicit effect of the transformation is to assign various weights to each feature dimension depending on their relevance to discriminate each class.

However a matrix singularity problem can arise in the discriminant projection if the number of samples is lower than the feature dimension and/or if some feature dimensions have near-zero variance (typically histogram bins). To circumvent this drawback Swets and Weng [4] proposed projection of the data with Principal Component Analysis (PCA) before discriminant projection. Although our discriminant projection differs from theirs we use their idea of performing a prior PCA projection into an intermediate space, denoted as $\mathcal{F}_{PCA}$. In this section we give a brief overview of PCA and Fisher Discriminant Analysis (FDA) transformations and then present our C-FDA transformation to perform multiclass discriminant analysis. The SDM is introduced as combination of PCA and C-FDA transforms.

3.1. PCA and FDA transformations

Given a set of p-dimensional feature vectors $\{x_i\}$ in $\mathcal{F}$, Principal Component Analysis finds a subspace $\mathcal{F}_{PCA}$ of dimension $p' \leq p$ whose basis vectors correspond to the maximum variance directions of $X = [x_1...x_p]$. If we call M_P the $p \times p'$ matrix of the linear transformation from $\mathcal{F}$ to $\mathcal{F}_{PCA}$, then the projection of x into this new space is $M_P^t x$. The columns of transformation matrix $M_P = [e_1...e_p]$ consist of the eigenvectors e_i obtained from the eigenvalue decomposition $\lambda_i e_i = (XX^t)e_i$, where XX^t is the covari-

ance matrix and λ_i the eigenvalue associated with the eigenvector e_i. The λ_i are sorted in decreasing order such that e_1 corresponds the direction of maximum variance of X and e_p to the one of minimum variance. We choose to keep the first p' dimensions such that $\sum_{i=1}^{p'} \lambda_i \geq 0.9999 \sum_{i=1}^{p} \lambda_i$. We denote the PCA transformation as $M_P : \mathcal{F} \mapsto \mathcal{F}_{PCA}$.

Unlike PCA, which finds the directions that capture most of the feature variance, FDA finds the direction w that best separate the features [3]. In the two class case, we assume the features x belong either to the class $\mathcal{C}$ or $\mathcal{C}'$. An optimal vector w must maximise the following criterion J:

$$J(w) = \frac{\mid w^t S_B w \mid}{\mid w^t S_W w \mid} \tag{1}$$

where $\mid . \mid$ denotes the matrix determinant, S_B the between-class scatter matrix and S_W the within-class scatter matrix defined as follows:

$$S_B = (m - m')(m - m')^t \tag{2}$$

$$S_W = \operatorname*{Mean}_{x \in \mathcal{C}}(x-m)(x-m)^t + \operatorname*{Mean}_{x \in \mathcal{C}'}(x-m')(x-m')^t \tag{3}$$

where $m = \operatorname*{Mean}_{x \in \mathcal{C}}[x]$ and $m' = \operatorname*{Mean}_{x \in \mathcal{C}'}[x]$ are the respective means of features in classes $\mathcal{C}$ and $\mathcal{C}'$. A solution of this maximisation problem is given by:

$$w = S_W^{-1}(m - m') \tag{4}$$

In this case, FDA builds a mapping from the feature space to a 1-dimensional subspace.

3.2. The C-FDA transformation

We introduce a new transformation, called C-FDA, to perform discriminant analysis in the multiclass case and which, unlike Multiple Discriminant Analysis (MDA)[3][1], determines a new feature space in which each dimension corresponds to a class membership likelihood. C-FDA consists in performing an FDA C times to discriminate between each class $\mathcal{C}_d$ and the other classes. For each class $\mathcal{C}_d$, $\mathcal{C}$ designates $\mathcal{C}_d$ and $\mathcal{C}'$ designates $\cup_{d' \neq d} \mathcal{C}'_d$ with section 3.1 notations. Thus for each class $\mathcal{C}_d$ we get from formula (4) a vector w_d defining the direction of optimal separation between $\mathcal{C}_d$ and the other classes. In our approach, since the features are first transformed with PCA, C-FDA is applied in $\mathcal{F}_{PCA}$ of dimension p'. So the C-FDA transformation is defined by the $p' \times C$ matrix $M_F = [w_1...w_C]$. The output space $\mathcal{S}$ is C-dimensional and called the **semantic space**. The transformation of a vector y in $\mathcal{F}_{PCA}$ is then written $M_F^t y$ in $\mathcal{S}$.

---

[1]Although MDA deals with multiclass, it is not suitable for our problem since the dimensions in the transformed space are not directly associated with a class. MDA finds a linear transformation from $\mathcal{F}$ to a $C-1$ dimensional subspace which optimally separates the data *globally*.

