Combination of Classifiers on the Decision Level for Face Recognition

Bernard Achermann and Horst Bunke
Institut für Informatik und angewandte Mathematik
Universität Bern, Neubrückstrasse 10, CH-3012 Bern
email: ackerman,bunke@iam.unibe.ch

IAM-96-002

January 1996

Combination of Classifiers on the Decision Level for Face Recognition

Bernard Achermann and Horst Bunke
Institut für Informatik und angewandte Mathematik
Universität Bern, Neubrückstrasse 10, CH-3012 Bern
email: ackerman,bunke@iam.unibe.ch

January 1996

Abstract

This report is divided into two parts. In the first part we present an overview of sensor fusion. We analyze the single steps in a fusion process and describe in a systematic way several methods for combining pattern classifiers. The second part describes some practical experiments with classifier combination in the field of face recognition. We investigate the impact of decision combination on the recognition rate. Several combination strategies and their results are compared. The face classifiers used for the practical experiments are shortly presented, too. We used two full face classifiers (HMMs, eigenfaces) and a profile classifier (shape comparison). Thus, the combination is based on two completely different information sources.

CR Categories and Subject Descriptors: I.4.8 [Image Processing]: Scene Analysis; I.5.0 [Pattern Recognition]: General; I.5.2 [Pattern Recognition]: Design Methodology.
General Terms: Algorithms, Design.
Additional Key Words: Sensor Fusion, Classifier Design, Classifier Combination.

Combination of Classifiers on the Decision Level for Face Recognition

Bernard Achermann and Horst Bunke
Institut für Informatik und angewandte Mathematik
Universität Bern, Neubrückstrasse 10, CH-3012 Bern
email: ackerman,bunke@iam.unibe.ch

January 1996

Abstract

This report is divided into two parts. In the first part we present an overview of sensor fusion. We analyze the single steps in a fusion process and describe in a systematic way several methods for combining pattern classifiers. The second part describes some practical experiments with classifier combination in the field of face recognition. We investigate the impact of decision combination on the recognition rate. Several combination strategies and their results are compared. The face classifiers used for the practical experiments are shortly presented, too. We used two full face classifiers (HMMs, eigenfaces) and a profile classifier (shape comparison). Thus, the combination is based on two completely different information sources.

1 Introduction

The aim of this work is to present a systematic overview of different methods in the field of sensor fusion and to verify the advantage of classifier combination with practical experiments in face recognition. We will provide a description of methods which are generally applicable to sensor fusion problems on the decision level (for a more detailed discussion of the different levels in the fusion process see Section 2). The different steps in this process will be analyzed and discussed. Then some practical experiments based on two full face and a profile classifier are presented. The profit of the previously discussed combination strategies will be investigated.

Experience in pattern recognition showed that different classifiers (e.g., for numerals or characters) have their particular advantages and disadvantages. Depending on the features or on the input data, a certain classifier is better suited for the recognition of certain patterns than another one. The latter instead may yield better results under other circumstances. Therefore, it is desirable to combine the results of classifiers into a joint result, which of course should take into account the weak and the strong properties of every classifier involved in the fusion process. The crucial point of this process, however, is to find a smart combination method which meets these requirements.

In the last years there has been a growing interest in sensor fusion methods. Experiments in character recognition, word recognition and robotics have shown promising results. Furthermore, we implemented a heuristic approach to combine two different face classifiers (for details about our previous work see [20]) which also showed good results. Therefore, we decided to expand our investigations in a systematic way on different combination methods and on the combination of more than two classifiers. Since we are relying on the results of face classifiers which yield a ranking of decisions as output, we concentrated on the combination on the decision level.

This report is organized as follows. In Section 2 we give a brief survey on sensor fusion and address systematically the single steps in the combination process (score transformation, combination set reduction, and combination strategies). Practical experiments are presented in Section 3, where we describe our face classifiers and discuss the experimental results. Finally, we present conclusions and possible future directions of our research in Section 4.

2 Overview of Sensor Fusion

2.1 General Remarks

The field of sensor fusion has been growing rapidly since the second half of the eighties, though first investigations in the field date back to the sixties. Since about 1988, there has been an explosive growth of publications and conferences. Because it is a rather new discipline within computer science, however, there is no widely recognized theory around yet. Good textbooks covering the important topics of sensor fusion are those by Waltz and Llinas [19], and Hall [5]. An extended bibliography and a collection of fundamental papers are provided by Dasarathy [2]. Basically, there are two research communities with high interest in sensor fusion: developers of military applications (automatic target recognition, tactical decision systems), and researchers in non-military areas (robotics, pattern recognition) with a more general interest in sensor fusion. Ref. [19] has a strong emphasis on military questions, whereas [5] and [2] are concerned with general sensor fusion. Other researchers, like Ho [6] in word recognition, have their focus mainly on the application field and are interested in sensor fusion only in order to get better recognition rates for their specific problem. In the following paragraphs, we summarize the key points of combining classifiers. It is mostly based on [19] and [5].

The idea of combining multiple inputs to infer informations about the actual environment is very natural. It is done by humans in everyday life, since we are all the time combining acoustic, visual, tactile, olfactory and thermal informations to react on the world around us. Biological systems seem to adopt such combination schemes very often to get more reliable knowledge about the environment. Sometimes it is even not possible to derive the information needed for a particular task from one single sensor, but in a joint effort it may be done. Furthermore there is no perfect sensor, so it is reasonable to make use of the favourable properties of a sensor and to suppress the disadvantages by applying a smart combination scheme.

A number of sensors observing the same phenomena (scene or objects) are required for a sensor fusion process. The sensors may be of different types (e.g., infrared and radar) and they may have different views of the scene, but it is required that there is a certain overlap in the observation data.

The sensor fusion process consists of several steps:

- **Data acquisition**

  As stated above several sensors observe the same phenomena. They produce raw data (signals), which require further processing.

- **Signal processing**

  Often it is necessary to apply filters on the signals for noise reduction, to reduce the amount of data, or to transform the data to another format.

- **Classification**

  A classification step takes place to identify the observed phenomena.

- **Combination**

  The information extracted until this moment in the process is used for a fused decision. The term "fusion" may be understood in a rather wide sense, since the collected informations may be incorporated into a larger context in order to plan further actions.

The use of sensor fusion is motivated by the expectation of better performance. In our case, we expect a higher recognition rate and a higher reliability of the results. Hall [5] gives a list of benefits of data fusion:

- Robust operational performance
- Extended spatial coverage
- Extended temporal coverage
- Increased confidence
- Reduced ambiguity
- Improved detection
- Enhanced spatial resolution
- Improved system reliability
- Increased dimensionality

Some interesting rules of thumb are given by Nahan and Pokoski (cited according to Hall [5]):

1. Combining data from multiple inaccurate sensors (having an individual probability of correct inference less than 0.5) does not provide a significant overall advantage.
2. Combining data from multiple highly accurate sensors (having an individual probability of correct inference of greater than 0.95) does not provide a significant increase in inference accuracy.
3. When the number of sensors becomes large (e.g., greater than 8 or 10), adding additional identical sensors does not provide a significant improvement in inference accuracy. Note, however, that adding a new sensor type may have a very significant impact in inference capability, because of an added dimensionality of observational data.
4. The greatest marginal improvement in sensor fusion occurs for a moderate number of sensors (i.e., one to seven), each having a reasonable probability of correct identification.

These points encouraged us in our intention to make use of sensor fusion, since we want to combine several classifiers which are based on different kinds of data (grey level images, binary images and range images of human faces) and on different views of the same scene (profile and full face view of faces). Sensor fusion is often applied in character recognition to combine the results of different classifiers. We make a broader use of sensor fusion in our approach, because in contrast to character recognition our classifiers are based on different input data. Thus we integrate more information into our fusion process.

The description of the fusion step in the list above has been significantly simplified. In fact, it is possible to realize sensor fusion on different levels: sensor data level fusion, feature vector level fusion, and decision level fusion. One may also think of hybrid approaches, especially when a larger number of sensors is involved. In our case, the main interest is in fusion on the decision level, since our classifiers yield decision rankings as results. Data fusion on sensor level or on feature level is not possible with the current implementation of our classifiers. We rather use them as black boxes yielding a ranking together with a score function, so we have no possibility to fuse the information earlier than on the decision level. Furthermore we think that fusion on the decision level has much more potential applications, since most of the classifiers implemented so far return decisions as result.

Another point is the sequence of the combination processes. Generally, there is the possibility of parallel, serial or hybrid architectures. Combination schemes may also be organized as multi

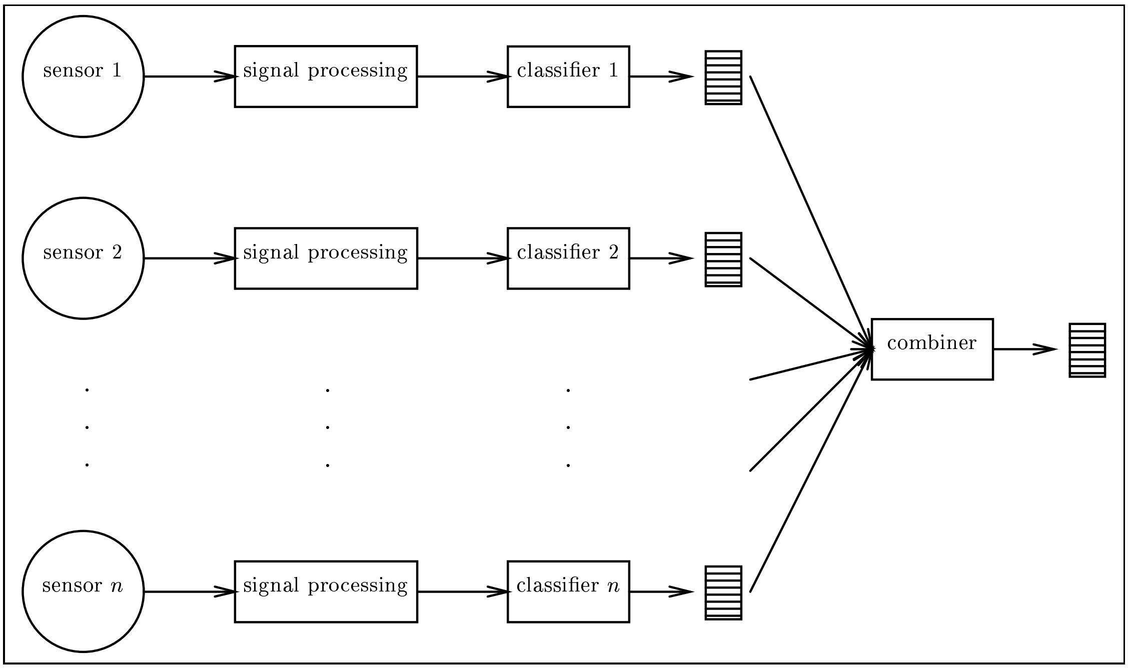


Figure 1: Topology of a simple combination scenario

stage processes. In this report, we restrict our investigations to a parallel combination of the input classifiers which is "naturally" given by the problem we want to solve. Figure 1 shows the topology of such a simple combination scenario.

Another point to mention is the redundancy of classifiers. The third rule of thumb of Nahan and Pokoski refers to this phenomenon. If two classifiers always yield the same or nearly the same results, they are obviously very dependent one on the other. In this case the second classifier does not contribute very much additional information to the fusion process, so it is highly redundant. Furthermore, there is even an increased danger of misclassifications in the combination result if several redundant classifiers are integrated in the fusion process. If they yield misclassifications, the plausibility for this class might be overweighted by the combiner. Unfortunately, it is in practice often not very easy to determine to which degree two classifiers are redundant.

A further aspect of sensor fusion is the rejection of a combined result. A combiner is expected to yield very reliable results, but sometimes this requirement cannot be met, because the input is rather uncertain or inconsistent. In such cases the combiner preferably rejects a joint decision.

2.2 Formal Problem Description

Our aim is to combine m classifiers $C_i (i = 1, \ldots, m)$ with a combination classifier K. A look at Figure 1 makes clear that the combiner is nothing else but an additional classifier working with the output of the classifiers C_i as input features, whereas the classifiers C_i are based on the data of the sensors. To make a clear distinction based on the type of input we will use from now on the terms *pattern classifier* for the classifiers C_i and *combination classifier* for K.

A *classifier* is a function which assigns a class to an input pattern based on a set of features. Of course, the decision of the classifier may be correct or wrong. Some classifiers reject the input if there is not enough evidence for the pattern belonging to one of the possible classes.

Classifiers may also be distinguished depending on the output information they generate. There are three possible types of classifiers:

- Type 1 classifiers. They yield a ranking with a score value for every class. Type 1 classifiers are the basic type for all other classifiers, which may be interpreted as special cases.
- Type 2 classifiers. They generate a ranking without score values. The rank itself, however, may be interpreted as a score function.
- Type 3 classifiers. They yield a decision for not more than one class. This type may be regarded as a classifier which yields a ranking consisting of only the first rank.

Generally, the ranking of a pattern classifier consists of the name of the class and the value of the score function. The class name is an element of a set of possible classes. The score function is also called discriminant and is dependent on the classifier's internal evaluation of the features it extracted from the input.

Formally we denote the ranking of a classifier C_i in the following way:

$$R_{C_i} = ((n_1, s_1), (n_2, s_2), \dots, (n_N, s_N)) \tag{1}$$

with N as the number of the ranks returned by C_i, $n_j (j = 1, \dots, N)$ as the class name on rank j, and s_j as the value of the score function on rank j.

The order in the ranking is established using the score function. For convention, the ranking is established in ascending order of the score function. A lower score value s_i means a better classification result for the presented pattern, i.e.

$$s_i \le s_{i+1}, i = 1, \dots, N-1 \tag{2}$$

The combination classifier is based on the output of the pattern classifiers. It yields a new ranking with the same properties as described above.

The *fusion process* is divided into the following steps:

- **Transformation of the score values**

  If the score functions yield values which are not directly comparable, a transformation step is required.

$$R'_{C_i} = \mathrm{Trans}(R_{C_i}) = ((n_1, \mathrm{Trans}(s_1)), \dots, ((n_N, \mathrm{Trans}(s_N))) \tag{3}$$

- **Reduction of the combination set**

  The *combination set* S is the set of classes which are involved in the combination step. Basically it might be every class that is a result of at least one of the pattern classifiers C_i. But mostly a reduction step takes place because it is not reasonable to take every class into account.

$$S = \mathrm{Red}(\{(n_j, s_j) | (n_j, s_j) \in R_{C_i}, i = 1, \dots, m\}) \tag{4}$$

- **Combination and reordering**

  For every class in the combination set, a *combination rule* is applied and the classes are reordered in order to get a new ranking.

$$R_K = \mathrm{Com}(S) \tag{5}$$

A more detailed discussion of these steps is presented in the following subsections.

2.3 Transformation of the Scores

For combination schemes based on score functions it is necessary in most cases to make transformations of the score values of the involved classifiers in order to make them comparable. The score functions, however, can be of different types. Therefore, different strategies have been developed to avoid the direct manipulation and use of the scores (e.g., voting strategies or rank based combination schemes; for more details see Section 2.5).

The score functions may be of the following type:

- Distance measurements
- Probabilities
- Quality measurements of the decision

It is a crucial point in the fusion process to find smart transformations for the score, in order to combine classifiers with a wide variety of score functions.

The score function of a classifier gives more information than only the ranking. A score is also an index for the quality of the decision and shows more details about the relation between the classes. It is, however, important to have knowledge about the distribution of the score values and the discrimination ability based on the score for a certain classifier in order to be able to construct an efficient combination classifier. Ideally, there should be a large gap between the score value of the first and of the second rank. In this case, one may assume that the classifier was able to designate a clear favourite class. Therefore, there is high evidence for a correct classification. A profound discussion of class separability and the behaviour of classifiers may be found in Fukunaga [3].

Among others, there exist the following types of transformations:

- Linear
- Logarithmic
- Exponential
- Logistic

All these transformations do not affect the order of the transformed classifier ranking, because they are based on monotone functions. This is a very desirable property, since any modification of the rank order should take place in the combination step and nowhere else. The purpose of these transformations is, first, to map the scores to the same range of values, and, second, to change the distribution of the scores. For example, the logarithmic transformation puts strong emphasis on the top ranks, whereas lower ranked scores, which are transformed to very high values, have a quickly decreasing influence. The behaviour of the transformations is shown in Figure 2.

In the next paragraphs, we will use the following notation conventions: s is an arbitrary value resulting from the score function of a classifier. The range of the score function is denoted by $I_s = [s_{min}, s_{max}]$. A value in the target range $I_{s'} = [s'_{min}, s'_{max}]$ of a transformation is denoted by s'.

A *linear transformation* maps the interval I_s into $I_{s'}$ by interpolation:

$$s' = s'_{min} + \frac{s - s_{min}}{s_{max} - s_{min}}(s'_{max} - s'_{min}) \tag{6}$$

Under a *logarithmic transformation* the score values are first linearly transformed into $I_{s''} = [0.0, 100.0]$ (normalization). Then the transformed score is computed as:

$$s' = \log(1 + s'') \tag{7}$$

The *exponential transformation* is defined similarly. After the normalization step the score is calculated as:

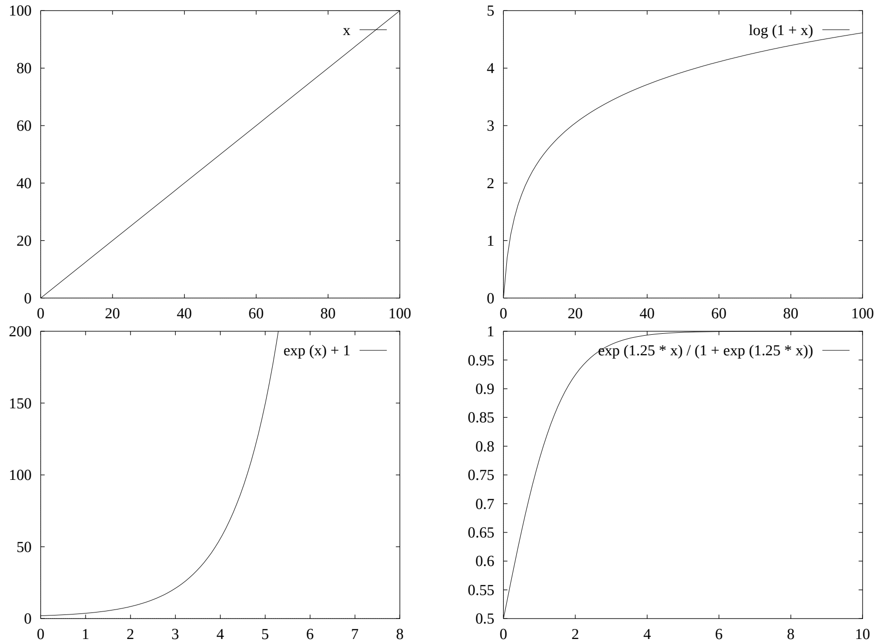


Figure 2: Score transformation types (linear and logarithmic transformation in the upper row, exponential and logistic transformation in the lower row)

$$s' = \exp(s'') - 1 \tag{8}$$

A *logistic transformation* is done in the following manner. First, a linear transformation to the interval $[0.0, 100.0]$ is computed. Then the new score is:

$$s' = \frac{\exp(\alpha + \beta s)}{1 + \exp(\alpha + \beta s)} \tag{9}$$

The parameters α and β have to be determined empirically; α defines where the function intersects the x axis and β controls the slope.

2.4 Combination Set Reduction

For the decision fusion of pattern classifiers it is required to determine for which classes a joint decision has to be computed. This step is called set reduction, because its aim is to determine which classes are reasonable to be involved in the fusion process. The resulting set of classes is the so called combination set. Especially for a large number of classes it is inefficient and computationally expensive to calculate a joint score for every possible class in the data base.

The set reduction step is defined by two parameters: reduction criterion (reduction on classifier level) and reduction operation (reduction on combination level). The reduction criterion specifies which classes of the input data are used in the fusion process, and the reduction operation determines how these input classes are combined.

The reduction criterion may be based on the rank or on the score information of the pattern classifiers. For instance, the top k decisions of every classifier ranking may be exclusively taken into account, since it is supposed that the classes ranked below k are of no relevance to the fusion process. Another possibility is to exclude classes with a score value above a certain threshold. The goal of this step is to get rid of candidate classes which have a low evidence to be the correct class.

The result of this step is an intermediate set per classifier:

$$R'_{C_i} = \{(n_j, s_j)|j < t\} \tag{10}$$

if highly ranked classes should be excluded, and

$$R'_{C_i} = \{(n_j, s_j)|s_j < t\} \tag{11}$$

if classes with a bad score should be eliminated. In both cases t is a predefined threshold value.

After the reduction on classifier level, the combination set S has to be determined. Basically there are the following possibilities: no reduction, intersection, or union, i.e.

$$S = S', \tag{12}$$

$$S = \bigcap_{i=1}^{m} R'_{C_i}, \text{ or} \tag{13}$$

$$S = \bigcup_{i=1}^{m} R'_{C_i}. \tag{14}$$

2.5 Combination Strategies

There exist many possibilities to combine classifier decisions, and there is nearly no limit in inventing more or less sophisticated combination rules. There are, however, very special combination rules, which may be applied to one particular problem, and general combination schemes with a wide applicability. We are rather interested in such general rules.

The methods for combining classifiers on the decision level presented in this paper may be roughly divided into two categories: heuristic approaches, like voting strategies or rank sums, and methods based on probability theory and statistics, like Bayesian inference or Dempster-Shafer theory.

In the following paragraphs we describe some of these methods.

2.5.1 Voting Strategies

There are some strategies in data fusion which are motivated by the way humans are making decisions, especially when there is a group involved in the decision process. Each sensor is in the position of a human expert with one vote. The resulting decision is depending on the majority of the votes. If the voting ends in a draw, the combination classifier is unable to decide for a certain class. In this case a reject is returned.

