Parsing radiographs by integrating landmark set detection and multi-object active appearance models

Albert Montillo*, Qi Song*, Xiaoming Liu, James V. Miller

GE Global Research Center One Research Circle, Niskayuna, NY, 12309 USA
*Authors contributed equally

ABSTRACT

This work addresses the challenging problem of parsing 2D radiographs into salient anatomical regions such as the left and right lungs and the heart. We propose the integration of an automatic detection of a constellation of landmarks via rejection cascade classifiers and a learned geometric constellation subset detector model with a multi-object active appearance model (MO-AAM) initialized by the detected landmark constellation subset. Our main contribution is twofold. First, we propose a recovery method for false positive and negative landmarks which allows to handle extreme ranges of anatomical and pathological variability. Specifically we (1) recover false negative (missing) landmarks through the consensus of inferences from subsets of the detected landmarks, and (2) choose one from multiple false positives for the same landmark by learning Gaussian distributions for the relative location of each landmark. Second, we train a MO-AAM using the true landmarks for the detectors and during test, initialize the model using the detected landmarks. Our model fitting allows simultaneous localization of multiple regions by encoding the shape and appearance information of multiple objects in a single model. The integration of landmark detection method and MO-AAM reduces mean distance error of the detected landmarks from 20.0mm to 12.6mm. We assess our method using a database of scout CT scans from 80 subjects with widely varying pathology.

Keywords: automatic landmark localization, organ localization, image parsing, radiograph, active appearance model, rejection cascade

1. INTRODUCTION

Parsing anatomical images entails the identification of scan content and localization of salient structures. Fully automated parsing is a critical first step that facilitates subsequent finer scale analyses, such as precise segmentation. Our goal is to enable the automatic parsing of 2D radiographs from ubiquitous *routine* clinical scans. We hypothesize that the integration of anatomical landmark subset detection and multi-object active appearance models which learn complementary local information and different global information will be well suited to the task and will improve parsing as measured by refined landmark extraction accuracy. We further hypothesize that both false positive as well as false negative landmarks can be corrected by learning a geometric landmark constellation model for subsets of landmarks across a training database. Our database consists of 80 subjects from whom a 2D anterior-posterior projection scout image was acquired with a computed tomography (CT) scanner. The subjects vary in age (18 to 75 years), gender, and have widely variable pathology including obesity, lung cancer, cardiomyopathy, and liver diseases. Additional pathological variability includes metallic implants: cardiac stents, hip and knee implants, vertebrae screws and cardiac pacemakers. These images gathered from multiple clinical sites have been acquired with widely variable protocols, including large variation in the Z range (body coverage) included in the scan, with image dimensions from 219 to 1357mm in height and 484mm in width. Representative images are shown in Fig. 1.

Landmark detection has been used to previously to parse radiographs[1] however false negatives are not inferred nor are the detections refined with an active appearance model. We show that these steps substantially improve accuracy. Methods for 3D CT volume parsing based on landmark detection have been presented.[2,3] In[2] landmark detections are refined by a search over exemplar cross-correlation maps while in[3] detections are refined by an active shape model (ASM). Neither is directly applicable to radiographs where projective image formation causes

Send correspondence to Albert Montillo. E-mail: montillo@ge.com Telephone: (518)387-4791

Medical Imaging 2013: Image Processing, edited by Sebastien Ourselin, David R. Haynor,
Proc. of SPIE Vol. 8669, 86690H · © 2013 SPIE · CCC code: 1605-7422/13/$18
doi: 10.1117/12.2007138

multiple structures to overlap confounding direct application of ASMs and where non-Hounsfield pixel intensities render cross-correlation maps problematic. Lastly, compared to other SIFT and part-based parsing methods,[4] our method directly models the relationship between detections and the global appearance variation via the MO-AAM.

