A MIXTURE OF GATED EXPERTS OPTIMIZED USING SIMULATED ANNEALING FOR 3D FACE RECOGNITION

Wael Ben Soltana [1], Di Huang[1], Mohsen Ardabilian[1], Liming Chen[1], and Chokri Ben Amar[2]

[1]Université de Lyon, CNRS, Ecole Centrale Lyon, LIRIS UMR5205, 69134, Ecully, France
[2]University of Sfax, Ecole Nationale d'Ingénieur de Sfax, REGIM, 3038, Sfax, Tunisia

ABSTRACT

A commonly accepted fact in the biometrics related domain is that fusing multiple classifiers for decision making generally leads to improved recognition performance. Meanwhile, the search for an optimal fusion strategy remains extraordinarily complex since the cardinality of the space of possible fusion schemes is exponentially proportional to the number of competing classifiers. In this paper, we propose a mixture of gated experts for 3D face recognition using an ensemble of 24 different classifiers. The mixture of gated experts is optimized using a Simulated Annealing-based algorithm. It automatically selects and fuses the most relevant similarity measurements. The experimental results of 3D face recognition achieved on the FRGC v2.0 dataset illustrate the effectiveness and stability of the proposed method. Additionally, as a learning-based method, it also has a good robustness to the variations of training database.

Index Terms— Fusion, mixture of gated experts, simulated annealing, similarity score selection, face recognition.

1. INTRODUCTION

A commonly accepted fact in the biometrics related domain is that fusing multiple classifiers for decision making generally leads to improved classification results [1]. The reason for combining multiple classifiers to solve a given classification problem is that the classifiers to be combined use various feature measurements or they are trained using different training databases, thereby providing different expertise and complementary information for decision making. In the literature, there are several types of strategies for multiple classifier combination, including parallel, cascading and hierarchical fusion [8]. The parallel fusion scheme, which consists of independently invoking individual classifiers and combining their results through a combination rule for final decision making, is preferred by most of researchers, especially in the biometrics field. We can still distinguish three parallel combination rules: (1) fusion at the matching score level [2], [3] where similarity scores achieved by different classifiers are combined by various approaches [4], e.g., sum rule, product rule, etc; (2) fusion at the rank level where sorted lists calculated by classifiers are merged based on different approaches such as Borda Count and Logistic Regression [5]; (3) fusion at the decision level [6] where all the candidates of the classifiers are fused by adopting some techniques [7], e.g., Majority Vote.

In the literature of 2D, 3D and multimodal face recognition, score level fusion, making use of simple sum rule, was extensively investigated [17]. This popularity can be greatly explained by the fact, as evidenced by Kittler et al. [5], that the sum rule is much less sensitive to the error of individual classifier when estimating posterior class probability, thereby outperforms other rules (e.g. product rule) in most experimental comparisons. In this paper, we investigate a mixture of gated experts (classifiers) which is a generalization of the simple sum rule for the purpose of 3D face recognition. Given a set of N similarity measurements of different classifiers on 3D faces, the basic idea is to add up the similarity scores delivered by a subset of the best classifiers while optimizing the whole recognition accuracy. The trouble is that the space of possible combination strategies using simple mixture of gated experts is exponential. In this work, we consider N = 24 different measurements, thus leading to 2^{24} = 16 777 216 possible fusion strategies. As an exhaustive search in such a space is intractable, one needs to resort to a kind of heuristic optimization technique. Thus, a Simulated Annealing-based method is employed to find a (sub) optimal mixture of gated experts to fuse the set of different similarity measurements for improved recognition results. We further study the stability of the searched optimal mixture of gated experts as learning data is increased. The proposed strategy was evaluated on the FRGC v2.0 database for 3D face recognition and the experimental results clearly demonstrate the effectiveness of our approach.

Section 2 presents fusion schemes by mixtures of gated experts. Section 3 describes the simulated annealing for stochastic search of a (sub) optimal mixture for 3D face recognition. Section 4 discusses the experimental results. Section 5 concludes the paper.

2. MIXTURES OF GATED EXPERTS

2.1. Similarity measurements

Given two 3D face scans A and B, a similarity measurement $s_f(A, B)$ is a positive value between 0 and 1 computed based on a given feature f, giving evidence on how the scans A and B represent the same subject. In this work, we made use of 24 different similarity measurements computed from a set of multi-scale extended local binary patterns (*MS-eLBP*) depth maps which provide a comprehensive representation of local shape variations of 3D facial surfaces [15]. As illustrated in Fig. 1, the *MS-eLBP* depth maps encode the information of a 3D facial range image at several scales, and contain not only the relative differences of gray values between a central pixel and its neighbors, as the LBP based depth map [16] does, but also their exact differences through some additional layers.