The majority criterion for the voting may be more or less strict. If there is an absolute majority required for a joint decision, the strategy is called *majority voting*. If a relative majority is sufficient, we speak of *consensus voting*. Let $v(n)$ be a function that returns the number of classifiers with class n on the first rank:

$$v(n) = |\{C_i \mid n_1 = n, n_1 \in R_{C_i}\}| \tag{15}$$

The decision rule for the consensus voting is defined as

$$R_K = \{n_i \mid v(n_i) = \max\{v(n_j), n_j \in S\}\} \tag{16}$$

For the majority voting there is the additional restriction that $\max\{|v(n_j)|\} > \lfloor \frac{m}{2} \rfloor$. A rejection will be returned, if $|S| \neq 1$. In this case either no class fulfills the criteria (S is an empty set) or more than one class has the same number of votes (S has several elements).

Variants of these strategies are weighted voting methods, where each expert has its individual weight.

$$v(n) = \sum_{n=n_1} w_{C_i}, n_1 \in R_{C_i} \tag{17}$$

The weights w_{C_i} have to be defined before the fusion process is applied. They may be found empirically or by computing the recognition rates of the sensors with a test set. Those recognition probabilities may be used as weights. An alternative is described by Ho [6, 7] where the weights are computed by fitting a regression plane to the data. Recently, Lam and Suen [12] presented an approach where the weights are optimized using a genetic algorithm. All these methods choose the weights dependent on the performance of the classifiers.

Voting methods only take into account a small part of the classifier's result, since they are exclusively based on the first rank. Often there is a very high probability to end up in a draw.

For more details on voting strategies see Mandler and Schürmann [14] or Lam and Suen [10, 11].

2.5.2 Rank Based Strategies

As stated previously it isn't always possible to deal with the scores of the individual classifiers, since they may not be comparable. In order to avoid the disadvantages of transformations and nevertheless use more information than only the top decision of a classifier, one may rely on the rankings of the involved classifiers. Alike voting strategies the ranking loses information, too, but it is not as coarse as just being based on the top decisions. Rank based strategies are a generalization of the simple voting methods.

The simplest approach is to compute the sum of the rank for every class in the combination set. The class with the lowest rank sum will be the first choice of the combination classifier. Let $r(n, R_{C_i})$ be the rank of the class with name n in the ranking R_{C_i}. The combination is computed in the following way:

$$R_K = \left\{ \left(n_j, \sum_{i=1}^{m} r(n_j, R_{C_i}) \right), n_j \in S \right\} \tag{18}$$

An equivalent formulation of this method is the so called *Borda count*. It is the sum of the number of classes ranked below class n, which will be reordered in descending order of the Borda count. For a constant number of ranks the Borda count yields the same results as the rank sum.

Though the score transformation is primarily thought for the manipulation of the score based strategies, it may be applied to the ranks (interpreted as scores in this case) too. In this way it is possible to change the influence of the ranks significantly. Depending on the chosen transformation a certain rank will have increased or reduced influence compared to the ranks above or below it.

In analogy to the voting methods it is, of course, possible to apply an individual weight for each classifier.

2.5.3 Score Based Strategies

If the score functions are directly comparable or if there exists at least an acceptable transformation scheme to make the involved classifiers comparable, score based strategies are a good way for decision fusion. This is reasonable because the score functions provide the combiner with some additional information which isn't available in the voting and the rank based methods described above. Of course, the quality of the score function is of eminent importance in this case.

The simplest way to combine classifiers using the score is to compute the sum of the score functions. The ranking is then reordered following the ascending scores of R_K.

$$R_K = \left\{ \left(n_j, \sum_{i=1}^{m} s_j \in R'_{C_i} \right), n_j \in S \right\} \tag{19}$$

2.5.4 Bayesian Inference

Bayes's theorem may be used for the inference of the joint probability of the input classifiers: $F_j(j = 1, \ldots, N)$ denotes the event that a pattern of class j has been shown to the classifiers C_i, with N being the number of classes. H_k instead represents the fact that the classifier identified the input as class k (hypothesis). The Bayes's theorem states:

$$P(F_j|H_k) = \frac{P(F_j)P(H_k|F_j)}{\sum_l P(F_l)P(H_k|F_l)} \tag{20}$$

If we have no information that event F_j happens more often than $F_k(j \neq k)$, we set

$$P(F_j) = \frac{1}{N} \tag{21}$$

The probabilities $P(H_k|F_j)$ have to be estimated. Bayes's theorem requires the condition

$$\sum_k P(F_j)P(H_k|F_j) = 1 \tag{22}$$

Since we have no further information we assume that given a pattern F_j the classification is correct with the probability p. Because of (22) the remaining probability values are equally set to the remainder of the probability.

$$P(H_k|F_j) = \begin{cases} p & k = j \\ \frac{1-p}{N-1} & k \neq j \end{cases} \tag{23}$$

The best way to determine p is to make use of a training set. With the aid of this set the recognition rate given a certain input pattern may be computed. The test set, however, should be large enough to get reliable results. In this case it is possible to rely on statistical data, but subjective estimation of the probabilities being based on experiences with the respective sensors may be used, too.
For a joint decision we have to compute:

$$P(F_j|\bigcap_l H_k^{(l)}) = \frac{P(F_j)\left(\prod_l P(H_k^{(l)}|F_j)\right)}{\left(\sum P(F_l)\right)\left(\prod_l P(H_k^{(l)}|F_j)\right)}, j = 1, \ldots, N \tag{24}$$

As joint decision the F_j with the highest joint probability will be elected.

The major drawback and the crucial point of this method is the determination of the a-priori-probabilities $P(H_k|F_j)$. These probabilities have to be defined before the fusion process may be applied. However, this flexible determination allows a very sophisticated tuning of the combination classifier, especially for large training sets. This is certainly an advantage of this method.

2.5.5 Method of Yu

This method is in detail described in our paper [20]. It is a rank based classifier, where it is additionally tried to integrate the computation of weights based on the expected quality and the discrimination ability of the two involved pattern classifiers C_1 (HMM classifier) and C_2 (profile classifier). In order to fulfill these goals the combiner was constructed in the following way:

- **Determination of the reduction threshold**

  In the first step, it is determined which part of the ranking of each classifier is used for the fusion process. For the classifier C_1 this value is set statically to the uppermost $k = 12$ ranks. This value was found heuristically. For the classifier C_2 the value is computed dynamically:

$$t_P = \left\lfloor \alpha \left(\frac{s_1}{s_2} + \ldots + \frac{s_1}{s_k} \right) - 1 \right\rfloor, \text{ where } (n_i, s_i) \in R_{C_2} \tag{25}$$

Based on training data, a suitable value $\alpha \approx 4$ was found. If $t_P < 1$ or $t_P > m$, then $t_P = 1$ and $t_P = m$, respectively, were chosen.

- **Reduction operation**

  With the classes fulfilling the reduction criterion, the combination set is computed by means of an intersection operation. If the combination set is empty, the combination classifier returns a rejection.

- **Combination and reordering**

  The combination score is calculated as a weighted rank sum:

$$s_i = w \cdot u + (1 - w) \cdot v \tag{26}$$

  where $n_i = n_u = n_v, (n_i, s_i) \in S, (n_u, s_u) \in R_{C_1}, (n_v, s_v) \in R_{C_2}$.
  The weight is computed as

$$w = \beta \left(\frac{s_1}{s_2} + \ldots + \frac{s_1}{s_m} \right), \text{ where } (n_i, s_i) \in R_{C_2} \tag{27}$$

  Experimentally a suitable value $\beta = 0.36$ was determined. The ranking is reordered in ascending order.

2.5.6 Other Strategies

As stated above, the pattern classifiers may be regarded as feature extractors, and the combiner works as a classifier based on these features. With this background in mind, it is clear that any statistical classification method, including artificial neural networks, may be applied to the problem. For additional information concerning the use of neural networks refer to Huang and Suen [8], or Lee and Srihari [13]. Other approaches are based on decision trees ([9]), production rules, fuzzy logic ([1]), blackboard architectures ([4]) and regression analysis.

3 Practical Combination Experiments in Face Recognition

3.1 Face Classifiers

The results of the practical experiments presented in this report are based on the combination of classifiers for face recognition. First, we briefly describe the face classifiers. Since other classifiers for face recognition are currently under construction, we are planning to integrate them later in our combination scheme.
In this work we used the following face classifiers:

1. Classification of full face images based on hidden Markov models (see Subsection 3.1.2)
2. Classification of full face images based on eigenfaces (see Subsection 3.1.3)
3. Classification of profile images based on the shape (see Subsection 3.1.4).

Note that we have two completely different kinds of input data.
Figure 3 shows a graphical representation of the recognition rates.

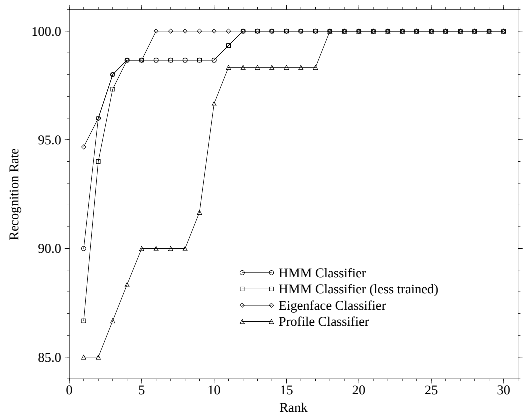


Figure 3: Recognition rates of the pattern classifiers

Figure 4: Example of one person in the data base (full face images in the upper two rows, binarized profile images in the last row)

3.1.1 Data Collection

Since it was our intention to combine classifiers working with different views of human faces, we collected a small data base which is divided into two parts:

1. **Frontal images**

   Included are 10 grey level images of each person with varying head positions (2 looking straight into the camera, 2 looking to the left, 2 looking to the right, 2 downwards, and 2 upwards).

2. **Profile images**

   We took 5 images of each person with varying size and orientation of the head.

The lighting conditions during image acquisition were carefully controlled. The data base contains images of 30 persons. An example of one person is shown in Figure 4.

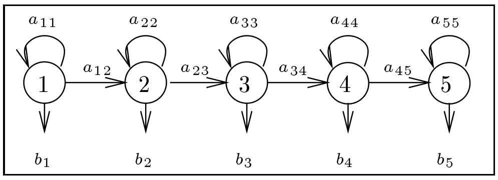


Figure 5: Linear left-right human face model

3.1.2 Full Face Classification with Hidden Markov Models

This classifier for full face images is based on hidden Markov models (HMMs). The approach we used is similar to the method of Samaria described in [17] and is in detail presented in Nyffenegger's thesis [15]. An introduction to hidden Markov models is provided by Rabiner in [16].

Generally, HMMs are working on one-dimensional signals or feature vectors. Images instead contain two-dimensional information. In order to make HMMs applicable to images, we reduce the image information to one-dimensional vectors by applying a sliding window. This window is moving from the top of the image to the bottom and covers the whole width of the image. The step size is chosen so that two successive windows have a certain overlap. The reason for the overlap is to avoid the cutting of significant face features and bring some context information into the process. The values in the sliding window are given (after some preprocessing steps) as feature vector to the HMM.

The idea underlying this method is the following. Intuitively, a human face consists of a number of regions like forehead, eyes, nose, mouth, chin and so on. These regions remain identifiable for a human observer, even when the face image is cut into sliding windows as described above. With the HMM we try to make use of this property. A human face is represented by a linear left-right model consisting of five states. These states correspond to the face parts ("1" for the forehead, "2" for the eyes, "3" for the nose, "4" for the mouth and "5" for chin). The feature vectors for the HMM are the intensity values in each window. For a graphical representation see Figure 5.

The input images were preprocessed by means of a histogram equalization and have a normalized width. The determination of the relevant image part has been done by hand, since it is very important that the face is precisely centered in the vertical direction in the image.
The classifier includes the following processing steps:

- **Code book design**

  The feature extraction with the sliding window is run on the test set of face images. All these feature vectors are used to build a code book.

- **Vector quantization**

  The feature vectors are quantized based on the code book built before.

- **Training of the model**

  For every person in the data base the parameterization of the hidden Markov model is calculated. In our case there are five training images per person. Furthermore, this step is iterated a few times in order to allow a fine adjustment of the HMM parameters.

- **Recognition**

  In the test step images of human faces are presented to the HMM which have not been used in the steps 1 to 3. These test images undergo the same preprocessing, feature extraction and vector quantization as the training images. Then the probability for a person in the data base given a certain model is computed for every trained face. The classifier returns a ranking of the possible persons in descending order of the probabilities of the models.

In Figures 3 and 37 the recognition results of the HMM classifier for full face images are given. With a rate of 90.0% correct classifications, the results are fairly good.

As mentioned above, training is an iterative process. For an optimal number of training iterations the classifier will yield the best results. Of course, the recognition rate for a less trained or overtrained HMM is lower. We also used a less trained HMM in order to have a larger test base and to test the performance of the combiner with suboptimal input. The results of the less trained classifier are listed in Figures 3 and 38.

3.1.3 Full Face Classification with Eigenfaces

This full face classifier is an implementation of the eigenface approach proposed by Turk and Pentland [18]. The basic idea is to represent a face as a linear combination of a small number of base vectors. It is supposed that the coefficients of this linear combination are characteristic for the face of a certain person. The base vectors are determined by applying principal component analysis (Karhunen-Loève transformation) to a training set of human faces.

In contrast with the HMM approach for full face recognition, no preprocessing of the images is required due to the fact that the images are already roughly standardized in lighting and position.

- **Determination of the face space**

  The face space is spanned by a set of base vectors which are termed "eigenfaces" by Turk and Pentland. They are computed in the following manner. The images of human faces in the test set with a size of $n \times n$ pixels are interpreted each as a vector $\Gamma_i, i = 1, \ldots, m$ of n^2 dimensions. On this set a principal component analysis is performed. Only the k most meaningful eigenvectors are chosen as base vectors. An eigenvector is considered more meaningful than another if its eigenvalue is higher. These k base vectors u_j are the so called eigenfaces and span the face space.

- **Building a gallery**

  The training set is used to build a gallery of known faces. Every face may be represented as a linear combination of the eigenfaces: $\Gamma_i \approx \sum \omega_{ij} u_j$ with $\Omega_i = (\omega_{i1}, \omega_{i2}, \ldots)$. Since exclusively the most meaningful eigenvectors are used, the loss of data is negligible. Turk and Pentland compute an average weight vector $\bar{\Omega}_l = (\bar{\omega}_{l1}, \bar{\omega}_{l2}, \ldots)$ with all the training images for a certain person. As we are dealing with a small database, we instead keep every representation Ω_i of a person in the gallery.

- **Recognition**

  For each test image Γ the weight vector $\Omega = (\omega_1, \omega_2, \ldots)$ is computed. Then it is determined which Ω_i in the gallery has the smallest Euclidean distance to Ω. If the distance is lower than a certain threshold, the test image is recognized as a picture of the person in the gallery. Otherwise, the image is rejected.

The results of the eigenface classifier are shown in Figures 3 and 39. The figures show that the method performs very well.

3.1.4 Profile Classification

This classifier is based on a comparison of the profile shape of human faces. The details are described in our papers [21] and [20]. The method consists of the following processing steps:

- **Determination of fiducial points**

  The profile line of a given image of a human side view is extracted. On this line some feature points, so called fiducial points, are determined: nose tip, chin, starting point of the nose above the lips, forehead, start of the profile line at the beginning of the hair and end of the profile line at the throat. The extraction of the fiducial points is based on the local curvature of the profile line (for further details see [21]). Additionally, a line L is computed with minimal quadratic distance to the points between the chin and the forehead point.

- **Canonical position**

  Based on the fiducial points a canonical position is computed for the presented profile line. The nose point is moved to the origin and the whole profile is rotated such that L becomes a horizontal line. The profile may be interpreted now as a function depending on x.

- **Computation of the model parameters**

  The model parameters are computed based on the profile images in a training set. For every face in the data base the interval between the maximal and the minimal y value in every point of the profile line is stored.

- **Comparison of the test image with the models**

  The test images undergo the same preprocessing as the training images: extraction of the fiducial points and transformation to the canonical position. The transformation, however, is computed not only for the fiducial points but for a region around these points in order to reduce the impact of an imprecise determination of the fiducial marks (25 variations in the current implementation). Additionally a scaling factor is computed based on the distances between the fiducial points and is applied to the profile line of the test image variations. This profile lines are compared to every model stored in the data base. The comparison is based on a distance measurement of the test profile line to the model profiles. A low distance measure indicates a highly similar test and profile line, so the model with the lowest score is chosen as matching model.

The results of the profile classifier are shown in Figures 3 and 40. As we see, it has a lower recognition rate than the full face image classifiers. This is not astonishing, since the profile contains less information than the frontal face view of a person.

3.2 General Remarks

We have implemented the combination strategies discussed in Section 2 in order to set up some practical experiments with decision fusion.

There exist five test images per person for the frontal face classifiers and two images for the profile classifiers. The data base contains images of 30 people. Each profile image of a person can be combined with each full face image of the same individual. Therefore, there exist totally $30 \cdot 5 \cdot 2 = 300$ possible input combinations for the combiner.

As stated above we have two different recognition results for the HMM classifier, because of a different number of training iterations. In the following sections of this report, we will refer to the results of the optimally trained HMM classifier as HMM1. HMM2 instead will be used for the less trained HMM classifier.

Of course, it is possible to construct very sophisticated combiners and to tune them for a special test set. This has not been the intention of this work. We are rather concerned with generally applicable methods. Therefore, we made no tuning in order to get the best possible results.

The parameters α and β for the logistic transformation are set to $\alpha = 0.0$ and $\beta = 1.25$. These values have been determined empirically.

In the appendix a complete listing of the results of all experiments is given. In these tables for every rank the number of correctly recognized cases is listed. In parenthesis the recognition rate which is calculated as the number of correct recognitions divided by the number of test cases is given. Additionally, the reliability reports the number of correct recognitions divided by the number of test cases minus the rejected cases.

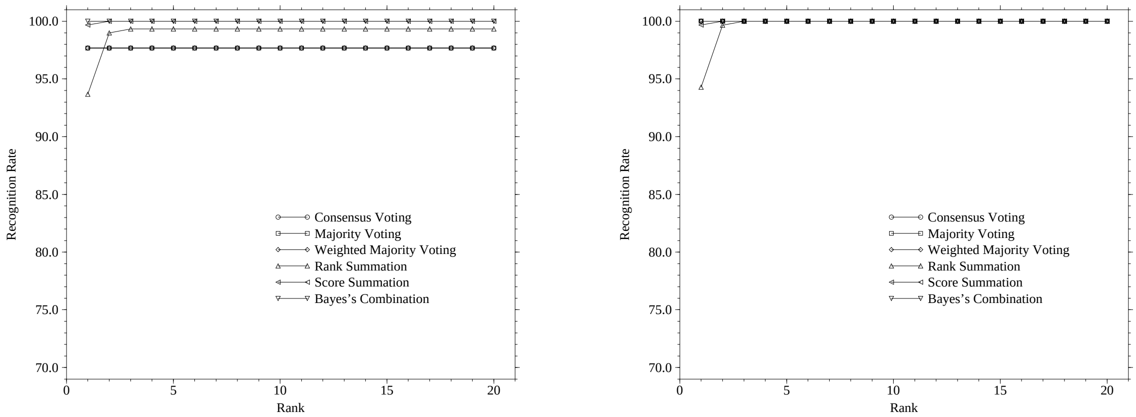


Figure 6: Recognition rates (left) and reliability (right) for the combination of the profile, HMM and eigenface classifier

3.3 Combination of the Profile, the HMM and the Eigenface Classifier

In a first experiment we combined all available face classifiers. The following combination methods were used:

- Consensus voting
- Majority voting
- Weighted majority voting
- Rank summation
- Score summation
- Bayes's combination rule

A graphical representation of the results is shown in Figure 6. The result tables are given in Figures 41 to 47.

All applied methods yield recognition rates clearly higher than 90%. Nearly 100% of the correct classes are ranked within the first three ranks. Thus, the combination of the three classifiers yields very good results which additionally have high reliability. Furthermore, the combiners reach very quickly their optimal performance (within the first three ranks in our example). Even for very simple combinations like the voting methods we get very good results (97.667%). Though the recognition rates of the HMM classifier (90.000%) and of the eigenface classifier (94.667%) are very high there is a significant enhancement of the recognition rate for the combiner.

3.4 Combination of the Profile and the HMM Classifier

In another experiment we combined the profile and the HMM classifier. The same methods as mentioned in Subsection 3.3 were applied. Additionally, we applied the combination method of Yu. The results are presented in Figures 7 and 8. The result tables are shown in Figures 48 to 63.

