Robust Facial Landmark Detection and Face Tracking in Thermal Infrared Images using Active Appearance Models

Marcin Kopaczka, Kemal Acar and Dorit Merhof
Institute of Imaging and Computer Vision, RWTH Aachen University, Templergraben 55, Aachen, Germany

Keywords: Thermal Infrared, Face Tracking, Facial Landmark Detection, Active Appearance Model.

Abstract: Long wave infrared (LWIR) imaging is an imaging modality currently gaining increasing attention. Facial images acquired with LWIR sensors can be used for illumination invariant person recognition and the contactless extraction of vital signs such as respiratory rate. In order to work properly, these applications require a precise detection of faces and regions of interest such as eyes or nose. Most current facial landmark detectors in the LWIR spectrum localize single salient facial regions by thresholding. These approaches are not robust against out-of-plane rotation and occlusion. To address this problem, we therefore introduce a LWIR face tracking method based on an active appearance model (AAM). The model is trained with a manually annotated database of thermal face images. Additionally, we evaluate the effect of different methods for AAM generation and image preprocessing on the fitting performance. The method is evaluated on a set of still images and a video sequence. Results show that AAMs are a robust method for the detection and tracking of facial landmarks in the LWIR spectrum.

1 INTRODUCTION

Algorithms for the analysis of face images are a key research area in computer vision. A large number of methods for detection, tracking, recognition and expression analysis of faces have been published in the last years. While most of the methods introduced in this field are aiming at regular photographs and videos, it is known that several frequency ranges outside the visual spectrum allow interesting applications that cannot be realized using visible light. Long wave infrared (LWIR) imaging is one of the domains that have gained increased attention in recent years. This subband is referred to as *thermal infrared* as the human body emits most of its heat in this range of the electromagnetic spectrum. This allows LWIR sensors to work independently from lighting conditions and to operate even in complete darkness. Besides of having the advantage of being invariant to illumination, LWIR sensors also allow the extraction of information from an image that is not easily detectable by sensors in the visible domain. In face images they reveal information on the subcutaneous vascular structure (Zhu et al., 2008) or vital signs such as respiratory rate (Lewis et al., 2011) and heart rate (Gault and Farag, 2013). These key properties form the basis for two major applications of LWIR imaging: Biometric face recognition for person identification and the extraction of temperature signals for medical purposes and affection state analysis. A number of publications has addressed both topics in recent years, an overview of each area can be found in the recent surveys by (Ghiass et al., 2014) and (Lahiri et al., 2012). However, many of the authors of studies that included analysis of face images have acquired their data under strongly controlled conditions that restrict head movement. The main reason for such controlled environments is the lack of established and robust face tracking methods in the thermal infrared. Therefore, recorded persons are required to minimize head movement in order to allow undisturbed data extraction from defined regions of interest (ROIs). The lack of tracking solutions can be attributed to the fact that the appearance of faces in the LWIR spectrum differs strongly from their appearance in the visual domain. LWIR images generally have lower contrast and do not reproduce any skin texture, so that many well-established tracking algorithms developed for the visual domain do not perform well when applied directly to LWIR data. Therefore, head movement is often restricted from the beginning. Methods that attempt face tracking in LWIR data are currently limited to the tracking of exclusive salient regions such as the nose or the inner corners of the eyes, which are

Kopaczka, M., Acar, K. and Merhof, D.
Robust Facial Landmark Detection and Face Tracking in Thermal Infrared Images using Active Appearance Models.
DOI: 10.5220/0005716801500158
In *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2016) - Volume 4: VISAPP*, pages 150-158
ISBN: 978-989-758-175-5

usually both easy to find due to their temperature signature but at the same time allow only limited robustness towards partial occlusion and out-of-plane rotation. To the best of our knowledge, no author has ever proposed a holistic approach to track whole faces and a complete set of facial landmarks in the thermal infrared.

Real-world scenarios require robust methods for precise facial landmark detection. Several authors have mentioned that advanced tracking methods would increase the range of possible applications of their LWIR image processing algorithms. Furthermore, it is known from research in the visual domain that holistic approaches offer wider support and are therefore more robust to unexpected pose changes, fast head movement and partial occlusion than single-ROI trackers. In this work, we therefore introduce a LWIR face tracker based on an active appearance model (AAM). We describe the face database used for tracker training as well as several approaches to improve fitting precision in low-contrast LWIR data by applying contrast-enhancing preprocessing to the images. We extensively evaluate the fitting performance of the AAM using different state-of-the-art fitting algorithms together with recent improvements in feature-based AAM training. To prove the versatility of our approach, we additionally show that our AAM can be used to track a previously unseen face in the LWIR in a video sequence with large head movement.

