Are Your Eyes Smiling?
Detecting genuine smiles with support vector machines and Gabor wavelets

Gwen Littlewort-Ford,
Marian Stewart Bartlett, and Javier R. Movellan
Institute for Neural Computation, UCSD

Abstract

We investigate the problem of computer recognition of "Duchenne" vs. non-Duchenne smiles. Duchenne smiles include the contraction of the orbicularis oculi, the sphincter muscles that circle the eyes. Genuine, happy smiles can be differentiated from posed, or social smiles by the contraction of this muscle (Ekman, Friesen, and O'Sullivan, 1988). This is a difficult visual discrimination task. Previously published performance of computer vision systems on this task is in the low 80%'s for a two-alternative forced choice. Gabor wavelet representations have been found to be highly effective for image recognition, including facial identity recognition (Lades et al., 1993), and facial expression analysis (Bartlett, 2000). Here we investigate the performance of Gabor wavelet representations in combination with support vector machines (SVM's) on the task of discriminating Duchenne from non-Duchenne smiles. A multiscale Gabor representation defined by Lades et al. (1993) comprised the inputs to support vector machines. The SVMs performed significantly better on this task than previously published systems. We analyze the reasons underlying the success of our approach. Linear SVM kernels did not perform as well as polynomial or Gaussian kernels, suggesting that the classes are not linearly separable, despite the high dimensionality of the Gabor vectors. SVM's on unfiltered difference images did not perform as well as the SVM's on the Gabor filtered images, demonstrating that the Gabor filters did contribute to overall classification performance.

1 Choice of classifier and task

Why use Support Vector Machines on Duchenne Smiles? Support vector machines, introduced by V. Vapnik,(Boser et al, 1992) are good classifiers for distinguishing between two classes of vectors. The theory can be extended in various ways to cope with multiclass tasks for later applications (e.g. Lee, Lin, and Wahba, 2001). Since the core of the SVM algorithm entails a quadratic optimization, it is usually quick and easy to use.

By effectively embedding vectors in a higher dimensional space, where they may be linearly separable, SVM's can solve highly non-linear problems. The computational complexity of the problem depends on the number of training examples, not on the dimension of the embedding space. This is ideal for many face emotion recognition problems, which tend to involve a relatively small set of training examples (order hundred), with each vector representation vector being very long (order million dimensions).

In this paper we investigate the problem of computer recognition of "Duchenne" vs. "non-Duchenne" smiles. Duchenne smiles involve the contraction of the orbicularis oculi, the sphincter muscles that circle the eyes. In Ekman's facial action coding system (Ekman et al 1988) , contraction of this muscle is coded as facial action unit 6 (FACS-6). Figure 1 shows examples of smiles in which FACS-6 is present (left side image) and absent (right side image). The presence of this action unit can differentiate genuine, happy smiles from posed, or social smiles (Ekman, Friesen, and O'Sullivan, 1988). Automatic recognition of this action unit is difficult, with previously published performance of computer vision systems on this task is in the low 80%'s for a two-alternative forced choice (e.g. Cohn et al. 1999).

Figure 1: Example images. a. Duchenne b. Non-Duchenne.

2 Overview of SVMs

For a full introduction, see, for example, the book "Support Vector Machines" by Cristianini and Shawe-Taylor. The objective in SVMs is to find a hyperplane that separates the data points (training vectors $\vec{X}_i$) correctly into two groups (training classes $y_i = \pm 1$) with as much distance as possible between the data points on either side. To begin with, consider the case where the two groups are linearly separable and the plane lies in the training vector space.