In this experiment we used both data sets, HMM1 and HMM2. We ran the test with HMM2 in order to compare the combination of two classifiers with recognition rates around 85% with the combination of classifiers where a least one yields better results (HMM1 has a recognition rate

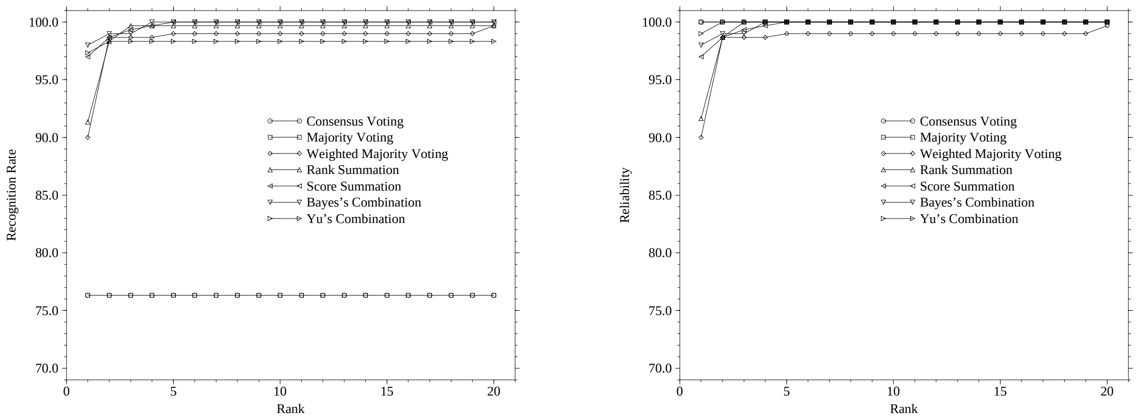


Figure 7: Recognition rates (left) and reliability (right) for the combination profile and HMM classifier (HMM1)

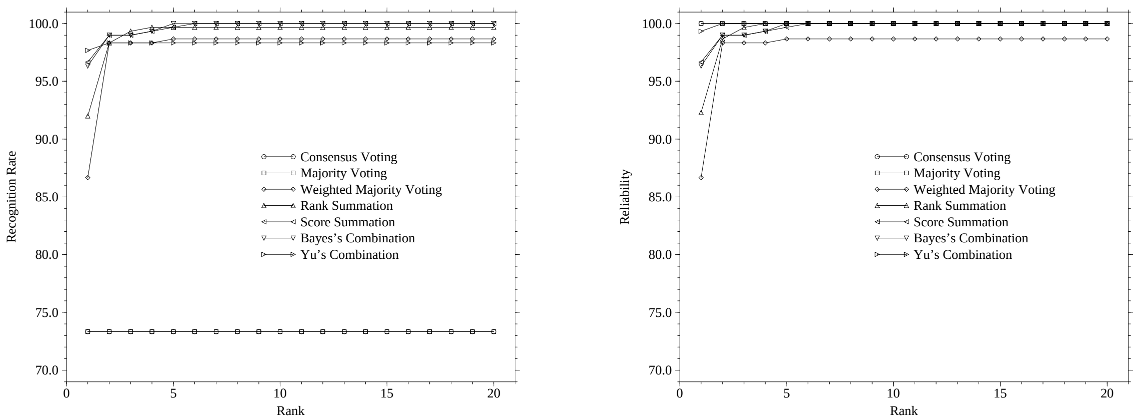


Figure 8: Recognition rates (left) and reliability (right) for the combination profile and HMM classifier (HMM2)

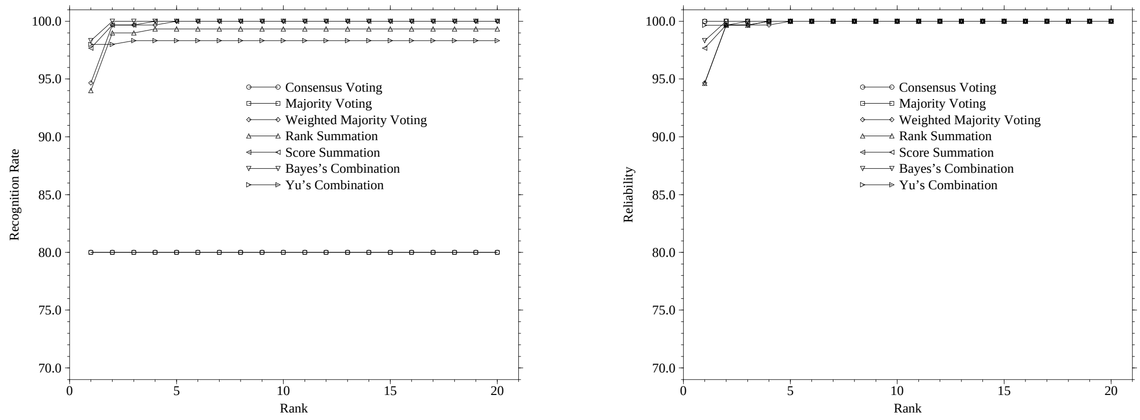


Figure 9: Recognition rates (left) and reliability (right) for the combination profile and eigenface classifier

of 90.0%). It is astonishing that the combined results do not differ very much. The recognition rate for the first rank is generally higher for HMM1 and the combiners reach faster their optimal performance, but most of the results for HMM2 are clearly above 90%. It seems that the combination of classifiers with lower recognition rates yields good results. In this case methods like Bayes's combination and score summation yield clearly better results than simpler techniques (cf. Subsection 3.7 for more information).

3.5 Combination of the Profile and the Eigenface Classifier

In this experiment we combined the profile classifier with the eigenface classifier. We applied the same methods as in the experiment of Subsection 3.4. The results are shown graphically in Figure 9 and in the result tables of Figures 64 to 71.

The results we received for this experiment are very similar to the results for the combination of the profile and the HMM classifier. Remarkable are the good results of the combiner of Yu which originally has been developed for the combination of profile and HMM classifier. Again, te voting strategies perform poorly (see Subsection 3.7 for an explanation).

3.6 Combination of the HMM and the Eigenface Classifier

With this experiment we investigated the combination of the HMM and the eigenface classifier. In contrast with the experiments mentioned above it is only a combination of classifiers working on one single information source. Thus, this combination scheme relies exclusively on images of frontal faces. We applied the same methods as in the experiment of Subsection 3.3. The results are shown in Figure 10 and Figures 72 to 78.

The results are slightly better than those for the combinations of the profile classifier with the HMM or the eigenface classifier. Obviously they supplement each other very well. The voting strategies yield the worst results. This phenomenon is discussed in Subsection 3.7.

3.7 Consensus Voting and Majority Voting

The results of the consensus and the majority voting are shown in the Figures 11 and 12. In our case both combiners yielded always identical results.

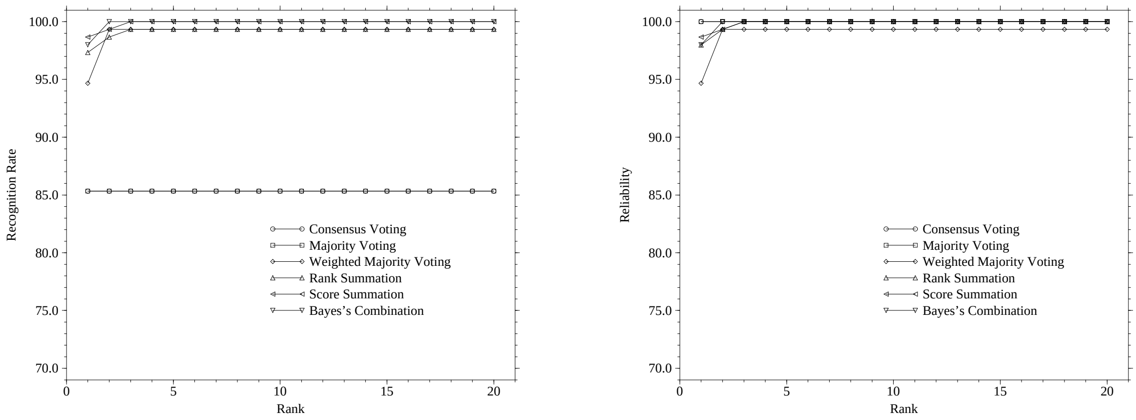


Figure 10: Recognition rates (left) and reliability (right) for the combination HMM and eigenface classifier

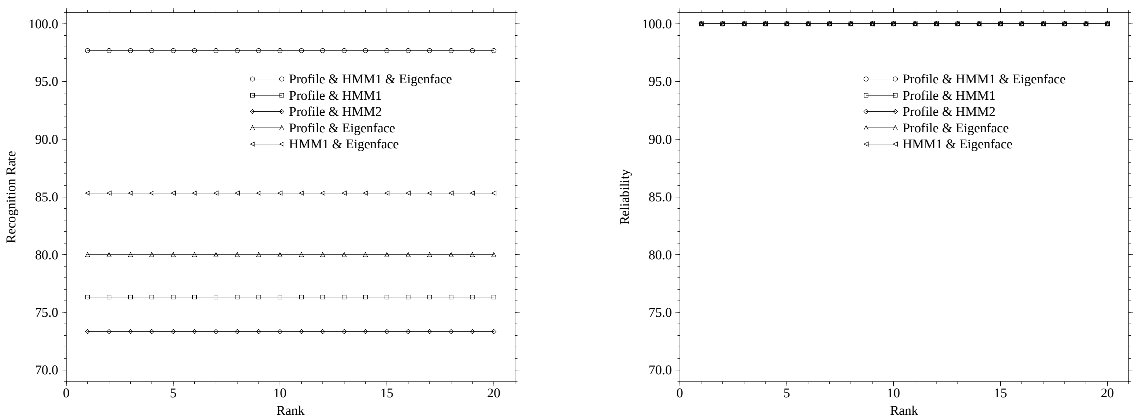


Figure 11: Recognition rates (left) and reliability (right) for consensus voting

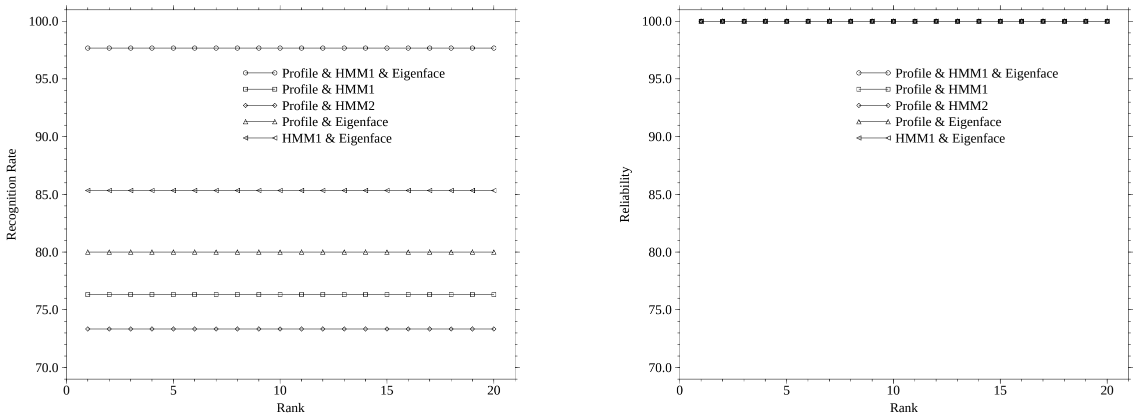


Figure 12: Recognition rates (left) and reliability (right) for majority voting

If we combine all three classifiers, the voting strategies yield very good results (97.667%). If only two classifiers are combined the recognition rate drops significantly (around 75% in the worst cases) and there is a high rejection rate (e.g., 71 cases for the combination of profile and HMM classifier with HMM1). In fact this combiner yields always a reject for the combination of two classifiers if the top decisions of the input classifiers are not identical. Let us state, however, the high reliability (100.0% in all of our cases) of such combination schemes, which is also true when even more classifiers are involved in the fusion process. To yield a misclassification several classifiers must make the same mistake, which is most of the time not true (unless one combines redundant classifiers, which is not recommended anyway). We observe a strong relationship between the reliability and the recognition rate. In order to realize a higher reliability there are more rejections needed, which results in a lower recognition rate. This fact is already well known for single classifiers, and the same is obviously true for combination classifiers.

3.8 Weighted Majority Voting

The weight for the voting is set to 0.85 for the profile classifier, to 0.90 for the HMM classifier and to 0.95 for the eigenface classifier, in accordance to the performance of each classifier. The results are shown in Figure 13.

Unlike unweighted majority voting there are much less rejected cases. And the results are slightly better than the recognition rates of the individual classifiers. The recognition rate on the first rank is similar to the rate of pattern classifiers involved. On the ranks below, instead, the recognition rate is slightly better than for the single pattern classifiers. Especially the combiner reaches its optimal performance faster.

3.9 Rank Summation

We expected the results of rank based strategies to be better than those of the voting strategies, since there is more information about the classifier decision available. The practical experiments proved this assumption to be true under certain circumstances. A graphical representation of the results is shown in Figure 14. We used the logarithmic transformation (the rank is interpreted as score value in this case; see Section 2.5) and no reduction criterion for these test runs. Other transformations yielded worse results.

An improvement of the recognition rate may be achieved by applying weights to each classifier.

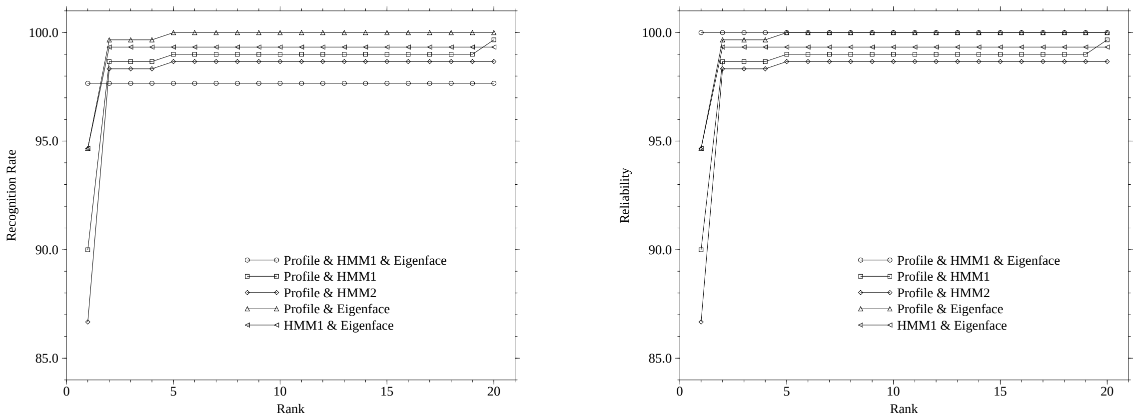


Figure 13: Recognition rates (left) and reliability (right) for weighted majority voting

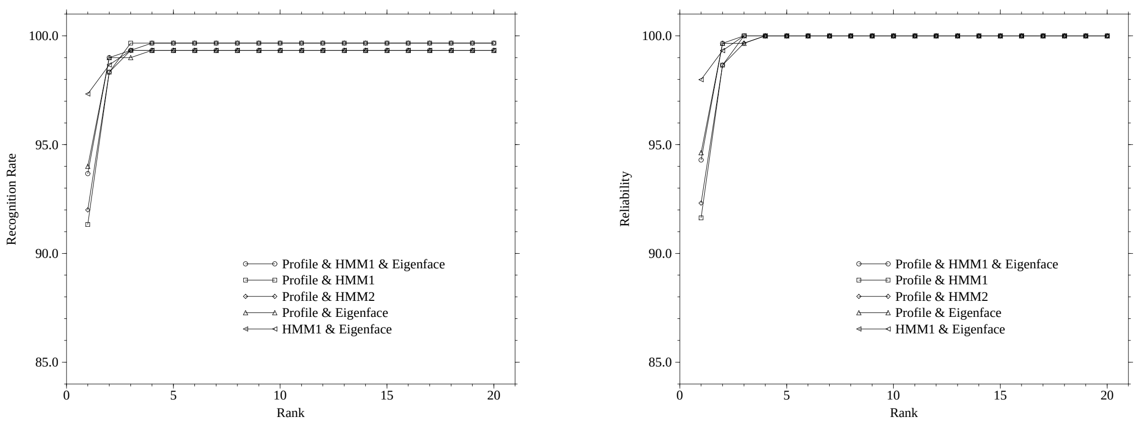


Figure 14: Recognition rates (left) and reliability (right) for rank summation

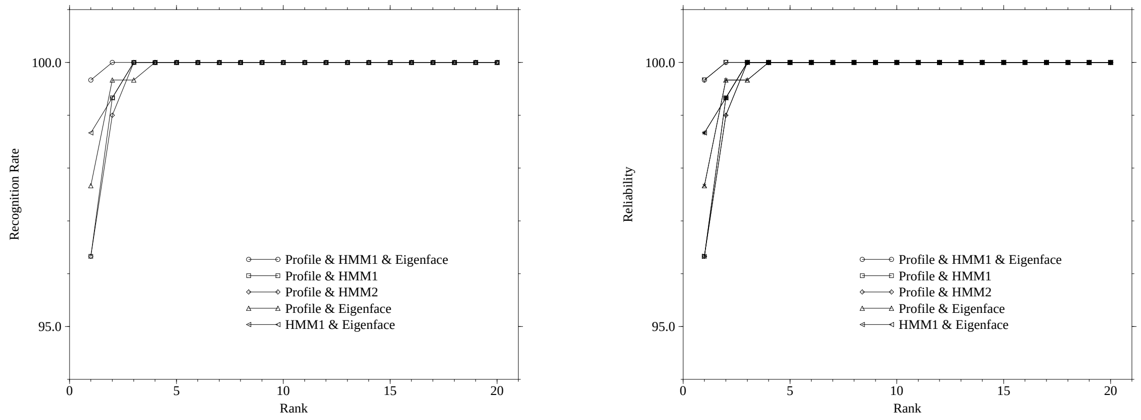


Figure 15: Recognition rates (left) and reliability (right) for score summation, logarithmic transformation

By reducing the influence of the profile classifier (e.g., 0.5 for the profile classifier and 1.00 for the full face classifier) it is possible to rise the recognition rate up to 94.0%.

Since there are nearly no rejections with the rank summation, the reliability is only slightly higher than the recognition rate.

3.10 Score Summation

Score based strategies yielded even better results than rank based methods, what was expected based on theoretical considerations. Of course, a transformation of the score is required. We got the best results for the logarithmic and the logistic transformation. The results are presented graphically in Figures 15 and 16. The differences between logarithmic and logistic transformation are rather small. Both behave very good in the top rankings. All the correct classifications are within the uppermost five ranks. And both have very high recognition rates, too (always more than 95.0%). The method with the logarithmic transformation behaves slightly better in concentrating the classifications in the uppermost ranks, whereas the logistic transformation yields the higher recognition rate.

The score summation yields only very seldom a rejection. In our test set we have no rejection at all. Therefore, the reliability is identical with the recognition rate.

3.11 Bayes's Combination Rule

The critical point of Bayes's combination rule is, as we already stated, the determination of the probabilities $P(H_k|F_j)$. Optimally there is a large training set used to estimate p by the relative frequency of correct classifications. In our case we estimated the probabilities based on the recognition results of each classifier. We applied the combiner with this parameterization to the test sets. Unfortunately, the probabilities p could only be roughly estimated due to the small set of images (especially in the case of the profile classifier). To investigate the importance of a fine determination of these probabilities, we ran extensive tests where we changed these p in a wide range (± 0.1). We found out that a coarse estimation of the p is sufficient for the combiner to yield good results. Of course, it is better to have completely separated test and training sets, but our data base is at the moment too small to do this. One method often used in such contexts is the leave-one-out method, but since we have only two images for the profile classifier the variance of p gets to large.

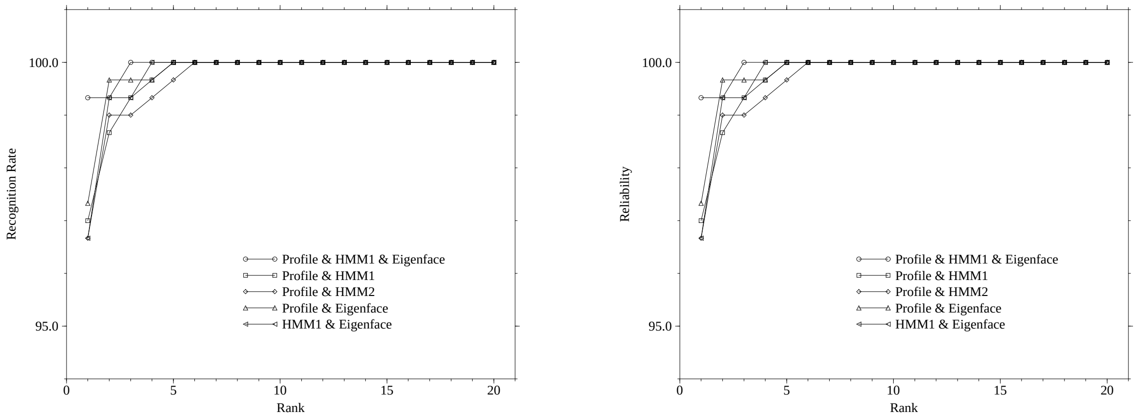


Figure 16: Recognition rates (left) and reliability (right) for score summation, logistic transformation

The results are listed in Figure 17. Obviously the combiner performs very well. It yields a high recognition rate and all correct classes are found within the first five ranks.

We ran also test with a more conservative choice of the probabilities p. In order to express a higher uncertainty for correct classifications we reduced p by 0.1 for every class. The performance of the combiner dropped slightly in this case.

Similar to the score summation, the combination with Bayes's rule produces nearly no rejections. So the recognition rate and the reliability are identical.

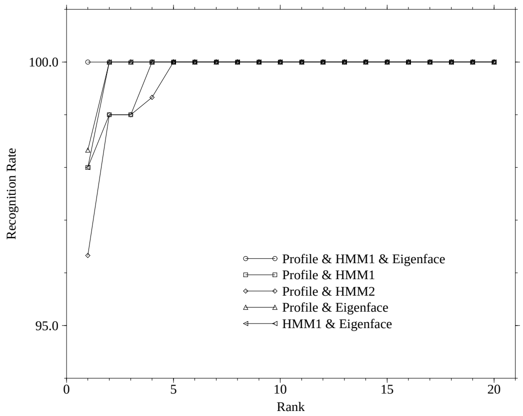
3.12 Method of Yu

The parameters for this method are exactly set as it is described in Section 2.5. This setup yielded very good results for both the combination with the HMM and with the eigenface classifier which are shown graphically in Figure 18. Though this combiner has a very high recognition rate it never reaches 100% due to its rejection criterion.

