Detecting Acromegaly: Screening for Disease with a Morphable Model

Erik Learned-Miller[1], Qifeng Lu[1], Angela Paisley[2], Peter Trainer[2], Volker Blanz[3], Katrin Dedden[4], and Ralph Miller[5]

[1] University of Massachusetts, Amherst, Massachusetts, USA
{elm, qlu}@cs.umass.edu
[2] Christie Hospital, Manchester, UK
peter.trainer@manchester.ac.uk, anpaisley@doctors.org.uk
[3] Universität Siegen, Germany
blanz@mpl-sb.mpg.de
[4] Max-Planck Institut für Informatik
[5] University of Kentucky Medical Center, Lexington, Kentucky, USA
remill0@uky.edu

Abstract. Acromegaly is a rare disorder which affects about 50 of every million people. The disease typically causes swelling of the hands, feet, and face, and eventually permanent changes to areas such as the jaw, brow ridge, and cheek bones. The disease is often missed by physicians and progresses beyond where it might if it were identified and treated earlier. We consider a semi-automated approach to detecting acromegaly, using a novel combination of support vector machines (SVMs) and a morphable model. Our training set consists of 24 frontal photographs of acromegalic patients and 25 of disease-free subjects. We modelled each subject's face in an analysis-by-synthesis loop using the three-dimensional morphable face model of Blanz and Vetter. The model parameters capture many features of the 3D shape of the subject's head from just a single photograph, and are used *directly* for classification. We report encouraging results of a classifier built from the training set of real human subjects.

1 Introduction

Figure 1 shows two men of the same age. For most observers, seeing either one of them alone on the street would not prompt any particular reaction. The man on the left would possibly be identified as slightly overweight but otherwise unremarkable while the man on the right appears perfectly healthy. In fact these men are identical twins. This photograph appeared recently in the New England Journal of Medicine as one of the journal's periodic "Medical Mysteries" with the caption, "Which twin is the patient?" [10]. The man on the left has a condition known as *acromegaly*, which is difficult for most laypeople, and even for most physicians, to diagnose.

Acromegaly is a rare disorder of the endocrine system which affects roughly 50 of every million people in the general population. Early detection is important in treating the disease successfully, but it is often missed because the signs are

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4191, pp. 495–503, 2006.

Fig. 1. These men are identical twins. Which one is sick, and what is his condition? By fitting the 3D morphable face model of Blanz and Vetter [2] to each of these faces (see Figure 2), the system described correctly identified the man on the left as an acromegalic, and the man on the right as healthy. This "Medical Mystery" first appeared in the New England Journal of Medicine [10].

subtle and the condition is rare. Since many of its symptoms affect facial appearance, the disease can be detected by specialists (endocrinologists) in many cases from a normal frontal photograph. If a patient's appearance suggests presence of the disease, blood tests may be performed to confirm its presence. Because the tests are expensive and time consuming, it would clearly be valuable to have an inexpensive and automatic prescreening method.

The ultimate goal of our research is to develop an automated, voluntary prescreening system for acromegaly. For example, when obtaining a photograph for a driver's license, one could voluntarily choose to be screened for various conditions such as acromegaly. We believe that such systems could make a signficant contribution to early detection (and hence to the timely treatment) of this disease [5].

In this work, we present a new way of combining morphable models [2] and support vector machines for supervised learning. We use this method in a semi-automated approach to detecting the presence of acromegaly in people, using a supervised learning paradigm. We built a training set of images by taking 24 frontal photographs of acromegalic patients and 25 frontal photographs of disease-free subjects. Because the recognition of acromegaly is dependent upon subtle features which are difficult to detect locally in the image, we decided upon a global method of modelling that uses information across the entire photograph. In particular, we modelled each subject's face in an analysis-by-synthesis loop using the morphable models of Blanz and Vetter [2]. The parameters of the morphable model capture many features of the three-dimensional shape of a subject's head from just a single photograph and are excellent features with which to classify subjects as either acromegalic or not. We use the parameters recovered for each test face in an SVM to classify a subject as having acromegaly or not. We report encouraging results of a classifier built from the training set of real human subjects.

