Noname manuscript No.
(will be inserted by the editor)

Reduced Analytic Dependency Modeling: Robust Fusion for Visual Recognition

Andy J Ma · Pong C Yuen

Received: date / Accepted: date

Abstract This paper addresses the robustness issue of information fusion for visual recognition. Analyzing limitations in existing fusion methods, we discover two key factors affecting the performance and robustness of a fusion model under different data distributions, namely 1) data dependency and 2) fusion assumption on posterior distribution. Considering these two factors, we develop a new framework to model dependency based on probabilistic properties of posteriors without any assumption on the data distribution. Making use of the range characteristics of posteriors, the fusion model is formulated as an analytic function multiplied by a constant with respect to the class label. With the analytic fusion model, we give an equivalent condition to the independent assumption and derive the dependency model from the marginal distribution property. Since the number of terms in the dependency model increases exponentially, the Reduced Analytic Dependency Model (RADM) is proposed based on the convergent property of analytic function. Finally, the optimal coefficients in the RADM are learned by incorporating label information from training data to minimize the empirical classification error under regularized least square criterion, which ensures the discriminative power. Experimental results from robust non-parametric statistical tests show that the proposed RADM method statistically significantly outperforms eight state-of-the-art score-level fusion methods on eight image/video datasets for different tasks of Digit, Flower, Face, Human Action, Object, and Consumer Video recognition.

Keywords Robustness · Score-Level Fusion · Probabilistic Constraints · Dependency Modeling · Visual Recognition

1 Introduction

With the challenges of small inter-class and large intra-class variations, many algorithms have been developed to extract local or global discriminative features (Oliva and Torralba, 2001; Lowe, 2004; Mikolajczyk and Schmid, 2004; Dalal and Triggs, 2005; He et al, 2005). While single feature may not provide sufficient information for robust recognition performance in many computer vision applications, information from multiple sources could give complementary cues for better prediction. Thus, fusion has been proposed and many encouraging results have been obtained (Kittler et al, 1998; Prabhakar and Jain, 2002; Toh et al, 2004b; Dass et al, 2005; Jain et al, 2005; Ross et al, 2006; Zhang et al, 2007; Nandakumar et al, 2008; Terrades et al, 2009; Gehler and Nowozin, 2009; He et al, 2010; Scheirer et al, 2010; Mittal et al, 2011; Awais et al, 2011; Ye et al, 2012; Liu et al, 2012; Natarajan et al, 2012; Fernando et al, 2012; Yuan et al, 2012; Ma et al, 2013a; Liu et al, 2013; Tang et al, 2013; Wang et al, 2013; Lan et al, 2014; Oh et al, 2014). Robustness is an important issue in the fusion process. Many robust statistical methods have been developed to estimate location, scale and parameters in general probability distribution function (Comaniciu, 2003; Chen and Meer, 2005; Huber and Ronchetti, 2009). Nevertheless, it receives little attention to study the fusion robustness for visual recognition due to the difficulty in discovering the relationship between the fusion model and recognition performance for robustness analysis.

In Scheirer et al (2010), a robust score normalization method was proposed by employing extreme value theory. While the normalized scores can be combined by the commonly used classifier combination rules (Kittler et al, 1998),

A. J. Ma
Department of Computer Science, Hong Kong Baptist University, Hong Kong
E-mail: jhma@comp.hkbu.edu.hk

P. C. Yuen
Department of Computer Science, Hong Kong Baptist University, Hong Kong, and BNU-HKBU United International College, Zhuhai
E-mail: pcyuen@comp.hkbu.edu.hk

it is a general assumption that classification scores are conditionally independent distributed. This independent assumption could simplify the fusion problem, but may not be valid in many practical pattern recognition applications. As such, the fusion performance will deteriorate. Instead of utilizing the conditionally independent assumption, lots of classifier fusion methods (Ueda, 2000; Demiriz et al, 2002; Toh et al, 2004b; Gehler and Nowozin, 2009) study the relationship between scores in order to improve the prediction performance. Since the probabilistic interpretation of these methods are not clear, it was shown by Ma et al (2013a) that modeling dependency explicitly give better and more robust performance. Therefore, data dependency modeling is one of the elements to ensure the fusion robustness for visual recognition.

Although explicit dependency modeling shows some superiorities over the independent and non-probabilistic fusion methods, existing dependency modeling algorithms (Dass et al, 2005; Terrades et al, 2009; Ma et al, 2013a) are developed under different assumptions on the distribution of posteriors. Copula function with multivariate normal assumption was used to model the score dependency in Dass et al (2005). On the other hand, Terrades et al (2009) proposed to combine classifiers in a non-Bayesian framework by linear combination under the Dependent Normal (DN) assumption as in Dass et al (2005). Besides the normal assumption, the linear dependency modeling method (Ma et al, 2013a) was proposed under the assumption that posteriors will not deviate very much from the priors. Since the derivation of these methods is based on specific assumptions on the posterior distribution, the fusion robustness in different recognition tasks cannot be guaranteed.

Based on the above analysis, the key factors affecting the visual recognition performance and robustness of a fusion model under different data distributions are summarized as follows: 1) data dependency and 2) fusion assumption on posterior distribution. Considering these two factors in the fusion process, we develop a novel fusion framework for robust visual recognition. And, a Reduced Analytic Dependency Modeling (RADM) algorithm is proposed in this paper. The contributions are highlighted as follows:

- We develop a new framework for dependency modeling without any assumption on the posterior distribution. Making use of the range characteristics of posteriors, we formulate the posterior of all the features as the multiplication of analytic function on posteriors of each feature and a constant with respect to the class label. With the analytic fusion model, an equation system is derived from the marginal distribution property. And, an equivalent condition to the independent assumption is given by the situation that the solution to the derived equation system is trivial. Since there may be infinite number of undetermined coefficients in analytic function, the dependency model is defined by setting non-trivial solution to the first N order equations. In order to deal with the problem of exponentially increasing number of terms, the Reduced Analytic Dependency Model (RADM) is proposed based on the convergent property of analytic function.
- We propose a novel RADM learning algorithm for robust score-level fusion with applications in visual recognition. Considering the dependency constraint, we empirically calculate the values in the equations derived from the marginal distribution property by the training data. In order to enhance discriminability in the RADM, we minimize the empirical classification error by the label information and formulate a new unconstrained quadratic programming problem based on the regularized least square criterion. The optimal model is obtained by setting the first derivative of the objective function to zeros. Since no assumption on the posterior distribution is imposed in the proposed model as well as the learning method, the proposed RADM method can achieve good and robust performance for different visual recognition tasks.

The preliminary version of this paper has been reported in Ma and Yuen (2012). Different from the previous version, this paper discusses the robustness issues for information fusion. In this paper, the theory has been revised to clarify how the proposed method can model dependency and the learning method has been further refined to remove the assumption about posteriors in the conference version paper. In addition, more experimental results including results on additional challenging datasets and analysis of statistical significance for robustness evaluation are added. Besides, the review section is further revised and enhanced in this paper.

The rest of this paper is organized as follows. We first review related works on score-level fusion methods. Section 3 reports the proposed method. Experimental results and conclusion are given in Section 4 and Section 5, respectively.

2 Related Works

Generally speaking, fusion of multiple pieces of information can be performed at five levels namely sensor, feature, score, rank and decision levels (Ross et al, 2006). Since classification scores contain moderate quantity of information and are easier to be accessed, score-level fusion is the most commonly used approach in many applications (Kuncheva, 2004; Ross et al, 2006). Thus, this paper focuses on the fusion process in score level. This section reviews existing score-level fusion methods, which can be categorized into probabilistic and non-probabilistic approaches. The important symbols used in this paper can be found in Table 1.

L	Number of classes
ω_l	Label for the l-th class
M	Number of features
f_m	The m-th feature descriptor
$\boldsymbol{x}_m$	The m-th feature vector constructed by f_m
s_{lm}	Posterior probability $\Pr(\omega_l\|\boldsymbol{x}_m)$
$\boldsymbol{s}_l$	Vector of posteriors $(s_{l1},\cdots,s_{lM})^T$
n	Variable order or dependency order
$\boldsymbol{n}$	Vector of variable orders for $\boldsymbol{s}_l$
h_l	Fusion function defined on $\boldsymbol{s}_l$
$\tilde{\boldsymbol{s}}_{lm}$	Vector of posteriors without s_{lm}
g_{lmn}	Analytic function defined on $\tilde{\boldsymbol{s}}_{lm}$
G_{lmn}	Integration of g_{lmn} over feature vectors except $\boldsymbol{x}_m$
$\boldsymbol{a}_l$	Weighting coefficient vector
$c_{lmn\boldsymbol{n}}$	Empirical integration estimation
q_{jl}	Difference between genuine and imposter scores
J	Number of training samples
y_j	Class label of the j-th sample

Table 1 Symbols used in this paper

2.1 Probabilistic Score Level Fusion

According to Bayesian theory (Feller, 1968), under the conditionally independent assumption, the posterior probability is given by

$$\Pr(\omega_l|\boldsymbol{x}_1,\cdots,\boldsymbol{x}_M)=\frac{P_0}{\Pr(\omega_l)^{M-1}}\prod_{m=1}^{M}\Pr(\omega_l|\boldsymbol{x}_m) \tag{1}$$

where ω_l denotes the label, M is the number of feature representations, $\boldsymbol{x}_m$ is the m-th feature representation and $P_0=\prod_{m=1}^{M}\Pr(\boldsymbol{x}_m)/\Pr(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_M)$. Product rule was then derived as equation (1) by Kittler et al (1998). Moreover, with the assumption that posterior probabilities of each classifier will not deviate dramatically from the priors, Sum rule was induced (Kittler et al, 1998). Based on Product and Sum rules, Kittler et al (1998) justified that the commonly used classifier combination rules, i.e. Max, Min, Median and Majority Vote, can be derived. Besides these combination rules developed under Bayesian framework (Kittler et al, 1998), Terrades et al (2009) tackled the classifier combination problem using a non-Bayesian probabilistic framework. Under the assumptions that classifiers can be combined linearly and the scores follow independent normal distribution, the Independent Normal (IN) combination rule was derived by Terrades et al (2009).

Since the independent assumption may deteriorate the fusion performance due to its invalidity in practical applications, the posterior probability can be computed by joint distribution estimation to model the dependency, which improves the performance and robustness. For example, Prabhakar and Jain (2002) used Parzen window density estimation to estimate the joint density of posterior probabilities by a selected set of classifiers. Since it needs numerous data to ensure the robustness in estimating the joint distribution (Silverman, 1986), the dependency between matching scores was considered by employing copula models in Dass et al (2005). With the copula function under multivariate normal distribution assumption, the joint density of matching scores can be modeled and used to compute the likelihood ratio statistics for score fusion. Terrades et al (2009) also made use of normal distribution assumption, and proposed to fuse classifiers by a linear combination model. When features are not conditionally independent, the covariance matrix in the normal distribution is not diagonal. In this case, the Dependent Normal (DN) rule (Terrades et al, 2009) was formulated into a constrained quadratic programming problem, which can be solved by nonlinear programming techniques (Luenberger and Ye, 2008). Removing normal distribution assumption on scores, Ma et al (2013a) proposed to add dependency terms to each posterior probability, and expand the product formulation as the Linear Classifier Dependency Model (LCDM) by neglecting high order terms, i.e.

$$\begin{aligned}&\Pr(\omega_l|\boldsymbol{x}_1,\cdots,\boldsymbol{x}_M)\\&\quad\approx P_0[\sum_{m=1}^{M}a_{lm}\Pr(\omega_l|\boldsymbol{x}_m)+(1-M)\Pr(\omega_l)]\end{aligned} \tag{2}$$

where $a_{l1},\cdots,a_{lM}$ are the dependency weights. Then, the optimal LCDM was learned by solving a standard linear programming problem, which maximizes the margins between genuine and imposter posterior probabilities.

2.2 Non-Probabilistic Score Level Fusion

Besides the probabilistic score-level fusion methods (Kittler et al, 1998; Prabhakar and Jain, 2002; Dass et al, 2005; Terrades et al, 2009; Ma et al, 2013a), the optimal weighting method (OWM) (Ueda, 2000), LPBoost algorithms (Demiriz et al, 2002; Gehler and Nowozin, 2009) and reduced multivariate polynomial model (RM) (Toh et al, 2004b) can be used to combine classifiers with multiple features by minimizing the empirical error of the training data. OWM and LPBoost methods aimed at determining the correct weighting for linear combination by minimizing the least square error and L_1 norm soft margin error, respectively. Since the linear algorithms are not robust to nonlinear data distribution, the Reduced Multivariate polynomial (RM) was introduced by Toh et al (2004b). However, the number of terms will increase exponentially with the model order in the multivariate polynomial. Toh et al (2004b) proposed to approximate the full polynomial by modified lumped multinomial. Then, the optimal RM model was learned by a weight-decay regularization problem in Toh et al (2004b).

