

## NGO Chong Wah

School of Computing and Information Systems  
Singapore Management University (SMU)  
80 Stamford Road  
Singapore 178902

Email: cwngo@smu.edu.sg  
Office Phone: 6826 4818



## Education

PhD, Hong Kong University of Science & Technology, 2001  
Master of Applied Science, Nanyang Technological University, Singapore, 1997  
Bachelor of Applied Science, Nanyang Technological University, Singapore, 1994

## Academic Appointments

Lee Kong Chian Professor of Computer Science, School of Computing and Information Systems, SMU, Jul 2024 - Present  
Professor of Computer Science, School of Computing and Information Systems, SMU, Apr 2021 - Jun 2024  
Professor of Information Systems, School of Computing and Information Systems, SMU, Jan 2021 - Mar 2021  
Professor, City University of Hong Kong, China, Jul 2015 - Dec 2020  
Associate Professor, City University of Hong Kong, China, Mar 2009 - Jun 2015  
Assistant Professor, City University of Hong Kong, China, Jan 2002 - Feb 2009

## Academic Administrative Positions

ResWORK Fellow, SMU Resilient Workforces Institute, SMU, Oct 2025 - Present  
Director, Human-Machine Collaborative Systems Cluster, School of Computing and Information Systems, SMU, Jul 2021 - Present

## Other Positions and Affiliations

Visiting Research Professor, NExT++ Research Center, National University of Singapore, Singapore, Dec 2018 - Nov 2020

## Awards and Honors

Best Student Paper, Multimedia Modeling, 2017  
President Award, City University of Hong Kong, 2017  
Best Student Paper, ACM Multimedia, 2016

Best Poster, ACM Multimedia , 2014

Best Paper, International Conference on Internet Multimedia Computing and Service, 2013

Grant Challenge Award (2nd Prize) , ACM Multimedia , 2011

## Professional Memberships

Senior Member, IEEE, 2017

Distinguished Scientist , ACM, 2016

## RESEARCH

---

### Research and Project Areas

Multimedia Search, Analysis and Mining; Video Analytics; Lifestyle and Wellness Intelligence; Computer Vision and Image Processing; Machine Learning

## Publications

### Journal Articles [Refereed]

CookingDiffusion: Cooking procedural image generation with Stable Diffusion, by WANG, Yuan; ZHU, Bin; HAO, Yanbin; NGO, Chong-Wah; TAN, Yi; WANG, Xiang. (2025). *ACM Transactions on Multimedia Computing, Communications and Applications*, 1-23. <https://doi.org/10.1145/3771995> (Advance Online)

Towards multimodal emotional support conversation systems, by CHU, Yuqi; LIAO, Lizi; ZHOU, Zhiyuan; NGO, Chong-Wah; HONG, Richang. (2025). *IEEE Transactions on Multimedia*, 278276-8287. <https://doi.org/10.1109/TMM.2025.3604951> (Published)

FoodLMM: A versatile food assistant using large multi-modal model, by YIN, Yuehao; QI, Huiyan; ZHU, Bin; CHEN, Jingjing; JIANG, Yu-Gang; NGO, Chong-Wah. (2025). *IEEE Transactions on Multimedia*, 1-38. <https://doi.org/10.48550/arXiv.2312.14991> (Advance Online)

Stream-ViT: Learning streamlined convolutions in Vision Transformer, by PAN, Yingwei; LI, Yehao; YAO, Ting; NGO, Chong-Wah; MEI, Tao. (2025). *IEEE Transactions on Multimedia*, 273755-3765. <https://doi.org/10.1109/TMM.2025.3535321> (Published)

PosMLP-Video: Spatial and temporal relative position encoding for efficient video recognition, by HAO, Yanbin; ZHOU, Diansong; WANG, Zhicai; NGO, Chong-Wah; HE, Xiangnan; WANG, Meng. (2024). *International Journal of Computer Vision*, 132(12), 5820-5840. <https://doi.org/10.21203/rs.3.rs-3485088/v1> (Published)

Learning temporal dynamics in videos with image transformer, by SHU, Yan; QIU, Z; LONG, Fuchen; YAO, Ting; NGO, Chong-Wah; MEI, Tao. (2024). *IEEE Transactions on Multimedia*, 26 8915-8927. <https://doi.org/10.1109/TMM.2024.3383662> (Published)

FoodMask: Real-time food instance counting, segmentation and recognition, by NGUYEN, Huu-Thanh; CAO, Yu; NGO, Chong-Wah; CHAN, Wing-Kwong. (2024). *Pattern Recognition*, 146 1-11. <https://doi.org/10.1016/j.patcog.2023.110017> (Published)

(Un)likelihood training for interpretable embedding, by WU, Jiaxin; NGO, Chong-Wah; CHAN, Wing-Kwong; HOU, Zhijian. (2023). *ACM Transactions on Information Systems*, 42(3), <https://doi.org/10.1145/3632752> (Published)

SibNet: Food instance counting and segmentation, by NGUYEN, Huu-Thanh.; NGO, Chong-Wah; CHAN,

Wing-Kwong. (2022). *Pattern Recognition*, 124 1-11. <https://doi.org/10.1016/j.patcog.2021.108470> (Published)

Approximate *k*-NN Graph Construction: A Generic Online Approach, by ZHAO, Wan-Lei; WANG, Hui; NGO, Chong-Wah. (2022). *IEEE Transactions on Multimedia*, 24 1909-1921. <https://doi.org/10.1109/TMM.2021.3073811> (Published)

Learning from web recipe-image pairs for food recognition: Problem, baselines and performance, by ZHU, Bin; NGO, Chong-Wah; CHAN, Wing-Kwong. (2022). *IEEE Transactions on Multimedia*, 24 1175-1185. <https://doi.org/10.1109/TMM.2021.3123474> (Published)

Interactive Search vs. Automatic Search: An Extensive Study on Video Retrieval, by NGUYEN, Phuong-Anh; NGO, Chong-Wah. (2021). *ACM Transactions on Multimedia Computing, Communications and Applications*, 17(2), 1-24. <https://doi.org/10.1145/3429457> (Published)

Mixed Dish Recognition With Contextual Relation and Domain Alignment, by DENG, Lixi; CHEN, Jingjing; NGO, Chong-Wah; SUN, Qianru; TANG, Sheng; ZHANG, Yongdong; CHUA, Tat-Seng. (2022). *IEEE Transactions on Multimedia*, 24 2034-2045. <https://doi.org/10.1109/TMM.2021.3075037> (Published)

Learning to match anchor-target video pairs with dual attentional holographic networks, by HAO, Yan Bin; NGO, Chong-Wah; ZHU, Bin. (2021). *IEEE Transactions on Image Processing*, 30 8130-8143. (Published)

A study of multi-task and region-wise deep learning for food ingredient recognition, by CHEN, Jingjing; ZHU, Bin; NGO, Chong-Wah; CHUA, Tat-Seng; JIANG, Yu-Gang. (2021). *IEEE Transactions on Image Processing*, 30 1514-1526. <https://doi.org/10.1109/TIP.2020.3045639> (Published)

A Hybrid Approach for Detecting Prerequisite Relations in Multi-Modal Food Recipes, by PAN, Liangming; CHEN, Jingjing; LIU, Shaoteng; NGO, Chong-Wah; KAN, Min-Yen; CHUA, Tat-Seng. (2021). *IEEE Transactions on Multimedia*, 23 4491-4501. <https://doi.org/10.1109/TMM.2020.3042706> (Published)

Neighbourhood structure preserving cross-modal embedding for video hyperlinking, by HAO, Yanbin; NGO, Chong-Wah; HUET, Benoit. (2020). *IEEE Transactions on Multimedia*, 22(1), 188-200. <https://doi.org/10.1109/TMM.2019.2923121> (Published)

A Fine Granularity Object-Level Representation for Event Detection and Recounting, by ZHANG, Hao; NGO, Chong-Wah. (2019). *IEEE Transactions on Multimedia*, 21(6), 1450-1463. <https://doi.org/10.1109/TMM.2018.2884478> (Published)

Cross-modal recipe retrieval with stacked attention model, by CHEN, Jing-Jing; PANG, Lei; NGO, Chong-Wah. (2018). *Multimedia Tools and Applications*, 77(22), 29457-29473. (Published)

*k*-means: A revisit, by ZHAO, Wan-Lei; DENG, Cheng-Hao; NGO, Chong-Wah. (2018). *Neurocomputing*, 291 195-206. (Published)

PageSense: Toward Stylewise Contextual Advertising via Visual Analysis of Web Pages, by MEI, Tao; LI, Lusong; TIAN, Xinmei; TAO, Dacheng; NGO, Chong-Wah. (2018). *IEEE Transactions on Circuits and Systems for Video Technology*, 28(1), 254-266. (Published)

Semantic Reasoning in Zero Example Video Event Retrieval, by DE BOER, M. H. T.; LU, Yi-Jie; ZHANG, Hao; SCHUTTE, Klamer; NGO, Chong-Wah; KRAAIJ, Wessel. (2017). *ACM Transactions on Multimedia Computing, Communications and Applications*, 13(4), 1-18. (Published)

Blind late fusion in multimedia event retrieval, by DE BOER, Maaike H. T.; SCHUTTE, Klamer; ZHANG, Hao; LU, Yi-Jie; NGO, Chong-Wah; KRAAIJ, Wessel. (2016). *International Journal of Multimedia Information Retrieval*, 5(4), 203-217. (Published)

Hierarchical Visualization of Video Search Results for Topic-Based Browsing, by JIANG, Yu-Gang; WANG, Jiajun; WANG, Qiang; LIU, Wei; NGO, Chong-Wah. (2016). *IEEE Transactions on Multimedia*, 18(11), 2161-2170. (Published)

Fast Covariant VLAD for Image Search, by ZHAO, Wan-Lei; NGO, Chong-Wah; WANG, Hanzi. (2016). *IEEE Transactions on Multimedia*, 18(9), 1843-1854. (Published)

Hyperlink-Aware Object Retrieval, by ZHANG, Wei; NGO, Chong-Wah; CAO, Xiaochun. (2016). *IEEE Transactions on Image Processing*, 25(9), 4186-4198. (Published)

Automatic Hookworm Detection in Wireless Capsule Endoscopy Images, by WU, Xiao; CHEN, Honghan;

GAN, Tao; CHEN, Junzhou; NGO, Chong-Wah; PENG, Qiang. (2016). *IEEE Transactions on Medical Imaging*, 35 (7), 1741-1752. (Published)

On the use of commonsense ontology for multimedia event recounting, by TAN, Chun-Chet; NGO, Chong-Wah. (2016). *International Journal of Multimedia Information Retrieval*, 5 (2), 73-88. (Published)

Opinion Question Answering by Sentiment Clip Localization, by PANG, Lei; NGO, Chong-Wah. (2016). *ACM Transactions on Multimedia Computing, Communications and Applications*, 12 (2), 1-19. (Published)

Detection of bird nests in overhead catenary system images for high-speed rail, by WU, Xiao; YUAN, Ping; PENG, Qiang; NGO, Chong-Wah; HE, Jun-Yan. (2016). *Pattern Recognition*, 51 242-254. (Published)

Object pooling for multimedia event detection and evidence localization, by ZHANG, Ho; NGO, Chong-Wah. (2016). *ITE Transactions on Media Technology and Applications*, 4 (3), 218-226. (Published)