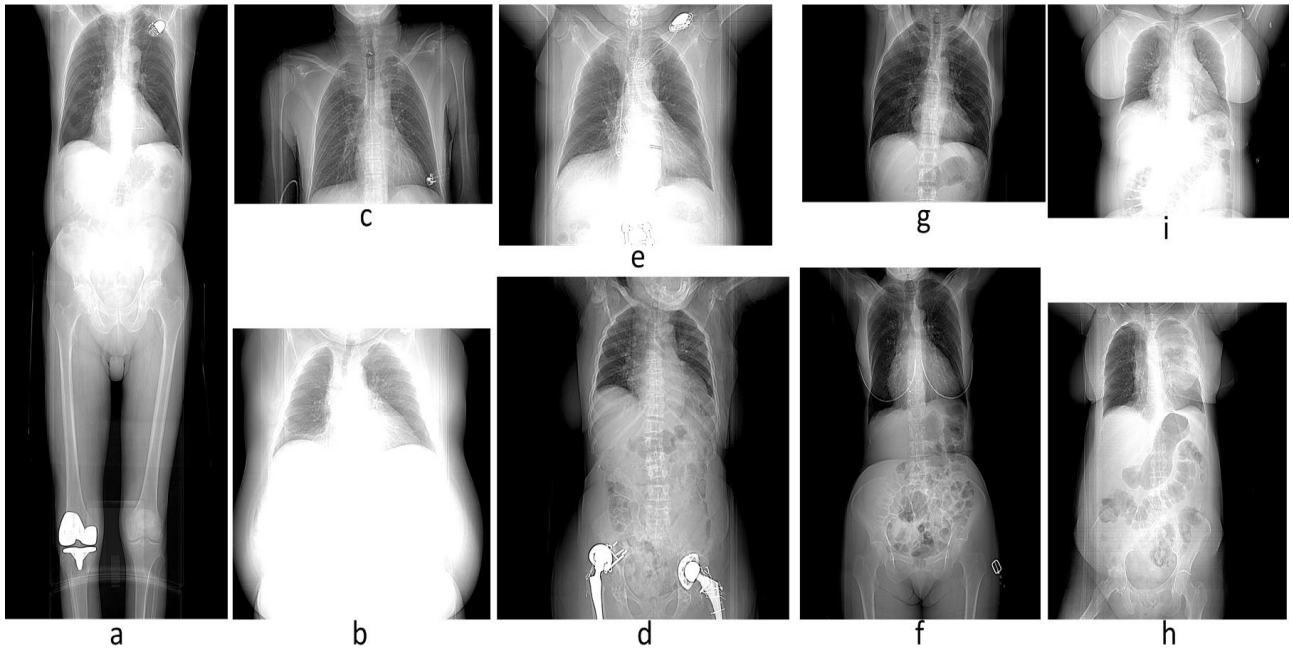


Figure 1. Scan variability. Scan range variability: (a) full body, (b,i) chest+abdomen, (d,f,h) chest through pelvis, (c,e,g) chest. Pathological variability includes: (a,c,e,d) implants, (i,b,h) obesity and lung disease. Imaging variability: (b,c) low and high contrast.

2. METHODS

Our method consists of two steps. First, a *discriminative+generative* landmark and constellation subset detection model is applied, which provides initial locations of salient landmarks. Then, an active appearance model approach is applied, in which both the shape and the appearance information of multiple organs and their relative context are encoded in a *single* model.

2.1 Landmark Detection

We adapt the discriminative rejection cascade classifier framework[5] to detect anatomical landmarks in radiographs. One rejection cascade classifier is independently trained for each landmark via supervised learning. Our cascades are built using Gentle Adaboost[6] for feature selection and classification. Each cascade is then applied as a sliding window classifier to determine if and where a particular landmark is present in a novel image.

To train the detectors, the images are annotated with landmarks manually as shown in Fig. 2(a), including lung landmarks: lower corner (3,8), top of diaphragm (2,9), lung left side (17) and right side (18), intersection of top rib with lung boundary (4,7), top-left point for left lung (19), and top point (5,6); and heart landmarks: corner of heart in right lung (1), mid-heart in right (14) and left (10) lung, top of heart in right (13) and left lung (11), bottom of heart along spine (15, 16) and sternum just above heart (12). To define positive patches, square patches are cropped from each image for each landmark that are large enough to include visible anatomical structure. To define negative patches, square patches not overlapping the positive exemplar are randomly cropped. Features are computed using an extended set of Haar templates, shown in Fig. 2(b), by computing the difference of the sum of pixel intensities in the black sub-rectangle(s) and the sum of the pixels in the white sub-rectangle(s).