Different from the supervised learning methods (Ueda, 2000; Demiriz et al, 2002; Toh et al, 2004b; Gehler and

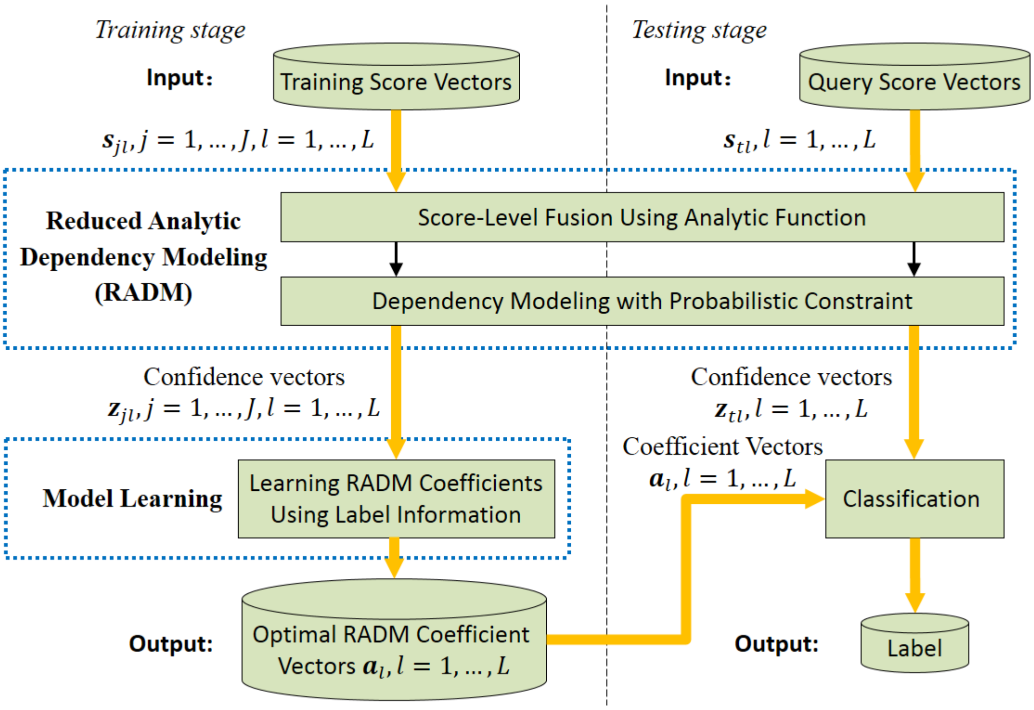

Training stage
Input:
Training Score Vectors
$s_{jl}, j = 1, \ldots, J, l = 1, \ldots, L$
Testing stage
Input:
Query Score Vectors
$s_{tl}, l = 1, \ldots, L$
Reduced Analytic Dependency Modeling (RADM)
Score-Level Fusion Using Analytic Function
Dependency Modeling with Probabilistic Constraint
Confidence vectors
$z_{jl}, j = 1, \ldots, J, l = 1, \ldots, L$
Confidence vectors
$z_{tl}, l = 1, \ldots, L$
Coefficient Vectors
$a_l, l = 1, \ldots, L$
Model Learning
Learning RADM Coefficients Using Label Information
Classification
Output:
Optimal RADM Coefficient Vectors $a_l, l = 1, \ldots, L$
Output:
Label

Fig. 1 Overview of the proposed Reduced Analytic Dependency Modeling (RADM) framework

Nowozin, 2009), the Signal Strength-based Combination (S-SC) (He and Cao, 2012) and Robust Late Fusion (RLF) (Ye et al, 2012) methods can discover the score relationship based on the testing data without the help of label information. SSC was derived based on the signal strength concept and uncertainty degree for ensemble learning. With the marginal distribution graph analysis in He and Cao (2012), it was shown that SSC could increase the margin to support the final decision. On the other hand, Ye et al (2012) proposed to convert the score vectors from each feature into pairwise score relation matrices, whose entries represent the comparative relationships of scores between any two test samples. Under the implicit assumption that the relation matrices can be decomposed into a shared rank-two matrix plus sparse errors, the fused score vector was obtained by fitting the recovered low-rank score relation matrix. Based on the low-rank and sparse properties, the noise components could be eliminated. In order to cooperate with the feature-level information, a graph based regularization term is added in the low-rank and sparse model. Then, the graph-regularized robust late fusion (GRLF) method was proposed (Ye et al, 2012).

3 Reduced Analytic Dependency Modeling

After a brief review on existing score-level fusion algorithms, this section presents the proposed Reduced Analytic Dependency Modeling (RADM) method for robust fusion. We first give an overview of the proposed method in Section 3.1. Then, the detailed derivations of each step are discussed in Sections 3.2-3.5. At last, we summarize the advantages of the proposed method over existing fusion algorithms by analyzing the fusion robustness in Section 3.6. Since the derivation of the proposed method contains a number of mathematical symbols, Table 1 summarizes the important ones to be used in this paper.

3.1 Overview of the Proposed Method

The block diagram of the proposed method is illustrated in Fig. 1 and consists of two stages: training and testing. In the training stage, the input data is a set of classification score vectors which can be constructed as follows. Given a set of J training samples from each class $\omega_l, l = 1, \cdots, L$, multiple features and classification scores for each sample can be calculated. Each training sample can then be represented by score vectors $s_{jl}, l = 1, \cdots, L$ for $j = 1, \cdots, J$. The training stage comprises two major steps: dependency modeling and model learning. In the dependency modeling step, a new and general score-level fusion model using analytic function is proposed. In order to model the dependency, probabilistic constraint is applied to the proposed score-level analytic fusion model. Since the number of terms in the analytic fusion model increases exponentially with the dependency order, convergent property of the power series is employed to reduce the number of terms in the analytic fusion model, and thus the Reduced Analytic Dependency Model (RADM) is developed. With the score vectors $s_{jl}, j = 1, \cdots, J, l =$

$1,\cdots,L$, the RADM generates a set of confidence vectors $z_{jl}, j = 1,\cdots,J, l = 1,\cdots,L$. Details of the score-level fusion model using analytic function and dependency modeling with probabilistic constraint are discussed in Sections 3.2 and 3.3, respectively. In the model learning step, the calculated confidence vectors $z_{jl}, j = 1,\cdots,J, l = 1,\cdots,L$ are used to determine the optimal RADM coefficient vectors $\boldsymbol{a}_l, l = 1,\cdots,L$. We propose to learn the optimal reduced model by minimizing the least square error to approximate the dependency modeling constraint. At the same time, to enhance discriminability, the objective function is further refined by making use of the label information to ensure that the genuine posterior is greater than the imposter ones. The detailed derivations of the optimization problem to learn the optimal coefficient vectors in the RADM are presented in Section 3.4 while the algorithm for solving the optimization problem is given in Section 3.5. In the testing stage, when score vectors of a query object/pattern is presented, query confidence vectors $z_{tl}, l = 1,\cdots,L$ are constructed using the proposed RADM. In the classification step, the posterior probability of each enrolled class is calculated by multiplying z_{tl} and $\boldsymbol{a}_l$. The label of the query object/pattern is assigned to the class with the maximum posterior probability.

3.2 Score-Level Fusion Using Analytic Function

Let us consider a combination problem that, there are M distinct feature descriptors $f_1,\cdots,f_M$ for any sample O. Denote feature representations $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_M$ as $\boldsymbol{x}_m = f_m(O)$. The objective of feature fusion is to estimate the posterior probability $\Pr(\omega_l|\boldsymbol{x}_1,\cdots,\boldsymbol{x}_M)$ for better and robust classification. We consider combining feature vectors by posterior probabilities of each feature, $\Pr(\omega_l|\boldsymbol{x}_m)$. Let us denote the posterior vector as $\boldsymbol{s}_l = (s_{l1},\cdots,s_{lM})^T$, where $s_{lm} = \Pr(\omega_l|\boldsymbol{x}_m)$. Since prior probabilities are not related to feature representations, prior $\Pr(\omega_l)$ is a positive constant p_l with respect to $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_M$. With these notations, the Product rule in equation (1) and LCDM in equation (2) can be rewritten as equations (3) and (4), respectively.

$$\begin{aligned}&\Pr(\omega_l|\boldsymbol{x}_1,\cdots,\boldsymbol{x}_M)\\&\quad= P_0\frac{\prod_{m=1}^M s_{lm}}{p_l^{M-1}} = P_0\cdot h_{\text{Product}}(\boldsymbol{s}_l)\end{aligned}\tag{3}$$

$$\begin{aligned}&\Pr(\omega_l|\boldsymbol{x}_1,\cdots,\boldsymbol{x}_M)\\&\quad\approx P_0(\sum_{m=1}^M a_{lm}s_{lm} + (1-M)p_l) = P_0\cdot h_{\text{LCDM}}(\boldsymbol{s}_l)\end{aligned}\tag{4}$$

As mentioned in Section 2, the Product rule is given by the independent assumption, while the LCDM is derived under the assumption that posterior probabilities of each classifier will not deviate dramatically from the priors. With equations (3) and (4), the Product rule and LCDM can be formulated as two different functions h_{Product} and h_{LCDM} on posterior probabilities $s_{l1},\cdots,s_{lM}$. This implies that different fusion assumptions result in different fusion functions on the posteriors. Generally speaking, the score-level fusion model can be given by the multiplication of a function h_l for class ω_l on $s_{l1},\cdots,s_{lM}$ and a constant with respect to the class label. Since equations (3) and (4) indicate that the constant is equal to P_0, we define the general fusion model as the following equation,

$$\Pr(\omega_l|\boldsymbol{x}_1,\cdots,\boldsymbol{x}_M) = P_0\cdot h_l(s_{l1},\cdots,s_{lM})\tag{5}$$

In order to explicitly write out the fusion function h_l, we propose to determine h_l by general series. If h_l is represented by divergent series, the fusion score could be infinite or swing between different values. Thus, h_l needs to be determined by convergent series for robustness. Considering the characteristics of probability, the estimated posteriors s_{lm} for $m = 1,\cdots,M$ and $l = 1,\cdots,L$ are positive after score normalization and the summation of s_{lm} over the class index l is equal to one for fixed m. Thus, the range of each posterior s_{lm} is larger than zero and less than one. For the convergence of the series to ensure robustness, we employ power series (analytic function) to define the function h_l, since multivariate power series converges in the range of $(0,1)$ for each variable provided that the coefficients are bounded, according to mathematical analysis (Rudin, 1976). With the definition of multivariate power series (Krantz and Parks, 2002), the analytic function h_l can be expressed explicitly as the following equation,

$$h_l(\boldsymbol{s}_l;\boldsymbol{a}_l) = \sum_{k=0}^{\infty}\sum_{|\boldsymbol{n}|=k} a_{l\boldsymbol{n}}\boldsymbol{s}_l^{\boldsymbol{n}}\tag{6}$$

where $\boldsymbol{n} = (n_1,\cdots,n_M)^T$ is the vector of variable orders, variable orders $n_1,\cdots,n_M$ are non-negative integers, $|\boldsymbol{n}| = n_1+\cdots+n_M$, $\boldsymbol{s}_l^{\boldsymbol{n}} = \prod_{m=1}^M s_{lm}^{n_m}$ and $\boldsymbol{a}_l = (a_{l\boldsymbol{0}},\cdots,a_{l\boldsymbol{n}},\cdots)^T$ is the weighting coefficient vector in which $\boldsymbol{0} = (0,\cdots,0)^T$.

3.3 Dependency Modeling with Probabilistic Constraint

With the analytic fusion model given in equation (6), we further investigate the model constraint and derive the dependency model from probabilistic aspect.

According to Bayes' rule (Feller, 1968), the posterior probabilities satisfy the following equations,

$$\begin{aligned}\Pr(\omega_l|\boldsymbol{x}_m) &= \frac{\Pr(\boldsymbol{x}_m|\omega_l)\Pr(\omega_l)}{\Pr(\boldsymbol{x}_m)}\\\Pr(\omega_l|\boldsymbol{x}_1,\cdots,\boldsymbol{x}_M) &= \frac{\Pr(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_M|\omega_l)\Pr(\omega_l)}{\Pr(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_M)}\end{aligned}\tag{7}$$

By the marginal property of joint density (Feller, 1968), the conditional probability $\mathrm{Pr}(\boldsymbol{x}_m|\omega_l)$ can be calculated by integrating the conditional joint density $\mathrm{Pr}(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_M|\omega_l)$ over random measurements except $\boldsymbol{x}_m$, i.e.

$$\begin{aligned}&\mathrm{Pr}(\boldsymbol{x}_m|\omega_l) = \\ &\quad\int \mathrm{Pr}(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_M|\omega_l)d\boldsymbol{x}_1\cdots d\boldsymbol{x}_{m-1}d\boldsymbol{x}_{m+1}\cdots d\boldsymbol{x}_M\end{aligned} \tag{8}$$

On the other hand, rewriting the conditional joint density given label ω_l by equations (5) and (7), and $P_0 = \frac{\prod_{m=1}^M \mathrm{Pr}(\boldsymbol{x}_m)}{\mathrm{Pr}(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_M)}$ as mentioned in Section 2.1, it becomes

$$\mathrm{Pr}(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_M|\omega_l) = \frac{\prod_{m=1}^M \mathrm{Pr}(\boldsymbol{x}_m)}{\mathrm{Pr}(\omega_l)} h_l(\boldsymbol{s}_l;\boldsymbol{a}_l) \tag{9}$$

With notations of the posteriors $s_{lm} = \mathrm{Pr}(\omega_l|\boldsymbol{x}_m)$, substituting the conditional probability $\mathrm{Pr}(\boldsymbol{x}_m|\omega_l)$ in (7) and joint density $\mathrm{Pr}(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_M|\omega_l)$ in (9) into equation (8), we get the model constraint as follows:

$$\begin{aligned}&s_{lm} = \\ &\quad\int \prod_{m'\neq m} \mathrm{Pr}(\boldsymbol{x}_{m'})h_l(\boldsymbol{s}_l;\boldsymbol{a}_l)d\boldsymbol{x}_1\cdots d\boldsymbol{x}_{m-1}d\boldsymbol{x}_{m+1}\cdots d\boldsymbol{x}_M\end{aligned} \tag{10}$$

