Registration Invariant Representations for Expression Detection

Patrick Lucey^{1,2}, Simon Lucey³ and Jeffrey F. Cohn^{1,2}

Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA¹

Department of Psychology, University of Pittsburgh, Pittsburgh, PA, USA²

ICT Center, CSIRO, Sydney, Australia³

plucey@pitt.edu, simon.lucey@csiro.au, jeffcjohn@cs.cmu

Abstract

Active appearance model (AAM) representations have been used to great effect recently in the accurate detection of expression events (e.g., action units, pain, broad expressions, etc.). The motivation for their use, and rationale for their success, lies in their ability to: (i) provide dense (i.e. 60- 70 points on the face) registration accuracy on par with a human labeler, and (ii) the ability to decompose the registered face image to separate appearance and shape representations. Unfortunately, this human-like registration performance is isolated to registration algorithms that are specifically tuned to the illumination, camera and subject being tracked (i.e. “subject dependent” algorithms). As a result, it is rare, to see AAM representations being employed in the far more useful “subject independent” situations (i.e., where illumination, camera and subject is unknown) due to the inherent increased geometric noise present in the estimated registration. In this paper we argue that “AAM like” expression detection results can be obtained in the presence of noisy dense registration through the employment of registration invariant representations (e.g., Gabor magnitudes and HOG features). We demonstrate that good expression detection performance can still be enjoyed over the types of geometric noise often encountered with the more geometrically noisy state of the art generic algorithms (e.g., Bayesian Tangent Shape Models (BTSM), Constrained Local Models (CLM), etc). We show these results on the extended Cohn-Kanade (CK+) database over all facial action units.

1. Introduction

Central to the success of an automatic facial expression detector is the face alignment/registration algorithm and the visual features derived from it. As expressions can be subtle, high accuracy is desired so that the correspondence between various facial features and muscles contracting and

Figure 1. This figure depicts AAM representations employed in the current state-of-the-art expression detection algorithms. Column (a) depicts the initial scenario in which all shape and appearance is preserved. In (b) geometric similarity is removed from both the shape and appearance; and in (c) shape (including similarity) has been removed leaving the average face shape and what we refer to as the canonical appearance. Features derived from the representations in columns (b) and (c) are used in AAM expression detection systems. Two central questions addressed in this paper are: (i) how sensitive are AAM representations to registration noise, and (ii) are their alternate representations that can give greater invariance?

controlling the face can be maintained, enhancing the ability of a classifier to detect the facial expression correctly. To facilitate this, active appearance models (AAMs) [7] have been widely used in the field of affective computing as they provide dense registration accuracy (i.e. 60-70 points on the face) so that these correspondences are kept allowing comparisons of the relevant areas to be performed [2, 3, 13, 15].

It has been well established [2, 15] when performing expression detection using AAM derived representations (i.e. decoupled shape and appearance features, see Figure 1) that: (i) dense registration is preferable to coarse registration, and (ii) improved alignment accuracy is correlated with improved detection performance. This is a desired result if we have an automatic dense face alignment algorithm that can exhibit “human like” accuracy (i.e. performance

that is indistinguishable from the error seen across multiple human labelers). Unfortunately, this type of accuracy is still not a reality for dense facial registration algorithms where one has to fit to a wide number of subjects, with high variability in expression, pose, camera conditions and illumination. Approaches such as Constrained Local Models [18] and Bayesian Tangent Shape Models (BTSM) have recently demonstrated impressive dense alignment performance, however, their performance is still poor for subtle facial deformations.

Due to these limitations it is common practice that *subject dependent* AAM models for registration/tracking are used for facial expression detection. These subject dependent models are tuned specifically to the subject, camera conditions, and illumination of the target image sequence to be tracked [2, 13, 15] and are able to exhibit “human like” accuracy. This tuning is accomplished through the judicious hand labeling of key frames in the target image sequence. Anywhere up to 5% of images in a given sequence need to be manually labelled so that appearance variabilities such as illumination, appearance, camera and pose are suitably countered for. In applications in the fields of behavioral science and others where time can be taken to gain an accurate and objective measure, this is viable solution. However, for commercial applications such as marketing, security/law enforcement [17], driver safety [19], health-care [14] and consumer electronics (e.g. digital cameras) [20] (e.g. Figure 2), a more generic or subject independent face alignment approach is required as: (i) the face needs to be registered quickly, and (ii) it also needs to generalize across the target population and amongst a host of different imaging conditions (e.g., illumination, pose, camera, etc.).