2 PREVIOUS WORK

Basic methods for face tracking in the LWIR domain are face detection and segmentation algorithms such as (Filipe and Alexandre, 2013). They often rely on the fact that LWIR sensors measure heat radiation and therefore faces are often easy to locate using thresholding and basic morphological operators. Despite their low computational complexity, these algorithms often perform well for face segmentation tasks in the thermal domain. However, they are not able to perform precise landmark detection on facial regions such as eyes or mouth. More advanced approaches employ a basic segmentation and add a feature detection step to the process. Current algorithms often perform landmark detection by locating temperature maxima which can usually be found in the inner corner of the eyes (Alkali et al., 2014). These methods generally assume a frontal view of the face and their performance degrades quickly when confronted with out-of-plane rotation.

Another group of current trackers suitable for landmark tracking in the thermal infrared are single-ROI trackers, either general-purpose algorithms such as TLD (Kalal et al., 2012) or complex approaches developed especially for thermal IR tracking, for example (Zhou et al., 2013). While showing good performance in scenarios with little movement, the limited support area of single-ROI trackers is a downside that leads to poor tolerance in case of ROI occlusion or fast movement. Only very little research has been published on the use of multi-point trackers that could counter these downsides. (Dowdall et al., 2007) demonstrated the use of a coalitional multi-point tracker to track faces in the LWIR domain. (Ghiass et al., 2013) was the first and up to now only person to use AAM in the thermal infrared domain, however for face recognition and not for tracking purposes. Instead, the research presented there was focused on algorithms that allow increasing image contrast and extract person-specific biometric information from the data. The AAM was trained for recognition tasks on single images, therefore it was not investigated if it could be used for robust face tracking. Furthermore, the proposed methods are focused on the identification of known persons and therefore contain no information on the model's ability to generalize towards unseen faces. To account for these issues, we show in our work that our AAM is well able to track faces in a largely unconstrained setting and that the model can robustly adapt to unseen individuals, both of which have not been addressed in existing publications on thermal AAM so far.

One of the current research areas for thermal imaging is the extraction of biosignals from face images. Several methods for measuring respiratory action (Lewis et al., 2011), blood vessel location (Zhu et al., 2008) and cardiac pulse (Gault and Farag, 2013) have been published in recent years. A common approach to measure the respiratory rate is temperature monitoring of the nostril area as respiration-induced temperature changes can be measured in this region with great reliability. As in the general case, current approaches here also either rely on single-ROI trackers or on an acquisition protocol that prohibits head movement.

In the following, we will introduce and evaluate methods for training AAMs in the thermal infrared with the focus on robust facial landmark detection and face tracking in unconstrained video sequences. In our work we introduce novel image preprocessing approaches and also evaluate the applicability of state-of-the art fitting methods and recent advances in feature-based AAM representations to LWIR images. We will show that our approach is able to reliably detect and track facial ROIs even in challenging scenarios including fast movement and significant out-of-

plane rotation. The method's ability to robustly generalize towards unseen faces will be demonstrated as well. To the best of our knowledge, our work demonstrates for the first time that AAMs are a viable solution for robust LWIR face tracking and at the same time it is the first solution that allows precise detection of facial landmarks at such a wide range of realistic and arbitrary head poses in thermal infrared face images.

3 METHODS

Generative methods such as AAM require an extensive set of training images to allow modeling of unseen faces. In this section, we will therefore first describe our steps to create a thermal face database that can be used to train facial landmark detectors. We then describe how we used the database and applied dense image features combined with contrast-enhancing preprocessing to train an AAM-based facial landmark detection system for the thermal infrared domain.

3.1 Thermal Face Database

We have created a face database with LWIR video sequences of currently 31 (25 male, 6 female) subjects. The persons were asked to perform a set of defined and arbitrary actions that cover a large range of poses and facial expressions. All videos were taken with a microbolometer-based LWIR camera with a relative thermal resolution of 0.03K and acquired at the sensor's native spatial resolution of 1024x768 pixels. Each subject was recorded and had to pose for at least two recordings of 40 seconds. During the first recording the participants followed a defined head movement to cover a wide range of head poses as shown in Figure 1. In the second recording, the volunteers were asked to perform arbitrary head movement and facial expressions (Figure 2). From these videos we extracted a total of 695 frames. In a next step, all selected frames were manually annotated with a 68-point template to precisely indicate the position of facial regions such as mouth, eyes and nose.

3.2 Image Preprocessing

It has been suggested by (Ghiass et al., 2013) that contrast-enhancing preprocessing of LWIR images could have a positive impact on the fitting performance of an AAM. The method proposed there was based on smoothing the input image with an anisotropic diffusion filter and subtracting the diffusion result from the original, thereby enhancing the edges and high-frequency components. We extend the suggested ideas by implementing and evaluating a group of sharpening filters based on the unsharp mask concept.

