Thermal Face Recognition – Matching Algorithms Performance

Jan Váňa, Martin Drahanský, Radim Dvořák

ivanajan@fit.vutbr.cz, drahan@fit.vutbr.cz, idvorak@fit.vutbr.cz

Faculty of Information Technology
Brno University of Technology
Brno, Czech Republic

Abstract

Face recognition based on thermal images belongs to less known biometric methods. The method works with images of human face captured in near infrared light spectrum. Infrared face recognition can be either use in stand-alone applications, or as a part of multimodal biometric system in combination with other methods (e.g. face recognition based on 2D or 3D images in visible light spectrum). This article presents different methods of normalization and matching algorithms and shows how they affect the performance of the biometric system. Results of all described methods are summarized to see, which combination of normalization and matching algorithms have the best performance.

Keywords: biometric systems, human recognition, thermal face recognition.

1 Introduction

For a human, the most natural characteristic for the person recognition is a human face. Therefore, it is logical that the scientists try to find a way for automated person recognition according to the face.

The most common method is the processing of the face images that are in the visible color spectrum. However, it brings many problems such as the different conditions of illumination, the face rotation, the change of the face during the lifetime and the easiness of fake. These problems can be solved more or less successfully, nevertheless the other possibilities appear. One of them is the usage of the three dimensional shape of the face – the real devices already appeared in the market. The alternative way is to use the thermal characteristics of the face. In recent time we do the extensive research in this field.

Like any other biometric methods, this one has its own pros and cons. The big advantage is the independence of the external lighting. Thermal image can be captured with the same quality even at night. The second advantage is minimizing the possibility of spoofing an input image of capturing device by a non-authorized person.

The paper is organized as follows. The general image processing pipeline is described in section 2 and it focuses on different methods of normalization and matching algorithm. Section 3 presents performed tests and achieved results of all methods proposed in section 2. Last section contains short conclusion of the paper.

2 Thermal face recognition process

Main goal of any biometric system is to confirm a human identity. The standard face recognition based systems use image processing pipeline to fulfill this goal. This pipeline usually consists of face detection, face normalization and matching score evaluation.

2.1 Image Capturing

The common digital camera is sufficient for the capturing of the data in human visible color spectrum. The purchase costs of these devices are minimal in recent time. It is necessary to use the thermal camera (Figure 1) to capture the thermal images (the thermal camera works in certain part of the infrared light). Unfortunately, these devices are very rare (they are used only in several branches of industry), thus their purchase costs are much higher.

The parameters of surrounding environment (e.g. humidity, temperature) and the information of the examined object (e.g. material, distance, etc.) are needed for the best possible quality of thermal-image data [1].

Figure 1: FLIR ThermaCAM EX300 [8].

2.2 Image processing

If we passed raw thermal face images to matching algorithm without any pre-processing, we would get unacceptable results. It is caused by high within-class variability of these images. There are a lot of aspects that contribute to this negative property such a different surrounding temperature, different face position and orientation, distance from camera etc. Image processing tries to deal with all these aspect and decrease the within-class variability as much as possible.

2.2.1 Face detection

Detection of face is a first step of the image processing phase. The output of this part is not only locating the face itself, but also to find other important parts of the human face, such as locations of eyes, nose, mouth etc. Several approaches that solve this problem can be found in [1, 2, 9]. Face area and location of important points are manually annotated for purposes of this work.

2.2.2 Face normalization

The main part of this article deals with the description of the normalization methods and its impact on a system reliability. This section describes the different types of normalization.

Intensity normalization

Comparison of thermal images in absolute scale usually does not lead to the best results. The absolute temperature of the human face varies with the environment temperature, physical activity and emotional state of person. Some testing databases even do not contain information how to map pixel intensity on the temperature.

Intensity normalization is necessary because of these aspects. It can be achieved most easily by image histogram equalization (Figure 2).