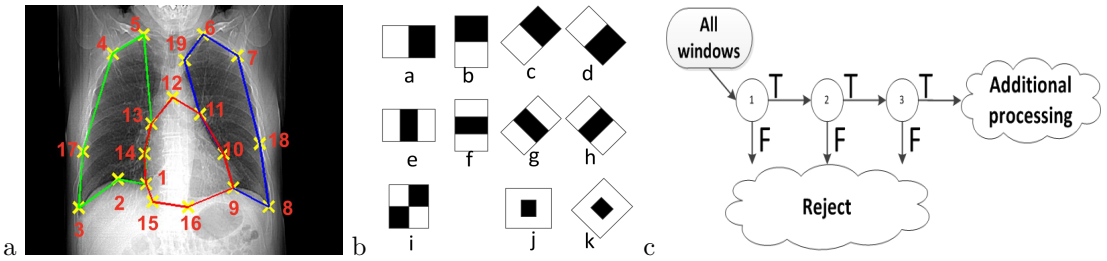


Figure 2. (a) Training landmarks for approximate regions of right lung (green), left lung (blue) and heart (red). (b) Haar features include (a-d) edge features, (e-h) line features, (i) diagonal line feature (j-k) center surround, (c) Rejection cascade.

We build a rejection cascade, as illustrated in Fig. 2(c), such that each cascaded stage achieves a high true positive rate of 99.5% with a reasonable false positive rate of 50%. The first stage uses all of the positives and negatives. Subsequent stages are trained using examples which pass through *all* previous stages (classified as positive). Stages are appended until an overall true/false positive rate is achieved or a maximum number of stages (14) is reached.

Applying the set of individual discriminative landmark detectors described above yields a set of candidate detections, C. To correct false negative and positive candidates we build a generative model of the geometric configuration of the landmarks. Given our complete set of landmarks, S, (Fig 2(a)) for every subset, $s_i \subset S$, of two distinct landmarks, we learn the parameters, $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ of the multi-variate Gaussian distribution for a third landmark, $q \notin s_i$ across the set of training images using maximum likelihood estimation.

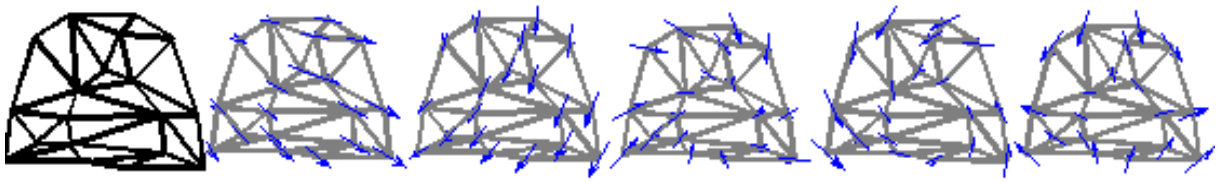
For a landmark t with multiple candidate detections, denoted $D = \{d_j\} \subset C$, we retain only the single candidate d^* with the lowest uncertainty estimated by its median Mahalanobis distance from the expected location of each size-2 subset, c_k, of distinct landmark candidates in $S \setminus t$. That is $c_k \subset C$ and c_k is comprised of two candidates $c_k = \{c_{m_1}, c_{m_2}\}$ where c_{m_1} is a candidate for landmark ℓ_1 and c_{m_2} is a candidate for landmark ℓ_2 while $\ell_1, \ell_2 \in S \setminus t$, and $\ell_1 \neq \ell_2$.

To further facilitate subsequent processing we also use our model to infer missing landmarks. Assuming normal basic human anatomy (two lungs and a heart), missing landmarks are those not found in an image by our detectors. Given that C is the set of candidate detections, and denoting the set of landmarks spanned by C as L, then the missing landmarks are $M = S \setminus L$. For each missing $m \in M$, we estimate its location, $\mathbf{x}$, based on predictions from the detected candidates. For each subset $c_k \subset L$ of 2 candidates for distinct landmarks, we infer one predicted location $\mathbf{x}$ using the mean offset, $\boldsymbol{\mu}$, from c_k learned from the training data. We estimate the final location using the trimmed mean of the central 50% for each element of all the predicted locations of m.