Fig. 1. A set of *eLBP* based depth maps of a facial range image with 8 neighboring points (P = 8), different radius value R from 1 to 8 (from left to right), and 3 layers (L = 3)

Specifically, we denote an *eLBP* depth map by $eDM(P, R, L)$; where P is the number of neighboring points around a central pixel, R is radius of the circular neighborhood thus describing the scale, and L is the L^{th} layer of the *eLBP* depth map. It should be noted that the first one (L=1) is the original LBP based depth map [16]. Given two 3D face scans and parameters (P, R, L), we can hence generate two *eLBP* depth maps, eDM_A and eDM_B, on which SIFT based local features are extracted and then matched. The similarity between eDM_A and eDM_B is the number of their matched keypoints. By varying the parameters P, R, and L, we can thus obtain different similarity measures s_i for each pair of 3D faces. Refer to [15] for more technique details. In this paper, P is set to 8; R varies from 1 to 8; i = {1, 2, 3}. We thus generate 24 similarity measurements s_k, k in 1, 2, …, 24.

2.2. The space of combination schemes based on the mixture of gated experts

In its most general form, the mixture of gated experts is defined by

$$p(t/x_1 \cdots x_K) = \sum_{k=1}^{K} \pi_k(x_k) p_k(t/x_k) \qquad (1)$$

where $p_k(t|x_k)$ is the posterior probability of the k^{th} classifier, or called *experts* given the feature x_k; while the mixture coefficients $\pi_k(x_k)$ are known as *gating functions* [19].

In 3D face recognition, the so-called experts are the different similarity measures. If we take all the gating functions to the constant 1, we can obtain the well-known sum rule for fusing classifiers. This work considers a generalized form of the simple sum rule by permitting some of these gating functions taking the value of 0. As a result, given a set of similarity measurement s_k, we propose a mixture of simplified gated experts defined as:

$$s(A,B) = \sum_{k=1}^{K} \pi_k s_k(x_k^A, x_k^B) \qquad (2)$$

where π_k is a 0 or 1 binary indicator function while s_k is the k^{th} similarity measurement calculated on feature x_k extracted from the two subjects A and B, respectively.

Now the problem appears that the search for the optimal scheme combining K similarity measures according to (2) is not easy as the space of possible combination schemes is 2^K!. A heuristic search strategy needs to be implemented.

3. OPTIMIZING THE MIXTURE OF GATED EXPERTS BY SIMULATED ANNEALING

Simulated annealing (SA) is a stochastic search methodology which was inspired by annealing in physics. SA was first proposed by Metropolis et al. [12], and later popularized by Kirkpatrick et al. [13], and has shown its effectiveness in a wide range of problems [10, 11] such as timetabling, traveling salesman [9], and communications systems [10].

3.1. SA algorithm

The basic SA algorithm maintains both a status represented by the current solution s and a temperature T. In each iteration, a new solution s' is produced in the neighborhood $N(s)$ of current solution s. The quality of both solutions (s and s') is evaluated by using an energy function f. Better solutions with respect to the objective (Δ>=0, $\Delta = f(s') - f(s)$) are always accepted. Moves to lower solutions ($\Delta < 0$) are allowed but their frequency is governed by a probability function $exp(-\Delta/T) < R$, which depends on the temperature T and the magnitude of the increase Δ with R, a random value drawn from the interval of [0, 1]. For a given temperature T, a number *n_rep* of iterations is performed. The temperature T is gradually decreased after each iteration depending on $T = \alpha(T)$ named *cooling schedule*. This decreases the acceptance rate of lower solutions progressively. The algorithm stops if the temperature reaches a pre-defined low threshold or the system becomes stable.

3.2. Optimizing the mixture of gated experts for 3D face recognition

Using binary valued gating function, a combination scheme s of N different similarity measurements can be simply encoded by a N-vector of binary values: "1" at the position i indicating that the i^{th} similarity is active or selected in the combination scheme s while "0" indicating its absence. For instance, a fusion scheme *s*, [1, 0, 1, 1], with N=4 indicates the 1^{st}, 3^{rd} and 4^{th} similarity are selected in simple sum rule.

