THERMAL FACE VERIFICATION BASED ON SCALE-INVARIANT FEATURE TRANSFORM AND VOCABULARY TREE

Application to Biometric Verification Systems

David Crespo, Carlos M. Travieso and Jesús B. Alonso
Signals and Communications Department, Institute for Technological Development and Innovation in Communications
University of Las Palmas de Gran Canaria, Campus Universitario de Tafira
s/n, 35017, Las Palmas de Gran Canaria, Spain

Keywords: Thermal Face Verification, Face Detection, SIFT Parameters, Vocabulary Tree, K-means, Image Processing, Pattern Recognition.

Abstract: This paper presents a comprehensible performance analysis of a thermal face verification system based on the Scale-Invariant Feature Transform algorithm (SIFT) with a vocabulary tree, providing a verification scheme that scales efficiently to a large number of features. The image database is formed from front-view thermal images, which contain facial temperature distributions of different individuals in 2-dimensional format, containing 1,476 thermal images equally split into two sets of modalities: face and head. The SIFT features are not only invariant to image scale and rotation but also essential for providing a robust matching across changes in illumination or addition of noise. Descriptors extracted from local regions are hierarchically set in a vocabulary tree using the k-means algorithm as clustering method. That provides a larger and more discriminatory vocabulary, which leads to a performance improvement. The verification quality is evaluated through a series of independent experiments with various results, showing the power of the system, which satisfactorily verifies the identity of the database subjects and overcoming limitations such as dependency on illumination conditions and facial expressions. A comparison between head and face verification is made, obtaining success rates of 97.60% with thermal head images in relation to 88.20% in thermal face verification.

1 INTRODUCTION

Human recognition through distinctive facial features supported by an image database is still an appropriate subject of study. We may not forget that this problem presents various difficulties. What will occur if the individual's haircut is changed? Is make-up a determining factor in the process of verification? Would it distort significantly facial features?

The use of thermal cameras originally conceived for military purpose has expanded to other fields of application such as control process in production lines, detection/monitoring of fire or applications of security and Anti-terrorism. Therefore, we consider its use in human identification tasks in scenarios where the lack of light restricts the operation of conventional cameras. Different looks of the main role from the film The Saint are shown in figure 1.

Figure 1: Facial changes of the character played by Val Kilmer in the film The Saint.

Val Kilmer modifies his look in this film spectacularly in order to not to be recognised by the enemy.

A correct matching between the test face and that stored in the image database is expected, although it may seem a hard problem to solve even if natural distortion effects such as illumination changes or interference are not considered. The recognition problem should be split in some stages, that is, acquisition of facial images for testing, features

Crespo D., Travieso C. and Alonso J..
THERMAL FACE VERIFICATION BASED ON SCALE-INVARIANT FEATURE TRANSFORM AND VOCABULARY TREE - Application to Biometric Verification Systems.
DOI: 10.5220/0003708104750481
In *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing* (MPBS-2012), pages 475-481
ISBN: 978-989-8425-89-8

extraction from specific facial regions and finally, verification of the individual's identity (Soon-Won et al., 2007).

In this context, the aim of this work is to propose and evaluate a facial verification system, applying the SIFT algorithm and obtaining local distinctive descriptors from each face. The construction of the vocabulary tree enables to have these descriptors hierarchically organised and ready to carry out a search to find a specific object.

This paper is organised as follows. The related work is described in section 2. The proposed system is presented in section 3. Description of experiments and results are outlined in section 4. Finally, discussions and conclusions are given in section 5.

2 RELATED WORK

Currently, computational face analysis is a very lively research field, in which we observe that new interesting possibilities are being studied. For example, we can quote an approach for improving system performance when working with low resolution images (LR) and decreasing computational load.

In (Huang, 2011), it is presented a facial recognition system, which works with LR images using nonlinear mappings to infer coherent features that favour higher recognition of the nearest neighbour (NN) classifiers for recognition of single LR face image. It is also interesting to cite the approach of (Imtiaz, 2011), in which a multi-resolution feature extraction algorithm for face recognition based on two-dimensional discrete wavelet transform (2D-DWT) is proposed. It exploits local spatial variations in a face image effectively obtaining outstanding results with 2 different databases.

The images of subjects are often taken in different poses or with different modalities, such as thermographic images, presenting different stages of difficulty in their identification.

In (Socolinsky, 2004) results on the use of thermal infrared and visible imagery for face recognition in operational scenarios are presented. These results show that thermal face recognition performance is stable over multiple sessions in outdoor scenarios, and that fusion of modalities increases performance.