There are, however, some significant disadvantages. The combiner can not combine more than two classifiers and is mostly based on the profile classifier. The combination size and the weight of the pattern classifiers are determined using the score function of the profile classifier. Informations of the full face classifier instead are not taken into account. There is a lack of flexibility and universality for a broad applicability of this method. Though this combiner yields very good results for our case, we doubt that it may be used for other problems.

3.13 Score Transformation

It is very important to find smart transformations for the score, in order to combine classifiers with different types of score functions. The pattern classifiers used in this work, for instance, yield various scores. The profile classifier computes the score as the distance (in pixels) between the presented profile line and the intervals in the models (for further details refer to our previous papers [21] and [20]). Theoretically this function may have values between 0 and ∞. Experimentally we found values in the range of 0 and 30000. The HMM classifier instead yields a score $s = -2 \log p$, which is derived from a probability. The value may theoretically range from 0 to ∞ again. But in practice we found only values between 300.0 and 1000.0. The formula makes also clear that the score is based on a logarithmic scale. Again another measurement is used by the eigenface classifier. Its score is a Euclidean distance in a vector space which ranges from about



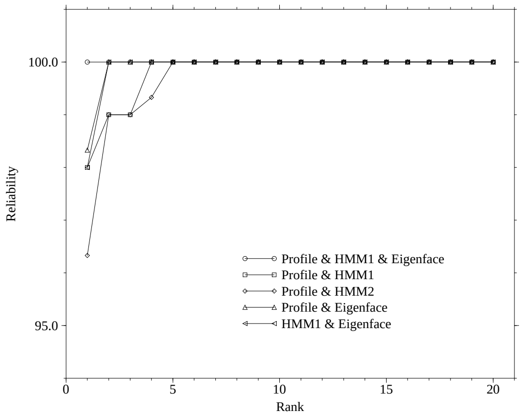
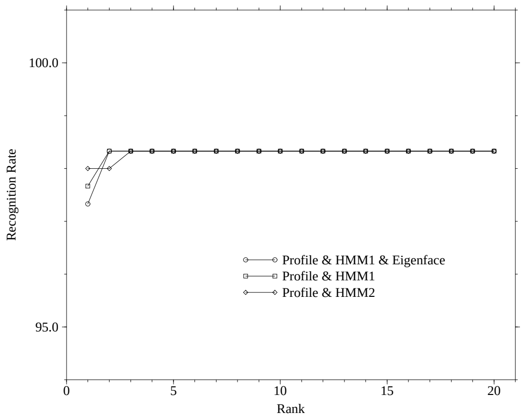


Figure 17: Recognition rates (left) and reliability (right) for Bayes's combination rule



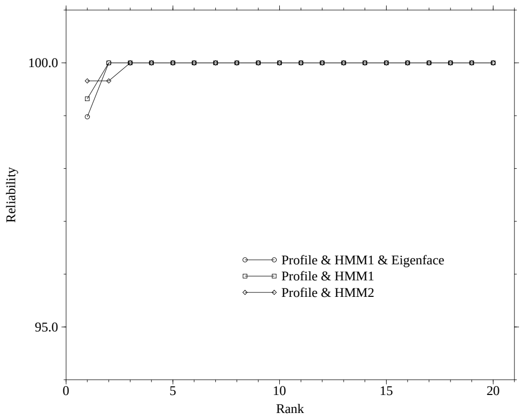


Figure 18: Recognition rates (left) and reliability (right) for the combiner of Yu

900.0 to 30000.0. Theoretically it may also yield values between 0 and ∞. Note, however, that not only the value of the score function but also the distribution of the values may be very different. The distributions of the score values for the profile and the HMM classifier are shown in Figures 19 and 20, respectively. It is evident that the distribution of the values is completely different (Note also that the scaling of the axes differs in the graphs).

An examination showed that all three classifiers we used do not behave ideally concerning the discrimination ability. In Figures 22 to 27 the distances between the first and the second rank of our test cases are shown. In fact, if the gap is very large, the decision of the classifier is very reliable. But there are also wrong decisions where the gap is large, and furthermore there are even correct decisions with a small distance between the score of the first and the second rank. As a tendency, however, our supposition is certainly right, but we have to take into account that there is a certain "zone of doubt". Another indicator for the quality of the classifier's decision might be the score of the first rank. A very low score means a very good decision. In Figures 28 to 33 the values of the first ranks of our pattern classifiers are shown. It is evident that for the profile classifier the score of the top decision might serve as a quality index, but for the HMM classifier and the eigenface classifier this is not true.

To test the consequences of the transformation applied to the score functions, we established the following test scenario: The combination method was set to score summation. Then the combiner was applied to the test data set with all five transformation methods (no transformation, linear, logarithmic, exponential and logistic transformation) in order to combine the profile and the HMM classifier. The recognition rates are listed in Figures 79, 80, 81, 82 and 83; a graphical representation is given in Figure 34.

It is not astonishing that the experiment without transformation yields the worst results. As we already pointed out, the score functions of our pattern classifiers are too different to be combined without modification. They have a different range and lie in another interval. It is remarkable, however, that the performance is always higher than or equal to the rate of the profile classifier.

The results of the other experiments are very good: On the first rank all methods accomplish recognition rates higher than 95.0%. Clearly the best method is the logistic transformation, because it yields very good reordering results. The correct class is always among the first three ranks.

In another experiment we investigated whether the recognition rate is improved if we apply different transformations for every pattern classifier (like, e.g., linear transformation for the profile classifier and logarithmic for the HMM classifier). We tried all the possible combinations with the profile and the HMM classifier. We found that there is no significant improvement of the recognition in this case. Note, however, that this statement is only true for these classifiers and that this might change if classifiers with very different properties are involved in the fusion process.

Another point to mention is the impact of transformations applied to rank based combination strategies, when the rank is interpreted as a score value. The choice of an appropriate transformation leads to an improved recognition rate.

3.14 Set Reduction

Similarly to the score transformation, we tested the impact of set reduction on the recognition rate. To keep the experiment simple we only examined the HMM and the profile classifier. Several tests were ran, where the reduction criterion was a limit of 5, 10, 15, 20 or 25 classes per pattern classifier involved in the fusion process. The transformation scheme is logarithmic, and the combination method is again score summation, because we obtained good results for this test scenario. For the first test set we used the intersection operation to build the combination set, for the second we applied the union operation. The results are graphically shown in Figures 35 and 36, and in Figures 84 to 93 the detailed results are given.

We observe that the test runs with intersection yielded generally better results for the first ranks unless the classifier limit was chosen rather small (5 classes per classifier). Another advantage of the intersection method is that the classifications tend to be ranked very high if the correct class

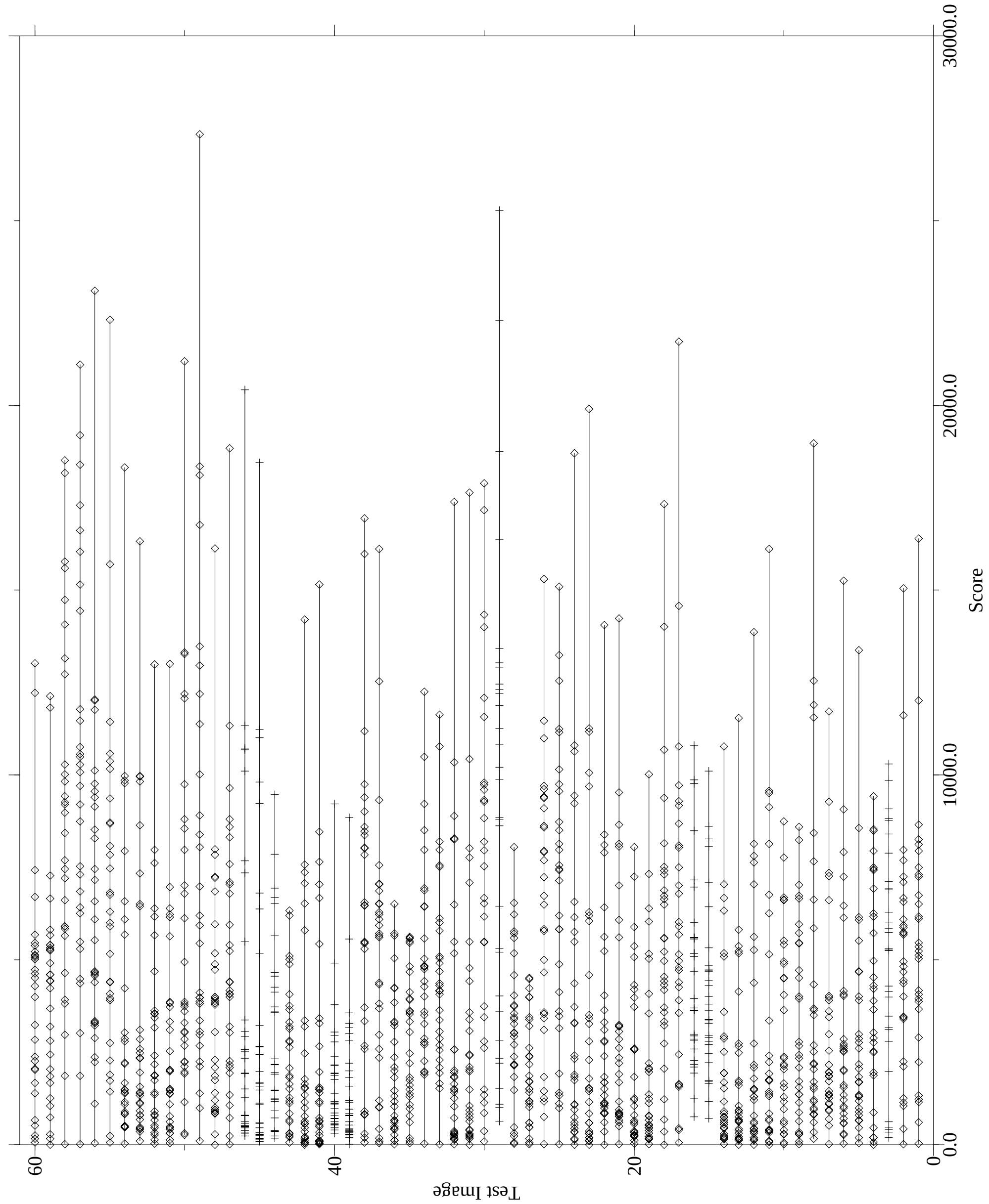


Figure 19: Distribution of the score of the profile classifier (correct classifications in the lower, misclassifications in the upper part)

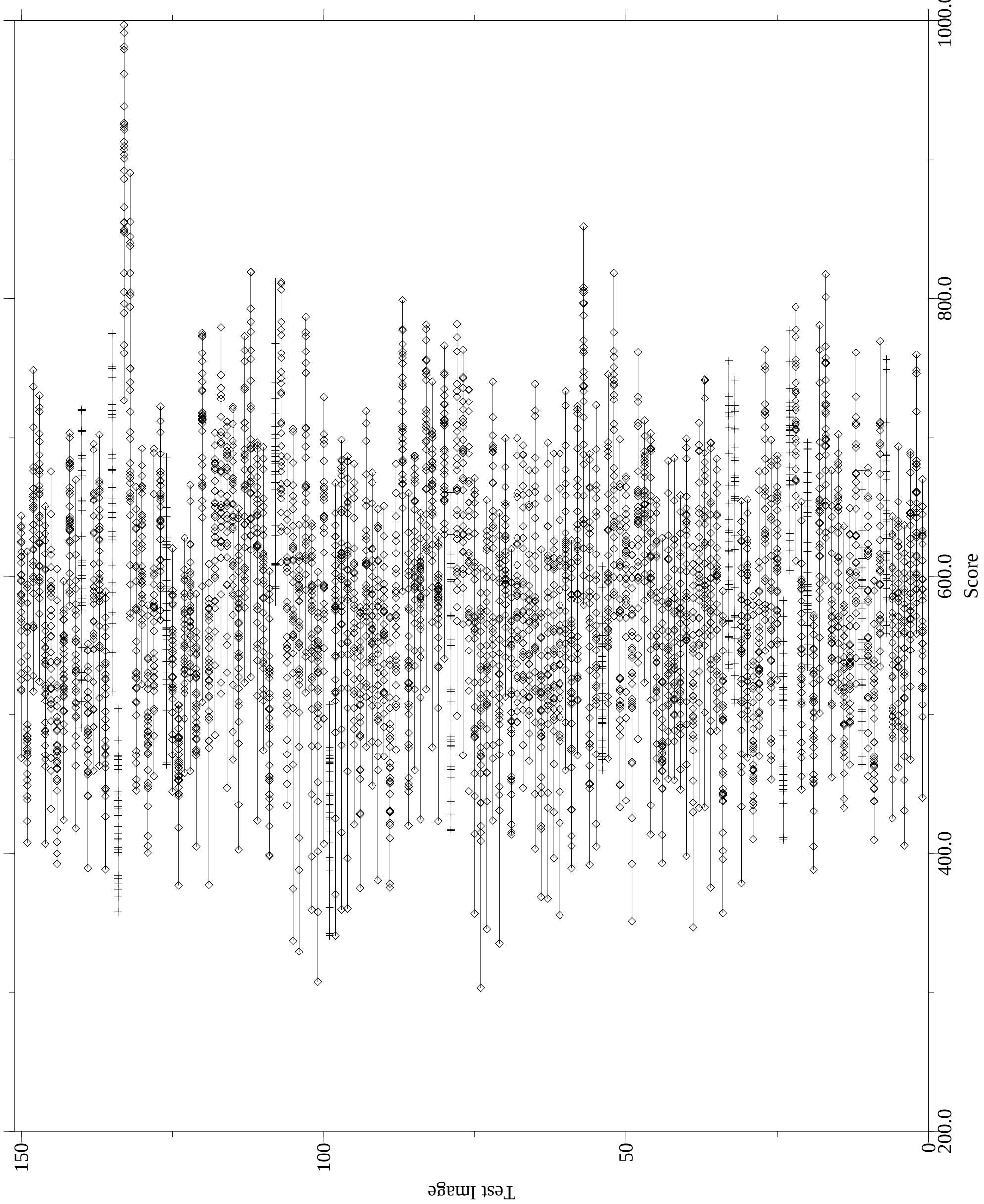


Figure 20: Distribution of the score of the full face HMM classifier with HMM1 (correct classifications in the lower, misclassifications in the upper part)

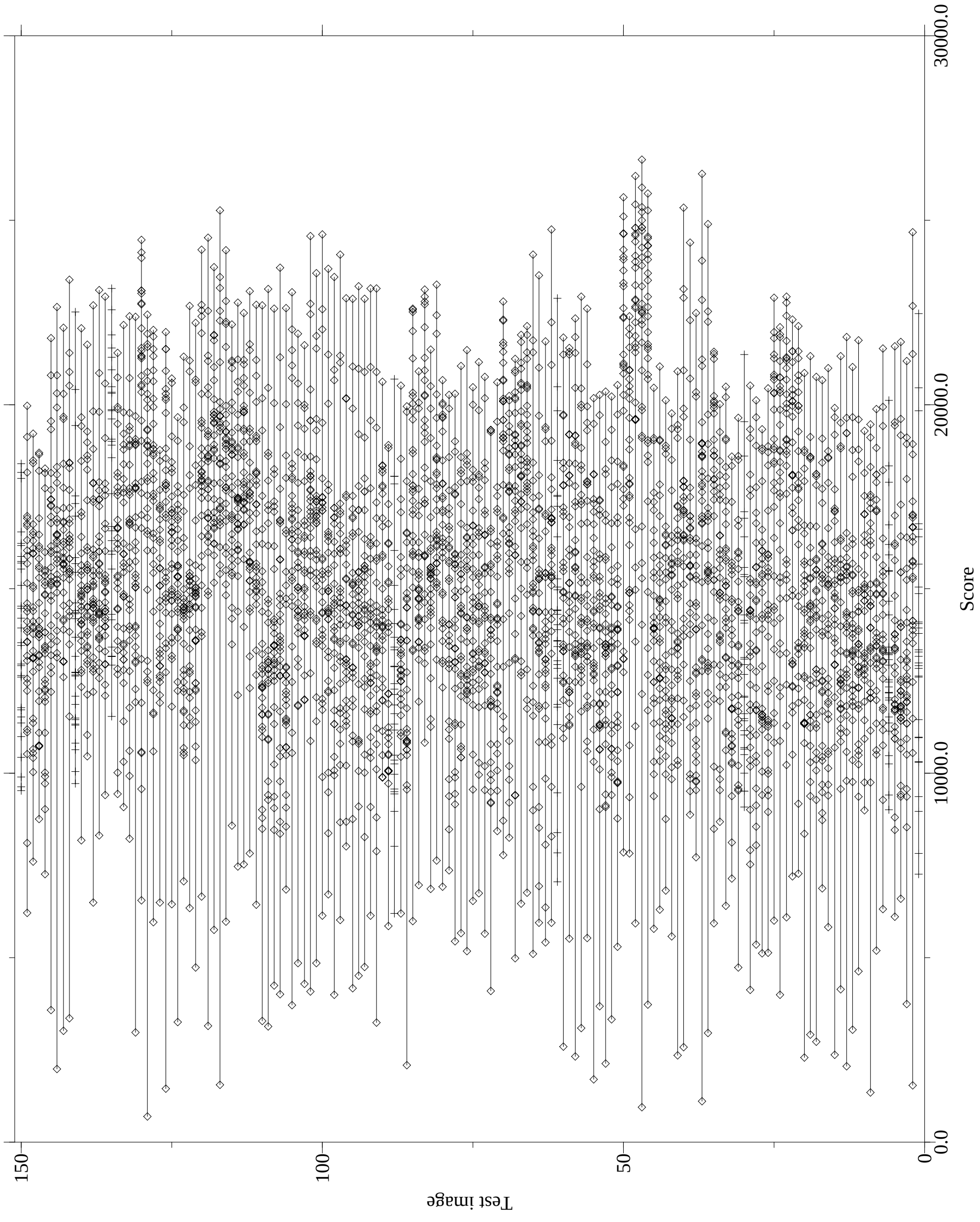


Figure 21: Distribution of the score of the full face eigenface classifier with HMM1 (correct classifications in the lower, misclassifications in the upper part)

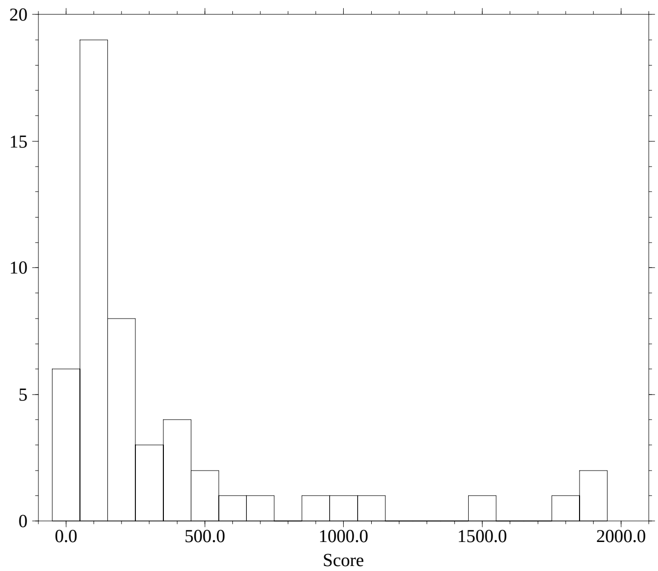


Figure 22: Distance between the first and the second rank of the profile classifier (correct classifications)

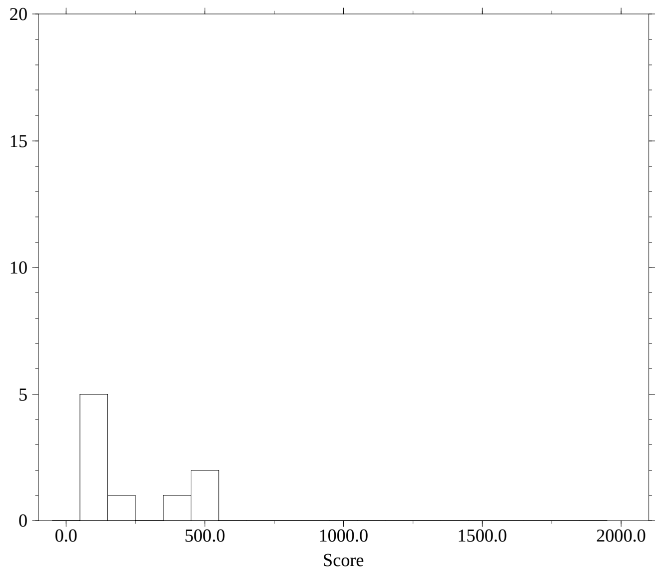


Figure 23: Distance between the first and the second rank of the profile classifier (misclassifications)

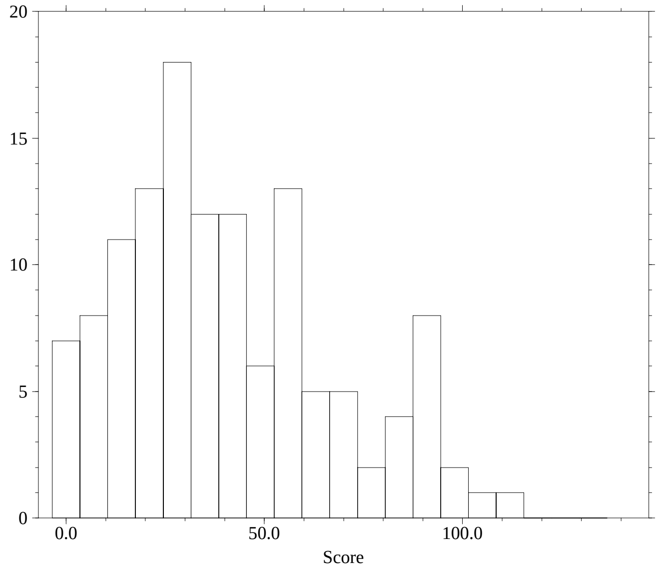


Figure 24: Distance between the first and the second rank of the full face HMM classifier (correct classifications)

