Cross-Modal Facial Attribute Recognition with Geometric Features

Chloe Bradley, Terrance E. Boult, and Jonathan Ventura
Department of Computer Science, University of Colorado Colorado Springs

Abstract— **We propose a purely geometric approach to facial attribute recognition which has better cross-modal performance than a state-of-the-art appearance-based method. While labeled color imagery is plentiful for facial attribute learning, labeled imagery in other modalities such as infrared is comparatively rare. Because face appearance is significantly altered in infrared imagery, standard attribute recognition methods trained on color imagery may not transfer well. To address this problem, we propose attribute recognition based purely on geometric information, i.e. geometric relationships derived from a facial landmark detector. We show that our method outperforms a state-of-the-art appearance-based method in attribute recognition when both are trained on color images and tested on infrared images.**

I. INTRODUCTION

Facial attribute recognition is a useful tool for visual search tasks. For example, in person search, such as in surveillance video, attribute recognition can greatly speed up the search process by winnowing the field of suspects with general search terms such as "male" or "wears glasses" [21]. More specific search terms such as descriptive facial attributes [11] and comparative soft biometrics [1] can narrow the field even further.

Attribute recognition is typically cast as a multi-label binary classification problem, and almost all existing facial attribute classifiers are appearance-based [12], [13], [4], [16], [20], meaning that they consider color and texture to determine attributes. However, real-world surveillance tasks often rely on imagery from outside the visible spectrum such as infrared imagery, because the subject was captured at night. Multi-modal or cross-modal attribute classification has not been well-studied, and few datasets exist.

It is unclear whether attribute classifiers learned on appearance in the visible spectrum (RGB) will perform well on other modalities, such as infrared. Because of the lack of large-scale labeled datasets in other modalities, learning modality-specific attribute classifiers is also infeasible currently.

In this work, we describe and evaluate an approach to attribute recognition which is designed to improve cross-modality performance. In particular, we avoid any use of appearance information in the classifier. Instead of appearance (photometric) information, we use purely geometric information to train the classifier. Namely, we use geometric relationships between fiducial points on the face to form a feature vector for attribute recognition.

We train our approach on a large-scale color image dataset, and test on both color and infrared imagery. We compare our work against a state-of-the-art appearance-based method [20] trained on the same color imagery. Our evaluation shows that our method outperforms the state-of-the-art on infrared imagery without any cross-modal retraining.

In the following, we summarize related work (Section II), describe our approach to attribute classification (Section III), detail our evaluation procedure and results (Section IV) and discuss our contributions and directions for future work (Section VI).

II. RELATED WORK

A. Facial Attribute Recognition

Most existing approaches to facial attribute recognition use classifiers trained on appearance features such as color and edge histograms. FaceTracer [12], one of the earliest approaches, splits the face into regions and trains support vector machines (SVMs) on several feature types. The authors manually selected face regions (e.g. cheeks, nose, etc.) and create a local SVM per area and feature type (RGB, HSV, edge magnitude, etc.). AdaBoost is used to determine the set of features to use for attribute recognition.

In followup work, Kumar et al. [13] use low level features such as gradients, HSV, and edges from each region of the face. It borrows the AdaBoost idea of FaceTracer, except that feature selection is changed to reduce error rates more. SVMs are also used in a similar way. It furthermore compares subjects to a small number of people outside the dataset, such that one could compare noses with Brad Pitt.

Liu et al. [16] use multiple convolutional neural networks (CNNs), which they call LNets, to find a face within an image. They then trains another CNN called ANet to determine facial attributes. Instead of using the CNN directly for classification, the features generated by the CNN are given to SVMs for the final classification.

Rudd et al. [20] accomplish two major objectives in their own CNN-based method. The first is to train for all attributes at once, rather than separately. The second is to address bias in data. It does both via a special loss function, which mixes predictions accuracy and correlations in data. This method is shown to largely surpass independently trained CNNs.