Fig. 2. A 3D morphable model [2] was adapted in an analysis-by-synthesis loop to match each of the men in Figure 1. Although the models are not perfect replicas of the photographs, they capture important properties of the faces which prove to be effective in distinghishing between patients that do or do not have acromegaly. For example, the 3D model captured the swelling of the left man's nose, a strong indicator of acromegaly which is difficult to detect as an image feature using traditional feature detectors operating on the image. Our classifier correctly classified the man on the left as acromegalic and the man on the right as negative for the disease.

While in a previous paper [7], Huang et al. use a morphable model in conjunction with SVMs for classification, this was a signficantly different approach than our own to combining these two tools. In that paper, the morphable model is used to generate additional sample images of a subject's face under varying lighting conditions and poses. These additional images can improve recognition accuracy in tasks like subject identification. Unlike our work, the previous paper uses the SVM to analyze these synthetic *images* of subjects' faces. In our approach, the SVM operates on the *parameters* of the morphable model directly. This is a key point, as the parameters of the morphable model, being natural parameters of face shape variation, are likely to be more discriminative than pixel-valued parameters as used in the previous work. A separate paper [3] uses model parameters directly for subject identification, but using a nearest neighbor classifier. Thus, to the best of our knowledge, this paper is the first to use SVM classification directly on morphable model parameters. In addition, this is the first paper to use such techniques in medical diagnosis.

2 The Training Data

Because acromegaly is uncommon, photographs of acromegalics are not readily available. Thus, a major portion of the effort in this work was in acquiring a database of images of 24 acromegaly patients, and a matching set of images from 25 healthy subjects. Three acromegalics from our database are shown in Figure 3. The leftmost patient shows clear signs of the disease. Features such as a large jaw, protruding brow, frontal bossing (protrusion of the forehead), swollen nose, prominent cheekbones, enlarged lips, and prominent naso-labial

Fig. 3. Examples of acromegalics from our database. The symptoms decrease in severity from left to right. (See text.)

folds (creases in the skin of the cheek) make such a patient a strong candidate for further evaluation by an endocrinologist. While no one of these symptoms would indicate acromegaly, taken together they are a strong indication of disease. The middle patient in the figure still has clear, although more subtle, signs of acromegaly. He is an example of a patient that would likely not be detected by a lay person, and might often be missed by a physician who is not an acromegaly specialist. The patient has an enlarged jaw, mild nasio-labial folds, and slight enlargment of the cheekbones. On the right of the figure is a patient with very minor, if any, signs of the disease visible even to the expert. It seems unlikely that any visual test will diagnose the most difficult cases; our goal rather is to flag patients in the second category.

One of the dangers in building a binary classifier is that details of the image acquisition environment will be leveraged by the classifier to "recognize" a condition such as acromegaly. That is, if some simple feature of the acquisition environment differs between one class and another, this may be used improperly to aid the classification. This was an especially important concern for us since we acquired images of acromegalics in one location and images of normals in another location. To minimize the probability that some exogenous factor affected the classification performance, we developed a protocol for taking photos. Our protocol specified the camera make and model, the background for the photograph, the expression of the subject (relaxed and neutral expression, including a closed mouth and open eyes), the orientation of the patient (front-facing), and the general lighting conditions. Ten pictures of each patient were taken so that images with closed eyes, accidental non-neutral facial expressions, blur due to movement, and other anomalies could be removed. One of the defect-free photographs of each subject was chosen manually for inclusion in the final database. The color and texture of a patient's face was *not* used directly in the final classification. This minimizes two important effects which could cause biases in the classification, that of lighting which is difficult to control carefully, and that of skin tone, which happens to be biased slightly toward darker tones in our set of acromegalics relative to our set of normals.