In this paper we argue that expression detection results, comparable to state of the art AAM expression detection methods with “human like” registration, can be obtained using local spatial invariant representations. Our central contributions in this paper are:-

- Explore and motivate the employment of (i) Gabor magnitudes and (ii) histograms of orientated gradients (HOG) as methods for encoding local spatial invariance. (Section 2)
- Review current state of the art AAM representations [2, 3, 13, 15] for expression detection, and quantify how these representations perform when poor face alignment is encountered. (Section 3 and 5)
- Demonstrate the natural registration invariant properties of our proposed representations for the task of AU detection on the extended Cohn-Kanade (CK+) dataset using histogram of orientated gradients (HOG) and Gabor magnitudes. (Section 5)

Figure 2. Examples of where detecting facial expressions are being used: (clockwise from top left) (a) health-care such (e.g. pain detection), (b) marketing (e.g. reaction to product), (c) driver safety (e.g. driver intent/fatigue detection), and (d) consumer electronics (e.g. smile detection on digital cameras). In these scenarios a subject independent approach to face alignment is required for practical reasons.

2. Local Distribution Features

Under the assumption that there will always be some degree of registration error in a target face image it is useful to explore features that give invariance to registration. Holistic invariant features are difficult to derive as one rarely has prior knowledge of how the image geometrically deforms holistically. Instead, it is simpler to adopt a strategy where a single complex holistic deformation in an image, such as those found in facial expressions, can always be broken down into multiple simple deformations (e.g., optical flow where a single complex deformation can be defined as multiple, one for each pixel, locally constrained translations). Representing an image as a “super vector” of concatenated local region features that are invariant to simple deformations (e.g., translation), an argument can then be made that this super vector will exhibit invariance to more complex holistic registration errors.

Many different techniques for describing local image regions have been proposed in literature. The simplest feature is a vector of raw appearance pixels. However, if an unknown error in registration occurs there is an inherent variability associated with that *true* (i.e. correctly registered) local image appearance. Due to this variability, an argument can be made that these local pixel appearances are more aptly described by a distribution rather than a static observation point. We investigate two popular methods in vision for obtaining distribution features that exhibit good local spatial invariance: (i) Histogram of Oriented Gradients (HOG), and (ii) Gabor magnitudes.

Histogram of Orientated Gradients: Histogram of Oriented Gradient grids (HOG) [8], are a close relation of the descriptor in Lowe’s seminal SIFT approach [12] to code visual appearance. Briefly, the HOG method tiles the input image with a dense grid of cells, with each cell containing a local histogram over orientation bins. At each pixel, the image gradient vector is calculated and converted to an angle, voting into the corresponding orientation bin with a vote weighted by the gradient magnitude. The orientation bins were evenly spaced over $0^\circ - 180^\circ$ (unsigned gradient). Histograms were obtained at different discrete scales using a Gaussian gradient function (in x- and y-) with the variance parameter σ^2 defining the scale. These scale specific histograms are all concatenated into a single feature vector. Shift invariance is naturally encoded in this type of feature through the size of the cell from which the histograms are derived. The larger the cell size, the greater the shift invariance. In this work we used a cell size of 12×12 over 3 frequencies and 4 rotations.

Gabor Magnitudes:

In the 2D spatial domain, a Gabor wavelet is a complex exponential modulated by a Gaussian,

$$g_{\omega, \theta}(x, y) = \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{x'^2 + y'^2}{2\sigma^2} + j\omega x'\right\} \quad (1)$$

where $x' = x \cos(\theta) + y \sin(\theta)$, $y' = -x \sin(\theta) + y \cos(\theta)$, x and y denote the pixel positions, ω represents the centre frequency, θ represents the orientation of the Gabor wavelet, while σ denotes the standard deviation of the Gaussian function. Please refer to [9] on strategies for spacing the filters in the 2D spatial frequency domain for a fixed number of scales and orientations. These filters are in quadrature where the real part of the filter is even symmetric and the imaginary part of the filter is odd symmetric. When convolved with an input image the the scalar magnitude value of the resultant complex response can be interpreted as the correlation matrix (i.e. distribution) of the local region (defined by the σ), for the image components resonating with the central frequency (defined by ω) in the direction of (θ). Like for HOG features the magnitude values for each orientation and central frequency are concatenated into a vector. In this paper, we used 8 different rotations and 8 different frequencies.