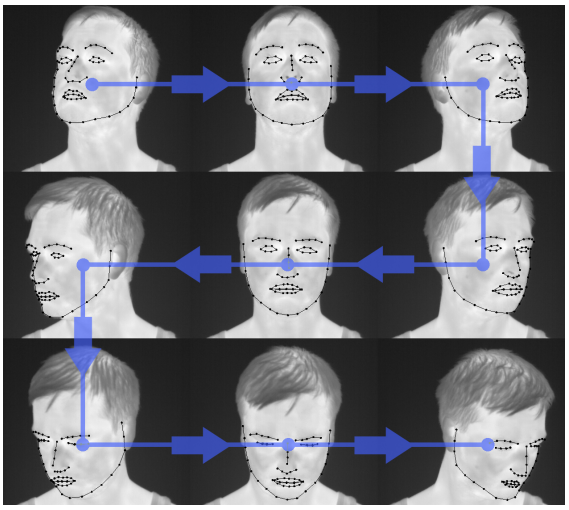


Figure 1: Posed head poses example.

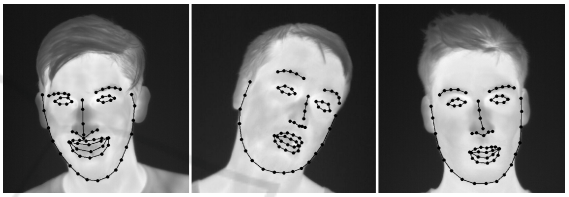


Figure 2: Spontaneus head pose and facial expression examples.

3.2.1 Unsharp Mask

In unsharp masking (USM), a high-pass filter is implemented by smoothing the image $I(x,y)$ with a low-pass filter $G(x,y)$ and subsequently subtracting the filter result from the original image, leaving only the image's high-frequency components:

$$I_{filtered}(x,y) = I(x,y) - I(x,y) * G(x,y) \quad (1)$$

Subsequently, the final sharpened image I_s is obtained by adding the filtered image I_f with with a weight factor k to the original image I:

$$I_s(x,y) = I(x,y) + k\,I_f(x,y) \quad (2)$$

The lowpass is commonly implemented using a Gaussian kernel:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (3)$$

Since the general concept of unsharp masking is not restricted to Gaussian kernels, we introduce two addi-

tional kernels based on anisotropic diffusion and bilateral filtering. The kernels are applied to the image and the results are fed into the USM algorithm according to Equations 1 and 2.

3.2.2 USM with Anisotropic Diffusion

Anisotropic diffusion filters offer edge-preserving image smoothing by blurring the image along edges. A commonly applied anisotropic diffusion filter is defined by

$$\frac{\partial \mathrm{I}}{\partial t} = \mathrm{div}\left(g\left(\parallel \nabla \mathrm{I} \parallel\right) \nabla \mathrm{I}\right), \tag{4}$$

with

$$g\left(\parallel \nabla \mathrm{I} \parallel\right) = e^{-\left(\frac{\parallel \nabla \mathrm{I} \parallel}{K}\right)^2}, \tag{5}$$

where K is a parameter controlling the sensitivity to edges in the image. The smoothed result is then subtracted from the original image as in Equation 1.

3.2.3 USM with Bilateral Filtering

Just as anisotropic difusion, the bilateral filter is another class of edge-preserving smoothing operation. In contrast to regular (unilateral) Gaussian blurring, the smoothing term of a bilateral filter depends not only on the spatial pixel distance, but also on the intensity difference between pixels. This means that the appearance of the filter kernel depends on the local image content. The filter to compute the new image intensity I_f at pixel coordinate x from the original image I is defined as

$$\mathrm{I}_f(x) = \frac{\sum_{x_i \in \Omega} \mathrm{I}(x_i)\, G_r(\parallel \mathrm{I}(x_i) - \mathrm{I}(x) \parallel)\, G_s(\parallel x_i - x \parallel)}{\sum_{x_i \in \Omega} G_r(\parallel \mathrm{I}(x_i) - \mathrm{I}(x) \parallel)\, G_s(\parallel x_i - x \parallel)}, \tag{6}$$

where Ω is the kernel window centered in x and G_r, G_s are Gaussian filters that are applied in the range and position domain. Again, we apply unsharp masking by performing the smoothing operation and subtracting the result from the original input image.

3.3 Active Appearance Models (AAM)

Active Appearance Models were originally introduced by (Cootes et al., 2001) and substantially extended in the work by (Matthews and Baker, 2004) with the introduction of the inverse compositional (IC) fitting algorithm. AAMs are a state-of-the-art method for landmark detection and mostly used to model faces in photographs or anatomical structures in medical images. An AAM is a generative method to model an instance of an object which makes it possible to detect object landmarks and at the same time acquire information on the properties of the modeled object.