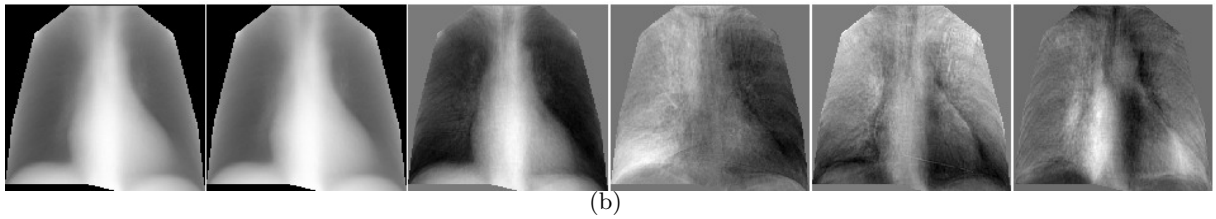
2.2 Multi-object Active Appearance Model (MO-AAM)

Active appearance model (AAM) based approaches have been applied for many computer vision applications.[7,8] We adapt the AAM to localize multiple organs based on our initial landmark detections. A single model is constructed encoding the shape and appearance information for both lungs and heart, which allows simultaneous localization of multiple regions. Our AAM approach has two parts: model learning and model fitting. In model learning, one shape model and associated appearance model are trained for the multiple objects based on the manually-labeled radiographs. Our model vertices are the same landmarks used to train the landmark detectors in Section 2.1. Given n landmarks for each training image, the shape is represented by vector $s = [x_1, y_1, x_2, y_2, \cdots, x_n, y_n]^T$, where (x_i, y_i) is the coordinate of the i^{th} landmark. The shape model is then defined by a $2n$ dimensional Gaussian distribution of landmarks. After applying a Principal Component Analysis (PCA), any shape can be represented by $s = s_0 + \sum_{i=1}^{n} p_i s_i$, where s_0 is the mean shape, s_i is the i^{th} shape basis vector and p_i is the corresponding shape coefficient. Fig. 3(a) shows the trained shape model.

After the shape model is trained, a warping function $W(x, y; p)$ is defined, which takes the pixel (x, y) in the mean shape s_0 and maps it to the location $W(x, y; p)$ in the image observation based on the learned shape coefficients $p = [p_1, p_2, \ldots, p_n]^T$. Given the learned shape model, each training image is warped to the mean shape based on the above warping function. A second PCA analysis is then applied for transformed appearances



(a)



(b)

Figure 3. The mean and top 5 basis vectors of (a) the shape model and (b) the appearance model sequenced by corresponding vector coefficients.

from all training images. Any appearance A can be represented by $A(\mathbf{x}) = A_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i A_i(\mathbf{x})$, where $\mathbf{x}$ is the set of all pixel coordinates inside the mean shape s_0, A_0 is the mean appearance, A_i is the i^{th} appearance basis vector and λ_i is the corresponding appearance coefficient. Fig. 3(b) illustrates the trained appearance model.

In the model fitting step, we apply the learned shape and appearance model to fit the test radiographs. This is achieved by finding the optimal shape and appearance coefficients such that the difference between current appearance estimation and the target image is minimized. We initialize the MO-AAM's vertices to the detected landmarks, then fit the model by minimizing the expression

$$\sum_{\mathbf{x}} \left\| A_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i A_i(\mathbf{x}) - I(W(\mathbf{x};p)) \right\|^2 \tag{1}$$

with respect to the shape coefficients $p = [p_1, p_2, \ldots, p_n]^T$ and appearance coefficients $\lambda = [\lambda_1, \lambda_2, \ldots, \lambda_m]^T$. Here $I(W(\mathbf{x};p))$ denotes the warped image observation and the expression defines the squared error between the synthesized appearance instance and the warped observation.

The optimization can be solved by the Simultaneous Inverse Compositional method (SIC).[9] Compared to the classical gradient descent based optimization, the key idea of SIC method is to change the role of appearance model and the image observation, which allows the pre-computation of the time-consuming steps for parameter estimation while still retaining the fitting quality.

3. EXPERIMENTS AND RESULTS

The proposed method was validated on 80 radiographs using a 4-fold cross validation with for each fold, 60 datasets used for training and remainder for testing. In qualitative evaluation we observe that our landmark detection method handles the extreme variability seen in clinical scans including field of view variations (Fig. 4(a-c) and anatomical and pathological variations (dense lungs, Fig. 4(b), obesity, Fig. 4(c), multiple implants Fig. 4(a), as well as healthy subjects Fig. 4(e)). Ground truth landmarks are dark blue while our method's detections are in green and yellow. In total, 145 cases of multiple detections per landmark (false positives) were corrected. The candidate selected from the multiple detections was 15.6mm closer to the ground truth on average than the next best candidate. In total, 128 missing (false negative) landmarks were recovered using the constellation model with a mean distance error of 20.1mm. Examples of missing landmarks are shown in light blue. Including the 1392 detected landmarks, the overall mean landmark distance error of the landmark detection stage is 20.0mm. This provides an excellent initialization for our MO-AAM.