Based on the model constraint given by equation (10), we further develop the dependency model by expanding it. Let us consider the scores obtained by the m-th feature and rewrite the fusion function (6) according to the order of s_{lm} as,

$$h_l(\boldsymbol{s}_l;\boldsymbol{a}_l) = \sum_{n=0}^{\infty} g_{lmn}(\tilde{\boldsymbol{s}}_{lm};\boldsymbol{a}_{lmn})s_{lm}^n \tag{11}$$

where $\tilde{\boldsymbol{s}}_{lm} = (s_{l1},\cdots,s_{l(m-1)},s_{l(m+1)},\cdots,s_{lM})^T$ and g_{lmn} is the coefficient of s_{lm}^n given by an analytic function of $\tilde{\boldsymbol{s}}_{lm}$ with coefficient vector $\boldsymbol{a}_{lmn}$ as defined by the following equation,

$$g_{lmn}(\tilde{\boldsymbol{s}}_{lm};\boldsymbol{a}_{lmn}) = \sum_{k=n}^{\infty} \sum_{|\boldsymbol{n}|=k,n_m=n} a_{l\boldsymbol{n}} \prod_{m'\neq m} s_{lm'}^{n_{m'}} \tag{12}$$

Substituting equation (11) into (10) and calculating the integration term by term, equation (10) becomes

$$s_{lm} = \sum_{n=0}^{\infty} G_{lmn}(\boldsymbol{a}_{lmn})s_{lm}^n \tag{13}$$

In equation (13), the integration function $G_{lmn}(\boldsymbol{a}_{lmn})$ is given by

$$\begin{aligned}&G_{lmn}(\boldsymbol{a}_{lmn}) \\ &\quad= \int \prod_{m'\neq m} \mathrm{Pr}(\boldsymbol{x}_{m'})g_{lmn}d\boldsymbol{x}_1\cdots d\boldsymbol{x}_{m-1}d\boldsymbol{x}_{m+1}\cdots d\boldsymbol{x}_M\end{aligned} \tag{14}$$

Comparing the left and the right hand sides in (13), we have the following equations for $l = 1,\cdots,L$ and $m = 1,\cdots,M$,

$$G_{lm1}(\boldsymbol{a}_{lm1}) = 1 \tag{15}$$

$$G_{lm0}(\boldsymbol{a}_{lm0}) = 0, G_{lm2}(\boldsymbol{a}_{lm2}) = 0, G_{lm3}(\boldsymbol{a}_{lm3}) = 0,\cdots \tag{16}$$

According to the definition in (12), $g_{lmr}(\tilde{\boldsymbol{s}}_{lm};\boldsymbol{a}_{lmr})$ is an analytic function similar to $h_l(\boldsymbol{s}_l;\boldsymbol{a}_l)$ in equation (6) and the score vector $\tilde{\boldsymbol{s}}_{lm}$ can be considered as mappings from feature representations $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{m-1},\boldsymbol{x}_{m+1},\cdots,\boldsymbol{x}_M$ to their posterior probabilities. Therefore, the integration of the function $\prod_{i\neq m}\mathrm{Pr}(\boldsymbol{x}_i)g_{lmr}(\tilde{\boldsymbol{s}}_{lm};\boldsymbol{a}_{lmr})$ over feature representations except $\boldsymbol{x}_m$, which is denoted by $G_{lmr}(\boldsymbol{a}_{lmr})$, is a linear function on the coefficient vector $\boldsymbol{a}_{lmr}$. Without calculating the integration, we can observe that $\boldsymbol{a}_{lm0} = \boldsymbol{0}, \boldsymbol{a}_{lm2} = \boldsymbol{0}, \boldsymbol{a}_{lm3} = \boldsymbol{0},\cdots$ is a trivial solution to the equation system (16) for $m = 1,\cdots,M$. Substituting this trivial solution into (11), we have the following proposition. (Please refer to Appendix A for the proof of this proposition.)

Proposition 1. *Conditionally independent condition given by equation* (3) *is equivalent to the situation that the solution to equation system* (16) *is trivial, i.e.*

$$\begin{aligned}&h_l(\boldsymbol{s}_l;\boldsymbol{a}_l) = p_l^{1-M}\prod_{m=1}^{M} s_{lm} \\ &\quad\Leftrightarrow \boldsymbol{a}_{lm0} = \boldsymbol{0}, \boldsymbol{a}_{lm2} = \boldsymbol{0}, \boldsymbol{a}_{lm3} = \boldsymbol{0},\cdots\end{aligned} \tag{17}$$

This proposition gives an equivalent condition to the independent assumption from the structure of the solution to equation system (16). Considering the negative and inverse-negative propositions to the proposition (17), if the solution to the equation system (16) is non-trivial, the dependency between scores can be modeled. Consequently, we propose to model the dependency by setting non-trivial solution to equation system (16). In order to deal with the problem that there is infinite number of coefficients in the dependency model, the solution to the first N equations in (16) is set as non-trivial, i.e.

$$\begin{aligned}&h_l(\boldsymbol{s}_l;\boldsymbol{a}_l) = h_l(\boldsymbol{s}_l;\boldsymbol{a}_l;N) = \sum_{k=0}^{MN}\sum_{|\boldsymbol{n}|=k} a_{l\boldsymbol{n}}\boldsymbol{s}_l^{\boldsymbol{n}} \\ &\text{s.t. } G_{lm0}(\boldsymbol{a}_{lm0}) = 0, G_{lm1}(\boldsymbol{a}_{lm1}) = 1, \\ &\quad G_{lm2}(\boldsymbol{a}_{lm2}) = 0,\cdots,G_{lmN}(\boldsymbol{a}_{lmN}) = 0, \\ &\quad m = 1,\cdots,M, \boldsymbol{n} = (n_1,\cdots,n_M)^T, 0 \le n_m \le N\end{aligned} \tag{18}$$

where N is a positive integer representing the dependency order. The fusion function given by equation (18) can model dependency between posteriors of order up to N.

The number of terms in the dependency model (18) is $(N+1)^M$, which increases exponentially with N, so that it may suffer from the problem of curse of dimension. Thus, we reduce the dependency model by the convergent property of the power series. According to the definition of convergence of series (Rudin, 1976), for any positive number ε, there exists a positive integer K, such that the reminder of the series is very small, i.e. $|\sum_{k=K+1}^{\infty}\sum_{|\boldsymbol{n}|=k} a_{l\boldsymbol{n}}\boldsymbol{s}_l^{\boldsymbol{n}}| \le \varepsilon$. If ε

Algorithm 1 Construct confidence vector z_l

```
Input: Posterior probability scores $s_{l1}, \cdots, s_{lM}$ and model parameters $K, N$;
Set $D = (0, 1, \cdots, N)^T$ and $z_l = (1, s_{l1}, s_{l1}^2, \cdots, s_{l1}^N)^T$;
for $m = 2, 3, \cdots, M$ do
    Set $\tilde{D} = (D, \mathbf{0})$, where $\mathbf{0} = (0, \cdots, 0)^T$ with the same column dimension of $D$;
    for $n = 1, 2, \cdots, N$ do
        Update $\tilde{D} = (\tilde{D}; (D, n\mathbf{1}))$ which is the column concatenation of $\tilde{D}$ and $(D, n\mathbf{1})$, where $\mathbf{1} = (1, \cdots, 1)^T$ with the same dimension of $D$;
        Update $z_l = (z_l; s_{lm}^n z_l)$ which is the column concatenation of $z_l$ and $s_{lm}^n z_l$;
        Delete the rows in $\tilde{D}$ and corresponding elements in $z_l$ such that the summations of the rows in $\tilde{D}$ are larger than $K$;
    end for
    Set $D = \tilde{D}$;
end for
Output: Confidence vector $z_l$ and index set $D$.
```

tends to zero, the analytic function can be approximated by the following equation,

$$h_l(\boldsymbol{s}_l; \boldsymbol{a}_l) \approx h_l(\boldsymbol{s}_l; \boldsymbol{a}_l; K) = \sum_{k=0}^{K} \sum_{|\boldsymbol{n}|=k} a_{l\boldsymbol{n}} \boldsymbol{s}_l^{\boldsymbol{n}} \tag{19}$$

where K denotes the model order. Combining equations (18) and (19), the Reduced Analytic Dependency Model (RADM) is given by

$$\begin{aligned} & h_l(\boldsymbol{s}_l; \boldsymbol{a}_l) \approx h_l(\boldsymbol{s}_l; \boldsymbol{a}_l; N, K) = \sum_{k=0}^{K} \sum_{|\boldsymbol{n}|=k} a_{l\boldsymbol{n}} \boldsymbol{s}_l^{\boldsymbol{n}} \\ \text{s.t. } & G_{lm0}(\boldsymbol{a}_{lm0}) \approx 0, G_{lm1}(\boldsymbol{a}_{lm1}) \approx 1, \\ & G_{lm2}(\boldsymbol{a}_{lm2}) \approx 0, \cdots, G_{lmN}(\boldsymbol{a}_{lmN}) \approx 0, \\ & m = 1, \cdots, M, \boldsymbol{n} = (n_1, \cdots, n_M)^T, 0 \le n_m \le N \end{aligned} \tag{20}$$

Denote the confidence vector $z_l = (\boldsymbol{s}_l^{\mathbf{0}}, \cdots, \boldsymbol{s}_l^{\boldsymbol{n}}, \cdots)^T$, where $\boldsymbol{s}_l^{\mathbf{0}}, \cdots, \boldsymbol{s}_l^{\boldsymbol{n}}, \cdots$ are the terms in equation (20). With these notations, the fusion function in the RADM given by equation (20) can be written as $h_l(\boldsymbol{s}_l; \boldsymbol{a}_l; N, K) = \boldsymbol{a}_l^T z_l$. The algorithmic procedure to obtain z_l in the RADM for class ω_l is presented in Algorithm 1.

3.4 Learning Optimal RADM Coefficients Using Label Information

Given J training samples with labels $y_1, \cdots, y_J$, the conditional probability s_{jlm} given label ω_l can be calculated for the j-th sample with the m-th feature. In order to estimate the integration function G_{lmn} in the reduced model (20), we substituting $g_{lmn}(\tilde{\boldsymbol{s}}_{lm}; \boldsymbol{a}_{lmn})$ in equation (12) into equation (14) and get the following equation,

$$\begin{aligned} & G_{lmn}(\boldsymbol{a}_{lmn}) \\ & = \sum_{k=n}^{\infty} \sum_{|\boldsymbol{n}|=k, n_m=n} a_{l\boldsymbol{n}} \prod_{m' \neq m} \int \Pr(\boldsymbol{x}_{m'}) s_{lm'}^{n_{m'}} d\boldsymbol{x}_{m'} \end{aligned} \tag{21}$$

The integration in equation (21) can be empirically computed by the summation of posteriors from the training samples $j = 1, \cdots, J$, i.e.

$$\int \Pr(\boldsymbol{x}_{m'}) s_{lm'}^{n_{m'}} d\boldsymbol{x}_{m'} \approx \frac{1}{J} \sum_{j=1}^{J} s_{jlm'}^{n_{m'}} \tag{22}$$

According to the reduced model (20) and the empirical estimation (22), the integration function G_{lmn} in equation (21) becomes

$$\begin{aligned} & G_{lmn}(\boldsymbol{a}_{lmn}) \approx \sum_{k=0}^{K} \sum_{|\boldsymbol{n}|=k} a_{l\boldsymbol{n}} c_{lmn\boldsymbol{n}} \\ & \text{s.t. } c_{lmn\boldsymbol{n}} = \begin{cases} \frac{1}{J^{M-1}} \prod_{m' \neq m} \sum_{j=1}^{J} s_{jlm'}^{n_{m'}}, & n_m = n \\ 0, & n_m \neq n \end{cases} \end{aligned} \tag{23}$$

In equation (23), $c_{lmn\boldsymbol{n}}$ is the culmulation of terms with corresponding order vector $\boldsymbol{n}$ in the analytic function g_{lmn} over training samples $j = 1, \cdots, J$. Since the terms for $n_m \neq n$ are not in the analytic function g_{lmn}, the culmulative value $c_{lmn\boldsymbol{n}}$ is set to be zero for $n_m \neq n$.

According to the derivation in Section 3.3, the probabilistic constraint given by the equation systems (15) (16) becomes the approximation in the reduced model (20). In order to learn the optimal coefficients in the RADM, we approximate the probabilistic constraint by minimizing the following normalized least square error

$$E(\boldsymbol{a}) = \frac{1}{2LM(N+1)} \sum_{l=1}^{L} \sum_{m=1}^{M} \sum_{n=0}^{N} (\boldsymbol{a}_l^T \boldsymbol{c}_{lmn} - b_n)^2 \tag{24}$$

where $\boldsymbol{a}$ is the column concatenation of $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_L$, $b_1 = 1$, $b_n = 0$ for $n \neq 1$, and $\boldsymbol{c}_{lmn}$ is the cumulative confidence vector with elements defined in equation (23), i.e. $\boldsymbol{c}_{lmn} = (c_{lmn\mathbf{0}}, \cdots, c_{lmn\boldsymbol{n}}, \cdots)^T$.

