Context-Patch for Difficult Face Recognition

Anonymous ICB 2012 submission

Abstract

Multiple research has shown the advantage of patch-based or local representations for face recognition. This paper builds on a novel way of putting the patches in contex, using a foveated representation, and shows this improves performance in difficult conditions. While humans focus on local regions and move between them, they always see these regions in "context". We hypothesize that using foveated context can improve performance of local region or patch based recognition techniques.

The face images captured in uncontrolled environment suffer greatly due to blur, scale, resolution and illumination. In such situations, a facial patch by itself does not provide highly discriminative information for the patch fusion algorithm to perform well. Correct patches may have higher intra-subject variation and incorrect patches may have lower inter-subject distance. To overcome this issue, we define a context-patch which is a face region that contains more information about a particular region and some contextual information about the rest of the face region. We divide the face images into multiple regions, and construct a context-patch for each region. The low-resolution context is tolerant of intra-subject variations but still responds to many inter-subject differences.

We build multi-class SVMs per patch, with a face model combining many independent context-patch classifiers. The per-patch margins are normalized using Z-score normalization and simple sum method is applied for fusion. We show that by using the context-patch decision level fusion, the identification as well as verification performance of face recognition system can be greatly improved, especially in the case of highly degraded images. We conducted the experiments on the Remote Face Database and show the improvement over state of the art algorithms and the standalone patch fusion algorithm. While patch-based techniques are often touted for handling occlusion, our testing is on all the availble face data, with only natural occlusions such as tree branches. We show that context patches, even with our simple features and fusion, provide state of the results on these difficult problems.

1. Introduction

Face Recognition has received a lot of attention in the biometrics systems because of the ease of acquisition (even at a distance), availability, identifiability, human usability/verifiability, and its accuracy on cooperative faces. However, the face images captured in an uncontrolled environment suffer greatly due to blur, illumination, pose , occlusion, scale, resolution etc. A lot of work has been done in the literature to address such issues. One of the challenges in such an uncontrolled face recognition if how to get the discriminative information from such degraded faces for recognition. Discriminative analysis of facial regions has been used to boost the performance of face recognition in the past. The few ways of making use of the facial patches for face recognition are using the feature level fusion, score level fusion. The real question considered in this paper is, how to improve facial patches individuality for the fusion of scores or decision level fusion, especially for difficult problems. While we do not want to mimic humans, it can be useful to look at the human visual system for inspiration and ideas, as it does very well at face recognition. The human visual system is a foveated system, with areas of very high resolution and a broader area around it with less resolution. It also views small regions/features then moves (formally saccades) to another region. While there has been work in modeling the saccades and visual search for faces [16, 20], these models suggest that multi-resolution representations and context play an important role. So a useful question is how we can incorporate these ideas into automated face-recognition. We choose to do this using what we call contextual patches.

In this paper, we demonstrate the power of a context-patch approach for face recognition in medium blurred and severely blurred images. A context patch is a facial region which contains more information from a specific facial regions and some contextual information from rest of the facial regions. We present one way of defining a facial contextual patch, but expect readers can develop many useful variations on that idea. We use multiple context patches from the face images and construct the multi-class SVMs. The decision scores from the individual context-patch classification are normalized and fused together for the final classification.

While many others have analyzed patch-based techniques with respect to things like occlusion, our experimental evaluation is focused on a real problem of face in difficult settings. In particular we show that the context-patches improve the face recognition performance on non-cooperative recognition on blurry and degraded images, and outperform the best reported results on that data.

The rest of the paper is formatted as follows: Chapter 2 gives the summary of related works in the area of discriminative analysis and fusion for face recognition. Chapter 3 defines the proposed context-patch and the motivation behind it. Chapter 4 describes the construction of individual classifiers, score generation, normalization and fusion. We show our experimental analysis and results in chapter 5 and finally conclude in chapter 6.

2. Related Works

Extracting the meaningful and discriminative information from the face images is still an interesting research topic in Face Recognition, with litterly 100s of related papers; we discuss only a few most relvent papers. The appearance based features such as global features (e.g. suspaces with PCA, LDA,) or local features such as Gabor features [14], Local Binary Patterns [1] have been popular in face recognition research. After the extraction of such features, learning methods are often applied to analyze the discriminative features with multiple groups declaring the most discriminative part in the face, are discriminant feature are around eyes and central facial regions. Rather than fixing certain regions or deciding a priori what is descrimative, we will use multiple local patches and let the system fuse them. We focus our related work on patch-based and decision fusion for face recognition.

