Correlation-based Re-ranking for Semantic Concept Detection

Hsin-Yu Ha, Fausto C. Fleites, Shu-Ching Chen
School of Computing and Information Sciences
Florida International University, Miami, FL 33199, USA
{hha001,fflei001,chens}@cs.fiu.edu

Min Chen
Computing and Software Systems School of STEM
University of Washington Bothell, WA 98011, USA
minchen2@u.washington.edu

***Abstract*—Semantic concept detection is among the most important and challenging topics in multimedia research. Its objective is to effectively identify high-level semantic concepts from low-level features for multimedia data analysis and management. In this paper, a novel re-ranking method is proposed based on correlation among concepts to automatically refine detection results and improve detection accuracy. Specifically, multiple correspondence analysis (MCA) is utilized to capture the relationship between a targeted concept and all other semantic concepts. Such relationship is then used as a transaction weight to refine detection ranking scores. To demonstrate its effectiveness in refining semantic concept detection, the proposed re-ranking method is applied to the detection scores of TRECVID 2011 benchmark data set, and its performance is compared with other state-of-the-art re-ranking approaches.**

I. Introduction

With the explosive growth of multimedia applications, the ability to effectively index and retrieve multimedia data becomes increasingly important. Semantic concept detection is widely considered an essential yet challenging step to achieve this goal [1], [2], [3], [4], [5], and has attracted numerous research attentions. One of the typical driven forces is the creation of the TRECVID benchmark by National Institute of Standards and Technology, which aims to boost the researches in semantic media analysis by offering a common video corpus and a common evaluation procedure [6].

Among all the existing work in this area, re-ranking method has been proven effective to improve detection performance when well-designed [7], [8], [9]. Its state-of-the-art process is depicted in Fig. 1. As can be seen, the first step is to preprocess raw multimedia data. Generally, it involves video segmentation, key frame identification, and low-level feature extraction. In the second step, various classification models are trained on training data set and applied to testing data. Then in the third step, the results from different classification models are fused with ranking scores indicating how likely semantic concepts can be detected from each testing instance. Finally, re-ranking process is performed using auxiliary information such as concept ontology to refine final ranking scores. For example, in [7], [8], Concept Association Network (CAN) is used for re-ranking, which captures strong associations among different concepts based on association rule mining (ARM). In [9], co-occurrence among semantic concepts is used to enhance the re-ranking process.

In this work, we propose to leverage the implication among semantic concepts in the re-ranking process. For example,

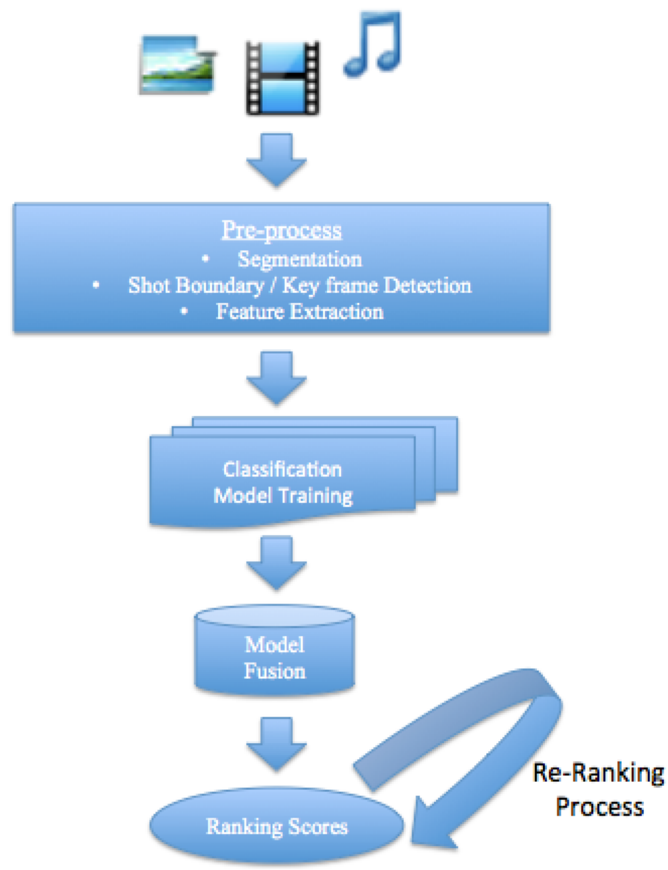

Fig. 1: A general semantic concept detection process

concept “forest” may help indicate the presence of concepts “outdoor” and “plants” instead of “telephone.” In other words, concept correlation is used in our re-ranking method to refine ranking scores produced by other classification models.

