Enhancing Rare Class Mining in Multimedia Big Data by Concept Correlation

Yilin Yan and Mei-Ling Shyu
Department of Electrical and Computer Engineering
University of Miami, Coral Gables, FL, USA
y.yan4@umiami.edu, shyu@miami.edu

***Abstract*—The development in information science has enabled an explosive growth of data, which attracts more and more researchers to engage in the field of big data analytics. Noticeably, in many real-world applications, large amounts of data are imbalanced data since the events of interests occur infrequently. However, the detection of these events is such an important research problem and has attracted significant research efforts as lots of real-world big data sets have skewed class distributions. Despite extensive research efforts, rare class mining remains one of the most challenging problems in information science, especially for multimedia big data. Though inter-concept correlations have been utilized to address this issue recently, the very small number of instances in the minority class often lead to the detection of imprecise correlations and unsatisfactory classification results. This paper proposes a novel concept correlation analysis strategy framework using the correlations between the retrieval scores and labels. By integrating the correlation information, the proposed framework can help imbalance data classification and enhance rare class (or concept) mining even with trivial scores from the minority class. Experimental results on the TRECVID multimedia big benchmark data set demonstrate the effectiveness of the proposed framework with promising performance.**

***Keywords*-Imbalanced data; Multimedia big data; Rare class mining; Concept correlation; Information integration**

I. Introduction

Massive amounts of multimedia data are generated and available on different kinds of devices via Internet everyday, while the problem of mining new and useful knowledge from these big data efficiently has attracted significant research efforts [1][2][3][4]. Among them, rare event mining from imbalanced data has become more and more important as lots of applications do not have uniform class distributions [5][6]. That is, the majority of the data instances belong to one class and far fewer data instances belong to the others. In such a data set, the classes with more data instances are called the majority classes; while the ones with far fewer data instances are called the minority classes. In many real-world applications, the data instances in a minority class actually represent the concept of interest like unusual events in surveillance, intrusions in network security, etc.

Since most classifiers are modeled by exploring data statistics, as a result, they may be biased towards the majority classes and hence show very poor classification accuracy on the minority classes. However, compared to the data instances of the majority class, the data instances of the minority class are usually more important and more interesting in a wide range of applications, including the detection of cancer cells, fault prevention of components, metal fatigue detection, etc. The same observation applies to multimedia semantic concept mining and retrieval [7][8][9][10], which is one of the centric research tasks in content-based information retrieval [11][12][13][14]. It focuses on mining semantic concepts such as “person”, “vehicle”, and “sky” from the multimedia data directly.

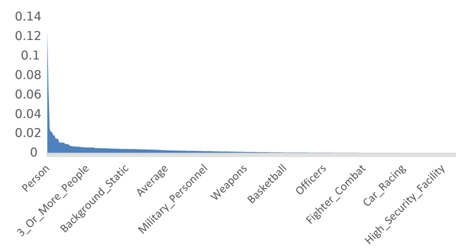

Figure 1. Positive to Negative (P/N) Ratios for some rare classes/concepts in the TRECVID data set

As mentioned earlier, most of the important events of interests are rare classes/concepts. For example, the TREC Video Retrieval (TRECVID) data set [15] includes a large amount of videos collected from the Internet and other sources by National Institute of Standards and Technology (NIST). This multimedia big data set has many concepts that are considered imbalanced and its average P/N (positive to negative) ratio is only 0.003 (as shown in Figure 1). An extreme example is the concept “Yasser Arafat” which contains only one keyframe (video shot), making semantic concept retrieval extremely hard. The purpose of this paper is to effectively detect such rare semantic concepts from multimedia big data sets.

In addition, a video shot usually contains multiply concepts which are correlated in real-world multimedia data sets, either positively or negatively. In other words, some concepts co-occur more frequently, e.g., sea and whale; while others rarely co-occur, e.g., sky and meeting. Such correlations can provide important context cues to help detect the concepts [16][17][18]. By building a semantic concept hierarchy and using inter-concept correlations, we propose a novel rare concept detection framework in this paper.

Experimental results on TRECVID 2015 semantic indexing (SIN) data set demonstrate that the proposed framework gives promising performance, comparing to several state-of-the-art approaches.