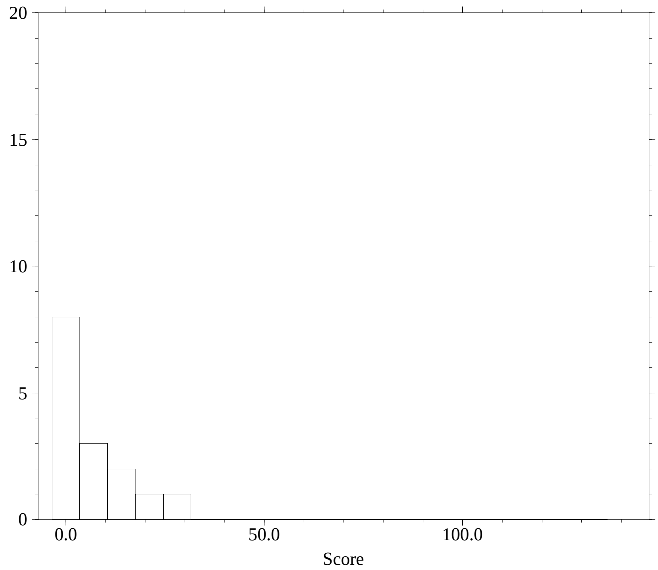


Figure 25: Distance between the first and the second rank of the full face HMM classifier (misclassifications)

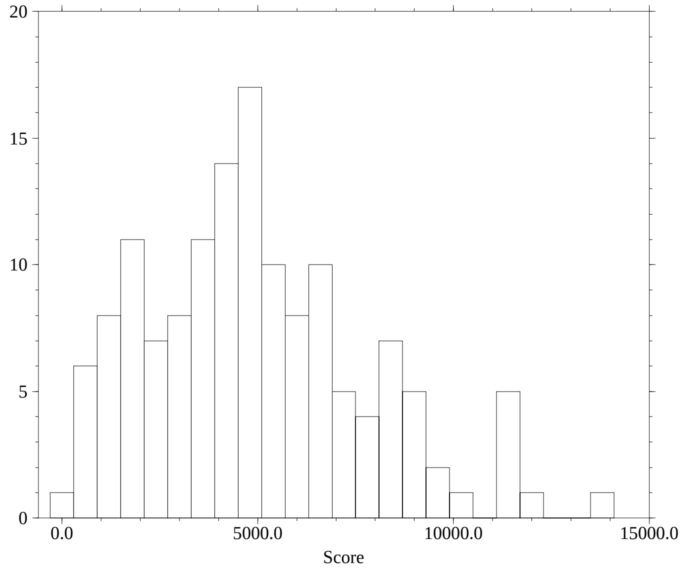


Figure 26: Distance between the first and the second rank of the full face eigenface classifier (correct classifications)

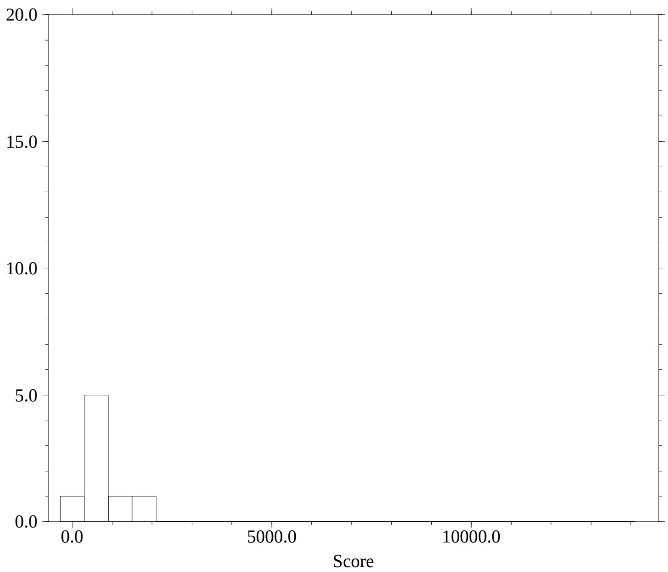


Figure 27: Distance between the first and the second rank of the full face eigenface classifier (misclassifications)

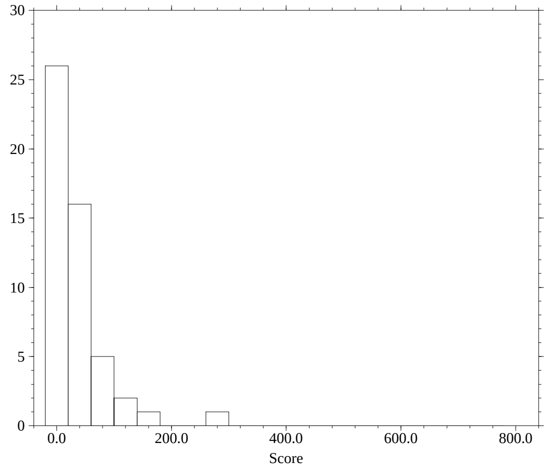


Figure 28: Score of the first rank of the profile classifier (correct classifications)

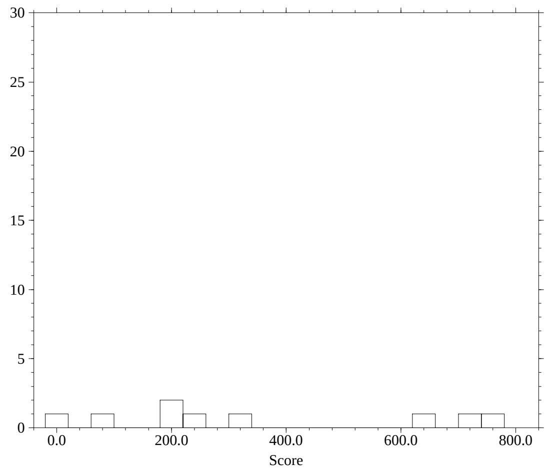


Figure 29: Score of the first rank of the profile classifier (misclassifications)

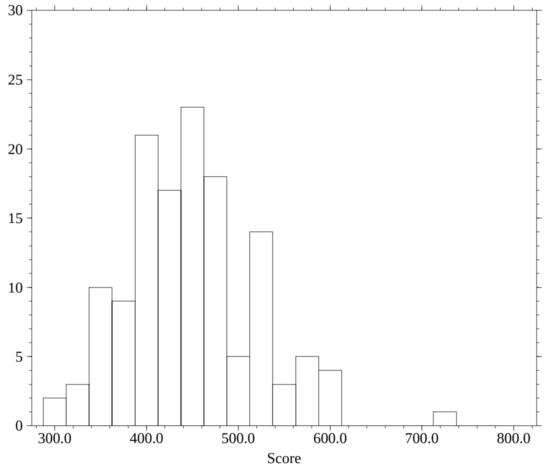


Figure 30: Score of the first rank of the full face HMM classifier (correct classifications)

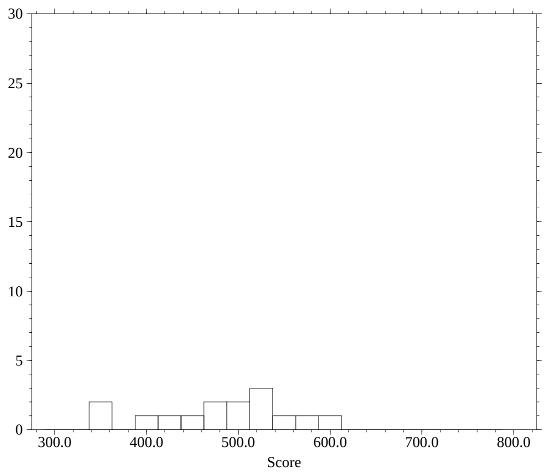


Figure 31: Score of the first rank of the full face HMM classifier (misclassifications)

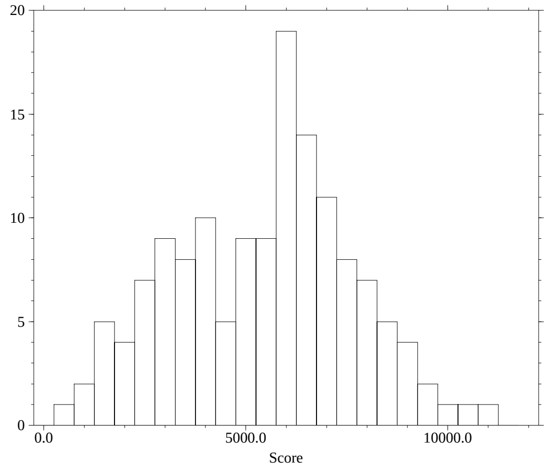


Figure 32: Score of the first rank of the full face eigenface classifier (correct classifications)

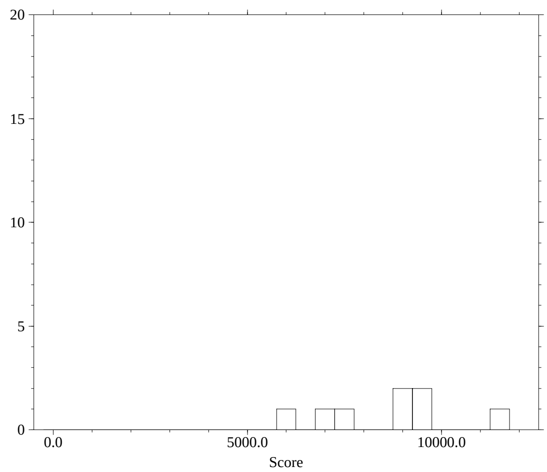


Figure 33: Score of the first rank of the full face eigenface classifier (misclassifications)

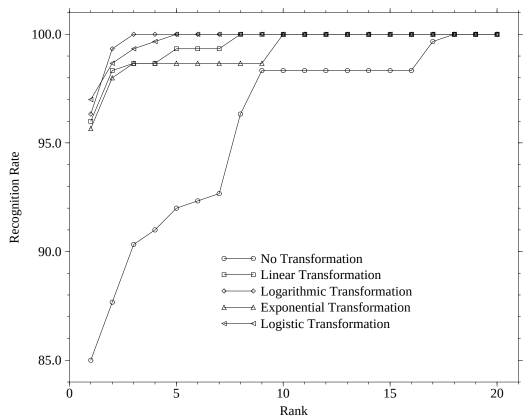


Figure 34: Recognition rates with different transformations

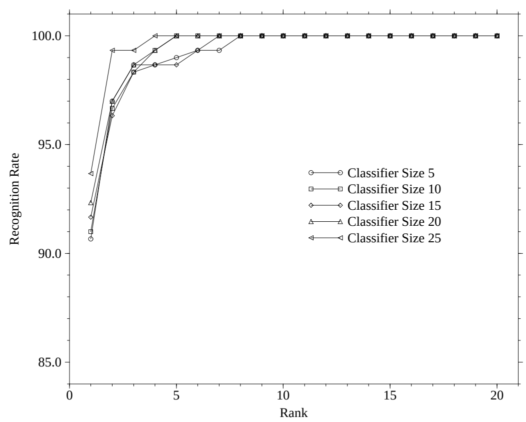


Figure 35: Recognition rates with union as reduction operation

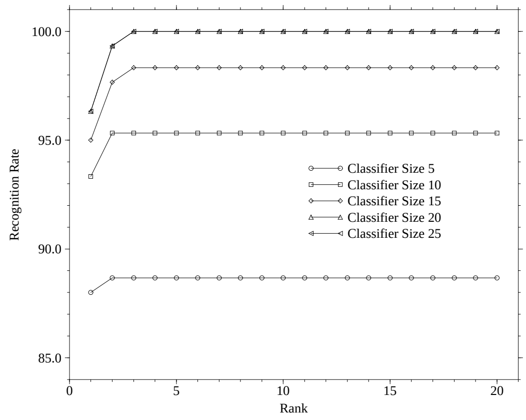


Figure 36: Recognition rates with intersection as reduction operation

is in the combination set. In this case the correct ones are mostly classified within the first three ranks.

It is a disadvantage, however, that correct classes may "disappear" in the fusion process with the intersection scheme. Classes ranked very low in one of the classifier rankings won't be integrated in the combination set. Therefore, the combination will not yield a correct classification. This is also the reason why reduction methods with intersections do not converge to a recognition rate of 100%. Combinations with the union scheme, instead, always reach a recognition rate of 100% in the lower ranks. It is somehow comparable to search methods. The combination with the intersection scheme is similar to a depth-first search, whereas the combination with the union scheme corresponds rather to a breadth-first search method.

We ran some test series with another reduction criterion (thresholds for the score of 20, 30, 40, 50, 60, 70 and 80), too. We found that the same is true for this reduction scheme as for the reduction based on ranks. Additionally we stated a higher dependency on the transformation used. In this sense reduction based on ranks seems to be more robust.

4 Conclusions

A comparison of the results of all combining strategies tested in our work shows that the performance of the combined classifiers exceeds those of the single classifiers. Even very simple combination methods like, e.g., weighted voting strategies yield significantly better results. Therefore, we state that decision fusion is a very useful tool to improve recognition and reliability, not exclusively for face recognition or optical character recognition.

The quality of combination, however, is certainly dependent on a multitude of properties of the single classifiers involved in the process. We state that an in-depth analysis of these classifiers is required in order to be able to design a useful combination scheme.

More classifiers obviously have an impact on the recognition rate. We got the best results for the combination of all three classifiers. And though the results of the eigenface and the HMM classifier were already very good we could realize a significant improvement of the recognition rate by applying combination methods.

We got the best results with the method of Bayes and the score summation for our pattern classifiers. Since the method of Bayes depends heavily on the probability matrix, we state that there should be available a training set with a reasonable size that allows to reliably estimate the required probabilities.

Generally we found that the improvement of the recognition rate is higher for the second test set (HMM2). It is a trivial statement that it is more difficult to improve good results than mediocre ones. Obviously this is also true for combination schemes.

In the near future we want to include more pattern classifiers into the fusion process with the methods presented in this document. With an increasing number of classifiers problems like redundancy and weighting of the single classifier will become more important. At the moment, we only have classifiers for frontal and profile images. We want to extend our set of classifiers in order to be able to work on other input data (like, e.g., range images). But we also want to have more pattern classifiers for each type of input data (e.g., two classifiers for profile images). With an increasing number of pattern classifiers we don't get just a broader fusion process, but we will also be able to investigate alternative architectures for decision fusion (e.g. processing of the information in several stages where the first stage combines the results of the pattern classifiers for each kind of input data and the second stage makes a combination of these intermediate results).

References

[1] J. Cao, M. Shridhar, and M. Ahmadi. Fusion of Classifiers with Fuzzy Integrals. In *Proceedings 3d International Conference on Document Analysis and Recognition, Montréal, Canada*, volume 1, pages 108–111. IEEE Computer Society Press, August 1995.

[2] B.V. Dasarathy. *Decision Fusion.* IEEE Computer Society Press, 1994.

[3] K. Fukunaga. *Introduction to Statistical Pattern Recognition.* Academic Press, 2 edition, 1990.

[4] A. Garzotto. *Vollautomatische Erkennung von Schriftzeichen in gedrucktem Schriftgut.* PhD thesis, Universität Zürich, 1994.

[5] D.L. Hall. *Mathematical Techniques in Multi-Sensor Data Fusion.* Artech House, Inc., 1992.

[6] T.K. Ho. *A Theory of Multiple Classifier Systems and Its Application to Visual Word Recognition.* PhD thesis, Departement of Computer Science, State University of New York at Buffalo, May 1992.

[7] T.K. Ho, J.J. Hull, and S.N. Srihari. Decision Combination in Multiple Classifier Systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(1):66–75, January 1994.

[8] Y.S. Huang and C.Y. Suen. A Method of Combining Multiple Classifiers – A Neural Network Approach. In *Proceedings 12th IAPR International Conference on Pattern Recognition (ICPR), Jerusalem, Israel*, volume 2, pages 473–475. IEEE Computer Society Press, October 1994.

[9] Y. Ito, T. Ohashi, and T. Ejima. Considerations on Designing a Decision-Tree with Multiple Features. In *Proceedings 4th International Workshop on Frontiers in Handwriting Recognition (IWFHR), Taipei, Taiwan*, pages 362–369, December 1994.

[10] L. Lam and C.Y. Suen. A Theoretical Analysis of the Application of Majority Voting to Pattern Recognition. In *Proceedings 12th IAPR International Conference on Pattern Recognition (ICPR), Jerusalem, Israel*, volume 2, pages 418–420. IEEE Computer Society Press, October 1994.

[11] L. Lam and C.Y. Suen. Increasing Experts for Majority Vote in OCR: Theoretical Considerations and Strategies. In *Proceedings 4th International Workshop on Frontiers in Handwriting Recognition*, pages 245–254, 1994.

[12] L. Lam and C.Y. Suen. Optimal Combinations of Pattern Classifiers. *Pattern Recognition Letters*, 16:945–954, 1995.

[13] D.-S. Lee and S.N. Srihari. A Theory of Classifier Combination: The Neural Network Approach. In *Proceedings 3d International Conference on Document Analysis and Recognition. Montréal, Canada*, volume 1, pages 42–45. IEEE Computer Society Press, August 1995.

[14] E. Mandler and J. Schürmann. Combining the Classification Results of Independant Classifiers Based on the Dempster-Shafer Theory of Evidence. In E.S. Gelsema and L.N. Kanal, editors, *Pattern Recognition and Artificial Intelligence*, pages 381–393. Elsevier, 1988.

[15] C. Nyffenegger. Gesichtserkennung mit Hidden-Markov-Modellen. Master's thesis, Institut für Informatik und angewandte Mathematik, Universität Bern, September 1995.

[16] L.R. Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, 77(2):257–286, February 1989.

[17] F.S. Samaria. *Face Recognition Using Hidden Markov Models.* PhD thesis, Trinity College, October 1994.

[18] M.A. Turk and A.P. Pentland. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[19] E. Waltz and J. Llinas. *Multisensor Data Fusion.* Artech House, Inc., 1990.

[20] K. Yu, B. Achermann, C. Nyffenegger, X.-Y. Jiang, H. Bunke, and E.G. Schukat-Talamazzini. Kombination von Frontal- und Profilanalyse menschlicher Gesichter. In G. Sagerer, S. Posch, and F. Kummert, editors, *Mustererkennung 1995. Verstehen akustischer und visueller Information. 17. DAGM-Symposium, Bielefeld*, pages 327–334. Springer-Verlag, September 1995.

[21] K. Yu, X.-Y. Jiang, and H. Bunke. Face Recognition by Facial Profile Analysis. In M. Bichsel, editor, *Proceedings International Workshop on Automatic Face and Gesture Recognition, Zürich*, pages 208–213, June 1995.

