Classification of Biological Objects using Active Appearance Modelling and Color Cooccurrence Matrices

Anders Bjorholm Dahl[†‡], Henrik Aanæs[†], Rasmus Larsen[†], and Bjarne Kjær Ersbøll[†]

abd@imm.dtu.dk

[†] Informatics and Mathematical Modelling, Technical University of Denmark
[‡] Dralle A/S - Cognitive Systems, Copenhagen, Denmark (www.dralle.dk)

Abstract. We use the popular active appearance models (AAM) for extracting discriminative features from images of biological objects. The relevant discriminative features are combined principal component (PCA) vectors from the AAM and texture features from cooccurrence matrices. Texture features are extracted by extending the AAM's with a textural warp guided by the AAM shape. Based on this, texture cooccurrence features are calculated. We use the different features for classifying the biological objects to species using standard classifiers, and we show that even though the objects are highly variant, the AAM's are well suited for extracting relevant features, thus obtaining good classification results. Classification is conducted on two real data sets, one containing various vegetables and one containing different species of wood logs.

1 Introduction

Object recognition is one of the fundamental problems in computer visions, and plays a vital role in constructing 'intelligent' machines. Our initial motivation for this work is the construction of an automated forestry system, which needs to keep track of wood logs. Many of the objects in our daily environment in general, and in our motivating problem in particular, are biological, and pose special challenges to a computer vision system. The origin of these challenges are the high degree of intraclass variation, which we as humans are very good at dealing with, e.g. consider the multitudes of ways a face or a potato can look. To enable biological variation to be handled in a classification system, we have to find methods for extracting discriminative features, from the depicted objects. AAM's have proven very well suited for addressing the challenge of handling biological variation in the case of image registration, cf. [4]. It is thus highly interesting if this property of the AAM's also proves well for classification of objects, and how this should be implemented. Therefore, we have investigated AAM's for extracting discriminative features by conducting the following experiments:

1. Classification based on **Multiple AAM's**, i.e. building an AAM for each class and assigning images of unknown objects to the most probable AAM.

2. Classification based on a **global AAM**, i.e. building one single AAM and using model parameters for assigning images of unknown objects to the most probable class.
3. Identify relevant discriminative patches from the use of an AAM. The object is identified by shape alignment from the **AAM and texture** is extracted and used for second order texture analysis.

Two data sets have been investigated in this paper, one containing vegetables and one containing wood logs. Experiment 1 and 2 have been conducted for both data sets and experiment 3 has been conducted only for the wood log data set.

1.1 Related work

The environment plays a vital role in solving object recognition problems. In a natural environment objects may be seen from many different angles, they may be occluded, light may change etc. Efforts on solving this type of problem have been put in identifying local object features invariant to the changing conditions, cf. e.g. [16, 14, 13], and the way to match these features to a model, cf. e.g. [5, 6].

Controlling the environment in some way, gives the opportunity of easing the flexibility constraints of the object recognition system. In some situation object recognition on whole objects is a reasonable approach, giving the option of e.g. extracting global PCA features. This is done for face recognition by e.g. Turk & Pentland [20] with the eigenface, and Belhumeuer *et al.* [1] for their fisherface based on Fishers Linear Discriminant Analysis. No shape information is included with these methods. Edwards *et al.* [7] introduces the use of an AAM for face recognition based on both shape and texture information. Fagertun *et al.* [9] improves this method by the use of canonical discriminant analysis. AAM's have been used for related recognition problems, e.g. eye tracking by Stegmann [18] and Hansen *et al.* [10].

Pure texture has also been used for object recognition. The second order texture statistics based on cooccurrence matrices, was originally developed by Haralick *et al.* [11] for pixel classification. This method has been extended to object recognition by e.g. Chang & Krumm [3] using color cooccurrence histograms. Palm [15] does classification of different textures, including wood bark textures, using color cooccurrence matrices. He extends from gray level to color images and improves the classification.

In this paper we focus on object recognition in an environment with some degree of controlled conditions. We use a black background, controlled lighting, and we make sure that the whole object is visible.

2 AAM and texture features

In the following we describe the methods for extracting the discriminative features used in the three experiments.

