Towards Exaggerated Image Stereotypes

Chen Chen*, Francois Lauze*, Christian Igel*, Aasa Feragen*, Marco Loog† and Mads Nielsen*
*Department of Computer Science, University of Copenhagen
Email:{chen, francois, igel, aasa, madsn}@diku.dk
†Pattern Recognition Laboratory, Delft University of Technology
Email: m.loog@tudelft.nl

***Abstract*—Given a training set of images and a binary classifier, we introduce the notion of an exaggerated image stereotype for some image class of interest, which emphasizes/exaggerates the characteristic patterns in an image and visualizes which visual information the classification relies on. This is useful for gaining insight into the classification mechanism. The exaggerated image stereotypes results in a proper trade-off between classification accuracy and likelihood of being generated from the class of interest. This is done by optimizing an objective function which consists of a discriminative term based on the classification result, and a generative term based on the assumption of the class distribution. We use this idea with Fisher's Linear Discriminant rule, and assume a multivariate normal distribution for samples within a class. The proposed framework has been applied on handwritten digit data, illustrating specific features differentiating digits. Then it is applied to a face dataset using Active Appearance Model (AAM), where male faces stereotypes are evolved from initial female faces.**

I. Introduction

Classification is one of the most important task in computer vision, and a variety of classification algorithms have been proposed to improve the classification accuracy. The methodology of these classifiers has been deeply studied and may be well explained technically, however, it is far from simple to tell what these classifiers are looking for. Since the relation between the input image and the output (the classification result as a label, binary integer or probability) is, in most cases, difficult to characterize. In some applications, the acceptance of these classifiers might suffer from this. Our motivation is to visualize the mechanism of classification, and help to gain the insight into the classifiers. This is done by producing the representative images of the class of interest based on the trained classifier.

Synthesizing or modeling representative images for different purposes has been deeply discussed from different points of view [1]. Chang et al. [2] and Knight et al. [3] proposed prototype learning algorithms, which search for fewer but more significant samples in the training set in order to reduce the computational complexity of forming the decision boundary while to maintain high classification accuracy. Schölkopf et al. [4]–[8] proposed to build pre-images in feature space using kernel based principle component analysis to solve the image denoising problem. Lillholm and Nielsen [9], [10] proposed a method to reconstruct a representative image for metametric class based on the relation between features and images. Zhu et al. [11] proposed the filters, random fields and maximum entropy (FRAME) algorithm used for texture modeling, which combines the filtering theory and statistical modeling and characterizes the ensemble of the images of the same textures.

This paper focuses on proposing a framework to synthesize/build the images, called *exaggerated stereotypes* or simply *stereotypes*, which are representative of the class of interest and also distinguished from other class(es) according to the trained classifier. Corresponding to the two properties defined above, the synthesis of the exaggerated stereotypes is mainly influenced by two factors, which are based on discriminative model (classifier) and generative model (samples distribution within a class) respectively. We also introduce a concept of typicality measurement in feature space and image space, and form the task of building the exaggerated stereotype as an optimization problem, solved by a gradient descent method, which allows us to visualize the evolving path in image space.

The paper is organized as follows: in Section II, we explain the general ideas of the exaggerated stereotype problem, and define our typicality measurement in feature space and image space. Then in Section III, the framework for computation of stereotypes is introduced, leading to an optimization problem and methods to solve it. In Section IV, the proposed framework is applied to handwritten digits and a dataset of human faces based on Active Appearance Model (AAM) [12], and finally we conclude and introduce the future work in Section V.

II. Typicality Measurement

An exaggerated stereotype, as we introduced in the previous section, refers to an image having two main properties: the first is to be representative of the class of interest, which is related with the sample structure or data distribution; the second is to be distinct from the other class(es), which leads to a discriminative property. In order to evaluate how an image resembles an exaggerated stereotype with both properties, one needs a notion of typicality measurement for images. Since feature extraction is often necessarily performed in image analysis, the concept of typicality measurement in feature space is addressed first.

A. Typicality measurement in feature space

We first fix a few notations. Let D be the image domain and let $I : D \to \mathbb{R}$ be the image intensity function. Features $F \in \mathcal{F}$ are extracted from image I, this is described by the feature extraction function f:

$$F = f(I). \tag{1}$$

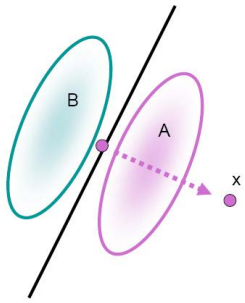


Fig. 1: The illustration of using only a discriminative term to form the typicality measurement: the feature point would be pushed too far away from decision boundary to be representative of class A.