In the same year 2004, L. Jiang proposed in (Jiang, 2004) an automated thermal imaging system that is able to discriminate frontal from non-frontal face views with the assumption that at any one time, there is only 1 person in the field of view of the camera and no other heat-emitting objects are present. In this approach, the distance from centroid (DFC) shows its suitability for comparing the degree of symmetry of the lower face outline.

The use of correlation filters in (Heo, 2005) has shown its adequacy for face recognition tasks using thermal infrared (IR) face images due to the invariance of this type of images to visible illumination variations. The results with Minimum Average Correlation Energy (MACE) filters and Optimum Trade-off Synthetic Discriminant Function (OTSDF) in low resolution images (20x20 pixels) prove their efficiency in Human Identification at a Distance (HID).

Scale Invariant Feature Transform (SIFT) algorithm (Lowe, 1999) are widely used in object recognition. In (Soyel, 2011) SIFT has appeared as a suitable method to enhance the recognition of facial expressions under varying poses over 2D images It has been demonstrated how affine transformation consistency between two faces can be used to discard SIFT mismatches.

Gender recognition is another lively research field working with SIFT algorithm. In (Jian-Gang, 2010), faces are represented in terms of dense-Scale Invariant Feature Transform (d-SIFT) and shape. Instead of extracting descriptors around interest points only, local feature descriptors are extracted at regular image grid points, allowing dense descriptions of face images.

However, systems generate large number of SIFT features from an image. This huge computational effort associated with feature matching limits its application to face recognition. An approach to this problem has been developed in (Majumdar, 2009), using a discriminating method. Computational complexity is reduced more than 4 times and accuracy is increased in 1.00% on average by checking irrelevant features.

Constructing methods that scale well with the size of a database and allow to find one element of a large number of objects in acceptable time is an avoidable challenge. This work is inspired by Nister and Stewenius (Nister, 2006), where object recognition through a k-means vocabulary tree is presented. Efficiency is proved by a live demonstration that recognised CD-covers from a database of 40000 images. The vocabulary tree showed good results when a large number of distinctive descriptors form a large vocabulary. Many different approaches to this solution have been developed in the last few years (Ober, 2007) and (Slobodan, 2008), showing its competency

organising several objects. Having regard to these good results, this solution will be tested in this paper with SIFT descriptors in a vocabulary tree.

3 PROPOSED THERMAL FACE RECOGNITION SYSTEM

The proposed approach use SIFT descriptors to extract information from thermal images in order to verify the identity of a test subject. Local distinctive descriptors are obtained from each face in the database and are used to build a vocabulary tree, through the use of the k-means function. For each test image, only its new descriptors are calculated and used to search through the hierarchical tree in order to build a vote matrix, in which the most similar image of the database can be easily identified. This approach mixes the singularity of the SIFT descriptors to perform reliable matching between different views of a face and the efficiency of the vocabulary tree for building a high discriminative vocabulary.

A description of the system is provided in the next subsections.

3.1 System Outline

The system is composed by four stages: Face segmentation, SIFT descriptors calculator, vocabulary tree construction and matching module.

While face segmentation is executed manually, the matching module searches in the vocabulary tree the best correspondence between the test descriptors and those of the database. Therefore, firstly the explanation is focused on the SIFT parameters and tree classification, and secondly a brief description of the matching module is given. A block diagram of the system is shown in figure 2.

3.2 Scale-invariant Feature Transform Parameters

The SIFT descriptors calculator uses most part of the results achieved by D. Lowe in (Lowe, 2004) as a guideline, only determinant parameters are modified in order to adapt the algorithm to the system. Keypoints are detected using a cascade filtering, searching for stable features across all possible scales. The scale space of an image, $L(x, y, \sigma)$ is produced from the convolution of a variable-scale Gaussian, $G(x, y, \sigma)$ with an input image, $I(x, y)$:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y), \quad (1)$$

where * is the convolution operation in x and y, and

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

Following (Lowe, 2004), scale-space in the Difference-of-Gaussian function (DoG) convolved with the image, $D(x, y, \sigma)$ can be computed as a difference of two nearby scales separated by a constant factor k:

$$D(x, y, \sigma) = \big(G(x, y, k\sigma) - G(x, y, \sigma)\big) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma). \quad (3)$$

From (Mikolajczyk, 2002), it is stated that the maxima and minima of the scale-normalised Laplacian of Gaussian (LGN), $\sigma^2\nabla^2 G$ produce the most stable image features in comparison with other functions, such as the gradient or Hessian. The relationship between D and $\sigma^2\nabla^2 G$ is:

$$G(x, y, k\sigma) - G(x, y, \sigma) \approx (k - 1)\sigma^2\nabla^2 G. \quad (4)$$