There are many patch based techniques, e.g [18, 3, 19, **?**, **?**] Local patch-based methods seek descrimative patches, e.g. [**?**, 18] used gabor features from the local patches for face recognition. In order to chose the most discriminant features from the facial regions, they extracted the patches from different regions of the face and use greedy algorithm to progressively incorporate the patches in the larger subset of the patches. The weights are learnt offline with the analysis of interclass and intraclass discriminability. [3] Extract the candidate patches at multiple spatial resolution. So a patch contains both high resolution as well as a low resolution (contextual) information. [19] used nearest neighbor classifier for the individual facial patches classification. The results from each such classifier are fused for the recognition. Basically, the two ways of making use of the facial patches for face recognition are using the feature level fusion and decision level fusion. In feature level fusion, the descriptive features of the patches are combined in a way it represents the global face[17]. In the decision level fusion, the classifiers that classify the corresponding patches in the query and target images are combined or fused together to boost the performance of face recognition system[5],[11],[2],[9], [21], [19].

3. Context Patches

3.1. Motivation

Face recognition can use global or local features, and this paper is focused on using local features or patches and fusing the results. Local features are more robust to misalignment, occlusion and other image degredations, and to our knowledge, all of the modern state-of-the art systems use local features. The question then is what features and how to fuse them. Our motivation for this work not to define the "best" feature, but rather to address our hypothesis that local feature "context", such as in a foveated system, can improve performance. That concept could be used ontop of almost any feature set, offering the potential for improvements in other systems. While looking for a feature descriptor to extract information at multiple scale, we came across a LBP-like feature descriptor which is called GRAB [17]. Because of the claimed superiority of LBP-Like features(GRAB features) [17] over the standard LBP features on low scale images, we decided to use the GRAB features in our experiments. One of the advantages of these GRAB features compared to the standard LBP features is that it provides the multi-scale description of the images. We use these relatvively simple LBP-like features in rectangular patches, and show how context and fusion can improve these simple features to provide performance to state of the art levels.

Information fusion can improve the performance of biometrics system, [15], if it can combine useful independent information. Classification error diversity and individual classifier accuracy are the keys to performance in case of information fusion methods [4] [10]. The fusion of multiple classifier is generally superior to the single classifier when the predictions of its component weak classifiers have enough diversity. At the same time when the weak classifiers have reasonable accuracy. Poor weak classifiers which do not have better performance than random guess can not improve the performance of a fused classification system. The two dominant ways of creating diversity in face recognition are: 1. Using the different features from the face images and construct different learning algorithms to generate the decision score. 2. Manipulating the training data while using the same features and learning algorithm. We focus on the just later as it makes the hypothesis of context easier to directly test.

A patch based face recognition using one kind of features which uses the scores from different facial regions for the final classification employs the diversity obtained by manip-

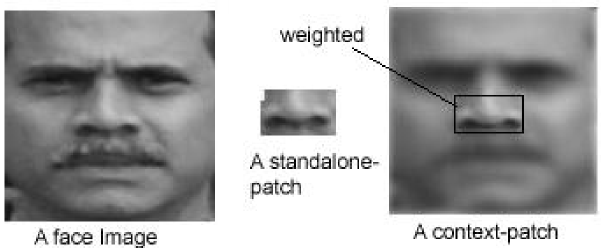

Figure 1. From Left to Right: A face image, an example of a standalone patch of nose area, an example of a context-patch with the nose area is sharp with more weight and rest of the area is blurred with less weight.

ulating the training data. Since each facial patch contains a different information, the weak classifiers built by learning this information provide the error diversity for fusion. But, it is important to know whether this error diversity comes with classification accuracy as well.