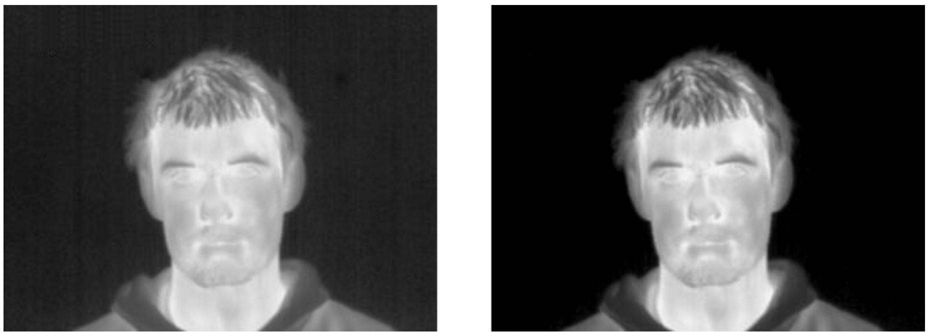

Figure 2: Intensity normalization by histogram equalization.

Geometric normalization

Biometric systems based on face recognition do not demand strict constraint for a head position in front of camera. The task of geometric normalization is to transform captured face to some default position (front view without any rotation). Fulfilling this task is one of the biggest challenges for 2D face recognition technologies. It is obvious that perfect solution cannot be achieved by 2D technology; however the variance caused by different positioning should be minimized as much as possible.

Geometric normalization often needs information about position of some important points within the human face. These points are usually image coordinates of eyes, nose and mouth. If they are correctly located, image can be aligned to a default template.

Basic methods of geometric normalization are based on the affine transformation, which is usually realized by transformation matrix T. Each point $p = [x, y]$ of original image I is converted to homogenous coordinates $p_h = [x, y, 1]$. All these points are multiplied by matrix T to get the new coordinates p'. This transformation is illustrated in Figure 3 [10].

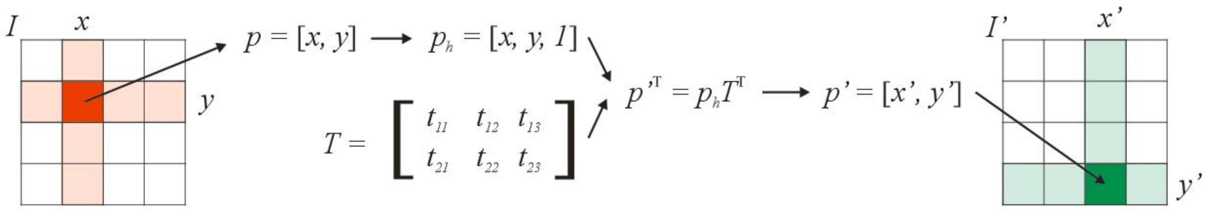

Figure 3: Image affine transformation illustration.

Methods of geometric normalization differ by computation complexity of determining transformation matrix:

- *Cropping:* This method crops original image I to region containing the face. This region is defined by rectangle $R = [x, y, w, h]$, where point $lc = [x, y]$ is position of top left corner and $[w, h]$ is size of the rectangle. Transformation is realized by matrix T which shifts original image by $[-x, -y]$. Obtained image I' is then cropped by window w x h (Figure 4a).
- *Scaling:* Transformation matrix T is composed by shift and scale matrices. The scale ratios are computed from the size of detected bounding box or from positions of important image points (Figure 4b).
- *General affine transformation:* This method is based on mapping of three different important points in original image I to their expected positions in new image I'. Transformation matrix coefficients are computed by solving simple set of linear algebraic equations (Figure 4c) [10].