The rest of this paper is organized as follows. In Section 2, previous work on imbalance data classification is discussed and various types of correlations are introduced as well. In Section 3, we show how to calculate the correlations between concepts as well as how to connect them in a hierarchy. Section 4 describes a novel idea of enhancing rare concept detection using the correlation between the retrieval scores and labels. Section 5 shows how to setup the framework and compares the results of the proposed system on the TRECVID data set. Finally, Section 6 draws the conclusion.

II. Related work

A. Imbalanced Data Classification

Sampling-based approaches are the most popular classification algorithms for imbalanced data sets. Among them, oversampling and undersampling methodologies have received significant attentions to counter the effect of imbalanced data sets and present the viewpoints on the usefulness of oversampling versus downsampling [19] for imbalanced data sets, though sometimes they are conflicting. The ideas are simple in oversampling, i.e., more positive data instances are somehow generated to make the data set balanced. The problem of oversampling is its tendency to overfit. Comparatively, downsampling is to select a part of the majority data instances to build a model with a similar number of positive samples. Although downsampling is somewhat efficient as it uses only a subset of the majority class, many data instances in the majority class are ignored and may result in the loss of information.

Other than sampling-based approaches, algorithm oriented approaches mainly focused on studying and modifying the training algorithms to achieve better performance in imbalanced data classification. For instance, cost-sensitive learning methods try to maximize the loss functions associated with a data set to improve the classification performance. These learning methods are motivated by the observation that most real-world applications do not have uniform costs for misclassifications. The actual costs associated with each kind of errors are typically unknown, so these methods need to determine the cost matrix based on the data and apply it to the learning stage. A closely related idea to cost-sensitive learners is to shift the bias of a machine to favor the minority class [20]. Though some studies have shown their potential in improving classification performance on imbalanced data, they are far from extensive or systematic.

B. Building Hierarchies for Classes

Organizing hierarchies for semantic concept retrieval and event detection has been investigated by many research groups. Wang et al. [21] proposed a hierarchical context model that simultaneously exploits contexts at all three levels and systematically incorporate them into event recognition. To tackle the learning and inference challenges that were brought in by the model hierarchy, they developed complete learning and inference algorithms for their hierarchical context model based on the variational Bayes method. The problem of general event classification from uncontrolled YouTube videos was investigated [22]. They proposed a text mining pipeline to automatically discover a collection of video event categories, and employed the WordNet hierarchy to refine the category selection.

Some recent researches on event detection focus on relations among concept classes to re-rank the concept scores. The selection of event-specific concepts based on the similarity to a textual event description had shown to yield effective event detection results without positive examples [23]. Tao et al. [24] proved that concept-concept association can be effective in bridging the semantic gap in multimedia data [25][26][27]. They proposed a concept-concept association information integration and multi-model collaboration framework to enhance high-level semantic concept detection from the multimedia data.

III. Building the Hierarchical Model

While some previous work considers that connectivity between classes provides information about their semantic relationship, most of them utilize the hierarchical relationship from the data provider for combining classes to generate the re-organized hierarchies [28][29][30]. These relationships are mainly generated manually, which may lead to biases and not suitable for big data sets. Though some big data sets are provided with the hierarchies (such as the WordNet hierarchy), it is almost impossible to include all existing classes. In addition, when there are new classes, building a new hierarchy or updating the current one is inevitable.

To build a hierarchical model for all concepts in our framework, conditional probabilities are first computed. Let C_{parent} be a parent concept, C_{child} represent a child concept, C^{+}_{parent} denote the positive collection of C_{parent}, and C^{+}_{child} represent the positive collection of C_{child}. Furthermore, let $P(.)$ denote the probability. If C_{parent} is the parent of C_{child}, the occurrence of C_{child} should imply the occurrence of C_{parent}. For instance, if a video shot contains the concept "boy", it definitely includes the concept "person" as well, unless the ground truth is incorrect. In this example, "boy" is a child concept while "person" is a parent concept.

