Facial Component Detection in Thermal Imagery

Brais Martinez

Department of Information and Communication Technologies

Universitat Pompeu Fabra, Barcelona, Spain

`brais.martinez@upf.edu`

Xavier Binefa

Department of Information and Communication Technologies

Universitat Pompeu Fabra, Barcelona, Spain

`xavier.binefa@upf.edu`

Maja Pantic

Computing Department / EEMCS

Imperial College University, London / Twente University

`m.pantic@imperial.ac.uk`

Abstract

This paper studies the problem of detecting facial components in thermal imagery (specifically eyes, nostrils and mouth). One of the immediate goals is to enable the automatic registration of facial thermal images. The detection of eyes and nostrils is performed using Haar features and the GentleBoost algorithm, which are shown to provide superior detection rates. The detection of the mouth is based on the detections of the eyes and the nostrils and is performed using measures of entropy and self similarity. The results show that reliable facial component detection is feasible using this methodology, getting a correct detection rate for both eyes and nostrils of 0.8. A correct eyes and nostrils detection enables a correct detection of the mouth in 65% of closed-mouth test images and in 73% of open-mouth test images.

1. Introduction

Thermal Images are receiving greater attention lately due to the appearance of cheaper sensors and the wide range of applications of such imagery. These applications include surveillance and security, medical imaging, biometrics, detection of deceit, elder care and military research. Furthermore, thermal imagery is perceived as non intrusive in terms of privacy, in contrast to regular (rgb) cameras.

In the case of facial thermal images, and as far as the problem of facial component detection is concerned, it is important to note the following: thermal images have typically a low signal to noise ratio, edges are diffuse and therefore less important than in color images (e.g. for eyebrows and lips contour detection) and the knowledge applicable for color face images is not, in general, applicable to thermal images. More specifically on the latter, the iris is hardly visible and there is no contrast with the sclera, the eyebrows are not consistently visible since this depends on their density, and the lips are in many cases undistinguishable and therefore the mouth is hardly distinguishable if it is closed. Furthermore, the face aspect is not affected by illumination changes. Although the images can change because of, for example, the heat in the room, these changes are subtle and not comparable to those introduced by variations in lightning in the case of color images. Some general considerations about thermal images can be found in [6].

Facial component detection in thermal imagery is relevant for face registration, tracking initialization (*e.g.* in [2] and [11] the tracking is initialized manually), face recognition ([9],[5]), and expression recognition ([10], [4]).

To the best of our knowledge, the only reported effort on facial component detection in thermal imagery is that presented in [10]. In here, a feature point extractor based on Harris features is applied. Then clustering using k-means is performed, under the assumption that the clusters will be coincident with the facial components. The objective is to detect both eyes and mouth. In general, this approach does not attain high precision; the nose is not detected at all and the detection of the mouth fails whenever the mouth is closed.

1

Our aim is to detect both eyes, both nostrils and the mouth. We divide the problem into the detection of eyes and nostrils and the detection of the mouth. This is due to the different nature of the problems and that the mouth detection is more challenging. Therefore, we consider the eyes-and-nostrils detection auxiliary to the mouth detection.

For the eyes-and-nostrils detection, we follow a general-to-specific approach in which the face region is subdivided into patches, each of which is further analyzed in terms of whether it contains the target facial component or not. The analysis is performed by two successive classifiers, called a detector and an identifier, following the notation in [1]. This process is designed to increase the robustness and alleviate the computational expense of an exhaustive search. The analysis is performed by means of a GentleBoost classifier [3]. We also tested and compared other classifiers including Support Vector Machines (SVM), Gaussian Mixture Models (GMM), and Subclass Discriminant Analysis (SDA) [14]. We show experimentally that all these provide lower performance when compared to GentleBoost. When discussing the classification problem, the risk of overfitting should be considered, due to the lack of extensive thermal imagery databases and due to the properties of thermal images. More precisely, facial thermal images are consistent with respect to intrasubject variation (they are considered as a very reliable biometric), mostly due to the different facial vascularization, and factors as fat deposits, age, etc. We also define an adequate procedure in order to obtain a valid training set for the case of thermal images, specially defined to deal with the large amount of clutter present in such images. We call it a two-stage training set. Finally, a set of anthropomorphical restrictions are applied to obtain the final eyes-and-nostrils detection. This process selects the most consistent set of the positives provided by the classifiers, based on the consistency of their spatial relations, which are modeled in a training stage by means of a set of gaussian distributions.

