Scale Normalization for the Distance Maps AAM.

Denis GIRI, Maxime ROSENWALD, Benjamin VILLENEUVE, Sylvain LE GALLOU and Renaud SÉGUIER
Supélec, IETR-SCEE Team
Avenue de la boulaie, BP 81127,
35 511 Cesson-Sévigné, France
Email: {denis.giri, maxime.rosenwald, benjamin.villeneuve, sylvain.legallou, renaud.seguier}@supelec.fr

Abstract— **The Active Apearence Models (AAM) are often used in Man-Machine Interaction for their ability to align the faces. We propose a new normalization method for AAM based on distance map in order to strengthen their robustness to differences in illumination. Our normalization do not use the photometric normalization protocol classically used in AAM and is much more simpler to implement. Compared to Distance Map AAM performances of [1] and other AAM implementation which use CLAHE [2] normalization or gradient information, our proposition is at the same time much robust to illumination and AAM initialization. The tests have been drive in the context of generalization: 10 persons with frontal illumination from M2VTS database [3] were considered to build the AAM, and 17 persons under 21 different illuminations from CMU database [4] were used for the testing base.**

I. INTRODUCTION

In the context of Man-Machine Interaction, face alignment (precise detection of key elements of the face: eyes, nose and mouth) is a necessary step for face recognition, gesture analysis and lip reading. The Active Appearance Models (AAMs) initially proposed by Cootes in [5] allow to detect finely the outlines of those key elements. To do so, the AAMs combine shape information with texture information. Indeed, the AAMs are based on a priori knowledge of shapes (points of interests connected to each other) and shape-free textures of a training database. They can thus be used to generate a set of plausible representations of shapes and textures of the learned objects. They also allow the search for objects in images by an optimization process on model parameters, in order to match the model as well as possible on the image zone containing the object.

The targeted applications are meant to be used on embedded systems (mobile phones, game consoles) in both indoors and outdoors environments. However, this implies a perfect robustness to illumination variations, and since the AAMs are partially based on texture analysis, they are quite sensitive to illumination conditions. We therefore propose a new normalization procedure dedicated to the AAM which use the Distance Map instead of gray level textures and compare it to standard methods that have already been proposed to address the specific problem of illumination robustness.

In the next section, we will present the different methods proposed by the scientific community to improve the AAM robustness illumination. In section III, different well known

techniques and our proposition will be explain. The section IV will presents comparison and we will conclude in section V.

II. STATE OF THE ART

This illumination problem in AAM is far from new. Recently, [6] have proposed a two step method. They first used an edge-filtering pre-processing, using a Gabor transform, which allows the AAMs to roughly converge to the targeted face. And then, they used a patch filtering pre-processing, some kind of adaptive histogram equalization method, to refine the convergence. [1] and [7] used Distance Map instead of grey-level image texture. Distance Map (DM) are pictures in which each pixel level is equal to its distance to the nearest edge pixel. In [7], six DM per person were made from six different grayscale pictures. This allowed to create an average DM that extracts the constant features of a face, using Hausdorff Distance [8]. [1] clearly pointed out the qualities of DM when used with AAMs. Nevertheless, they implemented the classical AAM photometric normalization method which was time consuming and used some persons from CMU database to learn the model, which was tested on other persons from the same database. Therefore the AAM learned CMU background information which facilitates the convergence on the tests images.

In the next section, different techniques will be exposed before the presentation of our specific normalization.

III. PRE-PROCESSES

After having introduced the AAM, we will rapidly propose the different pre-processes to which we will compare our proposition explain in section III-C.

A. Active Appearance Models

The AAMs are a pattern recognition method proposed in the 90's (see [5] and [9]) for medical image processing and face analysis. They works in two phases:

- In the 1st phase, they learn how a human face looks like by creating a mathematical model, using a database of annotated face pictures. That model includes the memorization of the mean shape and the mean texture of a human face, and their main variation modes throw Principal Component Analysis (PCA), according to the picture learning database. A PCA on a shape training

base and a PCA on a shape-free texture training base are applied respectively in order to create the statistical shape and texture models given by the formulas:

$$x_i = x_{moy} + \Phi_x * b_x \ and \ g_i = g_{moy} + \Phi_g * b_g \quad (1)$$

with x_i and g_i are respectively the synthesized shape and texture, x_{moy} and g_{moy} the mean shape and the mean texture, Φ_x and Φ_g the matrices of eigenvectors of shape and texture covariance matrices and b_x and b_g the controlling vectors of the synthesized shape and texture. Another PCA is then applied on several examples of b which is the concatenation of b_x and b_g in order to obtain the appearance parameter c:

$$b = \Phi * c \quad (2)$$

with Φ the matrix of PCA eigenvectors. c is a vector controlling b_x and b_g (equation 2) at the same time, that is to say the shape (first part of equation 1) and texture (second part of equation 1) of the model. Then, AAMs create a regression matrix by exploring the possibilities of the model it has just created, allowing the algorithm to predict how to adapt to any situation.

- In the 2nd phase, the AAM seeks a face in a picture by minimizing the difference of the model's texture to the texture it is placed on. That minimization is done through a number of iterations of the following iterated algorithm. First, the AAMs use the regression matrix to determine which parameter has to be changed, and then, the model goes through a few tests in order to find the best value for that parameter.

If a same illumination is used with all the pictures presented to this algorithm, then it will work perfectly well. However, due to the fact that the AAMs is based on textures, any slight modification in illumination will cause a huge loss in efficiency. To solve this problem, a photometric normalization has been introduced in that algorithm. It makes sure that each picture luminosity has the same first an second order statistics [9], thus improving the efficiency of the AAMs. However, this doesn't solve the main problem, which is the sensitivity to the angle of illumination. In order to solve this problem, following solutions have been formulated: adaptive histogram equalization, gradient texture, and distance maps.

B. Pre-processing for robustness to illumination variations

1) Adaptive histogram: As Fig 1 shows it, the classical AAMs works on the original grayscale pictures and need few iterations to converge. Like [6], we perform a Contrast Limited Adaptive Histogram Equalization (CLAHE) [2]. This pre-process divides the picture into 64 tiles, creates the mapping functions associated to these 64 tiles. Each pixel is then processed using the 4 adjoining tile's mapping functions, the color of the resulting pixel being a bilinear interpolation of the results given by these 4 mapping functions. The result is a smooth picture with a much more homogeneous illumination. As it can be seen in Fig 2, the AAM preceded by a CLAHE equalization need few iterations to converge

Fig. 1. Example of a face search with the standard AAM.

Fig. 2. Example of a face search with the AAM and CLAHE process.

perfectly.

2) Gradient texture: CLAHE equalization improve the results in the context of profil or semi-profil illumination. For that reason, we extract the gradient norm of the CLAHE images to produce the model and the tests images illustrated in Fig 3. However, since the gradient-processed pictures are not subject to a high variability of luminosity, the photometric normalization described above might not be necessary. To evaluate this assumption, two sets of tests will be made: with and without normalization.

3) Distance Map: Finally, we propose to test the usefulness of Distance Maps (DM) in AAMs based face alignment methods (see [1] and [7]). These DM are created first by extracting the edges out of a CLAHE processed picture, and then by assigning to each pixel a level corresponding to its distance to the nearest edge pixel. The DM is specified according to this equation :

$$DM[i,j] = \min_{\{k,l\} \in \{E\}} (\sqrt{(i-k)^2 + (j-l)^2}) \quad (3)$$

with $DM[i,j]$ the value of the DM in $[i,j]$ coordinates pixel, and E the set of edges pixels coordinates. The idea is to provide the AAMs with edge information, which isn't sensitive to illumination. However, since the AAMs need textures to converge with a high efficiency, we provide them with a synthetic texture filed only with edge information. And although this idea is indeed related to the use of a gradient pre-process, the edge information is not limited to the pixels close to an edge, but spread throughout the entire picture. Therefore, the AAMs have a lot more informative texture to work with and can converge more easily. See Fig 4 for an example of how the AAMs converge easily on a face in a DM-processed picture.
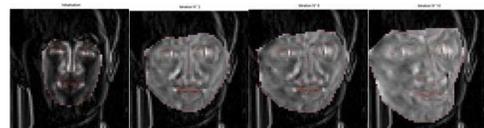

Fig. 3. Example of a face search with the AAM and Gradient process.