Considering the fact that videos in data sets are labeled either manually or by certain automatic labeling techniques, the ground truth is not always correct. Also, in many cases, the concept pairs have the parent-child relationships (such as "outdoor" and "rowboat"), which should be included. Therefore, a threshold is used to determine such a relationship. Based on our empirical studies, the threshold is set to

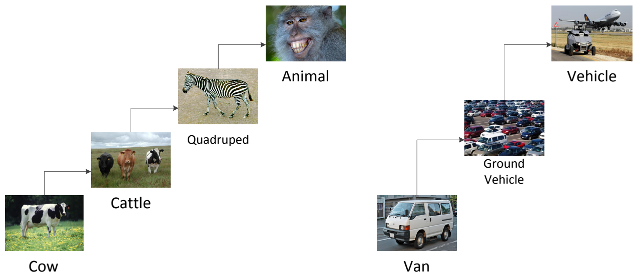

Figure 2. Hierarchies examples - Parents.

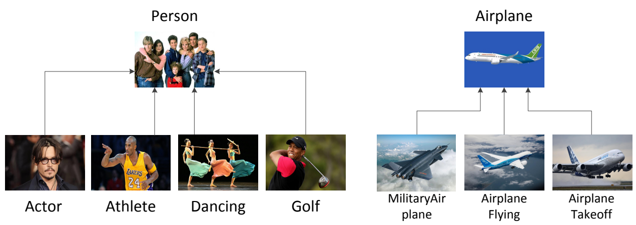

Figure 3. Hierarchy examples - Siblings.

0.9 in our proposed framework, which can be adjusted with different data sets and the number of positive instances of the rare concept for better performance.

$$P(C^{+}_{parent}|C^{+}_{child}) > 0.9. \quad (1)$$

After filtering all the parent-child concept pairs, we start to build the hierarchy model of all concepts in a bottom-up manner, i.e., from the leaf nodes. If a concept has no child but at least one parent, it is considered as a leaf node and is added to the initial model. Then, its "direct" parent is included as the parent node. For examples, a lizard is a reptile and an animal. Although the occurrence of a lizard implies the occurrence of an animal, the concepts "reptile" and "animal" also have the parent-child relationship. Thus, "reptile" is first included as a parent node and then follows by "animal". If a parent concept has no parent, it will be finally considered as a head (root) node. These operations are shown in Figure 2.

Afterward, we can combine the branches into a tree and thus find the siblings of the child concepts as given in Figure 3. Please note that different hierarchies would be generated based on different data sets and we build a tree for each target concept. In the aforementioned instance, if the concept "reptile" is removed, "animal" could be the parent of "lizard" in the updated hierarchy. The more concepts included, the more complete the hierarchy model would be. Though the model can never be perfect, it is suitable for the particular data set. Since only rare concepts are targeted, the cases when the concepts of multiple labels overlapping with each other do not occur in our model.

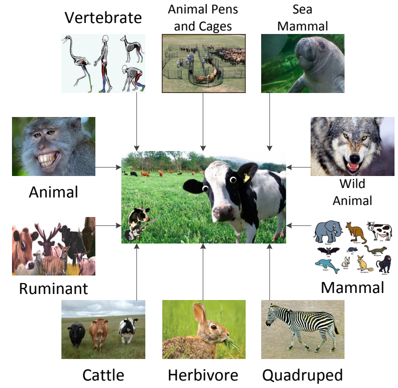

Figure 4. Top ten related concepts that support the rare concept "cow".

IV. Enhancing Prediction Scores of Rare Concepts

A. Score-based Correlation Generation

To our best knowledge, all previous studies including the aforementioned conditional probability approaces calculate the correlations and build the hierarchical models using the label information in the training data, i.e., the occurrence or non-occurrence of the concepts. While some groups used such information to leverage the retrieval scores, one main problem remains. That is, the high correlations between concepts do not necessarily lead to the high correlation between the concepts and prediction (detection) scores, since the scores are not always good, especially for rare concepts. This means that using those concepts' correlation information for enhancement may cause contrary results. As an example, the concept "flood" should have a positive correlation with the concept "natural disaster". However, with the bad prediction scores, "flood" does not really help the retrieval of "natural disaster" in the current data set.

