Information Fusion in Multimedia Information Retrieval

Jana Kludas, Eric Bruno, and Stephane Marchand-Maillet

University of Geneva, Switzerland
jana.kludas@cui.unige.ch
http://viper.unige.ch/

Abstract. In retrieval, indexing and classification of multimedia data an efficient information fusion of the different modalities is essential for the system's overall performance. Since information fusion, its influence factors and performance improvement boundaries have been lively discussed in the last years in different research communities, we will review their latest findings. They most importantly point out that exploiting the feature's and modality's dependencies will yield to maximal performance. In data analysis and fusion tests with annotated image collections this is undermined.

1 Introduction

The multi modal nature of multimedia data creates an essential need for information fusion for its classification, indexing and retrieval. Fusion has also great impact on other tasks such as object recognition, since all objects exist in multi modal spaces. Information fusion has established itself as an independent research area over the last decades, but a general formal theoretical framework to describe information fusion systems is still missing [14].

One reason for this is the vast number of disparate research areas that utilize and describe some form of information fusion in their context of theory. For example, the concept of data or feature fusion, which forms together with classifier and decision fusion the three main divisions of fusion levels, initially occurred in multi-sensor processing. By now several other research fields found its application useful. Besides the more classical data fusion approaches in robotics, image processing and pattern recognition, the information retrieval community discovered some years ago its power in combining multiple information sources [23].

The roots of classifier and decision fusion can be found in the neural network literature, where the idea of combining neural network outputs was published as early as 1965 [10]. Later its application expanded into other fields like econometrics as forecast combining, machine learning as evidence combination and also information retrieval in page rank aggregation [23].

In opposite to the early application areas of data, classifier and decision fusion, researchers were for a long time unclear about which level of information fusion is to be preferred and more generally, how to design an optimal information fusion strategy for multimedia processing systems.

N. Boujemaa, M. Detyniecki, and A. Nürnberger (Eds.): AMR 2007, LNCS 4918, pp. 147–159, 2008.

This can be seen in recently published approaches that solve similar tasks and nevertheless use different information fusion levels. Examples using classifier fusion are multimedia retrieval [28], multi-modal object recognition [12], multibiometrics [15] and video retrieval [29]. Concerning data fusion, the applications that can be named are multimedia summarization [1], text and image categorization [7], multi-modal image retrieval [27] and web document retrieval [19]. Other problems of interest are the determination of fusion performance improvement compared to single source systems and the investigation of its suspected influence factors like dependency and accuracy of classifiers and data.

Compared to other application fields of information fusion, there is in multimedia a limited understanding of the relations between basic features and abstract content description [26]. Many scientists have approached this problem in the past empirically and also attempted to justify their findings in theoretical frameworks. Lately the information fusion community did important progress in fusion theory that have not yet been considered for multimedia retrieval tasks.

In this paper we give first (Section 2) a review on information fusion in a generic context and what is important in practice for fusion system design. The section includes a discussion on influence factors of fusion effectiveness and the design of an optimal fusion system. In section 3, the task of semantic classification of keyword annotated images is analyzed in order to suggest ways of future research for an appropriate fusion strategy in accordance with the latest findings in information fusion. The paper also confirms this with experimental results.

2 General Overview on Information Fusion

The JDL working group defined information fusion as "an information process that associates, correlates and combines data and information from single or multiple sensors or sources to achieve refined estimates of parameters, characteristics, events and behaviors" [13]. Several classification schemes for information fusion systems have been proposed in literature, whereby the one from the JDL group is probably the most established. This functional model represents components of a fusion system in 4 core levels of fusion (L0-L3) and 2 extension levels (L4,L5). Here L0-L3 are the data or feature association and there above the object, situation, and impact refinement. The extension levels (L4, L5) consist of the process and the user refinement.