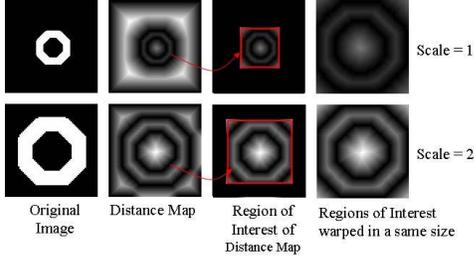Fig. 4. Example of a face search with the AAM and Distance Map process.

Fig. 5. Visualization of the problem of the Distance Map without the scale normalization.

Fig. 6. The new normalization of Distance Map.

Fig. 7. The 10 faces from M2VTS database used in the training phase of AAMs.

C. Scale Normalization

In the standard use of AAMs, photometric normalization is an important step of the algorithm. Indeed, it permits to compare all texture images despite their different gray level dynamics. At the same time, this normalization is done at every computation of the residual error which introduce complexity and increase time calculation. In addition, Distance Maps AAM of [1] have been tested with the photometric normalization while gray level dynamics only vary according to the scale of the model. Indeed, informations contained in DM are distances between edges which are larger and smaller when the scale of the model is respectively larger and smaller. So we propose to replace the standard photometric normalization by a new normalization which consists only to multiply all DM pixels by a scalar. This factor is the scale of the model which is the first element of the pose parameter. This new normalization is faster than the photometric normalization, because mean and variance of texture is no longer computed. Instead of photometric normalization, the DM is multiplied by the current model scale given by the pose parameter T:

$$DM = (T[1] + 1) * DM, \qquad (4)$$

The Fig 5 presents the standard Distance Map creation of [1] on a simple image in two scale. In the part of the figure 5 called "Regions of interest warped in a same size" we put the model in scale of one and two in the same size to have a comparison of the input image of the residual error when the same model is in scale of one and in a scale of two. We can easily see the problem of the normalization in the Region of interest: the gray level dynamics of DM of the same objet but in two different scale are very different. So, residual error is hedged.

In the Fig 6 we can see improvements of Distance Maps with the new normalization : the gray level dynamics of Distance Maps with the new normalization of the same objet but in two different scale are noticeably the same. In this example the s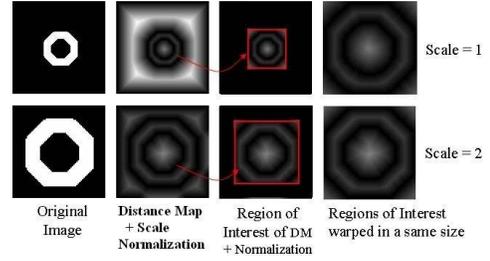cale of the model (first line of Fig 6) is multiplied by 2 (second line of Fig 6), so the new normalization consists to multiply the Distance Map of second line of Fig 6 by 0.5 because to compare this Distance Map to the Distance Map of the model (first line of Fig 6) we have a scale of 0.5.

IV. RESULTS

Our AAM algorithm creates the face model in a 64 pixels*64 pixels window so that the mean texture contains 1430 pixels.

We will compare our proposition to the pre-processes presented in section III-B in the context of generalization : the AAM will use ten persons with frontal illumination from M2VTS database (Fig 7) to learn a face model and tests will be made on seventeen persons from CMU database (Fig 8) under twenty one different illuminations leading to 357 test images.

The robustness to the illumination angle will first be illustrated, followed by some tests focused on initialization sensitivity.

A. Sensitivity to the illumination

All following graphs represent the convergence percentage versus the angle of illumination. We consider that a model has converged if the localization error of the eyes and mouth center and the nose (as pointed in Fig 9) are under a threshold of Deye/4, Deye being the distance between the eyes.

Figure 10 shows the CMU acquisition system: each person is recorded under 21 different illuminations created by a "flash system" laid out from the left to right of faces. In the next graphs, an illumination of 1 means the light came from the left of the subject, and an illumination of 21, that it came from its right; a central illumination being around 10-12.

Fig. 8. The 17 faces from PIE database used in the searching phase of AAMs displayed under the 21 different illuminations.
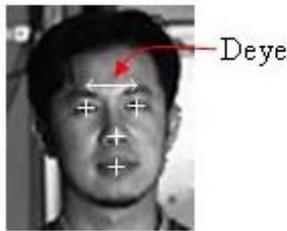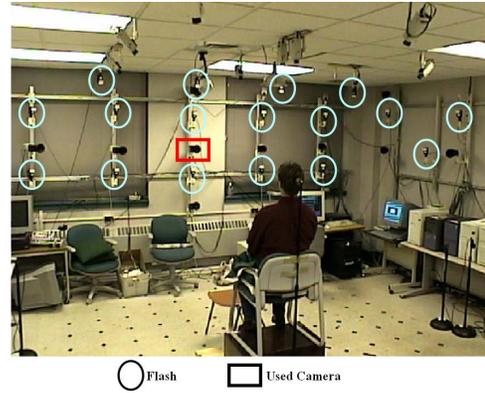

Fig. 9. Error localization.