3.2. Foveated and Weighted Context Patch

We explored ways of constructing the context-patches which provide classification error diversity as well as accuracy. Inspired by the Human Visual System which gives weight to different human face regions via the information obtained by fovea and parafovea at multiple fixations [7] [12], we define a foveated context patch. A foveated context patch is a foveated face image, where the patch under consideration is foveal and rest of the face image is parafoveal. We simulate this using two ways. One way of simulating this idea is using a high scale information for patch under consideration and a low scale information for the rest of the image area.

Another way of constructing the context patch is by using the higher weight on the patch under consideration while giving less weight to the rest of the images and construct the feature vector. We tried both approaches, and decided a combined approach, with both spatial resolution changes and added spatial weighting was the best.

Figure 1 shows an example of a foveated and a weighted context patch. In this paper, we refer context-patch as a region with the high scale as well as more weight on that region with scale and less weight contextual information from the rest of the face image. While we could use “smooth” region boundaries and guassian weighting, this was simpler to implement.

4. Classification and Fusion

We build a separate classifer for each patch and fuse the results. In particular a Support Vector Machine(SVM) [8] is used for per-patch classification followed by z-score normalization of the margins and sum fusion of the resulting scores.

4.1. Patch Classification

Since Face Recognition is a M-class problem, we use One-Against-All multi class SVM with the i_{th} SVM separating class i from the rest of the classes. It constructs k SVM models where k is the number of classes. We divide a face image into N patches, construct N sets of One-Against-All SVMs and obtain N sets of scores from the decision functions. Suppose, s_{ij} where, $i = 1, \ldots k$, $j = 1, \ldots N$ is the decision value of j_{th} patch for i_{th} class. Since these N sets of SVMs are constructed using different training examples, the distribution of the scores is different. Before, fusing the scores for final classification, a normalization step is necessary. We use Z-score normalization for that purpose.

Figure **??** show the accuracy of 64 classifiers built by using 64 face patches on a the degraded subsets of the Remote Face Database. As shown in the figure, though the diversity error is likely to be achieved from the classifiers from standalone patches, the accuracy from each patch is very low. To solve this problem in difficult face recognition, we propose an idea of a context patch. A context patch is a facial region which contains more information about a particular region with some information about the rest of the facial regions. This way, we can create the diversity in the classification error at the same time, do not have to compromise with the individual classifier accuracy.

4.2. Normalization and Fusion

We use Z-score normalization technique to normalize the decision scores (margins) obtained from the SVM patch classification. Suppose $S_j = (s_{1j}, \ldots s_{kj})$ is the score vector obtained for patch j where k is the number of classes. Suppose S_{jn} is the normalized scores after Z-score normalization. The Z-score normalized score is given by:

$$S_{jn} = \frac{S_j - mean(S_j)}{std(S_j)} \tag{1}$$

The fused scores for final decision are obtained by summing up the normalized scores. The sum is simple yet robust way of fusing the scores. [10] $S_{fusion} = ((s_{11} + s_{12} + \ldots + s_{1N}), (s_{21} + s_{22} + \ldots + s_{2N}), \ldots, (s_{k1} + s_{k2} + \ldots + s_{kN}))$ A test face image is given a label i^{**}, which has the largest values in S_{fusion}.

5. Experimentas and Results

5.1. Datasets

The database for the experimental analysis for the proposed approach is taken from the work of Jie et al. [13] and is called Remote Face Database which we obtain from those authors. The database was acquired in an unconstrained outdoor environment at a distance from 5m