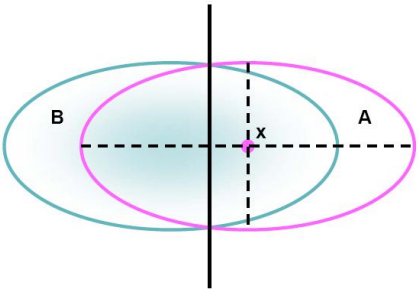


Fig. 2: The illustration of using only a generative term to form the typicality measurement: the classification becomes ambiguous when the feature point locates close to the mode of distribution of class A but also in the overlapped area.

The typicality for a feature vector should reflect the typicality for an image where the feature is extracted. Therefore, a typicality measurement of the feature vector F is thus built as a combination of two terms reflecting the output of classifier and the class distribution respectively.

We define the first term of the typicality measurement of a feature vector F, the discriminative term $E_{\mathrm{d}}^{\mathcal{F}}$, as the posterior probability of belonging to class A,

$$E_{\mathrm{d}}^{\mathcal{F}} = p(A \,|\, F), \tag{2}$$

or alternatively the likelihood ratio of the two classes

$$E_{\mathrm{d}}^{\mathcal{F}} = \frac{p(A \,|\, F)}{p(B \,|\, F)}, \tag{3}$$

which helps to discriminate between classes and quantifies the distance from the classification boundary. In this paper, we use the posterior probability as the discriminative term.

The second term of typicality measurement, the generative term $E_{\mathrm{g}}^{\mathcal{F}}$, is defined as (a function of) the conditional density function of features given some class of interest A

$$E_{\mathrm{g}}^{\mathcal{F}} = p(F \,|\, A) \tag{4}$$

to keep the stereotypes not far from the mode of class distribution. Various methods can be used to model the class density, for instance a mixture of Gaussians approach or a Parzen estimator [1], [13], [14].

On one side, since the classifiers are meant for discriminating two classes, it is not enough to rely only on classifiers to represent the typicality of any single class. As illustrated in Fig. 1, a feature point on the A side of the decision boundary, can be classified very well, but risks being far away from the mode(s) of class A, which gives an extremely low probability of belonging to class A. In term of generative model, it cannot be representative of being typical feature of class A.

On the other side, using only a generative term might not provide stereotypical enough images/features as well. If there is no or very little overlapping area between two classes, taking the features close to the modes won't lead to much ambiguity, and in this case purely using generative term can perform quite well. However, in many practical situations, the classes largely overlap, and using a purely generative term may fail to return features distinct enough from the other class(es). As illustrated in Fig.2, the resulting features might be located probably in or near the overlapping area. The choice of belonging to a given class may be ambiguous, making it not acceptable to take these features as the most typical features of the class of interest.

B. Typicality measurement in image space

Using our definition of the typicality measurements in feature space in Eq. (2,3,4) and the relation between the image and the corresponding features in Eq. (1), an image discriminative term is built from the feature discriminative term $E_{\mathrm{d}}^{\mathcal{F}}$ via the feature extraction function f. With a slight abuse of notation, it is defined by $E_{\mathrm{d}}(I) := E_{\mathrm{d}}^{\mathcal{F}}(f(I))$, and a similar approach holds for the image generative term $E_{\mathrm{g}}(I) := E_{\mathrm{g}}^{\mathcal{F}}(f(I))$.

We assemble these two terms, weighted them using a pair of exponents, into a typicality measurement $T(I)$,

$$T(I) = E_{\mathrm{d}}(I)^{\lambda_{\mathrm{d}}} E_{\mathrm{g}}(I)^{\lambda_{\mathrm{g}}}, \quad \lambda_{\mathrm{d}}, \lambda_{\mathrm{g}} \geq 0. \tag{5}$$

We then define an *exaggerated stereotype image* is an image that maximizes this weighted combination $T(I)$.

In this paper, we have restricted ourselves to simple linear classifier and data distribution assumption, using Fisher's Linear Discriminant as the classification rule, our discriminative model, and multivariate normal distribution as the generative model to form the two terms of typicality measurement.

