

of informative concept banks by performing three experiments on the unconstrained web video collection from the TRECVID 2011 multimedia event detection task using a total of 1,346 concept detectors.

The result of experiment 1 gives an indication that large banks of concept detectors are important for covering a variety of complex events, as they may appear in unconstrained video. In general, the event detection accuracy increases with the number of concept detectors in the bank. However, it also shows that some concept banks are more informative than others for specific events, and this may result in improved event detection accuracy. The results of experiment 2, and the same experiment on the dataset provided by Merler *et al.* [21], show that event detection using an informative concept bank outperform banks using all concepts, and always contains significantly less detectors. Finally, experiment 3 reveals that our informative concept bank outperforms both a bank using all concepts and a bag-of-words for small amounts of training examples. What is more the concepts in the informative concept bank appear to have a semantic relation with the events they model. We conclude that for video event detection using concept banks it pays to be informative.

Acknowledgments This research is supported by the STW STORY project, the BeeldCanon project, the Dutch national program COMMIT, and by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center contract number D11PC20067. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the U.S. Government.

6. REFERENCES

- [1] T. Althoff, H. O. Song, and T. Darrell. Detection bank: An object detection based video representation for multimedia event recognition. In *ACM Multimedia*, 2012.
- [2] S. Ayache and G. Quénot. Video corpus annotation using active learning. In *ECIR*, 2008.
- [3] L. Ballan, M. Bertini, A. D. Bimbo, and G. Serra. Video event classification using string kernels. *MTAP*, 48(1), 2010.
- [4] L. Ballan, M. Bertini, A. Del Bimbo, L. Seidenari, and G. Serra. Event detection and recognition for semantic annotation of video. *MTAP*, 51, 2011.
- [5] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A large-scale hierarchical image database. In *CVPR*, 2009.
- [6] L. Duan, D. Xu, I. W.-H. Tsang, and J. Luo. Visual event recognition in videos by learning from web data. *TPAMI*, 34(9), 2012.
- [7] S. Ebadollahi, L. Xie, S.-F. Chang, and J. R. Smith. Visual event detection using multi-dimensional concept dynamics. In *ICME*, 2006.
- [8] N. Gkalelis, V. Mezaris, and I. Kompatsiaris. High-level event detection in video exploiting discriminant concepts. In *CBMI*, 2011.
- [9] N. Haering, R. Qian, and I. Sezan. A semantic event-detection approach and its application to detecting hunts in wildlife video. *TCSVT*, 2000.
- [10] A. G. Hauptmann, M. G. Christel, and R. Yan. Video retrieval based on semantic concepts. *Proceedings of the IEEE*, 2008.
- [11] B. Huurnink, K. Hofmann, and M. de Rijke. Assessing concept selection for video retrieval. In *ACM MIR*, 2008.
- [12] N. Inoue et al. TokyoTech+Canon at TRECVID 2011. In *NIST TRECVID Workshop*, 2011.
- [13] Y. A. Ivanov and A. F. Bobick. Recognition of visual activities and interactions by stochastic parsing. *TPAMI*, 22(8), 2000.
- [14] Y.-G. Jiang, J. Yang, C.-W. Ngo, and A. Hauptmann. Representations of keypoint-based semantic concept detection: A comprehensive study. *TMM*, 12(1), 2010.
- [15] Y.-G. Jiang, X. Zeng, G. Ye, S. Bhattacharya, D. Ellis, M. Shah, and S.-F. Chang. Columbia-ucf trecvid2010 multimedia event detection: Combining multiple modalities, contextual concepts, and temporal matching. In *NIST TRECVID Workshop*, 2010.
- [16] G. Lavee, E. Rivlin, and M. Rudzsky. Understanding video events: A survey of methods for automatic interpretation of semantic occurrences in videos. *TSMC*, 39(5), 2009.
- [17] L.-J. Li, H. Su, E. P. Xing, and L. Fei-Fei. Object bank: A high-level image representation for scene classification & semantic feature sparsification. In *NIPS*, 2010.
- [18] X. Li, E. Gavves, C. G. M. Snoek, M. Worring, and A. W. M. Smeulders. Personalizing automated image annotation using cross-entropy. In *ACM Multimedia*, 2011.
- [19] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60, 2004.
- [20] S. Maji, A. C. Berg, and J. Malik. Classification using intersection kernel support vector machines is efficient. In *CVPR*, 2008.
- [21] M. Merler, B. Huang, L. Xie, G. Hua, and A. Natsev. Semantic model vectors for complex video event recognition. *IEEE Trans. Multimedia*, 14(1), 2012.
- [22] M. R. Naphade, J. R. Smith, J. Tešić, S.-F. Chang, W. Hsu, L. S. Kennedy, A. G. Hauptmann, and J. Curtis. Large-scale concept ontology for multimedia. *IEEE MultiMedia*, 13(3), 2006.
- [23] P. Natarajan, S. Wu, S. N. P. Vitaladevuni, X. Zhuang, S. Tsakalidis, U. Park, R. Prasad, and P. Natarajan. Multimodal feature fusion for robust event detection in web videos. In *CVPR*, 2012.
- [24] R. Y. Rubinstein and D. P. Kroese. *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning*. Springer, 2004.
- [25] S. Sadeanand and J. J. Corso. Action bank: A high-level representation of activity in video. In *CVPR*, 2012.
- [26] S. Shalev-Shwartz, Y. Singer, N. Srebro, and A. Cotter. Pegasos: primal estimated sub-gradient solver for svm. *Math. Program.*, 127(1), 2011.
- [27] A. F. Smeaton, P. Over, and W. Kraaij. Evaluation campaigns and TRECVID. In *ACM MIR*, 2006.
- [28] C. G. M. Snoek and M. Worring. Concept-based video retrieval. *FnTIR*, 2(4), 2009.
- [29] A. Tamrakar, S. Ali, Q. Yu, J. Liu, O. Javed, A. Divakaran, H. Cheng, and H. S. Sawhney. Evaluation of low-level features and their combinations for complex event detection in open source videos. In *CVPR*, 2012.
- [30] TRECVID Multimedia Event Detection Evaluation Track, 2011. <http://www.nist.gov/itl/iad/mig/med.cfm>.
- [31] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek. Evaluating color descriptors for object and scene recognition. *TPAMI*, 32(9), 2010.
- [32] X.-Y. Wei, C.-W. Ngo, and Y.-G. Jiang. Selection of concept detectors for video search by ontology-enriched semantic spaces. *TMM*, 10(6), 2008.
- [33] L. Xie, H. Sundaram, and M. Campbell. Event mining in multimedia streams. *Proceedings of the IEEE*, 96, 2008.
- [34] D. Xu and S.-F. Chang. Video event recognition using kernel methods with multilevel temporal alignment. *TPAMI*, 30(11), 2008.