Fig. 10. The CMU system of acquisition: positions of 17 of 21 flashes (4 left flashes are not visible in this view) and the camera used for creating facial images.
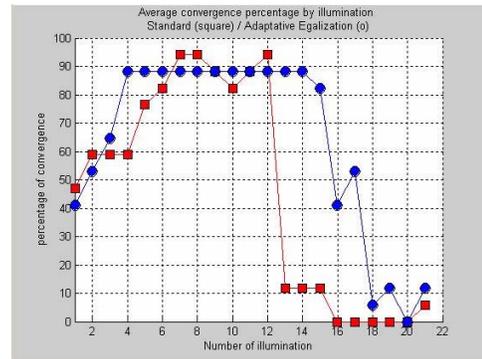

Fig. 11. Convergence comparison between the standard AAMs (square) and AAMs with CLAHE pre-process (o).
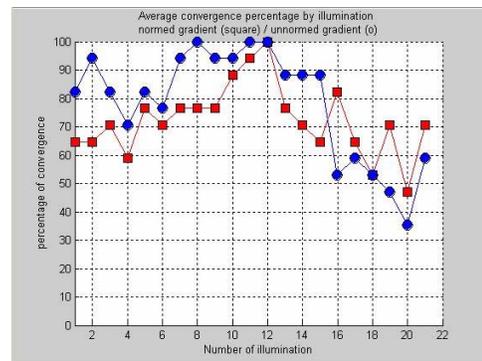

Fig. 12. Convergence comparison between AAMs with Gradient pre-process (square) and AAMs with Gradient pre-process but without photometric normalization (o).
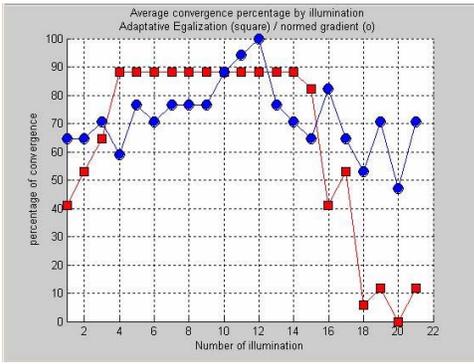
1) Grayscale pictures: Fig 11 shows how much a CLAHE pre-processing brings to the AAM's robustness to the angle of illumination. We can indeed see that, with CLAHE, the AAMs get a 90% rate of successful convergence for the 10 most central illuminations, whereas without this pre-process, the AAMs only get a 90% success rate on 4 illuminations.

2) Gradient texture: On Fig 12, we compare the use of photometric normalization for gradient-processed pictures, and as we can see, it is fairly difficult to tell which one is best. We however retain the normed-gradient pre-process for the rest of the study for its slightly better results on average.

We can also see on Fig 13 that although a gradient pre-process brings a huge improvement for the extreme

Fig. 13. Convergence comparison between AAMs with CLAHE pre-process (square) and AAMs with Gradient pre-process (o).

Fig. 14. Convergence comparison between AAMs with CLAHE pre-process (square) and AAMs with Distance Map pre-process with the new photometric normalization (o).
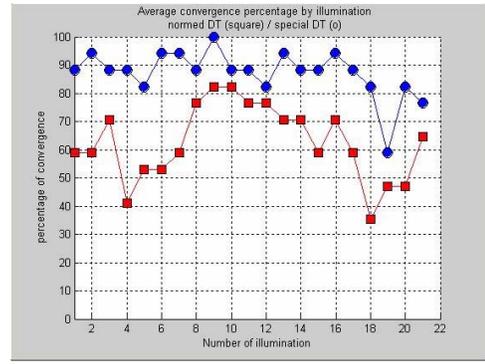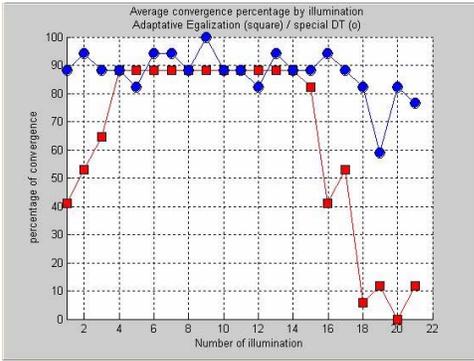