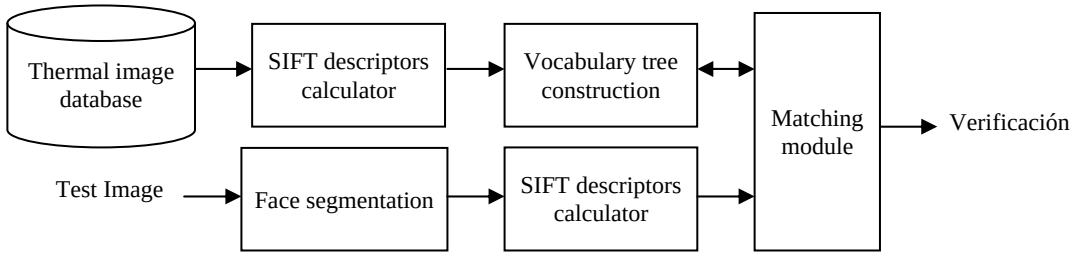

Figure 2: Diagram of the proposed thermal face recognition system.

The factor $(k-1)$ is a constant over all scales and does not influence strong location. A significant difference in scales has been chosen, $k=\sqrt{2}$, which has almost no impact on the stability and the initial value of $\sigma=1.6$ provides close to optimal repeatability according to (Lowe, 2004).

After having located accurate keypoints and removed strong edge responses of the DoG function, orientation is assigned. There are two important parameters for varying the complexity of the descriptor: the number of orientations and the number of the array of orientation histograms. Throughout this paper a 4x4 array of histograms with 8 orientations is used, resulting in characteristic vectors with 128 dimensions. The results in (Lowe, 2004) support the use of these parameters for object recognition purposes since larger descriptors have been found more sensitive to distortion.

3.3 Tree Classification

The verification scheme used in this paper is based on (Nistér, 2006). Once the SIFT descriptors are extracted from the image database, it's time for organizing them in a vocabulary tree. A hierarchically verification scheme allows to search selectively for a specific node in the vocabulary tree, decreasing search time and computational effort.

The k-means algorithm is used in the initial point cloud of descriptors for finding centroids through the minimum distance estimation so that a centroid represents a cluster of points.

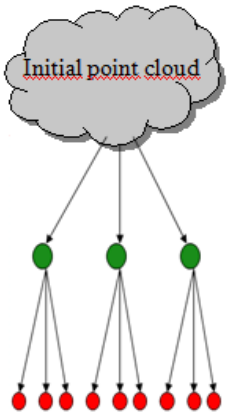

Figure 3: Two levels of a vocabulary tree with branch factor 3.

The k-means algorithm is applied iteratively, since the calculation of the centroid location can vary the associated points. The algorithm converges if centroids location does not vary. Each tree level represents a node division of the nearby superior stage.

The initial number of clusters is defined by 10, with 5 tree levels. These values have shown good results, working with the actual database.

A model of a vocabulary tree with 2 levels and 3 initial clusters is shown in figure 3.

4 EXPERIMENTS AND RESULTS

4.1 Database

Results are obtained using an image database built up by the Digital Signal Processing Division of the IDeTIC. This database contains 1476 thermal images of 704×756 pixels and 24 bit per pixel, from false thermal colour given for the sensor, in .png format equally split in two sets of 738 images with different modalities: face and head, corresponding to 18 images per subject in each set of 41 subjects. Therefore, the images are divided into categories depending on the type of information they provide:

- Heads: Images of full heads of subjects. (738 images).
- Faces: Images of facial details. (738 images).

The following figures present some examples of thermal images, heads (in figure 4) and faces (in figure 5) with the specified format.

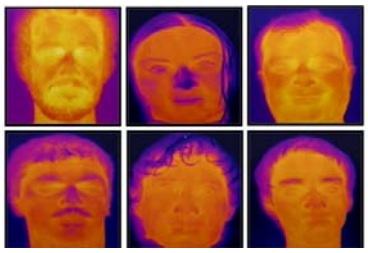

Figure 4: 6 thermal head images of the database. The examples show additional facial features such as head shape, hair and chin.