Although the coefficient vector $\boldsymbol{a}$ in the RADM can be determined by minimizing the error function (24), the learnt model is obtained only based on the probabilistic constraint and may not be able to classify all the training samples correctly. To enhance discriminability, it must satisfy the condition that the posterior of the true label is larger than those of the others, i.e. $\Pr(y_j | \boldsymbol{x}_{j1}, \cdots, \boldsymbol{x}_{jM}) > \Pr(\omega_l | \boldsymbol{x}_{j1}, \cdots, \boldsymbol{x}_{jM})$ for $\omega_l \neq y_j$. Since the posteriors are computed by equation (5) and P_0 is a constant with respect to class label ω_l, the discriminant condition is given by the analytic functions as $h_{y_j}(\boldsymbol{s}_{jy_j}; \boldsymbol{a}_{y_j}) > h_l(\boldsymbol{s}_{jl}; \boldsymbol{a}_l)$ for $\omega_l \neq y_j$, where $\boldsymbol{s}_{jl}$ represents the posterior vector of class label ω_l given the j-th sample, i.e. $\boldsymbol{x}_{jm}$ is the m-th feature vector for the j-th sample, $s_{jlm} = \Pr(\omega_l|\boldsymbol{x}_{jm})$, and $\boldsymbol{s}_{jl} = (s_{jl1}, \cdots, s_{jlM})^T$.

Denote the differences between the genuine and imposter scores in the analytic function for the j-th sample as $q_{jl} = h_{y_j}(\boldsymbol{s}_{jy_j}; \boldsymbol{a}_{y_j}) - h_l(\boldsymbol{s}_{jl}; \boldsymbol{a}_l)$ for $\omega_l \neq y_j$. To optimize the classification performance, we propose to maximize the summation of the genuine and imposter differences q_{jl} over $j = 1, \cdots, J$ and $\omega_l \neq y_j$. On the other hand, the reduced model $h_l(\boldsymbol{s}_l; \boldsymbol{a}_l; K, N)$ in equation (20) approximates but is not exactly equal to $h_l(\boldsymbol{s}_{jl}; \boldsymbol{a}_l)$. Thus, we minimize the least square error between the true differences q_{jl} and the differences of the genuine and imposter scores in the reduced model at the same time. Then, the objective function incorporating label information from training data is defined as the following equation,

$$
\begin{aligned}
E_{\text{Dis}}(\boldsymbol{a}, \boldsymbol{q}) = &-\frac{1}{J(L-1)} \sum_{j=1}^{J} \sum_{\omega_l \neq y_j} q_{jl} \\
&+ \frac{1}{2J(L-1)} \sum_{j=1}^{J} \sum_{\omega_l \neq y_j} ((\boldsymbol{a}_{y_j}^T \boldsymbol{z}_{jy_j} - \boldsymbol{a}_l^T \boldsymbol{z}_{jl}) - q_{jl})^2
\end{aligned}
\tag{25}
$$

where $\boldsymbol{q}$ represents the vector of score differences q_{jl}, and $\boldsymbol{z}_{jl}$ is the confidence vector of label ω_l given the j-th sample as in equation (20), i.e. $\boldsymbol{z}_{jl} = (\boldsymbol{s}_{jl}^{\boldsymbol{0}}, \cdots, \boldsymbol{s}_{jl}^{\boldsymbol{n}}, \cdots)^T$.

With the objective functions (24) for the dependency constraint and (25) for the discriminative constraint, we propose to minimize the weighted combination of them to learn the optimal RADM coefficients. On the other hand, a squared regularized term is added in the objective function, so that the fusion model suffers less from over-fitting problem. Therefore, the final optimization problem becomes,

$$
\min_{\boldsymbol{a}, \boldsymbol{q}} [E(\boldsymbol{a}) + \lambda E_{\text{Dis}}(\boldsymbol{a}, \boldsymbol{q}) + \frac{1}{2}\mu(\boldsymbol{a}^T \boldsymbol{a} + \boldsymbol{q}^T \boldsymbol{q})] \tag{26}
$$

where λ and μ are positive parameters balancing the discriminative constraint and the regularizaton term.

3.5 Model Optimization

To solve the optimization problem (26), we convert the objective function in (26) to matrix formulation. We first consider the objective function $E_{\text{Dis}}(\boldsymbol{a}, \boldsymbol{q})$ in equation (25) and rewrite it as

$$
\begin{aligned}
E_{\text{Dis}}(\boldsymbol{a}, \boldsymbol{q}) = &-\frac{1}{J(L-1)} \sum_{l=1}^{L} \sum_{l' \neq l} \sum_{y_j = \omega_l} q_{jl'} \\
&+ \frac{1}{2J(L-1)} \sum_{l=1}^{L} \sum_{l' \neq l} \sum_{y_j = \omega_l} ((\boldsymbol{a}_l^T \boldsymbol{z}_{jl} - \boldsymbol{a}_{l'}^T \boldsymbol{z}_{jl'}) - q_{jl'})^2
\end{aligned}
\tag{27}
$$

Let us set the vector of differences between genuine and imposter scores as $\boldsymbol{q}_{ll'} = (q_{j_1 l'}, \cdots, q_{j_{J_l} l'})^T$, and the confidence matrix as $Z_{ll'} = (\boldsymbol{z}_{j_1 l'}, \cdots, \boldsymbol{z}_{j_{J_l} l'})$ for $y_j = \omega_l$ in equation (27), where J_l denotes the number of samples for class ω_l. With these notations, the objective function $E_{\text{Dis}}(\boldsymbol{a}, \boldsymbol{q})$ derived by making use of the label information becomes (please refer to Appendix B for the detailed derivation)

$$
\begin{aligned}
E_{\text{Dis}}(\boldsymbol{a}, \boldsymbol{q}) = &\frac{1}{2}\boldsymbol{a}^T H_{\text{Dis}} \boldsymbol{a} \\
&+ \theta \sum_{l=1}^{L} (\frac{1}{2}\boldsymbol{q}_l^T \boldsymbol{q}_l - \boldsymbol{a}^T Z_l \boldsymbol{q}_l - \boldsymbol{q}_l^T \boldsymbol{1}), \\
\text{s.t. } H_{\text{Dis}} = &\ \theta \sum_{l=1}^{L} Z_l Z_l^T, \\
\boldsymbol{q}_l = &\ (\boldsymbol{q}_{l1}, \cdots, \boldsymbol{q}_{l(l-1)}, \boldsymbol{q}_{l(l+1)}, \cdots, \boldsymbol{q}_{lL})^T \\
Z_l = &\begin{pmatrix}
-Z_{l1} & & & & & \\
 & \ddots & & & & \\
 & & -Z_{l(l-1)} & & & \\
Z_{ll} \quad \cdots & & Z_{ll} & Z_{ll} & \cdots & Z_{ll} \\
 & & & -Z_{l(l+1)} & & \\
 & & & & \ddots & \\
 & & & & & -Z_{lL}
\end{pmatrix}
\end{aligned}
\tag{28}
$$

where θ denotes the normalization factor $\frac{1}{J(L-1)}$. Considering the other objective function (24), denote the matrix of the cumulative confidence vectors with different dependency orders as $C_{lm} = (\boldsymbol{c}_{lm0}, \cdots, \boldsymbol{c}_{lmN})$ and let $\boldsymbol{b} = (b_0, \cdots, b_N)^T$. The error function $E(\boldsymbol{a})$ in equation (24) related to the dependency constraint can be reformulated as (please refer to Appendix C for the detailed derivation)

$$
\begin{aligned}
E(\boldsymbol{a}) = &\frac{1}{2}\boldsymbol{a}^T H \boldsymbol{a} - \boldsymbol{a}^T \boldsymbol{f} + \frac{1}{2LM(N+1)} \sum_{l=1}^{L} \boldsymbol{b}^T \boldsymbol{b}, \\
\text{s.t. } H_l = &\frac{1}{LM(N+1)} \sum_{m=1}^{M} C_{lm} C_{lm}^T, \\
H = &\begin{pmatrix} H_1 & & \\ & \ddots & \\ & & H_L \end{pmatrix}, \boldsymbol{f} = \frac{1}{LM(N+1)} \begin{pmatrix} C_{l1}\boldsymbol{b} \\ \vdots \\ C_{lM}\boldsymbol{b} \end{pmatrix}
\end{aligned}
\tag{29}
$$

With equations (28) and (29), we solve the optimization problem (26) by computing the first derivatives with respect to the model coefficient vector $\boldsymbol{a}$ and the vector $\boldsymbol{q}_l$ of differences between the genuine and imposter scores of class ω_l, respectively, and setting the derivatives to zeros. Then, we get the following equations,

$$
(H + \lambda H_{\text{Dis}} + \mu I)\boldsymbol{a} - \boldsymbol{f} - \lambda\theta \sum_{l=1}^{L} Z_l \boldsymbol{q}_l = 0 \tag{30}
$$

$$
(\lambda\theta + \mu)\boldsymbol{q}_l - \lambda\theta(Z_l^T \boldsymbol{a} + \boldsymbol{1}) = 0 \tag{31}
$$

Algorithm 2 Learning coefficient vector $\boldsymbol{a}$ in RADM.

```
Input: Scores $s_{111}, \cdots, s_{JLM}$, labels $y_1, \cdots, y_J$, and positive
    parameters $\lambda, \mu$;
Construct confidence vector $z_{jl}$ and corresponding index set $D$ by Algorithm 1 for $j = 1, \cdots, J$ and $l = 1, \cdots, L$;
Construct cumulative vector $\boldsymbol{c}_{lmn}$ as defined in equations (23) (24) with the index set $D$ for $y_j = \omega_l$, $l = 1, \cdots, L$, and $m = 1, \cdots, M$;
Set $H_{\text{Dis}} = 0, \boldsymbol{f}_{\text{Dis}} = 0$ and $\theta = \frac{1}{J(L-1)}$;
for $l = 1, \cdots, L$ do
    Compute $H_l$ in equation (29);
    Compute $H_{\text{Dis}} = H_{\text{Dis}} + \theta Z_l Z_l^T$ in equation (28);
    Compute $\boldsymbol{f}_{\text{Dis}} = \boldsymbol{f}_{\text{Dis}} + \theta Z_l \mathbf{1}$ in equation (33);
end for
Combine $H_1, \cdots, H_L$ to obtain $H$ by equation (29);
Construct $\boldsymbol{f}$ by equation (29);
Obtain the optimal solution $\boldsymbol{a}$ by equation (33);
Output: Optimal coefficient vector $\boldsymbol{a}$.
```

where I is the unit matrix with the same dimension as the coefficient vector $\boldsymbol{a}$. Denote the transformation matrix for the coefficient vector $\boldsymbol{a}$ in equation (30) as $\Phi = H + \lambda H_{\text{Dis}} + \mu I$. Substituting $\boldsymbol{q}_l$ by equation (31) into (30), it has

$$\boldsymbol{a} = \left(\Phi - \frac{\lambda^2\theta^2}{\lambda\theta+\mu}\sum_{l=1}^{L} Z_l Z_l^T\right)^{-1}\left(\boldsymbol{f} + \frac{\lambda^2\theta^2}{\lambda\theta+\mu}\sum_{l=1}^{L} Z_l \mathbf{1}\right) \qquad (32)$$

With the definition of H_{Dis} and θ in equation (28), the solution to the RADM is given by the following equation,

$$\boldsymbol{a} = \left(H + \frac{\lambda\mu}{\lambda\theta+\mu} H_{\text{Dis}} + \mu I\right)^{-1}\left(\boldsymbol{f} + \frac{\lambda^2\theta}{\lambda\theta+\mu}\boldsymbol{f}_{\text{Dis}}\right) \qquad (33)$$

where $\boldsymbol{f}_{\text{Dis}} = \theta \sum_{l=1}^{L} Z_l \mathbf{1}$ with $\theta = \frac{1}{J(L-1)}$.

At last, the algorithmic procedure to train the optimal coefficient vector $\boldsymbol{a}$ in the RADM is summarized in Algorithm 2.

3.6 Robustness Analysis of Fusion Performance

In this section, we summarize the advantages of the proposed RADM over existing score-level fusion methods for visual recognition by analyzing the fusion robustness as follows:

- Compared with the independent and non-probabilistic fusion methods (Kittler et al, 1998; Ueda, 2000; Demiriz et al, 2002; Toh et al, 2004b; Gehler and Nowozin, 2009), the proposed method is derived from probabilistic properties and models dependency explicitly by setting non-trivial solution to equation systems (15) (16). Since the probabilistic properties are valid regardless of the data distribution, the proposed method can better model dependency to ensure the robustness.
- Compared with the fusion algorithms under normal assumption (Dass et al, 2005; Terrades et al, 2009) and the assumption that posteriors will not deviate very much from the priors (Ma et al, 2013a), the RADM combines classification scores without these assumptions. Consequently, it can achieve more robust recognition performance, when these assumptions are not valid.
- Compared with the unsupervised fusion algorithms (Ye et al, 2012; He and Cao, 2012), the RADM utilizes the label information to train a discriminative model. While the unsupervised methods are unguided, the learnt score relationship may not be able to minimize the empirical classification error. Thus, the recognition performance of the unsupervised methods is unpredictable and the fusion robustness cannot be guaranteed. Since the proposed method is developed without any assumption on the posterior distribution, it can achieve strong generalization ability. Thus, the testing performance of the RADM can be robustly improved by minimizing the empirical classification error.

4 Experiments

In this section, we statistically evaluate the performance and robustness of the proposed RADM for visual recognition by comparing it with eight state-of-the-art score-level fusion algorithms on eight datasets for six different domains of recognition problems namely 1) Digit Recognition, 2) Flower Classification, 3) Face Recognition, 4) Human Action Recognition, 5) Object Categorization and 6) Consumer Video Understanding. The datasets and settings are introduced in Section 4.1. The experimental results are reported in Section 4.2 and Section 4.3. In order to evaluate the robustness of the proposed method, we employ robust nonparametric statistical tests to analyze the recognition performance in Section 4.4. At last, we demonstrate that our method can robustly improve the recognition performance by combining with a more-discriminative feature for action recognition in Section 4.5.