C. Discriminative Model: Fisher's Linear Discriminant

In Fisher's Linear Discriminant, all the data are projected down to one dimension. The optimal choice of the direction for projection is given by $\omega \propto S_w^{-1}(\mu_B - \mu_A)$, where μ_A and μ_B are the empirical means of the two classes, S_w is the total within-class covariance matrix of all the feature points x in

the training data belonging to classes A and B

$$S_w = \sum_{x \in C_A} (x - \mu_A)(x - \mu_A)^T + \sum_{x \in C_B} (x - \mu_B)(x - \mu_B)^T. \tag{6}$$

A new data point F is projected down to one dimension onto the optimal direction as $\omega^T F$ and a threshold c is used to classify the new point, where $c = \frac{1}{2}\omega^T(\mu_A + \mu_B)$. The new point will be classified as class A if $\omega^T F \leq c$ and as class B if $\omega^T F > c$.

The corresponding discriminative function describing the posterior probability is a step function (Heaviside), which however prevents the use of smooth optimization methods, so we relax it as the modified sigmoid function:

$$p(A \mid F) = \frac{1}{1 + e^{-\frac{\alpha\omega^T(F-\mu_A)}{|\omega^T\mu_A - c|}}}, \tag{7}$$

where α, ω and c are parameters. The sigmoid function evaluates 0.5 at the decision boundary, and increases when points move to the area belonging to class A, decreases when points move to the other side. We take this modified sigmoid function as the discriminative term of the typicality measurement.

D. Generative Model: Multivariate Normal Distribution

Assuming the feature points belonging to the class A are normally distributed, the value of the class density function is

$$p_A(F \mid A) = \frac{1}{(2\pi)^{M/2}|\Sigma_A|^{1/2}} e^{-\frac{1}{2}(F-\mu_A)^T\Sigma_A^{-1}(F-\mu_A)}, \tag{8}$$

where the parameters μ_A and Σ_A of the normal distribution (i.e., its mean and covariance matrix) can be estimated empirically from the training data. We use it as our generative term.

III. Building Exaggerated Stereotypes

Instead of determining an image stereotype as a maximiser of typicality measure $T(I)$ in Eq. (5), we minimize its negative logarithm $-\log T(I)$, i.e.,

$$E(I) = -\log T(I) = -\lambda_{\mathrm{d}} \log\left(E_{\mathrm{d}}(I)\right) - \lambda_{\mathrm{g}} \log\left(E_{\mathrm{g}}(I)\right). \tag{9}$$

With the terms chosen in the previous section, $E(I)$ is smooth, and we can compute an evolution/morphing of an initial image I_0 into a stereotypical one by a steepest gradient descent, or, in its discrete form.

$$I^{n+1} = I^n - \Delta t \nabla_I E, \quad I^0 = I_0, \quad \Delta t > 0, \tag{10}$$

where $\nabla_I E$ denotes the gradient of the objective function with respect to the image $\partial E/\partial I$. Assuming that we use only one feature, then by the chain rule, the gradient can be written as

$$\frac{\partial E}{\partial I} = \frac{\partial E}{\partial F}\frac{\partial F}{\partial I}. \tag{11}$$

A. Combined Fisher's Linear Discriminant and Multivariate Normal Distribution

We now write down explicitly the objective function $E(I)$ that derives from the choices that we have made for the discriminative and generative parts in the previous section: the sigmoid function in Eq. (7) to model Fisher's Linear Discriminant, and the density function of multivariate normal distribution as the generative term. Our objective function is the weighted combination of these two terms:

$$\begin{aligned} E = &-\lambda_{\mathrm{d}} \log \frac{1}{1 + e^{-\frac{\alpha\omega^T(F-\mu_A)}{|\omega^T\mu_A - c|}}} \\ &- \lambda_{\mathrm{g}} \log \frac{1}{(2\pi)^{M/2}|\Sigma_A|^{1/2}} e^{-\frac{1}{2}(F-\mu_A)^T\Sigma_A^{-1}(F-\mu_A)}, \end{aligned} \tag{12}$$

and the first term of the gradient $\dfrac{\partial E}{\partial F}$ as decomposed in Eq.(11) is

$$\frac{\partial E}{\partial F} = -\lambda_{\mathrm{d}} \frac{\alpha}{|\omega^T(F - \mu_A)|}(1 - s_A)\omega + \lambda_{\mathrm{g}}(F - \mu_A)^T\Sigma_A^{-1}. \tag{13}$$

The second part of the gradient is computed from the feature extraction function. Specific examples will be presented in the next section.