Another issue is that the detection scores of rare concepts can be relatively imprecise. In the TRECVID data set, there are only 6 out of the total of 137,272 video shots that include the concept "cow". Most of the classifiers cannot get acceptable prediction scores for these rare concepts albeit with such a big training data set. Therefore, in this paper, we propose a framework to enhance the prediction scores of the rare concepts using Pearson correlations (ρ) from both the label information and score information [31].

Let C_T be the label information of a target (rare) concept, and S_R be the prediction score of a related (support) concept. Take the aforementioned example. In order to enhance the prediction score of the rare concept "cow", all $\rho_{(C_T, S_R)}$ are calculated and ranked, where T is the concept "cow" and $R = 1, 2, \cdots, N$ (where N is the number of concepts). The top ten related concepts are shown in Figure 4, which means the prediction scores of these concepts are helpful to enhance the prediction score of the concept "cow".

As shown in Figure 4, the top ten related concepts are “Herbivore”, “Ruminant”, “Mammal”, “Quadruped”, “Wild Animal”, “Vertebrate”, “Animal”, “Animal Pens And Cages”, “Sea Mammal”, and “Cattle”, respectively. Clearly, most of them are reasonable at the first glance expect “Sea Mammal”. Nevertheless, the shapes of some sea mammals are similar to those of the cows. Especially, one common kind of sea mammal, manatee, is known as “sea cow”. This highlights another advantage of the proposed framework, which can find the potentially related concepts. Figure 4 also implies the prediction score of “cow” itself is imprecise and will not be integrated for the enhancement.

B. Score Integration

Figure 5 shows an overview of our proposed framework, including a training phase and a testing phase. The data set is first split into two parts, namely a training set and a testing set. For the training phase, conditional probabilities are calculated to build a hierarchical model for all concepts from the training label information. Next, for all the video shots and N concepts, N concept detection models are trained such that for each video shot, the n^{th} model outputs a score measuring the likelihood that concept n exists in that video shot. For this part, all kinds of classifiers can be employed to generate different prediction scores, which may lead to different score-based correlations from the same data set. After ranking them by their correlation values, for a target concept C_T, only the related concepts connecting to this target concept in the hierarchical model (built in Section III) are kept. In this paper, the target concepts are the rare concepts. The scores of the top ten related concepts are used to train an integration model for re-ranking using a discriminant analysis classifier, which is based on a Gaussian mixture model (GMM).

In the testing phase, each testing video shot of the target concept is plugged into all concept detection models to generate the corresponding scores for the related concepts chosen. These scores are then input to the trained score integration model to generate a new set of re-ranked scores. Please note that the scores of the target concept may or may not be used, as shown in Figure 4, depending on whether they are chosen in the training phase or not. Finally, the new output scores are evaluated.

V. Experiments

A. Data Set

To setup the experiment, the IACC.1.B data set from the TRECVID 2015 benchmark [32] for the semantic indexing (SIN) task is used, which aims to detect the semantic concepts contained within a video shot. The data set is drawn from the IACC.1 collection with videos whose durations are between 10 seconds and 3.5 minutes with a total number of 137,272 video shots. It is a multimedia big data set [33], and essential for the retrieval, categorization, and other video exploitations. As mentioned earlier, there are challenges in the SIN task, such as data imbalance, scalability, and semantic gap [34][35].

Given the multimedia big data set, the master shot reference, and the concept/feature definitions, the proposed framework is employed for score integration and enhancement. Totally, 346 concepts are selected including many common semantic concepts like “cloud”, “island”, and “ocean”. The list of concepts and detailed explanations can be found in [15]. In this paper, we directly use the detection scores from the DVMM Lab of Columbia University [36], who ranked the first several years in the TRECVID competition.

B. Experimental Results

In this experiment, 20 most rare concepts are chosen with an average P/N ratio of 0.0001. Among the 68,663 video shots in the training data set, each of them does not have more than 10 video shots in the data set. These 20 concepts are: “Car Crash”, “Cigar Boats”, “Crustacean”, “High Security Facility”, “Helicopter Hovering”, “Cetacean”, “Military Buildings”, “Rpg”, “Prisoner”, “Police Truck”, “Colin Powell”, “Earthquake”, “Oil Drilling Site”, “Rescue Helicopter”, “Dolphin”, “Security Checkpoint”, “Fire Truck”, “Whale”, “Cows”, and “Yasser Arafat”.