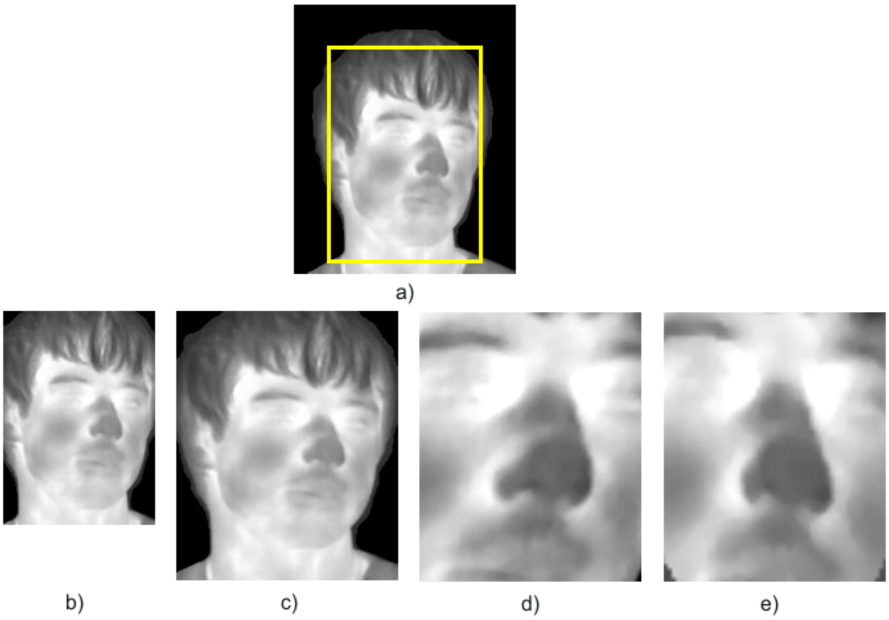

Figure 4: Geometric normalization; a) Original image with detected face area; b) Cropping; c) Scaling; d) General affine transformation; e) Mapping original image to 3D face model.

The next group of geometric normalization methods is based on expected shape of human face. The only tested method from this group is an algorithm based on ***mapping original image to 3D face model*** (Figure 4d).

Let M be the 3D model of the face. This model is manually annotated (positions of important face points are marked on the model surface) A_M. The input of the algorithm is the original image I and at least three locations of important face points in this image A_I. All corresponding points A_I have to be marked in model annotations A_M. The algorithm tries to find the best alignment of the 3D model to the face in the original image in a first step. Model transformation matrix T_M consists of translation, rotation and scale [10]. Model transformed by this matrix T_M projects all its annotated points A_{Mi} to corresponding locations in original image A_{Ii}. Coefficients of matrix T_M are found iteratively.

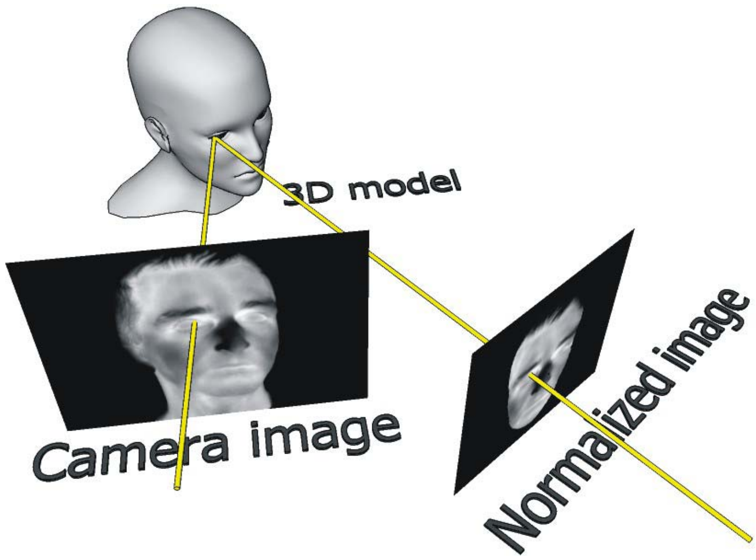

Figure 5: Original (camera) image geometric normalization by projecting texture to 3D model and re-projecting back to normalized image.

Textured 3D face model is obtained by an projection of the original image I to the aligned model M. Textured model is then transformed to its initial position and re-projected back to image I' (Figure 5).

This normalization considers 3D shape of the human face. However, the static (unchangeable) model is the biggest weakness of this method. There are more advanced techniques that solve this problem by using *3D Morphable Face Model* [3].

Anatomical normalization

Anatomical normalization takes into account a shape of the face and a variability of temperature emission within different parts of the face. The output image of geometric normalization has usually a rectangular shape. The main purpose of the anatomical normalization is to mark an area where are the most important face data in terms of unique characteristic. This normalization is usually done by multiplying the original image by some mask.

- ***Binary ellipse mask:*** Human face has approximately an ellipse shape. Binary ellipse mask is therefore the simplest working solution of the anatomical normalization problem (Figure 6a).
- ***Smooth ellipse mask***: There is no step change in the mask on the edge between expected face points and background points (Figure 6b).
- ***Smooth weighted ellipse mask***: Practical experiments show that the human nose is the most unstable in terms of temperature emissivity. Therefore, the final mask has lower weight within expected nose position area (Figure 6c).