Deep Multimodal Learning for Affective Analysis and Retrieval, by PANG, Lei; ZHU, Shuai; NGO, Chong-Wah. (2015). *IEEE Transactions on Multimedia*, 17 (11), 2008-2020. (Published)

Human Action Recognition in Unconstrained Videos by Explicit Motion Modeling, by JIANG, Yu-Gang; DAI, Qi; LIU, Wei; XUE, Xiangyang; NGO, Chong-Wah. (2015). *IEEE Transactions on Image Processing*, 24 (11), 3781-3795. (Published)

Topological Spatial Verification for Instance Search, by ZHANG, Wei; NGO, Chong-Wah. (2015). *IEEE Transactions on Multimedia*, 17 (8), 1236-1247. (Published)

Unsupervised Celebrity Face Naming in Web Videos, by PANG, Lei; NGO, Chong-Wah. (2015). *IEEE Transactions on Multimedia*, 17 (6), 854-866. (Published)

Click-boosting multi-modality graph-based reranking for image search, by YANG, Xiaopeng; ZHANG, Yongdong; YAO, Ting; NGO, Chong-Wah; MEI, Tao. (2015). *Multimedia Systems*, 21 (2), 217-227. (Published)

Name-Face Association in Web Videos: A Large-Scale Dataset, Baselines, and Open Issues, by CHEN, Zhi-Neng; NGO, Chong-Wah; ZHANG, Wei; CAO, Juan; JIANG, Yu-Gang. (2014). *Journal of Computer Science and Technology*, 29 (5), 785-798. (Published)

Video Event Detection Using Motion Relativity and Feature Selection, by WANG, Feng; SUN, Zhanhu; JIANG, Yu-Gang; NGO, Chong-Wah. (2014). *IEEE Transactions on Multimedia*, 16 (5), 1303-1315. (Published)

Collaborative error reduction for hierarchical classification, by ZHU, Shuai; WEI, Xiao-Yong; NGO, Chong-Wah. (2014). *Computer Vision and Image Understanding*, 124 79-90. (Published)

Placing Videos on a Semantic Hierarchy for Search Result Navigation, by TAN, Song; JIANG, Yu-Gang; NGO, Chong-Wah. (2014). *ACM Transactions on Multimedia Computing, Communications and Applications*, 10 (4), 1-120. (Published)

A Hamming Embedding Kernel with Informative Bag-of-Visual Words for Video Semantic Indexing, by WANG, Feng; ZHAO, Wen-Lei; NGO, Chong-Wah; MERALDO, Bernard. (2014). *ACM Transactions on Multimedia Computing, Communications and Applications*, 10 (3), 1-20. (Published)

Visual Typo Correction by Collocative Optimization: A Case Study on Merchandise Images, by WEI, Xiao-Yong; YANG, Zhen-Qun; NGO, Chong-Wah; ZHANG, Wei. (2014). *IEEE Transactions on Image Processing*, 23 (2), 527-540. (Published)

Multimedia modeling, by NGO, Chong-Wah; SCHOEFFMANN, Klaus; ANDREOPoulos, Yiannis; BREITENEDER, Christian. (2014). *Multimedia Tools and Applications*, 71 (1), 331-332. (Published)

Web-Scale Near-Duplicate Search: Techniques and Applications, by NGO, Chong-Wah; XU, Changsheng; KRAAIJ, Wessel; EL SADDIK, Abdulmotaleb. (2013). *IEEE MultiMedia*, 20 (3), 10-12. (Published)

Near-Duplicate Video Retrieval: Current Research and Future Trends, by LIU, Jiajun; HUANG, Zi; CAI, Hongyun; SHEN, Heng Tao; NGO, Chong-Wah; WANG, Wei. (2013). *ACM Computing Surveys*, 45 (4), 1-23. (Published)

Circular Reranking for Visual Search, by YAO, Ting; NGO, Chong-Wah. (2013). *IEEE Transactions on Image Processing*, 22 (4), 1642-1653. (Published)

Flip-Invariant SIFT for Copy and Object Detection, by ZHAO, Wan-Lei; NGO, Chong-Wah. (2013). *IEEE Transactions on Image Processing*, 22(3), 980-991. (Published)

Guest editorial: selected papers from ICIMCS 2011, by NGO, Chong-Wah; XU, Changsheng; WU, Xiao; EL SADDIK, Abdulmotaleb. (2013). *Multimedia Systems*, 19(1), 1-2. (Published)

Sampling and Ontologically Pooling Web Images for Visual Concept Learning, by ZHU, Shuai; NGO, Chong-Wah; JIANG, Yu-Gang. (2012). *IEEE Transactions on Multimedia*, 14(4), 1068-1078. (Published)

Fast Semantic Diffusion for Large-Scale Context-Based Image and Video Annotation, by JIANG, Yu-Gang; DAI, Qi; WANG, Jun; NGO, Chong-Wah. (2012). *IEEE Transactions on Image Processing*, 21(6), 3080-3091. (Published)

Boosting web video categorization with contextual information from social web, by WU, Xiao; NGO, Chong-Wah; ZHU, Yi-Ming; PENG, Qiang. (2012). *World Wide Web*, 15(2), 197-212. (Published)

Summarizing Rushes Videos by Motion, Object, and Event Understanding, by WANG, Feng; NGO, Chong-Wah. (2012). *IEEE Transactions on Multimedia*, 14(1), 1-12. (Published)

Tracking Web Video Topics: Discovery, Visualization, and Monitoring, by CAO, Juan; NGO, Chong-Wah; ZHANG, Yong-Dong; LI, Jin-Tao. (2011). *IEEE Transactions on Circuits and Systems for Video Technology*, 21(12), 1835-1846. (Published)

Beyond Search: Event-Driven Summarization for Web Videos, by HONG, Richard; TANG, Jinhui; TAN, Hung-Khoon; NGO, Chong-Wah; YAN, Shuicheng; CHUA, Tat-Seng. (2011). *ACM Transactions on Multimedia Computing, Communications and Applications*, 7(4), 1-21. (Published)

Mining Event Structures from Web Videos, by WU, Xiao; LU, Yi-Jie; PENG, Qiang; NGO, Chong-Wah. (2011). *IEEE MultiMedia*, 18(1), 38-51. (Published)

Concept-Driven Multi-Modality Fusion for Video Search, by WEI, Xiao-Yong; JIANG, Yu-Gang; NGO, Chong-Wah. (2011). *IEEE Transactions on Circuits and Systems for Video Technology*, 21(1), 62-73. (Published)

Efficient Mining of Multiple Partial Near-Duplicate Alignments by Temporal Network, by TAN, Hung-Khoon; NGO, Chong-Wah; CHUA, Tat-Seng. (2010). *IEEE Transactions on Circuits and Systems for Video Technology*, 20(11), 1486-1498. (Published)

Data-Driven Approaches to Community-Contributed Video Applications, by WU, Xiao; NGO, Chong-Wah; ZHAO, Wan-Lei. (2010). *IEEE MultiMedia*, 17(4), 58-68. (Published)

On the Annotation of Web Videos by Efficient Near-Duplicate Search, by ZHAO, Wan-Lei; WU, Xiao; NGO, Chong-Wah. (2010). *IEEE Transactions on Multimedia*, 12(5), 448-461. (Published)

Representations of Keypoint-Based Semantic Concept Detection: A Comprehensive Study, by JIANG, Yu-Gang; YANG, Jun; NGO, Chong-Wah. (2010). *IEEE Transactions on Multimedia*, 12(1), 42-53. (Published)

Localized matching using Earth Mover's Distance towards discovery of common patterns from small image samples, by TAN, Hung-Khoon; NGO, Chong-Wah. (2009). *Image and Vision Computing*, 27(10), 1470-1483. <https://doi.org/10.1016/j.imavis.2009.01.002> (Published)

Visual word proximity and linguistics for semantic video indexing and near-duplicate retrieval, by JIANG, Yu-Gang; NGO, Chong-Wah. (2009). *Computer Vision and Image Understanding*, 113(3), 405-414. (Published)

Real-Time Near-Duplicate Elimination for Web Video Search With Content and Context, by WU, Xiao; NGO, Chong-Wah; HAUPTMANN, Alexander G.. (2009). *IEEE Transactions on Multimedia*, 11(2), 196-207. (Published)

Scale-Rotation Invariant Pattern Entropy for Keypoint-Based Near-Duplicate Detection, by ZHAO, Wan-Lei; NGO, Chong-Wah. (2009). *IEEE Transactions on Image Processing*, 18(2), 412-423. (Published)

Selection of Concept Detectors for Video Search by Ontology-Enriched Semantic Spaces, by WEI, Xiao-Yong; NGO, Chong-Wah; JIANG, Yu-Gang. (2008). *IEEE Transactions on Multimedia*, 10(6), 1085-1096. (Published)

Structuring low-quality videotaped lectures for cross-reference browsing by video text analysis, by WANG, Feng; NGO, Chong-Wah; PONG, Ting-Chuen. (2008). *Pattern Recognition*, 41 (10), 3257-3269. (Published)

Simulating a smartboard by real-time gesture detection in lecture videos, by WANG, Feng; NGO, Chong-Wah; PONG, Ting-Chuen. (2008). *IEEE Transactions on Multimedia*, 10 (5), 926-935. (Published)

Measuring novelty and redundancy with multiple modalities in cross-lingual broadcast news, by WU, Xiao; HAUPTMANN, Alexander G.; NGO, Chong-Wah. (2008). *Computer Vision and Image Understanding*, 110 (3), 418-431. (Published)

Multimodal news story clustering with pairwise visual near-duplicate constraint, by WU, Xiao; NGO, Chong-Wah; HAUPTMANN, Alexander G.. (2008). *IEEE Transactions on Multimedia*, 10 (2), 188-199. (Published)

OM-based video shot retrieval by one-to-one matching, by PENG, Yuxin; NGO, Chong-Wah; XIAO, Jianguo. (2007). *Multimedia Tools and Applications*, 34 (2), 249-266. (Published)

Near-duplicate keyframe identification with interest point matching and pattern learning, by ZHAO, Wan-Lei; NGO, Chong-Wah; TAN, Hung-Khoon; WU, Xiao. (2007). *IEEE Transactions on Multimedia*, 9 (5), 1037-1048. (Published)

Moving-object detection, association, and selection in home videos, by PAN, Zailiang; NGO, Chong-Wah. (2007). *IEEE Transactions on Multimedia*, 9 (2), 268-279. (Published)

Lecture video enhancement and editing by integrating posture, gesture, and text, by WANG, Feng; NGO, Chong-Wah; PONG, Ting-Chuen. (2007). *IEEE Transactions on Multimedia*, 9 (2), 397-409. (Published)

Introduction: special issue for the selected papers in the fourth international conference on Intelligent Multimedia Computing and Networking (IMMCN) 2005, by NGO, Chong-Wah; LEONG, Hong-Va. (2007). *Multimedia Tools and Applications*, 34 (2), 137-138. (Published)

Gestalt-based feature similarity measure in trademark database, by JIANG, Hui; NGO, Chong-Wah; TAN, Hung-Khoon. (2006). *Pattern Recognition*, 39 (5), 988-1001. (Published)