The works mentioned here all train and test on color images. These images are usually from datasets of "in-the-wild" face images collected from the Internet such as CelebA [16]. Because these methods make extensive use

This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA R&D Contract No. 2014-14071600012. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

Fig. 1. Illustration of geometric feature vectors considered in our work. *Left:* Normalized landmark coordinates. *Middle:* Angles/distances between landmarks. *Right:* Triangle areas.

of brightness and color cues, they may not transfer well to images from other modalities such as infrared. In this work we build classifiers from only geometric information, and we train on color images but test on both color and infrared images.

B. Geometric Face Analysis

Face alignment and face landmark detection are well-studied problems. One of the most successful approaches is the Active Appearance Model [6] which represented both face appearance and face shape in a joint linear model. The earlier Active Shape Model models only the face shape and not the appearance [7]. Later works in face alignment also adopt an iterative approach to face alignment but replace the joint linear model with more complicated representations (e.g. [19], [16], [22]).

Some previous works have used geometric information for expression recognition [14], [15], [2]. The attributes under consideration in this paper include some expressions (e.g. smiling) but also other properties such as gender.

Also related is action unit (AU) recognition, which is recognition of movements of facial muscle groups and thus might be recognized by a combination of appearance-based and geometric features. For example, OpenFace [5] is a toolkit that collects a variety of data: facial landmarks, Histogram of Oriented Gradients (HOG) [8] features, yaw pitch and roll, and most relevantly, it predicts AUs. In doing so, it largely uses the methods of Baltrusaitis et al. [4]. HOG features are extracted from aligned faces and the dimensions are reduced via Principal Components Analysis (PCA), then given alongside facial landmarks to a linear SVM. To determine the intensity, a linear support vector regression was used. Zhao et al. [23] utilize landmark patches, as opposed to a uniform grid across the face. These patches use dense SIFT [17] features located at each of these landmarks. To identify AUs, they take determine of the correlations between them using SVM. That data is then then used in further SVMs for AUs themselves.

In our work we rely on existing facial landmark detectors [3], [5] for face alignment. We then use only geometric relationships between the landmarks for facial attribute recognition. We target only facial attributes which might reasonably be determined by geometric relationships and avoid clearly appearance-based attributes such as hair color. To the best of our knowledge, ours is the first work to use purely geometric information to recognize facial attributes other than expressions or action units, and the first to apply this idea to cross-modal attribute recognition.

III. METHOD

Our system takes as input an image of a face and predicts binary attributes as output. We assume a suitable system for face detection and facial landmark detection which will provide the 2D locations of the landmarks. We assume that the face image will be aligned to compensate for head rotation.

A. Feature Vectors

The landmark locations are transformed into a feature vector and fed into a separate classifier for each attribute. We considered several feature vector transformations, illustrated in Figure 1 and described here:

- **Locations.** The landmark locations, normalized to the range $[0\ 1] \times [0\ 1]$ according to the bounding box output by the face detector.
- **Distances – Partial.** The distance between each landmark and a central landmark at the bridge of the noise.
- **Distances – Complete.** The distance between every pair of landmarks.
- **Angles – Partial.** The angle between each landmark and a central landmark at the bridge of the nose. The angle between the two inner eye corners is subtracted from all angles in the feature vector to compensate for any head roll not removed by the face alignment.
- **Angles – Complete.** The angle between every pair of landmarks, with the angle between the inner eye corners subtracted from all angles.
- **Triangles.** The area of every triangle in a Delaunay triangulation of the landmarks. The same triangulation is used for all faces, which is determined by selecting the face image with the least yaw, pitch and roll, as

Fig. 2. Sample images from CelebA.

well as no severe occlusions such as sunglasses, and performing Delaunay triangulation on it.

Each feature vector is normalized to unit length before being passed as input to a classifier.

For each type of feature vector we train a multi-layer perceptron for each binary attribute. The network contained three hidden layers with 200, 100, and 75 neurons, respectively. We used ReLU activations and binary logistic loss for training. Training was performed using ADAM [10] with a constant learning rate of 0.001, L2 regularization with $\alpha = 0.0001$, batch size of 200, and a stopping criterion of change tolerance 0.0001.