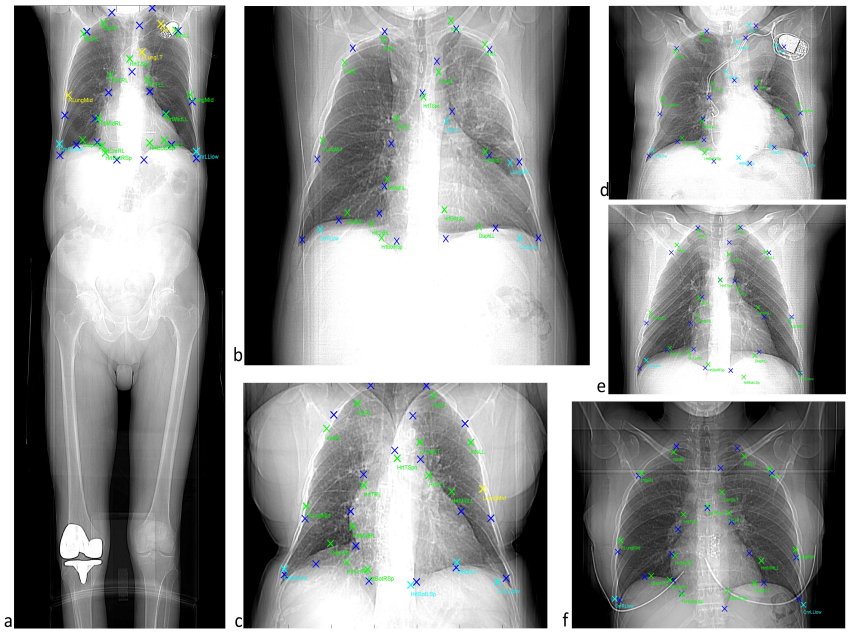


Figure 4. Landmark detection results. Automatically detected landmarks in green and yellow, inferred landmarks in cyan, manually identified landmarks (ground truth) in blue. Proposed method handles: (a) large field of view scan with pacemaker and knee implant (b) lung disease and (c) obese subjects. (d-f) Close-up of detections in three additional subjects.

The MO-AAM fitting results are shown in Fig. 5, with initial mesh from landmark detections in black dashed lines and final fitting results in light blue. Approximate regions for right and left lungs and heart are indicated by green, blue and red lines, respectively. We observe significant improvement of landmark detections (arrows) due to the incorporation of global shape and appearance.

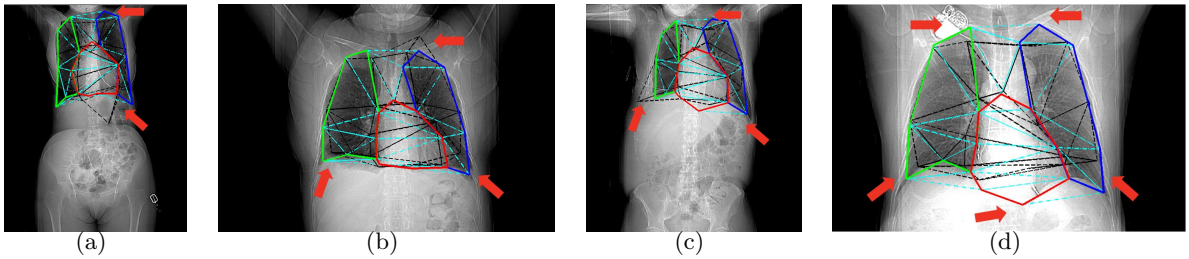


Figure 5. Typical examples of automatic organ localization results for the right lung (green), the left lung (blue) and the heart (red). Both initial landmark detection results (black dash) and AAM model fitting results (cyan dash) are shown.