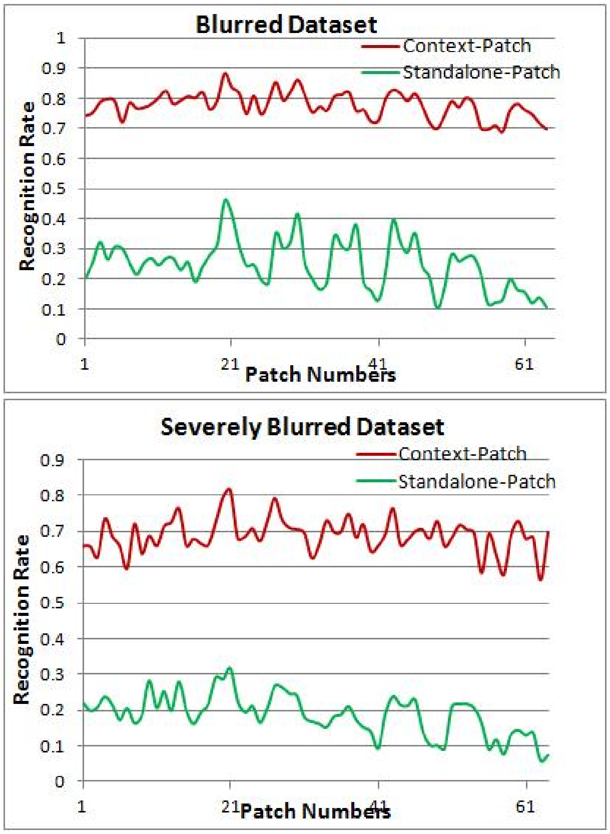

Figure 2. Patch Scores Distribution on subset 1(blured) on the top and subset 2 (Severly degraded) on the bottom. The graphs shows the recognition rate while using stand-alone patches and context-patches for the recognition. It is clear from the graph that the diversity of the recognition rate is similar in both stand-alone and context patch cases. While the performance of the standalone-patch by itself is much lower than the performance of the context-patch.

to 250m. The face images in this database suffer from variations due to blur, poor illumination, pose and occlusion. The database contains 688 clear images, 85 partially occluded images, 37 severely occluded images, 540 images with medium blur/degradation, 245 images with severe blur/degradation and 244 with poor illumination conditions. In this paper, we use 540 images with medium blur as our first probe set and 245 images with severe blur as the second probe set. The gallery images are from the set of clear images. All the images are cropped and resized to a fixed size of 120x120 pixels.

Fig. 3 shows the variations in some of the sample images from the database. The database contains 17 different subjects, but signficant real-world difficult probe settings over hundreds of conditions. The scores of the baseline comparison algorithm have been presented in prior papers allowing direct comparison on hard data with many images per person.

This is a narrow experiment to test the "context patch" hypothesis. While the data may seem small, by focusing only on a smaller number of subjects, under hundred of conditions, the experiment is more focused on the impact of context-patch under lots of conditions to show the effect. Our goal is not claiming this is the best recognition system, which would take testing on a much larger gallery. Before such larger testing, more experiments are needed to select the the "best" features and patch geometry. But these experiments do strongly suggests the hypothesis is confirmed.

5.2. Experimental setup

The first set of experiments consists of the procedure as mentioned [13]. A face image is divided into different 64 facial regions, and GRAB features at multiple scale are extracted and concatenated to form a global and multi-scale facial description. The feature vectors thus obtained are classified using the One-Against-All SVMs. This approach is applied in the medium blurred and severely blurred partitions of the Remote Face Database. For each database, the gallery images consist of the images from the clear subset of the database. Figure 4 shows the results. Recognition results are obtained by randomly sampling 3 images and 5 images respectively from gallery. We repeated the experiments several times and took the average to arrive at the final recognition results. The experimental protocol is similar to the one mentioned in [13]. The baseline algorithm presented in [13] consists of an algorithm involving KPCA, LDA and SVM. The comparison of this baseline algorithm, GRAB feature based algorithm and V1-Like algorithm is given in figure 4

The comparison used with GRAB features from standalone-patches followed by SVM classification and fusion of the scores. In this technique, we divide a face into multiple standalone-patches, extract the multi-scale GRAB features from each patch and use multi-class SVMs for patch classification. The decision scores thus obtained are fused together for the final classification as mentioned in section 4.2 . For the gallery set, use the same process for the context-patch process. this way because we directly compare the results of the feature level fusion, standalone-patch fusion and context-patch fusion methods on the same gallery images. We used one particular set of gallery images which consisted of 5 images. We did this experiment on medium blurred and severely blurred partitions of Remote Face Database.

The recognition rate is given in figure 5 and roc curve is given in figure 6 and 7.

The thrid algorithm was using GRAB features from context-patches followed by SVM classification and fusion of the scores. This algorithm is used similar to the experimental setup described in other experimental setup except instead of standalone-patches, we use the GRAB features from the context-patches and use multi-class SVMs for classification. The scores thus obtained are normalized and fused for final decision. The gallery images used for this experiments are same as described earlier. The recognition rate is compared to the GRAB feature based recognition, standalone-patch recognition and V1-like algorithm, in figure 5. The roc curve is in figure 6 and 7.