3. AAM Representations

Active Appearance Models (AAMs) have been shown to be a good method of aligning a pre-defined linear shape model that also has linear appearance variation, to a previously unseen source image containing the object of interest. In general, AAMs fit their shape and appearance components through a gradient-descent search, although other

optimization methods have been employed with similar results [7].

The shape s of an AAM [7] is described by a 2D triangulated mesh. In particular, the coordinates of the mesh vertices define the shape $s = [x_1, y_1, x_2, y_2, \dots, x_n, y_n]$, where n is the number of vertices. These vertex locations correspond to a source appearance image, from which the shape was aligned. Since AAMs allow linear shape variation, the shape s can be expressed as a base shape s_0 plus a linear combination of m shape vectors s_i :

$$s = s_0 + \sum_{i=1}^m p_i s_i \quad (2)$$

where the coefficients $\mathbf{p} = (p_1, \dots, p_m)^T$ are the shape parameters. These shape parameters can typically be divided into rigid similarity parameters $\mathbf{p}_s$ and non-rigid object deformation parameters $\mathbf{p}_o$, such that $\mathbf{p}^T = [\mathbf{p}_s^T, \mathbf{p}_o^T]$. Similarity parameters are associated with a geometric similarity transform (i.e. translation, rotation and scale). The object-specific parameters, are the residual parameters representing non-rigid geometric variations associated with the determining object shape (e.g., mouth opening, eyes shutting, etc.). Procrustes alignment [7] is employed to estimate the base shape s_0 .

Keyframes within each video sequence were manually labelled, while the remaining frames were automatically aligned using a gradient descent AAM fitting algorithm described in [16]. Once we have tracked the subject’s face by estimating the shape and appearance AAM parameters, we can use this information to derive the following features:

- **SPTS:** The similarity normalized shape, s_n , refers to the 68 vertex points in s_n for both the x - and y - coordinates, resulting in a raw 136 dimensional feature vector. These points are the vertex locations after all the rigid geometric variation (translation, rotation and scale), relative to the base shape, has been removed. The similarity normalized shape s_n can be obtained by synthesizing a shape instance of s , using Equation 2, that ignores the similarity parameters $\mathbf{p}$.
- **SAPP:** The similarity normalized shape, a_n , refers to the where all the rigid geometric variation (translation, rotation and scale) has been removed. It achieves this by using s_n calculated above and warps the pixels in the source image with respect to the required translation, rotation and scale. This is the type of approach is employed by most researchers [4, 20], as only coarse registration is required (i.e. just face and eye locations). When out-of-plane head movement is experienced some of the face is partially occluded which can affect performance, also some non-facial information is included due to occlusion.

Figure 3. In our first experiment we compared the SAPP and CAPP features from the AAM across various geometric noise levels, which is symptomatic of poor registration in subject independent algorithms: (a) ideal tracking, (b) 5 RMS-PE, (c) 10 RMS-PE, (d) 15 RMS-PE, (e) 20 RMS-PE, (f) 25 RMS-PE, and (g) 30 RMS-PE. From this you can see when the amount of noise is increased, the piece-wise affine warp which synthesizes the CAPP image causes significant deformation to the face which is much noisier representation than the SAPP image.

- **CAPP:** The canonical normalized appearance a_0 refers to where all the non-rigid shape variation has been normalized with respect to the base shape s_0 . This is accomplished by applying a piece-wise affine warp on each triangle patch appearance in the source image so that it aligns with the base face shape. In previous work [1], it was shown by removing the rigid shape variation, poor performance was gained.

In this paper, we are interested in analyzing the change in performance of the different appearance features (SAPP and CAPP) when face alignment is poor, as is sometimes experienced in using subject independent or generic methods.

4. Experimental Setup

In this paper, we conducted experiments for the task of facial action unit (AU) detection. These experiments were set up to compare the difference between subject dependent (e.g. AAM) and subject independent (e.g. Viola-Jones, CLM) face alignment algorithms, and their subsequent visual features. We had two main interests here: 1) comparing different AAM pixel representations across noise levels (SAPP vs CAPP), and 2) comparing these pixel representations against shift invariant features (i.e. HOG and Gabor magnitudes).