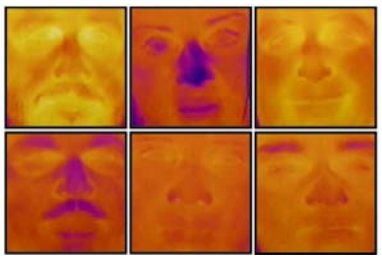

Figure 5: 6 thermal face images of the database from the same subjects in figure 4. The examples only show basic facial features such as eyes, lips and nose, representing the minimum information needed to verify a subject in the system.

The images were taken indoors by a SAT-S280 infrared camera in 3 different sessions with different facial expressions such as happiness, sadness or anger, various facial orientations and distinctive changes in the haircut or facial hair of subjects.

In order to assure the independence of results, both sets of images are equally divided into 2 subsets in a second stage: test and training. For each modality, 369 test images and 369 training images are available for the experiments.

The set of head images collects interesting details for recognition tasks, such as ear shape, haircut and chin. On the other hand, the set of facial images provides the minimum information that is nose, mouth and eyes areas.

Face detection and segmentation were manually realized in order to not to lose details in the process. One by one, from each image head or face was segmented and stored in separate files.

4.2 Experiments and Results

The aim of the experiments was to find how important is the extra information provided by head shape for human verification. Additionally, a comparison between head and facial verification results is carried out. The proposed methodology consisting on the matching of facial features of thermal images is compared with the matching of thermal images of heads.

For each subject, an equally random division of the image database is made so that 9 images per individual are used for testing and the remaining 9 for training purposes. As previously commented, 369 test images and 369 training images randomly chosen are available for the experiments in each modality. This division is carried out 41 times that is subject by subject in 41 iterations.

The process of face/head verification for a subject is the following. Firstly, the previously stated division of the database is made. Secondly, each of the 9 images of the test subject is compared with the 369 training images and results are obtained. Once these 9 images are processed, the database is joined together again and the process restarts with the next subject until the 41 subjects of the database are processed.

The parameters that take part in the experiments are the False Rejection Rate (FRR), False Acceptance Rate (FAR) and Equal Error Rate (EER), commonly used in biometric studies. Mean times are also considered in this study. These parameters are collected in form of vectors depending on a variable, the histogram threshold.

Since the verification process finishes, a histogram with the contributions of each image in the database is obtained. The database image that best fits the test image shows the biggest value in the histogram. In a second stage, histogram values are normalised with regard to the biggest value, from 1 to -1. A threshold is used during the experiments for considering only the contributions of images that are above this limit; those below are discarded in that moment. The histogram threshold descends from value 1 to -1 in order to consider different samples each time. In figure 6 and 7 FRR and FAR are shown in relation to the histogram threshold. X-axis represents the threshold variation and Y-axis shows FRR and FAR values.

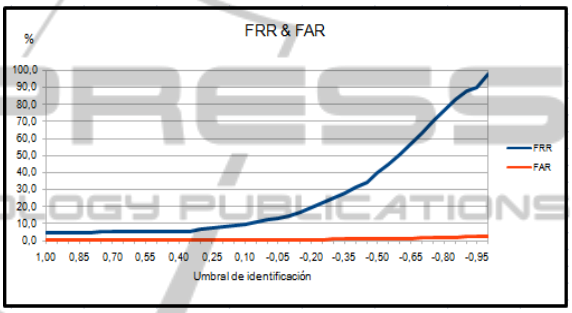

Figure 6: FRR (blue line) and FAR (red line) in terms of the histogram threshold in head verification.

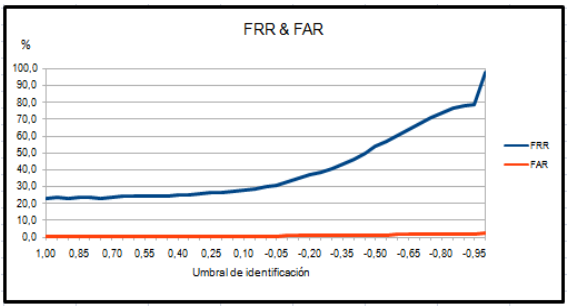

Figure 7: FRR (blue line) and FAR (red line) in terms of the histogram threshold in facial verification.

In practical terms, the threshold fall represents how the system becomes less demanding, taking more samples in account, increasing the FRR and FAR, since the additional samples do not belong to the test subject.

In table 1, a table with mean times can be observed. Although the verification time remains the same, the database updating (model building time) with head images is substantially higher as these images possess more information than facial images and therefore, consume more time and need more computational effort.

Table 1: Mean times of head and face images verification during the experiment.