In this paper, we use the mean average precision (MAP) value as a metric which is widely used in multimedia concept/information retrieval. The experimental results are shown in Table I. The “Base” one is calculated using the raw scores directly from the classifiers in [36]. Although these scores are the best prediction scores of the data set, they are still bad because of the extremely skewed distributions. Different classifiers including Naive Bayes, Random Forrest, and Discriminant Analysis Classifier are used to re-rank these scores and all target on these imbalanced concepts. The results clearly show that if the target concepts are extremely rare, re-ranking the scores directly will not help enhance the results. Comparatively, albeit with the imprecise raw scores, the propose framework can successfully enhance the results, as demonstrated in Table I.

VI. Conclusion and Future work

It is challenging to obtain reasonable classification accuracies when the target concept is rare, since the data instances in the majority classes usually overshadow those in the minority classes. In this paper, a novel concept correlation analysis strategy framework using the correlation between the retrieval scores and labels is proposed to enhance rare class/concept mining. The experimental results on a multimedia big data set clearly show the effectiveness of the proposed framework and how it can successfully enhance the prediction scores of the chosen rare concepts.

In the future, the co-existence of multiple tree hierarchies when building them will be considered. Furthermore, under

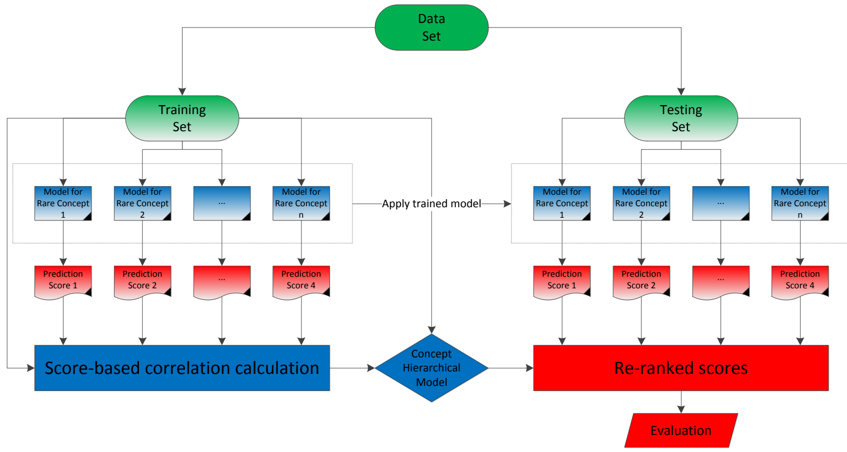

Figure 5. The proposed framework.

Table I
EXPERIMENTAL RESULTS.

Framework	MAP@10	MAP@20	MAP@50	MAP@100	MAP@200	MAP@500	MAP@1000
Base	0.04464	0.04382	0.03123	0.03175	0.03220	0.03020	0.02034
Naive Bayes	0.00556	0.00572	0.00723	0.00577	0.00577	0.00518	0.00528
Random Forrest	0.03750	0.04083	0.02972	0.02154	0.02154	0.02154	0.02154
Discriminant Analysis	0.02429	0.03815	0.03361	0.03074	0.03017	0.02688	0.02593
Proposed Framework	**0.11298**	**0.11298**	**0.08560**	**0.07329**	**0.07113**	**0.06142**	**0.05553**

the scenario of manually labeling the videos, user subjectivity can lead to totally different tree structures. Modeling such user subjectivity is needed to make the proposed framework more robust and will be investigated.

REFERENCES

[1] S.-C. Chen, M.-L. Shyu, and R. Kashyap, "Augmented transition network as a semantic model for video data," *International Journal of Networking and Information Systems*, vol. 3, no. 1, pp. 9–25, 2000.

[2] S.-C. Chen, M.-L. Shyu, and C. Zhang, "An intelligent framework for spatio-temporal vehicle tracking," in *Proceedings of the 4th IEEE International Conference on Intelligent Transportation Systems*, August 2001, pp. 213–218.