In [2], an overview of information fusion scenari in the context of multi modal biometrics is given, which can be easily adapted to general tasks. Hence in information fusion the settings that are possible are: (1) single modality and multiple sensors, (2) single modality and multiple features, (3) single modality and multiple classifiers and (4) multi modalities. Where in the latter case for each modality one of the combinations (1)-(3) can be applied. The multi modal fusion can be done serial, parallel or hierarchical. For completeness reasons, we add a scenario found in [15]: single modality and multiple sample, which is of importance in information retrieval approaches like bagging.

The gain of information fusion is differentiated in [14]: "By combining low level features it is possible to achieve a more abstract or a more precise representation of the world". This difference in the fusion goal is also covered in the Durrant-Whyte classification of information fusion strategies [13], which refers to complementary, cooperative and competitive fusion. In the first case, the information gain results from combining multiple complementary information sources to generate a more complete representation of the world. Here, the overall goal is to exploit the sources diversity or complementarity in the fusion process. The cooperative and competitive fusion provide a reduced overall uncertainty and hence also increased robustness in fusion systems by combining multiple information sources or multiple features of a single source respectively. These latter strategies utilize the redundancy in information sources. Since the sum of complementarity and redundancy of a source equals a constant, it is only possible to optimize a fusion system in favor of the one or the other [3].

In general the benefit of fusion, presuming a proper fusion method, is that the influence of unreliable sources can be lowered compared to reliable ones [18]. This is of a high practical relevance, because during system design it is often not clear how the different features and modalities will perform in real world environments.

Further aspects of information fusion like the system architecture (distributed, centralized) and utilization of certain mathematical tools (probability and evidence theory, fuzzy set and possibility theory, neural networks, linear combination) can be found in an older review on information fusion [11], but their detailed presentation is out of the scope of this paper.

2.1 Information Fusion System Design

Based on the theory presented before in the practice of system design the following points have to be considered: sensors or sources of information, choice of features, level, strategy and architecture of fusion processing and if further background or domain knowledge can be comprised [14]. The choice of sensors and information sources is normally limited by the application. The available sources should be considered in regard to their inhered noise level, cost of computation, diversity in between the set and its general ability to describe and distinguish the aimed at patterns.

During feature selection, one must realize that feature values of different modalities can encounter a spectrum of different feature types: continuous, discrete and even symbolic. That is why modality fusion is more difficult and complex [30], i.e. especially for joint fusion at data level, where a meaningful projection of the data to the result space has to be defined. But also in the case of only continuous features observed from different modalities the information fusion is not trivial, because nonetheless an appropriate normalization has to be applied [2].

The most common location of fusion are at data/feature, classifier/score level and decision level. Hence, a decision between low level or high level fusion must be taken, but also hybrid algorithms that fuse on several levels are possible. An exception is presented in [25], where the authors fuse kernel matrices. In [14] the

authors proved, with the help of their category theory framework, that classifier and decision fusion are just special cases of data fusion. Furthermore they stated that a correct fusion system is always at least as effective as any of its parts, because, due to fusion several sources, more information about the problem is involved. The emphasize is here on 'correct', so inappropriate fusion can lead to a performance decrease. Attention should be payed to the difference between data fusion and data concatenation. The latter is circumventing the data alignment problem and thus is not having the power of data fusion. But it can be an easy and sufficient solution for compatible feature sets.

Many publications so far treated the topic data versus decision fusion by investigating the pros and conc of each fusion type. Data fusion can, due to the data processing inequality, achieve the best performance improvements [17], because at this early stage of processing the most information is available. Complex relations in data can be exploited during fusion, provided that their way of dependence is known. Drawbacks in data and feature fusion are problems due to the 'curse of dimensionality', its computationally expensiveness and that it needs a lot of training data.

The opposite is true for decision fusion. It can be said to be throughout faster because each modality is processed independently which is leading to a dimensionality reduction. Decision fusion is however seen as a very rigid solution, because at this level of processing only limited information is left.