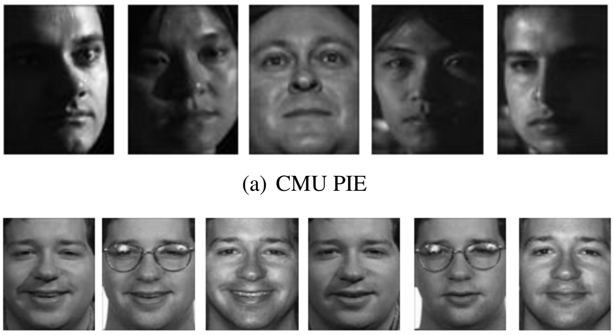
4.1 Datasets and Settings

Since limited training data are available in practice, we do not have sufficient data to learn individual models and the fusion model on two independent training sets. To avoid biased estimation, we follow the two-step training scheme in Gehler and Nowozin (2009). In the first step, cross validation (CV) is performed to select the best parameters for individual models. Then, the best individual models are learned by the training data with the best parameters. In the second step, the CV outputs (instead of scores from the best individual models trained using all the training data) corresponding to the best parameters selected in the first step are used to train the fusion model. If a validation set is not available, CV is performed again with the CV outputs in the first step to select the best fusion parameters. Otherwise, the fusion parameters are selected by the validation set. This training scheme ensures that each input for training the fusion model is a prediction of an individual model which was not trained using that sample. Since the testing data are out of the training set, this property in the two-step training scheme will provide better generalization ability for testing as mentioned in Gehler and Nowozin (2009).

Multiple feature Digit dataset (Breukelen et al, 1998) contains ten digits from 0 to 9, and 200 examples for each digit. Six features, namely Fourier coefficients, profile correlations, Karhunen-Love coefficients, pixel averages, Zernike moments and morphological features, are extracted and available on the website[1]. The experiments on this dataset was performed by randomly selecting 20 samples of each digit for training and the rest for testing. Five-fold CVs were used to select the best parameters, and train the weights for classifier combination by the CV outputs.

Oxford 17 Flowers (Nilsback and Zisserman, 2006) dataset contains 17 categories of flowers with 80 images per category. Seven features including shape, color, texture, HSV, HoG, SIFT internal, and SIFT boundary, were extracted using the methods reported in (Nilsback and Zisserman, 2006, 2008). Distance matrices of these features and three predefined splits of the dataset (17×40 for training, 17×20 for validation, and 17×20 for testing) are available on the website[2]. Outputs of the five-fold CVs were used to trained the fusion model. The best parameters were selected by the validation set. This experiment was repeated three times using the predefined splits of this dataset (Nilsback and Zisserman, 2006).

For face recognition, two publicly available face datasets, CMU PIE (Sim et al, 2003) and FERET (Phillips et al, 2000), were used for experiments. CMU PIE face dataset contains 68 subjects with 41,368 images captured under varying pose, illumination and expression. We used 105 near frontal-view face images for each individual, randomly selecting six for training, four for validation and the rest for testing. In FERET dataset, we selected 72 individuals with six near frontal-view face images per person under different face expressions. Six images for each individual were randomly separated into training, validation and testing sets with equal size, i.e. two images for each set. The selected images with different variations in CMU PIE and FERET datasets are shown in Fig. 2(a) and Fig. 2(b), respectively. Four types of features, Eigenfaces (Belhumeur et al, 1997), Fisherfaces (Belhumeur et al, 1997), Laplacianfaces (He et al, 2005) and local binary patterns (LBP) (Ahonen et al, 2004) were extracted in both two datasets. Parameters introduced from these features were determined as suggested in their papers (Belhumeur et al, 1997; Ahonen et al, 2004; He et al, 2005). The best parameters of the fusion methods were selected by the validation set and experiments were repeated ten times on CMU PIE and three times on FERET dataset with six-fold CVs.


(a) CMU PIE

(b) FERET

Fig. 2 Example images in CMU PIE and FERET Face datasets

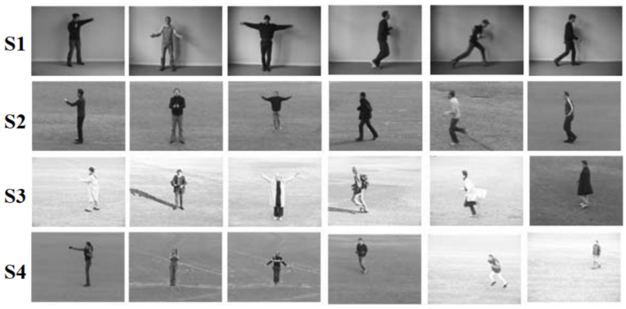


Fig. 3 Example images showing the six actions and four scenarios in KTH dataset

For human action recognition, we compared the fusion algorithms on Weizmann (Gorelick et al, 2007) and KTH (Schuldt et al, 2004) human action datasets. Weizmann dataset contains 93 videos from nine persons, each performing ten actions. Eight out of the nine persons in this dataset were used for training, and the remaining one was used for evaluation. This was repeated nine times and the recognition rates were averaged. On the other hand, there are 25 subjects performing six actions under four scenarios (as illustrated in Fig. 3) in KTH dataset. We followed the common setting in Schuldt et al (2004) to separate the video set into training (8 persons), validation (8 persons), and testing (9 persons) sets. In order to evaluate the robustness of the fusion algorithms to different variations, we also performed experiments under each scenario. Eight features including

[1] http://archive.ics.uci.edu/ml/datasets/Multiple+Features

[2] http://www.robots.ox.ac.uk/~vgg/data/flowers/17/index.html

intensity, intensity difference, HoF, HoG, HoF2D, HoG2D, HoF3D and HoG3D, were extracted from videos as reported in Ma et al (2013a). In these two datasets, eight-fold CVs were performed on the training data, and the CV outputs were used to train the weights for classifier fusion. The best parameters were selected by the CV outputs on Weizmann and validation set on KTH dataset, respectively.

PASCAL VOC 2007 (Everingham et al, 2007) is one of the benchmark datasets for visual object recognition in realistic scenes. There are are around 10,000 consumer images of 20 different object categories, which are collected from the photo-sharing website. With the default split, 5,011 images were used for training, while 4,952 images for testing. Eight features reported in Guillaumin et al (2010), including RGB, HSV, LAB, dense SIFT, Harris SIFT, dense HUE and Harris HUE with 3 × 1 horizontal decomposition of images as well as GIST descriptor, are available on the website[3] and employed in the experiments. Five-fold CVs were performed on the training data and the classifier fusion models. The best parameters were determined by the CV results. In order to reduce the impact of wiggles in the recall and precision curve, the fusion methods were evaluated by the interpolated average precision as reported in Everingham et al (2010).

Columbia Consumer Video (CCV) dataset (Jiang et al, 2011) is a recently developed benchmark for consumer video analysis. This dataset contains 9,317 YouTube videos over 20 semantic categories, in which around half of the videos are used for training and the other half for testing. In this experiment, we used three online available features[4], including SIFT visual feature, spatial-temporal interest point (STIP) visual feature, and Mel-frequency cepstral coefficients (MFCC) audio features reported in Jiang et al (2011), to evaluate the fusion methods. We selected the best parameters, trained the classifiers and fusion models by five-fold CVs. Following Jiang et al (2011), the average precision was calculated by the the uninterpolated recall and precision curve to evaluate the fusion methods.

Since the probabilities are hard to determine accurately due to the problem of limited training samples, we used Support Vector Machines (SVM) (Canu et al, 2005) for each feature and normalized the classification outputs by the double sigmoid method (Jain et al, 2005) to approximate the probabilities. Following the settings in Gehler and Nowozin (2009), Guillaumin et al (2010) and Jiang et al (2011), kernel SVMs were used for Oxford 17 Flowers, VOC 2007 and CCV datasets, and the kernel matrices were defined as exponential function $\exp(-d(\boldsymbol{x},\boldsymbol{x}')/\eta)$, where d is the distance and η is the mean of pairwise distances. Linear SVMs were employed for the other datasets. The parameter C introduced in the soft margin SVMs was selected from $\{10^{-3},\cdots,10^{3}\}$. Positive parameters λ and μ in equation (26) or (33) were selected from $\{10^{-4},\cdots,10^{4}\}$, while the dependency order N was selected from one to three and the model order K was selected from one to eight with one step increment.

Eight state-of-the-art score-level fusion algorithms were used for comparison, including Sum (Kittler et al, 1998), IN (Terrades et al, 2009), DN (Terrades et al, 2009) and LCDM (Ma et al, 2013a), LP-B (Gehler and Nowozin, 2009), RM (Toh et al, 2004b), SSC (He and Cao, 2012), GRLF (Ye et al, 2012). We used the implementations of IN, DN, LCDM[5] and SSC[6] provided by respective authors. RM was trained using the code in Toh et al (2004a), while other methods have been re-implemented. Parameters in LCDM, LP-B, RM, SSC and GRLF were selected as follows. The soft margin parameter ν in LCDM and LP-B are selected from 0.1 to 0.9 with 0.1 increment. The logistic filter parameter in SSC and the regularization parameter in RM and GRLF were selected from $\{10^{-4},\cdots,10^{4}\}$, while the model order in RM was selected from one to eight with one step increment.

4.2 Comparison Results on Multi-Class Recognition Performance

The recognition accuracies of the best single feature (BestFea) and different combination methods on Digit, Flower, CMU PIE, FERET, Weizmann and KTH datasets are shown in Table 2. Comparing the performance between the best feature and the fusion methods, we can see that recognition accuracies of all the fusion methods are higher than that of the best single feature on each dataset. This convince that the recognition performance can be improved by combining different pieces of information used in this experiment. On the other hand, the standard deviation (Std) of accuracies over the six datasets reported in the last column of Table 2 show that more robust results can be obtained by fusing multiple complementary features.

Comparing the fusion algorithms, the proposed RADM achieves the highest accuracies on all the six datasets with the smallest standard deviation. Since the data distribution varies with different datasets for different recognition tasks, these results indicate that the RADM outperforms other fusion methods and give more robust performance on different datasets for different recognition tasks.

While results in Table 2 show the rank-one accuracies, the Cumulative Match Characteristic (CMC) curve is another measure to evaluate the multi-class recognition systems. CMC curves of the top four methods on CMU PIE and FERET face datasets are plotted in Fig. 4(a) and Fig. 4(b), respectively. From these two figures, it can be seen that the RADM achieves not only the highest rank-one accuracy but

[3] http://lear.inrialpes.fr/pubs/2010/GVS10/

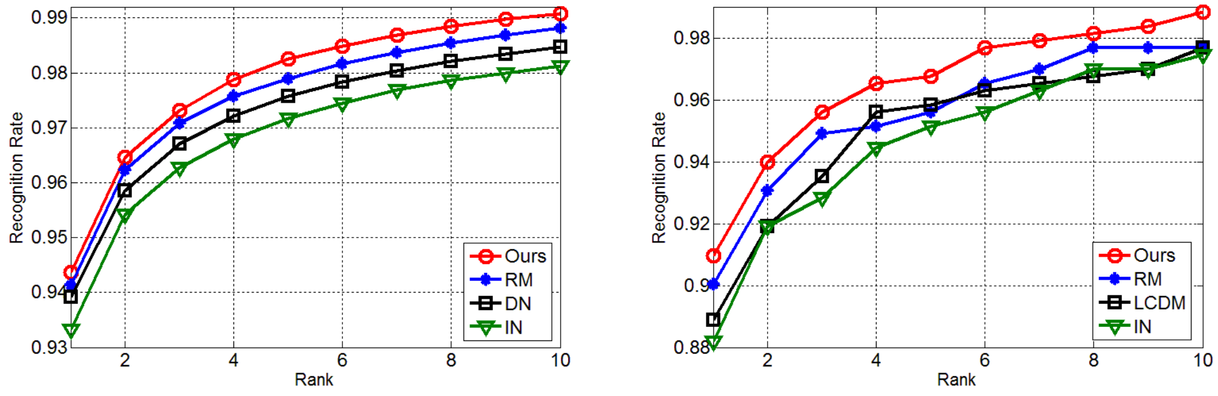
[4] http://www.ee.columbia.edu/ln/dvmm/CCV/

[5] http://www.comp.hkbu.edu.hk/~jhma/

[6] http://www.ele.uri.edu/faculty/he/

Dataset Method	Digit	Flower	CMU PIE	FERET	Weizmann	KTH	Mean ± Std
BestFea	94.77	70.39	88.87	83.33	82.22	78.70	83.05 ± 8.38
Sum	96.23	85.39	91.21	86.11	84.44	84.72	88.02 ± 4.73
IN	95.63	85.49	93.31	88.19	**85.56**	84.26	88.74 ± 4.68
DN	94.93	84.22	93.91	87.73	84.44	83.80	88.17 ± 5.05
LCDM	96.79	86.27	93.01	88.89	**85.56**	85.19	89.29 ± 4.69
LP-B	96.57	85.78	92.00	87.65	84.44	85.19	88.61 ± 4.75
RM	96.51	85.49	94.14	90.05	84.44	88.89	89.92 ± 4.73
SSC	96.88	86.08	91.80	87.04	84.44	84.26	88.42 ± 4.97
GRLF	96.28	85.98	90.72	84.03	83.33	83.80	87.36 ± 5.15
Ours	**96.98**	**87.75**	**94.34**	**90.97**	**85.56**	**90.28**	**90.98 ± 4.19**

Table 2 Recognition accuracies (%) of different methods on six datasets



(a) CMC curve on CMU PIE

(b) CMC curve on FERET

Fig. 4 CMC curves of the top four fusion methods on CMU PIE and FERET Face datasets

Method	S1	S2	S3	S4	Mean ± Std
Sum	90.74	92.59	79.63	74.07	84.26 ± 8.88
IN	90.74	92.59	81.48	72.22	84.26 ± 9.38
DN	90.74	92.59	88.89	74.07	86.57 ± 8.47
LCDM	92.59	92.59	90.74	77.78	88.43 ± 7.15
LP-B	85.19	92.59	90.74	77.78	86.58 ± 6.65
RM	88.89	92.59	88.89	77.78	87.04 ± 6.41
SSC	90.74	92.59	83.33	74.07	85.18 ± 8.42
GRLF	**96.30**	**94.44**	83.33	77.78	87.96 ± 8.88
Ours	94.44	**94.44**	**92.59**	**83.33**	**91.20 ± 5.32**

Table 3 Recognition accuracies (%) of different methods on KTH dataset with different scenarios

also the highest accuracies with different numbers of ranks. This indicates that the proposed RADM gives the best and robust results with different performance measures for multi-class recognition.