2.1 AAM

The AAM model - in 2D as it will be used here - is a description of an object in an image via it's contour or shape and it's texture. Each of these entities can be represented as a vector, i.e. $\mathbf{s}_i$ and $\mathbf{t}_i$ respectively, where the subscript, i, denotes an instance (image). The parameters of the AAM is, however, a lower dimensional vector, $\mathbf{c}_i$, and a specific AAM consists of an linear mapping for $\mathbf{c}_i$ to $\mathbf{s}_i$ and $\mathbf{t}_i$, i.e.

$$\mathbf{m}_i = \begin{bmatrix} \mathbf{Ws} \\ \mathbf{t} \end{bmatrix}_i = \mathbf{\Phi c}_i \ , \tag{1}$$

where $\mathbf{\Phi}$ is a matrix representing the linear map. The AAM or $\mathbf{\Phi}$ is estimated from an annotated training set. By optimizing the AAM to a new depicted object, an image close to the original is synthesized, and the model parameters $\mathbf{c}_i$ is a vector of principal components describing the unknown object with regards to the shape and texture of the object. The interested reader is referred to Cootes and Taylor [4] for a more detailed description and Stegmann *et al.* [19] for a detailed description of the model implementation.

Features from multiple AAM's In this case an AAM, $\mathbf{\Phi}_j$, is fitted to each class $\mathcal{C}_j$, i.e. the training set is divided into its component classes, and one AAM is fitted to each.

Here there is a feature vector $\mathbf{c}_i$ i specific for each model, and these features are not comparable, because they belong to a specific model and can not be used directly for classification.

Given an AAM for each class $\mathcal{C}_j$, you would expect the optimization of an image i to perform best for the class that the object belongs to. Therefore, a goodness of fit would be a reasonable measure for classifying the object. For a given unknown object image textural difference between the object texture $g_{i_{obj}}$ and the model instance $\bar{g}_{mod}$ is calculated:

$$E = \sum_{i=1}^{n} (g_{i_{obj}} - \bar{g}_{mod})^2 = ||g_{i_{obj}} - \bar{g}_{mod}||_2^2, \tag{2}$$

where E is the difference between the model image and the measured image by the squared 2-norm.

Features from a global AAM In this case a single global AAM, $\mathbf{\Phi}$, is fitted to instances of all classes. Following this, the $\mathbf{c}_i$ are calculated for each instance in the training set. The elements of $\mathbf{c}_i$, containing both shape and texture information, are used in a linear classifier, see section 2.3.

Textural warp The basis for making a textural warp is knowledge of the log localization in the image. This comes from the AAM shape alignment. The warp is done by sampling pixels along elliptic curves on the surface of the logs using

bilinear interpolation, see Figure 1. The elliptic curves are calculated from the shape of the end face of the wood log, and guided by the shape of the sides of the log. A square image of the bark texture is the result of the warp, which is illustrated in Figure 2. One end of the log is usually smaller than the other, resulting in a difference in the sampling intensity in the warped bark image. Other shape variations may result in the same kind of sampling variation. These small variations have not been considered as a problem for the texture analysis.
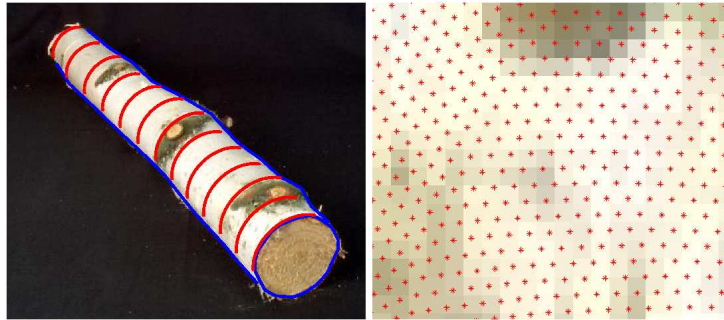

Fig. 1. Illustration of bark texture warp. Left is an image of a Birch log shown with a few of the elliptic sampling curves shown in red. Blue lines show the AAM shape alignment. The right image is a close up of sampling points.

2.2 Color cooccurrence matrices

As mentioned above, the AAM clasification, is extended by texture clasification, where the texture is obtained via texture warp as described in Section 2.1. This classification is done via second order textural statistics in the form of cooccurrence matrices (CM) cf. [2, 11, 15]. The fundamental element of calculating CM's is the probability of a certain change in pixel intensity classes (k, l) given a certain pixel displacement $\mathbf{h}$ equivalent to $Pr(k, l|\mathbf{h})$. The CM's can be extended to color, by calculating the displacements in each band and across bands. The CM's have proven useful for classification cf. [15]. Sample CM's for the relevant bark textures is shown in Figure 2. In this paper the textures have been preprocessed by Gaussian histogram matching, in order to increase robustness to lighting conditions, cf. [2]. In this paper we use the following CM classes: contrast, dissimilarity, homogeneity, energy, entropy, maximum, correlation, diagonal moment, sum average, sum entropy, and sum variance.