The fusion strategy is mostly determined by the considered application. For example, all sensor integration, image processing, multi modal tracking tasks and the like execute cooperative fusion since they exploit temporal or spatial co-occurrence of feature values.

However, for information retrieval systems the situation is not as trivial. For example, three different effects in rank aggregation tasks can be exploited with fusion [4]:

(1) Skimming effect: the lists include diverse and relevant items
(2) Chorus effect: the lists contain similar and relevant items
(3) Dark Horse effect: unusually accurate result of one source

According to the theory presented in the last section, it is impossible to exploit all effects within one approach because the required complementary (1) and cooperative (2) strategy are contradictory.

The task of multi modal information retrieval and classification, e.g. joint processing of images aligned with texts or annotated with keywords, was approached in the past with success using cooperative strategies like LSI, which uses feature co-occurence matrices, or mixture models, which exploit the feature's joint probabilities [19]. The same is true for complementary ensemble methods, that train classifiers for each modality separately and fuse them afterwards [5].

2.2 Performance Improvement Boundaries

The lack of a formal theory framework for information fusion caused a vibrant discussion in the last years about the influences on fusion results and especially

on theoretical achievable performance improvement boundaries compared to single source systems. Early fusion experiments have shown thorough performance improvement. Later publications accumulated that reported about ambivalent fusion results, mostly, where ensemble classifier were outperformed by the best single classifier. So the information fusion community began to empirically investigate suspected influence factors such as diversity, dependency and accuracy of information sources and classifiers. Based on the experiments explanations for the fusion result ambiguity and mostly application specific upper and lower bounds of performance improvements were found. This section will summarize their findings.

First investigations of these problems were undertaken in competitive fusion on behalf of decorrelated neural network ensembles [20], that outperformed independently trained ones. The overall reduced error is achieved due to negative correlated errors[1] in the neural networks, that average out in combination. [6] confirmed that more diverse classifiers improve the ensemble performance.

The bias-variance decomposition of the mean square error of the fusion result serves as theoretical explanation: more training lowers the bias, but gives rise to variance of the fusion result. The bias-variance-covariance relation is an extension of the former decomposition [16], that shows theoretically that dependencies between classifiers increase the generalization error compared to independent ones. So this strategy achieves a more precise expectation value in the result due to averaging over the inputs.

A theoretical study on complementary fusion [15] applied to multibiometrics found its lower bound of performance improvement for highly correlated modalities and the upper bound for independent ones. This strategy works only for truly complementary tasks, which means that it is aimed at independent patterns in the data, as in a rank aggregation problem where the combined lists contain a significant number of unique relevant documents [22]. Here, the opposite of the bias-variance decomposition for averaging applies: more training rises the bias (ambiguity) and lowers the variance of the result. But the influence of the classifier's bias, variance and their number, affects the fusion result not as much as dependency [10]. For this fusion strategy high level or late fusion is most efficient, since there are no dependencies that can be exploited at data level.

In practice often independence between the fusion inputs is assumed for simplicity reasons or in reference to the diversity of involved modalities. But this is not true i.e. for modalities in multibiometrics [15] and most certainly also not for modalities of other applications, even though it may contain only small dependencies. Applying in this situations high level fusion will hence never yield the maximum theoretical possible fusion performance improvement, because the information reduction caused by processing makes it impossible to exploit data dependencies completely in late fusion.

[1] Negative correlated errors are here referred to as being signals with an opposite developing of their values, not that only negative correlation coefficients are found between the signals.

On the other hand, data level fusion can have blatant disadvantages in practice due to the 'curse-of-dimensionality' [12] and perform badly towards generalization, modeling complexity, computational intensity and need of training data. A solution for the trade off data versus classifier fusion, can be a hybrid system fusing on several levels.

Empirical tests that investigate optimal features in an approach to fuse visual clues from different sources, hence cooperative fusion, showed that they should be redundant in their values, but complementary in their errors [8]. In [17] fusion performance is investigated on behalf of a multi modal binary hypothesis testing problem as e.g. used in multibiometrics. Considering the error exponents for the dependence and independence assumption for the modalities, it is found that the dependent case gives the upper performance improvement bound, and the case of independence the lower.