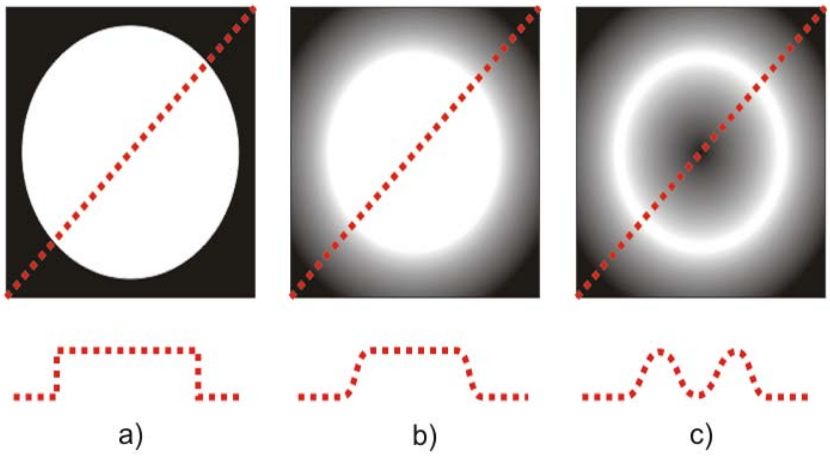

Figure 6: Masks of anatomical normalization and their intensity profiles.
a) Binary ellipse mask; b) Smooth ellipse mask; c) Smooth weighted ellipse mask.

Other possibilities of normalization are the noise removal, edge enhancement, etc.

2.3 Matching

The task of the matching algorithm is to evaluate the similarity of the normalized images. The difference between matching algorithm output score and some threshold determines if the input images match. The value of threshold is set in a training (learning) phase, as well as other parameters.

Tested matching algorithms are described in following sections.

2.3.1 Matching based on vector distance measurement

Most of the methods presented in this paper compute the matching score from distance between two vectors in n–dimensional space R_n. Used methods differs in:

- Dimensionality of space R_n.
- Transforming input image to the space R_n.
- Using different metric for distance computation.

All presented methods assume given input images in resolution w x h pixels.

Naive matching

This algorithm is almost never used in practice; however, it can be used for evaluation of the normalization methods performance. Dimensionality of the space n is equal to w x h. Image is transformed to the vector in this space using a simple concatenation of all pixels intensities row by row. The matching score is calculated by computing Euclidian distance between vectors representing input images.

PCA

PCA (*Principal Component Analysis*) is statistical method that transforms the number of potentially correlated variables into a (smaller) number of uncorrelated variables called principal components. The original space of correlated variables R_c is created in the same way as the method of *Naive matching*. Transformation between R_c and space of uncorrelated variables R_n is computed in the training phase. The dimensionality of R_n is given by r x w x h where r (reduction ratio) satisfy $0 < r <= 1$. The space R_c is reduced according to descending variability of the principal components. Vectors distance in R_n is computed using Mahanalobis distance. Details can be found in [4, 6].

LDA

LDA (*Linear Discriminant Analysis*) method is based on a similar principle as the PCA with the difference that this method takes into account the input data class membership (class = identity of individual). The new space R_n is formed to separate all classes as much as possible (minimize 'within-class' variance and maximize 'between-class' variance) [5, 6].

PCA and LDA

The combination of PCA and LDA methods sometimes gives better results than the methods alone. Original space R_c is transformed and reduced by PCA to temporary space R_{tmp}. This space is used as an input for the LDA method, which forms the final space R_n where the classes are well separable.

2.3.2 Phase Correlation

Phase correlation is the method commonly used in many matching problems. Proposed method distinguishes between template record *Tpl* and currently evaluated record *Rec*. It means that template processing differs from current image processing.

The phase image is obtained by DFT (*Discrete Fourier transform*) from normalized image. Then, in case of the template *Tpl*, only the *r* % of most significant frequencies is taken into account. In case of current record image *Rec* the all image frequencies are used for later comparison. The correlation of these two images is then computed by

$$C_{phase} = \frac{DFT_{\mathrm{Rec}} \cdot DFT^{*}_{Tpl}}{\left| DFT_{\mathrm{Rec}} \cdot DFT^{*}_{Tpl} \right|} \tag{1}$$

where DFT means discrete Fourier transform, DFT* means the complex conjugate of the DFT (DFT consists of two parts, real and imaginary ones). The multiplication is done element-wise.