A Result Tables

A.1 Pattern Classifiers

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	135	(90.000%)	135	(90.000%)	90.000%
2	9	(6.0000%)	144	(96.000%)	96.000%
3	3	(2.0000%)	147	(98.000%)	98.000%
4	1	(0.66667%)	148	(98.667%)	98.667%
5	0	(0.0000%)	148	(98.667%)	98.667%
6	0	(0.0000%)	148	(98.667%)	98.667%
7	0	(0.0000%)	148	(98.667%)	98.667%
8	0	(0.0000%)	148	(98.667%)	98.667%
9	0	(0.0000%)	148	(98.667%)	98.667%
10	0	(0.0000%)	148	(98.667%)	98.667%
11	1	(0.66667%)	149	(99.333%)	99.333%
12	1	(0.66667%)	150	(100.00%)	100.00%
13	0	(0.0000%)	150	(100.00%)	100.00%
14	0	(0.0000%)	150	(100.00%)	100.00%
15	0	(0.0000%)	150	(100.00%)	100.00%
16	0	(0.0000%)	150	(100.00%)	100.00%
17	0	(0.0000%)	150	(100.00%)	100.00%
18	0	(0.0000%)	150	(100.00%)	100.00%
19	0	(0.0000%)	150	(100.00%)	100.00%
20	0	(0.0000%)	150	(100.00%)	100.00%

Total of test cases:	150
Rejected cases:	0

Figure 37: Results full face image HMM classifier

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	130	(86.667%)	130	(86.667%)	86.667%
2	11	(7.3333%)	141	(94.000%)	94.000%
3	5	(3.3333%)	146	(97.333%)	97.333%
4	2	(1.3333%)	148	(98.667%)	98.667%
5	0	(0.0000%)	148	(98.667%)	98.667%
6	0	(0.0000%)	148	(98.667%)	98.667%
7	0	(0.0000%)	148	(98.667%)	98.667%
8	0	(0.0000%)	148	(98.667%)	98.667%
9	0	(0.0000%)	148	(98.667%)	98.667%
10	0	(0.0000%)	148	(98.667%)	98.667%
11	1	(0.66667%)	149	(99.333%)	99.333%
12	1	(0.66667%)	150	(100.00%)	100.00%
13	0	(0.0000%)	150	(100.00%)	100.00%
14	0	(0.0000%)	150	(100.00%)	100.00%
15	0	(0.0000%)	150	(100.00%)	100.00%
16	0	(0.0000%)	150	(100.00%)	100.00%
17	0	(0.0000%)	150	(100.00%)	100.00%
18	0	(0.0000%)	150	(100.00%)	100.00%
19	0	(0.0000%)	150	(100.00%)	100.00%
20	0	(0.0000%)	150	(100.00%)	100.00%

Total of test cases:	150
Rejected cases:	0

Figure 38: Results full face image HMM classifier (less trained)

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	142	(94.667%)	142	(94.667%)	94.667%
2	2	(1.3333%)	144	(96.000%)	96.000%
3	3	(2.0000%)	147	(98.000%)	98.000%
4	1	(0.66667%)	148	(98.667%)	98.667%
5	0	(0.0000%)	148	(98.667%)	98.667%
6	2	(1.3333%)	150	(100.00%)	100.00%
7	0	(0.0000%)	150	(100.00%)	100.00%
8	0	(0.0000%)	150	(100.00%)	100.00%
9	0	(0.0000%)	150	(100.00%)	100.00%
10	0	(0.0000%)	150	(100.00%)	100.00%
11	0	(0.0000%)	150	(100.00%)	100.00%
12	0	(0.0000%)	150	(100.00%)	100.00%
13	0	(0.0000%)	150	(100.00%)	100.00%
14	0	(0.0000%)	150	(100.00%)	100.00%
15	0	(0.0000%)	150	(100.00%)	100.00%
16	0	(0.0000%)	150	(100.00%)	100.00%
17	0	(0.0000%)	150	(100.00%)	100.00%
18	0	(0.0000%)	150	(100.00%)	100.00%
19	0	(0.0000%)	150	(100.00%)	100.00%
20	0	(0.0000%)	150	(100.00%)	100.00%

Total of test cases:	150
Rejected cases:	0

Figure 39: Results full face image eigenface classifier

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	51	(85.000%)	51	(85.000%)	85.000%
2	0	(0.0000%)	51	(85.000%)	85.000%
3	1	(1.6667%)	52	(86.667%)	86.667%
4	1	(1.6667%)	53	(88.333%)	88.333%
5	1	(1.6667%)	54	(90.000%)	90.000%
6	0	(0.0000%)	54	(90.000%)	90.000%
7	0	(0.0000%)	54	(90.000%)	90.000%
8	0	(0.0000%)	54	(90.000%)	90.000%
9	1	(1.6667%)	55	(91.667%)	91.667%
10	3	(5.0000%)	58	(96.667%)	96.667%
11	1	(1.6667%)	59	(98.333%)	98.333%
12	0	(0.0000%)	59	(98.333%)	98.333%
13	0	(0.0000%)	59	(98.333%)	98.333%
14	0	(0.0000%)	59	(98.333%)	98.333%
15	0	(0.0000%)	59	(98.333%)	98.333%
16	0	(0.0000%)	59	(98.333%)	98.333%
17	0	(0.0000%)	59	(98.333%)	98.333%
18	1	(1.6667%)	60	(100.00%)	100.00%
19	0	(0.0000%)	60	(100.00%)	100.00%
20	0	(0.0000%)	60	(100.00%)	100.00%

Total of test cases:	60
Rejected cases:	0

Figure 40: Results profile image classifier

A.2 Combination Profile, HMM and Eigenface Classifier

Rank	**Hits**	(per rank)	**Hits**	(accum.)	Reliability
1	293	(97.667%)	293	(97.667%)	100.00%

Total of test cases:	300
Rejected cases:	7

Figure 41: Results consensus voting (HMM1)

Rank	**Hits**	(per rank)	**Hits**	(accum.)	Reliability
1	293	(97.667%)	293	(97.667%)	100.00%

Total of test cases:	300
Rejected cases:	7

Figure 42: Results majority voting (HMM1)

Rank	**Hits**	(per rank)	**Hits**	(accum.)	Reliability
1	293	(97.667%)	293	(97.667%)	100.00%
2	0	(0.0000%)	293	(97.667%)	100.00%
3	0	(0.0000%)	293	(97.667%)	100.00%
4	0	(0.0000%)	293	(97.667%)	100.00%
5	0	(0.0000%)	293	(97.667%)	100.00%
6	0	(0.0000%)	293	(97.667%)	100.00%
7	0	(0.0000%)	293	(97.667%)	100.00%
8	0	(0.0000%)	293	(97.667%)	100.00%
9	0	(0.0000%)	293	(97.667%)	100.00%
10	0	(0.0000%)	293	(97.667%)	100.00%
11	0	(0.0000%)	293	(97.667%)	100.00%
12	0	(0.0000%)	293	(97.667%)	100.00%
13	0	(0.0000%)	293	(97.667%)	100.00%
14	0	(0.0000%)	293	(97.667%)	100.00%
15	0	(0.0000%)	293	(97.667%)	100.00%
16	0	(0.0000%)	293	(97.667%)	100.00%
17	0	(0.0000%)	293	(97.667%)	100.00%
18	0	(0.0000%)	293	(97.667%)	100.00%
19	0	(0.0000%)	293	(97.667%)	100.00%
20	0	(0.0000%)	293	(97.667%)	100.00%

Total of test cases:	300
Rejected cases:	7

Figure 43: Results weighted majority voting (HMM1)

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	281	(93.667%)	281	(93.667%)	94.295%
2	16	(5.3333%)	297	(99.000%)	99.664%
3	1	(0.33333%)	298	(99.333%)	100.00%
4	0	(0.0000%)	298	(99.333%)	100.00%
5	0	(0.0000%)	298	(99.333%)	100.00%
6	0	(0.0000%)	298	(99.333%)	100.00%
7	0	(0.0000%)	298	(99.333%)	100.00%
8	0	(0.0000%)	298	(99.333%)	100.00%
9	0	(0.0000%)	298	(99.333%)	100.00%
10	0	(0.0000%)	298	(99.333%)	100.00%
11	0	(0.0000%)	298	(99.333%)	100.00%
12	0	(0.0000%)	298	(99.333%)	100.00%
13	0	(0.0000%)	298	(99.333%)	100.00%
14	0	(0.0000%)	298	(99.333%)	100.00%
15	0	(0.0000%)	298	(99.333%)	100.00%
16	0	(0.0000%)	298	(99.333%)	100.00%
17	0	(0.0000%)	298	(99.333%)	100.00%
18	0	(0.0000%)	298	(99.333%)	100.00%
19	0	(0.0000%)	298	(99.333%)	100.00%
20	0	(0.0000%)	298	(99.333%)	100.00%

Total of test cases:	300
Rejected cases:	2

Figure 44: Results ranks summation (HMM1)

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	299	(99.667%)	299	(99.667%)	99.667%
2	1	(0.33333%)	300	(100.00%)	100.00%
3	0	(0.0000%)	300	(100.00%)	100.00%
4	0	(0.0000%)	300	(100.00%)	100.00%
5	0	(0.0000%)	300	(100.00%)	100.00%
6	0	(0.0000%)	300	(100.00%)	100.00%
7	0	(0.0000%)	300	(100.00%)	100.00%
8	0	(0.0000%)	300	(100.00%)	100.00%
9	0	(0.0000%)	300	(100.00%)	100.00%
10	0	(0.0000%)	300	(100.00%)	100.00%
11	0	(0.0000%)	300	(100.00%)	100.00%
12	0	(0.0000%)	300	(100.00%)	100.00%
13	0	(0.0000%)	300	(100.00%)	100.00%
14	0	(0.0000%)	300	(100.00%)	100.00%
15	0	(0.0000%)	300	(100.00%)	100.00%
16	0	(0.0000%)	300	(100.00%)	100.00%
17	0	(0.0000%)	300	(100.00%)	100.00%
18	0	(0.0000%)	300	(100.00%)	100.00%
19	0	(0.0000%)	300	(100.00%)	100.00%
20	0	(0.0000%)	300	(100.00%)	100.00%

Total of test cases:	300
Rejected cases:	0

Figure 45: Results score summation, logarithmic transformation (HMM1)

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	298	(99.333%)	298	(99.333%)	99.333%
2	0	(0.0000%)	298	(99.333%)	99.333%
3	2	(0.66667%)	300	(100.00%)	100.00%
4	0	(0.0000%)	300	(100.00%)	100.00%
5	0	(0.0000%)	300	(100.00%)	100.00%
6	0	(0.0000%)	300	(100.00%)	100.00%
7	0	(0.0000%)	300	(100.00%)	100.00%
8	0	(0.0000%)	300	(100.00%)	100.00%
9	0	(0.0000%)	300	(100.00%)	100.00%
10	0	(0.0000%)	300	(100.00%)	100.00%
11	0	(0.0000%)	300	(100.00%)	100.00%
12	0	(0.0000%)	300	(100.00%)	100.00%
13	0	(0.0000%)	300	(100.00%)	100.00%
14	0	(0.0000%)	300	(100.00%)	100.00%
15	0	(0.0000%)	300	(100.00%)	100.00%
16	0	(0.0000%)	300	(100.00%)	100.00%
17	0	(0.0000%)	300	(100.00%)	100.00%
18	0	(0.0000%)	300	(100.00%)	100.00%
19	0	(0.0000%)	300	(100.00%)	100.00%
20	0	(0.0000%)	300	(100.00%)	100.00%

Total of test cases: 300
Rejected cases: 0

Figure 46: Results score summation, logistic transformation (HMM1)

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	300	(100.00%)	300	(100.00%)	100.00%
2	0	(0.0000%)	300	(100.00%)	100.00%
3	0	(0.0000%)	300	(100.00%)	100.00%
4	0	(0.0000%)	300	(100.00%)	100.00%
5	0	(0.0000%)	300	(100.00%)	100.00%
6	0	(0.0000%)	300	(100.00%)	100.00%
7	0	(0.0000%)	300	(100.00%)	100.00%
8	0	(0.0000%)	300	(100.00%)	100.00%
9	0	(0.0000%)	300	(100.00%)	100.00%
10	0	(0.0000%)	300	(100.00%)	100.00%
11	0	(0.0000%)	300	(100.00%)	100.00%
12	0	(0.0000%)	300	(100.00%)	100.00%
13	0	(0.0000%)	300	(100.00%)	100.00%
14	0	(0.0000%)	300	(100.00%)	100.00%
15	0	(0.0000%)	300	(100.00%)	100.00%
16	0	(0.0000%)	300	(100.00%)	100.00%
17	0	(0.0000%)	300	(100.00%)	100.00%
18	0	(0.0000%)	300	(100.00%)	100.00%
19	0	(0.0000%)	300	(100.00%)	100.00%
20	0	(0.0000%)	300	(100.00%)	100.00%

Total of test cases: 300
Rejected cases: 0

Figure 47: Results Bayes's combination rule (HMM1)

A.3 Combination Profile and HMM Classifier

Rank	**Hits**	(per rank)	**Hits**	(accum.)	Reliability
1	229	(76.333%)	229	(76.333%)	100.00%

Total of test cases:	300
Rejected cases:	71

Figure 48: Results consensus voting (HMM1)

Rank	**Hits**	(per rank)	**Hits**	(accum.)	Reliability
1	220	(73.333%)	220	(73.333%)	100.00%

Total of test cases:	300
Rejected cases:	80

Figure 49: Results consensus voting (HMM2)

Rank	**Hits**	(per rank)	**Hits**	(accum.)	Reliability
1	229	(76.333%)	229	(76.333%)	100.00%

Total of test cases:	300
Rejected cases:	71

Figure 50: Results majority voting (HMM1)

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	220	(73.333%)	220	(73.333%)	100.00%

Total of test cases:	300
Rejected cases:	80

Figure 51: Results majority voting (HMM2)

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	270	(90.000%)	270	(90.000%)	90.000%
2	26	(8.6667%)	296	(98.667%)	98.667%
3	0	(0.0000%)	296	(98.667%)	98.667%
4	0	(0.0000%)	296	(98.667%)	98.667%
5	1	(0.33333%)	297	(99.000%)	99.000%
6	0	(0.0000%)	297	(99.000%)	99.000%
7	0	(0.0000%)	297	(99.000%)	99.000%
8	0	(0.0000%)	297	(99.000%)	99.000%
9	0	(0.0000%)	297	(99.000%)	99.000%
10	0	(0.0000%)	297	(99.000%)	99.000%
11	0	(0.0000%)	297	(99.000%)	99.000%
12	0	(0.0000%)	297	(99.000%)	99.000%
13	0	(0.0000%)	297	(99.000%)	99.000%
14	0	(0.0000%)	297	(99.000%)	99.000%
15	0	(0.0000%)	297	(99.000%)	99.000%
16	0	(0.0000%)	297	(99.000%)	99.000%
17	0	(0.0000%)	297	(99.000%)	99.000%
18	0	(0.0000%)	297	(99.000%)	99.000%
19	0	(0.0000%)	297	(99.000%)	99.000%
20	2	(0.66667%)	299	(99.667%)	99.667%

Total of test cases:	300
Rejected cases:	0

Figure 52: Results weighted majority voting (HMM1)

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	260	(86.667%)	260	(86.667%)	86.667%
2	35	(11.667%)	295	(98.333%)	98.333%
3	0	(0.0000%)	295	(98.333%)	98.333%
4	0	(0.0000%)	295	(98.333%)	98.333%
5	1	(0.33333%)	296	(98.667%)	98.667%
6	0	(0.0000%)	296	(98.667%)	98.667%
7	0	(0.0000%)	296	(98.667%)	98.667%
8	0	(0.0000%)	296	(98.667%)	98.667%
9	0	(0.0000%)	296	(98.667%)	98.667%
10	0	(0.0000%)	296	(98.667%)	98.667%
11	0	(0.0000%)	296	(98.667%)	98.667%
12	0	(0.0000%)	296	(98.667%)	98.667%
13	0	(0.0000%)	296	(98.667%)	98.667%
14	0	(0.0000%)	296	(98.667%)	98.667%
15	0	(0.0000%)	296	(98.667%)	98.667%
16	0	(0.0000%)	296	(98.667%)	98.667%
17	0	(0.0000%)	296	(98.667%)	98.667%
18	0	(0.0000%)	296	(98.667%)	98.667%
19	0	(0.0000%)	296	(98.667%)	98.667%
20	0	(0.0000%)	296	(98.667%)	98.667%

Total of test cases: 300
Rejected cases: 0

Figure 53: Results weighted majority voting (HMM2)

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	274	(91.333%)	274	(91.333%)	91.639%
2	21	(7.0000%)	295	(98.333%)	98.662%
3	4	(1.3333%)	299	(99.667%)	100.00%
4	0	(0.0000%)	299	(99.667%)	100.00%
5	0	(0.0000%)	299	(99.667%)	100.00%
6	0	(0.0000%)	299	(99.667%)	100.00%
7	0	(0.0000%)	299	(99.667%)	100.00%
8	0	(0.0000%)	299	(99.667%)	100.00%
9	0	(0.0000%)	299	(99.667%)	100.00%
10	0	(0.0000%)	299	(99.667%)	100.00%
11	0	(0.0000%)	299	(99.667%)	100.00%
12	0	(0.0000%)	299	(99.667%)	100.00%
13	0	(0.0000%)	299	(99.667%)	100.00%
14	0	(0.0000%)	299	(99.667%)	100.00%
15	0	(0.0000%)	299	(99.667%)	100.00%
16	0	(0.0000%)	299	(99.667%)	100.00%
17	0	(0.0000%)	299	(99.667%)	100.00%
18	0	(0.0000%)	299	(99.667%)	100.00%
19	0	(0.0000%)	299	(99.667%)	100.00%
20	0	(0.0000%)	299	(99.667%)	100.00%

Total of test cases: 300
Rejected cases: 1

Figure 54: Results rank summation (HMM1)

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	276	(92.000%)	276	(92.000%)	92.308%
2	19	(6.3333%)	295	(98.333%)	98.662%
3	3	(1.0000%)	298	(99.333%)	99.666%
4	1	(0.33333%)	299	(99.667%)	100.00%
5	0	(0.0000%)	299	(99.667%)	100.00%
6	0	(0.0000%)	299	(99.667%)	100.00%
7	0	(0.0000%)	299	(99.667%)	100.00%
8	0	(0.0000%)	299	(99.667%)	100.00%
9	0	(0.0000%)	299	(99.667%)	100.00%
10	0	(0.0000%)	299	(99.667%)	100.00%
11	0	(0.0000%)	299	(99.667%)	100.00%
12	0	(0.0000%)	299	(99.667%)	100.00%
13	0	(0.0000%)	299	(99.667%)	100.00%
14	0	(0.0000%)	299	(99.667%)	100.00%
15	0	(0.0000%)	299	(99.667%)	100.00%
16	0	(0.0000%)	299	(99.667%)	100.00%
17	0	(0.0000%)	299	(99.667%)	100.00%
18	0	(0.0000%)	299	(99.667%)	100.00%
19	0	(0.0000%)	299	(99.667%)	100.00%
20	0	(0.0000%)	299	(99.667%)	100.00%

Total of test cases:	300
Rejected cases:	1

Figure 55: Results rank summation (HMM2)

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	289	(96.333%)	289	(96.333%)	96.333%
2	9	(3.0000%)	298	(99.333%)	99.333%
3	2	(0.66667%)	300	(100.00%)	100.00%
4	0	(0.0000%)	300	(100.00%)	100.00%
5	0	(0.0000%)	300	(100.00%)	100.00%
6	0	(0.0000%)	300	(100.00%)	100.00%
7	0	(0.0000%)	300	(100.00%)	100.00%
8	0	(0.0000%)	300	(100.00%)	100.00%
9	0	(0.0000%)	300	(100.00%)	100.00%
10	0	(0.0000%)	300	(100.00%)	100.00%
11	0	(0.0000%)	300	(100.00%)	100.00%
12	0	(0.0000%)	300	(100.00%)	100.00%
13	0	(0.0000%)	300	(100.00%)	100.00%
14	0	(0.0000%)	300	(100.00%)	100.00%
15	0	(0.0000%)	300	(100.00%)	100.00%
16	0	(0.0000%)	300	(100.00%)	100.00%
17	0	(0.0000%)	300	(100.00%)	100.00%
18	0	(0.0000%)	300	(100.00%)	100.00%
19	0	(0.0000%)	300	(100.00%)	100.00%
20	0	(0.0000%)	300	(100.00%)	100.00%

Total of test cases:	300
Rejected cases:	0

Figure 56: Results score summation, logarithmic transformation (HMM1)

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	289	(96.333%)	289	(96.333%)	96.333%
2	8	(2.6667%)	297	(99.000%)	99.000%
3	3	(1.0000%)	300	(100.00%)	100.00%
4	0	(0.0000%)	300	(100.00%)	100.00%
5	0	(0.0000%)	300	(100.00%)	100.00%
6	0	(0.0000%)	300	(100.00%)	100.00%
7	0	(0.0000%)	300	(100.00%)	100.00%
8	0	(0.0000%)	300	(100.00%)	100.00%
9	0	(0.0000%)	300	(100.00%)	100.00%
10	0	(0.0000%)	300	(100.00%)	100.00%
11	0	(0.0000%)	300	(100.00%)	100.00%
12	0	(0.0000%)	300	(100.00%)	100.00%
13	0	(0.0000%)	300	(100.00%)	100.00%
14	0	(0.0000%)	300	(100.00%)	100.00%
15	0	(0.0000%)	300	(100.00%)	100.00%
16	0	(0.0000%)	300	(100.00%)	100.00%
17	0	(0.0000%)	300	(100.00%)	100.00%
18	0	(0.0000%)	300	(100.00%)	100.00%
19	0	(0.0000%)	300	(100.00%)	100.00%
20	0	(0.0000%)	300	(100.00%)	100.00%

Total of test cases: 300
Rejected cases: 0

Figure 57: Results score summation, logarithmic transformation (HMM2)

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	291	(97.000%)	291	(97.000%)	97.000%
2	5	(1.6667%)	296	(98.667%)	98.667%
3	2	(0.66667%)	298	(99.333%)	99.333%
4	1	(0.33333%)	299	(99.667%)	99.667%
5	1	(0.33333%)	300	(100.00%)	100.00%
6	0	(0.0000%)	300	(100.00%)	100.00%
7	0	(0.0000%)	300	(100.00%)	100.00%
8	0	(0.0000%)	300	(100.00%)	100.00%
9	0	(0.0000%)	300	(100.00%)	100.00%
10	0	(0.0000%)	300	(100.00%)	100.00%
11	0	(0.0000%)	300	(100.00%)	100.00%
12	0	(0.0000%)	300	(100.00%)	100.00%
13	0	(0.0000%)	300	(100.00%)	100.00%
14	0	(0.0000%)	300	(100.00%)	100.00%
15	0	(0.0000%)	300	(100.00%)	100.00%
16	0	(0.0000%)	300	(100.00%)	100.00%
17	0	(0.0000%)	300	(100.00%)	100.00%
18	0	(0.0000%)	300	(100.00%)	100.00%
19	0	(0.0000%)	300	(100.00%)	100.00%
20	0	(0.0000%)	300	(100.00%)	100.00%

Total of test cases: 300
Rejected cases: 0

Figure 58: Results score summation, logistic transformation (HMM1)

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	290	(96.667%)	290	(96.667%)	96.667%
2	7	(2.3333%)	297	(99.000%)	99.000%
3	0	(0.0000%)	297	(99.000%)	99.000%
4	1	(0.33333%)	298	(99.333%)	99.333%
5	1	(0.33333%)	299	(99.667%)	99.667%
6	1	(0.33333%)	300	(100.00%)	100.00%
7	0	(0.0000%)	300	(100.00%)	100.00%
8	0	(0.0000%)	300	(100.00%)	100.00%
9	0	(0.0000%)	300	(100.00%)	100.00%
10	0	(0.0000%)	300	(100.00%)	100.00%
11	0	(0.0000%)	300	(100.00%)	100.00%
12	0	(0.0000%)	300	(100.00%)	100.00%
13	0	(0.0000%)	300	(100.00%)	100.00%
14	0	(0.0000%)	300	(100.00%)	100.00%
15	0	(0.0000%)	300	(100.00%)	100.00%
16	0	(0.0000%)	300	(100.00%)	100.00%
17	0	(0.0000%)	300	(100.00%)	100.00%
18	0	(0.0000%)	300	(100.00%)	100.00%
19	0	(0.0000%)	300	(100.00%)	100.00%
20	0	(0.0000%)	300	(100.00%)	100.00%