Since there are four scenarios namely indoors (S1), outdoors (S2), outdoors with different clothes (S3) and outdoors with scale variations (S4) (as illustrated in Fig. 3) in the KTH dataset, we evaluate the fusion methods under these variations. The recognition accuracies under each scenario and their average are shown in Table 3. From Table 3, we can see that GRLF achieves the highest accuracies in scenarios one and two. However, the performance of GRLF degrades a lot with the clothes and scale variations in scenarios three and four. Although the recognition accuracy of our method also decreases under these two scenarios, the degree of the deterioration is much smaller. This indicates that the RADM gives more robust performance under different variations. Thus, it achieves the highest mean accuracy and lowest standard deviation under these four scenarios as shown in the last column of Table 3.

4.3 Comparison Results on Per-Class Recognition Performance

The multi-class recognition results reported in the previous section convince the performance and robustness of the proposed method. In this section, we evaluate the fusion methods by the per-class recognition performance on VOC 2007 and CCV datasets.

The average precision (AP) with corresponding rank of the fusion methods for each class in VOC 2007 and CCV datasets is recorded in Table 4 and Table 5, respectively. From these two tables, we can see that the proposed method outperforms the other eight fusion methods in classifying 17

	Sum	IN	DN	LCDM	LP-B	RM	SSC	GRLF	Ours
Airplane	68.01 (4)	61.75 (9)	65.29 (6)	66.30 (5)	69.06 (3)	69.53 (2)	62.60 (8)	64.63 (7)	**71.10** (1)
Bicycle	41.46 (6)	37.35 (8)	42.08 (5)	53.07 (2)	52.50 (4)	53.02 (3)	36.76 (9)	38.23 (7)	**55.51** (1)
Bird	39.28 (9)	39.79 (7)	42.81 (4)	43.85 (3)	42.13 (5)	47.72 (2)	39.68 (8)	42.05 (6)	**50.13** (1)
Boat	55.73 (8)	62.33 (3)	61.13 (6)	61.54 (5)	59.22 (7)	62.61 (2)	61.88 (4)	54.09 (9)	**64.10** (1)
Bottle	24.37 (3)	22.90 (7)	22.61 (9)	23.69 (5)	23.17 (6)	24.33 (4)	22.68 (8)	**27.34** (1)	25.40 (2)
Bus	42.07 (8)	44.60 (6)	47.83 (5)	54.51 (3)	54.40 (4)	55.10 (2)	43.24 (7)	41.98 (9)	**57.64** (1)
Car	61.73 (9)	67.79 (7)	68.48 (2)	68.37 (4)	68.34 (5)	68.38 (3)	68.01 (6)	63.04 (8)	**69.54** (1)
Cat	40.42 (7)	40.39 (8)	46.99 (5)	51.61 (2)	47.31 (4)	51.27 (3)	39.58 (9)	41.21 (6)	**52.63** (1)
Chair	41.78 (8)	41.50 (9)	44.63 (5)	46.79 (4)	47.10 (2)	47.04 (3)	42.13 (7)	43.51 (6)	**47.80** (1)
Cow	25.93 (9)	27.61 (6)	27.40 (7)	34.43 (2)	33.97 (3)	33.80 (4)	26.80 (8)	27.89 (5)	**38.28** (1)
Table	38.15 (6)	37.89 (8)	35.45 (9)	40.92 (4)	42.23 (3)	42.47 (2)	38.00 (7)	39.13 (5)	**45.58** (1)
Dog	33.13 (9)	39.63 (6)	39.68 (5)	41.35 (3)	38.15 (8)	41.45 (2)	39.61 (7)	41.07 (4)	**42.06** (1)
Horse	65.78 (9)	69.05 (8)	71.96 (4)	72.08 (3)	71.42 (5)	72.29 (2)	69.66 (7)	70.97 (6)	**73.37** (1)
Motorbike	52.44 (5)	44.57 (9)	47.33 (7)	54.60 (2)	53.54 (4)	54.44 (3)	45.55 (8)	48.36 (6)	**55.96** (1)
Person	77.75 (9)	78.84 (7)	80.44 (3)	80.40 (5)	80.41 (4)	80.39 (6)	78.05 (8)	**82.04** (1)	80.63 (2)
Plant	28.09 (4)	25.98 (8)	23.44 (9)	26.40 (5)	26.34 (6)	28.18 (3)	26.13 (7)	**31.29** (1)	28.61 (2)
Sheep	25.35 (9)	31.75 (7)	31.32 (8)	32.52 (4)	32.61 (3)	31.78 (5.5)	31.78 (5.5)	32.92 (2)	**35.72** (1)
Sofa	33.33 (8)	34.02 (6)	32.39 (9)	37.00 (3)	36.68 (4)	37.25 (2)	33.97 (7)	34.74 (5)	**38.96** (1)
Train	59.70 (9)	62.02 (6)	64.45 (5)	68.59 (2)	67.59 (4)	68.25 (3)	61.80 (7)	60.99 (8)	**70.23** (1)
Monitor	33.28 (9)	34.90 (6)	36.09 (5)	39.72 (4)	40.75 (2)	40.35 (3)	34.63 (7)	34.48 (8)	**40.98** (1)
MAP	44.39 (7.4)	45.23 (7.1)	46.59 (5.9)	49.89 (3.5)	49.35 (4.3)	50.48 (3.0)	45.13 (7.2)	46.00 (5.5)	**52.21** (1.2)

Table 4 Average precisions (%) with corresponding ranks of different methods on VOC 2007 dataset

	Sum	IN	DN	LCDM	LP-B	RM	SSC	GRLF	Ours
Basketball	76.80 (2)	73.84 (7)	72.21 (9)	76.45 (3)	73.08 (8)	76.43 (4)	73.99 (6)	75.63 (5)	**77.21** (1)
Baseball	**56.38** (1)	54.17 (5)	53.74 (6)	55.10 (3)	50.70 (8)	54.74 (4)	53.70 (7)	48.84 (9)	56.30 (2)
Soccer	63.68 (4)	60.52 (8)	59.91 (9)	63.58 (5)	60.68 (7)	63.79 (3)	61.32 (6)	64.33 (2)	**64.40** (1)
Skating	85.27 (3)	81.56 (6)	81.00 (9)	84.99 (4)	81.04 (8)	85.54 (2)	81.39 (7)	83.10 (5)	**87.46** (1)
Skiing	75.57 (4)	72.92 (8)	72.51 (9)	75.38 (5)	73.03 (7)	76.30 (2.5)	73.56 (6)	76.30 (2.5)	**77.83** (1)
Swimming	74.11 (2)	70.27 (6)	69.60 (9)	72.53 (4)	70.13 (7)	74.03 (3)	70.85 (5)	69.95 (8)	**76.29** (1)
Biking	45.38 (6)	44.25 (8)	41.61 (9)	48.42 (3)	46.78 (5)	48.78 (2)	45.25 (7)	47.05 (4)	**48.79** (1)
Graduation	44.03 (8)	44.45 (7)	42.01 (9)	48.70 (4)	47.38 (5)	49.21 (3)	46.60 (6)	**50.27** (1)	49.81 (2)
Birthday	45.89 (8)	46.54 (4)	46.08 (7)	46.51 (5)	47.29 (3)	46.37 (6)	**47.53** (1)	43.47 (9)	46.91 (3)
Reception	30.00 (9)	31.83 (8)	35.21 (4)	35.71 (3)	36.97 (1)	33.86 (7)	34.22 (6)	35.10 (5)	36.10 (2)
Ceremony	41.22 (9)	47.12 (6)	47.20 (5)	46.41 (7)	51.62 (2)	45.65 (8)	48.33 (4)	50.86 (3)	**55.59** (1)
Dance	51.80 (9)	52.92 (7)	57.08 (5)	**60.29** (1)	57.33 (3)	56.83 (6)	52.72 (8)	57.10 (4)	60.05 (2)
Music	23.98 (9)	31.61 (6)	32.19 (5)	29.41 (7)	**36.07** (1)	28.78 (8)	32.36 (4)	33.00 (3)	34.33 (2)
NonMusic	70.87 (4)	68.30 (6)	66.51 (9)	69.75 (5)	67.32 (8)	71.54 (2)	67.35 (7)	70.91 (3)	**72.09** (1)
Parade	67.10 (3)	65.45 (7)	64.73 (8)	67.22 (2)	66.43 (5)	66.97 (4)	65.86 (6)	63.39 (9)	**67.75** (1)
Cat	74.84 (4)	72.70 (5)	69.59 (8)	71.11 (7)	69.25 (9)	75.28 (3)	71.52 (6)	**75.67** (1)	75.49 (2)
Dog	67.05 (3)	65.00 (7)	63.81 (9)	67.00 (4)	65.11 (6)	**67.54** (1)	63.99 (8)	65.93 (5)	67.22 (2)
Bird	68.77 (5)	66.54 (8)	65.78 (9)	69.56 (4)	67.33 (6)	70.13 (2)	66.69 (7)	69.92 (3)	**70.74** (1)
Beach	73.78 (4)	70.76 (8)	70.81 (7)	74.46 (2)	71.20 (5)	73.99 (3)	70.06 (9)	71.18 (6)	**74.96** (1)
Playground	59.73 (5)	57.56 (9)	58.84 (6)	61.34 (2)	58.70 (7)	60.73 (3)	58.05 (8)	60.24 (4)	**61.58** (1)
MAP	59.81 (5.1)	58.92 (6.8)	58.52 (7.6)	61.20 (4.0)	59.87 (5.5)	61.32 (3.8)	59.27 (6.2)	60.61 (4.6)	**63.04** (1.5)

Table 5 Average precisions (%) with corresponding ranks of different methods on CCV dataset

out of 20 classes on VOC 2007 and 12 out of 20 classes on CCV dataset. Although other methods achieves the highest APs for some categories, the RADM still gives close performance for classification of these object or video classes. These results convince that the RADM improves the per-class recognition performance for most classes, so that the overall performance is improved. Therefore, the proposed RADM gives better and more robust performance in both multi-class and per-class recognition.

Comparing the mean average precision (MAP) recorded in the last rows of Table 4 and Table 5, the proposed RADM outperforms the independent assumption based fusion methods, Sum and IN, by a remarkable improvement of 6.98% in VOC 2007 dataset, while the improvement in CCV dataset by the RADM over these two methods is smaller than that in VOC 2007 dataset. Since features in CCV are more independent than those in VOC 2007 due to the different modalities of video and audio in CCV dataset, this indicate that the independent assumption based fusion methods could give

	Sum	IN	DN	LCDM	LP-B	RM	SSC	GRLF	Ours
Digit	96.23 (7)	95.63 (8)	94.93 (9)	96.79 (3)	96.57 (4)	96.51 (5)	96.88 (2)	96.28 (6)	96.98 (1)
Flower	85.39 (8)	85.49 (6.5)	84.22 (9)	86.27 (2)	85.78 (5)	85.49 (6.5)	86.08 (3)	85.98 (4)	87.75 (1)
CMU PIE	91.21 (8)	93.31 (4)	93.91 (3)	93.01 (5)	92.00 (6)	94.14 (2)	91.80 (7)	90.72 (9)	94.34 (1)
FERET	86.11 (8)	88.19 (4)	87.73 (5)	88.89 (3)	87.65 (6)	90.05 (2)	87.04 (7)	84.03 (9)	90.97 (1)
Weizmann	84.44 (6)	85.56 (2)	84.44 (6)	85.56 (2)	84.44 (6)	84.44 (6)	84.44 (6)	83.33 (9)	85.56 (2)
KTH	84.72 (5)	84.26 (6.5)	83.80 (8.5)	85.19 (3.5)	85.19 (3.5)	88.89 (2)	84.26 (6.5)	83.80 (8.5)	90.28 (1)
KTH S1	90.74 (5.5)	90.74 (5.5)	90.74 (5.5)	92.59 (3)	85.19 (9)	88.89 (8)	90.74 (5.5)	96.30 (1)	94.44 (2)
KTH S2	92.59 (6)	92.59 (6)	92.59 (6)	92.59 (6)	92.59 (6)	92.59 (6)	92.59 (6)	94.44 (1.5)	94.44 (1.5)
KTH S3	79.63 (9)	81.48 (8)	88.89 (4.5)	90.74 (2.5)	90.74 (2.5)	88.89 (4.5)	83.33 (6.5)	83.33 (6.5)	92.59 (1)
KTH S4	74.07 (7)	72.22 (9)	74.07 (7)	77.78 (3.5)	77.78 (3.5)	77.78 (3.5)	74.07 (7)	77.78 (3.5)	83.33 (1)
VOC2007	44.39 (9)	45.23 (7)	46.59 (5)	49.89 (3)	49.35 (4)	50.48 (2)	45.13 (8)	46.00 (6)	52.21 (1)
CCV	59.81 (6)	58.92 (8)	58.52 (9)	61.20 (3)	59.87 (5)	61.32 (2)	59.27 (7)	60.61 (4)	63.04 (1)
MeanRank	7.0	6.2	6.5	3.3	5.0	4.1	6.0	5.7	1.2

Table 6 Recognition accuracies and mean average precisions (%) with corresponding ranks of different methods on all datasets

comparable results, if the fusion assumption can be satisfied in some recognition tasks, e.g. CCV dataset. However, it cannot be guaranteed that the fusion assumption is valid for all applications. Therefore, the performance of Sum and IN degrades a lot in VOC 2007 dataset, when the independent assumption is not valid. Since the RADM is derived without any specific assumption, it robustly improves the recognition performance under different data distributions for different applications.