Based on this detection and after normalizing the face position in terms of translation, rotation, and scaling, a Region of Interest (ROI) for the mouth detection is selected. The aspect of the mouth greatly varies for the case of an opened mouth and a closed mouth. When the mouth is closed, the problem becomes very challenging since the mouth is sometimes barely distinguishable even to the naked eye. Furthermore, the interior of the mouth or the juncture of the lips can either appear as whiter or darker (hotter or colder). We thus divide the mouth detection into the opened-mouth case and the closed-mouth case. The first case is performed by using an entropy-based approach, while the latter is tackled adapting the concept of Self-similarity [8].

Section 2 describes the approach to eyes-and-nostrils detection. Section 3 contains the mouth detection procedure. Section 4 provides details of the conducted experimental studies. Finally, section 5 concludes the paper.

2. Eyes-and-Nostrils Detection

For the detection of eyes and nostrils, all possible image patches with intensity corresponding to possible skin temperatures are considered. Each patch is decomposed into a feature vector by means of Haar wavelets. Two successive classifiers are applied to each of the patches. The first one, called the detector, is coarser but less computationally demanding. The second one, called the identifier, is more precise and computationally expensive; it is only applied over the positives provided by the first classifier. A point is considered as a positive if it is a positive for both of the two successive classifiers. In section 2.1, the image patch representation by means of Haar wavelet decomposition is presented. Section 2.2 explains the procedure followed in order to obtain the training set used for training each of the classifiers. This procedure is the same for both the detector and the identifier. This section also contains the first tests conducted in order to select the best classifier for this task, showing the superior performance of SVM and GentleBoost. Finally, the process followed to first model and then apply the set of anthropomorphical restrictions is described.

2.1. Patch Representation Through Haar Wavelet Decomposition

A patch is a subwindow of an input image of a predetermined size. Each patch can be represented as a coefficient vector, where each coefficient is the result of applying a filter onto the patch. Our work uses Haar wavelets, which are templates containing ± 1 values that can be interpreted as coarse derivatives. As in [7], the filter bank used in this work contains 3 orientations (horizontal, vertical and diagonal) and is overcomplete. That means that each filter is applied to positions with 50% overlap (*e.g.*, a patch of size 8×8 is applied to positions with a displacement of 4 pixels).

We consider Haar features suitable for application to thermal imagery since: thermal imagery has a low signal-to-noise ratio and Haar features are robust to noise, Haar features respond to coarser aspects of the image (in thermal imagery there are no subtle textures as is the case in color images, which would be modeled more appropriately by, *e.g.*, Gabor filters) and they model well diffused edges, characteristic of thermal imagery. They are also computationally efficient [13].

As mentioned above, we use two successive classifiers with different level of detail. In the detector, we use a patch size of 16×16 and a filter bank containing wavelets of sizes 8×8, 4×4, and 2×2. The identifier uses a patch size of 32×32 and a filter bank containing wavelets of sizes

16×16, 8×8, 4×4, and 2×2.

Patch intensity normalization is a common pre-processing step. For color images it is used for handling the effects of lightning. In contrast to color images, the lightning conditions do not affect thermal imagery. Hence, the use of a patch intensity normalization is not needed and the results obtained with and without the intensity normalization are comparable, as shown in the experimental results.