For classification and browsing purposes we find it convenient to normalise the transformed values in $\mathcal{S}$ such that :

$$\underset{y \in \mathcal{C}_d}{\text{Mean}}[w_d^t y] = +1 \text{ and } \underset{y \notin \mathcal{C}_d}{\text{Mean}}[w_d^t y] = -1 \qquad (5)$$

This is achieved by applying a linear transformation, denoted as T_N, from $\mathcal{S}$ to itself which is straightforward to compute. T_N is calculated on the training data.

3.3. Semantic Discriminant Mapping

Now we define the Semantic Discriminant Mapping (SDM), denoted as s. It is the full linear transformation from the original feature space $\mathcal{F}$ onto the semantic space $\mathcal{S}$. It is the composition of the following transformations: PCA transformation, the discriminant transformation and the normalisation. Its expression is the following:

$$\begin{aligned} s : \mathcal{F} &\mapsto \mathcal{S} \\ x &\mapsto T_N(M_F^t \cdot M_P^t \cdot x) \end{aligned} \qquad (6)$$

where the original feature space $\mathcal{F}$ is of dimension p and the output semantic space $\mathcal{S}$ is of dimension C.

4. C-CLASS CLASSIFIER IN SEMANTIC SPACE

The multiclass classifier is built as a combination of scalar predictors in the semantic space $\mathcal{S}$. For each dimension d in $\mathcal{S}$, training data are optimally separated in the linear sense depending on whether they belong or not to class $\mathcal{C}_d$. We denote as $s_d(x)$, the dth component of the transform of x in $\mathcal{S}$. Given formula (5), if $s_d(x)$ is big then x is very likely to belong to $\mathcal{C}_d$. An image will be predicted in $\mathcal{C}_d$ if its feature x satisfies $s_d(x) \geq t_d$ where t_d is a threshold. We want to find the optimal thresholds $\{t_d\}$ which minimise the classification error $\epsilon_d(t)$. We define $\epsilon_d(t)$ as the sum of probabilities of detection of false negatives and false positives:

$$\begin{aligned} \epsilon_d(t) = &P(s_d(x) \geq t \mid x \notin \mathcal{C}_d) + \\ &P(s_d(x) < t \mid x \in \mathcal{C}_d) \end{aligned} \qquad (7)$$

To determine $\epsilon_d(t)$, training data are sorted with respect to their $s_d(x)$ values. For each x, the two probability terms in 7 are computed and the optimal threshold t_d is set at the value of $s_d(x)$ which minimises $\epsilon_d(s_d(x))$.

The multiclass classifier f is defined by considering the predictor with the highest classification score measured by the quantity $s_d(x) - t_d$. For each x in $\mathcal{F}$, f is defined by:

$$f(x) = \underset{d=1,...,C}{\text{argmax}}[s_d(x) - t_d] \qquad (8)$$

5. RESULTS

In this section we detail the procedure to build our database for training and testing. We will then show the classification results and present a browsing application.

5.1. Groundtruth database construction

Our 1040 aerial image database was constructed from seven large aerial images from Window on the UK[2]. We extracted 14575 64×64 pixel subimages which were manually assigned to one of the following eight classes: building, road, river, field, grass, tree, boat, vehicle (see figure 1). In order to have classes of the same size, we kept 130 images per class yielding 1040 images total. This image set was then split into a training set and a test set, consisting of 65 images per class. So both training and test database contain 520 images. We have made this database available online[3].

Fig. 1. Each column illustrates samples from a class. From left to right: building, road, river, field, grass, tree, boat, vehicle.

In the presented results, the SDM was determined on the *training* images. Classification and browsing experiments are conducted on the *test* database after transformation with the SDM.

5.2. Classification

The 520 images from the test database were classified in the semantic space as described in section 4 (with $p' = 154$). The classification performance is illustrated in table 1 with a confusion matrix on the eight classes. Since each class contains 65 images, a diagonal value of 65 in the matrix indicates a perfect classification for this class. From this matrix the average classification accuracy across all eight classes is 89.4% for our method. In the confusion matrix, we see that the confusion errors made by our classifier are consistent with our perception. Indeed, the highest errors (7 and 8) occur in the discrimination between perceptually ambiguous classes: grass versus field and vehicle versus building. Indeed field images are very similar to grass images except they have a slightly perceptible directional texture. On the other hand buildings and vehicles both relate to rectangular objects with heterogeneous colors when using low-level features. Another encouraging observation is that manmade classes (building, road, vehicle, boat) are rarely confused with natural classes (grass, river, tree, field) which is a desired property in remote sensing analysis.