4.4 Statistical Analysis on Fusion Robustness

When comparing algorithms over multiple datasets, we may use different measurements for different recognition tasks. For example, this paper used recognition accuracy for Digit, Flower, Face and Human Action datasets, while mean average precision (MAP) is used for PASCAL VOC 2007 and Columbia Consumer Video (CCV) datasets. Since recognition accuracy and MAP are with different commensurabilities, it may not be reasonable to compare the values of them directly. To solve this problem, rank based statistics are employed for comparison over multiple datasets instead. As mentioned in (Demšar, 2006), Friedman test (Friedman, 1937) with the corresponding post-hoc test (Dunn, 1961) is a robust and non-parametric test, so we employ it for statistical significance[7] comparison of different fusion algorithms over multiple datasets.

Comparison of Multiple Fusion Algorithms: In order to perform these tests, the rank statistics are calculated in Table 6. Before analyzing the statistical difference between the proposed RADM and other fusion methods, the Friedman test is used to check whether the average ranks of different fusion methods are statistically significantly different from the mean rank. If this difference is statistically significant, the post-hoc test namely Bonferroni-Dunn test (Dunn, 1961) is employed to determine whether the proposed method statistically significantly outperforms others. The statistics in these tests are computed in details as follows: (please refer to Demšar (2006) for more information about the computation of these statistics)

A. Evaluating whether the performance of all algorithms is the same: Since there are totally nine fusion algorithms for comparison, the mean rank is calculated by $(1+\cdots+9)/9 = 5$. In order to calculate the statistics in the Friedman test, we rank the fusion algorithms for each test set (including different scenarios in KTH dataset) separately as shown in Table 6, i.e. the best performing algorithm gets the rank of 1, the second best rank 2, and so on. In case of ties (like those in Flower, Weizmann and KTH datasets), average ranks are assigned. After that, the average rank of each algorithm is computed and recorded in the last row of Table 6. Under the null-hypothesis, which states that the performance of all the algorithms is the same and equal to a rank 5, the refined Friedman statistic $\mathscr{F}_F$ with nine algorithms and 12 test sets is calculated as follows:

$$\begin{aligned}\chi_F^2 &= \frac{12\times 12}{9\times(9+1)}[(7.0^2+6.2^2+6.5^2+3.3^2+5.0^2\\&\quad +4.1^2+6.0^2+5.7^2+1.2^2)-9\times 5^2] = 43.7 \qquad (34)\\ \mathscr{F}_F &= \frac{(12-1)\times\chi_F^2}{12\times(9-1)-\chi_F^2} = 9.2\end{aligned}$$

where χ_F^2 is the original Friedman statistic and $\mathscr{F}_F$ is the refined one. Since the refined Friedman statistic $\mathscr{F}_F$ follows the $\mathscr{F}$ distribution with $9-1=8$ and $(9-1)\times(12-1)=88$ degrees of freedom and the critical value of $\mathscr{F}(8,88)$ for significance level $\alpha = 0.05$ is $2.1 < 9.2$, we reject the null-hypothesis, which means that the fusion results of different methods are statistically significantly different from the average performance.

[7] It should be noticed that significance in this paper refers to the statistical significance, but not the degree of improvement. In statistics, a result is called statistically significant, if the difference in an experiment is unlikely to be obtained by chance alone and is likely to be the result of a genuine experimental effect (Sheskin, 2011).

Method	Rank Difference
LCDM	2.1
RM	2.9
LP-B	3.8
GRLF	4.5
SSC	4.8
IN	5.0
DN	5.3
Sum	5.8

Table 7 Difference of average ranks between the proposed method and other fusion algorithms

B. Evaluating whether the proposed RADM outperforms others statistically significantly: Since the the null-hypothesis in the Friedman test is rejected, the post-hoc test can be performed to compare the proposed method with others. The critical difference (CD), which measures whether the performance of any two fusion methods is statistically significantly different with each other in terms of the corresponding average ranks, is defined as follows:

$$\mathrm{CD} = c_\alpha \sqrt{\frac{9 \times (9+1)}{6 \times 12}} = 1.1 \times c_\alpha \tag{35}$$

where c_α is the critical values based on the Studentized range statistic divided by $\sqrt{2}$.

It can be found in Demšar (2006) that the critical value in the Bonferroni-Dunn test for statistical significance level $\alpha = 0.05$ is 2.7 when the number of algorithms is nine, so the corresponding CD is $1.1 \times 2.7 = 3.0$. The difference between average ranks of the proposed method and the other eight fusion algorithms is summarized in Table 7. From Table 7, we can see that the proposed RADM performs statistically significantly better than Sum, IN, DN, LP-B, SSC and GRLF, since the rank differences are larger the the CD value 3.0. As mentioned in Section 2, Sum and IN are derived under independent assumption. IN and DN utilize normal distribution, while SSC and GRLF are unsupervised methods in which the empirical classification error has not been minimized to train a discriminative model. Although LP-B is derived without independent assumption, it does not model dependency explicitly as LCDM (Ma et al, 2013a). Consequently, these statistical results show that modeling dependency, relaxing assumption on distribution function and utilizing label information can obtain better and more robust performance for visual recognition.

For significance level $\alpha = 0.10$, the critical value in the Bonferroni-Dunn test is 2.5 with nine comparing algorithms (Demšar, 2006). Thus, the corresponding CD is $1.1 \times 2.5 = 2.8$ for $\alpha = 0.10$. From Table 7, we can see that the proposed RADM performs better than the other fusion methods except LCDM for significance level $\alpha = 0.10$.

Comparing with LCDM: In order to further compare the proposed method with LCDM, we use the sign test in Demšar (2006) to compare the performance of two algorithms. According to Demšar (2006), when the number of test sets is 12, an classifier is significantly better than another, if the performance is better on at least 10 test sets for significance level $\alpha = 0.05$. Since the RADM outperforms LCDM on 11.5 (counting 0.5 for the tie on Weizmann dataset) out of 12 test sets, the improvement by the proposed method is statistically significant compared with that of LCDM by the sign test. These results suggest that the RADM gives better performance robust to different data distributions by better modeling dependency without assumption on the posterior distribution, compared with LCDM.

4.5 Fusion with Discriminative Feature

In this experiment, we evaluate the fusion methods by combining multiple less-discriminative features with a more-discriminative one. From Table 2, we can see that the highest fusion accuracy of 85.56% achieved on Weizmann and 90.28% on KTH dataset are not competitive compared with state-of-the-art methods for human action recognition, e.g. the supervised spatio-temporal neighborhood topology learning (SSTNTL) method (Ma et al, 2013b) achieved higher recognition accuracies of 100% on Weizmann and 94.44% on KTH dataset. On the other hand, as shown in Table 2 and Table 8, the SSTNTL remarkably outperforms the best interest-points based feature extracted in previous experiments, i.e. 82.22% on Weizmann and 78.70% on KTH. Therefore, the SSTNTL can be considered as a much more discriminative feature compared with the eight interest points based features.

The recognition accuracies for combining the SSTNTL and the eight interest-points based features are shown in Table 8. Comparing the results from Tables 2 and 8, we can see that the unsupervised fusion methods, Sum, SSC and GRLF cannot achieve remarkable improvements by combining with a more-discriminative feature, SSTNTL. On the other hand, with the help of label information for training, LCDM, LP-B and RM may discover the more-discriminative feature for fusion. However, their recognition accuracies are lower than that using SSTNTL only as shown in Table. 8. By probabilistically modeling dependency using label information without fusion assumptions, the proposed method not only outperforms other fusion algorithms, but also achieves 100% accuracy on Weizmann dataset together with the SSTNTL feature. On KTH dataset, our method is better than that with the best feature, even it is very discriminative. This convinces that the proposed method can robustly improve recognition performance with multiple more-discriminative and/or less-discriminative features.

	SSTNTL	Sum	IN	DN	LCDM	LP-B	RM	SSC	GRLF	Ours
Weizmann	**100.0**	86.67	87.78	91.11	98.98	98.98	96.67	86.67	85.56	**100.0**
KTH	94.4	85.19	86.57	87.04	92.13	92.13	92.13	86.11	85.65	**96.76**

Table 8 Recognition accuracy (%) comparison by fusing with a more-discriminative feature

5 Conclusion

In this paper, we have designed and proposed a new framework for score-level fusion, which can 1) model dependency and 2) combine multiple pieces of information without any assumption on the posterior distribution. According to the range characteristics of posteriors and convergence of power series, the fusion model is formulated as an analytic function on posteriors multiplied by a constant with respect to the class label. With the analytic fusion model, we give an equivalent condition to the independent assumption and derive the Reduced Analytic Dependency Model (RADM) by the marginal distribution property. Finally, the optimal coefficients in the RADM is learned by incorporating the label information from training data under the regularized least square criterion to ensure the discriminative power.

Since the RADM is developed by general probabilistic properties without assumption on the data distribution, experimental results show that the proposed method gives convincing and robust performance on eight datasets for different tasks of Digit, Flower, Face, Human Action, Object, and Consumer Video recognition. The robust non-parametric statistical tests also demonstrate that the RADM performs statistically significantly better than existing fusion methods for visual recognition. On the other hand, it is validated by the experiments that 1) dependency modeling and 2) fusion without assumption on the posterior distribution, are two elements to ensure that a fusion model can achieve better performance robust to different data distributions.

Although the RADM gives better and more robust performance for visual recognition compared with other score-level fusion methods, feature level contains more information according to the data processing inequality (Cover and Thomas, 2006). Therefore, we will further study the robust feature-level fusion method for visual recognition in the future. Along this direction, joint sparse representation methods (Yuan et al, 2012; Wang et al, 2013; Lan et al, 2014) will be investigated to formulate into a probabilistic dependency modeling framework. And, efficient methods, e.g. Wang et al (2012), could be employed to solved the optimization problem.

Acknowledgements This project was partially supported by the Science Faculty Research Grant of Hong Kong Baptist University, Hong Kong Research Grants Council General Research Fund 212313, National Science Foundation of China Research Grant 61172136. The authors would like to thank the editor and reviewers for their helpful comments which improve the quality of this paper.

Appendix A: Proof of Proposition 1

We first show that conditionally independent condition implies the solution to the equation system (16) is trivial, i.e. $\boldsymbol{a}_{lm0} = \boldsymbol{0}, \boldsymbol{a}_{lm2} = \boldsymbol{0}, \boldsymbol{a}_{lm3} = \boldsymbol{0}, \cdots$ is a trivial solution to equation system (16) for $m = 1, \cdots, M$. If feature representations are independent with each other given class label ω_l, the analytic function $h_l(\boldsymbol{s}_l; \boldsymbol{a}_l)$ becomes equation (3). Rewriting the analytic function in (3) according to the order of s_{lm}, we get

$$h_l(\boldsymbol{s}_l; \boldsymbol{a}_l) = g_{lm1}(\tilde{\boldsymbol{s}}_{lm}; \boldsymbol{a}_{lm1}) s_{lm} \tag{36}$$

where $g_{lm1}(\tilde{\boldsymbol{s}}_{lm}; \boldsymbol{a}_{lm1}) = p_l^{1-M} \prod_{m' \neq m} s_{lm'}$. This equation (36) means that $g_{lmn}(\tilde{\boldsymbol{s}}_{lm}; \boldsymbol{a}_{lmn}) \equiv 0$ or equivalently $\boldsymbol{a}_{lmn} = \boldsymbol{0}$ for $n \neq 1$, i.e. the solution to equation system (16) is trivial.

On the other hand, given the solution to equation system (16) is trivial, we need to show that the analytic function $h_l(\boldsymbol{s}_l; \boldsymbol{a}_l)$ is equal to equation (3). If $\boldsymbol{a}_{lmn} = \boldsymbol{0}$ for $n \neq 1$, then the analytic function $h_l(\boldsymbol{s}_l; \boldsymbol{a}_l)$ can be rewritten as equation (36) for $m = 1, \cdots, M$. This implies each term in the power series h_l contains all variables $s_{l1}, \cdots, s_{lM}$ and the order of each s_{lm} cannot be larger than one. In this case, there is only one non-zero term $\prod_{m=1}^{M} s_{lm}$ in the analytic function h_l. In addition, according to the normalization equation (15), the non-zero term $\prod_{m=1}^{M} s_{lm}$ is normalized by the prior. And the analytic function becomes equation (3). This complete the proof of this proposition.