Fig. 15. Convergence comparison between AAMs with Distance Map pre-process [1](square) and AAMs with Distance Map pre-process with the new photometric normalization (o).
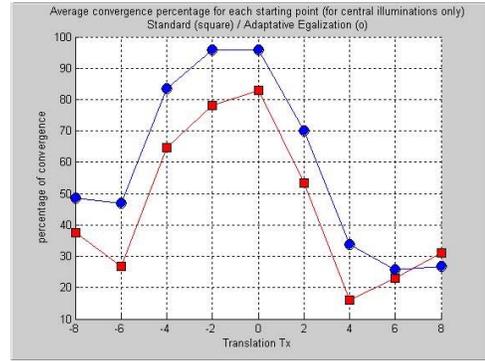

Fig. 16. Convergence comparison in translation between the standard AAMs (square) and AAMs with CLAHE pre-process (o) for only the 10 central illuminations of the PIE database faces.

illuminations, it also brings a slight loss of about 10% of successful convergence when the illumination is more central. The gain obviously comes from the fact that gradient is less sensitive to illumination. And the loss comes from the fact that the density of information is quite low since only the areas close to an edge have information, while every pixel of a grayscale picture contains information. Therefore, the AAMs will necessarily have some trouble converging precisely with gradient-processed pictures. That lead [6] to chose a two step algorithm: first close in roughly on the targeted face using gradient-processed pictures, then precisely using grayscale pictures.

3) Distance Map: However, if we use a DM pre-processing with our specific normalization, a 2 step algorithm is no longer necessary to obtain a precise convergence whatever the illumination, as we can see in Fig 14.

The mean convergence ratio is still of 90% for the central illuminations, and doesn't go below 75% for the extreme illuminations, except for 1 single illumination. But this result would not be possible without our special normalization. As we can see in the Fig 15, if we use a classical photometric normalization as in [1], we lose between 10% and 40% of successful convergence depending on the illumination considered.

B. Sensitivity to the initialization point

As we will see, these gains in term of illumination robustness come with a gain in terms of a less sensitivity to the initialization point. The X-axis in the following figures represent different initialization of the AAMs around the center of the face. We of course kept our results from the previous illumination tests, when the AAMs always started centered on the targeted face, but we also added 4 starting point on each side of that face, every 2 pixels. Cootes suggested in [10] that the model could be disrupted by 10% in translation. In our case, the face model is contained in a 64 pixels*64 pixels window so that the face model is around 40 pixels wide. So we expect to have an initialization robustness in translation of our algorithm at around 4 pixels.

As we can see in the Fig 16 and 17, a CLAHE pre-processing is as interesting concerning the problem of initialization sensitivity as it was concerning the angle of illumination. Fig 17 concerns all the 17 faces and all the 21 illuminations of the CMU database, whereas Fig 16 only deals with the 10 central illuminations where CLAHE-processed pictures had a 90% of successful convergence ratio (Illuminations 4 to 14, see Fig 11).

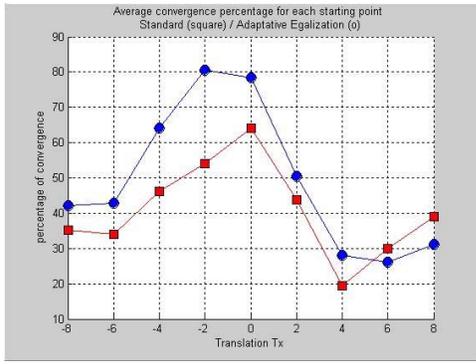Fig 18, which only deals with the central illuminations

Fig. 17. Convergence comparison in translation between the standard AAMs (square) and AAMs with CLAHE pre-process (o) for the 21 illuminations of the PIE database faces.
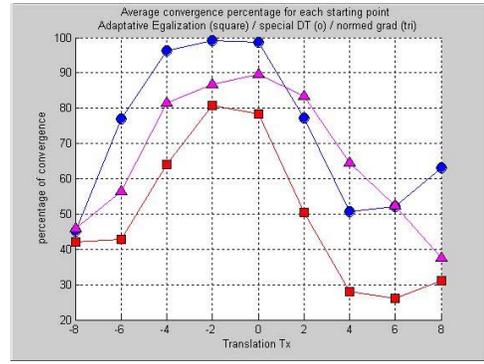

Fig. 18. Convergence comparison in translation between AAMs with CLAHE pre-process (square), AAMs with Distance Map pre-process with the new photometric normalization (o) and AAMs with Gradient pre-process (tri) for only the 10 central illuminations of the PIE database faces.