[3] L. Lin and M.-L. Shyu, "Weighted association rule mining for video semantic detection," *International Journal of Multimedia Data Engineering and Management*, vol. 1, no. 1, pp. 37–54, 2010.

[4] M.-L. Shyu, S.-C. Chen, and R. Kashyap, "Generalized affinity-based association rule mining for multimedia database queries," *Knowledge and Information Systems (KAIS): An International Journal*, vol. 3, no. 3, pp. 319–337, August 2001.

[5] Y. Yan, M. Chen, M.-L. Shyu, and S.-C. Chen, "Deep learning for imbalanced multimedia data classification," in *Proceedings of the 2015 IEEE International Symposium on Multimedia (ISM)*, December 2015, pp. 483–488.

[6] Y. Yan, Y. Liu, M.-L. Shyu, and M. Chen, "Utilizing concept correlations for effective imbalanced data classification," in *Proceedings of the IEEE 15th International Conference on Information Reuse and Integration*, August 2014, pp. 561–568.

[7] D. Liu, Y. Yan, M.-L. Shyu, G. Zhao, and M. Chen, "Spatio-temporal analysis for human action detection and recognition in uncontrolled environments," *International Journal of Multimedia Data Engineering and Management*, vol. 6, no. 1, pp. 1–18, Jan. 2015.

[8] S.-C. Chen, M.-L. Shyu, and C. Zhang, "Innovative shot boundary detection for video indexing," in *Video Data Management and Information Retrieval*, S. Deb, Ed. Idea Group Publishing, 2005, pp. 217–236.

[9] X. Li, S.-C. Chen, M.-L. Shyu, and B. Furht, "Image retrieval by color, texture, and spatial information," in *Proceedings of the 8th International Conference on Distributed Multimedia Systems*, September 2002, pp. 152–159.

[10] L. Lin, G. Ravitz, M.-L. Shyu, and S.-C. Chen, "Video semantic concept discovery using multimodal-based association classification," in *Proceedings of the IEEE International Conference on Multimedia & Expo*, July 2007, pp. 859–862.

[11] S.-C. Chen, M.-L. Shyu, C. Zhang, and R. L. Kashyap, "Identifying overlapped objects for video indexing and modeling in multimedia database systems," *International Journal on Artificial Intelligence Tools*, vol. 10, no. 4, pp. 715–734, 2001.

[12] X. Li, S.-C. Chen, M.-L. Shyu, and B. Furht, "An effective content-based visual image retrieval system," in *Proceedings*

of the Computer Software and Applications Conference, 2002, pp. 914–919.

[13] M.-L. Shyu, C. Haruechaiyasak, and S.-C. Chen, "Category cluster discovery from distributed www directories," *Information Sciences*, vol. 155, no. 3, pp. 181–197, 2003.

[14] L. Lin, G. Ravitz, M.-L. Shyu, and S.-C. Chen, "Effective feature space reduction with imbalanced data for semantic concept detection," in *Proceedings of the IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing*, 2008, pp. 262–269.

[15] A. F. Smeaton, P. Over, and W. Kraaij, "Evaluation campaigns and TRECVid," in *Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval*, October 2006, pp. 321–330.

[16] S.-C. Chen, S. Sista, M.-L. Shyu, and R. Kashyap, "Augmented transition networks as video browsing models for multimedia databases and multimedia information systems," in *Proceedings of the 11th IEEE International Conference on Tools with Artificial Intelligence*, November 1999, pp. 175–182.

[17] M.-L. Shyu, Z. Xie, M. Chen, and S.-C. Chen, "Video semantic event/concept detection using a subspace-based multimedia data mining framework," *IEEE Transactions on Multimedia*, vol. 10, no. 2, pp. 252–259, February 2008.

[18] Q. Zhu, L. Lin, M.-L. Shyu, and S.-C. Chen, "Feature selection using correlation and reliability based scoring metric for video semantic detection," in *Proceedings of the Fourth IEEE International Conference on Semantic Computing*, 2010, pp. 462–469.

[19] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *SIGKDD Explor. Newsl.*, vol. 6, no. 1, pp. 20–29, Jun. 2004.