[2] http://www.bnsc.org
[3] http://www.eng.cam.ac.uk/~jf330/GTDB/

class labels	bu.	r.	g.	r.	t.	f.	v.	bo.
building	52	4	0	0	1	0	4	4
road	3	57	0	0	1	0	4	0
grass	0	0	55	2	1	7	0	0
river	0	0	1	64	0	0	0	0
tree	2	0	0	0	63	0	0	0
field	0	0	4	0	0	61	0	0
vehicle	8	3	0	0	1	0	52	1
boat	1	1	0	1	0	0	1	61

Table 1. Confusion matrix for Semantic Discriminant Mapping classifier. A value of 65 on the diagonal corresponds to 100% correct classification.

To compare our classification method to a state of the art method we trained and tested a multiclass linear Support Vector Machine (SVM) on the same image sets using libsvm toolbox[4]. For the linear SVM the average accuracy across classes was 89.2%. In the same conditions, we also trained and tested a non-linear SVM using a gaussian kernel. The optimal hyperparameters of this classifier were found by grid-search. The average accuracy was 92.3% across classes. Note that our classifier works in the semantic space while both SVMs were tested in the original feature space. Our classifier has an equivalent performance to the linear SVM (89.4% versus 89.2%) but is outperformed by the gaussian SVM (92.3%). The performance gain obtained by the use of a non-linear kernel for an SVM motivates the future integration of a non-linear kernel in our approach.

5.3. Browsing

Unlike low-level feature dimensions, the individual dimensions of the semantic space $\mathcal{S}$ are meaningful to the user and are suitable for database browsing and visualisation. Depending on the user-selected axes of visualisation, images with the same semantics are grouped in the same region of the browsing space. The applications of this browsing tool could be to provide the overview of a database or to show the user the top-ranked results of an image search engine.

Visualisation can be performed in any 1-, 2- or 3- dimensional subspace of the C-dimensional semantic space. We found the 2-dimensional representation to be the most effective. In figure 2, we show the entire content of the 520 image test database along two of the eight axes: the road axis (horizontal) and the tree axis (vertical). The axes are centered at the threshold values (t_{road}, t_{tree}). Images corresponding to trees are on the upper part of the plane and images of roads on the right part. Images which neither correspond to tree or road lie in the lower left part. Note that the high classification performance in the semantic space reported in the previous section guarantees the consistency of the visualisation. Screenshots of projections along all axes are available online [5].
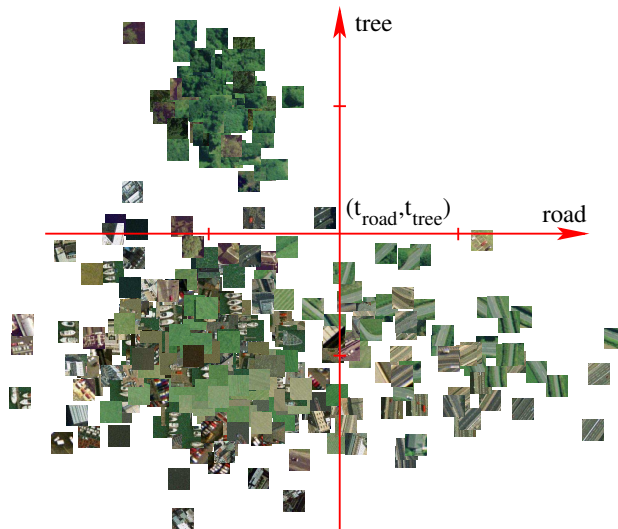

Fig. 2. The 520 test images are displayed in the road×tree subspace of the semantic space (see text).

6. CONCLUSION AND PERSPECTIVES

We presented the Semantic Discriminant Mapping which maps a feature space into a low dimensional subspace (the semantic space). In this new feature space a very simple multiclass classifier was proposed with equivalent performance to an SVM. Since each image coordinate corresponds to a class membership likelihood, a large set of images can be visualised in this space in a meaningful way for the user.

Futhermore if we view this semantic space as a new low dimensional feature space in which data are optimally separated, it should be suitable for other image analysis problems such as image retrieval. We are investigating the integration of a non-linear kernel in our SDM.

7. REFERENCES

[1] N.G. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *J. of Applied and Computational Harmonic Analysis*, vol. 10, 2001.

[2] R. Anderson, N.G. Kingsbury, and J. Fauqueur, "Coarse-level object recognition using interlevel products of complex wavelets," *IEEE ICIP*, 2005.

[3] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2nd ed.)*, Wiley, 2000.

[4] D. L. Swets and J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Trans. on PAMI*, vol. 18, no. 8, pp. 831–836, 1996.

---

[4]libsvm toolbox: http://www.csie.ntu.edu.tw/~cjlin/libsvm/

[5]http://www.eng.cam.ac.uk/~jf330/SDM/2Dshots/