To facilitate these goals, we added various amounts of geometric noise to the test images. To do this, the similarity normalized base template had an inter-ocular distance of 50 pixels. For a fair comparison, we took into account differing face scales between testing images. This is done by first removing the similarity transform between the estimated shape and the base template shape and then computing the RMS-PE between the 68 points. We obtained the poor initial alignment by synthetically adding

affine noise to the ground-truth coordinates of the face. We then perturbed these points with a vector generated from white Gaussian noise. The magnitude of this perturbation was controlled to give a desired root mean squared (RMS) pixel error (PE) from the ground-truth coordinates (which were the AAM tracked landmarks). During learning, the initially misaligned images were defined to have between 5-30 RMS-PE. This range of perturbation was chosen as it approximately reflects the range of alignment error that can be experienced using subject independent face alignment algorithms. Examples of the poor tracking is given in Figures 3 and 4. In our experiments all the training images were clean (i.e. zero noise) and they were tested across different noise levels (i.e. 5-30 RMS-PE). After all images were registered they were downsampled to 35×30 pixels¹.

Database:

In this paper we used the extended Cohn-Kanade (CK+) database [13], which contains 593 sequences from 123 subjects. The image sequence vary in duration (i.e. 10 to 60 frames) and incorporate the onset (which is also the neutral frame) to peak formation of the facial expressions. For the 593 posed sequences, full FACS coding of peak frames is provided. Approximately fifteen percent of the sequences were comparison coded by a second certified FACS coder. Inter-observer agreement was quantified with coefficient kappa, which is the proportion of agreement above what would be expected to occur by chance [10]. The mean kappa for inter-observer agreement were 0.82 for action units coded at apex and 0.75 for frame-by-frame coding.

Classification using Support Vector Machines:

¹In this work we varied the resolutions from 88×75 , 61×52 and 35×30 , with little change in performance in AU detection performance over these resolutions

Figure 4. In our second set of experiments we tested the histogram of orientated gradients (HOG), Gabor magnitudes and pixel features on the similarity normalized appearance images when the following geometric noise was encountered in the face alignment phase: (a) ideal tracking, (b) 5 RMS-PE, (c) 10 RMS-PE, (d) 15 RMS-PE, (e) 20 RMS-PE, (f) 25 RMS-PE, and (g) 30 RMS-PE.

Support vector machines (SVMs) have been proven useful in a number of pattern recognition tasks including face and facial action recognition. SVMs attempt to find the hyperplane that maximizes the margin between positive and negative observations for a specified class. A linear SVM classification decision is made for an unlabeled test observation $\mathbf{x}^*$ by,

$$\begin{aligned} \mathbf{w}^T \mathbf{x}^* &>_{true} b \\ &<_{false} \end{aligned} \quad (3)$$

where $\mathbf{w}$ is the vector normal to the separating hyperplane and b is the bias. Both $\mathbf{w}$ and b are estimated so that they minimize the structural risk of a train-set, thus avoiding the possibility of overfitting to the training data. Typically, $\mathbf{w}$ is not defined explicitly, but through a linear sum of support vectors. A linear kernel was used in our experiments due to its ability to generalize well to unseen data in many pattern recognition tasks [11]. LIBSVM was used for the

training and testing of SVMs [6].

For AU detection, we just used a linear one-vs-all two-class SVM (i.e. AU of interest vs non-AU of interest). For the training of the linear SVM for each of the AU detectors, all neutral and peak frames from the training sets were used. The frames which were coded to contain the AU were used as positive examples and all others were used as negative examples, regardless if the AU occurred alone or in combination with other AUs.

Benchmarking Protocol and Evaluation Metric:

To maximize the amount of training and testing data, we used the leave-one-subject-out cross-validation configuration. This means for AU detection, 123 different training and testing sets need to be used. The area underneath the receiver-operator characteristic (ROC) curve (A') was used as it has been used to assess the performance of automatic facial expression detection systems [4]. The A' metric ranges from 50 (pure chance) to 100 (ideal classification)².

²In literature, the A' metric varies from 0.5 to 1, but for this work we have multiplied the metric by 100 for improved readability of results

Results were averaged across these sets.

5. Results and Discussion

In our experiments we tested across all noise levels for 17 AUs³. Results are shown in Table 1. As you can see there is a gradual drop-off in performance as the amount of noise is increased. A clearer picture emerges when we analyze the average AU performance across these noise levels, which is shown in Figure 5⁴.