The remainder of the paper is organized as follows. Section II discusses the details about the proposed framework. Section III presents the experimental settings and performance evaluation, and Section IV concludes this paper.

II. Proposed Framework

The proposed framework aims to automatically refine classification ranking scores for semantic concept detection. As such ranking scores may be produced by any state-of-the-art classification models and may possibly be continuous data, we first apply a discretization process to convert them into nominal values. Then the concept correlation component is applied to discover how closely two concepts are semantically associated followed by the refined ranking process. The proposed framework is shown in Fig. 2.

A. Ranking Score Discretization

Given a training data set with N instances (i.e., images, video key frames, etc.) and M high-level semantic concepts (such as classroom, airplane, etc.), various classification models are trained and in the process each training instance is associated with a ranking score toward a concept. An example is shown in Table I. As can be seen from the table, the ranking score of training instance 1 toward concept 1 is -1.49 while that of instance 2 is -0.97 toward concept 1. We then use a supervised discretization method called minimum description length (MDL) to discretize the ranking scores into several intervals for each concept and correspondingly we define concept-value pair as follows:

Definition 1. *A **concept-value pair** C^i_j represents the j^{th} ranking score interval of the i^{th} concept, where $1 \leq i \leq M$ and the range of j is determined by the discretization results.*

TABLE I: Concept Ranking Scores

	Concept 1 Ranking Score	Concept 2 Ranking Score	...	Concept M Ranking Score
Instance 1	-1.49	1.08	...	-0.45
Instance 2	-0.97	-0.85	...	-1.32
...	...	...	...	...
Instance N	-0.48	-0.97	...	-1.01

For example, assume the range of ranking scores for concept 1 is discretized into three intervals. They are then denoted as three concept-value pairs: C^1_1, C^1_2, and C^1_3, respectively. Note a concept will be eliminated from further processing if it has only one concept-value pair as it fails to differentiate among instances. Once MDL is applied to all the M concepts, two options are provided to construct an indicator matrix I for a target concept.

- Option 1. Combine concept-value pairs of a single concept (e.g., Concept 1) with the ground truth information of a target concept (see example in Table II.
- Option 2. Combine concept value pairs of all concepts (i.e., Concept 1, Concept 2, ..., Concept M) with the ground truth information of a target concept (see example in TABLE III.

In both cases, number 1 or 0 represents true or false. In other words, the entry for instance 1 in C^1_1 is 1 that means instance 1's initial ranking score (-1.49 as in TABLE I) falls into C^1_1. Similarly, the entry for instance 1 in Column "Target Concept Positive" is 0 that means instance 1 is not labeled with the target concept.

TABLE II: Indicator Matrix of the Concept Ranking Scores for Single Concept

	Concept 1			Target Concept Positive	Target Concept Negative
	C^1_1	C^1_2	C^1_3		
Instance 1	1	0	0	0	1
Instance 2	0		0	1	0
...	...	...	...	...	...
Instance N	0	0	1	1	0

B. Concept Correlation

Multiple correspondence analysis (MCA) has been proven to perform well on many research topics, such as feature selection [10], discretization [11], video semantic concept detection [12], [13], [14], [15], and data pruning [16], which motivates us to apply it in capturing concept correlations for re-ranking process.