2.2. Two-stage Training Set

It is usual to build a training set by including examples of patches containing the target facial component, patches that do not or only partially contain the target facial component, and some random background patches. This procedure is effective when dealing with rgb images, since the patches likely to be confused with the positives examples are those partially containing the target facial component. For the case of thermal imagery, the situation is not the same. One of the most significant problems when it comes to facial component detection is the high level of clutter present in the image. Furthermore, in facial thermal imagery 'foreground' stands for parts of the face containing the target facial component, while the 'background' stands for regions of the face that do not contain any part of the target facial component. Thus, the clutter fully or almost fully belongs to the face region and therefore is less variable and can be more easily modeled than is the case for rgb imagery. Directly applying the usual procedure for the case of rgb images to thermal imagery produces a training set where the clutter is not represented properly.

To solve this issue and properly represent the clutter, we construct a two-stage training set. After constructing a first training set using the aforementioned procedure, a classifier is trained on it. Using this classifier, facial component detection is conducted using the complete training set and the false positives are kept. Then, a new training set is built so the negative example set is now composed by a 25% of randomly selected false negatives and a 75% of negative examples randomly selected from the original training set. The proportion between negative and positive examples is kept to 2:1. Figure 1, suggests a great reduction of false positives attained with the two-stage training set. We trained the two different classifiers with the same training set constructed as described above.

We considered the following classification methodologies, which range from generative to discriminative classifiers: SVM, GentleBoost, GMM, and SDA. A well-known characteristic of SVM is that it needs a feature selection procedure as inclusion of meaningless features in the training stage usually implies poor performance results. To tackle this issue, we compute the most discriminative features using GentleBoost, which has been reported to give good results as a feature selection method preceding SVM classifi-
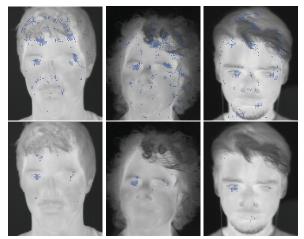

Figure 1. Performance of the classifier for the left eye constructed using single (top row) and two-stage training sets (bottom row).

	precision	recall	F1 score
SVM	0.946 \ 0.941	0.931 \ 0.910	0.934 \ 0.914
GMM	0.844 \ 0.732	0.853 \ 0.950	0.842 \ 0.818
SDA	0.879 \ 0.922	0.860 \ 0.879	0.857 \ 0.892
GentleBoost	0.921 \ 0.932	0.890 \ 0.926	0.902 \ 0.924

Table 1. Performance of each classifier for the left eye detection $without \backslash with$ intensity normalization. The experiment was run over the training set and using Leave-one-subject-out crossvalidation

cation [12]. Thus, whenever we refer to SVM we actually mean Gentle-SVM. In the case of GMM, the selection of the number of components of the GMM is decided experimentally, *i.e.*, by testing all possibilities from 1 to 15 and keeping the best result. The kernel used for the SVM is a RBF kernel. The parameter estimation for each gaussian is performed by means of an Expectation Maximization (EM) algorithm.

The evaluation of these different classification methods was first performed by means of leave-one-subject-out crossvalidation over the training set. The results show that the best performance is achieved by SVM when no patch intensity normalization is conducted ($F1$ score = 0.934). GentleBoost with intensity normalization achieves the second best performance, with a $F1$ score = 0.924. SDA and GMM showed poorer results, as shown in table 1. Given these results, we decided to focuss on both SVM and GentleBoost in further evaluation studies.

2.3. Selection of an Anthropomorphically-Consistent Eyes-and-Nostrils Configuration

The facial component detectors presented above can still produce false positives or fail to detect the target component. We proceed by exploiting the knowledge about

the morphology of the human face. More specifically, we model possible facial configurations (in terms of facial components) by using a probabilistic approach. We use the ground truth data about the centers of the facial components (manually annotated in the training set) to learn possible facial configurations. In here, a configuration is a set of facial component centers. From each such facial configuration it is possible to extract its shape. That is to say, the facial configuration can be normalized in terms of translation, rotation and scale. The shape is extracted by estimating and applying a procrustes transformation.