IV. Experiments and Results

We presented experiments with two different datasets. First, we applied our stereotype framework to a dataset of handwritten digits, 1s and 7s. Our second experiment used a dataset of human face images, and an Active Appearance Model as features.

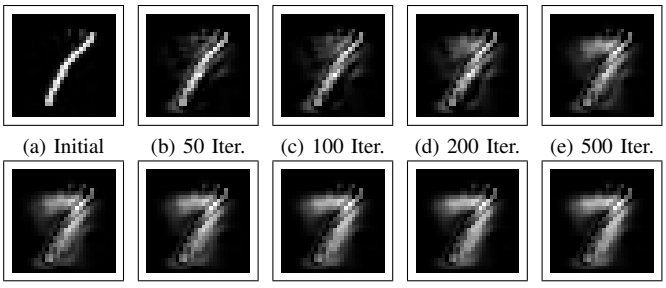
A. Handwritten Digits: 7 vs. 1

As a toy example, we applied the proposed algorithm to a subset of the MNIST handwritten digits database [15] . We used two categories of handwritten digits: class of digit 1s and class of digit 7s (the class of interest). There were 6,265 training samples of the digit 7 and 6,742 samples of the digit 1. Each image was of size 28×28 pixels and the intensity values were integer numbers in the range of $[0, 255]$. We considered the whole image as a feature.

The evolution from one sample of digit 1 towards the exaggerated stereotype of digit 7 is showed in Fig. 3. The weights of the objective function $E(I)$ (9) were set to $\lambda_d = 0.5$, $\lambda_g = 1$, and the step size for iteration was very large. The stopping criteria was that the value of the objective function was smaller than 10^{-3}. As shown in Fig. 3, during the evolution, we can see the digit inside the image got the shape more and more similar as digit seven instead of digit one.

B. Human Faces: Male vs. Female

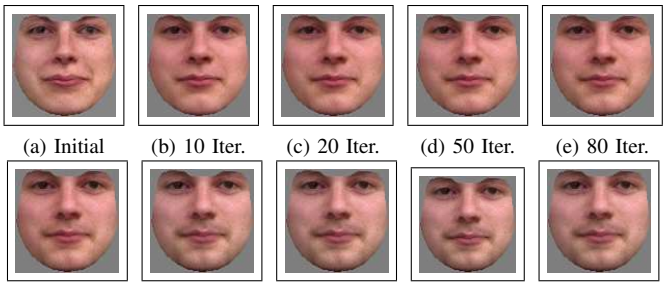
In the second experimemt, we applied the stereotype idea to a set of human face data [16]. There were 40 images representing 40 different individual faces divided into two classes, 7 female and 33 male. Each image was of size 640×480 pixels, and 58 landmarks were annotated along the



(a) Initial (b) 50 Iter. (c) 100 Iter. (d) 200 Iter. (e) 500 Iter.

(f) 800 Iter. (g) 1000 Iter. (h) 2000 Iter. (i) 3000 Iter. (j) Result

Fig. 3: The evolution starting from an image sample of digit 1 towards the exaggerated stereotype of digit 7: (a) the initial image; (b)-(i): the evolving images; (j): the result image of an exaggerated stereotype satisfying the stopping criteria.



(a) Initial (b) 10 Iter. (c) 20 Iter. (d) 50 Iter. (e) 80 Iter.

(f) 150 Iter. (g) 300 Iter. (h) 500 Iter. (i) 1000 Iter. (j) Result

Fig. 4: The evolution starting from a female face towards the exaggerated stereotype of male face: (a) the initial image; (b)-(i): the evolving images; (j): the result image of an exaggerated stereotype satisfying the stopping criteria.

eyebrows, eyes, nose, mouth and jaw. The Active Appearance Model (AAM) transform [12] was applied to each image, and we used the shape-free patches (with size of 100×100 pixels) as the training images. The parameters for combined shape and appearance data, as obtained via the AAM transform, were the features of the training images. In the experiment, we took the male faces as the class of interest, and implemented the proposed framework using the Fisher's linear discriminant and multivariate normal distribution as discriminative and generative model respectively. The weighting parameters for balancing the two models were set as $\lambda_d = 0.5$ and $\lambda_g = 1$, and the step size for iteration was set as $\Delta t = 4$.