Clip-based similarity measure for query-dependent clip retrieval and video summarization, by PENG, Yuxin; NGO, Chong-Wah . (2006). *IEEE Transactions on Circuits and Systems for Video Technology*, 16 (5), 612-627. (Published)

Threading and autodocumenting news videos, by WU, Xiao; NGO, Chong-Wah; LI, Qing. (2006). *IEEE Signal Processing Magazine*, 23 (2), 59-68. (Published)

Selective object stabilization for home video consumers, by PAN, Zailiang; NGO, Chong-Wah. (2005). *IEEE Transactions on Consumer Electronics*, 51 (4), 1074-1084. (Published)

Video text detection and segmentation for optical character recognition, by NGO, Chong-Wah; CHAN, Chi-Kwong. (2005). *Multimedia Systems*, 10 (3), 261-272. (Published)

Video summarization and scene detection by graph modeling, by NGO, Chong-Wah; MA, Yu-Fei; ZHANG, Hong-Jiang. (2005). *IEEE Transactions on Circuits and Systems for Video Technology*, 15 (2), 296-305. (Published)

Key technologies for content-based video retrieval, by PENG, Y.; PENG, Y.; NGO, C.W.; GUO, Z.; XIAO, J.. (2004). *Jisuanji Gongcheng/Computer Engineering*, 30 (1), 14-. (Published)

Approach for video retrieval by video clip, by PENG, Y.; PENG, Y.; NGO, C.W.; GUO, Z.; XIAO, J.. (2003). *Ruanjian Xuebao /Journal of Software*, 14 (8), 1409-1417. (Published)

Motion analysis and segmentation through spatio-temporal slices processing, by NGO, Chong-Wah; PONG, Ting-Chuen; ZHANG, Hong-Jiang. (2003). *IEEE Transactions on Image Processing*, 12 (3), 341-355. (Published)

On clustering and retrieval of video shots through temporal slices analysis, by NGO, Chong-Wah; PONG, Ting-Chuen; ZHANG, Hong-Jiang. (2002). *IEEE Transactions on Multimedia*, 4 (4), 446-458. (Published)

Motion-based video representation for scene change detection, by NGO, Chong-Wah; PONG, Ting-Chuen; ZHANG, Hong-Jiang. (2002). *International Journal of Computer Vision*, 50 (2), 127-142. (Published)

Exploiting image indexing techniques in DCT domain, by NGO, Chong-Wah; PONG, Ting-Chuen; CHIN,

Ronald T.. (2001). *Pattern Recognition*, 34 (9), 1841-1851. (Published)

Video partitioning by temporal slice coherency, by NGO, Chong-Wah; PONG, Ting-Chuen; CHIN, Roland T.. (2001). *IEEE Transactions on Circuits and Systems for Video Technology*, 11 (8), 941-953. (Published)

Recent advances in content-based video analysis, by NGO, Chong-Wah; PONG, Ting-Chuen; ZHANG, Hong-Jiang. (2001). *International Journal of Image and Graphics*, 1 (3), 445-468. (Published)

Motion tracking of human mouth by generalized deformable models, by CHAN, Syin; NGO, Chong-Wah; LAI, Kok F.. (1999). *Pattern Recognition Letters*, 20 (9), 879-887. (Published)

#### Journal Articles [Non-Refereed]

Empowering weight loss: a pragmatic randomized controlled trial of a theory-driven self-regulation mobile app for young adults with excess body weight, by CHEW, H. S. J.; NGOOI, J. W.; DU, R. C.; CHAN, P. Z.; JANSSON, M.; ZHU, B.; CAO, Y.; NGO, C. W.; FOO, R.; SHABBIR, A.; HO, D.; SEVDALIS, N.; NGIAM, K. Y.. (2025). *European Journal of Cardiovascular Nursing*, 24 1-1. <https://doi.org/10.1093/eurjcn/zvaf122.128> (Published)

Approach for shot retrieval by optimal matching in the bipartite graph, by PENG, Yu-Xin; NGO, Chong-Wah; XIAO, Jian-Guo. (2004). *Tien Tzu Hsueh Pao/Acta Electronica Sinica*, 32 (7), 1135-1139. (Published)

#### Editorials

Guest editorial: When multimedia meets food: Multimedia computing for food data analysis and applications, by MIN, Weiqing; JIANG, Shuqiang; RADEVA, Petia; PAVLOVIC, Vladimir; NGO, Chong-Wah; AIZAWA, Kiyoharu; LI, Wanqing. (2025). *IEEE Transactions on Multimedia*, 27 2708-2712. <https://doi.org/10.1109/TMM.2025.3566452> (Published)

Intelligent analytics for big video data, by LIU, Wu; MA, Huadong; NGO, Chong-Wah. (2020). *Multimedia Tools and Applications*, 79 (7-8), 4637-4637. (Published)

Special issue on multimedia recommendation and multi-modal data analysis, by HE, Xiangnan; LIU, Zhenguang; ZHANG, Hanwang; NGO, Chong-Wah; KARAMAN, Svebor; ZHANG, Yongfeng. (2019). *Multimedia Systems*, 25 (6), 591-592. <https://doi.org/10.1007/s00530-019-00639-3> (Published)

#### Conference Proceedings

Seeing culture: A benchmark for visual reasoning and grounding, by SATAR, Burak; MA, Zhixin; IRRAWAN, Patrick Amadeus; MULYAWAN, Wilfried Ariel; JIANG, Jing; LIM, Ee-Peng; NGO, Chong-Wah. (2025.0). *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, Suzhou, China, November 4-9*, (pp. 22238-22254) Suzhou: ACL. <https://doi.org/10.18653/v1/2025.emnlp-main.1131> (Published)

Mitigating cross-modal representation bias for multicultural image-to-recipe retrieval, by WANG, Qing; NGO, Chong-Wah; CAO, Yu; LIM, Ee-Peng. (2025.0). *MM '25: Proceedings of the 33rd ACM International Conference on Multimedia, Dublin, Ireland, October 27-31*, (pp. 6223-6231) New York : ACM. <https://doi.org/10.1145/3746027.3755583> (Published)

From holistic to localized: Local enhanced adapters for efficient visual instruction fine-tuning, by JIAO, Pengkun; ZHU, Bin; CHEN, Jingjing; NGO, Chong-Wah; JIANG, Yugang. (2025.0). *Proceedings of the 2025 International Conference on Computer Vision, Honolulu, Hawaii, October 19-23*, (pp. 1-10) Honolulu, Hawai'i, USA: (Published)

Efficient prompt tuning for hierarchical ingredient recognition, by GUI, Yinxuan; ZHU, Bin; CHEN, Jingjing; NGO, Chong-Wah. (2025.0). *Proceedings of the 2025 IEEE International Conference on Multimedia and Expo (ICME 2025), Nantes, France, June 30 - July 4*, (pp. 1-6) Piscataway, NJ: IEEE. <https://doi.org/10.48550/arXiv.2504.10322> (Presented)

Robust relevance feedback for interactive known-item video search, by MA, Zhixin; NGO, Chong-Wah. (2025.0). *ICMR '25: Proceedings of the 2025 International Conference on Multimedia Retrieval, Chicago, USA, June 30 - July 3*, (pp. 982-990) New York : ACM. <https://doi.org/10.1145/3731715.3733422> (Published)

Advancing food nutrition estimation via visual-ingredient feature fusion, by QI, Huiyan; ZHU, Bin; NGO, Chong-Wah; CHEN, Jingjing; LIM, Ee-Peng. (2025.0). *Proceedings of the 2025 International Conference on Multimedia Retrieval, Chicago, IL, USA, June 30 - July 3*, (pp. 1091-1099) New York: ACM. <https://doi.org/10.1145/3731715.3733269> (Published)

Worldcuisines: A massive-scale benchmark for multilingual and multicultural visual question answering on global cuisines, by WINATA, Genta Indra; et. al. (2025.0). *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), Albuquerque, New Mexico, April 29 - May 4*, (pp. 1-23) New Mexico: ACL. <https://doi.org/10.18653/v1/2025.nacl-long.167> (Published)

Retrieval augmented recipe generation, by LIU, Guoshan; YIN, Hailong; ZHU, Bin; CHEN, Jingjing; NGO, Chong-Wah; JIANG, Yu-Gang. (2025.0). *IEEE/CVF Winter Conference on Applications of Computer Vision 2025, Tucson, Arizona, February 28 - March 4*, (pp. 1-11) Tucson, Arizona: (Accepted)

RAGG: Retrieval-augmented grasp generation model, by TANG, Zhenhua; ZHU, Bin; HAO, Yanbin; NGO, Chong-Wah; HONG, Richang. (2025.0). *AAAI'25/IAAI'25/EAAI'25: Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence and Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence and Fifteenth Symposium on Educational Advances in Artificial Intelligence, Philadelphia, USA, February 25 - March 4*, (pp. 7311-7319) New York: ACM. <https://doi.org/10.1609/aaai.v39i7.32786> (Published)

LLMs-based augmentation for domain adaptation in long-tailed food datasets, by WANG, Qing; NGO, Chong-Wah; LIM, Ee Peng; SUN, Qianru. (2025.0). *MultiMedia Modeling: 31st International Conference on Multimedia Modeling, MMM 2025, Nara, January 8-10, Proceedings*, (pp. 282-295) Cham: Springer. [https://doi.org/10.1007/978-981-96-2061-6\\_21](https://doi.org/10.1007/978-981-96-2061-6_21) (Published)

Interactive video search with multi-modal LLM video captioning, by CHENG, Yu-Tong; WU, Jiaxin; MA, Zhixin; HE, Jiangshan; WEI, Xiao-Yong; NGO, Chong-Wah. (2025.0). *MultiMedia Modeling: 31st International Conference on Multimedia Modeling, MMM 2025, Nara, January 8-10, Proceedings*, (pp. 302-309) Cham: Springer. [https://doi.org/10.1007/978-981-96-2074-6\\_36](https://doi.org/10.1007/978-981-96-2074-6_36) (Published)

Hi3D: Pursuing high-resolution image-to-3D generation with video diffusion models, by YANG, Haibo; CHEN, Yang; PAN, Yingwei; YAO, Ting; CHEN, Zheneng; NGO, Chong-Wah; MEI, Tao. (2024.0). *MM '24: Proceedings of the 32nd ACM International Conference on Multimedia, Melbourne, Australia, October 28-November 1*, (pp. 6870-6879) New York: ACM. <https://doi.org/10.1145/3664647.3681634> (Published)

Navigating weight prediction with diet diary, by GUI, Yinxuan; ZHU, Bin; CHEN, Jingjing; NGO, Chong-Wah; JIANG, Yu-Gang. (2024.0). *MM '24: Proceedings of the 32nd ACM International Conference on Multimedia, Melbourne, Australia, October 28 - November 1*, (pp. 127-136) New York: ACM. <https://doi.org/10.1145/3664647.3680977> (Published)

Improving interpretable embeddings for ad-hoc video search with generative captions and multi-word concept bank, by WU, Jiaxin; NGO, Chong-Wah; CHAN, Wing-Kwong. (2024.0). *ICMR '24: Proceedings of the 2024 International Conference on Multimedia Retrieval, Phuket, Thailand, June 10-14*, (pp. 73-82) New York: ACM. <https://doi.org/10.1145/3652583.3658052> (Published)