AAMs are trained with a manually annotated database. For facial landmark detection, the database contains images of persons with added landmarks for facial regions such as eyes, nose or mouth. In the training stage, the images are normalized using Procrustes analysis and the key components of shape and appearance variation are extracted independently using principal component analysis (PCA). The training results in mean vectors for shape and appearance as well as vectors describing different orders of deviation from the mean shape and appearance respectively, described by the eigenvectors gained from PCA. Using the training data, new faces can be modeled as a linear combination of the mean shape and appearance combined with weighted shape and appearance vectors.

3.3.1 AAM Fitting

AAM Fitting is the iterative process of adapting the model parameters in order to minimize the difference between the modeled and the target face, a task for which several methods have been proposed since the introduction of AAMs. Despite being proposed by (Matthews and Baker, 2004) over a decade ago, the Inverse-Compositional (IC) algorithm is still a competitive method and the standard in many current applications. In their work describing the algorithm, Matthews and Baker have shown that major parts of the iterative computation can be moved outside the loop, allowing a drastic increase in computation speed without sacrificing fitting precision. We use the IC algorithm as baseline and compare its fitting performance on LWIR face images to two more recent extensions, namely simultaneous inverse compositional (SIC) as suggested by (Gross et al., 2005) and the alternating inverse compositional (AIC) fitting introduced in (Papandreou and Maragos, 2008).

3.3.2 Feature-based AAM

Traditionally, AAM fitting is performed on the unprocessed input images. However, with the introduction of feature descriptors such as SIFT and HOG as a powerful tool in image processing, several successful attempts to combine AAMs with descriptors have been published. When working with a feature-based AAM, the model is not directly trained on the input images, but instead on sets of extracted feature matrices with densely extracted features. It has been shown in (Antonakos et al., 2015) that using features for image description can significantly increase the accuracy of AAMs in the visual domain. To analyze the

suitability of feature-based AAMs for face fitting in LWIR images we have therefore compared the performance of regular models with those trained using dense SIFT and HOG features.

3.4 Proposed Processing Pipeline

We have combined the described processing steps into an image processing pipeline that allows to evaluate the performance of different algorithms for each step. In our implemented pipeline, the database images are first filtered using one of the preprocessing filters to enhance image contrast. In the next step, AAMs are trained using the database images and the manually annotated landmark positions. The tool chain allows training of traditional and feature-based AAMs using dense HOG and SIFT features utilizing functionality provided by the Menpo software package (Alabort-Imedina et al., 2014). Fitting of the AAM to test images can be performed using any of the different IC-based fitting algorithms described above. The starting position for the fitting process can be acquired from a user-selectable bounding box or by placing an initial shape on the bounding box of the image's ground truth landmarks, provided a ground truth is available. In case a ground truth exists, the software makes it possible to compare the results of the fitting process to the ground truth reference visually and quantitatively.

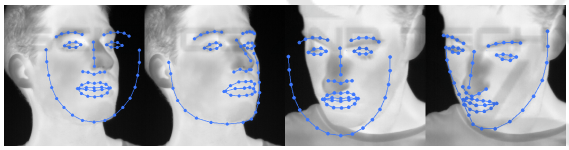


Figure 3: First and third image: Initial landmark location. Second and last image: Landmarks after fitting with a DSIFT-AAM and SIC without preprocessing.

4 RESULTS

In this section, we display quantitative results of the fitting performance for single frames of both seen and unseen individuals and an analysis of the different method's abilities to track a face in an unconstrained video session. Finally, vital sign extraction from a moving person's face is demonstrated.

We have exhaustively analyzed all possible combinations (45 in total) of the following algorithms:

- **Preprocessing:** No preprocessing, anisotropic diffusion highpass as in (Ghiass et al., 2013), USM with anisotropic diffusion, USM with bilateral filtering and traditional USM with a Gaussian kernel.
- **AAM:**Traditional Intensity-based AAM, AAM with dense HOG features, AAM with dense SIFT features.
- **Fitting Algorithm:** Project-out inverse compositional (PIC), alternating inverse compositional (AIC), simultaneous inverse compositional (SIC).

In a first step, we mirrored each image in the database to increase pose variation, resulting in a total of 1390 images. To evaluate the AAMs performance to model unseen faces, we have then split our annotated database into 1272 images of 28 persons for training and 118 images of 3 persons for testing. Fitting was initialized using the AAM's mean shape, scaled and translated to fit the bounding box of the ground truth. We performed fitting and compared the fitting result to the ground truth reference (Figure 3). To quantify the fitting error, we used the normalized error metric introduced in (Zhu and Ramanan, 2012) which minimizes the effect of face size and head pose on the final result, thereby allowing an efficient comparison of errors across different image sets and databases. The error metric E_i computed for each image I_i is the root mean squared distance in pixels between each horizontal and vertical fitted landmark position $x_{n,f}$, $y_{n,f}$ after AAM fitting and its corresponding ground truth landmark $x_{n,g}$, $y_{n,g}$, accumulated across all N landmarks in the image and normalized by the mean of face width w_i and height h_i:

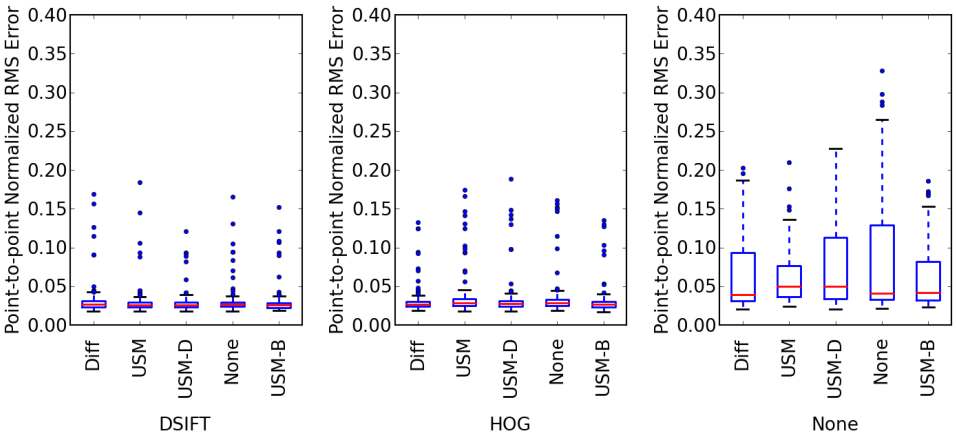
$$E_i = N_i \sqrt{\frac{\sum_{n=1}^{N} \left[(x_{n,f} - x_{n,g})^2 + (y_{n,f} - y_{n,g})^2 \right]}{2N}} \quad (7)$$

with

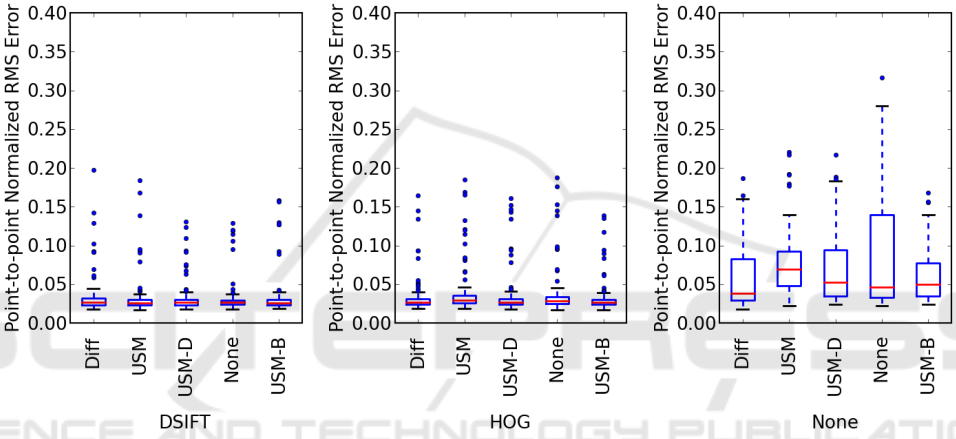
$$N_i = \left(\frac{1}{\frac{1}{2}(w_i + h_i)} \right) \quad (8)$$

Figure 4 shows a quantitative analysis of the performance of each combination. It can be seen that the introduced preprocessing algorithms result in an improvement of the AAM fitting performance for intensity-based AAM, especially when an advanced fitting algorithm such as AIC or SIC is used. Generally, AIC and SIC show comparable performance the tested data and outperform the traditional PIC in all direct comparisons, i.e. when preprocessing method and used feature remain unchanged and only the fitting algorithm is varied. At the same time, the two tested feature-based AAM combinations clearly outperform their intensity-based counterparts in all direct comparisons, with DSIFT slightly outperforming HOG in terms of fitting performance and outlier count. Notably, DSIFT and HOG performance is only minimally affected by any preprocessing.

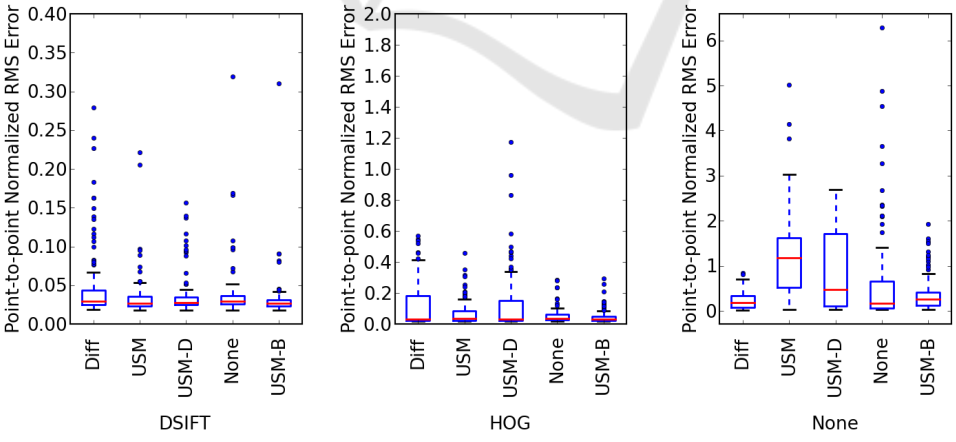
Since it has now been shown that preprocessing has only minimal impact on fitting performance