The mean distance error between the ground truth and each landmark across the 80 test images are shown in Fig. 6. Red bars show distance error using landmark detection, blue bars show error after MO-AAM fitting. Errors are significantly reduced for through the MO-AAM fitting; overall the mean distance error is reduced from 20.0mm to 12.6mm.

4. CONCLUSIONS

In this work we address the core task of parsing radiographs into salient structures. We learn *local* models of appearance and geometry using rejection cascade landmark detectors and *global* models of geometry using landmark constellation subset models. Meanwhile, we construct a multi-object AAM model, which learns *global* shape and appearance information for multiple regions jointly. We show how to use the constellation model for

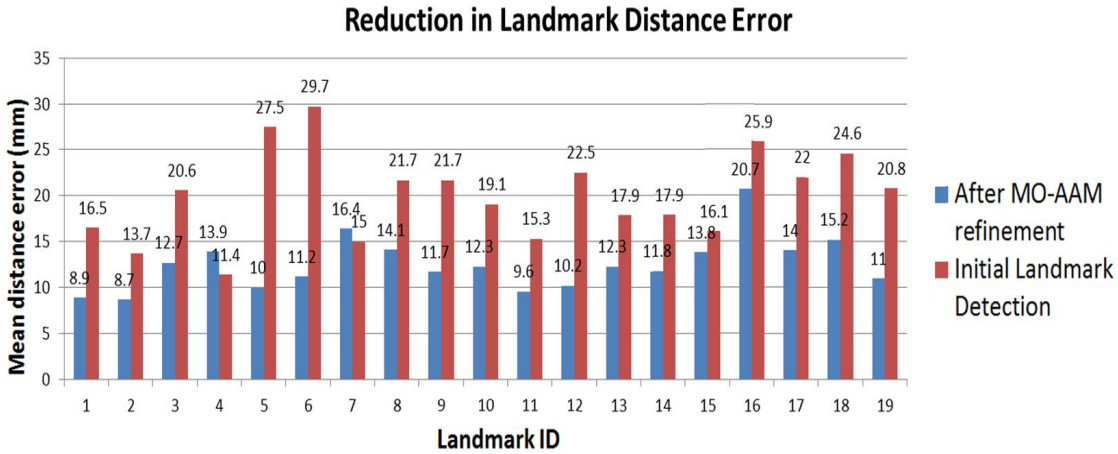


Figure 6. Combining landmark detection and MO-AAM reduces landmark distance error.

false positive and negative recovery which enables handling anatomical and pathological variability found in the clinic. Lastly, we describe how to fit our MO-AAM guided by the landmarks for simultaneous localization of multiple regions and show how the combination reduces overall distance error from 20.0mm to 12.6mm.

ACKNOWLEDGMENTS

This work was supported in part by NIH P41 EB015902.

REFERENCES

[1] Tao, Y., Peng, Z., Krishnan, A., and Zhou, X., "Robust learning-based parsing and annotation of medical radiographs," *Medical Imaging, IEEE Transactions on* **30**(2), 338–350 (2011).

[2] Potesil, V., Kadir, T., Platsch, G., and Brady, M., "Personalization of pictorial structures for anatomical landmark localization," in [*Information Processing in Medical Imaging*], 333–345, Springer (2011).

[3] Seifert, S., Barbu, A., Zhou, S., Liu, D., Feulner, J., Huber, M., Suehling, M., Cavallaro, A., and Comaniciu, D., "Hierarchical parsing and semantic navigation of full body ct data," *Medical Imaging* **7259**, 02:1–8 (2009).

[4] Toews, M. and Arbel, T., "A statistical parts-based model of anatomical variability," *Medical Imaging, IEEE Transactions on* **26**(4), 497–508 (2007).

[5] Viola, P. and Jones, M., "Robust real-time face detection," *International journal of computer vision* **57**(2), 137–154 (2004).

[6] Freund, Y. and Schapire, R., "Experiments with a new boosting algorithm," in [*Machine Learning: Proceedings of the Thirteenth International Conference*], 148–156, Morgan Kauman (1996).

[7] Cootes, T., Edwards, G., and Taylor, C., "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**(6), 681–685 (2001).

[8] Liu, X., "Discriminative face alignment," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **31**(11), 1941–1954 (2009).

[9] Baker, S. and Matthews, I., "Lucas-kanade 20 years on: A unifying framework," *International Journal of Computer Vision* **56**(3), 221–255 (2004).