Final comparison algoirthm is a state of the art algorithm (V1-like) mentioned in [14], who provided us code. This algorithm is tested on the medium blurred and severe blurred partitions of the Remote Face Database. We conduct the experiments with the gallery of random 3 images and 5 images respectively for both the partitions. We also conduct experiments on the fixed gallery size of 5 images for both the partitions. The recognition results are mentioned in 4 and 5. The roc curves are in figure 6 and 7.

5.3. Results and Analysis

The comparison of GRAB feature based method and V1-Like algorithm with the KPCA, LDA and SVM based baseline algorithm presented in [13] is given in figure 4. As shown in the graph, the rank 1 recognition rate of GRAB and the KPCA/LDA algorithms are superior to the V1-Like algorithm in medium as well as severely blurred images. The performance of GRAB features based algorithm is comparable to that of the baseline performance in medium blurred images and superior to baseline in severely blurred images. As in prior work, the results show that for all algorithms, in both medium and severely blurred images the performance increases as the number of gallery images per subject increases from 3 to 5. This experiment was done using randomly sampling 3 and 5 gallery images separately with the same probe images. This experiment shows that the performance of GRAB multi-scale features are good descriptors when the quality of the images is low. This is the reason we decided to pick these features and build a context-patch fusion based recognition system for better performance.

Figures 5 shows the recognition results of GRAB feature based method, standalone-patch fusion method and context-patch fusion method. The experiment was done on the same gallery which consisted of 5 images for all 3 algorithms. The results confirm our hypothesis that using context-patch based score level fusion will improve performance compared to basic standalone-patch score based fusion or SVM-based feature level fusion. The recognition gains for the most difficult, severely blurred images, is the most signfficant.

Figures 6 and 7 shows the ROC curves for verification results on the medium blurred and severely blurred partitions. It is clear that the context-fusion based method is superior to any of the other methods in low false accept rate.

From all the experiments and results, it is very clear that as the images get more degraded, and harder is it essential to make use of most of the information available in the images for better recognition and verification purposes. The proposed a way of describing the facial regions with the context information is one of the ways of making most out of the information available in the degraded images.

Figure 3. Sample images of 2 subjects from the Remote Face Database. From left to right:Clear images, medium blur images, severe blur images

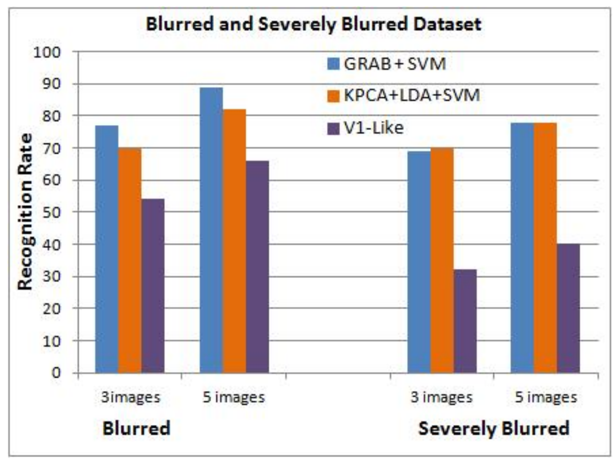

Figure 4. Recognition rates of 3 face recognition alrogithms on medium blurred and severely blurred images from Remote Face Database. The results show that the GRAB descriptors based face recognition method is superior to both baseline(KPCA+LDA+SVM) and V1-Like method. But in blurred dataset, the recognition results of GRAB feature based algorithms is comparable to that of the baseline algorithm.

6. Conclusion and future works

Current face recognition algorithms are limited when it comes to recognition in uncontrolled face recognition. While considering the information fusion for face recognition on images from such unconstrained situation, it is important to anylyze how much dicriminative information is available for fusion for the fused system to perform well. In this paper, we hypothesize that using contextual-patches, a simple type of foveated images, can improve perofmrance.