(a) Simlutaneous Inverse Compositional (SIC)



(b) Alternating Inverse Compositional (AIC)



(c) Project-Out Inverse Compositional (PIC)

Figure 4: Normalized residual error after performing fitting on a set of 118 unseen images of 3 persons. In each column, the AAM was trained using a different feature extraction method, while the fitting algorithm was different for each row. The figures represent an exhaustive overview of all tested combinations of features, fitting algorithms and preprocessing methods.

of feature-based AAMs, further analysis focuses on the precision of feature- and intensity-based methods on unfiltered images. Figure 6 displays the percentage of test images that meet given preci-

sion requirements, evaluated for PIC/SIC/AIC and intensity/DSIFT/HOG-based AAMs. Again, it is shown that feature-based AAMs allow more precise fitting than intensity-based models. Overall, DSIFT shows a better convergence behavior than HOG for LWIR images; the performance difference between both descriptors is higher than the difference for regular photographs reported in (Antonakos et al., 2015), indicating that DSIFT is a particularly well suited descriptor for LWIR face fitting. AIC and SIC perform comparably well and outperform the traditional PIC method; especially the original inverse-compositional AAM approch (intensity-based PIC) by (Matthews and Baker, 2004) shows lower fitting precision than all other competing combinations.

Figure 5 shows a performance comparison for seen and unseen faces. It displays box plots of the fitting performance of the AAM trained on 28 persons and evaluated on 118 images of 3 unseen persons, and for comparison plots of each fitting algorithms' performance trained with the same database and evaluated on a set of 125 seen images from the training database. For this comparison we used the well-performing combination of DSIFT with no preprocessing and compared all three implemented fitting algorithms. The results show that the trained model performs slightly better when confronted with seen faces.

AAM performance in face tracking was tested quantitatively on a 60 second video sequence taken at 30 fps, where each 5th frame was annotated with a total of 8 points located at the inner and outer eye corners, the outer mouth corners and the center edges of upper and lower lip, resulting in 359 annotated frames. During the sequence, the person was performing increasingly fast and complex head pose changes, starting with a slow controlled left-to-right movement and ending with fast and arbitrary head shakes and rotations. The tracker was initialized manually by defining the face's bounding box in the first frame. The landmark detection for all subsequent frames was performed fully automatically by the AAM fitting software using each frame's final fitting result as initial landmark positions for the next frame. The normalized error between the 8 ground truth points and the corresponding points detected by the AAM was computed using Equation 7. Figure 7 shows the normalized error values as well as the mean frame-to-frame change of the marker point positions acquired as normalized error between two consecutive frames to indicate the current movement speed of the person's head in each frame. Additionally, actual fitting results are shown. It can be seen that tracking precision changes only marginally during slow head

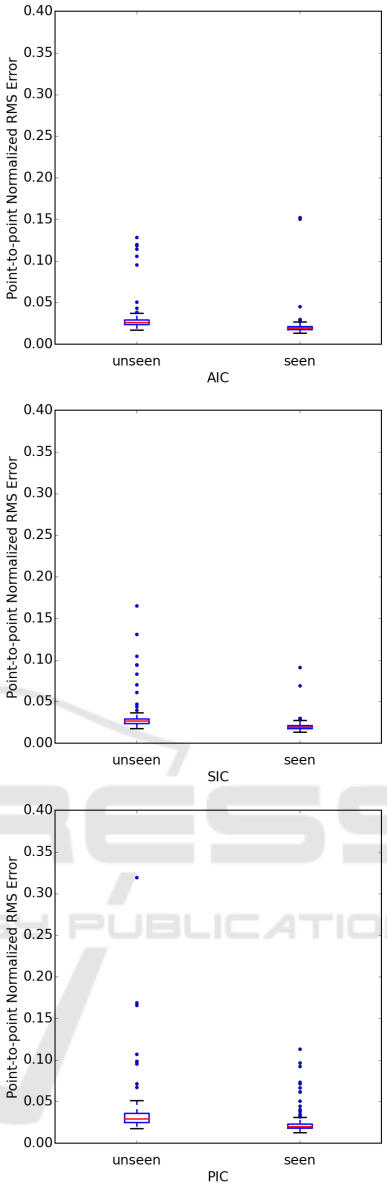
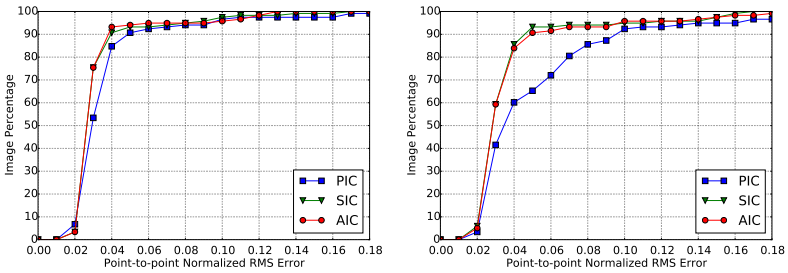