Specifically, with indicator matrix I produced in the previous step, a Burt matrix B is constructed as I^TI. The sum of all elements in matrix B, denoted as gt is then obtained using Equation 1, where L is number of columns in matrix I.

$$gt = \sum_{i=0}^{L}\sum_{j=0}^{L} B_{ij}; \tag{1}$$

Thus a normalized Burt matrix NB can be constructed as shown in Equation 2.

$$NB = B/gt; \tag{2}$$

Let $row = \{row_i, i = 1, 2, ...L\}$ and $col = \{col_j, j = 1, 2, ...L\}$ where $row_i = \sum_j NB_{ij}$ and $col_j = \sum_i NBij$, respectively, a centralized matrix Z can be generated following Equation 3.

$$Z = D_{row}^{-1/2}(NB - row * col^T)D_{col}^{-1/2}; \tag{3}$$

Here D_{row} and D_{col} are the diagonal matrices for row and col, respectively. With the application of Single Value Decomposition (SVD), we can then derive eigenvectors from the centralized matrix Z. Because more than 95% of the total variance can be captured by the top two principal components, a subspace is constructed using two eigenvectors with the largest eigenvalues, to which the concept-value pairs in matrix I are projected. Fig. 3 shows an example result using option 1 (i.e., only one concept's concept-value pairs are used), where Pos and Neg represents the positions of positive and negative classes, PC_1 and PC_2 are the x-axis and y-axis corresponding to the top two principal components. In contrast, an example result of option 2 is shown in Fig. 4, where concept-value pairs for all the concepts are used to build the indicator matrix I.

Two parameters are then proposed to represent the concept-value pair's correlation toward the target concept: similarity α and reliability β, which are defined as follows.

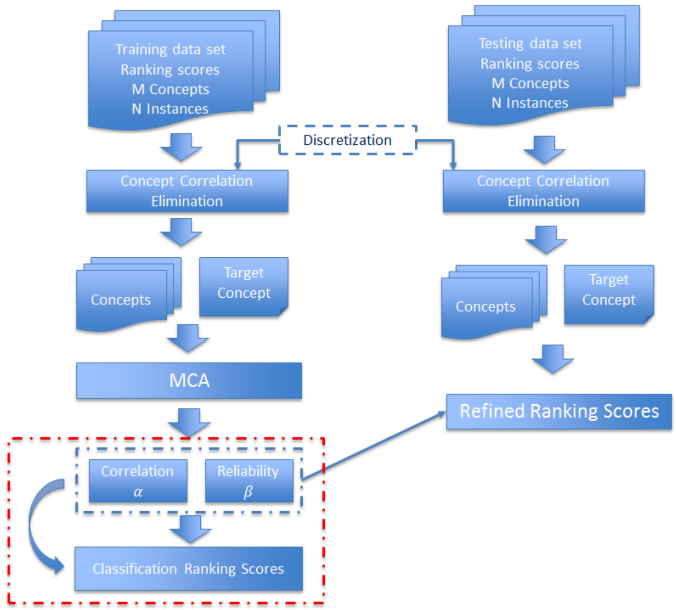

Fig. 2: Proposed Re-Ranking Framework

TABLE III: Indicator Matrix of the Concept Ranking Scores for All Concepts

	Concept 1			Concept 2		...	Concept M		Target Concept Positive	Target Concept Negative
	C_1^1	C_2^1	C_3^1	C_1^2	C_2^2	...	C_1^M	C_2^M		
Instance 1	1	0	0	1	0	...	1	0	0	1
Instance 2	0	1	0	0	1	...	0	1	1	0
...	...	...	...	...	...	...	...	...	...	...
Instance N	0	0	1	0	1	...	1	0	1	0