IV. EVALUATION

We evaluated and compared the feature vector types on a color image dataset and an infrared image dataset. We describe here the datasets and face alignment method used, evaluation procedures, and results.

A. Datasets

For color face imagery we used the CelebA dataset [16]. Sample images are shown in Figure 2. It contains 202,599 faces of 10,177 people. Each image is labeled with 40 binary attributes. We removed from consideration attributes clearly un-related to facial landmarks such as hair styles, wearing eyeglasses, facial hair. The fourteen attributes we retained are: arched eyebrows; attractive; big lips; big nose; chubby; double chin; high cheekbones; male; mouth slightly open; narrow eyes; oval face; pointy nose; smiling; young. Some of these are obviously geometric; others, such as chubby, are not as obvious and may benefit from non-geometric data.

This list of attributes unfortunately mixes categorical and comparative or relative attributes [18]. However, we used these attributes because they have been used in previous work [16], [20] and thus enable a fair comparison.

Data balancing is a significant issue in the CelebA dataset, meaning that some attributes are severely under- or over-represented. For example, most celebrities ares not chubby. Male, high cheekbones, smiling, and mouth slightly open have the most balanced representations, while chubby, double chin, and narrow eyes have very few positives.

For infrared imagery we used the SCFace dataset [9]. Sample images are shown in Figure 3. In total, the dataset contains 4,160 images of 130 subjects. The images are in various sizes, from 1600×1200 down to 224×168. SCFace includes images from five cameras at three different distances. Two of these cameras also have infrared functionality, creating a total of seven cameras. In addition, it contains less challenging mug shots, both in color and infrared, as well as an unused series that rotates each subject. SCFace is used solely for testing in our experiments, not for training.

Fig. 3. Sample images from SCFace, with infrared images on the bottom row. From left to right: distance 1 (largest distance and smallest image size); distance 2 (middle distance and image size); and distance 3 (smallest distance and largest image size).

SCFace provides the gender of each subject but no other useable attribute labeling. To compensate, we manually labeled each subject with a subset of the fourteen attributes mentioned above (we omitted attributes unrelated to the identity of the person). We note here that, similar to CelebA, SCFace is biased in terms of attribute balancing; for example, only 15 of the subjects are female, and the majority are young.

B. Face Alignment

For facial landmark detection we used the publicly available Chehra software package [3]. Chehra provides 2D landmarks and yaw, pitch, and roll estimates. We only made use of the 2D landmarks for attribute recognition. Chehra detects 49 face landmarks (see Fig. 1). In our preliminary evaluations we also experimented with OpenFace [5] but found that it led to a slight decrease in recognition accuracy.

C. Evaluations and Results

1) Feature Type Comparison: We first compare the performance of our proposed classifiers with state-of-the-art attribute recognizers when trained and tested on color imagery.

TABLE I

ERROR RATES (%) OF VARIOUS FEATURE VECTOR TYPES ON CELEBA

	Locations	**Distances – Simple**	**Distances – Complete**	**Angles – Simple**	**Angles – Complete**	**Triangles**
Arched Eyebrows	23.13	27.36	23.01	**22.29**	23.49	22.38
Attractive	**25.89**	30.96	26.25	29.49	26.22	27.16
Big Lips	32.42	31.96	31.85	**15.45**	32.04	30.71
Big Nose	18.74	21.62	19.89	21.70	18.98	**18.32**
Chubby	5.35	5.35	5.28	6.08	5.43	**5.28**
Double Chin	4.63	4.63	4.63	4.91	4.64	**4.57**
High Cheekbones	19.42	22.53	**18.26**	19.12	18.48	19.65
Male	**10.68**	19.94	12.38	15.74	11.76	11.44
Mouth Slightly open	11.48	11.14	**10.66**	17.48	10.77	10.79
Narrow Eyes	14.60	14.76	14.32	**7.43**	14.27	14.02
Oval Face	29.73	29.89	29.15	28.98	29.08	**28.15**
Pointy Nose	28.80	28.74	28.45	28.78	28.16	**28.10**
Smiling	11.07	15.60	11.82	12.57	**10.63**	12.76
Young	19.63	21.30	20.48	23.87	20.91	**19.37**
Average	18.26	20.41	18.32	18.14	18.20	**18.05**

We trained and tested our classifiers on the CelebA [16] dataset.