Appendix B: Derivation for $E_{\text{Dis}}(\boldsymbol{a}, \boldsymbol{q})$

$$\begin{aligned}
& E_{\text{Dis}}(\boldsymbol{a}, \boldsymbol{q}) \\
= & -\theta \sum_{l=1}^{L} \sum_{l' \neq l} \sum_{y_j = \omega_l} q_{jl'} \\
& + \frac{\theta}{2} \sum_{l=1}^{L} \sum_{l' \neq l} \sum_{y_j = \omega_l} ((\boldsymbol{a}_l^T \boldsymbol{z}_{jl} - \boldsymbol{a}_l'^T \boldsymbol{z}_{jl'}) - q_{jl'})^2 \\
= & -\theta \sum_{l=1}^{L} \sum_{l' \neq l} \boldsymbol{q}_{ll'}^T \boldsymbol{1} + \frac{\theta}{2} \sum_{l=1}^{L} \sum_{l' \neq l} \| (Z_{ll}^T \boldsymbol{a}_l - Z_{ll'}^T \boldsymbol{a}_l') - \boldsymbol{q}_{ll'} \|^2 \\
= & -\theta \sum_{l=1}^{L} \boldsymbol{q}_l^T \boldsymbol{1} + \frac{\theta}{2} \sum_{l=1}^{L} (\boldsymbol{a}^T Z_l - \boldsymbol{q}_l^T)(Z_l^T \boldsymbol{a} - \boldsymbol{q}_l) \\
= & \frac{1}{2} \boldsymbol{a}^T H_{\text{Dis}} \boldsymbol{a} + \theta \sum_{l=1}^{L} (\frac{1}{2} \boldsymbol{q}_l^T \boldsymbol{q}_l - \boldsymbol{a}^T Z_l \boldsymbol{q}_l - \boldsymbol{q}_l^T \boldsymbol{1})
\end{aligned}$$

Appendix C: Derivation of the Matrix Formulation for $E(\boldsymbol{a})$

$$
\begin{aligned}
&E(\boldsymbol{a})\\
=&\frac{\sum_{l=1}^{L}\sum_{m=1}^{M}\|\boldsymbol{a}_l^T(\boldsymbol{c}_{lm0},\cdots,\boldsymbol{c}_{lmN})-(b_0,\cdots,b_N)\|^2}{2LM(N+1)}\\
=&\frac{\sum_{l=1}^{L}\sum_{m=1}^{M}(\boldsymbol{a}_l^T C_{lm}-\boldsymbol{b}^T)(C_{lm}^T\boldsymbol{a}_l-\boldsymbol{b})}{2LM(N+1)}\\
=&\frac{\sum_{l=1}^{L}[\boldsymbol{a}_l^T(\sum_{m=1}^{M}C_{lm}C_{lm}^T)\boldsymbol{a}_l-2\boldsymbol{a}_l^T\sum_{m=1}^{M}C_{lm}\boldsymbol{b}+\boldsymbol{b}^T\boldsymbol{b}]}{2LM(N+1)}\\
=&\frac{1}{2}\sum_{l=1}^{L}\boldsymbol{a}_l^T H_l\boldsymbol{a}_l-\sum_{l=1}^{L}\boldsymbol{a}_l^T\boldsymbol{f}_l+\frac{1}{2LM(N+1)}\sum_{l=1}^{L}\boldsymbol{b}^T\boldsymbol{b}\\
=&\frac{1}{2}\boldsymbol{a}^T H\boldsymbol{a}-\boldsymbol{a}^T\boldsymbol{f}+\frac{1}{2LM(N+1)}\sum_{l=1}^{L}\boldsymbol{b}^T\boldsymbol{b}
\end{aligned}
$$

References

Ahonen, T., Hadid, A., and Pietikäinen, M. (2004). Face recognition with local binary patterns. In: *European Conference on Computer Vision*, *Lecture Notes in Computer Science*, vol 3021, pp 469–481

Awais, M., Yan, F., Mikolajczyk, K., and Kittler, J. (2011). Augmented kernel matrix vs classifier fusion for object recognition. In: *British Machine Vision Conference*, pp 60.1–60.11

Belhumeur, P. N., Hespanha, J. P., and Kriegman, D. J. (1997). Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720

Breukelen, M., Duin, R., Tax, D., and Hartog, J. (1998). Handwritten digit recognition by combined classifiers. *Kybernetika*, 34(4):381–386

Canu, S., Grandvalet, Y., Guigue, V., and Rakotomamonjy, A. (2005). SVM and kernel methods matlab toolbox. Perception Systèmes et Information, INSA de Rouen, Rouen, France

Chen, H., and Meer, P. (2005). Robust fusion of uncertain information. *IEEE Transactions on Systems, Man, and Cybernetics–Part B: Cybernetics*, 35(3):578–586

Comaniciu, D. (2003). Robust information fusion using variable-bandwidth density estimation. In: *International Conference of Information Fusion*, vol 2, pp 1303–1309

Cover, T. M., and Thomas, J. A. (2006). *Elements of information theory*. Wiley-Interscience

Dalal, N., and Triggs, B. (2005). Histograms of oriented gradients for human detection. In: *IEEE Conference on Computer Vision and Pattern Recognition*, vol 1, pp 886–8930

Dass, S. C., Nandakumar, K., and Jain, A. K. (2005). A principled approach to score level fusion in multimodal biometric systems. In: *International Conference on Audio- and Video-Based Biometric Person Authentication*, pp 1049–1058

Demiriz, A., Bennett, K. P., and Shawe-Taylor, J. (2002). Linear programming boosting via column generation. *Machine Learning*, 46(1-3):225–254

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30

Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association*, 56:52–64

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A. (2007). The PASCAL Visual Object Classes Challenge 2007 (VOC 2007) Results

Everingham, M., Gool, L. V., Williams, C. K. I., Winn, J., and Zisserman, A. (2010). The PASCAL Visual Object Classes (VOC) challenge. *International Journal of Computer Vision*, 88(2):303–338

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications, Volume I*. Wiley

Fernando, B., Fromont, E., Muselet, D., and Sebban, M. (2012). Discriminative feature fusion for image classification. In: *IEEE Conference on Computer Vision Pattern Recognition*, pp 3434–3441

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32:675–701

Gehler, P., and Nowozin, S. (2009). On feature combination for multiclass object classification. In: *IEEE International Conference on Computer Vision*, pp 221–228

Gorelick, L., Blank, M., Shechtman, E., Irani, M., and Basri, R. (2007). Actions as space-time shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(12):2247–2253

Guillaumin, M., Verbeek, J., and Schmid, C. (2010). Multimodal semi-supervised learning for image classication. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp 902–909

He, H., and Cao, Y. (2012). SSC: A classifier combination method based on signal strength. *IEEE Transactions on Neural Networks and Learning Systems*, 23(7):1100–1117

He, M., Horng, S.-J., Fan, P., Run, R.-S., Chen, R.-J., Lai, J.-L., Khan, M. K., and Sentosa, K. O. (2010). Performance evaluation of score level fusion in multimodal biometric systems. *Pattern Recognition*, 43(5):1789–1800

He, X., Yan, S., Hu, Y., Niyogi, P., and Zhang, H. J. (2005). Face recognition using laplacianfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):328–340

Huber, P. J., and Ronchetti, E. M. (2009). *Robust Statistics, Second Edition*. Wiley

Jain, A., Nandakumar, K., and Ross, A. (2005). Score normalization in multimodal biometric systems. *Pattern Recognition*, 38(12):2270–2285

Jiang, Y.-G., Ye, G., Chang, S.-F., Ellis, D., and Loui, A. C. (2011). Consumer video understanding: A benchmark database and an evaluation of human and machine performance. In: *ACM International Conference on Multimedia Retrieval*, pp 29:1–29:8

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

Krantz, S. G., and Parks, H. R. (2002). *A Primer of Real Analytic Functions*. Birkhäuser

Kuncheva, L. I. (2004). *Combining Pattern Classifiers: Methods and Algorithms*. Wiley

Lan, X., Yuen, P. C., and Ma, A. J. (2014). Multi-cue visual tracking using robust feature-level fusion based on joint sparse representation. In: *IEEE Conference on Computer Vision and Pattern Recognition*

Liu, D., Lai, K.-T., Ye, G., Chen, M.-S., and Chang, S.-F. (2013). Sample specific late fusion for visual category recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp 803–810

Liu, J., McCloskey, S., and Liu, Y. (2012). Local expert forest of score fusion for video event classification. In: *European Conference on Computer Vision*, *Lecture Notes in Computer Science*, vol 7576, pp 397–410

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

Luenberger, D. G., and Ye, Y. (2008). *Linear and Nonlinear Programming, Third Edition*. Springer

Ma, A. J., and Yuen, P. C. (2012). Reduced analytical dependency modeling for classifier fusion. In: *European Conference on Computer Vision*, *Lecture Notes in Computer Science*, vol 7574, pp 792–805

Ma, A. J., Yuen, P. C., and Lai, J.-H. (2013a). Linear dependency modeling for classifier fusion and feature combination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(5):1135–1148

Ma, A. J., Yuen, P. C., Zou, W. W., and Lai, J.-H. (2013b). Supervised spatio-temporal neighborhood topology learning for action recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 23(8):1447–1460

Mikolajczyk, K., and Schmid, C. (2004). Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86

Mittal, A., Zisserman, A., and Torr, P. (2011). Hand detection using multiple proposals. In: *British Machine Vision Conference*, pp 75.1–75.11

Nandakumar, K., Chen, Y., Dass, S. C., and Jain, A. K. (2008). Likelihood ratio based biometric score fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):342–347

Natarajan, P., Wu, S., Vitaladevuni, S., Zhuang, X., Tsakalidis, S., Park, U., Prasad, R., and Natarajan, P. (2012). Multimodal feature fusion for robust event detection in web videos. In: *IEEE Conference on Computer Vision Pattern Recognition*, pp 1298–1305

Nilsback, M.-E., and Zisserman, A. (2006). A visual vocabulary for flower classification. In: *IEEE Conference on Computer Vision and Pattern Recognition*, vol 2, pp 1447–1454

Nilsback, M.-E., and Zisserman, A. (2008). Automated flower classification over a large number of classes. In: *IEEE Indian Conference on Computer Vision, Graphics and Image Processing*, pp 722–729

Oh, S., McCloskey, S., Kim, I., Vahdat, A., Cannons, K., Hajimirsadeghi, H., Mori, G., Perera, A., Pandey, M., and Corso, J. (2014). Multimedia event detection with multimodal feature fusion and temporal concept localization. *Machine Vision and Applications*, 25(1):49–69

Oliva, A., and Torralba, A. (2001). Modeling the shape of the scene: a holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3):145–175

Phillips, P. J., Moon, H., Rizvi, S. A., and Rauss, P. J. (2000). The FERET evaluation methodology for face recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104

Prabhakar, S., and Jain, A. K. (2002). Decision-level fusion in fingerprint verification. *Pattern Recognition*, 35(4):861–874

Ross, A., Nandakumar, K., and Jain, A. K. (2006). *Handbook of Multibiometrics*. Springer

Rudin, W. (1976). *Principles of mathematical analysis*. McGraw-Hill

Scheirer, W., Rocha, A., Micheals, R., and Boult, T. (2010). Robust fusion: Extreme value theory for recognition score normalization. In: *European Conference on Computer Vision*, *Lecture Notes in Computer Science*, vol 6313, pp 481–495

Schuldt, C., Laptev, I., and Caputo, B. (2004). Recognizing human actions: A local SVM approach. In: *IEEE International Conference on Pattern Recognition*, vol 3, pp 32–36

Sheskin, D. J. (2011). *Handbook of Parametric and Nonparametric Statistical Procedures, Fifth Edition*. Chapman and Hall/CRC

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall

Sim, T., Baker, S., and Bsat, M. (2003). The CMU pose, illumination, and expression database. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1615–1618

Tang, K., Yao, B., Fei-Fei, L., and Koller, D. (2013). Combining the right features for complex event recognition. In: *IEEE International Conference on Computer Vision*

Terrades, O. R., Valveny, E., and Tabbone, S. (2009). Optimal classifier fusion in a non-bayesian probabilistic framework. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(9):1630–1644

Toh, K.-A., Tran, Q.-L., and Srinivasan, D. (2004a). Benchmarking a reduced multivariate polynomial pattern classifier. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6):740–755

Toh, K.-A., Yau, W.-Y., and Jiang, X. (2004b). A reduced multivariate polynomial model for multimodal biometrics and classifiers fusion. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(2):224–233

Ueda, N. (2000). Optimal linear combination of neural networks for improving classification performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(2):207–215

Wang, H., Nie, F., and Huang, H. (2013). Heterogeneous visual features fusion via sparse multimodal machine. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp 3097–3102

Wang, J., Kwon, S., and Shim, B. (2012). Generalized orthogonal matching pursuit. *IEEE Transactions on Signal Processing*, 60(12):6202–6216

Ye, G., Liu, D., Jhuo, I.-H., and Chang, S.-F. (2012). Robust late fusion with rank minimization. In: *IEEE Conference on Computer Vision Pattern Recognition*, pp 3021–3028

Yuan, X.-T., Liu, X., and Yan, S. (2012). Visual classification with multitask joint sparse representation. *IEEE Transactions on Image Processing*, 21(10):4349–4360

Zhang, J., Marszałek, M., Lazebnik, S., and Schmid, C. (2007). Local features and kernels for classification of texture and object categories: A comprehensive study. *International Journal of Computer Vision*, 73(2):213–238