A comparison of a complementary and cooperative classifier fusion applied to multibiometrics [21] showed a slight performance advantage for the cooperative fusion approach that exploits the modalities dependencies. Admittedly, this gain over the complementary fusion is small, which is due to the little dependencies between the modalities. Furthermore, it needs a lot of training data to estimate the correlation matrix. So in practice there is a trade off between performance improvement and computational cost, whereas the independence assumption often will achieve sufficient results. The authors [21] show as well that class-dependent training of classifiers can help to improve the system's ability to discriminate the classes and hence improve the over all performance.

After having reviewed the fundamentals of information fusion, the next section will analyze first the data of a multi modal classification task, specifically of keyword annotated images. Some simple fusion test undermine the presented information fusion theory and should lead the way to develop an efficient solution to the problem.

3 Data Analysis Towards Effective Multi-media Information Retrieval

Due to the high interest in multimedia processing in the past many multi modal collections have been made available to the research communities. Here 2 examples of them are investigated. The Washington database contains 21 image classes of locations like Australia and Italy, but also semantic concepts like football and cherry trees. Most of the images are manually annotated with a few keywords (1-10). Classes with no annotation were left out of the tests. So we experimented with 16 classes that contained in all 675 images, which are nearly equally distributed over the classes.

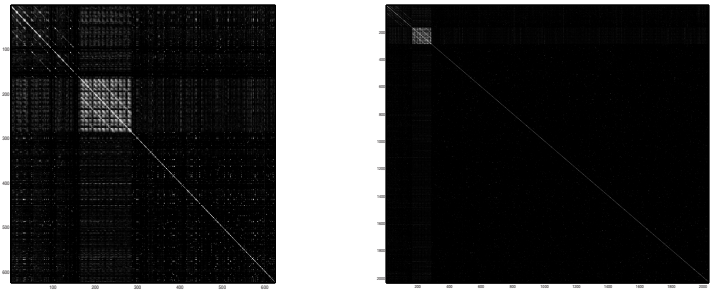
The second collection is a subset of the Corel database, for which [9] created keyword annotations. The final set contains 1159 images in 49 classes, where the images are unequally distributed over the classes. They form similar concepts as in the Washington collection. An important characteristic of this data collection

is that the annotations also include complete nonsense keyword sets, which makes it similar to what one would expect in real world data.

For preprocessing, GIFT features [24] (color and texture histograms) of all images and term-frequency vectors of their annotations were computed. Hence each data sample in the Washington collection is described by 624 features (166 color, 120 texture and 338 text) and in the Corel collection by 2035 features (166 color, 120 texture and 1749 text), where, of course, the textual features are very sparse.

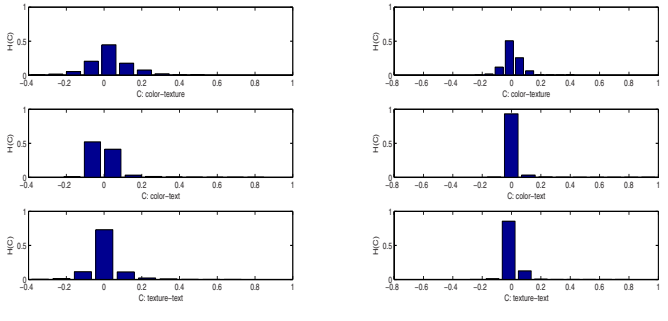
Figure 1 shows the absolute correlation matrices over the feature vectors of the Washington and Corel collection respectively. Bright areas represent feature pairs with high correlation (positive and negative) and darker areas low correlation and hence independence. The significantly correlated feature pairs can be numbered with 17% in the simpler Washington and only 3% in the Corel collection. This tendency of decrease in correlation we expect to be enforced in even noisier real world data.