Table I shows the accuracy of each feature vector type over all attributes when trained and tested on the CelebA dataset. The best feature type varied per attribute, and all had similar performance. Because **Triangles** resulted in the lowest average error rate, we use that feature type for all remaining experiments, and simply refer to **Triangles** as "Ours" in the other tables. Triangle areas have high invariance to scale and rotation, which may explain the good performance obtained here.

2) Method Comparison on Color Imagery: For detailed comparison, we compare the error rate of our geometric method for each attribute to the error rates for two state-of-the-art systems, LNets+ANet and MOON. Table II gives the results.

"Bias" refers to the error rate obtained by simply predicting the majority attribute for each image. For example, because the dataset contains more female than male faces, "Bias" will always predict female.

3) Method Comparison on Infrared Imagery: To test cross-modal attribute recognition, we tested our method and MOON on infrared imagery from the SCFace dataset. Both our method and MOON were trained on color imagery from CelebA – neither method had access to infrared imagery at testing time. To test infrared images with MOON, which expects three-channel color images as input, the infrared channel is replicated to the three input channels.

Our method can only be used when the face alignment method succeeds, whereas MOON can be used on any face image. For fair comparison, we computed error rates for both methods only on images where the face alignment method succeeded. Of the 910 infrared images available, 244 (27%) could be successfully aligned by Chehra, with the majority coming from the high quality camera 8 and from the images taken nearer to the subject. When face alignment did succeed, it was sometimes inaccurate because of the lack of contrast in the infrared images. Figure 4 illustrates the results of face alignment on the images.

Results are shown in Table III. Our method outperforms MOON on all attributes. However, in most cases, "Bias" is as good or better than either classifier. The low error rate of "Bias" reflects the lack of attribute variation in this dataset. The balance of attributes is different between the training dataset (CelebA) and the testing dataset (SCFace).

V. DISCUSSION

We considered several different geometric feature types and found that they are all roughly similar in performance. We selected triangle areas as our reference feature type since it had the best average accuracy across all attributes tested.

Our method, geometric attribute classifiers based on triangle areas, is not as accurate as the state-of-the-art MOON system when trained and tested on color images. However, it is indeed more accurate than competing methods FaceTracer and LNets+ANet for some attributes. In fact, the average accuracy across all attributes tested is better than FaceTracer, even though our method uses no photometric information for the classification.

More importantly, when trained on color images and tested on infrared images, our method significantly outperforms MOON, a state-of-the-art method. For all attributes tested, our method achieved a lower error rate. Our average error rate was 11.71% compared to 18.19% for MOON. This suggests that geometric information is better preserved when considering cross-modal data. Classifiers trained on color imagery may not fare well when presented with images from other modalities, because of their reliance on illumination and other effects present in the color images. The generality of this conclusion is limited by the fact that we were only able to test on the 216 infrared images where the facial alignment tool succeeded, and that some attributes were severely under-represented in this dataset.

A drawback of our approach is that it relies on a facial alignment tool in order to find facial landmarks. This tool itself will likely have been trained on color imagery, and its performance on other modalities may suffer as a result. Indeed, the facial alignment tool failed on many of the infrared images which were excessively blurry or low resolution. Despite this drawback, our method did outperform the

TABLE II

ERROR RATES (%) ON CELEBA DATASET

	Bias	FaceTracer	LNets+ANet	MOON	Ours
Arched Eyebrows	28.44	24	21	**17.74**	22.38
Attractive	49.58	22	19	**18.33**	27.16
Big Lips	32.70	36	32	**28.52**	30.71
Big Nose	21.20	26	22	**16**	18.32
Chubby	5.30	14	9	**4.56**	5.28
Double Chin	4.57	12	8	**3.68**	4.57
High Cheekbones	48.18	16	13	**12.99**	19.65
Male	38.65	9	2	**1.9**	11.44
Mouth Slightly Open	49.51	13	8	**6.46**	10.79
Narrow Eyes	14.87	18	19	**13.48**	14.02
Oval Face	29.56	36	34	**24.27**	28.15
Pointy Nose	28.57	32	28	**23.54**	28.10
Smiling	49.97	11	8	**7.4**	12.76
Young	24.29	20	13	**11.92**	19.37
Average	30.39	20.64	16.86	**13.63**	18.05