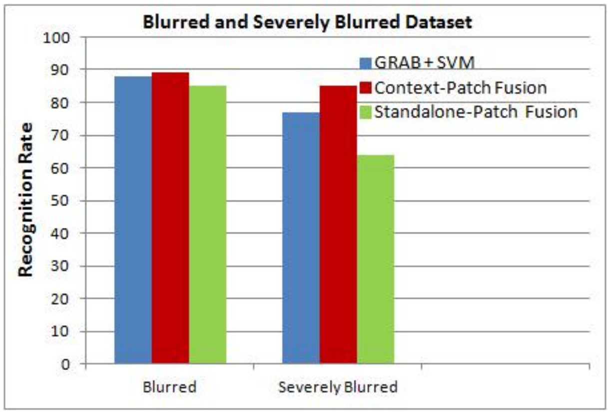

Figure 5. Comparision of Recognition Rate of GRAB feature based , Standalone-patch fusion and Context-patch fusion algorithms. All these methods use the same GRAB descriptors. The methods only differ in the way the patches are combined together. The first one uses the feature level fusion from patches, second uses the fusion of decision scores from standalone-patches and the third one uses the fusion of decision scores from context-patches. The context-patch fusion method outperforms the feature level and standalone patch fusion and the performance gain is especially high on severely blurred images

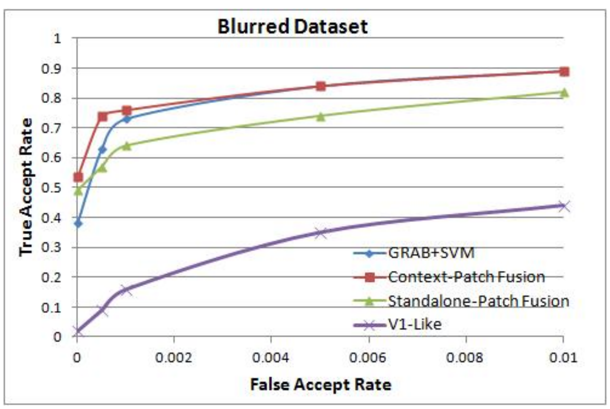

Figure 6. ROC curve, Medium Blurred Dataset: This shows the effectiveness of context-patch on very low false accept rate. The context-patch based fusion approach outperforms the GRAB feature based algorithm and a standalone-patch based algorithm as well as state of the art V1-Like algorithm.

Our experiments provide initial comfirmation of that hypothesis, though larger experimentation is still needed.

To compute define the patches we divided the facial image into multiple regions and defined the context-patches with more information from a particular spatial region associated, combining it with low-resolution data (the contextual information) from the rest of the face regions. Multi-class SVMs are used for the per-patch classification for the context-patches and the decision scores are z-score normalized and fused. Experiments on Remote Face Database demonstrated that the proposed method has promising results especially when the image quality is highly degraded. We compare our method to the state of the art algorithms and achieve better performance.

In future work, a detailed analysis on the value of context in patch-based face sould be done as well as much larger experiments. Such future work whould examine how much of the context information along with a specific patch information could be useful instead of the whole face images to reduce the computational complexity. Also, more robust decision fusion methods could be studies. In addition the choice and location of "local regions" should be examined. Researchers developing patch-based techniques are encouraged to try adding context to their methods.

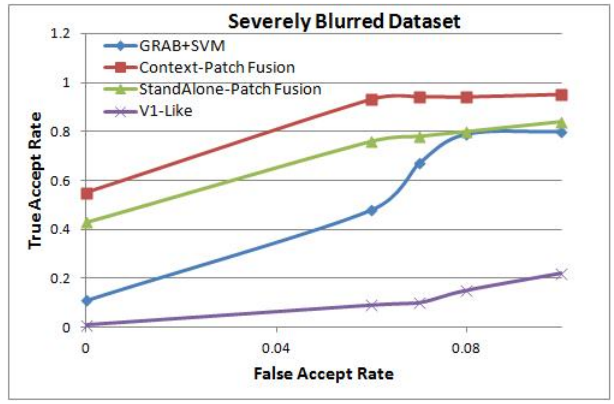

Figure 7. ROC curve, Severely Blurred Dataset:This shows the effectiveness of context-patch on very low false accept rate. The context-patch based fusion approach outperforms the GRAB feature based algorithm and a standalone-patch based algorithm as well as state of the art V1-Like algorithm

References

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE TPAMI*, 28(12):2037–2041, 2006.