Total of test cases:	300
Rejected cases:	0

Figure 59: Results score summation, logistic transformation (HMM2)

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	294	(98.000%)	294	(98.000%)	98.000%
2	3	(1.0000%)	297	(99.000%)	99.000%
3	0	(0.0000%)	297	(99.000%)	99.000%
4	3	(1.0000%)	300	(100.00%)	100.00%
5	0	(0.0000%)	300	(100.00%)	100.00%
6	0	(0.0000%)	300	(100.00%)	100.00%
7	0	(0.0000%)	300	(100.00%)	100.00%
8	0	(0.0000%)	300	(100.00%)	100.00%
9	0	(0.0000%)	300	(100.00%)	100.00%
10	0	(0.0000%)	300	(100.00%)	100.00%
11	0	(0.0000%)	300	(100.00%)	100.00%
12	0	(0.0000%)	300	(100.00%)	100.00%
13	0	(0.0000%)	300	(100.00%)	100.00%
14	0	(0.0000%)	300	(100.00%)	100.00%
15	0	(0.0000%)	300	(100.00%)	100.00%
16	0	(0.0000%)	300	(100.00%)	100.00%
17	0	(0.0000%)	300	(100.00%)	100.00%
18	0	(0.0000%)	300	(100.00%)	100.00%
19	0	(0.0000%)	300	(100.00%)	100.00%
20	0	(0.0000%)	300	(100.00%)	100.00%

Total of test cases:	300
Rejected cases:	0

Figure 60: Results Bayes's combination rule (HMM1)

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	289	(96.333%)	289	(96.333%)	96.333%
2	8	(2.6667%)	297	(99.000%)	99.000%
3	0	(0.0000%)	297	(99.000%)	99.000%
4	1	(0.33333%)	298	(99.333%)	99.333%
5	2	(0.66667%)	300	(100.00%)	100.00%
6	0	(0.0000%)	300	(100.00%)	100.00%
7	0	(0.0000%)	300	(100.00%)	100.00%
8	0	(0.0000%)	300	(100.00%)	100.00%
9	0	(0.0000%)	300	(100.00%)	100.00%
10	0	(0.0000%)	300	(100.00%)	100.00%
11	0	(0.0000%)	300	(100.00%)	100.00%
12	0	(0.0000%)	300	(100.00%)	100.00%
13	0	(0.0000%)	300	(100.00%)	100.00%
14	0	(0.0000%)	300	(100.00%)	100.00%
15	0	(0.0000%)	300	(100.00%)	100.00%
16	0	(0.0000%)	300	(100.00%)	100.00%
17	0	(0.0000%)	300	(100.00%)	100.00%
18	0	(0.0000%)	300	(100.00%)	100.00%
19	0	(0.0000%)	300	(100.00%)	100.00%
20	0	(0.0000%)	300	(100.00%)	100.00%

Total of test cases:	300
Rejected cases:	0

Figure 61: Results Bayes's combination rule (HMM2)

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	292	(97.333%)	292	(97.333%)	98.983%
2	3	(1.0000%)	295	(98.333%)	100.00%
3	0	(0.0000%)	295	(98.333%)	100.00%
4	0	(0.0000%)	295	(98.333%)	100.00%
5	0	(0.0000%)	295	(98.333%)	100.00%
6	0	(0.0000%)	295	(98.333%)	100.00%
7	0	(0.0000%)	295	(98.333%)	100.00%
8	0	(0.0000%)	295	(98.333%)	100.00%
9	0	(0.0000%)	295	(98.333%)	100.00%
10	0	(0.0000%)	295	(98.333%)	100.00%
11	0	(0.0000%)	295	(98.333%)	100.00%
12	0	(0.0000%)	295	(98.333%)	100.00%
13	0	(0.0000%)	295	(98.333%)	100.00%
14	0	(0.0000%)	295	(98.333%)	100.00%
15	0	(0.0000%)	295	(98.333%)	100.00%
16	0	(0.0000%)	295	(98.333%)	100.00%
17	0	(0.0000%)	295	(98.333%)	100.00%
18	0	(0.0000%)	295	(98.333%)	100.00%
19	0	(0.0000%)	295	(98.333%)	100.00%
20	0	(0.0000%)	295	(98.333%)	100.00%

Total of test cases:	300
Rejected cases:	5

Figure 62: Results combiner of Yu (HMM1)

Rank	**Hits**	(per rank)	**Hits**	(accum.)	Reliability
1	293	(97.667%)	293	(97.667%)	99.322%
2	2	(0.66667%)	295	(98.333%)	100.00%
3	0	(0.0000%)	295	(98.333%)	100.00%
4	0	(0.0000%)	295	(98.333%)	100.00%
5	0	(0.0000%)	295	(98.333%)	100.00%
6	0	(0.0000%)	295	(98.333%)	100.00%
7	0	(0.0000%)	295	(98.333%)	100.00%
8	0	(0.0000%)	295	(98.333%)	100.00%
9	0	(0.0000%)	295	(98.333%)	100.00%
10	0	(0.0000%)	295	(98.333%)	100.00%
11	0	(0.0000%)	295	(98.333%)	100.00%
12	0	(0.0000%)	295	(98.333%)	100.00%
13	0	(0.0000%)	295	(98.333%)	100.00%
14	0	(0.0000%)	295	(98.333%)	100.00%
15	0	(0.0000%)	295	(98.333%)	100.00%
16	0	(0.0000%)	295	(98.333%)	100.00%
17	0	(0.0000%)	295	(98.333%)	100.00%
18	0	(0.0000%)	295	(98.333%)	100.00%
19	0	(0.0000%)	295	(98.333%)	100.00%
20	0	(0.0000%)	295	(98.333%)	100.00%

Total of test cases:	300
Rejected cases:	5

Figure 63: Results combiner of Yu (HMM2)

A.4 Combination Profile and Eigenface Classifier

Rank	**Hits**	(per rank)	**Hits**	(accum.)	Reliability
1	240	(80.000%)	240	(80.000%)	100.00%

Total of test cases:	300
Rejected cases:	60

Figure 64: Results consensus voting (HMM1)

Rank	**Hits**	(per rank)	**Hits**	(accum.)	Reliability
1	240	(80.000%)	240	(80.000%)	100.00%

Total of test cases:	300
Rejected cases:	60

Figure 65: Results majority voting (HMM1)

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	284	(94.667%)	284	(94.667%)	94.667%
2	15	(5.0000%)	299	(99.667%)	99.667%
3	0	(0.0000%)	299	(99.667%)	99.667%
4	0	(0.0000%)	299	(99.667%)	99.667%
5	1	(0.33333%)	300	(100.00%)	100.00%
6	0	(0.0000%)	300	(100.00%)	100.00%
7	0	(0.0000%)	300	(100.00%)	100.00%
8	0	(0.0000%)	300	(100.00%)	100.00%
9	0	(0.0000%)	300	(100.00%)	100.00%
10	0	(0.0000%)	300	(100.00%)	100.00%
11	0	(0.0000%)	300	(100.00%)	100.00%
12	0	(0.0000%)	300	(100.00%)	100.00%
13	0	(0.0000%)	300	(100.00%)	100.00%
14	0	(0.0000%)	300	(100.00%)	100.00%
15	0	(0.0000%)	300	(100.00%)	100.00%
16	0	(0.0000%)	300	(100.00%)	100.00%
17	0	(0.0000%)	300	(100.00%)	100.00%
18	0	(0.0000%)	300	(100.00%)	100.00%
19	0	(0.0000%)	300	(100.00%)	100.00%
20	0	(0.0000%)	300	(100.00%)	100.00%

Total of test cases:	300
Rejected cases:	0

Figure 66: Results weighted majority voting (HMM1)

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	282	(94.000%)	282	(94.000%)	94.631%
2	15	(5.0000%)	297	(99.000%)	99.664%
3	0	(0.0000%)	297	(99.000%)	99.664%
4	1	(0.33333%)	298	(99.333%)	100.00%
5	0	(0.0000%)	298	(99.333%)	100.00%
6	0	(0.0000%)	298	(99.333%)	100.00%
7	0	(0.0000%)	298	(99.333%)	100.00%
8	0	(0.0000%)	298	(99.333%)	100.00%
9	0	(0.0000%)	298	(99.333%)	100.00%
10	0	(0.0000%)	298	(99.333%)	100.00%
11	0	(0.0000%)	298	(99.333%)	100.00%
12	0	(0.0000%)	298	(99.333%)	100.00%
13	0	(0.0000%)	298	(99.333%)	100.00%
14	0	(0.0000%)	298	(99.333%)	100.00%
15	0	(0.0000%)	298	(99.333%)	100.00%
16	0	(0.0000%)	298	(99.333%)	100.00%
17	0	(0.0000%)	298	(99.333%)	100.00%
18	0	(0.0000%)	298	(99.333%)	100.00%
19	0	(0.0000%)	298	(99.333%)	100.00%
20	0	(0.0000%)	298	(99.333%)	100.00%

Total of test cases:	300
Rejected cases:	2

Figure 67: Results rank summation (HMM1)

Rank	**Hits**	(per rank)	**Hits**	(accum.)	Reliability
1	293	(97.667%)	293	(97.667%)	97.667%
2	6	(2.0000%)	299	(99.667%)	99.667%
3	0	(0.0000%)	299	(99.667%)	99.667%
4	1	(0.33333%)	300	(100.00%)	100.00%
5	0	(0.0000%)	300	(100.00%)	100.00%
6	0	(0.0000%)	300	(100.00%)	100.00%
7	0	(0.0000%)	300	(100.00%)	100.00%
8	0	(0.0000%)	300	(100.00%)	100.00%
9	0	(0.0000%)	300	(100.00%)	100.00%
10	0	(0.0000%)	300	(100.00%)	100.00%
11	0	(0.0000%)	300	(100.00%)	100.00%
12	0	(0.0000%)	300	(100.00%)	100.00%
13	0	(0.0000%)	300	(100.00%)	100.00%
14	0	(0.0000%)	300	(100.00%)	100.00%
15	0	(0.0000%)	300	(100.00%)	100.00%
16	0	(0.0000%)	300	(100.00%)	100.00%
17	0	(0.0000%)	300	(100.00%)	100.00%
18	0	(0.0000%)	300	(100.00%)	100.00%
19	0	(0.0000%)	300	(100.00%)	100.00%
20	0	(0.0000%)	300	(100.00%)	100.00%

Total of test cases:	300
Rejected cases:	0

Figure 68: Results score summation, logarithmic transformation (HMM1)

Rank	**Hits**	(per rank)	**Hits**	(accum.)	Reliability
1	292	(97.333%)	292	(97.333%)	97.333%
2	7	(2.3333%)	299	(99.667%)	99.667%
3	0	(0.0000%)	299	(99.667%)	99.667%
4	0	(0.0000%)	299	(99.667%)	99.667%
5	1	(0.33333%)	300	(100.00%)	100.00%
6	0	(0.0000%)	300	(100.00%)	100.00%
7	0	(0.0000%)	300	(100.00%)	100.00%
8	0	(0.0000%)	300	(100.00%)	100.00%
9	0	(0.0000%)	300	(100.00%)	100.00%
10	0	(0.0000%)	300	(100.00%)	100.00%
11	0	(0.0000%)	300	(100.00%)	100.00%
12	0	(0.0000%)	300	(100.00%)	100.00%
13	0	(0.0000%)	300	(100.00%)	100.00%
14	0	(0.0000%)	300	(100.00%)	100.00%
15	0	(0.0000%)	300	(100.00%)	100.00%
16	0	(0.0000%)	300	(100.00%)	100.00%
17	0	(0.0000%)	300	(100.00%)	100.00%
18	0	(0.0000%)	300	(100.00%)	100.00%
19	0	(0.0000%)	300	(100.00%)	100.00%
20	0	(0.0000%)	300	(100.00%)	100.00%

Total of test cases:	300
Rejected cases:	0

Figure 69: Results score summation, logistic transformation (HMM1)

Rank	**Hits**	(per rank)	**Hits**	(accum.)	Reliability
1	295	(98.333%)	295	(98.333%)	98.333%
2	5	(1.6667%)	300	(100.00%)	100.00%
3	0	(0.0000%)	300	(100.00%)	100.00%
4	0	(0.0000%)	300	(100.00%)	100.00%
5	0	(0.0000%)	300	(100.00%)	100.00%
6	0	(0.0000%)	300	(100.00%)	100.00%
7	0	(0.0000%)	300	(100.00%)	100.00%
8	0	(0.0000%)	300	(100.00%)	100.00%
9	0	(0.0000%)	300	(100.00%)	100.00%
10	0	(0.0000%)	300	(100.00%)	100.00%
11	0	(0.0000%)	300	(100.00%)	100.00%
12	0	(0.0000%)	300	(100.00%)	100.00%
13	0	(0.0000%)	300	(100.00%)	100.00%
14	0	(0.0000%)	300	(100.00%)	100.00%
15	0	(0.0000%)	300	(100.00%)	100.00%
16	0	(0.0000%)	300	(100.00%)	100.00%
17	0	(0.0000%)	300	(100.00%)	100.00%
18	0	(0.0000%)	300	(100.00%)	100.00%
19	0	(0.0000%)	300	(100.00%)	100.00%
20	0	(0.0000%)	300	(100.00%)	100.00%

Total of test cases:	300
Rejected cases:	0

Figure 70: Results Bayes's combination rule (HMM1)

Rank	**Hits**	(per rank)	**Hits**	(accum.)	Reliability
1	294	(98.000%)	294	(98.000%)	99.661%
2	0	(0.0000%)	294	(98.000%)	99.661%
3	1	(0.33333%)	295	(98.333%)	100.00%
4	0	(0.0000%)	295	(98.333%)	100.00%
5	0	(0.0000%)	295	(98.333%)	100.00%
6	0	(0.0000%)	295	(98.333%)	100.00%
7	0	(0.0000%)	295	(98.333%)	100.00%
8	0	(0.0000%)	295	(98.333%)	100.00%
9	0	(0.0000%)	295	(98.333%)	100.00%
10	0	(0.0000%)	295	(98.333%)	100.00%
11	0	(0.0000%)	295	(98.333%)	100.00%
12	0	(0.0000%)	295	(98.333%)	100.00%
13	0	(0.0000%)	295	(98.333%)	100.00%
14	0	(0.0000%)	295	(98.333%)	100.00%
15	0	(0.0000%)	295	(98.333%)	100.00%
16	0	(0.0000%)	295	(98.333%)	100.00%
17	0	(0.0000%)	295	(98.333%)	100.00%
18	0	(0.0000%)	295	(98.333%)	100.00%
19	0	(0.0000%)	295	(98.333%)	100.00%
20	0	(0.0000%)	295	(98.333%)	100.00%

Total of test cases:	300
Rejected cases:	5

Figure 71: Results combiner of Yu (HMM1)

A.5 Combination HMM and Eigenface Classifier

Rank	**Hits**	(per rank)	**Hits**	(accum.)	Reliability
1	128	(85.333%)	128	(85.333%)	100.00%

Total of test cases: 150
Rejected cases: 22

Figure 72: Results consensus voting (HMM1)

Rank	**Hits**	(per rank)	**Hits**	(accum.)	Reliability
1	128	(85.333%)	128	(85.333%)	100.00%

Total of test cases: 150
Rejected cases: 22

Figure 73: Results majority voting (HMM1)

Rank	**Hits**	(per rank)	**Hits**	(accum.)	Reliability
1	142	(94.667%)	142	(94.667%)	94.667%
2	7	(4.6667%)	149	(99.333%)	99.333%
3	0	(0.0000%)	149	(99.333%)	99.333%
4	0	(0.0000%)	149	(99.333%)	99.333%
5	0	(0.0000%)	149	(99.333%)	99.333%
6	0	(0.0000%)	149	(99.333%)	99.333%
7	0	(0.0000%)	149	(99.333%)	99.333%
8	0	(0.0000%)	149	(99.333%)	99.333%
9	0	(0.0000%)	149	(99.333%)	99.333%
10	0	(0.0000%)	149	(99.333%)	99.333%
11	0	(0.0000%)	149	(99.333%)	99.333%
12	0	(0.0000%)	149	(99.333%)	99.333%
13	0	(0.0000%)	149	(99.333%)	99.333%
14	0	(0.0000%)	149	(99.333%)	99.333%
15	0	(0.0000%)	149	(99.333%)	99.333%
16	0	(0.0000%)	149	(99.333%)	99.333%
17	0	(0.0000%)	149	(99.333%)	99.333%
18	0	(0.0000%)	149	(99.333%)	99.333%
19	0	(0.0000%)	149	(99.333%)	99.333%
20	0	(0.0000%)	149	(99.333%)	99.333%

Total of test cases: 150
Rejected cases: 0

Figure 74: Results weighted majority voting (HMM1)

Rank	**Hits**	(per rank)	**Hits**	(accum.)	Reliability
1	146	(97.333%)	146	(97.333%)	97.987%
2	2	(1.3333%)	148	(98.667%)	99.329%
3	1	(0.66667%)	149	(99.333%)	100.00%
4	0	(0.0000%)	149	(99.333%)	100.00%
5	0	(0.0000%)	149	(99.333%)	100.00%
6	0	(0.0000%)	149	(99.333%)	100.00%
7	0	(0.0000%)	149	(99.333%)	100.00%
8	0	(0.0000%)	149	(99.333%)	100.00%
9	0	(0.0000%)	149	(99.333%)	100.00%
10	0	(0.0000%)	149	(99.333%)	100.00%
11	0	(0.0000%)	149	(99.333%)	100.00%
12	0	(0.0000%)	149	(99.333%)	100.00%
13	0	(0.0000%)	149	(99.333%)	100.00%
14	0	(0.0000%)	149	(99.333%)	100.00%
15	0	(0.0000%)	149	(99.333%)	100.00%
16	0	(0.0000%)	149	(99.333%)	100.00%
17	0	(0.0000%)	149	(99.333%)	100.00%
18	0	(0.0000%)	149	(99.333%)	100.00%
19	0	(0.0000%)	149	(99.333%)	100.00%
20	0	(0.0000%)	149	(99.333%)	100.00%

Total of test cases:	150
Rejected cases:	1

Figure 75: Results rank summation (HMM1)

Rank	**Hits**	(per rank)	**Hits**	(accum.)	Reliability
1	148	(98.667%)	148	(98.667%)	98.667%
2	1	(0.66667%)	149	(99.333%)	99.333%
3	1	(0.66667%)	150	(100.00%)	100.00%
4	0	(0.0000%)	150	(100.00%)	100.00%
5	0	(0.0000%)	150	(100.00%)	100.00%
6	0	(0.0000%)	150	(100.00%)	100.00%
7	0	(0.0000%)	150	(100.00%)	100.00%
8	0	(0.0000%)	150	(100.00%)	100.00%
9	0	(0.0000%)	150	(100.00%)	100.00%
10	0	(0.0000%)	150	(100.00%)	100.00%
11	0	(0.0000%)	150	(100.00%)	100.00%
12	0	(0.0000%)	150	(100.00%)	100.00%
13	0	(0.0000%)	150	(100.00%)	100.00%
14	0	(0.0000%)	150	(100.00%)	100.00%
15	0	(0.0000%)	150	(100.00%)	100.00%
16	0	(0.0000%)	150	(100.00%)	100.00%
17	0	(0.0000%)	150	(100.00%)	100.00%
18	0	(0.0000%)	150	(100.00%)	100.00%
19	0	(0.0000%)	150	(100.00%)	100.00%
20	0	(0.0000%)	150	(100.00%)	100.00%

Total of test cases:	150
Rejected cases:	0

Figure 76: Results score summation, logarithmic transformation (HMM1)

Rank	**Hits**	(per rank)	**Hits**	(accum.)	Reliability
1	145	(96.667%)	145	(96.667%)	96.667%
2	4	(2.6667%)	149	(99.333%)	99.333%
3	0	(0.0000%)	149	(99.333%)	99.333%
4	1	(0.66667%)	150	(100.00%)	100.00%
5	0	(0.0000%)	150	(100.00%)	100.00%
6	0	(0.0000%)	150	(100.00%)	100.00%
7	0	(0.0000%)	150	(100.00%)	100.00%
8	0	(0.0000%)	150	(100.00%)	100.00%
9	0	(0.0000%)	150	(100.00%)	100.00%
10	0	(0.0000%)	150	(100.00%)	100.00%
11	0	(0.0000%)	150	(100.00%)	100.00%
12	0	(0.0000%)	150	(100.00%)	100.00%
13	0	(0.0000%)	150	(100.00%)	100.00%
14	0	(0.0000%)	150	(100.00%)	100.00%
15	0	(0.0000%)	150	(100.00%)	100.00%
16	0	(0.0000%)	150	(100.00%)	100.00%
17	0	(0.0000%)	150	(100.00%)	100.00%
18	0	(0.0000%)	150	(100.00%)	100.00%
19	0	(0.0000%)	150	(100.00%)	100.00%
20	0	(0.0000%)	150	(100.00%)	100.00%

Total of test cases:	150
Rejected cases:	0

Figure 77: Results score summation, logistic transformation (HMM1)

Rank	**Hits**	(per rank)	**Hits**	(accum.)	Reliability
1	147	(98.000%)	147	(98.000%)	98.000%
2	3	(2.0000%)	150	(100.00%)	100.00%
3	0	(0.0000%)	150	(100.00%)	100.00%
4	0	(0.0000%)	150	(100.00%)	100.00%
5	0	(0.0000%)	150	(100.00%)	100.00%
6	0	(0.0000%)	150	(100.00%)	100.00%
7	0	(0.0000%)	150	(100.00%)	100.00%
8	0	(0.0000%)	150	(100.00%)	100.00%
9	0	(0.0000%)	150	(100.00%)	100.00%
10	0	(0.0000%)	150	(100.00%)	100.00%
11	0	(0.0000%)	150	(100.00%)	100.00%
12	0	(0.0000%)	150	(100.00%)	100.00%
13	0	(0.0000%)	150	(100.00%)	100.00%
14	0	(0.0000%)	150	(100.00%)	100.00%
15	0	(0.0000%)	150	(100.00%)	100.00%
16	0	(0.0000%)	150	(100.00%)	100.00%
17	0	(0.0000%)	150	(100.00%)	100.00%
18	0	(0.0000%)	150	(100.00%)	100.00%
19	0	(0.0000%)	150	(100.00%)	100.00%
20	0	(0.0000%)	150	(100.00%)	100.00%