TABLE III

ERROR RATE (%) ON SCFACE INFRARED DATASET

	Bias	MOON	Ours
Arched Eyebrows	12.94	13.05	**12.35**
Attractive	**10.00**	15.58	15.29
Big Lips	**1.18**	5.47	**1.18**
Big Nose	**1.18**	12.42	11.76
Chubby	**2.94**	18.74	**2.94**
Double Chin	**1.18**	2.74	**1.18**
Male	**12.35**	13.89	13.53
Oval Face	**20.00**	34.53	**20.00**
Pointy Nose	**1.18**	6.32	**1.18**
Young	**19.41**	59.16	37.65
Average	**8.24**	18.19	11.71

Fig. 4. Examples of face alignment results on SCFace subject 102. *Left:* Color image (Camera 5; Distance 3; Image size 168×224). *Center:* Infrared image (Camera 7; Distance 3; Image size 168×224). Note the inaccuracy in the eyebrow fiducials. *Right:* Infrared image failure case (Camera 7; Distance 2; Image size 108×144). Note the lack of contrast and sharpness due to infrared illumination and distance to the subject.

state-of-the-art on the infrared images for which the facial alignment tool produced an output.

VI. CONCLUSIONS AND FUTURE WORK

This work presents an approach to cross-modal facial attribute recognition which uses purely geometric information as input to the classifiers. Because we specifically avoid the use of photometric information in the classifier, we achieve better performance than the state-of-the-art when training on color imagery and testing on infrared imagery.

The ability to retain performance when training on one modality and testing on another is important given the relative scarcity of labeled cross-modal image datasets. Our geometric method provides one possible approach to successful cross-modal recognition with faces.

Attribute classification from geometric information performed surprisingly well, despite not using any form of color or texture information in the classifier. Geometric classification even performed well for attributes which are not obviously geometric, such as “Attractive,” “Male,” and ”Young.” We did not consider clearly color-rated attributes such as hair color or ethnicity.

We found that a feature vector of triangle areas gave the best average error rate in our preliminary tests. However, because different feature types performed better for certain attributes, it may be advantageous to combine different feature types in the classifier. This is a topic for future work.

A current limitation of our approach is its reliance on existing facial alignment tools for facial landmark detection. Existing tools are trained exclusively on color imagery and thus are prone to failure on some infrared images. Despite this limitation, we were able to outperform the state-of-the-art on infrared imagery using an existing tool. However, the development of facial alignment tools which can perform better on cross-modal imagery would be crucial to further advancement of our geometric approach.

In addition, existing facial alignment tools are typically trained on high resolution images and thus tend to fail when presented with small, blurry face images as commonly encountered in surveillance applications. The development of a facial alignment tool which is more robust to blur and low resolution would likely lead to an improvement for our system. 3D landmarks might also give more useful information for some attributes, such as “big nose.”

Attribute subjectivity is an important issues for attribute classification. Some attributes such as “young” or “attractive” are highly subjective and are likely to be unreliably labeled. This issue may affect confidence in the results of experiments on attribute classification.

The presence of severe dataset bias in existing datasets is also an issue, since it greatly influences classifier learning and testing. MOON [20] addresses dataset bias by explicitly compensating for it in a mixed-objective loss function. We expect that incorporating this weighted, mixed objective into our system will lead to an improvement in classification accuracy.