Since in fusion inter modal dependencies are of a special interest, the average correlation coefficients for both collections are given in table 1. Additionally, the



(a) in Washington (b) in Corel

Fig. 1. Absolute correlation matrices of features in Washington and Corel collection



(a) in Washington (b) in Corel

Fig. 2. Histograms of correlation matrices of Washington and Corel collection

respective histograms are pictured in figure 2. The average decencies turn out to be close to zero, since a nearly equal amount of positive and negative correlation is contained. The maximum correlation coefficients can be found between the histograms and textual features (Wash/Corel color-text: 0.85/0.99, texture-text: 0.74/0.76), whereas between color and texture histogram itself smaller dependencies appear (color-texture: 0.54/0.41). These dependencies between the modalities should be exploited in order to develop an efficient multi modal information fusion system.

Another interesting and not yet explained process is the propagation of the feature dependencies to the classifier outcome. Their dependencies are given in the bottom line of table 1. One can say that there is found again the stronger dependence between the histograms and the text modality. Even though here the dependency between color and texture is not as much smaller as in the feature set.

3.1 Information Fusion Experiments

For the fusion experiments we used a support vector machine (SVM) classifier with rbf kernel as described in [5]. First we compared several simple fusion approaches: (1) hierarchical SVM, which consists of one SVM classifier for each modality and then as well a SVM classifier to fuse their results, (2) concatenated data and SVM, which uses all modalities concatenated as classifier input, (3) averaging the classifier outputs of the modalities, (4) weighted sum, which is the same as (3), but weights the best modality (text) more than the others and (5) majority vote of the classifier outcomes.

The tests were run as one-against-all classifications, where 7 positive and 7 negative samples for the Washington collection and 5 positive and 7 negative samples for the Corel collection were used to train the classifiers. Thereafter their performance is evaluated by applying the classifiers to the remaining data samples. The classification error, false alarm rate (false negative) and miss rate (false positive) are given in percent and are averaged over all classes of each collection.

The experimental results for the Washington and Corel collection are given in table 2 and 3 respectively. For both collections the hierarchical fusion, a learning based approach, performs superior to all other approaches considering the overall classification error, but the classification results in the positive class (false alarm) are throughout better with the simpler fusion strategies. Here, majority vote is

Table 1. Average inter modal dependencies of Washington and Corel collection and dependency found between the classifier outcomes calculated on each modality

	Washington			Corel		
	color-texture	color-text	texture-text	color-texture	color-text	texture-text
av	0.026	0.002	0.0008	0.003	$6.0e-05$	0.001
class	0.284	0.528	0.244	0.111	0.179	0.163

Table 2. Fusion experiments on Washington collection

in %	color	texture	text	hier SVM	concat SVM	averaging	weight sum	major vote
classification error	39.8	41.1	33.2	9.9	44.9	37.9	36.5	35.2
false alarm	9.4	32.8	1.5	23.9	14.6	4.6	3.4	5.6
miss	41.4	41.5	34.9	9.1	46.1	39.6	38.3	36.7

Table 3. Fusion experiments on Corel collection

in %	color	texture	text	hier SVM	concat SVM	averaging	weight sum	major vote
classification error	46.5	47.3	44.5	10.3	56.4	46.5	46.1	45.0
false alarm	34.0	35.3	19.0	58.9	26.9	20.9	18.5	24.5
miss	46.7	47.4	44.9	9.4	56.8	46.9	46.5	45.3

best in the overall classification, whereas weighted sum fusion performs best considering the false alarm rate. The experiment shows as well that the feature concatenation is really a weak fusion strategy and hence performs worst.

The observed results can be said to be ambiguous between the preference of learning and simple fusion approaches. Because the better performance of the simple fusion in discriminating the positive classes is in favor of information retrieval, where it would lead to more relevant documents in the result list. But this better performance compared to the learning approach is more than compensated by the performance in distinguishing the negative class. Here, more tests with improved classifier performance of the modalities should show, if this decreases the miss rate. Then simple fusion methods would be a very interesting approach for large scale problems, because of its low computational complexity.