Figure 5: Comparison of the normalized fitting error for a set of 3 seen and 3 unseen images, tested for different fitting algorithms and using DSIFT with no preprocessing to generate the model.

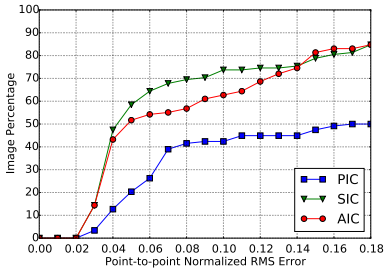
movement regardless of head pose. The tracker performance temporarily degrades in single parts of the sequence that show very fast head movement, however it can be seen that the model is able to automatically recover from high misalignment in single frames.

5 DISCUSSION

It can be seen that performing diffusion filtering as



(a) DSIFT (b) HOG



(c) Intensity

Figure 6: Percentage of Images that fall within given precision requirements, computed for images with no preprocessing applied.

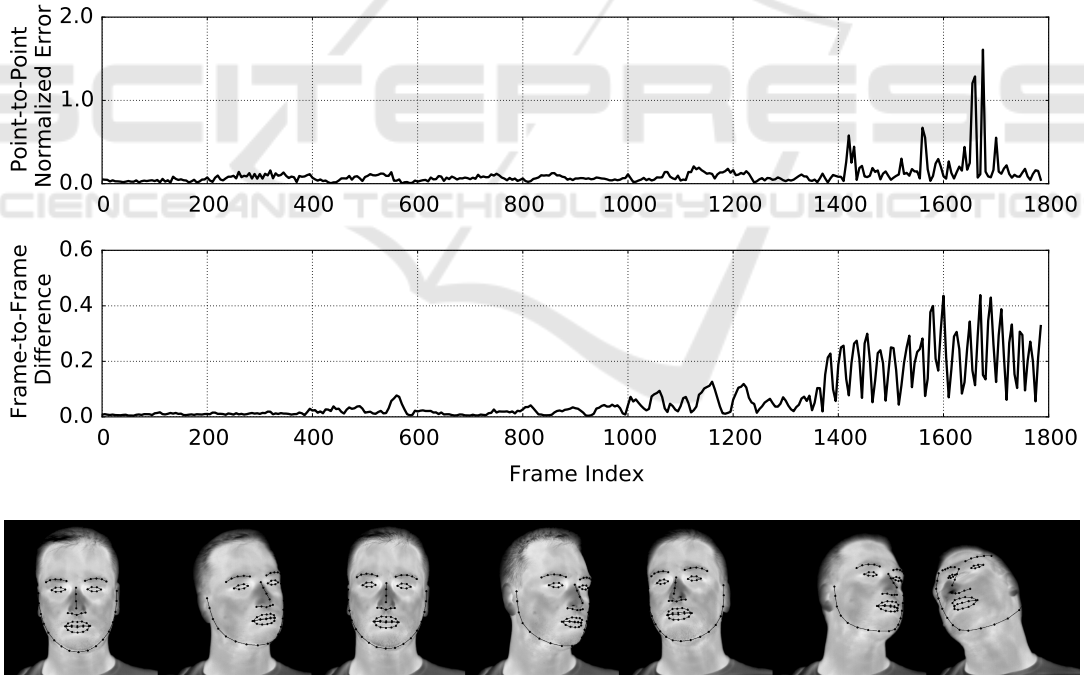


Figure 7: Analysis of AAM performance for tracking a face in an unseen video sequence. Top: Normalized error between ground truth and fitting result for each annotated frame. Center: Normalized difference between two consecutive frames to indicate head movement speed in different sections of the video sequence. Bottom: Fitting result examples, sorted by their final normalized error in ascending order. From left to right: Best, 2nd percentile, 25th percentile, median, 75th percentile, 98th percentile, worst.

preprocessing step improves the fitting accuracy of the AAM in case the project-out inverse compositional algorithm (PIC) is used as stated by (Ghiass et al., 2013). However, by employing the more recently introduced feature-based AAM, even better results are obtained. The fact that the final fitting errors

of intensity-based AAM cover a large span of values and that similar USM-based approaches lead to significantly different results suggests that intensity-based AAM and the analyzed preprocessing filters lack robustness and are prone to the bias introduced by initialization and preprocessing parameters. On the contrary, the two analyzed feature-based AAM proved to be more robust. Using DSIFT and HOG to train the model drastically improves fitting performance regardless of preprocessing; the fact that the results do not differ significantly for all preprocessing algorithms shows that the extracted features describe image content very robustly. The results suggest that using preprocessing for feature-based AAM does not result in a significant performance increase and that the preprocessing step can be omitted, especially when considering the additional computational requirements to run the preprocessing filter.