There are three properties of our database that are not ideal, and which we plan to address in the near future. As noted above, all of the acromegalic photos were taken in one locale and all of the normals were taken in another locale. Despite our precautions, this could possibly lead to hidden biases in the classifier. Another issue is that the photos of acromegalics were taken somewhat closer or with a slightly different zoom, on average, than the photos of normals. While the morphable model software explicitly incorporates estimates of distance and zoom into its estimate of 3D shape (minimizing the effect of differences in perspective or fish-eye distortion), these are nevertheless systematic differences in the database which should be eliminated in the future. Finally, our database of normals consists only of white males, while our acromegalics are a signficantly more diverse population. While we do not believe that these issues had a significant impact on the classifier, we plan to remedy them by improving the database in the future. Next, we turn to the task of modelling the faces in our database.

3 Modelling Faces

Many of the symptoms of acromegaly are difficult to capture using traditional local image features such as edges and image derivatives. Symptoms such as swelling of the nose and lips, protrusion of the brow and cheekbones, and growth of the jaw are very difficult to detect locally. Our initial work on this project focussed on measuring distances between various facial landmarks and computing the relative size of landmarks not correlated with disease, such as iris diameter, and landmarks like jaw width, that would be expected to be larger, on average, given that the patient had the disease.

One problem with this approach is choosing landmarks that lead to consistent measurements of a patient's face. It is difficult to define in a repeatable fashion measurements such as the "width of the jaw". Developing software to do this automatically is even more difficult. Another problem with the landmark method is that it does not use all of the information available in the photograph. A patient's nose may be a normal width, according to a landmark based specification, and yet it may be clear to any observer that the patient's nose is actually swollen. Thus, landmark-based methods may not capture a significant symptom which is obvious to an observer.

It became clear that a system which could model the true three-dimensional shape of each subject's head could alleviate many of these problems. The 3D morphable models of Blanz and Vetter [2] seemed to be an ideal tool to tackle such a problem. In previous work, Blanz and Vetter developed a linear statistical model of the 3D geometry and texture (or surface color) of human heads from a set of 3D Cyberware (TM) laser scans of 200 individuals (100 male and 100 female).

In the morphable face model, facial surface data that were recorded with the laser scanner as a triangular mesh are represented in shape vectors that combine x, y, and z coordinates of all vertices:

$$\mathbf{v} = (x_1, y_1, z_1, ..., x_n, y_n, z_n)^T \in \mathcal{R}^{3n}.$$

Sampled at a spacing of less than one millimeter, each surface is represented by $n = 75972$ vertices. Linear combinations of shape vectors will only produce realistic novel faces if corresponding points, such as the tip of the nose, are represented by the same vector components across all individual shape vectors. This is achieved by establishing dense correspondence between different scans, and forming vectors $\mathbf{v}_i$ in a consistent way. Along with shape, the morphable face model also represents texture, but texture is discarded in our work here.

There are two key features of the morphable model work that made it an ideal tool for our application. First, the 3D head model represents shape variations in terms of the common modes (principal components) of shape deformation of healthy human heads. Each parameter of the geometry-part of the model describes a common mode of variation in (densely) registered human heads. While parameters such as jaw size are not explicitly coded into the model, the statistical model must be able to represent such variability as linear combinations of its parameters in order to achieve good approximations of the subjects in the original Cyberware scan database. That is, as long as there were some significant relative differences in jaw size among the initial subjects, the model would need to encode this variability in some form to make good fits. We hypothesized that the analysis of these sorts of natural parameters of face variation would allow us to sort subjects into groups of acromegalics and healthy patients.