In the overall performance the simple fusion strategies have unsurprisingly trouble to cope with the badly performing classifier results of the modalities. Text and color based classification work better than the texture based one, but in general they achieve only unreliable results especially for the negative class (miss). Up till now there is no over all successful strategy of fusing dependent, weak classifiers, even though more sophisticated score function approaches like bagging and boosting have been developed and applied with a certain success.

In the second experiment, we investigated how the usage of dependent, by means of correlation, and de-correlated features influences the performance. Since we did not want to search for more or less correlated input features, we created feature subsets for each modality with especially correlated or uncorrelated features. In order to find the correlated ones, we chose features, that have at least once a correlation coefficient with another feature larger than $C > \beta$. As uncorrelated features were chosen that have a maximum absolute correlation coefficient with another feature smaller than the threshold $|C| < \gamma$.

Table 4. Fusion of dependent and independent modalities on Washington collection

in %	color	texture	text	full	dependent	de-correlated
dep: $C > 0.85$ (8/166,58/120,80/338), de-cor: $\|C\| < 0.5$ (16/166, 1/120, 77/330)						
classification error	40.2	43.8	34.8	12.1	20.5	25.2
false alarm	10.4	35.1	5.3	22.5	29.1	36.5
miss	41.7	44.3	36.4	11.5	20.1	24.6
dep: $C > 0.75$ (58/166,71/120,186/338), de-cor: $\|C\| < 0.7$ (38/166, 26/120, 131/330)						
classification error	41.8	45.9	32.8	10.4	15.9	16.5
false alarm	10.3	30.4	4.8	24.3	27.1	30.1
miss	43.5	46.8	34.3	9.5	15.3	15.3

Table 5. Fusion of dependent and independent modalities on Corel collection

in %	color	texture	text	full	dependent	de-correlated
dep: $C > 0.75$ (11/166,65/120,176/1749), de-cor: $\|C\| < 0.5$ (16/166, 1/120, 232/1749)						
classification error	45.8	45.9	45.9	9.0	19.7	19.1
false alarm	32.9	40.6	22.2	60.9	59.3	60.1
miss	46.0	46.1	46.4	8.2	19.1	18.5

The tables 4 and 5 show the results for the Washington and Corel collection respectively, where the number of features selected for each modality (color, texture, text) is given in brackets. To make the results of the correlated and uncorrelated feature subset comparable, a near equality of the over all number of features was tried to achieve, since their number in determining the performance heavily.

The experiments above show a performance advantage for the correlated features in the experiments for the Washington collection. This can also be caused by the different number of features involved in each of the correlated and uncorrelated case. To investigate this further experiments are necessary. But in general both approaches are able to perform a strong dimensionality reduction using a different subset of features (subsets intersect in only up to 10 features). The result of experiment for the Corel collection is not this clear. Here both cases work equally good with a slight advantage for the uncorrelated features.

Concerning the Corel collection another point is interesting to see: the fusion based on the correlated and independent subsets performs in the false alarm rate better than the normal hierarchical SVM. This phenomenon was never observed for the Washington collection. For now we have not a satisfying explanation for this, but we will investigate this further in future.

More extensive tests with e.g. truly differently correlated input features or even artificially created data sets have to be done to prove the influence of correlation and independence to performance improvement of information fusion

Table 6. Correlation class dependent and uncorrelated fusion

in %	color	texture	text	full	dependent	uncorrelated
corel dep: $C > 0.75$, de-cor: $\|C\| < 0.5$ per cur class						
classification error	41.2	44.1	33.9	11.4	14.1	15.5
false alarm	14.7	33.1	3.7	26.5	38.5	40.8
miss	42.5	44.6	35.6	10.4	12.8	14.2

systems. Furthermore other measures of dependency such as mutual information and its influence on the fusion system result should be investigated. Finally fusion approaches that exploit more explicit the features dependencies like LSA and those that consider the accuracy of modalities will be interesting to compare when applied to this problem.