2.3 Classifier

The classifiers used in experiments 1 to 3 are as follows:
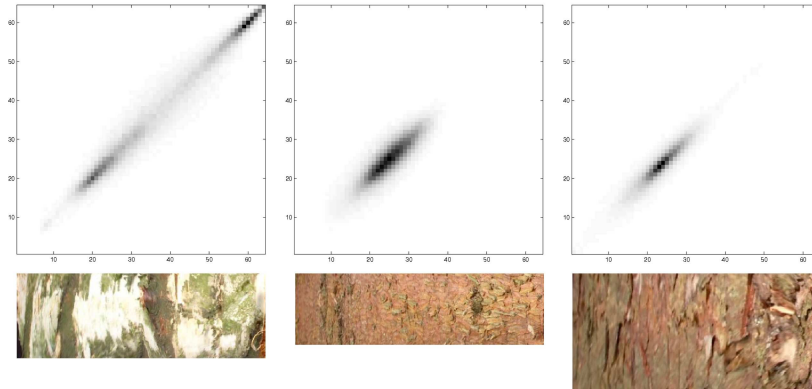
Fig. 2. Illustration of cooccurrence matrix (*top*) of bark texture (*bottom*). Higher intensity illustrates larger cooccurrence. From left to right: Birch, Spruce, and Pine. The displacement is $(1, 1)$ in a 64 level image.

1. **Multiple AAM's.** The model minimizing (2) is chosen.
2. **Global AAM.** Here three different classifications schemes based on the AAM feature vector are evaluated. These are: Bayes classifier, Canonical discriminant analysis, and LARS-EN classifier cf. [8, 12, 21].
3. **AAM and Texture.** Here LARS-EN is applied to the texture, obtained via the AAM based warp described in Section 2.1.

3 Data

Experiments were conducted for two groups of biological objects: vegetables, cf. Figure 3 and wood logs cf. Figure 1. The vegetables are are apples, carrots and potatoes and consist of 189 images totally where 27 are used for training the models, i.e. 9 from each group. The wood log data consists of the three species Scotch Pine, Norway Spruce and Birch. There was a total of 164 wood log images, 18 from each group were used for training. Also a reduced wood log data set, consisting of the 30 most characteristic logs from each group (90 in all) was used.
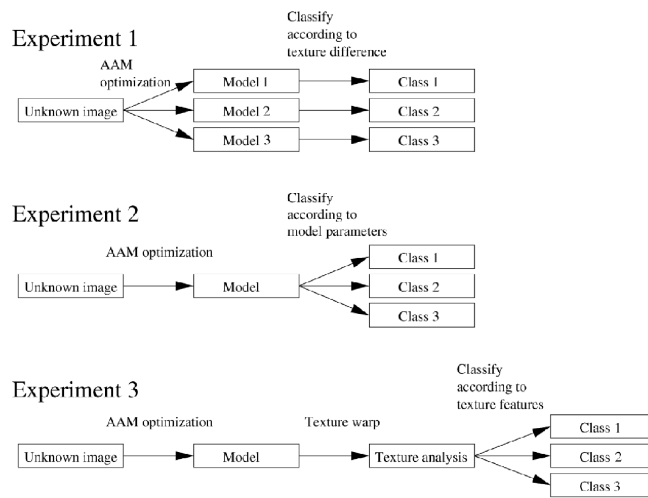
4 Experiments

All three experiments are illustrated in Figure 4. The procedure is as follows.

Experiment 1 For each class there is built an AAM based on the training images. All models are matched to each of the test images giving model textures for all classes. The model texture is then compared to the original image by calculating the texture difference, see section 2.1. Classification is done by assigning the test image in question to the model giving the least texture difference.
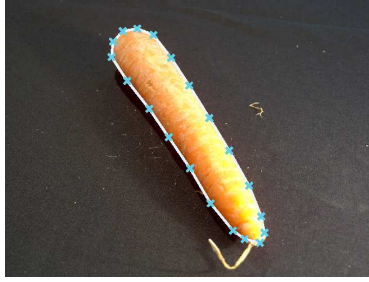
Fig. 3. Illustration of AAM alignment. The light blue line illustrates the shape and the blue crosses mark the annotated points.

Fig. 4. Schematic representation of the three experiments.

Experiment 2 In this experiment one AAM is built based on training images from all classes. The parameters from matching the model to a test image are used for classification. Based on these parameters the image is assigned to the most probable class as described in section 2.3.

Experiment 3 As in experiment 2, one AAM is matched to a test image, but here the alignment is used for extracting texture features. To enable a calculation of cooccurrence features from the bark texture, a warp of the bark area of the image is conducted.