The margin m is the distance from the hyperplane to the closest points either side, that is, to the (parallel) marginal hyperplanes. Weights $\vec{W}$ and threshold b specify the hyperplane $\vec{W}.\vec{X} + b = 0$ you wish to find. For point $\vec{X}_s$ on the marginal hyperplane $\vec{W}.\vec{X} + b = 1$ and point $\vec{X}_0$ in the hyperplane,

$$ m = \frac{(\vec{X}_s - \vec{X}_0) \cdot \vec{W}}{|\vec{W}|} = \frac{(1 - b) - (-b)}{|\vec{W}|} = \frac{1}{|\vec{W}|}. $$

Thus maximizing the margin is equivalent to minimizing the weight vector length subject to the following constraints

$$\text{when } y_i = 1 \qquad \vec{W} \cdot \vec{X}_i + b \geq 1$$
$$\text{when } y_i = -1 \qquad \vec{W} \cdot \vec{X}_i + b \leq -1$$

Written together:

$$\mathcal{C}_i = y_i(\vec{W} \cdot \vec{X}_i + b) - 1 \geq 0 \tag{1}$$

All vectors are correctly classified and none lie inside the margin. This motivates the cost function $.5W^2 - \vec{\alpha}.\vec{\mathcal{C}}$ with Lagrange multipliers α

$$
\begin{aligned}
\mathcal{L}_p(W, b, \alpha) &= \frac{1}{2}\sum_{k=1}^{d} W_k^2 - \sum_{i=1}^{l}\alpha_i[y_i(\sum_{k=1}^{d} W_k X_{ik} + b) - 1] \\
&= \frac{1}{2}\sum_{k=1}^{d} W_k^2 - \sum_{k=1}^{d} W_k(\sum_{i=1}^{l}\alpha_i y_i X_{ik}) - b\sum_{i=1}^{l}\alpha_i y_i + \sum_{i=1}^{l}\alpha_i \tag{2}
\end{aligned}
$$

Minimizing with respect to W and b yields an optimal weight vector W^* that is a linear combination of certain of the training vectors $\vec{X}_i$'s called support vectors

$$\frac{\partial \mathcal{L}_p}{\partial W_k} = 0 = W_k - \sum_{i=1}^{l}\alpha_i y_i X_{ik} \quad \longrightarrow \quad \vec{W}^* = \sum_{i=1}^{l}\alpha_i y_i \vec{X}_i \tag{3}$$

$$\text{Equality constraints: } \frac{\partial \mathcal{L}_p}{\partial b} = 0 \quad \longrightarrow \quad \sum_{i=1}^{l}\alpha_i y_i = 0 \tag{4}$$

Inserting optimal weights $\vec{W}^*$, the Dual Lagrangian is a function of the multipliers, and can be optimized numerically to find the α's which in turn determine the support vectors.

$$
\begin{aligned}
\mathcal{L}_D(X, y, \alpha) &= \frac{1}{2}\sum_{k=1}^{d}(W_k^*)^2 - \sum_{k=1}^{d} W_k^*(W_k^*) - b\sum_{i=1}^{l}\alpha_i y_i + \sum_{i=1}^{l}\alpha_i \\
&= -\frac{1}{2}\sum_{k=1}^{d}(W_k^*)^2 + \sum_{i=1}^{l}\alpha_i \\
&= \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{l}\alpha_i\alpha_j y_i y_j \vec{X}_i \cdot \vec{X}_j \tag{5}
\end{aligned}
$$

$$\text{The dual } \mathcal{L}_D(X, y, \alpha) = \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{l}\alpha_i\alpha_j y_i y_j K_{ij} \tag{6}$$

In this last form of the objective function, only the SVkernel matrix $K_{ij} = \vec{X}_i.\vec{X}_j$ is needed, not the original vectors. Since testing the SVM classifier on new data can also be formulated in terms of the SVkernels (because the test output $X_{new}.W^*$ is a

known linear combination of $K_{new,sv}$), SVM's can be generalized to non-linearly separable problems by defining SVkernels which are non-linear (positive semi-definite) functions of covariance or distance for each pair of vectors. For more on the possible types and restrictions of kernel functions see Chapter 3 of Cristianini et al.