In the last experiment we changed the feature selection rule to chose the feature subsets not according to the correlation coefficients of the whole collection, but according to the correlation found in the currently to distinguish class. With this dynamic, class-dependent selection the feature subsets should contain those features, that are especially helpful in discriminating this class from the negative samples. The results for the Washington collection are presented in table 6.

As it can be seen the class-dependent features selection is not achieving a significant performance improvement, even though the theory presented in section 2 suggests this. We will still investigate this approach further by searching for more efficient and robust ways to adapt the features to the classes, since it is from the sound of theory an appealing approach.

4 Conclusions and Future Work

In retrieval, indexing and classification of multimedia data an efficient information fusion of the different modalities is essential for the system's overall performance. Since information fusion, its influence factors and performance improvement boundaries have been lively discussed in the last years in different research communities, our summarization of their findings will be helpful for all fusion system designers in future.

In our experiments we compared the utilization of correlated and uncorrelated features, because new findings in information theory advises that the better fusion performance can be achieved only with correlated feature. We were able to show that a correlated feature subset for this problem perform slightly better than explicitly de-correlated features. More extensive tests are necessary to underpin these preliminary results and the theoretical findings.

Another promising way to achieve better information fusion performance is to utilize class-dependent classifier settings. This helps in discriminating the positive from the negative classes. Our experiments for now have shown no real improvement in performance.

In general we like to experiment with artificial data where the correlation, diversity and accuracy of each modality as well of their contained features can be set. In this framework a better understanding of the influence factors to the fusion result could be obtained.

References

1. Benitez, A.B., Chang, S.F.: Multimedia knowledge integration, summarization and evaluation. In: Workshop on Multimedia Data Mining, pp. 23–26 (2002)
2. Ross, A., Jain, A.K.: Multimodal biometrics: An overview. In: EUSIPCO Proc. of 12th European Signal Processing Conference (EUSIPCO), pp. 1221–1224 (2004)
3. Fassinut-Mombot, B., Choquel, J.B.: A new probabilistic and entropy fusion approach for management of information sources. Information Fusion 5, 35–47 (2004)
4. Vogt, C.C., Cottrell, G.W.: Fusion via a linear combination of scores. Information Retrieval 1(3), 151–173 (1999)
5. Bruno, E., Moenne-Loccoz, N., Marchand-Maillet, S.: Design of multimodal dissimilarity spaces for retrieval of multimedia documents. IEEE Transaction on Pattern Analysis and Machine Intelligence (to appear, 2008)
6. Brown, G., Yao, X.: On the effectiveness of negative correlation learning. In: First UK Workshop on Computational Intelligence (UKCI 2001) (2001)
7. Chechik, G., Tishby, N.: Extracting relevant structures with side information. In: Advances in Neural Information Processing Systems, vol. 15 (2003)
8. Taylor, G., Kleeman, L.: Fusion of multimodal visual cues for model-based object tracking. In: Australasian Conference on Robotics and Automation (2003)
9. Barnard, K., Johnson, M.: Word sense disambiguation with pictures. Artificial Intelligence 167, 13–30 (2005)
10. Tumer, K., Gosh, J.: Linear order statistics combiners for pattern classification. Combining Artificial Neural Networks, 127–162 (1999)
11. Valet, L., Bolon, P., Mauris, G.: A statistical overview of recent literature in information fusion. In: Proceedings of the Third International Conference on Information Fusion, vol. 1, pp. MOC3/22 – MOC3/29 (2000)
12. Wu, L., Cohen, P.R., Oviatt, S.L.: From members to team to committee - a robust approach to gestural and multimodal recognition. Transactions on Neural Networks 13 (2002)
13. Llinas, J., Bowman, C., Rogova, G., Steinberg, A., Waltz, E., White, F.: Revisiting the jdl data fusion model II. Information Fusion, 1218–1230 (2004)
14. Kokar, M.M., Weyman, J., Tomasik, J.A.: Formalizing classes of information fusion systems. Information Fusion 5, 189–202 (2004)
15. Poh, N., Bengio, S.: How do correlation and variance of base-experts affect fusion in biometric authentication tasks? IEEE Transactions on Acoustics, Speech, and Signal Processing 53, 4384–4396 (2005)
16. Ueda, N., Nakano, R.: Generalization error of ensemble estimators. IEEE International Conference on Neural Networks 1, 90–95 (1996)
17. Koval, O., Pun, T., Voloshynovskiy, S.: Error exponent analysis of person identification based on fusion of dependent/independent modalities. In: Proceedings of SPIE-IS&T Electronic Imaging 2007, Security, Steganography, and Watermarking of Multimedia Contents IX (2007)
18. Aarabi, P., Dasarathy, B.V.: Robust speech processing using multi-sensor, multi-source information fusion - an overview of the state of the art. Information Fusion 5, 77–80 (2004)