Although quantitative analysis has shown a measurable difference in fitting performance on seen and unseen images the ability of the AAM to model untrained faces still allows for precise landmark detection in unseen images. The model has been shown to be robust enough to track an unseen face during a series of challenging head pose changes in a video sequence with the ability to recover even after phases of fast head movement or extreme out-of-plane rotation.

6 CONCLUSION

In this paper we have shown that AAMs are a viable approach for face tracking in the thermal infrared domain. Using a suitable database and a well-performing combination of algorithms comprising DSIFT for modeling and SIC for fitting yields stable and robust results. It has been shown that AAMs can be used for robust single-frame initialized LWIR face tracking.

REFERENCES

Alabort-I-medina, J., Antonakos, E., Booth, J., Snape, P., and Zafeiriou, S. (2014). Menpo: A comprehensive platform for parametric image alignment and visual deformable models. In *ACM International Conference on Multimedia*, MM '14, pages 679–682, Orlando, Florida, USA. ACM.

Alkali, A. H., Saatchi, R., Elphick, H., and Burke, D. (2014). Eyes' corners detection in infrared images for real-time noncontact respiration rate monitoring. In *WCCAIS 2014*, pages 1–5. IEEE.

Antonakos, E., i medina, J. A., Tzimiropoulos, G., and Zafeiriou, S. (2015). Feature-based lucas-kanade and active appearance models. *IEEE Transactions on Image Processing*, 24(9):2617–2632.

Cootes, T. F., Edwards, G. J., and Taylor, C. J. (2001). Active appearance models. *IEEE PAMI*, 23(6):681–685.

Dowdall, J., Pavlidis, I. T., and Tsiamyrtzis, P. (2007). Coalitional tracking. *Comput. Vis. Image Underst.*, 106(2-3):205–219.

Filipe, S. and Alexandre, L. A. (2013). Thermal infrared face segmentation: A new pose invariant method. In *Pattern Recognition and Image Analysis*, pages 632–639. Springer.

Gault, T. R. and Farag, A. A. (2013). A fully automatic method to extract the heart rate from thermal video. In *CVPRW 2013*, pages 336–341. IEEE.

Ghiass, R. S., Arandjelović, O., Bendada, A., and Maldague, X. (2014). Infrared face recognition: A comprehensive review of methodologies and databases. *Pattern Recognition*, 47(9):2807–2824.

Ghiass, R. S., Arandjelovic, O., Bendada, H., and Maldague, X. (2013). Vesselness features and the inverse compositional aam for robust face recognition using thermal ir. *arXiv preprint arXiv:1306.1609*.

Gross, R., Matthews, I., and Baker, S. (2005). Generic vs. person specific active appearance models. *Image and Vision Computing*, 23(12):1080–1093.

Kalal, Z., Mikolajczyk, K., and Matas, J. (2012). Tracking-learning-detection. *PAMI*, 34(7):1409–1422.

Lahiri, B., Bagavathiappan, S., Jayakumar, T., and Philip, J. (2012). Medical applications of infrared thermography: a review. *Infrared Physics & Technology*, 55(4):221–235.

Lewis, G. F., Gatto, R. G., and Porges, S. W. (2011). A novel method for extracting respiration rate and relative tidal volume from infrared thermography. *Psychophysiology*, 48(7):877–887.

Matthews, I. and Baker, S. (2004). Active appearance models revisited. *IJCV*, 60(2):135–164.

Papandreou, G. and Maragos, P. (2008). Adaptive and constrained algorithms for inverse compositional active appearance model fitting. In *CVPR 2008*, pages 1–8. IEEE.

Zhou, Y., Tsiamyrtzis, P., Lindner, P., Timofeyev, I., and Pavlidis, I. (2013). Spatiotemporal smoothing as a basis for facial tissue tracking in thermal imaging. *IEEE Trans. Biomed. Engineering*, 60(5):1280–1289.

Zhu, X. and Ramanan, D. (2012). Face detection, pose estimation, and landmark localization in the wild. In *CVPR 2012*, pages 2879–2886. IEEE.

Zhu, Z., Tsiamyrtzis, P., and Pavlidis, I. (2008). The segmentation of the supraorbital vessels in thermal imagery. In *AVSS 2008*, pages 237–244.