Fig. 19. Convergence comparison in translation between AAMs with CLAHE pre-process (square), AAMs with Distance Map pre-process with the new photometric normalization (o) and AAMs with Gradient pre-process (tri) for the 21 illuminations of the PIE database faces.
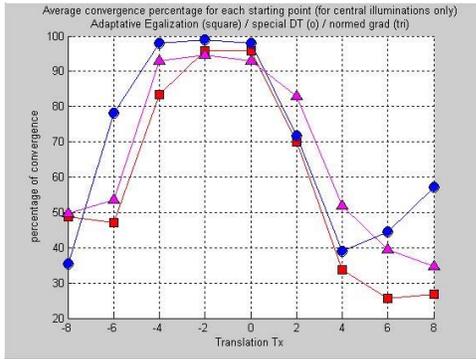
shows that our specifically normalized DM are always better than CLAHE-processed pictures and almost always better than gradient-processed pictures. In Fig 19, which deals with all illuminations, we can see this tendency clearly confirmed. We can interpret the dissymmetry of the graph by noticing that the presence of several objects in the CMU background (on the right of each persons) are quite damaging for the DM-based AAMs. As it can be noticed in Fig 8, the background of the learning base M2VTS from witch the AAM was created is uniform, the regression matrices had then not the possibility to learn the specific errors coming from the CMU background.

V. Conclusion

In this article, we have described a new normalization method for the AAM which use Distance Map instead of grey level pixels texture to improve the illumination robustness. This proposition was compared to other well known techniques dedicated to illumination sensitivity in the context of generalization, meaning that face Active Appearance Model was extracted from a specific database (M2VTS), tests being made on other persons in different environments (CMU). Our proposition is more robust both in terms of illumination angle (around 85% of successful convergence whatever the illumination), and in terms of initialization (it has an average of

10% more of successful convergence than a grayscale method, and this only on the central illuminations, where the grayscale methods still work). In the future, we might want to try and explore the possibility of having better results using an oriented edge extraction method and of course the Distance Map and specific normalization we described here.

References

[1] S. Le Gallou, G. Breton, C. Garcia, and R. Séguier, "Distance maps : A robust illumination preprocessing for active appearance models," in *VISAPP'06, International Conference on Computer Vision Theory and Applications*, 2006.

[2] K. Zuiderveld, *Contrast Limited Adaptive Histogram Equalization*. Graphics Gems IV, 1994.

[3] S. Pigeon, "M2VTS," in *www.tele.ucl.ac.be/PROJECTS/M2VTS/m2fdb.html*, 1996.

[4] T. Sim, S. Baker, and M. Bsat, "The cmu pose, illumination, and expression (pie) database," in *FG'02, IEEE International Conference on Automatic Face and Gesture Recognition*, 2002.

[5] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," in *ECCV'98, European Conference on Computer Vision*, vol. 2, 1998, pp. 484–498.

[6] Y. Huang, S. Lin, S. Z. Li, H. Lu, and H. Y. Shum, "Face alignment under variable illumination," in *FGR'04, Automatic Face and Gesture Recognition*, 2004.

[7] A. Pujol, J. J. Villanueva, and J. L. Alba, "A supervised modification of the Hausdorff distance for visual shape classification," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 16, no. 3, pp. 349–359, 2002.

[8] D. Huttenlocher, G. Klanderman, and W. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 850–863, 1993.

[9] M. B. Stegmann, "Active appearance models: Theory, extensions and cases," Master's thesis, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Aug. 2000.

[10] T. F. Cootes and C. J. Taylor, "Statistical models of appearance for computer vision," Imaging Science and Biomedical Engineering, University of Manchester, www.isbe.man.ac.uk, Tech. Rep., March 2004.