When the dual Lagrangian is optimized (maximized) numerically, most of the multipliers α are zero. Further analysis of this type of problem leads to Karush-Kuhn-Tucker conditions which clarify how to interpret the numerical solutions.

$$\text{KKT} : \alpha_i[y_i(\sum_{k=1}^{d} W_k X_{ik} + b) - 1] = 0 \tag{7}$$

These conditions imply that either $\alpha = 0$, or α_i is non-zero and the original inequality constraints above $\mathcal{C}_i$ are satisfied with equality (points in the marginal hyperplanes). The support vectors which make up the optimal weight vector $\vec{W}^*$ have coefficients α_i, so only those training vectors with non-zero α contribute, thus all support vectors must lie in one of the two parallel marginal hyperplanes. It is precisely those vectors which are hardest to classify, being closest to the opposite side in the projected hyperspace, which determine the dividing hyperplane.

In cases where there are outliers in the wrong camp, or where some training vectors may be incorrectly labeled, it is useful to use soft margins. This allows some support vectors in the training set to be on the wrong side of the appropriate marginal hyperplane. This effectively puts an upper limit on the α's. For the particular task in this paper, soft margins have not proved useful.

3   Data Sources

Duchenne and non-Duchenne face images were collected by J. Cohn at Pittsburgh (Pitts) and by P. Ekman and J. Hager at UCSF (Ekman-Hager). The Pittsburgh data was collected by asking naive subjects to pose a smile. The facial behavior, including the presence or absence of FACS-6, was then scored by two certified FACS coders. The Ekman-Hager dataset was collected by directing expert subjects to contract specific facial muscles. The facial actions were then verified by three certified FACS coders . A set of genuine/posed smiles were collected by C. Riley of the BBC. Twenty actors were videotaped while watching a segment of "Faulty Towers" and again when asked to demonstrate their best smile. Two smile intensity levels and a neutral were digitized for each subject.

3.1   Preprocessing of images

The centers of the eyes and mouth were manually located in each frame. The faces were linearly warped and rotated so that the center of the eyes and mouth occupied the same pixel position across the entire database. The images were then cropped to 66×96 for the upper face and then the pixel values where linearly rescaled to minimum and maximum values of 0 and 255. Difference images were obtained by subtracting the neutral expression in the first image of each sequence from the subsequent images in the sequence.

Gabor functions are sine wave gratings modulated by a Gaussian. Banks of Gabor filters at multiple scales and orientations model the response properties of primary visual cortical cells in primates (Daugman 1988). Gabor representations of the smile

images were obtained by convolving difference images with a family of Gabor filters at 5 spatial scales and 8 orientations, increasing the dimensionality of each vector by a factor of 40 (Lades et al. 1993; Donato et al., 1999).

3.2 Constructing SVkernels

Support vector kernels were constructed for the Gabor filter representations of the images as well as for the plain gray scale images and the difference images. All kernels were constructed using either the matrix of covariance between images C or the squared distance matrix d^2, thus dispensing with the high dimensionality of the filtered images. Functions applied (component-wise) to these matrices included polynomials of form $(C + 1)^n$, Gaussian radial basis functions of d^2, exponentials, Laplacians, reciprocals of the form $\frac{1}{D+1}$.

3.3 SVM training: leave-'one'-out validation

The data was divided into test sets, which were omitted, one at a time, from the kernel matrix during training. The support vector coefficients α calculated in the numerical optimization were then used to classify the members of the test set.

4 Performance on classifying smiles

4.1 BBC data

The top row of Figure 2 illustrates 'genuine' smiles from the BBC set, in which while subjects responded to an episode of Faulty Towers. The bottom row illustrates 'posed' smiles in which the actors were told to look happy. In pilot studies, human subjects including delegates to the Joint Conference on Neural Computation scored about 60 percent correct on guessing which people were which. Using a linear SVM on normalized Gabor filter outputs, 75 percent of the 40 test images (low and high intensity) were correctly classified.

Figure 3 shows the upper face difference images for 3 examples from each class using images correctly classified by the SVM (top and bottom rows). These are the same images as Figure 2. The two extra images are examples of support vectors from each class respectively.

4.2 Pittsburgh data

Performance of leave-one-out validation on medium amplitude smiles from 90 sequences containing AU 6,12 (Duchenne) or AU 12 only (non-Duchenne) was 87% for a Gaussian SVkernel ($\sigma = 0.9$) on gray-scale distance, and 90% for a linear combination of SVM outputs of plain Gray and Gabor filtered SVM's.