The first thing to note is the performance of both the AAM representations (SAPP=red, CAPP=blue). As you can see from these two curves, the performance is very similar. This begs the question, what is the benefit of the CAPP approach? The answer to this is quite simple, when there is large amounts of head motion and out-of-plane head rotation, the CAPP features are far more superior than the SAPP features as they can project all features into a uniform view. We will be investigating the problem of out-of-plane head motion in our future work.

What is interesting though, is that both the shift invariant features (which were processed using the SAPP images, as this is what occurs when using a coarse face alignment approach such as Bartlett et al. [5]), remain somewhat invariant to this type of noise. This gives an insight in why the method of Bartlett et al. [5] only employs a very coarse registration method (i.e. Viola-Jones face detector followed by eye detector), in conjunction with their Gabor filter approach. In most of the work in which this system is applied on, there is very little head motion and so having coarse registration (noise from 0-15 RMS-PE), you will suffer little degradation in performance by employing shift-invariant features. So in this paradigm, there is no real benefit gained from a more sophisticated synthesized view such as the CAPP features. However, as mentioned earlier, this approach can not be used when there is a lot of head motion.

6. Conclusions and Future Work

In this paper we showed that when we have perfect alignment or very close to it, similar to the alignment we get when we employ an subject dependent approach to face alignment such as AAMs, there is little benefit in employing shift-invariant features such as Gabor magnitudes or histogram of orientated gradients (HOG). However, when there is noise present in the alignment, which is the case when utilizing subject independent or generic algorithms such as Viola-Jones or CLM, a robust solution is to use these shift-invariant features. But the main point is that these shift

Table 1. Results showing the area underneath the ROC curve for the various feature representations (F) – (i.e. AAM pixel representations: similarity normalized appearance (S), and canonical normalized appearance (C), and different shift invariant features: histogram of gradients (H), and Gabor magnitudes (G)) for selected AUs of interest. N is the number of positive examples for each AU.

AU	N	F	Geometric Noise Added						
			0	5	10	15	20	25	30
1	173	S	89.0	83.6	75.0	65.5	63.0	58.6	55.2
		C	89.9	83.4	72.1	60.4	67.2	57.1	53.2
		H	92.2	90.1	86.3	81.1	77.4	75.4	65.6
		G	82.4	79.4	73.4	73.6	67.5	65.4	64.9
2	116	S	96.0	92.9	84.8	73.5	70.2	59.5	63.8
		C	94.0	86.8	76.6	70.3	75.4	66.3	61.2
		H	91.4	89.6	87.4	80.3	81.1	78.7	72.6
		G	89.7	89.1	89.4	83.2	78.4	74.1	75.6
4	191	S	84.6	79.6	74.3	65.5	59.5	63.2	52.7
		C	85.4	77.4	68.1	59.1	61.1	57.7	51.8
		H	72.5	72.8	69.7	63.2	64.6	60.5	58.8
		G	87.9	86.0	82.8	81.5	75.0	78.0	70.9
6	122	S	94.4	88.7	75.8	66.9	57.7	53.4	56.6
		C	93.5	87.5	72.6	59.9	57.2	54.8	54.0
		H	86.1	84.2	79.7	74.1	63.5	61.4	60.6
		G	92.9	93.4	92.5	88.9	89.1	78.8	80.1
7	119	S	80.7	77.1	65.0	64.0	66.6	52.7	58.8
		C	80.8	75.2	61.2	58.0	52.7	56.3	57.8
		H	65.9	68.3	66.2	67.8	66.9	59.2	62.2
		G	82.2	81.9	78.5	74.7	71.3	69.7	68.5
9	74	S	99.4	96.4	80.3	73.0	60.6	73.9	45.4
		C	98.1	92.4	71.6	62.6	65.0	55.8	51.5
		H	97.9	97.5	89.5	82.2	77.6	74.4	67.1
		G	98.4	98.0	94.5	92.5	85.4	89.2	81.4
12	111	S	93.9	91.2	81.7	77.6	72.9	61.4	52.5
		C	93.3	89.8	85.5	69.7	65.9	53.9	53.8
		H	95.0	94.5	90.1	87.0	85.3	80.4	75.1
		G	92.2	91.3	90.6	90.3	87.2	78.4	78.3
17	196	S	86.7	82.7	70.7	66.6	60.4	57.8	52.4
		C	88.3	79.8	66.1	59.6	62.9	53.9	53.7
		H	83.7	81.8	72.9	68.3	65.3	59.8	60.3
		G	85.2	84.3	76.4	72.1	67.0	60.2	62.2

invariant features do not gain you an improvement in performance, just robustness when there is misalignment.