In order to model the valid configurations, the set of all possible shapes, $\{S_j = (s_j^1, s_j^2, s_j^3, s_j^4)\}_{j=1,\ldots,n}$, is considered. Then, the possible positions of each facial component i, $\{s_j^i\}_{j=1,\ldots,n}$ ($i = 1,\ldots 4$ denoting left eye, right eye, left nostril, right nostril), is modeled by fitting one gaussian ($\{\mu_i, \Sigma_i\}$) to it. We considered the use a Gaussian Mixture Model, but the point distributions was already very close to a Gaussian distribution.

To asses whether a facial configuration obtained as the output of the two-stage classification procedure explained above is a possible facial configuration, the shape $S_0 = (s_1, s_2, s_3, s_4)$ is extracted based on the detected centers of the facial components. The score for this particular facial configuration is given by $p(S_0) = \prod_{i=1,\ldots,4} N(s_i, \mu_i, \Sigma_i)$. Here N denotes the normal distribution. Thus, the probability of each facial component position is considered independently.

Since the two-stage classifier is applied on multiple image patches, several image patches are usually classified as positives for a given facial component. After evaluating all facial configurations formed by all positives found for the four components, the one having the highest $p(S_0)$ is kept. If this score exceeds a fixed value, the facial configuration is accepted as the final output. Note that this procedure is exhaustive. More specifically, if the number of positives is n_1, n_2, n_3, n_4 for the four facial components, then there are $n_1 \cdot n_2 \cdot n_3 \cdot n_4$ combinations to be evaluated. Despite of this exhaustive procedure, the low number of false positives resulting from the sequential application of the detector and the identifier, and the use of single gaussians to model the probabilities, keep the number of combinations and the computational cost low.

In the case that the score ($p(S_0)$) is under a fixed value, the absence of a correct detection for at least one of the facial components is assumed. We then proceed with testing all hypotheses in order to find out which facial component was inaccurately detected. To this end, we proceed by computing all scores leaving one facial component out. Since it is possible to conduct normalization in terms of translation, rotation and scaling with just 3 points (we are forced to assume that only one of the points is incorrectly detected, otherwise the shape cannot be extracted), the process men-
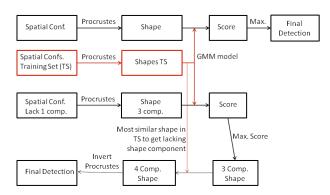

Figure 2. Flow chart of how the best facial configuration is selected. Red indicates training.

tioned above is repeated: the shape is extracted with just 3 components and $p(S_0)$ score is computed. After repeating this process for all components, the maximum score is used to decide on the component that was incorrectly detected. Given the shape that the positions of the three accurately detected components form, the most similar shape is extracted from the training set. Based on this data, the position of the inaccurately detected facial component is retrieved. Finally the inverse of the procrustes transformation is applied to recover the spatial configuration.

The result of an effective application of this tactic is shown in figure 5, where the detection of the left nostril was inaccurate.

3. Mouth Detection

Mouth detection is a very complicated issue in facial thermal imagery. Two very different situations can be distinguished depending on whether the mouth is open or closed. The detection in the first case is simpler since there is a big difference in the temperature inside and outside the mouth, although it can still be considered as challenging due to the great variability of the mouth aspect, specially for the case of expressive faces, and the fact that the interior of the mouth can appear as hotter or colder than the skin. Furthermore, teeth, tongue, and the rest of the interior of the mouth can present different temperatures as well. When the mouth is closed, the detection is very difficult as it considers distinguishing the lips from the rest of the face, which is sometimes very complicated even to the naked eye. Any approach based on edges and contours is inapplicable because of the lack of edges in thermal imagery and because lips are very often indistinguishable from the rest of the skin. The use of interest points is similarly inapplicable and the patch representation using wavelets is not feasible due to the extreme differences in the appearance of the mouth. These differences are due to flexible movements of the mouth as well as due

to differences in the physiognomy of subjects. We thus present a methodologically new approach, based on the concept of self-similarity [8].