After the correlation in the phase space of a template *Tpl* and a current record *Rec*, the inverse DFT is done to the C_{phase} image, to obtain C_{color} image in color space, which is finally normalized to get the scores in interval (0, 1). The goal now is to look for a peak (global maximum) in C_{color} image. Based on its height with respect to the mean value, the final score is obtained (2).

$$score = 1 - \frac{1}{Max - Mean} \tag{2}$$

The best matching score is then 0. C_{color} images with obvious peaks generate scores that are better separated from these with lower peaks or without peaks at all. In another words, the genuine images are better separated from impostors [7].

3 Testing and results

Tests were performed on two databases: *Database I* (300 images, 30 persons) and *Database II* (300 images, 28 persons). Both databases were divided into two parts – reference and testing part. Records in *Database II* have bigger variability in head orientation (see Figure 7).

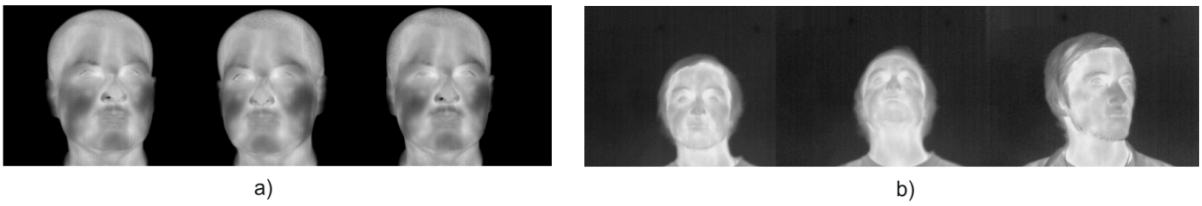

Figure 7: Samples of records in databases; a) *Database I*; b) *Database II*.

Firstly, the list of authorized users was created from reference parts of the databases. Then the matching method parameters were learned (if necessary) using the data from this records list. Finally, all records from testing part were sequentially compared with all of authorized user's reference records.

3.1 Normalization evaluation

All proposed methods of normalization were tested in all possible combinations. Since the total number of combinations was enormously high, only small and simple preliminary test was performed for each combination to get the best candidates. This simple test used only small subset of *Database I* (only 50 records of 12 persons). A simple metric was chosen to compare performance of each combination. It was minimal FRR (*False Rejection Rate*) while preserving zero FAR (*False Acceptance Rate*) [11].

All tested possibilities for each type of normalization are summarized in following list:

- Anatomical normalization: No, Binary ellipse (BE), Smooth ellipse (SE), Smooth weighted ellipse (SWE)
- Geometric normalization: Crop and Scale (CS), General affine transformation - eyes, mouth (GEM), General affine transformation - eyes, nose (GEN)
- Intensity normalization: No, Equalized (EQ)
- Noise normalization: No, Gaussian smooth 3x3 (G3), Gaussian smooth 5x5 (G5)
- Template size: 32x32, 48x48, 64x64 pixels
- Template margin: Small, Middle, Big; Small margin means that important points (eyes, mouth) are placed near the edge within the template, big margin means that they are more in the centre of the template (bigger part of face is preserved).

Several candidates were chosen for serious evaluation. Results are summarized in following table (Table 1). There are FRR of simple test and FRR and EER (Equal Error Rate) of serious test in results column. Both FRR values are minimal while preserving zero FAR [11].

Table 1: Performance of normalization methods evaluated by naive matching algorithm.

#	Normalization				Template	Results			
	Anatomical	Geometric	Intensity	Noise	Size	Margin	FRR (Simple)	FRR (Serious)	EER (Serious)
#n1	SWE	CS	EQ	G5	32x32	Big	**12.5%**	**61.6%**	14.6%
#n2	NO	CS	EQ	G5	32x32	Small	15.0%	74.7%	14.5%
#n3	SWE	CS	EQ	G3	32x32	Big	15.8%	**61.9%**	14.9%
#n4	SWE	CS	EQ	G5	32x32	Small	18.3%	78.1%	14.4%
#n5	NO	CS	EQ	G3	48x48	Small	20.0%	73.2%	14.3%
#n6	BE	CS	EQ	G5	32x32	Big	20.8%	68.1%	13.8%
#n7	NO	CS	EQ	G3	32x32	Middle	21.6%	68.4%	14.7%
#n8	SWE	CS	EQ	NO	64x64	Big	21.6%	63.6%	15.3%
#n9	BE	GEM	EQ	G5	32x32	Small	28.3%	73.3%	**12.0%**
#n10	NO	GEM	EQ	G3	64x64	Small	30.0%	70.7%	13.7%