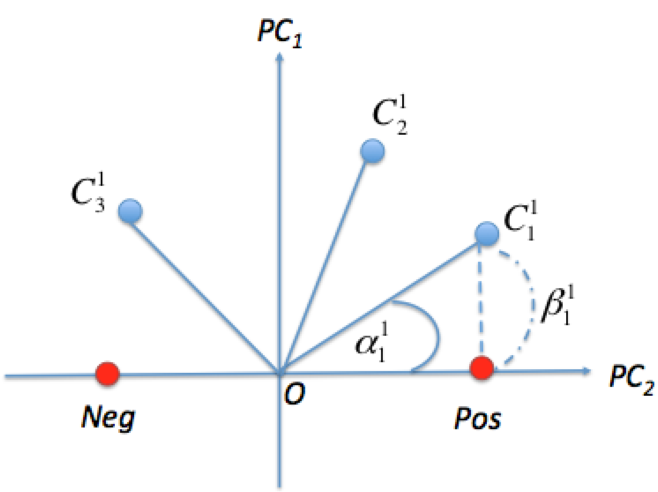

Fig. 3: Projected concept-value pairs using option 1

Definition 2. *Concept Similarity: This is the cosine value of the angle between a concept-value pair C_j^i and the positive class P_OS. Mathematically, it is computed in Equation 4.*

$$\alpha_j^i = \frac{\vec{C_j^i} \cdot \vec{Pos}}{\left|\vec{C_j^i}\right| \left|\vec{Pos}\right|}; \tag{4}$$

Definition 3. *Concept Reliability: This is the Euclidean distance between a concept-value pair C_j^i and the positive class P_OS, which is denoted as β_j^i. It is proposed to take into consideration the dependence between a concept-value pair and the positive class.*

For example, the correlation between concept-value pair C_1^1 and the target concept are captured by α_j^i and β_j^i, as marked in Fig. 3

C. Ranking Score Refinement

To refine the ranking scores, similarity α and reliability β obtained above are used to compute the transaction weight as follows.

Definition 4. $Transaction Weight per concept-value pair:$ *It indicates the correlation between this concept-value pair and the target semantic concept, and is computed in Equation 5. The higher the transaction weight of a concept-value pair C_j^i is, the more likely a target concept will be detected from*

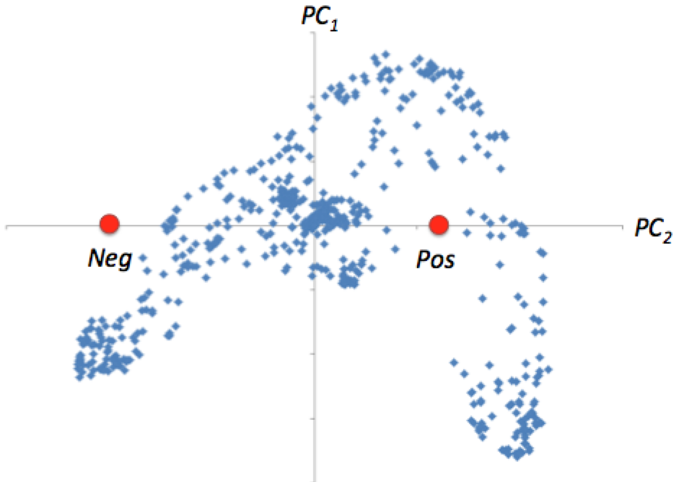

Fig. 4: Projected concept-value pairs using option 2

a testing instance when its ranking score for concept i is discretized into C_j^i.

$$TW_j^i = \alpha_j^i w_j^i + \beta_j^i (1 - w_j^i) \tag{5}$$

Here, w_j^i is a weighting factor. The value of a weighting factor has the range between 0 and 1 with an increment of 0.2, i.e., $w_j^i \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$. The weighting factor corresponding to the highest MAP on the training set is automatically selected for the testing set.

Definition 5. *$TransactionWeightperinstance$: This is considered as the refined ranking score for the testing instances, which is basically the accumulation of the transaction weight for each concept-value pair as depicted in Equation 6.*

$$TransactionWeight_k = \sum_{i=0}^{M} TW_j^i \tag{6}$$

Here, M is the total number of concepts, k represents the index of the testing instance, and j indicates the index of the concept-value pair for concept i.