Leveraging LLMs and generative models for interactive known-item video search, by MA, Zhixin; WU, Jiaxin; NGO, Chong Wah. (2024.0). *MultiMedia Modeling, MMM 2024: International Conference, Amsterdam, January 29 - February 2: Proceedings*, (pp. 1-7) Cham: Springer . [https://doi.org/10.1007/978-3-031-53302-0\\_35](https://doi.org/10.1007/978-3-031-53302-0_35) (Published)

OVFoodSeg : Elevating open-vocabulary food image segmentation via image-informed textual representation, by WU, Xiongwei; YU, Sicheng; LIM, Ee-Peng; NGO, Chong-Wah. (2024.0). *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2024) : Seattle, WA, USA, June 16-22*, (pp. 4144-4153) Seattle, USA: IEEE Computer Society. <https://doi.org/10.1109/CVPR52733.2024.00397> (Published)

ObjectFusion: Multi-modal 3D object detection with object-centric fusion, by CAI, Q.; PAN, Y.; YAO, T.; NGO, C. W.; MEI, T.. (2023.0). *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 2023, October 4-6*, (pp. 18067-18076) Paris: (Published)

CONE: An efficient COarse-to-fINE alignment framework for long video temporal grounding, by HOU, Zhijian; ZHONG, Wanjun; JI, Lei; GAO, Difei; YAN, Kun; CHAN, Wing-Kwong; NGO, Chong-Wah; SHOU, Mike Z.; DUAN, Nan.. (2023.0). *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Toronto, Canada, July 9-14*, (pp. 8013-8028) Texas, USA: Association

for Computational Linguistics. <http://dx.doi.org/10.18653/v1/2023.acl-long.445> (Published)

Adaptive split-fusion transformer, by SU, Zixuan; CHEN, Jingjing; PANG, Lei; NGO, Chong-Wah; JIANG, Yu-Gang. (2023.0). *Proceedings of the 2023 IEEE International Conference on Multimedia and Expo, Brisbane, Australia, July 10-14*, (pp. 1169-1174) New Jersey: IEEE . <https://doi.org/10.1109/ICME55011.2023.00204> (Published)

Reinforcement learning enhanced PicHunter for interactive search, by MA, Zhixin; WU, Jiaxin; LOO, Weixiong; NGO, Chong-Wah. (2023.0). *MultiMedia Modeling: 29th International Conference MMM 2023, Bergen, Norway, January 9-12: Proceedings*, (pp. 690-696) Cham: Springer. [https://doi.org/10.1007/978-3-031-27077-2\\_60](https://doi.org/10.1007/978-3-031-27077-2_60) (Published)

Wave-ViT: Unifying wavelet and transformers for visual representation learning, by YAO, Ting; PAN, Yingwei; LI, Yehao; NGO, Chong-Wah; MEI, Tao. (2022.0). *Computer Vision ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23-27: Proceedings*, (pp. 328-345) Cham: Springer. [https://doi.org/10.1007/978-3-031-19806-9\\_19](https://doi.org/10.1007/978-3-031-19806-9_19) (Published)

Dynamic temporal filtering in video models, by LONG, Fuchen; QIU, Zhaofan; PAN, Yingwei; YAO, Ting; NGO, Chong-Wah; MEI, Tao. (2022.0). *Computer Vision ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23-27: Proceedings*, (pp. 475-492) Cham: Springer. [https://doi.org/10.1007/978-3-031-19833-5\\_28](https://doi.org/10.1007/978-3-031-19833-5_28) (Published)

Interactive video corpus moment retrieval using reinforcement learning, by MA, Zhixin; NGO, Chong-Wah. (2022.0). *Proceedings of the 30th ACM International Conference on Multimedia, Lisboa, Portugal, 2022 October 10-14*, (pp. 296-306) Lisbon, Portugal: ACM. <http://doi.org/10.1145/3503161.3548277> (Published)

Long-term leap attention, short-term periodic shift for video classification, by ZHANG, Hao; CHENG, Lechao; HAO, Yanbin; NGO, Chong-Wah. (2022.0). *Proceedings of the 30th ACM International Conference on Multimedia, Lisboa, Portugal, 2022 October 10-14*, (pp. 5773-5782) Lisbon, Portugal: ACM. <http://doi.org/10.1145/3503161.3547908> (Published)

Cross-lingual adaptation for recipe retrieval with mixup, by ZHU, Bin; NGO, Chong-Wah; CHEN, Jingjing; CHAN, Wing-Kwong. (2022.0). *ICMR '22: Proceedings of the 2022 International Conference on Multimedia Retrieval, Newark, NJ, June 27-30*, (pp. 258-267) New York: ACM. <https://doi.org/10.1145/3512527.3531375> (Published)

Group contextualization for video recognition, by HAO, Yanbin; ZHANG, Hao; NGO, Chong-Wah; HE, Xiangnan. (2022.0). *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, June 18-24: Proceedings*, (pp. 918-928) Piscataway, NJ: IEEE. <https://doi.org/10.1109/CVPR52688.2022.00100> (Published)

MLP-3D: A MLP-like 3D architecture with grouped time mixing, by QIU, Zhaofan; YAO, Ting; NGO, Chong-Wah; MEI, Tao. (2022.0). *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, June 18-24*, (pp. 3052-3062) New Orleans, USA: Computer Vision Foundation. <http://doi.org/10.1109/CVPR52688.2022.00307> (Published)

Reinforcement learning-based interactive video search, by MA, Zhixin; WU, Jiaxin; HOU, Zhijian; NGO, Chong-Wah. (2022.0). *MMM 2022: Proceedings of the 28th International Conference, Phu Quoc, June 6-10*, (pp. 549-555) Cham: Springer. [https://doi.org/10.1007/978-3-030-98355-0\\_53](https://doi.org/10.1007/978-3-030-98355-0_53) (Published)

K-sums clustering: A stochastic optimization approach, by ZHAO, Wan-Lei; LAN, Shi Ying; CHEN, Run-Qing; NGO, Chong-Wah. (2021.0). *CIKM '21: Proceedings of the 30th ACM International Conference on Information & Knowledge Management, Virtual Conference, November 1-5*, (pp. 2679-2687) New York : ACM. <https://doi.org/10.1145/3459637.3482359> (Published)

Token shift transformer for video classification, by ZHANG Hao; HAO, Yanbin.; NGO, Chong-Wah . (2021.0). *Proceedings of the 29th ACM International Conference on Multimedia, October 20-24*, (pp. 917-925) New York : Association for Computing Machinery, Inc. (Published)

CONQUER: Contextual query-aware ranking for video corpus moment retrieval, by HOU, Zhijian; NGO, Chong-Wah; CHAN, W. K.. (2021.0). *Proceedings of the 29th ACM International Conference on Multimedia, MM 2021, Virtual Conference, 2021 October 20-24*, (pp. 3900-3908) Virtual Conference: Association for Computing Machinery, Inc. (Published)

AIxFood'21: 3rd workshop on AIxFood, by GUERRERO, Ricardo; SPRANGER, Michael; JIANG, Shuqiang; NGO, Chong-Wah. (2021.0). *Proceedings of the 29th ACM International Conference on Multimedia, MM*

2021, *Virtual Conference, October 20-24*, (pp. 5688-5689) Virtual Conference: Association for Computing Machinery, Inc. (Published)

Condensing a sequence to one informative frame for video recognition, by QIU, Zhaofan; YAO, Ting; SHU, Yan; NGO, Chong-Wah; MEI, Tao . (2021.0). *Proceedings of the IEEE/CVF International Conference on Computer Vision, Virtual Online, October 11-17*, (pp. 16311-16320) New York : IEEE . (Published)

Optimization planning for 3D ConvNets, by QIU, Zhaofan; YAO, Ting; NGO, Chong-Wah; MEI, Tao. (2021.0). *Proceedings of the 38th International Conference on Machine Learning, Virtual Conference, 2021 July 18-24*, (pp. 8726-8736) Virtual Conference: PMLR . (Published)

Boosting video representation learning with multi-faceted integration, by QIU, Zhaofan; TING, Yao; NGO, Chong-Wah; ZHANG, Xiao-Ping; WU, Dong; MEI, Tao . (2021.0). *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 20-25*, New York : IEEE . (Published)

SQL-like interpretable interactive video search, by WU, Jiaxin; NGUYEN, Phuong Anh; MA, Zhixin; NGO, Chong-Wah. (2021.0). *Multimedia Modeling, MMM 2021: 27th International Conference, Prague, June 22-24: Proceedings*, (pp. 391-397) Cham: Springer. [https://doi.org/10.1007/978-3-030-67835-7\\_34](https://doi.org/10.1007/978-3-030-67835-7_34) (Published)

Terrace-based food counting and segmentation, by NGUYEN, Huu-Thanh; NGO, Chong-Wah. (2021.0). *Proceedings of the 35th AAAI Conference on Artificial Intelligence, Virtual Conference, 2021 February 2-9*, (pp. 2364-2372) USA: AAAI. (Published)

Interpretable embedding for ad-hoc video search, by WU, Jiaxin; NGO, Chong-Wah. (2020.0). *Proceedings of the 28th ACM International Conference on Multimedia, MM 2020, Seattle, October 12-16*, (pp. 3357-3366) Virtual Conference: Association for Computing Machinery, Inc. (Published)

Multi-modal cooking workflow construction for food recipes, by PAN, Liangming; CHEN, Jingjing; WU, Jianlong; LIU, Shaoteng; NGO, Chong-Wah; KAN, Min-Yen; JIANG, Yugang; CHUA, Tat-Seng. (2020.0). *Proceedings of the 28th ACM International Conference on Multimedia, MM 2020, Seattle, October 12-16*, (pp. 1132-1141) Virtual Conference: Association for Computing Machinery, Inc. (Published)

Compact bilinear augmented query structured attention for sport highlights classification, by HAO, Yanbin; ZHANG, Hao; NGO, Chong-Wah; LIU, Qing; HU, Xiaojun. (2020.0). *Proceedings of the 28th ACM International Conference on Multimedia, MM 2020, Seattle, October 12-16*, (pp. 628-636) Virtual Conference: Association for Computing Machinery, Inc. (Published)

Person-level action recognition in complex events via TSD-TSM networks, by HAO, Yanbin; LIU, Zi-Niu; ZHANG, Hao; ZHU, Bin; CHEN, Jingjing; JIANG, Yu-Gang; NGO, Chong-Wah. (2020.0). *Proceedings of the 28th ACM International Conference on Multimedia, MM 2020, Seattle, October 12-16*, (pp. 4699-4702) Virtual Conference: Association for Computing Machinery, Inc. (Published)

Cross-domain cross-modal food transfer, by ZHU, Bin; NGO, Chong-Wah ; CHEN, Jingjing. (2020.0). *Proceedings of the 28th ACM International Conference on Multimedia, MM 2020, Seattle, October 12-16*, (pp. 3762-3770) Virtual Conference: Association for Computing Machinery, Inc. (Published)

Hyperbolic visual embedding learning for zero-shot recognition, by LIU, Shaoteng; CHEN, Jingjing; PAN, Liangming; NGO, Chong-Wah; CHUA, Tat-Seng; JIANG, Yu-Gang. (2020.0). *Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, June 13-19*, (pp. 9270-9278) Virtual Conference: IEEE Computer Society. (Published)

CookGAN: Causality based Text-to-Image Synthesis, by ZHU, Bin; NGO, Chong-Wah. (2020.0). *Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, June 13-19*, (pp. 5518-5526) Virtual Conference: IEEE Computer Society. (Published)