19. Zhao, R., Grosky, W.I.: Narrowing the semantic gap - improved text-based web document retrieval using visual features. IEEE Transactions on Multimedia 4(2), 189–200 (2002)
20. Rosen, B.E.: Ensemble learning using decorrelated neural networks. Connections Science 8, 373–384 (1996)
21. Dass, S.C., Jain, A.K., Nandakumar, K.: A principled approach to score level fusion in multimodal biometric systems. In: Proceedings of Audio- and Video-based Biometric Person Authentication (AVBPA), pp. 1049–1058 (2005)
22. Beitzel, S.M., Chowdury, A., Jensen, E.C.: Disproving the fusion hypothesis: An analysis of data fusion via effective information retrieval strategies. In: ACM symposium on Applied computing, pp. 823–827 (2003)
23. Wu, S., McClean, S.: Performance prediction of data fusion for information retrieval. Information Processing and Management 42, 899–915 (2006)
24. Squire, D.M., Müller, W., Müller, H., Raki, J.: Content-based query of image databases, inspirations from text retrieval: inverted files, frequency-based weights and relevance feedback. Pattern Recognition Letters (Selected Papers from The 11th Scandinavian Conference on Image Analysis SCIA 1999) 21(13-14), 1193–1198 (2000)
25. Joachims, T., Shawe-Taylor, J., Cristianini, N.: Composite kernels for hypertext categorization, pp. 250–257. Morgan Kaufmann, San Francisco (2001)
26. Kolenda, T., Winther, O., Hansen, L.K., Larsen, J.: Independent component analysis for understanding multimedia content. Neural Networks for Signal Processing, 757–766 (2002)
27. Westerveld, T., de Vries, A.P.: Multimedia retrieval using multiple examples. In: Enser, P.G.B., Kompatsiaris, Y., O'Connor, N.E., Smeaton, A.F., Smeulders, A.W.M. (eds.) CIVR 2004. LNCS, vol. 3115, Springer, Heidelberg (2004)
28. Wu, Y., Chen-Chuan Chang, K., Chang, E.Y., Smith, J.R.: Optimal multimodal fusion for multimedia data analysis. In: MULTIMEDIA 2004: Proceedings of the 12th annual ACM international conference on Multimedia, pp. 572–579. ACM Press, New York (2004)
29. Yan, R., Hauptmann, A.G.: The combination limit in multimedia retrieval. In: MULTIMEDIA 2003: Proceedings of the eleventh ACM international conference on Multimedia, pp. 339–342. ACM Press, New York (2003)
30. Li, C., Biswas, G.: Unsupervised clustering with mixed numeric and nominal data - a new similarity based agglomerative system. In: International Workshop on AI and Statistics, pp. 327–346 (1997)