3.2 Matching algorithm evaluation

Matching algorithms were not evaluated with all possible combinations of normalization, but again only with the best candidates coming from normalization test. Most of the algorithms can be parameterized by reduction rate r in interval (0, 1). Meaning of this number is described in chapter 2.3. Following algorithms were tested:

- Naive matching
- PCA (r)
- LDA (r)
- PDA (r) + LDA (1.0)
- Phase Correlation (r)

Tests were performed on the *Database I* and then on both of the databases together. The achieved results are summarized in Table 2. We can see that all advanced matching methods can improve the final performance only slightly.

Table 2: Results of tested matching algorithms.

#	Method	Normalization #	*Database I + II*		*Database I*	
			FRR	EER	FRR	EER
#ma1	PCA(0.35) + LDA(1.0)	#n3	**53.9%**	**10.8%**	67.3%	7.6%
#ma2	PCA(0.35) + LDA(1.0)	#n6	56.5%	11.8%	53.3%	6.3%
#ma3	PCA(0.45)	#n3	58.9%	14.7%	66.7%	7.8%
#ma4	PCA(0.35) + LDA(1.0)	#n2	61.1%	12.3%	**31.0%**	**5.5%**
#ma5	Naive matching	#n1	61.6%	14.6%	59.0%	9.1%
#ma6	Phase correlation	#n8	63.9%	16.3%	56.0%	10.3%

4 Conclusions

The standard face recognition procedure from the thermal images was described. Several methods of face normalization were introduced, which could help to reduce within-class variability. Well known algorithms like PCA, LDA or Phase correlation were chosen to solve the matching problem.

All proposed solutions have been implemented and tested on the database of 600 shots of 60 various persons. The performed tests show that the selection of appropriate methods of normalization leads to a results improvement. Right choice of appropriate matching algorithm may also improve the reliability of the face system. Unfortunately, the best achieved results are not good enough despite the large variety of tested methods. Equal error rate is around 10%, which is not acceptable in practical applications. Improving reliability is the subject of further research.

Acknowledgements

This work is partially supported by the grant “Information Technology in Biomedical Engineering”, GA102/09/H083 and the research plan “Security-Oriented Research in Information Technology”, MSM0021630528.

References

[1] Socolinsky, D. A., Selinger, A. Thermal Face Recognition Over Time, In: Proceedings of ICPR'04, IEEE CS, 2004, pp. 187-190, ISBN 0-7695-2128-2.

[2] Chen, X., Flynn, P. K., Bowyer, K. W. PCA-Based Face Recognition in Infrared Imagery: Baseline and Comparative Studies amfg, pp.127, IEEE International Workshop on Analysis and Modeling of Faces and Gestures, 2003.

[3] Blanz, V., Vetter, T.: A Morphable Model for the Synthesis of 3D Faces, Proc. of the SIGGRAPH'99, Los Angeles, USA, pp. 187-194, 08-13 August 1999.

[4] Smith, L. A tutorial on Principal Components Analysis. Cornell University, USA, 2002.

[5] Fukunaga K., Introduction to Statistical Pattern Recognition, second ed. Academic Press, 1990.

[6] Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 711-720, July 1997.

[7] Ahmed J., Jafri, M. N.: Improved Phase Correlation Matching, Lecture Notes in Computer Science, 2008, Volume 5099, Image and Signal Processing, Pages 128-135, 2008.

[8] FLIR system, ThermaCAM Reporter - user's manual, Professional Edition. Version 8.1, publ. No. 1558567, Language English, 2007.

[9] Buddharaju, P., Pavlidis, I.T., Tsiamyrtzis, P., Bazakos, M.: Physiology-Based Face Recognition in the Thermal Infrared Spectrum, Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.29, no.4, pp.613-626, April 2007.

[10] Weisstein, E. W. "Affine Transformation." From MathWorld--A Wolfram Web Resource. http://mathworld.wolfram.com/AffineTransformation.html.

[11] Jain, A. K., Flynn, P., Ross, A.: Handbook of Biometrics, Springer, 2007.