In our future work, we hope to conduct these comparisons on spontaneous data which contains subtle expressions, illumination as well as head motion variations. We intend to use the CLM and AAM so that a full comparison can be made. In this paper, we have seen the benefit of using shift-invariant features in the spatial domain. In future, we also plan to look into the problem of making the features/classifier shift invariant in the temporal domain which has the potential to improve AU and expression detection performance. Once these areas can be fully explored and

³These AUs were chosen as they had more than 20 examples in the CK+ dataset. These AUs were: 1, 2, 4, 5, 6, 7, 9, 11, 12, 15, 17, 20, 23, 24, 25 and 27

⁴The average is across all 17 AUs and it is a weighted one based on the number of positive examples (N).

quantified, a better understanding on which approach can be best used for a particular application can be made.

References

- [1] A. Ashraf, S. Lucey, J. Cohn, T. Chen, Z. Ambadar, K. Prkachin, P. . Solomon, and B.-J. Theobald. The painful face: pain expression recognition using active appearance models. In *Proceedings of the 9th international conference on Multimodal interfaces*, pages 9–14, Nagoya, Aichi, Japan, 2007. ACM.
- [2] A. Ashraf, S. Lucey, J. Cohn, K. M. Prkachin, and P. Solomon. The Painful Face II– Pain Expression Recognition using Active Appearance Models. *Image and Vision Computing*, 27(12):1788–1796, 2009.
- [3] A. Asthana, J. Saragih, M. Wagner, and R. Goecke. Evaluating AAM Fitting Methods for Facial Expression Recognition. In *Proceedings of the International Conference on Affective Computing and Intelligent Interaction*, 2009.
- [4] M. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan. Automatic Recognition of Facial Actions in Spontaneous Expressions. *Journal of Multimedia*, 2006.
- [5] M. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan. Fully automatic facial action recognition in spontaneous behavior. In *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, pages 223–228, 2006.
- [6] C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [7] T. Cootes, G. Edwards, and C. Taylor. Active Appearance Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.
- [8] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE International Conference on Computer Vision and Pattern Recognition*, 2005.
- [9] D. J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, 4(12):2379–2393, 1987.
- [10] J. Fleiss. *Statistical Methods for Rates and Proportions*. Wiley, N.Y, 1981.
- [11] C. Hsu, C. C. Chang, and C. J. Lin. A practical guide to support vector classification. Technical report, 2005.
- [12] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, November 2004.
- [13] P. Lucey, J. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews. The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. In *Proceedings of the IEEE Workshop on CVPR for Human Communicative Behavior Analysis*, 2010.
- [14] P. Lucey, J. Cohn, S. Lucey, I. Matthews, S. Sridharan, and K. Prkachin. Automatically Detecting Pain Using Facial Actions. In *Proceedings of the International Conference on Affective Computing and Intelligent Interaction*, pages 1–8, 2009.
- [15] S. Lucey, I. Matthews, C. Hu, Z. A. F. de la Torre, and J. Cohn. AAM derived face representations for robust facial action recognition. In I. Matthews, editor, *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, pages 155–160, 2006.
- [16] I. Matthews and S. Baker. Active appearance models revisited. *International Journal of Computer Vision*, 60(2):135–164, 2004.
- [17] A. Ryan, J. Cohn, S. Lucey, J. Saragih, P. Lucey, F. D. la Torre, and A. Rossi. Automated Facial Expression Recognition System. In *Proceedings of the International Carnahan Conference on Security Technology*, pages 172–177, 2009.
- [18] J. Saragih, S. Lucey, and J. Cohn. Face Alignment through Subspace Constrained Mean-Shifts. In *Proceedings of the International Conference on Computer Vision (ICCV)*, 2009.
- [19] E. Vural, M. Cetin, A. Ercil, G. Littlewort, M. Bartlett, and J. Movellan. Automated Drowsiness Detection for Improved Driver Safety– Comprehensive Databases for Facial Expression Analysis. In *Proceedings of the International Conference on Automotive Technologies*, 2008.
- [20] J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett, and J. Movellan. Towards Practical Smile Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(11):2106–2111, 2009.

Figure 5. Plot showing the difference in AU detection performance across all AUs across different levels of alignment noise for pixels, Gabor Magnitudes and Histogram of Gradient features.