[20] C. P. Unsworth and G. Coghill, "Excessive noise injection training of neural networks for markerless tracking in obscured and segmented environments," *Neural Comput.*, vol. 18, no. 9, pp. 2122–2145, Sep. 2006.

[21] X. Wang and Q. Ji, "A hierarchical context model for event recognition in surveillance video," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, June 2014, pp. 2561–2568.

[22] B. Ni, Y. Song, and M. Zhao, "Youtubeevent: On large-scale video event classification," in *Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, November 2011, pp. 1516–1523.

[23] L. Jiang, S.-I. Yu, D. Meng, Y. Yang, T. Mitamura, and A. G. Hauptmann, "Fast and accurate content-based semantic search in 100m internet videos," in *Proceedings of the 23rd ACM International Conference on Multimedia*. New York, NY, USA: ACM, 2015, pp. 49–58.

[24] T. Meng and M.-L. Shyu, "Concept-concept association information integration and multi-model collaboration for multimedia semantic concept detection," *Information Systems Frontiers*, pp. 1–13, 2013.

[25] M.-L. Shyu, S.-C. Chen, M. Chen, and C. Zhang, "A unified framework for image database clustering and content-based retrieval," in *Proceedings of the 2nd ACM International Workshop on Multimedia Databases*. New York, NY, USA: ACM, 2004, pp. 19–27.

[26] M.-L. Shyu, S.-C. Chen, M. Chen, C. Zhang, and K. Sarinnapakorn, "Image database retrieval utilizing affinity relationships," in *Proceedings of the 1st ACM International Workshop on Multimedia Databases*. New York, NY, USA: ACM, 2003, pp. 78–85.

[27] M.-L. Shyu, T. Quirino, Z. Xie, S.-C. Chen, and L. Chang, "Network intrusion detection through adaptive sub-eigenspace modeling in multiagent systems," *ACM Trans. Auton. Adapt. Syst.*, vol. 2, no. 3, Sep. 2007.

[28] P. Mettes, D. C. Koelma, and C. G. M. Snoek, "The imagenet shuffle: Reorganized pre-training for video event detection," *CoRR*, vol. abs/1602.07119, 2016.

[29] H. J. Escalante, M. Montes, and L. E. Sucar, "Word co-occurrence and markov random fields for improving automatic image annotation," in *Proceedings of the 18th British Machine Vision Conference*, 2007, pp. 600–609.

[30] H. J. Escalante, M. Montes, and E. Sucar, "Multimodal indexing based on semantic cohesion for image retrieval," *Information Retrieval*, vol. 15, no. 1, pp. 1–32, 2012. [Online]. Available: http://dx.doi.org/10.1007/s10791-011-9170-z

[31] T. Meng, Y. Liu, M.-L. Shyu, Y. Yan, and C.-M. Shu, "Enhancing multimedia semantic concept mining and retrieval by incorporating negative correlations," in *Proceedings of the IEEE International Conference on Semantic Computing*, June 2014, pp. 28–35.

[32] P. Over, G. Awad, M. Michel, J. Fiscus, G. Sanders, W. Kraaij, A. F. Smeaton, G. Quenot, and R. Ordelman, "Trecvid 2015 - an overview of the goals, tasks, data, evaluation mechanisms and metrics," in *Proceedings of TRECVID 2015*. NIST, USA, 2015.

[33] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Washington, DC, USA: IEEE Computer Society, 2014, pp. 1725–1732.

[34] Q. Zhu, L. Lin, M.-L. Shyu, and D. Liu, "Utilizing context information to enhance content-based image classification," *International Journal of Multimedia Data Engineering and Management*, vol. 2, no. 3, pp. 34–51, 2011.

[35] L. Lin, C. Chen, M.-L. Shyu, and S.-C. Chen, "Weighted subspace filtering and ranking algorithms for video concept retrieval," *IEEE Multimedia*, vol. 18, no. 3, pp. 32–43, March 2011.

[36] Y.-G. Jiang, "Prediction scores on TRECVID 2010 data set," http://www.ee.columbia.edu/ln/dvmm/CU-VIREO374/, 2010, last accessed on September 2011.