The second appealing feature of the morphable model work is that its developers showed how an analysis-by-synthesis method could produce the approximate 3D shape and texture of a person's head from a single photograph. After a manual initialization process (described below), a fully automatic procedure adapts the parameters of the morphable model until a rendered image of the model matches a given photograph as closely as possible under a soft constraint that makes the parameters of the resulting 3D head as likely as possible under the statistical morphable model. That is, the analysis-by-synthesis loops strives to minimize an error of the form

$$E(\alpha) = \sum_{x,y} ||\mathbf{I}_{input}(x, y) - \mathbf{I}_{model}(x, y; \alpha)||^2 + f(\alpha),$$

where α is a vector of parameters estimated for the particular head under consideration and $f(\alpha)$ is the negative log likelihood of the parameters under the original linear statistical model of heads.

We used the software of Blanz and Vetter to develop 3D models from our database photographs. To intialize the model fitting process, the user clicks on a number (seven to 12) of feature points in the image and the corresponding features on the 3D model using an interactive tool. These point correspondences are enforced in the first iterations of the algorithm. Their weight is gradually reduced to zero during the optimization. The pose, illumination, shape and texture coefficients are all set to the default values at the beginning of the fitting process. Additional details of the initialization and fitting process are described elsewhere [3].

In addition to global estimates of head shape, a secondary procedure was used in which smaller parts of the face were extracted and estimated separately. This procedure gives greater detail and accuracy for the followings facial regions, or groups of regions: the nose; the eyes, eyebrows, and brow; the mouth; and all other features, such as the cheeks, chin, forehead, ears, and neck. Thus, in addition to a single global geometric shape estimate for each head, there were four additional "parts" estimates, yielding a total of five individual geometric 3D shape estimates per face. In our experiments, these additional shapes improved the performance of the classifier significantly.

4 Experiments

After fitting the morphable model to each face in our database, and to each of the subregions described above, we retrieved 199 "geometry" parameters for each of the five pieces of the head estimate, for a total of 995 parameters per photograph. Parameters measuring texture were discarded as it was decided in advance that the small benefit they might add would be outweighed by the general increase in variance of the results.

We used the first 99 parameters from each part of the geometric head shape estimate, for a total of 495 parameters per head in our final experiments. These parameters represent the components of shape with greatest variability in the original database of Blanz and Vetter. To see how this is used in acromegaly classification, consider a feature such as jaw size. Normal variation in jaw size will be expressed by the variability of some parameter of the original "normal" morphable model. If an acromegalic has an enlarged jaw, then fitting the morphable model will give a large coefficient for this particular parameter. *By examining these coefficients over the full set of acromegalics we can detect trends in these unusually large coefficients and select them as features for classification, rather than as simply statistical outliers.*

With only 24 examples from the acromegalic class and 25 examples from the normal class, we needed to mitigate overfitting as much as possible. As a result we decided on a leave-one-out classification paradigm, in which we train on all examples except one, and then use the trained classifier to classify the remaining single sample. *It is important to note that we are not using leave-one-out* **cross-validation** *[4], across multiple testing points, to train our classifier.* In other words, the classification results on one data point have no influence on the training of the classifier for another data point. To do this would be "fitting to the test data", and would invalidate our results. While the variance of our results may be high due to the small amount of traning and test data, we have no reason to believe that we are overfitting or that our results are biased.

When working with such a small data set, it is important not only to control the capacity of the classifier being used, but also to limit the number of classifiers tried on the test data. Since we could not afford (given the limited training set size) to set aside a validation set with which to experiment, we chose almost all of the classifier parameters up front. We decided in advance to use 99 parameters from each part of the geometric shape estimate, and to use linear or quadratic SVMs. While we show results (in Figure 4) using varying numbers of model components to suggest a general trend, we chose, before seeing the training data, to select the one with 99 parameters as our final accuracy estimate. We

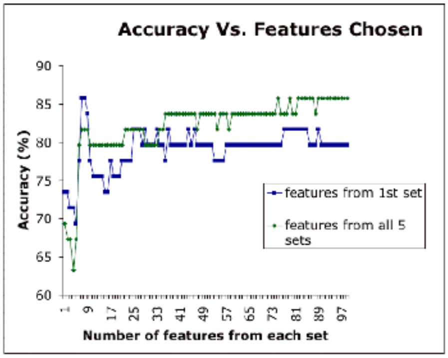

Fig. 4. Leave-one-out classification accuracy of a linear kernel SVM versus the number of morphable model features used. The blue curve shows performance using only the global estimate of the head, while the green curve shows the performance using components from each of the five geometric shape estimates (one global and four parts based estimates).

did allow ourselves the luxury of experimenting with quadratic and linear kernels, and found that linear kernels performed better.