Our goal was to build exaggerated stereotypes for male faces with different starting points. For comparison purpose, we took several female faces as initial points. A number of iterations have been taken towards the exaggerated stereotype of male faces. The stopping criterion was that the value of the objective function became smaller than 10^{-4}. In Fig. 4, one image sample of female face (the shape-free patch) is taken as the initial point for evolution towards the exaggerated male face, the gradient descent method helps to show the evolving path inside the image space. During the evolution started from the initial image, the evolving images have visualized the gradually appearing differences of faces between male and female.

In Fig. 5, the images in the top row are the initial samples of different female faces, the images in the second row are the result exaggerated stereotypes of male faces corresponding to the initial samples in the top row, and the images in the bottom row showed the differences between the initial samples and resulting stereotypes.

As we can see from Fig. 5, some characteristics of male faces in common become more obvious, such as the moustache around lip area, the shape of the jaw, and so on. The resulting images in the second row show that these shape-free patches with much higher probability of belonging to the class of male faces. In addition, these resulting images are getting similar to each other in some sense, but remaining differences tend to prove that they have evolved towards the exaggerated stereotypes of male faces through the different paths in the complex input space of faces.

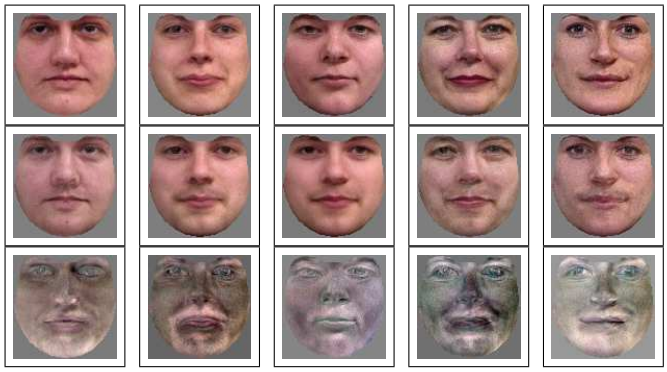


Fig. 5: The resulted exaggerated stereotypes for male faces with different starting points of female faces: the top row shows the initial female faces; the second row shows the resulted exaggerated stereotypes of the male faces; the bottom row shows the differences between the initial female faces and resulted stereotypes.

V. Discussions and Future Work

We proposed a method to build exaggerated image stereotypes for some image class based on the combination of generative and discriminative model. The resulted image stereotypes are achieved by optimizing an objective function using the gradient descent method. As our first try of building the exaggerated stereotypes, we restricted ourselves in using Fisher's linear discriminant and the multivariate normal distribution for making proof of concept, and more elaborate models will be utilized in the future work. The experiments on handwritten digits and human faces showed that the proposed framework well performed.

A. Discussions

In order to build the exaggerated stereotypes, we first defined the typicality of the class of interest. The objective function proposed in this paper was based on the definition of the typicality measurement including a generative term and a discriminative term. Using only the generative term could

lead to difficulties if there is large overlapping area in the distribution of two classes. Even if the resulted features locate near the mode of distribution, it is still ambiguous to label the features, and these features are not representative enough for the class of interest. However, in the case of only using the discriminative term, the evolution may drive the initial image too far from the stereotype. Therefore, a proper combination is a better choice.

Naturally, the built stereotype depends highly on the choices of the classifier and the assumption of class distribution, which lead to different formulation of the discriminative term E_{d} and the generating term E_{g} of the objective function. If the chosen classifier does not distinguish the two classes well, or the assumption of class distribution does not fit the data, it is impossible to form a reasonable objective function, and it is not able to offer any useful contribution to build the stereotype.

In the simple case when features only include intensities, as a result of using the combination of Fisher's Linear Discriminant and multivariate normal distribution, theoretically, the evolved feature points of exaggerated stereotype will move approaching the specific line which goes through the mode of the class of interest and also is orthogonal to the decision boundary (this line is indicated by the dashed arrow in Fig. 1). In other cases, feature extraction gives some implicit (spatial) constrains, which leads to more complicated image evolution and also avoids some anomalous situation. Furthermore, other choices of discriminative and generative models will lead to more general cases.