Transferring and regularizing prediction for semantic segmentation, by ZHANG, Yiheng; QIU, Zhaofan; YAO, Ting; NGO, Chong-Wah; LIU, Dong; MEI, Tao. (2020.0). *Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, June 13-19*, (pp. 9618-9627) Virtual Conference: IEEE Computer Society. (Published)

Exploring category-agnostic clusters for open-set domain adaptation, by PAN, Yingwei; YAO, Ting; LI, Yehao; NGO, Chong-Wah; MEI, Tao. (2020.0). *Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, June 13-19*, (pp. 13864-13872) Virtual Conference: IEEE Computer Society. (Published)

Zero-shot ingredient recognition by multi-relational graph convolutional network, by CHEN, Jingjing; PAN, Liangming; WEI, Zhipeng; WANG, Xiang; NGO, Chong-Wah; CHUA, Tat-Seng. (2020.0). *Proceedings of the 34th AAAI Conference on Artificial Intelligence, AAAI 2020, New York, February 7-12, (pp. 10542-10550)* New York: AAAI press. (Published)

VIREO @ Video Browser Showdown 2020, by NGUYEN, Phuong Anh; WU, Jiaxin; NGO, Chong-Wah; FRANCIS, Danny; HUET, Benoit. (2020.0). *MultiMedia Modeling: 26th International Conference, MMM 2020, Daejeon, January 5-8: Proceedings, (pp. 772-777)* Cham: Springer. [https://doi.org/10.1007/978-3-030-37734-2\\_68](https://doi.org/10.1007/978-3-030-37734-2_68) (Published)

Vireo-EuReCOM @ TRECVID 2019: Ad-hoc video search (AVS), by NGUYEN, Phuong Anh; WU, Jiaxin; NGO, Chong-Wah; DANNY, Francis; HUET, Benoit. (2019.0). *Proceedings of 2019 TREC Video Retrieval Evaluation, TRECVID 2019, Gaithersburg, MD, USA, November 12-13, (pp. 1-6)* Gaithersburg: National Institute of Standards and Technology (NIST). (Published)

Fusion of multimodal embeddings for ad-hoc video search, by FRANCIS, Danny; NGUYEN, Phuong Anh; HUET, Benoit; NGO, Chong-Wah. (2019.0). *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Korea, October 27-28, (pp. 1868-1872)* Seoul: Institute of Electrical and Electronics Engineers Inc. (Published)

Learning spatio-temporal representation with local and global diffusion, by QIU, Zhaofan; YAO, Ting; NGO, Chong-Wah; TIAN, Xinmei; MEI, Tao. (2019.0). *Proceedings of the 32nd IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, California, June 16-21, (pp. 12048-12057)* Long Beach: IEEE Computer Society. (Published)

Exploring object relation in mean teacher for cross-domain detection, by CAI, Qi; PAN, Yingwei; NGO, Chong-Wah; TIAN, Xinmei; DUAN, Lingyu; YAO, Ting. (2019.0). *Proceedings of the 32nd IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, California, June 16-21, (pp. 11449-11458)* Long Beach: IEEE Computer Society. (Published)

R2GAN: Cross-modal recipe retrieval with generative adversarial network, by ZHU, Bin; NGO, Chong-Wah; CHEN, Jingjing; HAO, Yanbin. (2019.0). *Proceedings of the 32nd IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, California, June 16-21, (pp. 11469-11478)* Long Beach: IEEE Computer Society. (Published)

Transferable prototypical networks for unsupervised domain adaptation, by PAN, Yingwei; YAO, Ting; LI, Yehao; WANG, Yu; NGO, Chong-Wah; MEI, Tao. (2019.0). *Proceedings of the 32nd IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, California, June 16-21, (pp. 2234-2242)* Long Beach: IEEE Computer Society. (Published)

DietLens-eout: Large scale restaurant food photo recognition, by WEI, Zhipeng; CHEN, Jingjing; MING, Zhaoyan; NGO, Chong-Wah; CHUA, Tat-Seng; ZHOU, Fengfeng. (2019.0). *Proceedings of the 2019 on International Conference on Multimedia Retrieval, ICMR 2019, Ottawa, Canada, June 10-13, (pp. 399-403)* Ottawa: ACM. (Published)

Mixed dish recognition through multi-label learning, by WANG, Yunan; CHEN, Jingjing; NGO, Chong-Wah; CHUA, Tat-Seng; ZUO, Wanli; MING, Zhaoyan. (2019.0). *CEA 2019: Proceedings of the 11th Workshop on Multimedia for Cooking and Eating Activities, Ottawa, Canada, June 10, (pp. 1-8)* New York: ACM. <https://doi.org/10.1145/3326458.3326929> (Published)

Vireo @ video browser showdown 2019, by NGUYEN, Phuong Anh; NGO, Chong-Wah; FRANCIS, Danny; HUET, Benoit. (2019.0). *MultiMedia Modeling: 25th International Conference, MMM 2019, Thessaloniki, Greece, January 8-11: Proceedings, (pp. 609-615)* Cham: Springer. [https://doi.org/10.1007%2F978-3-030-05716-9\\_54](https://doi.org/10.1007%2F978-3-030-05716-9_54) (Published)

Interpretable multimodal retrieval for fashion products, by LIAO, Lizi; HE, Xiangnan; ZHAO, Bo; NGO, Chong-Wah; CHUA, Tat-Seng. (2018.0). *Proceedings of the 26th ACM international conference on Multimedia, Seoul, 2018 October 22-26, (pp. 1571-1579)* Seoul: Association for Computing Machinery, Inc. (Published)

Deep understanding of cooking procedure for cross-modal recipe retrieval, by CHEN, Jingjing; NGO, Chong-Wah; FENG, Fu-Li; CHUA, Tat-Seng. (2018.0). *MM '18: Proceedings of the 26th ACM international conference on Multimedia, Seoul, October 22-26, (pp. 1020-1028)* New York: ACM. <https://doi.org/10.1145/3240508.3240627> (Published)

Interpretable multimodal retrieval for fashion products, by LIAO, Lizi; HE, Xiangnan; ZHAO, Bo; NGO,

Chong-Wah; CHUA, Tat-Seng. (2018.0). *Proceedings of the 26th ACM International Conference on Multimedia, Seoul, Republic of Korea, 2018 October 22 - 26*, (pp. 1571-1579) United States: Association for Computing Machinery. <https://doi.org/10.1145/3240508.3240646> (Published)

Food photo recognition for dietary tracking: System and experiment, by MING, Zhao-Yan; CHEN, Jingjing; CAO, Yu; FORDE, Ciarán; NGO, Chong-Wah; CHUA, Tat Seng. (2018.0). *MultiMedia Modeling: 24th International Conference, MMM 2018, Bangkok, February 5-7: Proceedings*, (pp. 129-141) Cham: Springer. [https://doi.org/10.1007/978-3-319-73600-6\\_12](https://doi.org/10.1007/978-3-319-73600-6_12) (Published)

Enhanced vireo KIS at VBS 2018, by NGUYEN, Phuong Anh; LU, Yi-Jie; ZHANG, Hao; NGO, Chong-Wah. (2018.0). *MultiMedia Modeling: 24th International Conference, MMM 2018, Bangkok, Thailand, February 5-7: Proceedings*, (pp. 407-412) Cham: Springer. [https://doi.org/10.1007/978-3-319-73600-6\\_42](https://doi.org/10.1007/978-3-319-73600-6_42) (Published)

Color-sketch simulator: A guide for color-based visual known-item search, by LOKOČ, Jakub; PHUONG, Anh Nguyen; VOMLELOVÁ, Marta; NGO, Chong-Wah. (2017.0). *Advanced Data Mining and Applications: 13th International Conference, ADMA 2017, Singapore, November 5-6: Proceedings*, (pp. 754-763) Cham: Springer. [https://doi.org/10.1007/978-3-319-69179-4\\_53](https://doi.org/10.1007/978-3-319-69179-4_53) (Published)

Cross-modal recipe retrieval with rich food attributes, by CHEN, Jingjing; NGO, Chong-Wah; CHUA, Tat-Seng. (2017.0). *Proceedings of the 25th ACM International Conference on Multimedia, MM 2017, Mountain View, California, October 23-27*, (pp. 1771-1779) Mountain View: Association for Computing Machinery, Inc. (Published)

PIC2DISH: A customized cooking assistant system, by AN, Yongsheng; CAO, Yu; CHEN, Jingjing; NGO, Chong-Wah; JIA, Jia; LUAN, Huanbo; CHUA, Tat-Seng. (2017.0). *Proceedings of the 25th ACM International Conference on Multimedia, MM 2017, Mountain View, California, October 23-27*, (pp. 1269-1273) Mountain View: Association for Computing Machinery, Inc. (Published)

On the selection of anchors and targets for video hyperlinking, by CHENG, Zhi-Qi; ZHANG, Hao; WU, Xiao; NGO, Chong-Wah. (2017.0). *Proceedings of the 17th ACM International Conference on Multimedia Retrieval, ICMR 2017*, (pp. 287-293) Bucharest: Association for Computing Machinery, Inc. (Published)

Concept-based interactive search system, by LU, Yi-Jie; NGUYEN, Phuong Anh; ZHANG, Hao; NGO, Chong-Wah. (2017.0). *MultiMedia Modeling: 23rd International Conference, MMM 2017, Reykjavík, Iceland, January 4-6: Proceedings*, (pp. 463-468) Cham: Springer. [https://doi.org/10.1007/978-3-319-51814-5\\_42](https://doi.org/10.1007/978-3-319-51814-5_42) (Published)

Cross-modal recipe retrieval: How to cook this dish?, by CHEN, Jingjing; PANG, Lei; NGO, Chong-Wah. (2017.0). *MultiMedia Modeling: 23rd International Conference, MMM 2017, Reykjavík, Iceland, January 4-6: Proceedings*, (pp. 588-600) Cham: Springer. [https://doi.org/10.1007/978-3-319-51811-4\\_48](https://doi.org/10.1007/978-3-319-51811-4_48) (Published)

VIREO @ TRECVID 2016: Multimedia event detection, ad-hoc video search, video-to-text description, by ZHANG, Hao; LU, Yi-Jie; NGO, Chong-Wah. (2016.0). *Proceedings of TRECVID 2016: Gaithersburg, November 14-16*, (pp. 1-15) Gaithersburg: National Institute of Standards and Technology. <https://www-nlpir.nist.gov/projects/tvpubs/tv16.papers/vireo.pdf> (Published)

Deep-based ingredient recognition for cooking recipe retrieval, by CHEN, Jingjing; NGO, Chong-Wah. (2016.0). *Proceedings of the 24th ACM International conference on Multimedia, MM 2016, Amsterdam, October 15-19*, (pp. 32-41) Amsterdam: ACM. (Published)

Serendipity-driven celebrity video hyperlinking, by YANG, Shujun; PANG, Lei; NGO, Chong-Wah; HUET, Benoit. (2016.0). *Proceedings of the 2016 ACM International Conference on Multimedia Retrieval, New York, USA, June 6-9*, (pp. 413-416) New York: Association for Computing Machinery, Inc. (Published)