We used the publicly available SVM package, SVM Light [8]. Using a leave-one-out paradigm, we trained an SVM using all but one of the training samples, and tested on the remaining sample. The accuracy with all 495 parameters from the five geometric parts estimates was 85.7%. Interestingly, none of the normals were classified as acromegalics, while seven of the acromegalics were classified as normals. The sensitivity of the test was hence $17/24 \approx 71\%$. Five of the seven misclassified patients had very subtle signs of the disease that might easily be missed by an expert. Two misclassified examples showed obvious signs of the disease. The morphable models for one of these was a poor fit to the patient's photograph and may explain one of the errors.

There are a number of efforts in the medical community to statistically analyze face shape (e.g. [6,1]). These methods require a specialized apparatus to acquire three-dimensional face information. Stereo methods [1] and full 3D scans [6] are common methods for such acquisitions. A key feature of our work is that while the original statistical morphable model required 3D laser scans, it can be applied to other databases, like ours, consisting only of regular two-dimensional photographs. Such systems could in theory be widely deployed, analysing standard photographs to screen for a variety of conditions. To make such a system practical, the remaining manual initialization of the morphable model to a photograph will probably need to be automated.

While there is an enormous literature on face recognition, it is less common to use face databases to categorize images into categories. An example is the classification of faces as male or female [9]. We are not aware of any previous work in trying to identify acromegaly. The visual nature of many of the symptoms and the benefits from early diagnosis [5] make this disease an ideal candidate for this type of analysis.

References

1. Nicola D'Appuzo, "Measurement and modeling of human faces from multiimages," *International Archives of Photogrammetry and Remote Sensing*, Volume 34, Number 5, Commission V, WG, V/6. Pages 241–246, 2003.
2. Volker Blanz and Thomas Vetter, "A morphable model for the synthesis of 3D faces." *SIGGRAPH'99 Conference Proceedings.* 1999.
3. Volker Blanz and Thomas Vetter, "Face recognition based on fitting a 3D morphable model." *Trans. on Pattern Analysis and Machine Intelligence*," Vol. 25, No. 9. pp. 1063–1074. 2003.
4. Richard Duda, Peter Hart, and David Stork. "Pattern Classification." Second edition. John Wiley and Sons, Inc., New York, 2001.
5. Pamela U. Freda, "Advances in the diagnosis of acromegaly." *The Endocrinologist,* Volume 10, pages 237-244, 2000.
6. Peter Hammond, Tim J. Hutton, Michael A. Patton, and Judith E. Allanson. "Delineation and visualization of congenital abnormality using 3D facial images." *Intelligent Data Analysis in Medicin and Pharmacology* at *MEDINFO 2001*, 2001.
7. Jennifer Huang, Volker Blanz, and Bernd Heisele. "Face recognition using component-based SVM classification and morphable models." Proceedings of the 4th International Conference on Audio- and Video-Based Biometric Person Authentication. surrey, UK, 2003.
8. Thorsten Joachims. "Learning to classify text using support vector machines: Methods, Theory, and Algorithms." Dissertation. Kluwer, 2002.
9. B. Moghaddam and M.-H. Yang. "Learning gender with support faces." *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, **24**, 5, pp. 707–711, 2002.
10. Pieters Nieuwlaat, "A medical mystery – Which twin is the patient?" *New England Journal of Medicine*, Vol. 351, page 68, 2004.