Figure 2: Examples from the BBC data set. Row 1: Genuine, Row 2: Posed.

Figure 3: Difference images from the BBC data set. Row 1: Genuine, Rows 2 and 3: Support Vectors, Row 4: Posed

	E+	P+	E-	P-
E+	1.00	0.64	0.42	0.38
P+	0.64	1.00	0.33	0.59
E-	0.42	0.33	1.00	0.32
P-	0.38	0.59	0.32	1.00

Table 1: Correlations between difference image vectors for the Ekman-Hager (E) and Pittsburgh (P) datasets. Duchenne and non-Duchenne classes are indicated by $+$ and $-$ respectively.

4.3   Ekman-Pittsburgh combined data

Two data sources were combined, resulting in 100 Ekman-Hager Duchennes, 57 Pittsburgh Duchennes, 40 Ekman-Hager non-Duchennes and 32 Pittsburgh non-Duchennes.

Correlations between difference image vectors from these four sets (Table 1) show that while there is some correlation between Duchennes from different sources, the non-Duchennes are more closely correlated with their own Duchennes than with other non-Duchennes from a different source. This illustrates the difficulty of the task.
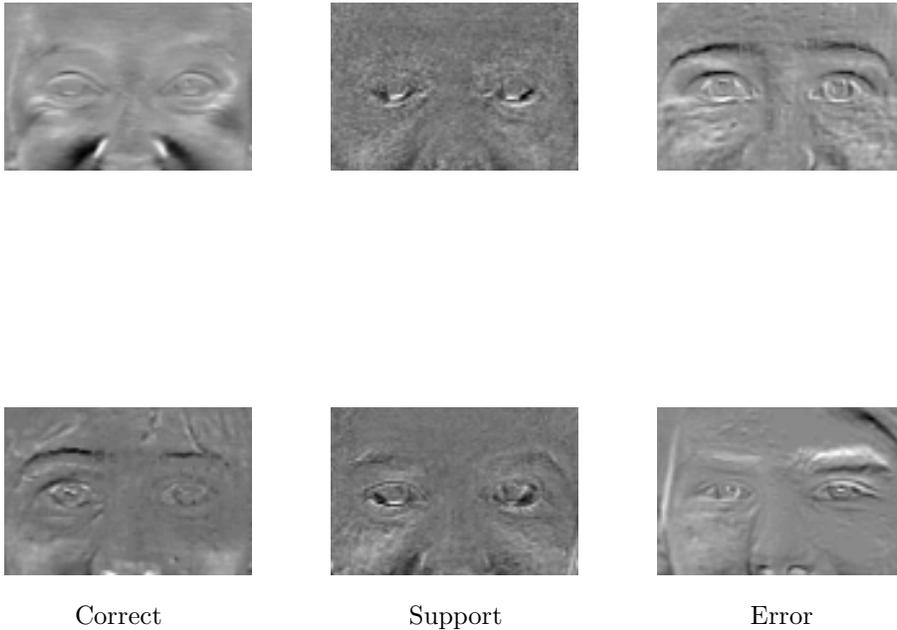

| Correct | Support | Error |

Figure 4: Ekman-Pittsburg Upper Face Difference Images. Top row: Duchenne, Bottom Row: Non-Duchenne, Left: SVM correct, Right: SVM in error, Middle: a Support Vector

Classification performance was 84% for the best individual SVM, and 87% for a linear combination of experts.

For a larger data set of 319 images including low and high intensities for each subject, the best SVM was $\frac{1}{0.5+d^2}$ which obtained 86% accuracy.

7

The support vectors (middle column, Figure 4) are similar to one another, since they are closest to the margin. The erroneous Duchenne example (top right) and the correct non Duchenne example (lower left) look similar.The SVM classified them both as non Duchenne.