4.1 Results and discussion

Results of our three experiments are presented in table 1 and 2.

Experiment 1 This experiment gave rather stable and good results. In the vegetable data only one image of a potato was misclassified as an apple, giving an average classification rate of 99.3%. The wood log data set gave stable classifications around 83% except for the whole log model, where many Spruce logs were misclassified as Pine logs and visa versa.

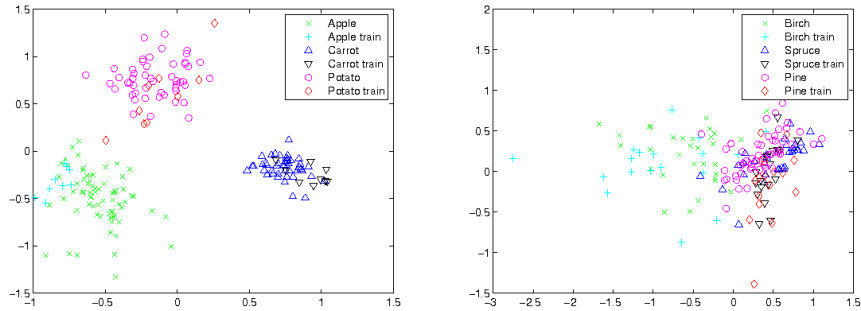| Experiment	1	2		
Model	Texture difference	Bayes	Canonical	LARS-EN
Vegetable	99.3%	100.0%	93.7%	100.0%
Log end	82.8%	70.2%	71.0%	75.5%
Large images	83.5%	64.1%	48.3%	82.1%
Whole log	67.8%	72.8%	71.1%	72.8%
Log end, reduced	82.5%	71.4%	38.1%	85.7%

Table 1. Classification rates of experiment 1 and 2 using the different classifiers. In the Vegetable and Whole log experiments, the shape covers the entire object, whereas in the rest, only the end part of the object is covered. Large images refer to the use of higher resolution images. Reduced refers to the reduced data set.

Experiment 2 For this experiment three different classifiers have been tried. The vegetable experiment obtains 100% correct classification with the Bayes classifier and LARS-EN but the canonical discriminant analysis gives only a classification rate of 93.7%. The canonical discriminant analysis is also very unstable for the wood log data set. The Bayes classifier gives around 70% correct classification and LARS-EN around 80%.

AAM results in a relatively large number of features, and therefore, it is necessary to have many observations for training a classifier. A limited number of observations could be one reason for the relatively poor performance of the classifiers.

LARS-EN gives a good indication of the importance of the features in a linear model. For both data sets, the first two principal components are the most discriminative features. But there is large difference in the discriminative capabilities of the parameters from the two data sets, which also would be expected when the features are plotted, see Figure 5. The rest of the principal components are selected somewhat randomly, showing that feature reduction using PCA is not necessarily in accordance with classification criterions.

A problem encountered using canonical discriminant analysis, is a good separation of the training data, but a poor performance of assigning the unknown objects to the right classes. For the vegetable data, where we have a very good separation, this becomes very clear. We would expect to retain the separation, but that is not the case because of variation in training data. This sensitivity towards variation is a clear limitation to the canonical discriminant analysis.

Fig. 5. Plot of the first two principal components from the AAM in the vegetable experiment (*left*) and wood log experiment (*right*).

The AAM's in the reduced data set is based on only 9 images of each class. This is probably too few to get good estimates of $\boldsymbol{\Phi}$, and could be a cause of the poor AAM classification performance.

Experiment 3 The best performance for the wood logs is achieved by the texture analysis in experiment 3, reaching close to 90% correct classification rates for the whole data set, and 96.8% correct classification in the best experiment of the reduced data set. This shows an accordance between what we see as humans and the predictions of the model.

Model	LARS-EN	
Data set	*whole*	*reduced*
Gray level	78.4%	93.7%
Gray level directional	89.9%	96.8%
Color	89.5%	90.5%

Table 2. Classification rates of experiment 3 using LARS-EN for classification based on texture features.

A varying number of features are calculated, because of differences in distances, directions, and across color bands in the three models. In the first model we have 33 features (*only different distances*), in the second 132 features (*different distance and direction*), and 528 features in the third (*varying all three*).

Looking at which features are selected by the LARS-EN classifier, we can see a pattern in the features selected for classification. The sum average and the diagonal moment, are the most frequently used features, even though, there is not a clear pattern in which features works best for classifying the wood logs. In contrast to Palms [15] investigations, the performance in our experiments is not improved by extending the analysis to include color images.