Event detection with zero example: Select the right and suppress the wrong concepts, by LU, Yi-Jie; ZHANG, Hao; DE BOER, Maaike; NGO, Chong-Wah. (2016.0). *Proceedings of the 6th ACM International Conference on Multimedia Retrieval, ICMR 2016, New York, June 6-9*, (pp. 127-134) New York: ACM. (Published)

Learning query and image similarities with ranking canonical correlation analysis, by YAO, Ting; MEI, Tao; NGO, Chong-Wah. (2015.0). *Proceedings of the 15th IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13*, (pp. 28-36) Santiago, Chile: Institute of Electrical and Electronics Engineers Inc. (Published)

Semi-supervised hashing with semantic confidence for large scale visual search, by PAN, Yingwei; YAO, Ting; LI, Houqiang; NGO, Chong-Wah; MEI, Tao. (2015.0). *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, Santiago, Chile, 2015 August 9-13*, (pp. 53-62) Santiago, Chile: ACM. (Published)

Multimodal learning with deep Boltzmann Machine for emotion prediction in user generated videos, by PANG, Lei; NGO, Chong-Wah. (2015.0). *Proceedings of the 5th ACM International Conference on Multimedia Retrieval, ICMR 2015, Shanghai, China, June 23-26*, (pp. 619-622) Shanghai: ACM. (Published)

Improving automatic name-face association using celebrity images on the Web, by CHEN, Zheneng; FENG, Bailan; NGO, Chong-Wah; JIA, Caiyan; HUANG, Xiangsheng. (2015.0). *Proceedings of the 5th ACM International Conference on Multimedia Retrieval, ICMR 2015, Shanghai, China, June 23-26*, (pp. 623-626) Shanghai: ACM. (Published)

Semi-supervised domain adaptation with subspace learning for visual recognition, by YAO, Ting; PAN, Yingwei; NGO, Chong-Wah; LI, Houqiang; MEI, Tao. (2015.0). *Proceedings of the 28th IEEE Conference on Computer Vision and Pattern Recognition: CVPR 2015, Boston, June 7-12*, (pp. 2142-2150) Boston: IEEE. (Published)

CeleLabel: An interactive system for annotating celebrities in web videos, by CHEN, Zheneng; BAI, Jinfeng; NGO, Chong-Wah; FENG, Bailan; XU, Bo. (2014.0). *Proceedings of the 22nd ACM international conference on Multimedia, MM14, Orlando, Florida, November 3-7*, (pp. 759-760) Orlando: ACM. (Published)

Organizing video search results to adapted semantic hierarchies for topic-based browsing, by WANG, Jiajun; JIANG, Yu-Gang; WANG, Qiang; YANG, Kuiyuan; NGO, Chong-Wah. (2014.0). *Proceedings of the 22nd ACM international conference on Multimedia, MM 2014, Orlando, Florida, November 3-7*, (pp. 845-848) Orlando: ACM. (Published)

Click-through-based subspace learning for image search, by PAN, Yingwei; YAO, Ting; TIAN, Xinmei; LI, Houqiang; NGO, Chong-Wah. (2014.0). *Proceedings of the 22nd ACM international conference on Multimedia, MM 2014, Orlando, Florida, November 3-7*, (pp. 233-236) Orlando: ACM. (Published)

Scalable visual instance mining with threads of features, by ZHANG, Wei; LI, Hongzhi; NGO, Chong-Wah; CHANG, Shih-Fu. (2014.0). *Proceedings of the 22nd ACM international conference on Multimedia, MM 2014, Orlando, Florida, November 3-7*, (pp. 297-306) Orlando: ACM. (Published)

Click-through-based cross-view learning for image search, by PAN, Yingwei; YAO, Ting; MEI, Tao; LI, Houqiang; NGO, Chong-Wah; RUI, Yong. (2014.0). *Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2014, Gold Coast, Australia, July 6-11*, (pp. 717-726) Gold Coast, Australia: ACM. (Published)

CeleBrowser: An example of browsing big data on small device, by TAN, Song; NGO, Chong-Wah; XU, Jun; RUI, Yong. (2014.0). *Proceedings of the 4th ACM International Conference on Multimedia Retrieval, ICMR 2014, Glasgow, United Kingdom, April 1-4*, (pp. 514-517) Glasgow: ACM. (Published)

Image search by graph-based label propagation with image representation from DNN, by PAN, Yingwei; TING, Yao; YANG, Kuiyuan; LI, Houqiang; NGO, Chong-Wah; WANG, Jingdong; MEI, Tao. (2013.0). *MM '13: Proceedings of the 21st ACM International Conference on Multimedia: October 21-25, Barcelona, Spain*, (pp. 397-400) Barcelona, Spain: ACM. (Published)

Error recovered hierarchical classification, by ZHU, Shuai; WEI, Xiao-Yong; NGO, Chong-Wah. (2013.0). *MM '13: Proceedings of the 21st ACM International Conference on Multimedia: October 21-25, Barcelona, Spain*, (pp. 697-700) Barcelona, Spain: ACM. (Published)

Annotation for free: Video tagging by mining user search behavior, by TING, Yao; MEI, Tao; NGO, Chong-Wah; LI, Shipeng. (2013.0). *MM '13: Proceedings of the 21st ACM International Conference on Multimedia: October 21-25, Barcelona, Spain*, (pp. 977-986) Barcelona, Spain: ACM. (Published)

Click-boosting random walk for image search reranking, by YANG, Xiaopeng; ZHANG, Yongdong; YAO, Ting; ZHA, Zheng-Jun; NGO, Chong-Wah. (2013.0). *Proceedings of the 5th International Conference on Internet Multimedia Computing and Service, ICIMCS 2013, Huangshan, Anhui, China, August 17-19*, (pp. 1-6) Huangshan, Anhui, China: ACM. (Published)

Video concept detection by learning from web images: A case study on cross domain learning, by ZHU, Shuai; YAO, Ting; NGO, Chong-Wah. (2013.0). *Proceedings of the 2013 IEEE International Conference on Multimedia and Expo Workshops, ICMEW 2013, San Jose, CA, July 15-19*, (pp. 1-6) San Jose, CA: IEEE.

(Published)

Unified entity search in social media community, by YAO, Ting; LIU, Yuan; NGO, Chong-Wah; MEI, Tao. (2013.0). *Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil, 2013 May13-17*, (pp. 1457-1466) Rio de Janeiro, Brazil: ACM. (Published)

Searching visual instances with topology checking and context modeling, by ZHANG, Wei; NGO, Chong-Wah. (2013.0). *Proceedings of the 3rd ACM International Conference on Multimedia Retrieval, ICMR 2013, Dallas, Texas, April 16-20*, (pp. 57-64) Dallas, Texas: ACM. (Published)

FashionAsk: Pushing community answers to your fingertips, by ZHANG, Wei; PANG, Lei; NGO, Chong-Wah. (2012.0). *Proceedings of the 20th ACM international conference on Multimedia, MM 2012, Nara, Japan, October 29 - November 2*, (pp. 1345-1346) Nara, Japan: ACM. (Published)

Video hyperlinking: Libraries and tools for threading and visualizing large video collection, by PANG, Lei; ZHANG, Wei; TAN, Hung-Khoon; NGO, Chong-Wah. (2012.0). *Proceedings of the 20th ACM international conference on Multimedia, MM 2012, Nara, Japan, October 29 - November 2*, (pp. 1461-1464) Nara, Japan: ACM. (Published)

Predicting domain adaptivity: Redo or recycle?, by YAO, Ting; NGO, Chong-Wah; ZHU, Shuai. (2012.0). *Proceedings of the 20th ACM international conference on Multimedia, MM 2012, Nara, Japan, October 29 - November 2*, (pp. 821-824) Nara, Japan: ACM. (Published)

Snap-and-ask: Answering multimodal question by naming visual instance, by ZHANG, Wei; PANG, Lei; NGO, Chong-Wah. (2012.0). *Proceedings of the 20th ACM international conference on Multimedia, MM 2012, Nara, Japan, October 29 - November 2*, (pp. 609-618) Nara, Japan: ACM. (Published)

Community as a connector: Associating faces with celebrity names in web videos, by CHEN, Zheneng; NGO, Chong-Wah; CAO, Juan; ZHANG, Wei. (2012.0). *Proceedings of the 20th ACM international conference on Multimedia, MM 2012, Nara, Japan, October 29 - November 2*, (pp. 809-812) Nara, Japan: ACM. (Published)

Trajectory-based modeling of human actions with motion reference points, by JIANG, Yu-Gang; DAI, Qi; XUE, Xiangyang; LIU, Wei; NGO, Chong-Wah. (2012.0). *Computer Vision: 12th European Conference, ECCV 2012, Florence, Italy, October 7-13: Proceedings*, (pp. 425-438) Cham: Springer. [https://doi.org/10.1007/978-3-642-33715-4\\_31](https://doi.org/10.1007/978-3-642-33715-4_31) (Published)

On the pooling of positive examples with ontology for visual concept learning, by ZHU, Shuai; NGO, Chong-Wah; JIANG, Yu-Gang. (2011.0). *Proceedings of the 19th ACM International Conference on Multimedia ACM Multimedia 2011, MM'11, Scottsdale, Arizona, November 28 - December 1*, (pp. 1045-1048) Scottsdale, Arizona: ACM. (Published)

Galaxy browser: Exploratory search of web videos, by PANG, Lei; TAN, Song; TAN, Hung-Khoon; NGO, Chong-Wah. (2011.0). *Proceedings of the 19th ACM International Conference on Multimedia ACM Multimedia 2011, MM'11, Scottsdale, Arizona, November 28 - December 1*, (pp. 803-804) Scottsdale, Arizona: ACM. (Published)

Context-based friend suggestion in online photo-sharing community, by YAO, Ting; NGO, Chong-Wah; MEI, Tao. (2011.0). *Proceedings of the 19th ACM International Conference on Multimedia ACM Multimedia 2011, MM'11, Scottsdale, Arizona, November 28 - December 1*, (pp. 945-948) Scottsdale, Arizona: ACM. (Published)

Towards textually describing complex video contents with audio-visual concept classifiers, by TAN, Chun Chet; JIANG, Yu-Gang; NGO, Chong-Wah. (2011.0). *Proceedings of the 19th ACM International Conference on Multimedia ACM Multimedia 2011, MM'11, Scottsdale, Arizona, November 28 - December 1*, (pp. 655-658) Scottsdale, Arizona: ACM. (Published)

Cross media hyperlinking for search topic browsing, by TAN, Song; NGO, Chong-Wah; TAN, Hung-Khoon; PANG, Lei. (2011.0). *Proceedings of the 19th ACM International Conference on Multimedia ACM Multimedia 2011, MM'11, Scottsdale, Arizona, November 28 - December 1*, (pp. 243-252) Scottsdale, Arizona: ACM. (Published)

Fusing heterogeneous modalities for video and image re-ranking, by TAN, Hung-Khoon; NGO, Chong-Wah. (2011.0). *Proceedings of the 1st ACM International Conference on Multimedia Retrieval: ICMR '11, Trento, Italy, April 17-20*, (pp. 1-8) Trento, Italy: ACM. (Published)

Topical summarization of web videos by visual-text time-dependent alignment, by TAN, Song; TAN, Hung-Khoon; NGO, Chong-Wah. (2010.0). *Proceedings of the 18th ACM International Conference on Multimedia ACM Multimedia, MM2010, Firenze, Italy, October 25-29*, (pp. 1095-1098) Firenze, Italy: ACM. (Published)