III. Experiments

In this section, the performance of the proposed re-ranking method is tested and compared with raw ranking scores (i.e., no re-ranking framework is applied so it is called "Baseline") as well as that of three re-ranking frameworks: "Aytar" [17], "DASD" [18], and "AAN" [19]. In Aytar et al., the correlation of a related concept to the target concept is represented by conditional probability and it is further leveraged to enhance the overall detection performance. In Jiang et al., "DASD" is proposed to model the correlation among the concepts as symmetric links with the weights obtained from training labels. Thus, a graphic model is formed to address the potential domain change problem. In Meng et al., "AAN" is built applying association rule mining method to capture the strong association among different concepts and later is used to refine the ranking scores. All three methods focus on applying concept correlation to better improve the re-ranking results without requirng domain knowledge. The performance is evaluated using Mean Average Precision (MAP), which is defined as the arithmetic mean of per-concept average precision and is commonly adopted to evaluate the effectiveness of semantic concept detection.

TABLE IV: Statistics of data set IACC.1.B

Dataset	IACC.1.B
TRECVID Year	2011
No. Concepts	346
No. Instances	137327
Average P/N Ratio	0.003
Average Pos No.	408.32

A. Experimental Setup

In this paper, IACC.1.B data set from TRECVID 2011 is used as testbed. Some basic statistics of the data set are depicted in TABLE IV

The detection scores produced by the Shinoda Lab at Tokyo Institute of Technology on this data set is used as baseline because it performed the best in TRECVID 2011 Semantic Indexing Task. In addition, three-fold cross validation is adopted and the performance is reported by averaging the MAPs obtained from three rounds of classification results.

B. Experimental Results

TABLE V shows the comparison results for all the 130 concepts in terms of MAP value. Different MAP value were calculated based on the numbers of retrieved instances, meaning Top10 MAP represents the MAP value of the top 10 retrieved instances after sorting the ranking scores in descending order. The last column "Overall" shows the MAP value of all the retrieved instances. The higher the MAP value is, the better the semantic concept detection performance are. The rows "Single Concept" and "All Concepts" indicate Option 1 and Option 2 of our proposed re-ranking method, whose performances are compared to "Baseline" (MAP value of the original ranking scores without any re-ranking process), "Aytar", "DASD", and "AAN". The rows "IR_1," "IR_2," ..., "IR_5" show the improvement rates between our "Single Concept" option and the other five methods: Baseline, Aytar, DASD, AAN, and "All Concept" option. Specifically, it is defined in Equation 7.

$$IR_i = \frac{(SingleConcept'sMAP - i^{th}Method'sMAP)}{i^{th}Method'sMAP} \tag{7}$$

and $i = 1, 2, ...5$. For example, IR_1 at TOP10 is computed as $(0.5677 - 0.5218)$ / $0.5218 = 4.59\%$.

As can be seen, Option 1 of our proposed method (i.e., "Single Concept") consistently improves the raw ranking scores and outperforms all the re-ranking methods across all different retrieved levels in terms of MAP value, as also depicted in Fig. 5. This clearly shows the effectiveness of

TABLE V: The MAP values of 130 concepts in IACC.1.B for different number of retrieved instances with the improvement rate of the proposed method using single concept against other re-ranking methods

MAP Retrieved Level	TOP10	TOP20	TOP40	TOP60	TOP80	TOP100	Overall
Baseline	0.5218	0.4898	0.4481	0.4212	0.3999	0.3845	0.1382
Aytar	0.4600	0.4304	0.4075	0.3925	0.3806	0.3654	0.1363
DASD	0.4637	0.4561	0.4240	0.4063	0.3903	0.3743	0.1397
AAN	0.5491	0.5143	0.4709	0.4428	0.4211	0.4051	0.1452
All Concepts	0.5154	0.4722	0.4256	0.3999	0.3801	0.3642	0.1665
Single Concept	0.5677	0.5349	0.4881	0.4658	0.4431	0.4207	0.1881
Improvement R1	4.59%	4.51%	4.00%	4.46%	4.32%	3.62%	4.99%
Improvement R2	10.77%	10.45%	8.06%	7.33%	6.25%	5.13%	5.18%
Improvement R3	10.40%	7.88%	6.41%	5.95%	5.28%	4.64%	4.84%
Improvement R4	1.86%	2.06%	1.72%	2.30%	2.20%	1.56%	4.29%
Improvement R5	5.23%	6.27%	6.25%	6.59%	6.30%	5.65%	2.16%