The first required process is the selection of a ROI for the mouth detection. To this end, the detected positions of eyes and nostrils are used to normalize the image to a frontal, unrotated and scale-normalized face image. Then a predefined region, defined to cover all the mouths present in the training set, is selected.

Then, the hypothesis of the presence of an opened mouth is tested. To this end, we use an entropy-based approach. We compute the intensity-based entropy in the ROI. The entropy is computed in patches of different sizes. In our experiments, the patches are of size 10x10 and 5x5, while the face image is resized so the ROI size is of 30x60 on average. If the entropy is high, the mouth is considered to be opened, otherwise, the mouth is considered to be closed. The entropy of patch P is formally computed as:

$$ent(P) = -\sum_i h_i(P) \log h_i(P) \qquad (1)$$

where $h_i(P)$ is the ith component of the greyscale intensity histogram of P. Some results of this process are illustrated in figure 6.

If the calculated entropy is low, then the mouth is considered to be closed. In this case we assume a particular appearance structure and try to take advantage of it. Due to the nature of thermal images, the lips can be sometimes undistinguishable from the rest of the face. Nonetheless, the juncture of the lips is commonly visible, although it can sometimes be hotter and sometimes colder than the lips (i.e., it appears as either a whiter or a darker area between the lips; see figure 3). Hence, we assume that the mouth is a rectangular area having similar intensity values along the x axis and variable intensity values along the y-axis, where the appearance of the area is approximately symmetric in respect to the x-axis. We then apply a similar approach to that of self-similarity: for any image position candidate to be the mouth center, a small patch centered at it is correlated with patches displaced along the x-axis. By multiplying the correlation coefficients, the output is a measure of how constant the image intensity is along the x-axis (S_{hor}). We repeat the process for the y-axis (S_{vert}) and for the symmetry constraint (S_{symm}). We then calculate a coefficient S_{tot} as:

$$S_{tot} = S_{hor} \cdot (S_{vert})^{-1} \cdot S_{symm} \qquad (2)$$

After repeating this process for each mouth position candidate, the one providing the maximum value for this coefficient is taken as the mouth position. The motivation for using this coefficient is visible in the mouth-detection ROI in figure 3.

4. Experimental Results

We used a database collected in our lab by recording 10 females and 12 male subjects of varying age (24-45) and being of European origin (the database will be made available online via www.mahnob-db.eu). For each subject, several videos were recorded on different days in an indoor environment. The experiments presented here are based on 78 images, 3-4 per subject. The position of the face varies to some extent, with moderate rotations inside and outside the image plane. The images were chosen as to picture expressionless as well as expressive faces. The expressive faces were recorded while subjects were watching stimulus material (mostly funny videos), or were asked to pose certain facial expressions (e.g. joy). Some images include background clutter. The annotation of the ground truth in terms of facial components was made manually using the annotations made by two different operators and averaging them, in terms of patch location enclosing a target facial component.

The evaluation of the proposed methodology has been conducted using leave-one-subject-out crossvalidation. In other words, images of all but one subject are used for training purposes while the images of the left out subject are used for testing. The procedure is repeated for all subjects in the dataset. The presented results are the average of results obtained for all subjects. We adopt this procedure since it guarantees good generalization performance of the method on unseen data.