A hard task using the LARS-EN algorithm is to find the right stopping criterion [21]. The results presented here is the number of features giving the best classification rates. Therefore, the LARS-EN algorithm will be problematic to implement in a real world application.

In these experiments we have used logs of young trees where some important biological characteristics have not yet developed, e.g. the colored core of Scotch Pine, which could improve the hard distinguishing of Pine and Spruce.

5 Conclusion

We have investigated the use of active appearance models (AAM's, cf. [4]) for classification of biological objects, and shown that this approach is well suited for different objects. Two data set, one of vegetables and one of wood logs have been investigated.

In experiment 1 an AAM is built for each class, and we obtain results close to 100% correct classification for the vegetable data, and around 80% classification rates for wood logs.

In experiment 2 one AAM for all classes is built, and model parameters for test images are used for classification. Most models gave 100% correct classification for the vegetables. On average the classification for the wood logs was not as good as experiment 1, and especially canonical discriminant analysis gave very poor results.

In experiment 3 LARS-EN has been used for classifying texture features, where only the wood log data is investigated. This experiment gave the best results classifying about 90% of the test set correct in the best cases of the whole data set, and up to 96.7% correct classification using a reduced data set.

It is hard to find a good stopping criterion for the LARS-EN model, which is problematic for classification. Therefore, we conclude that the most promising classification model is the texture difference used in experiment 1.

6 Acknowledgements

References

1. P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):711–720, 1997.
2. J. M. Carstensen. *Description and Simulation of Visual Texture*. PhD thesis, Institure of Mahematical Statistics and Operations Research, Technical University of Denmark, Kgs. Lyngby, 1992.

3. P. Chang and J. Krumm. *Object Recognition with Color Cooccurrence Histogram.* 1999.

4. T. F. Cootes and C. J. Taylor. Statistical models of appearance for medical image analysis and computer vision, 2004. In Proc. SPIE Medical Imaging.

5. M. Demirci, A. Shokoufandeh, S. Dickinson, Y. Keselman, and L. Bretzner. Manyto -many feature matching using spherical coding of directed graphs, 2004.

6. S. Dickinson, L. Bretzner, Y. Keselman, A. Shokoufandeh, and M. F. Demirci. Object recognition as many-to-many feature matching. *International Journal of Computer Vision*, 69(2):203–222, 2006.

7. G. J. Edwards, T. F. Cootes, and C. J. Taylor. Face recognition using active appearance models. *Computer Vision - ECCV'98. 5th European Conference on Computer Vision. Proceedings*, pages 581–95 vol.2, 1998.

8. B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *Annals of Statistics*, 32(2):407–451, 2004.

9. J. Fagertun, D. D. Gomez, B. K. Ersbøll, and R. Larsen. A face recognition algorithm based on multiple individual discriminative models. In Søren I. Olsen, editor, *Dansk Selskab for Genkendelse af Mønstre (Danish Pattern Recognition Society) DSAGM 2005*, DIKU Technical report 2005/06, pages 69–75, Universitetsparken 1, 2100 København Ø, aug 2005. DIKU, University of Copenhagen.

10. D. W. Hansen, M. Nielsen, J. P. Hansen, A. S. Johansen, and M. B. Stegmann. Tracking eyes using shape and appearance. In *IAPR Workshop on Machine Vision Applications - MVA*, pages 201–204, dec 2002.

11. R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Trans Syst Man Cybern*, SMC-3(6):610–621, 1973.

12. T. Hastie, J. Tibshirani, and J. Friedman. *The Elements of Statistical Learning, Data Mining, Inference and Prediction.* Springer, 2001.

13. D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

14. D.G. Lowe. Object recognition from local scale-invariant features. *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, 2:1150 –1157 vol.2, 1999.

15. C. Palm. Color texture classification by integrative co-occurrence matrices. *Pattern Recognition*, 37(5):965–976, 2004.

16. C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(5):530 –535, 1997.

17. K. Skoglund. The lars-en algorithm for elastic net regression - matlab implementation, 2006.

18. M. B. Stegmann. Object tracking using active appearance models. In Søren I. Olsen, editor, *Proc. 10th Danish Conference on Pattern Recognition and Image Analysis*, volume 1, pages 54–60, Copenhagen, Denmark, jul 2001. DIKU.

19. M. B. Stegmann, B. K. Ersbøll, and R. Larsen. FAME - a flexible appearance modelling environment. *IEEE Transactions on Medical Imaging*, 22(10):1319–1331, 2003.

20. M.A. Turk and A.P. Pentland. Face recognition using eigenfaces. *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR '91., IEEE Computer Society Conference on*, pages 586–591, 1991.

21. H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 67(2):301, 2005.