our proposed method in discovering the correlation between concepts and in using such correlation to help re-rank detection scores. On the other hand, Option 2 (i.e., "All Concepts") does not produce comparative results against all other methods except for "Aytar." It shows that when considering all the concept at the same time, the correlation might be affected by some irrelevant concepts. Nevertheless, as we can see, its overall MAP value is better than the other three re-ranking frameworks. It indicates it in fact greatly helps data elements with low classification scores (i.e., they do not appear in the TopK) to obtain correct class labels, which may be beneficial in some scenarios.

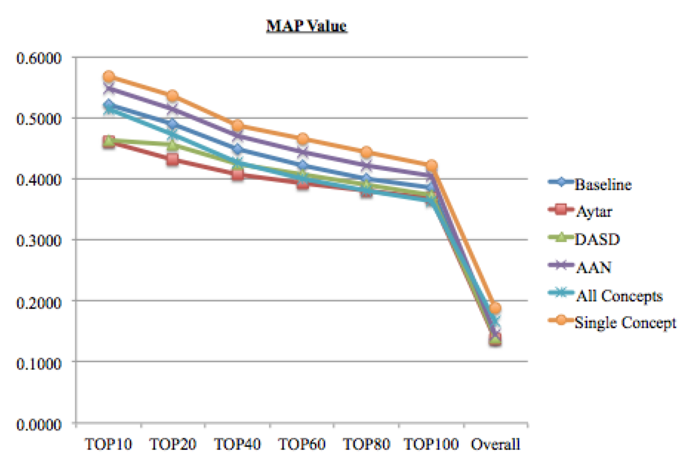

Fig. 5: The MAP values of 130 concepts in IACC.1.B for different number of retrieved instances using one concept against other re-ranking methods

IV. CONCLUSIONS

The paper proposes a re-ranking framework that utilizes concept correlation to automatically refine ranking scores for semantic concept detection. Specifically, multiple correspondence analysis (MCA) is applied to capture concept correlation. In the process, two parameters $similarity$ and $reliability$ are modeled to compute transaction weight which is then translated as the refined ranking score. In the experiments, the ranking scores of TRECVID 2011 IACC.1.B data set is used as baseline and the performance of the proposed methods is compared with that of three state-of-the-art re-ranking frameworks in terms of how well the ranking scores are refined. It shows that Option 1 of our proposed method outperforms other other re-ranking methods at all the retrieval levels while Option 2 is more suitable where all instances need to be retrieved.

In the future, we will first carefully identify the strength of the proposed two options. For example, how well each of them can handle concept with only few positive instances. Then, we will study the possibility of combining results from these two options to further enhance the final results. We will also work on negative correlation and identify concept subsets with higher correlation toward the target concept.

ACKNOWLEDGMENT

This research was supported in part by the U.S. Department of Homeland Security under grant Award Number 2010-ST-062-000039, the U.S. Department of Homeland Security's VACCINE Center under Award Number 2009-ST-061-CI0001, and NSF HRD-0833093.

REFERENCES

[1] Shu-Ching Chen, Mei-Ling Shyu, Chengcui Zhang, Lin Luo, and Min Chen, "Detection of soccer goal shots using joint multimedia features and classification rules," *MDM/KDD*, vol. 3, 2003.

[2] Shu-Ching Chen, Mei-Ling Shyu, Min Chen, and Chengcui Zhang, "A decision tree-based multimodal data mining framework for soccer goal detection," in *Multimedia and Expo, 2004. ICME'04. 2004 IEEE International Conference on*. IEEE, 2004, vol. 1, pp. 265–268.

[3] Shu-Ching Chen, Mei-Ling Shyu, Chengcui Zhang, and Min Chen, "A multimodal data mining framework for soccer goal detection based on decision tree logic," *International Journal of Computer Applications in Technology*, vol. 27, no. 4, pp. 312–323, 2006.