The training set was created using, as positive examples, the patch centered on the target facial component and patches being displaced for 2 pixels from it, keeping a total of 9 positive examples per image. These patches were taken from the resized thermal images, so that the facial component image is 16×16 pixels for the first classifier, 32×32 for training the second classifier). The negative examples were created by shifting the patch for 4 to 6 pixels away from the center of the target facial component. A random selection of the negative examples was performed in order to keep a good proportion (2 negative examples per positive example). An additional 10% of negative examples were included representing patches randomly selected from the background. Some of the positive examples for the left eye and left nostril appearing in our database are shown in figure 3. Each patch in the training set was then transformed, as described in section 2.1, to a 849 feature vector for the first classifier and to a 3732 feature vector for the second classifier. Both the GentleBoost and the Gentle-SVM reduce the dimensionality of this data to 40-70 features, and to 70-100 features respectively. The output of the classifiers is binary, i.e. with possible labels $\{\pm 1\}$. Examples of detected eyes and nostrils are shown in figure 5. Examples of both true positives and false positives are illustrated. The third example shows a faulty detection of the left nostril as

Figure 3. Examples of patches containing eyes, nostrils and mouths. Note that the juncture of the lips is sometimes whiter and sometimes darker than the lips.

	Eyes	Nostrils
GentleBoost ($err \leq 0.1\backslash0.15$)	0.769 \ 0.833	0.782 \ 0.820
SVM ($err \leq 0.1\backslash0.15$)	0.718 \ 0.756	0.679 \ 0.731

Table 2. Performance for LOO crossvalidation for 78 images of 22 subjects.

output by the cascade classifier. The fourth example shows the final result of the nostrils detection, corrected using the anthropomorphically consistent eyes-and-nostrils configuration selection.

To evaluate the performance of the proposed detectors, each of the automatically located centers of the target facial components was compared to the true (manually annotated) center. The displacement (error), normalized to account for variation in face scale is computed as:

$$err_{eye/nostril} = \frac{max(\|P_l - \hat{P}_l\|, \|P_r - \hat{P}_r\|)}{\|P_l - P_r\|} \quad (3)$$

where P_l is the true position of the left eye/nostril and $\hat{P}_l$ is the automatically detected position. The same holds for P_r and $\hat{P}_r$, where 'r' stands for 'right'. $\| \cdot \|$ stands for L_2 norm.

An automatically detected component the center of which is displaced in any direction from the true position so that $err < 0.15$ is regarded as success. Figure 7 shows the performance of GentleBoost and SVM classifiers. Figure 4 shows some examples of patches corresponding to the maximum error accepted, $err = 15$. The detection rates for eyes and nostrils are shown in table 2. In order to decide which
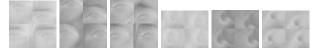

Figure 4. Image patches corresponding to $err = 0.15$.

of the two classification methods is more suited for the target problem (eyes and nostrils detection with $err < 0.15$) we perform a t-test as a criterium of statistical significance. That is to say, wether the differences in performance are, with a confidence margin of t, product of the particularities of our test set or the superior performance of one classifier. The results show that GentleBoost outperforms SVM

while $err < 0.15$ is statistically significant with margin 6% (t=0.06).

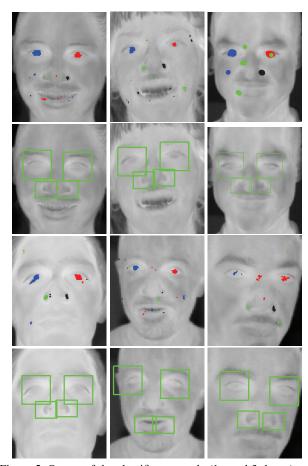The detection of the mouth is considered successful if the

Figure 5. Output of the classifiers cascade (1st and 3rd row) and final result (2nd and 4th row). The color code differentiates the target component. The examples shown include expressive faces (examples 1 and 2), small rotations out of the image plane (ex. 4 and 6), an error in the classifier cascade output accurately solved (ex. 3), and an erroneous classifier output (ex. 5).