In this work, we compared against existing, un-modified attribute recognizers trained on color imagery. This allowed us to consider the potential of existing systems for color-to-infrared learning transfer. It is likely that re-training existing systems on infrared would improve their performance. However, the lack of abundant infrared imagery precludes this approach currently. An interesting alternative, and a possible topic of future work, would be to re-train existing systems such as MOON [20] on only the red channel of color imagery, as an approximation to the infrared appearance. Existing facial landmark detectors may also perform better on infrared imagery when re-trained on red-channel images.

VII. ACKNOWLEDGMENTS

The authors gratefully acknowledge the anonymous reviewers for their comments and suggestions for improvement, and Manuel Günther and Ethan Rudd of the UCCS VAST Lab for their helpful suggestions and for testing MOON on the SCFace dataset.

REFERENCES

[1] N. Y. Almudhahka, M. S. Nixon, and J. S. Hare. Unconstrained human identification using comparative facial soft biometrics. In *Biometrics Theory, Applications and Systems (BTAS), 2016 IEEE 8th International Conference on*, 2016.

[2] B. Amberg, R. Knothe, and T. Vetter. Expression invariant 3D face recognition with a morphable model. In *2008 8th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2008*, 2008.

[3] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic. Incremental face alignment in the wild. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1859–1866, 2014.

[4] T. Baltrušaitis, M. Mahmoud, and P. Robinson. Cross-dataset learning and person-specific normalisation for automatic Action Unit detection. In *FG*, volume 06, pages 1–6, 2015.

[5] T. Baltrusaitis, P. Robinson, and L. P. Morency. OpenFace: An open source facial behavior analysis toolkit. In *2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016*, 2016.

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

[7] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active Shape Models-Their Training and Application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[8] N. Dalal and B. Triggs. Histogram of Oriented Gradients for Human Detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005*, volume 1, pages 886–893, 2005.

[9] M. Grgic, K. Delac, and S. Grgic. SCface - Surveillance cameras face database. *Multimedia Tools and Applications*, 51(3):863–879, 2011.

[10] D. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations*, pages 1–13, 2014.

[11] B. F. Klare, S. Klum, J. C. Klontz, E. Taborsky, T. Akgul, and A. K. Jain. Suspect identification based on descriptive facial attributes. In *IJCB 2014 - 2014 IEEE/IAPR International Joint Conference on Biometrics*, 2014.

[12] N. Kumar, P. Belhumeur, and S. Nayar. FaceTracer: A search engine for large collections of images with faces. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 5305 LNCS, pages 340–353, 2008.

[13] N. Kumar, A. Berg, P. N. Belhumeur, and S. Nayar. Describable visual attributes for face verification and image search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(10):1962–1977, 2011.

[14] X. Li, G. Mori, and H. Zhang. Expression-invariant face recognition with expression classification. In *Third Canadian Conference on Computer and Robot Vision, CRV 2006*, volume 2006, 2006.

[15] X. Li and H. Zhang. Adapting geometric attributes for expression-invariant 3D face recognition. In *Proceedings - IEEE International Conference on Shape Modeling and Applications 2007, SMI'07*, pages 21–30, 2007.

[16] Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 11-18-Dece, pages 3730–3738, 2015.

[17] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[18] D. Parikh and K. Grauman. Relative attributes. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 503–510, 2011.

[19] S. Ren, X. Cao, Y. Wei, and J. Sun. Face alignment at 3000 FPS via regressing local binary features. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1685–1692, 2014.

[20] E. M. Rudd, M. Günther, and T. E. Boult. MOON: A mixed objective optimization network for the recognition of facial attributes. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 9909 LNCS, pages 19–35, 2016.

[21] D. A. Vaquero, R. S. Feris, D. Tran, L. Brown, A. Hampapur, and M. Turk. Attribute-based people search in surveillance environments. In *2009 Workshop on Applications of Computer Vision, WACV 2009*, 2009.

[22] X. Xiong and F. De La Torre. Global supervised descent method. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 07-12-June, pages 2664–2673, 2015.

[23] K. Zhao, W. S. Chu, F. De La Torre, J. F. Cohn, and H. Zhang. Joint patch and multi-label learning for facial action unit and holistic expression recognition. *IEEE Transactions on Image Processing*, 25(8):3931–3946, 2016.