[4] Min Chen, Shu-Ching Chen, Mei-Ling Shyu, and Kasun Wickramaratna, "Semantic event detection via multimodal data mining," *Signal Processing Magazine, IEEE*, vol. 23, no. 2, pp. 38–46, 2006.

[5] Mei-Ling Shyu, Zongxing Xie, Min Chen, and Shu-Ching Chen, "Video semantic event/concept detection using a subspace-based multimedia data mining framework," *Multimedia, IEEE Transactions on*, vol. 10, no. 2, pp. 252–259, 2008.

[6] Alan F Smeaton, Paul Over, and Wessel Kraaij, "Evaluation campaigns and trecvid," in *Proceedings of the 8th ACM international workshop on Multimedia information retrieval*. ACM, 2006, pp. 321–330.

[7] Tao Meng and Mei-Ling Shyu, "Leveraging concept association network for multimedia rare concept mining and retrieval," in *Multimedia and Expo (ICME), 2012 IEEE International Conference on*. IEEE, 2012, pp. 860–865.

[8] Tao Meng and Mei-Ling Shyu, "Concept-concept association information integration and multi-model collaboration for multimedia semantic concept detection," *Information Systems Frontiers*, pp. 1–13, 2013.

[9] Chao Chen, Lin Lin, and Mei-Ling Shyu, "Utilization of co-occurrence relationships between semantic concepts in re-ranking for information retrieval," in *Multimedia (ISM), 2011 IEEE International Symposium on*. IEEE, 2011, pp. 53–60.

[10] Qiusha Zhu, Lin Lin, Mei-Ling Shyu, and Shu-Ching Chen, "Feature selection using correlation and reliability based scoring metric for video semantic detection," in *Semantic Computing (ICSC), 2010 IEEE Fourth International Conference on*. IEEE, 2010, pp. 462–469.

[11] Qiusha Zhu, Lin Lin, Mei-Ling Shyu, and Shu-Ching Chen, "Effective supervised discretization for classification based on correlation maximization," in *Information Reuse and Integration (IRI), 2011 IEEE International Conference on*. IEEE, 2011, pp. 390–395.

[12] Lin Lin, Mei-Ling Shyu, and Shu-Ching Chen, "Association rule mining with a correlation–based interestingness measure for video semantic concept detection," *International Journal of Information and Decision Sciences*, vol. 4, no. 2, pp. 199–216, 2012.

[13] Lin Lin and Mei-Ling Shyu, "Effective and efficient video high-level semantic retrieval using associations and correlations," *International Journal of Semantic Computing*, vol. 3, no. 04, pp. 421–444, 2009.

[14] Lin Lin and Mei-Ling Shyu, "Weighted association rule mining for video semantic detection," *International Journal of Multimedia Data Engineering and Management (IJMDEM)*, vol. 1, no. 1, pp. 37–54, 2010.

[15] Lin Lin, Guy Ravitz, Mei-Ling Shyu, and Shu-Ching Chen, "Correlation-based video semantic concept detection using multiple correspondence analysis," in *Multimedia, 2008. ISM 2008. Tenth IEEE International Symposium on*. IEEE, 2008, pp. 316–321.

[16] Lin Lin, Mei-Ling Shyu, and Shu-Ching Chen, "Enhancing concept detection by pruning data with mca-based transaction weights," in *Multimedia, 2009. ISM'09. 11th IEEE International Symposium on*. IEEE, 2009, pp. 304–311.

[17] Yusuf Aytar, Omer Bilal Orhan, and Mubarak Shah, "Improving semantic concept detection and retrieval using contextual estimates," in *Multimedia and Expo, 2007 IEEE International Conference on*. IEEE, 2007, pp. 536–539.

[18] Yu-Gang Jiang, Jun Wang, Shih-Fu Chang, and Chong-Wah Ngo, "Domain adaptive semantic diffusion for large scale context-based video annotation," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 1420–1427.

[19] Tao Meng, "Association affinity network based multi-model collaboration for multimedia big data management and retrieval," 2013.