Trajectory-based visualization of web video topics, by CAO, Juan; NGO, Chong-Wah; ZHANG, Yong-Dong; ZHANG, Dong-Ming; MA, Liang. (2010.0). *Proceedings of the 18th ACM International Conference on Multimedia ACM Multimedia, MM2010, Firenze, Italy, October 25-29*, (pp. 1639-1642) Firenze, Italy: ACM. (Published)

Automatic generation of semantic fields for annotating web images, by WANG, Gang; CHUA, Tat Seng; NGO, Chong-Wah; WANG, Yong Cheng. (2010.0). *Proceedings of the 23rd International Conference on Computational Linguistics, Coling 2010, Beijing, August 23-27*, (pp. 1301-1309) Beijing, China: ACM. (Published)

Co-reranking by mutual reinforcement for image search, by YAO, Ting; MEI, Tao; NGO, Chong-Wah. (2010.0). *Proceedings of the ACM International Conference on Image and Video Retrieval, ACM-CIVR 2010, Xi'an, China, July 5-7*, (pp. 34-41) Xi'an, China: ACM. (Published)

On the sampling of web images for learning visual concept classifiers, by ZHU, Shuai; WANG, Gang; NGO, Chong-Wah; JIANG, Yu-Gang. (2010.0). *Proceedings of the ACM International Conference on Image and Video Retrieval, ACM-CIVR 2010, Xi'an, China, July 5-7*, (pp. 50-57) Xi'an, China: ACM. (Published)

Coherent bag-of audio words model for efficient large-scale video copy detection, by LIU, Yang; ZHAO, Wan-Lei; NGO, Chong-Wah; XU, Chang-Sheng; LU, Han-Qing. (2010.0). *Proceedings of the ACM International Conference on Image and Video Retrieval, ACM-CIVR 2010, Xi'an, China, July 5-7*, (pp. 89-96) Xi'an, China: ACM. (Published)

Semantic context modeling with maximal margin conditional random fields for automatic image annotation, by XIANG, Yu; ZHOU, Xiangdong; LIU, Zuotao; CHUA, Tat-Seng; NGO, Chong-Wah. (2010.0). *Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2010, San Francisco, June 13-18*, (pp. 3368-3375) San Francisco: IEEE. (Published)

PageSense: Style-wise web page advertising, by LI, Lusong; MEI, Tao; NIU, Xiang; NGO, Chong-Wah. (2010.0). *Proceedings of the 19th International World Wide Web Conference, WWW2010, Raleigh, April 26-30*, (pp. 1273-1276) Raleigh: ACM. (Published)

Scalable detection of partial near-duplicate videos by visual-temporal consistency, by TAN, Hung-Khoon; NGO, Chong-Wah; HONG, Richang; CHUA, Tat-Seng. (2009.0). *Proceedings of the 17th ACM international conference on Multimedia, MM'09, Beijing, China, October 19-24*, (pp. 145-154) Beijing, China: ACM. (Published)

Localizing volumetric motion for action recognition in realistic videos, by WU, Xiao; NGO, Chong-Wah; LI, Jintao; ZHANG, Yongdong. (2009.0). *Proceedings of the 17th ACM International Conference on Multimedia, MM'09, Beijing China, 2009 October 19-24*, (pp. 505-508) Beijing, China: ACM. (Published)

Towards Google challenge: Combining contextual and social information for web video categorization, by WU, Xiao; ZHAO, Wan-Lei; NGO, Chong-Wah. (2009.0). *Proceedings of the 17th ACM international conference on Multimedia, MM'09, Beijing, China, October 19-24*, (pp. 1109-1110) Beijing, China: ACM. (Published)

Semantic context transfer across heterogeneous sources for domain adaptive video search, by JIANG, Yu-Gang; NGO, Chong-Wah; CHANG, Shih-Fu. (2009.0). *Proceedings of the 17th ACM international conference on Multimedia, MM'09, Beijing, China, October 19-24*, (pp. 155-164) Beijing, China: ACM. (Published)

Distribution-based concept selection for concept-based video retrieval, by CAO, Juan; JING, HongFang; NGO, Chong-Wah; ZHANG, YongDong. (2009.0). *Proceedings of the 17th ACM international conference on Multimedia, MM'09, Beijing, China, October 19-24*, (pp. 645-648) Beijing, China: ACM. (Published)

Domain adaptive semantic diffusion for large scale context-based video annotation, by JIANG, Yu-Gang; WANG, Jun; CHANG, Shih-Fu; NGO, Chong-Wah. (2009.0). *Proceedings of the 12th International Conference on Computer Vision, ICCV 2009, Kyoto, Japan, September 29 - October 2*, (pp. 1420-1427) Kyoto, Japan: IEEE. (Published)

Exploring inter-concept relationship with context space for semantic video indexing, by WEI, Xiao-Yong;

JIANG, Yu-Gang; NGO, Chong-Wah. (2009.0). *Proceedings of the ACM International Conference on Image and Video Retrieval, CIVR 2009, Santorini, July 8-10*, (pp. 108-115) Santorini: ACM. (Published)

Large-scale near-duplicate web video search: Challenge and opportunity, by ZHAO, Wan-Lei; TAN, Song; NGO, Chong-Wah. (2009.0). *Proceedings of 2009 IEEE International Conference on Multimedia and Expo, ICME 2009, New York, June 28 - July 3*, (pp. 1624-1627) New York: IEEE. (Published)

A revisit of generative model for automatic image annotation using markov random fields, by XIANG, Yu; ZHOU, Xiangdong; CHUA, Tat-Seng; NGO, Chong-Wah. (2009.0). *Proceedings of 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Miami, June 20-25*, (pp. 1153-1160) Miami: IEEE Computer Society. (Published)

Video event detection using motion relativity and visual relatedness, by WANG, Feng; JIANG, Yu-Gang; NGO, Chong-Wah. (2008.0). *Proceedings of the 16th ACM International Conference on Multimedia, MM '08, Vancouver, 2008 October 26-31*, (pp. 239-248) Vancouver, Canada: ACM. (Published)

Accelerating near-duplicate video matching by combining visual similarity and alignment distortion, by TAN, Hung-Khoon; WU, Xiao; NGO, Chong-Wah; ZHAO, Wan-Lei. (2008.0). *Proceedings of the 16th ACM International Conference on Multimedia, MM '08, Vancouver, 2008 October 26-31*, (pp. 861-864) Vancouver, Canada: (Published)

Modeling video hyperlinks with hypergraph for web video reranking, by TAN, Hung-Khoon; NGO, Chong-Wah; WU, Xiao. (2008.0). *Proceedings of the 16th ACM International Conference on Multimedia, MM '08, Vancouver, 2008 October 26-31*, (pp. 659-662) Vancouver, Canada: ACM. (Published)

Fusing semantics, observability, reliability and diversity of concept detectors for video search, by WEI, Xiao-Yong; NGO, Chong-Wah. (2008.0). *Proceedings of the 16th ACM International Conference on Multimedia, MM '08, Vancouver, 2008 October 26-31*, (pp. 81-90) Vancouver, Canada: ACM. (Published)

Ontology-based visual word matching for near-duplicate retrieval, by JIANG, Yu-Gang; NGO, Chong-Wah. (2008.0). *Proceedings of 2008 IEEE International Conference on Multimedia and Expo, ICME 2008, Hannover, Germany, June 23-26*, (pp. 125-128) Hannover, Germany: IEEE. (Published)

Bag-of-visual-words expansion using visual relatedness for video indexing, by JIANG, Yu-Gang; NGO, Chong-Wah. (2008.0). *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR 2008, Singapore July 20-24*, (pp. 769-770) Singapore: ACM. (Published)

Evaluating bag-of-visual-words representations in scene classification, by YANG, Jun; JIANG, Yu-Gang; HAUPTMANN, Alexander G.; NGO, Chong-Wah. (2007.0). *Proceedings of the international workshop on Workshop on multimedia information retrieval: MIR07, Augsburg, Bavaria, September 28-29*, (pp. 197-206) Augsburg, Bavaria: ACM. (Published)

Novelty detection for cross-lingual news stories with visual duplicates and speech transcripts, by WU, Xiao; HAUPTMANN, Alexander G.; NGO, Chong-Wah. (2007.0). *Proceedings of the 15th ACM International Conference on Multimedia, Augsburg Germany, 2007 September 25-29*, (pp. 168-177) New York: ACM. (Published)

Rushes video summarization by object and event understanding, by WANG, Feng; NGO, Chong-Wah. (2007.0). *Proceedings of the international workshop on TRECVID video summarization, TVS'07, Augsburg, Bavaria, September 28*, (pp. 25-29) Augsburg, Bavaria: ACM. (Published)

Ontology-enriched semantic space for video search, by WEI, Xiao-Yong; CHO, Chong-Wah. (2007.0). *Proceedings of the 15th ACM International Conference on Multimedia, MM2007, Augsburg, Bavaria, September 23-28*, (pp. 981-990) Augsburg, Bavaria: ACM. (Published)

Practical elimination of near-duplicates from Web video search, by WU, Xiao; HAUPTMANN, Alexander G.; NGO, Chong-Wah. (2007.0). *Proceedings of the 15th ACM International Conference on Multimedia, MM2007, Augsburg, Bavaria, September 23-28*, (pp. 218-227) Augsburg, Bavaria: ACM. (Published)

Near-duplicate keyframe retrieval with visual keywords and semantic context, by WU, Xiao; ZHAO, Wan-Lei; NGO, Chong-Wah. (2007.0). *Proceedings of the 6th ACM international conference on Image and video retrieval, CIVR 2007, Amsterdam, July 9 - 11*, (pp. 162-169) Amsterdam: ACM. (Published)

Towards optimal bag-of-features for object categorization and semantic video retrieval, by JIANG, Yu-Gang; NGO, Chong-Wah; YANG, Jun. (2007.0). *Proceedings of the 6th ACM international conference on*

*Image and video retrieval, CIVR 2007, Amsterdam, July 9 - 11, (pp. 494-501) Amsterdam: ACM. (Published)*

Efficient near-duplicate keyframe retrieval with visual language models, by WU, Xiao; ZHAO, Wan-Lei; NGO, Chong-Wah. (2007.0). *Proceedings of 2007 International Conference on Multimedia & Expo, Beijing July 2-5, (pp. 500-503) Beijing, China: IEEE Computer Society. (Published)*

Mining multiple visual appearances of semantics for image annotation, by TAN, Hung-Khoon; NGO, Chong-Wah. (2007.0). *Multimedia Modeling: 13th International Conference, MMM 2007, Singapore, January 9-12: Proceedings, (pp. 269-278) Cham: Springer. [https://doi.org/10.1007/978-3-540-69423-6\\_27](https://doi.org/10.1007/978-3-540-69423-6_27) (Published)*

Fast tracking of near-duplicate keyframes in broadcast domain with transitivity propagation, by NGO, Chong-Wah; ZHAO, Wan-Lei; JIANG, Yu-Gang. (2006.0). *Proceedings of the 14th ACM international conference on Multimedia, MM 2006, Santa Barbara, California, October 23-27, (pp. 845-854) Santa Barbara, California: ACM. (Published)*