5   Gabors and SVM's

Gabor filter representations of images have been found to be highly effective for face recognition (e.g. Lades et al., 1993) and expression analysis (e.g. Donato et al. 1999). See Fasel, Movellan, and Bartlett (2001) for an analysis of Gabor filter methods for face image processing. Aside from the similarity to visual cortical cell responses, the reason for the success of this representation is unclear. One hypothesis is that the bank of Gabor filters projects the images in to a high dimensional space where the classes are linearly separable in a manner analogous to SVM kernels. If that is the case, then a linear classifier should perform as well as a nonlinear classifier, each taking Gabor representations as input. In addition, if the action of the Gabor projection is equivalent to an SVM, then SVM's applied to the original images should perform as well as SVM's applied to Gabors, given that the correct kernel can be found.

For all Duchenne versus non-Duchenne discriminations (The Ekman-Hager, Pittsburgh, and combined datasets), linear kernels performed 5-10 percentage points lower than nonlinear kernels applied to the Gabor outputs. SVM's on unfiltered difference images typically performed only a few percentage points lower than SVM's on the Gabors, but required more complex kernels and a more extensive search for the right kernel.

Gabor filters also appeared to minimize the need for taking difference images. SVM's applied directly to the original grayelvel images were near chance, whereas SVM's applied to Gabor representations performed similarly, whether the Gabors were applied to the original graylevel images, or to the difference images. The difference images did augment the performance with Gabors, but typically by only a few percentage points.

6   Conclusions

The SVMs performed significantly better on this task than previously published systems. Linear SVM kernels did not perform as well as polynomial or Gaussian kernels, suggesting that the classes are not linearly separable, despite the high dimensionality of the Gabor vectors. SVM's on unfiltered difference images did not perform as well as the SVM's on the Gabor filtered images, demonstrating that the Gabor filters did contribute to overall classification performance. Using Gabor representations allows the use of simpler SVkernel functions, but the best performing SVkernel is sometimes a function of distance between difference images. Linear combinations of several experts can perform better than individual SVMs, provided that kernels from both plain gray-scaled images and Gabor filtered images are used.

References

Bartlett, M.S., Donato, G.L., Movellan, J.R., Hager, J.C., Ekman, P. and Sejnowski, T.J. (2000). Image representations for facial expression coding. In S. Solla, T. Leen, and K.-R. Muller (Eds.). Advances in Neural Information Processing Systems, Vol 12. MIT Press.

B.E.Boser. I.M. Guyon, V.N.Vapnik (1992). "A training algorithm for optimal margin classifiers. Proceedings of 5th Annual ACM Workshop on Computational Learning Theory", p144-152.

Cohn, J.F., Zlochower, A., Lien, J., & Kanade T., (1999). Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding. Psychophysiology, 35(1).

Cristianini, N. and Shawe-Taylor, J. (2000). Support Vector Machines. Cambridge University Press.

Daugman, J.G. (1988). Complete discrete 2D Gabor transform by neural networks for image analysis and compression. IEEE Transactions on Acoustics, Speech, and Signal Processing, 36, p. 1169-1179.

Donato, G.L., Bartlett, M.S., Hager, J.C., Ekman, P., and Sejnowski, T.J. (1999). Classifying Facial Actions. IEEE Transactions on Pattern Analysis and Machine Intelligence 21(10) p. 974-989.

P. Ekman, W. Friesen and M. O'Sullivan (1988). Smiles when lying, Journal of Personality and Social Psychology, 5445:414-420.

Fasel, I.R., Bartlett, M.S., and Movellan, J.R. (2001). A Comparison of Gabor Filter Methods for Automatic Detection of Facial Landmarks UCSD MPLab TR 2001.04, June 2001.

Lades M., Vorbruggen, J., Buhmann, J., Langer, J., Konen, W., von der Malsburg, C., and Wurtz., R. (1993). Distortion invariant object recognition in the dynamic link architecture. IEEE Transactions on Computers, 42(3):300-311.

Lee, Y., Lin, Y., and Wahba, G. (2001). Multicategory support vector machines. University of Wisconsin-Madison Statistics Department Technical Report, TR 1040.