A good objective function for building the stereotypes should be a proper trade-off between the chosen discriminative model and generative model. As shown in Eq.(9), the proper choices of λ_{d} and λ_{g} give a balance between the classifier and the class distribution. As a result, the weighting parameters should be optimized in order to achieve the optimal results.

Given a good objective function, the optimization should be done in an effective way. In this paper, we proposed to use gradient descent method. When we implemented our method on the face data, the dimensionality of the parameters for the AAM faces was 68 and the size of the shape-free image patches was 100×100, therefore, the input space for the AAM face data was comparatively simple and the gradient descent method worked well. However, if we consider other applications with more complicated feature extractions or with much larger image size, the structure of input space would be much more complicated in higher dimensional space. In order to deal with such applications, more effective optimization methods are needed.

B. Future Work

We are trying to build exaggerated stereotypes by formulating better objective function and using more effective optimization methods. We will apply other discriminative models. A variety of choices of popular classifiers, such as k nearest neighbors (k-NN) classifier and support vector machines (SVM) can be utilized in order to handle more complicated cases. We also would like to try other assumptions of the distribution of given data to achieve better generative term. Mixture of Gaussians or Parzen estimation could be possible choices. Besides the class density function, the distance measurement between the stereotype and the modes of distribution could also be a choice. Since the objective function also depends on the trade-off of the generative term and discriminative term, we will analyze the relation between the resulted stereotype and the choices of the weighting parameters λ_{d} and λ_{g}, and then optimize the weighting parameters in order to achieve a better trade-off. Instead of gradient descent method, other stochastic optimization approaches can be used in order to solve the optimization problem in more complicated cases like the non-linear scenario.

ACKNOWLEDGMENT

The authors would like to thank Dr. Pechin Lo and Dr. Lauge Sørensen for their constructive comments and suggestions.

REFERENCES

[1] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer Verlag, 2009.

[2] F. Chang, C. Lin, and C. Lu, "Adaptive prototype learning algorithms: theoretical and experimental studies," *The Journal of Machine Learning Research*, vol. 7, pp. 2125–2148, 2006.

[3] L. Knight and S. Sen, "PLEASE: A prototype learning system using genetic algorithms," in *Proceedings of the Sixth International Conference on Genetic Algorithms*, 1995, pp. 429–435.

[4] S. Mika, B. Schölkopf, A. Smola, K. Müller, M. Scholz, and G. Rätsch, "Kernel PCA and de-noising in feature spaces," *Advances in neural information processing systems*, vol. 11, no. 1, pp. 536–542, 1999.

[5] A. Teixeira, A. Tomé, and K. Stadlthanner, "KPCA denoising and the pre-image problem revisited," *Digital Signal Processing*, vol. 18, pp. 568–580, 2008.

[6] B. Schölkopf, S. Mika, C. Burges, P. Knirsch, K.-R. Müller, G. Ratsch, and A. Smola, "Input space versus feature space in kernel-based methods," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 1000–1017, 2002.

[7] B. Schölkopf, E. Smola, and K.-R. Müller, "Kernel PCA pattern reconstruction via approximate pre-images," *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.

[8] T. Abrahamsen and L. Hansen, "Input space regularization stabilizes pre-images for kernel PCA de-noising," in *IEEE International Workshop on Machine Learning for Signal Processing*, 2009, pp. 1–6.

[9] M. Nielsen and M. Lillholm, "What do features tell about images?" *Scale-Space and Morphology in Computer Vision*, pp. 39–50, 2001.

[10] M. Lillholm, M. Nielsen, and L. Griffin, "Feature-based image analysis," *International Journal of Computer Vision*, vol. 52, no. 2, pp. 73–95, 2003.

[11] S. Zhu, Y. Wu, and D. Mumford, "Filters, random fields and maximum entropy (FRAME): Towards a unified theory for texture modeling," *International Journal of Computer Vision*, vol. 27, no. 2, pp. 107–126, 1998.

[12] T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 6, pp. 681–685, 2001.

[13] C. Bishop, *Pattern recognition and machine learning*. Springer New York, 2006.

[14] R. Duda, P. Hart, and D. Stork, *Pattern classification*. Wiley Interscience, 2001.

[15] Y. LeCun and C. Cortes. (1998) MNIST handwritten digit database. http://yann.lecun.com/exdb/mnist/.

[16] M. M. Nordstrøm, M. Larsen, J. Sierakowski, and M. B. Stegmann, "The IMM face database - an annotated dataset of 240 face images," Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, may 2004.