The satisfactory degree of a scheme *s* is calculated based on an energy function *f*. In the case of 3D face recognition, different objective functions can be utilized depending upon the given application scenario, e.g. Equal Error Rate (EER), False Acceptance Rate (FAR), False Rejection Rate (FRR), etc. In this work, identification rate is adopted as the objective function. Therefore, we calculate the identification rate of each combination scheme *s*. The move from one solution *s* to another neighboring one *s'* (*N(s)*) is based on multiple random mutations. For example, given a fusion scheme *s* = [1, 0, 1, 1], we can obtain *s'* = [1, 0, 0, 0] with two mutations in the two last bits. If the system becomes stable or the temperature reaches a pre-defined low threshold, the solution *s* is delivered as the (sub) optimal fusion scheme (FS).

4. EXPERIMENTAL RESULTS AND EVALUATION

4.1. Database and Experimental Settings

The FRGC v2.0 [14] database was used for experiments. It contains 4007 3D face models of 466 different subjects with significant expression variations. We eliminated 56 subjects which possess only one 3D face scan from the dataset. Thus, 3951 scans of 410 subjects were finally utilized in our identification experiments.

Table 1. Experimental Settings

		Subject Number	Image Number
Database 1 177 Subjects 1778 images	Subset 1	95	885
	Subset 2	15	174
	Subset 3	16	178
	Subset 4	16	187
	Subset 5	17	176
	Subset 6	18	178
Database 2 233 Subjects 1763 images	Subset 1	104	849
	Subset 2	24	179
	Subset 3	24	183
	Subset 4	15	183
	Subset 5	31	181
	Subset 6	35	188

One 3D face scan with neutral expression was selected from each subject to make up a gallery of 410. The remaining 3D face scans (3541) were treated as probe and further divided randomly into two subsets i.e. Database 1 and Database 2. As illustrated in table 1, each dataset is composed of 5 subsets: Subset1 (50%), Subset2 (10%), Subset3 (10%), Subset4 (10%), Subset5 (10%), and Subset6 (10%).

The proposed approach was evaluated based on experiments designed for face recognition. Another objective is to assess the stability of the optimized mixture of gated experts when the training dataset is increased with new subjects. For this purpose, we generated the first fusion scheme based on Subset1. Next, we added separately other subsets to Subset1. In each step, we generated a new fusion scheme based on the previous simulated annealing step. Two experiments were carried out to cross-validate the proposed method. In the first experiment, Database 1 was used as learning set and the whole Database 2 was exploited as test set. In second experiment, the two databases were inverted.

4.2. Results and Analysis

The parameters employed in the proposed simulated annealing algorithm are shown in Table 2. They were empirically adjusted to have an optimal solution, and for other applications, their values probably change and need to be fixed specifically. The initial temperature was high (e.g. 500) in order to accept most or all solutions. The final temperature was set close to zero in order to make the search process stops when the possibility of accepting a better fusion strategy becomes negligible.

Table 2. Parameters used in the proposed SA algorithm

Parameter	Value
Initial Temperature	500
Final Temperature	1e-8
Cooling schedule (α(T))	0.8*T
Number of iterations (n_rep)	1600

Table 3. First Experiment

Database 1	Training RR	Test RR Full Dataset 2
(1) S1	97.51	95.97
(2) S1+S2	97.73	97.11
(3) S1+S2+S3	97.90	97.11
(4) S1+S2+S3+S4	97.92	97.11
(5) S1+S2+S3+S4+S5	97.94	96.99
(6) S1+S2+S3+S4+S5+S6	97.94	96.99

Table 4. Second Experiment

Database 2	Training RR	Test RR Full Dataset 1
(1) S1	97.33	97.41
(2) S1+S2	97.36	97.41
(3) S1+S2+S3	97.37	97.41
(4) S1+S2+S3+S4	97.42	97.41
(5) S1+S2+S3+S4+S5	97.45	97.68
(6) S1+S2+S3+S4+S5+S6	97.45	97.68

The first evaluation results are given in Table 3. We used all the subsets of Database1 for training and the whole Database 2 for test. Experiments (1) to (6) aim to assess the stability and convergence of the proposed searching algorithm for the best mixture of gated experts. From (2), (3) and (4), we obtained the same fusion strategy FS1 = [1, 0, 1, 0, 1, 1, 1, 1, 0, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1] at the training stage. The recognition rate at the test stage is 97.11%. From (5) and (6), we obtained the fusion strategy FS2, with some few different bits from FS1, FS2 = [0, 0, 1, 0, 0, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 0, 0]. This variation in combination scheme impacts the system behavior at the test stage, and the result became slightly lower. This can be explained by the fact that the generated combination scheme from experiment 5 and 6 is not the best one for Database2.