[2] M. Bengherabi, L. Mezai, F. Harizi, A. Guessoum, and M. Cheriet. Score fusion of svd and dct-rlda for face recognition. In *Image Processing Theory, Tools and Applications, 2008. IPTA 2008. First Workshops on*, pages 1 –8, nov. 2008.

[3] M. Bicego, E. Grosso, A. Lagorio, G. Brelstaff, L. Brodo, and M. Tistarelli. Distinctiveness of faces: A computational approach. *ACM Trans. Appl. Percept.*, 5:11:1–11:18, May 2008.

[4] G. Brown, J. Wyatt, R. Harris, and X. Yao. Diversity creation methods: a survey and categorisation. *Information Fusion*, 6(1):5 – 20, 2005. ¡ce:title¿Diversity in Multiple Classifier Systems¡/ce:title¿.

[5] O. Celiktutan, H. Akakin, and B. Sankur. Multi-attribute robust facial feature localization. In *Automatic Face Gesture Recognition, 2008. FG '08. 8th IEEE International Conference on*, pages 1 –6, sept. 2008.

[6] K. Etemad and R. Chellappa. Discriminant analysis for recognition of human face images. *Journal of Optical Society of America A*, 14:1724–1733, 1997.

[7] F. Fang, Z. Ma, L. Qing, J. Miao, X. Chen, and W. Gao. Face reconstruction using fixation positions and foveated imaging. In *Automatic Face Gesture Recognition, 2008. FG '08. 8th IEEE International Conference on*, pages 1 –6, sept. 2008.

[8] B. Heisele, P. Ho, and T. Poggio. Face recognition with support vector machines: global versus component-based approach. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 2, pages 688 –694 vol.2, 2001.

[9] D. Kisku, M. Tistarelli, J. Sing, and P. Gupta. Face recognition by fusion of local and global matching scores using ds theory: An evaluation with uni-classifier and multi-classifier paradigm. In *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*, pages 60 –65, june 2009.

[10] J. Kittler, M. Hatef, R. Duin, and J. Matas. On combining classifiers. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(3):226 –239, mar 1998.

[11] X. Lu, Y. Wang, and A. Jain. Combining classifiers for face recognition. *Multimedia and Expo, IEEE International Conference on*, 3:13–16, 2003.

[12] S. Minut, S. Mahadevan, J. M. Henderson, and F. C. Dyer. Face recognition using foveal vision, 2000.

[13] J. Ni and R. Chellappa. Evaluation of state-of-the-art algorithms for remote face recognition. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 1581 –1584, sept. 2010.

[14] N. Pinto, J. J. DiCarlo, and D. D. Cox. How far can you get with a modern face recognition test set using only simple features? In *IEEE CVPR*, 2009.

[15] A. A. Ross, K. Nandakumar, and A. K. Jain. *Handbook of Multibiometrics (International Series on Biometrics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[16] A. Salah, E. Alpaydin, and L. Akarun. A selective attention-based method for visual pattern recognition with application to handwritten digit recognition and face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(3):420–425, 2002.

[17] A. Sapkota, B. Parks, W. Scheirer, and T. Boult. FACE-GRAB: Face Recognition with General Region Assigned to Binary Operator. 2010.

[18] Y. Su, S. Shan, X. Chen, and W. Gao. Hierarchical ensemble of global and local classifiers for face recognition. *Computer Vision, IEEE International Conference on*, 0:1–8, 2007.

[19] B. Topcu and H. Erdogan. Decision fusion for patch-based face recognition. *Pattern Recognition, International Conference on*, 0:1348–1351, 2010.

[20] G. Van Belle, M. Ramon, P. Lefèvre, and B. Rossion. Fixation patterns during recognition of personally familiar and unfamiliar faces. *Frontiers in Psychology*, 1, 2010.

[21] M. Vatsa, R. Singh, A. Ross, and A. Noore. Likelihood ratio in a svm framework: Fusing linear and non-linear face classifiers. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08. IEEE Computer Society Conference on*, pages 1 –6, june 2008.

[22] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surv.*, 35:399–458, December 2003.