MEAN TIMES HEAD AND FACE IMAGES		
Concept	Head Time (sec.)	Face Time (sec.)
Model Building	121.56	102.55
Test Verification	0.26	0.26

The best result obtained in the experiments with thermal head images is 97.60% in relation to 88.20% in thermal face verification. It can be observed that the success rate with head images is higher in comparison with facial images.

5 DISCUSSIONS AND CONCLUSIONS

The main contribution of this work is the use, for the first time, of a head/facial verification system based on SIFT descriptors with a vocabulary tree. This work is a preliminary step in the development of face verification systems using SIFT descriptors in thermal images of subjects.

The two variants compared in this work have different performance for verification. The cause is the amount of information provided by each format. On the one hand, head images preserve important discriminative characteristics about the original thermal images for identifying a subject that facial images do not include. On the other hand, it becomes clear that in case of head images more SIFT descriptors are produced and therefore, more essential data for the verification process is extracted. Additionally, faces of different subjects have often common features that provide no discriminant information.

As discriminative SIFT parameters are being widely used, specialised methods can be developed in future works for increasing dramatically the face verification rates using thermal imaging systems.

As future work we would like to increase considerably the size of database, and to include outdoor images. The proposed approach will be validated in this extended database.

ACKNOWLEDGEMENTS

This work was partially supported by “Cátedra Telefónica - ULPGC 2010/11”, and partially supported by research Project TEC2009-13141-C033-01/TCM from Ministry of Science and Innovation from Spanish Government.

Special thanks to Jaime Roberto Ticay Rivas for their valuable help.

REFERENCES

Heo, J., Savvides, M., Vijayakumar, B. V. K., 2005. Performance Evaluation of Face Recognition using Visual and Thermal Imagery with Advanced Correlation Filters. In *CVPR’05, Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.* pp.9-15. ISSN: 1063-6919/05.

Huang, H., He, H., 2011. Super-Resolution Method for Face Recognition Using Nonlinear Mappings on Coherent Features. In *Neural Networks, IEEE Transactions*. vol.22-1. pp.121–130. ISSN:1045-9227.

Imtiaz, H., Fattah, S.A., 2011. A wavelet-domain local feature selection scheme for face recognition. In *ICCSP’11, 2011 International Conference on Communications and Signal Processing.* pp.448. Kerala, India.

Jiang, L., Yeo, A., Nursalim, J., Wu, S., Jiang, X., Lu, Z., 2004. Frontal Infrared Human Face Detection by Distance From Centroide Method. In *Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing.* pp.41-44. Hong Kong.

Jian-Gang, W., Jun, L., Wei-Yun, Y., Sung, E., 2010. Boosting dense SIFT descriptors and shape contexts of face images for gender recognition. In *CVPRW’10, 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops.* pp.96-102.

Lowe, D. G., 1999. Object recognition from local scale-invariant features. In *ICIP’99, Proceedings of the Seventh IEEE International Conference on Computer Vision.* vol.2. pp.1150-1157.

Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. In *International Journal of Computer Vision.* vol.60-2 (2004). pp.91-110.

Majumdar, A., Ward, R.K., 2009. Discriminative SIFT features for face recognition. In *CCECE '09, 2009 Canadian Conference on Electrical and Computer Engineering.* pp.27-30.

Mikolajczyk, K., 2002. Detection of local features invariant to affine transformations, Ph.D. thesis. *Institut National Polytechnique de Grenoble*, France.

Nister, D., Stewenius, H., 2006. Scalable Recognition with a Vocabulary Tree. In *CVPR’06, 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.* vol.2. pp.2161- 2168.

Ober, S., Winter, M., Arth, C., Bischof, H., 2007. Dual-Layer Visual Vocabulary Tree Hypotheses for Object Recognition. In *ICIP’07, 2007 IEEE International Conference on Image Processing.* vol.6. pp.VI-345-VI-348.

Slobodan, I., 2008. Object labeling for recognition using vocabulary trees. In *ICPR'08, 19th International Conference on Pattern Recognition.* pp.1-4.

Socolinsky, D. A., Selinger, A., 2004. Thermal Face Recognition in an Operational Scenario. In *CVPR'04, Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.* vol.2. pp.1012-1019. ISSN:1063-6919/04.

Soon-Won, J., Youngsung, K., Teoh, A. B. J., Kar-Ann, T., 2007. Robust Identity Verification Based on Infrared Face Images. In *ICCIT'07, 2007 International Conference on Convergence Information Technology.* pp.2066-2071.

Soyel, H., Demirel, H., 2011. Improved SIFT matching for pose robust facial expression recognition. In *FG'11, 2011 IEEE International Conference on Automatic Face & Gesture Recognition and Workshops.* pp.585-590.