overlap of the detected mouth region and manually annotated mouth region is at least 75%. This definition of success is different than the one provided for eyes and nostrils due to the variability of the size of the mouth patch. It is not possible to define a (scale-normalized) error in terms of pixels. Provided successful detection of the eyes and the nostrils, the proposed method achieves successful mouth detection in 73% of the cases of open mouth and in 65% of the cases of closed mouth. Since there is no training for the mouth detection, the whole dataset has been used as the test set. A graphic of the evolution of the successful mouth detection rates for varying error values is shown in figure 7.
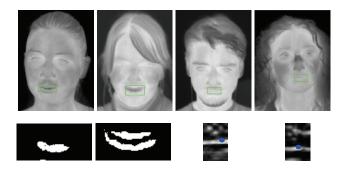
Figure 6. Mouth detection results for the case of entropy-based detection and self-similarity-based detection. The bottom row shows the thresholded entropy for the case of an opened mouth and the self-similarity surface for the case of a closed mouth
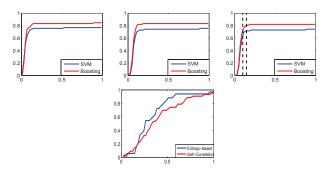

Figure 7. Detection results (from left to right): eyes detection, nostrils detection, eyes-and-nostrils, and mouth detection. Y-axis: percentage of successful detections. X-axis: err rate regarded as success. The vertical dashed lines mark err of 0.1 and 0.15.

5. Conclusions

In this article we present an automatic method for detecting facial components in facial thermal imagery. This work is a first step in applications such as face tracking, facial expression recognition, and face recognition. We show that detection of eyes and nostrils in thermal imagery is feasible and that the proposed method is reliable and rather accurate. The mouth detection is more complicated due to the nature of thermal images, but we present a novel method for obtaining an estimation of its position.

Acknowledgments

References

[1] P. Campadelli, R. Lanzarotti, and G. Lipori. *Face Recognition*, pages 367–394. I-Tech Education and Publishing, 2007. 2

[2] J. Dowdall, I. T. Pavlidis, and P. Tsiamyrtzis. Coalitional tracking. *Computer Vision and Image Understanding*, 106(2-3):205–219, 2007. 1

[3] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting. *Annals of Statistics*, 28(2):337–407, 2000. 2

[4] B. Hernández, G. Olague, R. Hammoud, L. Trujillo, and E. Romero. Visual learning of texture descriptors for facial expression recognition in thermal imagery. *Computer Vision and Image Understanding*, 106(2-3):258–269, 2007. 1

[5] S. G. Kong, J. Heo, B. R. Abidi, J. Paik, and M. A. Abidi. Recent advances in visual and infrared face recognition–a review. *Computer Vision and Image Understanding*, 97(1):103–135, 2005. 1

[6] N. Morris, S. Avidan, W. Matusik, and H. Pfister. Statistics of infrared images. *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–7, 2007. 1

[7] C. Papageorgiou and T. Poggio. Trainable pedestrian detection. In *International Conference on Image Processing*, pages 35–39, 1999. 2

[8] E. Shechtman and M. Irani. Matching local self-similarities across images and videos. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007. 2, 5

[9] D. Socolinsky and A. Selinger. Thermal face recognition in an operational scenario. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 2:II–1012–II–1019 Vol.2, 2004. 1

[10] L. Trujillo, G. Olague, R. Hammoud, and B. Hernandez. Automatic feature localization in thermal images for facial expression recognition. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, page 14, 2005. 1

[11] P. Tsiamyrtzis, J. Dowdall, D. Shastri, I. T. Pavlidis, M. G. Frank, and P. Ekman. Imaging facial physiology for the detection of deceit. *International Journal of Computer Vision*, 71(2):197–214, 2007. 1

[12] M. Valstar and M. Pantic. Combined support vector machines and hidden markov models for modeling facial action temporal dynamics. In *IEEE Int'l Workshop on HCI, LNCS*, pages 118–127. 3

[13] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:511, 2001. 2

[14] M. Zhu and A. M. Martinez. Subclass discriminant analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(8):1274–1286, 2006. 2