The similar experiments were carried out with all subsets of Database 2 for training and the whole Database1 for testing. Table 4 shows all the rank-one recognition rates. As we can see, now using more data for training gives better fusion scheme which displays a 97.68% recognition rate. All experimental results also indicate that the overall behavior of the searched fusion scheme is quite stable since addition of new persons does not affect the accuracy of the entire system.

Table 5. Identification Rate

	Test Recognition Rate
Best Expert	89.61
Mian et al. [18]	96.64
Gökberk et al. [17]	96.94
Proposed Approach	97.33

In Table 5, we can see that all fusion methods surpass the accuracy of the best one (89.61%) of the utilized 24 experts. The results of the proposed SA algorithm are also compared with another two methods in the literature. The first one is Sequential Floating Backward Search-based classifier selection proposed by Gökberk et al. [17] that applied the SFBS to find the near-optimal subset. The second one is an online method developed by Mian et al., [18] based on a weighted sum rule. We firstly computed similarity measurements from Database 1 and Database 2 separately; then calculated their average rate for experiment (6) of both the first and second test. It can be seen that the proposed SA algorithm provides the best result.

5. CONCLUSION

In this paper, we presented a mixture of gated experts, which is a generalized simple sum rule, to fuse different classifiers for 3D face recognition. To search the optimal fusing strategy in an exponential space of possible fusion strategies, we further proposed a Simulated Annealing (SA)-based stochastic search approach. The experimental results on the FRGC v2.0 database showed the effectiveness of the proposed approach both in terms of performance and stability.

In our future work, we will study the most general form of mixtures of gated experts for fusing.

ACKNOWLEDGEMENT

This work was supported in part by French FAR3D project supported by ANR under the grant ANR-07-SESU-004-03.

6. REFERENCES

[1] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: a review," *IEEE Trans. PAMI*, 22(1), 4-37, 2000.

[2] J. Cook, M. Cox, V. Chandran, and S. Sridharan, "Robust 3D face recognition from expression categorization," *ICB*, 2007.

[3] Y. Kim, K.-A. Toh, and A. B. J. Teoh, "An online learning algorithm for biometric scores fusion," *BTAS*, 2010.

[4] T. Faltemier, K. Bowyer, and P. Flynn, "3D face recognition with region committee voting," *3DPVT*, 2006.

[5] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. PAMI,* vol. 20, no. 3, 1998.

[6] A. Godil, S. Ressler and P. Grother, "Face recognition using 3D facial shape and color map information: comparison and combination," *Biometric Technology for Human Identification, SPIE*, 5404, 351–361, 2005.

[7] Md. M. Monwar and M. Gavrilova, "FES: a system for combining face, ear, and signature biometrics using rank level fusion," *International Conference on Information Technology: New Generations*, 922-927, 2008.

[8] E. Marasco, A. Ross, and C. Sansone, "Predicting identification errors in a multibiometric system based on ranks and scores," *BTAS*, 2010.

[9] V. J. Rayward-Smith, I. H. Osman, C. R. Reeves, and G. D. Smith, "*Modern Heuristic Search Methods*", Wiley, 1996.

[10] A. Sadegheih, "New methods for system planning," *Applied Soft Computing*, Vol. 11, no. 1, pp. 1098-1107, 2011.

[11] X. Zhao, "Simulated annealing algorithm with adaptive neighborhood," *Applied Soft Computing*, vol. 11, no. 2, pp. 1827-1836, 2011.

[12] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, " Equation of state calculations by fast computing machines," *J. Chem. Phys.*, vol. 21, no. 6, pp. 1087-1092, 1953.

[13] S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi, "Optimization by simulated annealing," *Science*, 1983.

[14] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Over-view of the face recognition grand challenge," *CVPR*, 2005.

[15] D. Huang, M. Ardabilian, Y. Wang, and L. Chen, "A novel geometric facial representation based on multiscale extended local binary patterns," *FG*, 2011.

[16] D. Huang, G. Zhang, M. Ardabilian, Y. Wang, and L. Chen, "3D face recognition using distinctiveness enhanced facial representations and local feature hybrid matching," *BTAS*, 2010.

[17] B. Gökberk , H. Dutagacı, L. Akarun, B. Sankur, "Representation plurality and fusion for 3D face recognition," *IEEE Trans. on SMC-Part B*, vol. 38, no. 1, pp. 155–173, 2008.

[18] A. S. Mian, M. Bennamoun, and R. Owens, "An efficient multimodal 2d-3d hybrid approach to automatic face recognition," *IEEE Trans. PAMI*, vol. 29, no. 11, pp. 1927–1943, 2007.

[19] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, "Adaptive mixtures of local experts," *Neural Computation*, vol. 3, no. 1, pp. 79-87, 1991.