Total of test cases:	150
Rejected cases:	0

Figure 78: Results Bayes's combination rule (HMM1)

A.6 Transformations

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	255	(85.000%)	255	(85.000%)	85.000%
2	8	(2.6667%)	263	(87.667%)	87.667%
3	8	(2.6667%)	271	(90.333%)	90.333%
4	2	(0.66667%)	273	(91.000%)	91.000%
5	3	(1.0000%)	276	(92.000%)	92.000%
6	1	(0.33333%)	277	(92.333%)	92.333%
7	1	(0.33333%)	278	(92.667%)	92.667%
8	11	(3.6667%)	289	(96.333%)	96.333%
9	6	(2.0000%)	295	(98.333%)	98.333%
10	0	(0.0000%)	295	(98.333%)	98.333%
11	0	(0.0000%)	295	(98.333%)	98.333%
12	0	(0.0000%)	295	(98.333%)	98.333%
13	0	(0.0000%)	295	(98.333%)	98.333%
14	0	(0.0000%)	295	(98.333%)	98.333%
15	0	(0.0000%)	295	(98.333%)	98.333%
16	0	(0.0000%)	295	(98.333%)	98.333%
17	4	(1.3333%)	299	(99.667%)	99.667%
18	1	(0.33333%)	300	(100.00%)	100.00%
19	0	(0.0000%)	300	(100.00%)	100.00%
20	0	(0.0000%)	300	(100.00%)	100.00%

Total of test cases:	300
Rejected cases:	0

Figure 79: Results without transformation of the score function (HMM1)

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	288	(96.000%)	288	(96.000%)	96.000%
2	7	(2.3333%)	295	(98.333%)	98.333%
3	1	(0.33333%)	296	(98.667%)	98.667%
4	0	(0.0000%)	296	(98.667%)	98.667%
5	2	(0.66667%)	298	(99.333%)	99.333%
6	0	(0.0000%)	298	(99.333%)	99.333%
7	0	(0.0000%)	298	(99.333%)	99.333%
8	2	(0.66667%)	300	(100.00%)	100.00%
9	0	(0.0000%)	300	(100.00%)	100.00%
10	0	(0.0000%)	300	(100.00%)	100.00%
11	0	(0.0000%)	300	(100.00%)	100.00%
12	0	(0.0000%)	300	(100.00%)	100.00%
13	0	(0.0000%)	300	(100.00%)	100.00%
14	0	(0.0000%)	300	(100.00%)	100.00%
15	0	(0.0000%)	300	(100.00%)	100.00%
16	0	(0.0000%)	300	(100.00%)	100.00%
17	0	(0.0000%)	300	(100.00%)	100.00%
18	0	(0.0000%)	300	(100.00%)	100.00%
19	0	(0.0000%)	300	(100.00%)	100.00%
20	0	(0.0000%)	300	(100.00%)	100.00%

Total of test cases:	300
Rejected cases:	0

Figure 80: Results with linear transformation of the score function (HMM1)

Rank	**Hits**	(per rank)	**Hits**	(accum.)	Reliability
1	289	(96.333%)	289	(96.333%)	96.333%
2	9	(3.0000%)	298	(99.333%)	99.333%
3	2	(0.66667%)	300	(100.00%)	100.00%
4	0	(0.0000%)	300	(100.00%)	100.00%
5	0	(0.0000%)	300	(100.00%)	100.00%
6	0	(0.0000%)	300	(100.00%)	100.00%
7	0	(0.0000%)	300	(100.00%)	100.00%
8	0	(0.0000%)	300	(100.00%)	100.00%
9	0	(0.0000%)	300	(100.00%)	100.00%
10	0	(0.0000%)	300	(100.00%)	100.00%
11	0	(0.0000%)	300	(100.00%)	100.00%
12	0	(0.0000%)	300	(100.00%)	100.00%
13	0	(0.0000%)	300	(100.00%)	100.00%
14	0	(0.0000%)	300	(100.00%)	100.00%
15	0	(0.0000%)	300	(100.00%)	100.00%
16	0	(0.0000%)	300	(100.00%)	100.00%
17	0	(0.0000%)	300	(100.00%)	100.00%
18	0	(0.0000%)	300	(100.00%)	100.00%
19	0	(0.0000%)	300	(100.00%)	100.00%
20	0	(0.0000%)	300	(100.00%)	100.00%

Total of test cases:	300
Rejected cases:	0

Figure 81: Results with logarithmic transformation of the score function (HMM1)

Rank	**Hits**	(per rank)	**Hits**	(accum.)	Reliability
1	287	(95.667%)	287	(95.667%)	95.667%
2	7	(2.3333%)	294	(98.000%)	98.000%
3	2	(0.66667%)	296	(98.667%)	98.667%
4	0	(0.0000%)	296	(98.667%)	98.667%
5	0	(0.0000%)	296	(98.667%)	98.667%
6	0	(0.0000%)	296	(98.667%)	98.667%
7	0	(0.0000%)	296	(98.667%)	98.667%
8	0	(0.0000%)	296	(98.667%)	98.667%
9	0	(0.0000%)	296	(98.667%)	98.667%
10	4	(1.3333%)	300	(100.00%)	100.00%
11	0	(0.0000%)	300	(100.00%)	100.00%
12	0	(0.0000%)	300	(100.00%)	100.00%
13	0	(0.0000%)	300	(100.00%)	100.00%
14	0	(0.0000%)	300	(100.00%)	100.00%
15	0	(0.0000%)	300	(100.00%)	100.00%
16	0	(0.0000%)	300	(100.00%)	100.00%
17	0	(0.0000%)	300	(100.00%)	100.00%
18	0	(0.0000%)	300	(100.00%)	100.00%
19	0	(0.0000%)	300	(100.00%)	100.00%
20	0	(0.0000%)	300	(100.00%)	100.00%

Total of test cases:	300
Rejected cases:	0

Figure 82: Results with exponential transformation of the score function (HMM1)

Rank	**Hits**	(per rank)	**Hits**	(accum.)	Reliability
1	291	(97.000%)	291	(97.000%)	97.000%
2	5	(1.6667%)	296	(98.667%)	98.667%
3	2	(0.66667%)	298	(99.333%)	99.333%
4	1	(0.33333%)	299	(99.667%)	99.667%
5	1	(0.33333%)	300	(100.00%)	100.00%
6	0	(0.0000%)	300	(100.00%)	100.00%
7	0	(0.0000%)	300	(100.00%)	100.00%
8	0	(0.0000%)	300	(100.00%)	100.00%
9	0	(0.0000%)	300	(100.00%)	100.00%
10	0	(0.0000%)	300	(100.00%)	100.00%
11	0	(0.0000%)	300	(100.00%)	100.00%
12	0	(0.0000%)	300	(100.00%)	100.00%
13	0	(0.0000%)	300	(100.00%)	100.00%
14	0	(0.0000%)	300	(100.00%)	100.00%
15	0	(0.0000%)	300	(100.00%)	100.00%
16	0	(0.0000%)	300	(100.00%)	100.00%
17	0	(0.0000%)	300	(100.00%)	100.00%
18	0	(0.0000%)	300	(100.00%)	100.00%
19	0	(0.0000%)	300	(100.00%)	100.00%
20	0	(0.0000%)	300	(100.00%)	100.00%

Total of test cases:	300
Rejected cases:	0

Figure 83: Results with logistic transformation of the score function (HMM1)

A.7 Reduction

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	272	(90.667%)	272	(90.667%)	90.667%
2	19	(6.3333%)	291	(97.000%)	97.000%
3	5	(1.6667%)	296	(98.667%)	98.667%
4	0	(0.0000%)	296	(98.667%)	98.667%
5	1	(0.33333%)	297	(99.000%)	99.000%
6	1	(0.33333%)	298	(99.333%)	99.333%
7	0	(0.0000%)	298	(99.333%)	99.333%
8	2	(0.66667%)	300	(100.00%)	100.00%
9	0	(0.0000%)	300	(100.00%)	100.00%
10	0	(0.0000%)	300	(100.00%)	100.00%
11	0	(0.0000%)	300	(100.00%)	100.00%
12	0	(0.0000%)	300	(100.00%)	100.00%
13	0	(0.0000%)	300	(100.00%)	100.00%
14	0	(0.0000%)	300	(100.00%)	100.00%
15	0	(0.0000%)	300	(100.00%)	100.00%
16	0	(0.0000%)	300	(100.00%)	100.00%
17	0	(0.0000%)	300	(100.00%)	100.00%
18	0	(0.0000%)	300	(100.00%)	100.00%
19	0	(0.0000%)	300	(100.00%)	100.00%
20	0	(0.0000%)	300	(100.00%)	100.00%

Total of test cases:	300
Rejected cases:	0

Figure 84: Results combination set as union of 5 classes

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	273	(91.000%)	273	(91.000%)	91.000%
2	17	(5.6667%)	290	(96.667%)	96.667%
3	5	(1.6667%)	295	(98.333%)	98.333%
4	3	(1.0000%)	298	(99.333%)	99.333%
5	2	(0.66667%)	300	(100.00%)	100.00%
6	0	(0.0000%)	300	(100.00%)	100.00%
7	0	(0.0000%)	300	(100.00%)	100.00%
8	0	(0.0000%)	300	(100.00%)	100.00%
9	0	(0.0000%)	300	(100.00%)	100.00%
10	0	(0.0000%)	300	(100.00%)	100.00%
11	0	(0.0000%)	300	(100.00%)	100.00%
12	0	(0.0000%)	300	(100.00%)	100.00%
13	0	(0.0000%)	300	(100.00%)	100.00%
14	0	(0.0000%)	300	(100.00%)	100.00%
15	0	(0.0000%)	300	(100.00%)	100.00%
16	0	(0.0000%)	300	(100.00%)	100.00%
17	0	(0.0000%)	300	(100.00%)	100.00%
18	0	(0.0000%)	300	(100.00%)	100.00%
19	0	(0.0000%)	300	(100.00%)	100.00%
20	0	(0.0000%)	300	(100.00%)	100.00%

Total of test cases:	300
Rejected cases:	0

Figure 85: Results combination set as union of 10 classes

Rank	**Hits**	(per rank)	**Hits**	(accum.)	Reliability
1	275	(91.667%)	275	(91.667%)	91.667%
2	14	(4.6667%)	289	(96.333%)	96.333%
3	6	(2.0000%)	295	(98.333%)	98.333%
4	1	(0.33333%)	296	(98.667%)	98.667%
5	0	(0.0000%)	296	(98.667%)	98.667%
6	2	(0.66667%)	298	(99.333%)	99.333%
7	2	(0.66667%)	300	(100.00%)	100.00%
8	0	(0.0000%)	300	(100.00%)	100.00%
9	0	(0.0000%)	300	(100.00%)	100.00%
10	0	(0.0000%)	300	(100.00%)	100.00%
11	0	(0.0000%)	300	(100.00%)	100.00%
12	0	(0.0000%)	300	(100.00%)	100.00%
13	0	(0.0000%)	300	(100.00%)	100.00%
14	0	(0.0000%)	300	(100.00%)	100.00%
15	0	(0.0000%)	300	(100.00%)	100.00%
16	0	(0.0000%)	300	(100.00%)	100.00%
17	0	(0.0000%)	300	(100.00%)	100.00%
18	0	(0.0000%)	300	(100.00%)	100.00%
19	0	(0.0000%)	300	(100.00%)	100.00%
20	0	(0.0000%)	300	(100.00%)	100.00%

Total of test cases:	300
Rejected cases:	0

Figure 86: Results combination set as union of 15 classes

Rank	**Hits**	(per rank)	**Hits**	(accum.)	Reliability
1	277	(92.333%)	277	(92.333%)	92.333%
2	14	(4.6667%)	291	(97.000%)	97.000%
3	5	(1.6667%)	296	(98.667%)	98.667%
4	2	(0.66667%)	298	(99.333%)	99.333%
5	2	(0.66667%)	300	(100.00%)	100.00%
6	0	(0.0000%)	300	(100.00%)	100.00%
7	0	(0.0000%)	300	(100.00%)	100.00%
8	0	(0.0000%)	300	(100.00%)	100.00%
9	0	(0.0000%)	300	(100.00%)	100.00%
10	0	(0.0000%)	300	(100.00%)	100.00%
11	0	(0.0000%)	300	(100.00%)	100.00%
12	0	(0.0000%)	300	(100.00%)	100.00%
13	0	(0.0000%)	300	(100.00%)	100.00%
14	0	(0.0000%)	300	(100.00%)	100.00%
15	0	(0.0000%)	300	(100.00%)	100.00%
16	0	(0.0000%)	300	(100.00%)	100.00%
17	0	(0.0000%)	300	(100.00%)	100.00%
18	0	(0.0000%)	300	(100.00%)	100.00%
19	0	(0.0000%)	300	(100.00%)	100.00%
20	0	(0.0000%)	300	(100.00%)	100.00%

Total of test cases:	300
Rejected cases:	0

Figure 87: Results combination set as union of 20 classes

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	281	(93.667%)	281	(93.667%)	93.667%
2	17	(5.6667%)	298	(99.333%)	99.333%
3	0	(0.0000%)	298	(99.333%)	99.333%
4	2	(0.66667%)	300	(100.00%)	100.00%
5	0	(0.0000%)	300	(100.00%)	100.00%
6	0	(0.0000%)	300	(100.00%)	100.00%
7	0	(0.0000%)	300	(100.00%)	100.00%
8	0	(0.0000%)	300	(100.00%)	100.00%
9	0	(0.0000%)	300	(100.00%)	100.00%
10	0	(0.0000%)	300	(100.00%)	100.00%
11	0	(0.0000%)	300	(100.00%)	100.00%
12	0	(0.0000%)	300	(100.00%)	100.00%
13	0	(0.0000%)	300	(100.00%)	100.00%
14	0	(0.0000%)	300	(100.00%)	100.00%
15	0	(0.0000%)	300	(100.00%)	100.00%
16	0	(0.0000%)	300	(100.00%)	100.00%
17	0	(0.0000%)	300	(100.00%)	100.00%
18	0	(0.0000%)	300	(100.00%)	100.00%
19	0	(0.0000%)	300	(100.00%)	100.00%
20	0	(0.0000%)	300	(100.00%)	100.00%

Total of test cases:	300
Rejected cases:	0

Figure 88: Results combination set as union of 25 classes

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	264	(88.000%)	264	(88.000%)	90.722%
2	2	(0.66667%)	266	(88.667%)	91.409%
3	0	(0.0000%)	266	(88.667%)	91.409%
4	0	(0.0000%)	266	(88.667%)	91.409%
5	0	(0.0000%)	266	(88.667%)	91.409%
6	0	(0.0000%)	266	(88.667%)	91.409%
7	0	(0.0000%)	266	(88.667%)	91.409%
8	0	(0.0000%)	266	(88.667%)	91.409%
9	0	(0.0000%)	266	(88.667%)	91.409%
10	0	(0.0000%)	266	(88.667%)	91.409%
11	0	(0.0000%)	266	(88.667%)	91.409%
12	0	(0.0000%)	266	(88.667%)	91.409%
13	0	(0.0000%)	266	(88.667%)	91.409%
14	0	(0.0000%)	266	(88.667%)	91.409%
15	0	(0.0000%)	266	(88.667%)	91.409%
16	0	(0.0000%)	266	(88.667%)	91.409%
17	0	(0.0000%)	266	(88.667%)	91.409%
18	0	(0.0000%)	266	(88.667%)	91.409%
19	0	(0.0000%)	266	(88.667%)	91.409%
20	0	(0.0000%)	266	(88.667%)	91.409%

Total of test cases:	300
Rejected cases:	9

Figure 89: Results combination set as intersection of 5 classes

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	280	(93.333%)	280	(93.333%)	93.333%
2	6	(2.0000%)	286	(95.333%)	95.333%
3	0	(0.0000%)	286	(95.333%)	95.333%
4	0	(0.0000%)	286	(95.333%)	95.333%
5	0	(0.0000%)	286	(95.333%)	95.333%
6	0	(0.0000%)	286	(95.333%)	95.333%
7	0	(0.0000%)	286	(95.333%)	95.333%
8	0	(0.0000%)	286	(95.333%)	95.333%
9	0	(0.0000%)	286	(95.333%)	95.333%
10	0	(0.0000%)	286	(95.333%)	95.333%
11	0	(0.0000%)	286	(95.333%)	95.333%
12	0	(0.0000%)	286	(95.333%)	95.333%
13	0	(0.0000%)	286	(95.333%)	95.333%
14	0	(0.0000%)	286	(95.333%)	95.333%
15	0	(0.0000%)	286	(95.333%)	95.333%
16	0	(0.0000%)	286	(95.333%)	95.333%
17	0	(0.0000%)	286	(95.333%)	95.333%
18	0	(0.0000%)	286	(95.333%)	95.333%
19	0	(0.0000%)	286	(95.333%)	95.333%
20	0	(0.0000%)	286	(95.333%)	95.333%

Total of test cases:	300
Rejected cases:	0

Figure 90: Results combination set as intersection of 10 classes

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	285	(95.000%)	285	(95.000%)	95.000%
2	8	(2.6667%)	293	(97.667%)	97.667%
3	2	(0.66667%)	295	(98.333%)	98.333%
4	0	(0.0000%)	295	(98.333%)	98.333%
5	0	(0.0000%)	295	(98.333%)	98.333%
6	0	(0.0000%)	295	(98.333%)	98.333%
7	0	(0.0000%)	295	(98.333%)	98.333%
8	0	(0.0000%)	295	(98.333%)	98.333%
9	0	(0.0000%)	295	(98.333%)	98.333%
10	0	(0.0000%)	295	(98.333%)	98.333%
11	0	(0.0000%)	295	(98.333%)	98.333%
12	0	(0.0000%)	295	(98.333%)	98.333%
13	0	(0.0000%)	295	(98.333%)	98.333%
14	0	(0.0000%)	295	(98.333%)	98.333%
15	0	(0.0000%)	295	(98.333%)	98.333%
16	0	(0.0000%)	295	(98.333%)	98.333%
17	0	(0.0000%)	295	(98.333%)	98.333%
18	0	(0.0000%)	295	(98.333%)	98.333%
19	0	(0.0000%)	295	(98.333%)	98.333%
20	0	(0.0000%)	295	(98.333%)	98.333%

Total of test cases:	300
Rejected cases:	0

Figure 91: Results combination set as intersection of 15 classes

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	289	(96.333%)	289	(96.333%)	96.333%
2	9	(3.0000%)	298	(99.333%)	99.333%
3	2	(0.66667%)	300	(100.00%)	100.00%
4	0	(0.0000%)	300	(100.00%)	100.00%
5	0	(0.0000%)	300	(100.00%)	100.00%
6	0	(0.0000%)	300	(100.00%)	100.00%
7	0	(0.0000%)	300	(100.00%)	100.00%
8	0	(0.0000%)	300	(100.00%)	100.00%
9	0	(0.0000%)	300	(100.00%)	100.00%
10	0	(0.0000%)	300	(100.00%)	100.00%
11	0	(0.0000%)	300	(100.00%)	100.00%
12	0	(0.0000%)	300	(100.00%)	100.00%
13	0	(0.0000%)	300	(100.00%)	100.00%
14	0	(0.0000%)	300	(100.00%)	100.00%
15	0	(0.0000%)	300	(100.00%)	100.00%
16	0	(0.0000%)	300	(100.00%)	100.00%
17	0	(0.0000%)	300	(100.00%)	100.00%
18	0	(0.0000%)	300	(100.00%)	100.00%
19	0	(0.0000%)	300	(100.00%)	100.00%
20	0	(0.0000%)	300	(100.00%)	100.00%

Total of test cases: 300
Rejected cases: 0

Figure 92: Results combination set as intersection of 20 classes

Rank	Hits	(per rank)	Hits	(accum.)	Reliability
1	289	(96.333%)	289	(96.333%)	96.333%
2	9	(3.0000%)	298	(99.333%)	99.333%
3	2	(0.66667%)	300	(100.00%)	100.00%
4	0	(0.0000%)	300	(100.00%)	100.00%
5	0	(0.0000%)	300	(100.00%)	100.00%
6	0	(0.0000%)	300	(100.00%)	100.00%
7	0	(0.0000%)	300	(100.00%)	100.00%
8	0	(0.0000%)	300	(100.00%)	100.00%
9	0	(0.0000%)	300	(100.00%)	100.00%
10	0	(0.0000%)	300	(100.00%)	100.00%
11	0	(0.0000%)	300	(100.00%)	100.00%
12	0	(0.0000%)	300	(100.00%)	100.00%
13	0	(0.0000%)	300	(100.00%)	100.00%
14	0	(0.0000%)	300	(100.00%)	100.00%
15	0	(0.0000%)	300	(100.00%)	100.00%
16	0	(0.0000%)	300	(100.00%)	100.00%
17	0	(0.0000%)	300	(100.00%)	100.00%
18	0	(0.0000%)	300	(100.00%)	100.00%
19	0	(0.0000%)	300	(100.00%)	100.00%
20	0	(0.0000%)	300	(100.00%)	100.00%

Total of test cases: 300
Rejected cases: 0

Figure 93: Results combination set as intersection of 25 classes