Audio similarity measure by graph modeling and matching, by PENG, Yuxin; NGO, Chong-Wah; FANG, Cuihua; CHEN, Xiaou; XIAO, Jianguo. (2006.0). *Proceedings of the 14th ACM international conference on Multimedia, MM 2006, Santa Barbara, California, October 23-27, (pp. 603-606) Santa Barbara, California: ACM. (Published)*

A mobile lecture slide organization tool for students, by CHOW, K.O.; NGO, C.W.; IP H.H.S.; CHEUNG, K.S.. (2006.0). *Proceedings of 2006 1st International Symposium on Pervasive Computing and Applications, Urumqi, China, August 3-5, (pp. 354-357) Urumqi, China: IEEE Computer Society. (Published)*

Keyframe retrieval by keypoints: Can point-to-point matching help?, by ZHAO, Wanlei; JIANG, Yu-Gang; NGO, Chong-Wah. (2006.0). *Image and Video Retrieval: 5th International Conference, CIVR 2006, Tempe, AZ, July 13-15: Proceedings, (pp. 72-81) Cham: Springer. [https://doi.org/10.1007%2F11788034\\_8](https://doi.org/10.1007%2F11788034_8) (Published)*

Hierarchical hidden Markov model for rushes structuring and indexing, by NGO, Chong-Wah; PAN, Zailiang; WEI, Xiaoyong. (2006.0). *Image and Video Retrieval: 5th International Conference on Image and Video Retrieval, CIVR 2006, Tempe, AZ, July 13-15: Proceedings, (pp. 241-250) Cham: Springer. [https://doi.org/10.1007/11788034\\_25](https://doi.org/10.1007/11788034_25) (Published)*

Prediction-based gesture detection in lecture videos by combining visual, speech and electronic slides, by WANG, Feng; NGO, Chong-Wah; PONG, Ting-Chuen. (2006.0). *Proceedings of the 2006 IEEE International Conference on Multimedia and Expo, ICME 2006, Toronto, Canada, July 9-12, (pp. 653-656) Toronto, Canada: IEEE. (Published)*

Exploiting self-adaptive posture-based focus estimation for lecture video editing, by WANG, Feng; NGO, Chong-Wah; PONG, Ting-Chuen. (2005.0). *Proceedings of the 13th ACM International Conference on Multimedia, MM 2005, Singapore, November 6-11, (pp. 327-330) Singapore: ACM. (Published)*

Common pattern discovery using earth mover's distance and local flow maximization, by TAN, Hung-Khoon; NGO, Chong-Wah. (2005.0). *Proceedings of the 10th IEEE International Conference on Computer Vision, ICCV 2005, Beijing, China, October 17-21, (pp. 1222-1229) Beijing, China: IEEE. (Published)*

EMD-based video clip retrieval by many-to-many matching, by PENG, Yuxin; NGO, Chong-Wah. (2005.0). *Proceedings of the 4th International Conference on Image and Video Retrieval, CIVR 2005, Singapore, July 20-22, (pp. 71-81) Singapore: Springer Verlag. (Published)*

Hot event detection and summarization by graph modeling and matching, by PENG, Yuxin; NGO, Chong-Wah. (2005.0). *Image and Video Retrieval: 4th International Conference, CIVR 2005, Singapore, July 20-22: Proceedings, (pp. 257-266) Cham: Springer. [https://doi.org/10.1007/11526346\\_29](https://doi.org/10.1007/11526346_29) (Published)*

Multibiometrics based on palmprint and handgeometry, by WEI, Xiao-Yong; XU, Dan; NGO, Chong-Wah. (2005.0). *Proceedings of the 4th Annual ACIS International Conference on Computer and Information Science, ICIS 2005, Jeju Island, South Korea, 2005 July 14-16, (pp. 495-500) Jeju Island, South Korea: IEEE. (Published)*

Co-clustering of time-evolving news story with transcript and keyframe, by WU, Xiao; NGO, Chong-Wah; LI, Qing. (2005.0). *Proceedings of 2005 IEEE International Conference on Multimedia and Expo, Amsterdam, July 6, (pp. 117-120) Amsterdam: IEEE. (Published)*

A robust method for recovering geometric proxy from multiple panoramic images, by WAN, Ada S. K.; SIU, Angus M. K.; LAU, Rynson W. H.; NGO, Chong-Wah. (2004.0). *Proceedings of 2004 International Conference on Image Processing, ICIP '04, Singapore October 24-27*, (pp. 1369-1372) Singapore: IEEE. (Published)

Deformable object model matching by topological and geometric similarity, by TAN, Kwok-Leung; LAU, Rynson W. H.; NGO, Chong-Wah. (2004.0). *Proceedings of Computer Graphics International, GI 2004, Crete, June 19*, (pp. 335-342) Crete: IEEE. (Published)

Clip-based similarity measure for hierarchical video retrieval, by PENG, Yuxin; NGO, Chong-Wah. (2004.0). *Proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval, New York, October 15-16*, (pp. 53-60) New York: ACM. (Published)

Structuring home video by snippet detection and pattern parsing, by PAN, Zailiang; NGO, Chong-Wah. (2004.0). *Proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval, New York, 2004 October 15-16*, (pp. 69-76) New York: Association for Computing Machinery (ACM). (Published)

Indexing and matching of polyphonic songs for query-by-singing system, by LEUNG, Tat-Wan; NGO, Chong-Wah. (2004.0). *Proceedings of the 12th ACM International Conference on Multimedia, New York, 2004 October 10-16*, (pp. 308-311) New York: ACM. (Published)

Novel seed selection for multiple objects detection and tracking, by PAN, Zailiang; NGO, Chong-Wah. (2004.0). *Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004, Cambridge, UK, 2004 August 23-26*, (pp. 744-747) Cambridge: IEEE. (Published)

Graph based image matching, by JIANG, Hui; NGO, Chong-Wah. (2004.0). *Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004, Cambridge, UK, 2004 August 23-26*, (pp. 658-661) Cambridge: IEEE. (Published)

ICA-FX features for classification of singing voice and instrumental sound, by LEUNG, Tat-Wan; NGO, Chong-Wah; LAU, Rynson W. H. (2004.0). *Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004, Cambridge, UK, 2004 August 23-26*, (pp. 367-370) Cambridge: IEEE. (Published)

Deformable geometry model matching by topological and geometric signatures, by TAM, Kwok-Leung; LAW, Rynson W.H.; NGO, Chong-Wah. (2004.0). *Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004, Cambridge, UK, 2004 August 23-26*, (pp. 910-913) Cambridge: IEEE. (Published)

Gesture tracking and recognition for lecture video editing, by WANG, Feng; NGO, Chong-Wah; PONG, Ting-Chuen. (2004.0). *Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004, Cambridge, UK, 2004 August 23-26*, (pp. 934-937) Cambridge: IEEE. (Published)

Structuring lecture videos for distance learning applications, by NGO, Chong-Wah; WANG, Feng; PONG, Ting-Chuen. (2003.0). *Proceedings of 5th IEEE International Symposium on Multimedia Software Engineering, ISMSE 2003, Taichung, Taiwan, December 10-12*, (pp. 215-222) Taichung: IEEE. (Published)

Synchronization of lecture videos and electronic slides by video text analysis, by WANG, Feng; NGO, Chong-Wah; PONG, Ting-Chuen. (2003.0). *Proceedings of the 11th ACM International Conference on Multimedia, MM'03, Berkeley, California, November 2-8*, (pp. 315-318) Berkeley, California: ACM. (Published)

A robust dissolve detector by support vector machine, by NGO, Chong-Wah. (2003.0). *Proceedings of the 11th ACM International Conference on Multimedia, MM'03, Berkeley, California, November 2-8*, (pp. 283-286) Berkeley, California: ACM. (Published)

Automatic video summarization by graph modeling, by NGO, Chong-Wah; MA, Yu-Fei; ZHANG, Hong-Jiang. (2003.0). *Proceedings of the 9th IEEE International Conference on Computer Vision, Nice, France, 2003 October 13-16*, (pp. 104-109) Nice, France: IEEE. (Published)

Detection of documentary scene changes by audio-visual fusion, by VELIVELLI, Atulya; NGO, Chong-Wah; HUANG, Thomas S.. (2003.0). *Image and Video Retrieval: 2nd International Conference, CIVR 2003, Urbana-Champaign, IL, July 24-25: Proceedings*, (pp. 227-237) Cham: Springer. [https://doi.org/10.1007/3-540-45113-7\\_23](https://doi.org/10.1007/3-540-45113-7_23) (Published)

Trifocal morphing, by SIU, Angus M. K.; WAN, Ada S. K.; LAU, Rynson W. H.; NGO, Chong-Wah. (2003.0).

*Proceedings of the 7th International Conference on Information Visualization, IV 2003, London, July 16-18*, (pp. 24-29) London: Institute of Electrical and Electronics Engineers Inc. (Published)

Video clip retrieval by maximal matching and optimal matching in graph theory, by PENG, Yu-Xin; NGO, Chong- Wah; DONG, Qing-Jie; GUO, Zong-Ming; XIAO, Jian-Guo. (2003.0). *Proceedings of 2003 International Conference on Multimedia and Expo, Baltimore, Maryland, July 6-9*, (pp. 317-320) Baltimore: IEEE Computer Society. (Published)

Motion retrieval by temporal slices analysis, by NGO, Chong-Wah; PONG, Ting-Chuen; ZHANG, Hong-Jiang. (2002.0). *Proceedings of the 16th International Conference on Pattern Recognition, Quebec City, Quebec, Canada, 2002 August 11-15*, (pp. 64-67) Quebec City, Quebec, Canada: IEEE. (Published)

Detection of slide transition for topic indexing, by NGO, Chong-Wah; PONG, Ting-Chuen; HUANG, Thomas S.. (2002.0). *Proceedings of 2002 IEEE International Conference on Multimedia and Expo, ICME 2002 Lausanne, Switzerland, August 26-29*, (pp. 533-536) Lausanne: Institute of Electrical and Electronics Engineers Inc. (Published)

Integrating color and spatial features for content-based video retrieval, by LIN, Tong; NGO, Chong-Wah; ZHANG, Hong-Jiang; SHI, Qing-Yun. (2001.0). *Proceedings 2001 International Conference on Image Processing, Thessaloniki, Greece, October 7-10*, (pp. 592-595) Thessaloniki, Greece: IEEE. (Published)

On clustering and retrieval of video shots, by NGO, Chong-Wah; PONG, Ting-Chuen; ZHANG, Hong-Jiang. (2001.0). *Proceedings of the 9th ACM International Conference on Multimedia, Ottawa, Canada, 2001 September 30 - October 5*, (pp. 51-60) Ottawa, Canada: Association for Computing Machinery (ACM). (Published)

Motion-based video representation for scene change detection, by NGO, Chong-Wah; PONG, Ting-Chuen; ZHANG, Hong-Jiang; CHIN, Roland T.. (2000.0). *Proceedings 15th International Conference on Pattern Recognition, ICPR-2000, Barcelona, Spain, September 3-7*, (pp. 827-830) Barcelona, Spain: IEEE. (Published)

Motion characterization by temporal slices analysis, by NGO, Chong-Wah; PONG, Ting-Chuen; ZHANG, Hong-Jiang; CHIN, Roland T.. (2000.0). *2000 IEEE Conference on Computer Vision and Pattern Recognition CVPR: Hilton Head, SC, June 15: Proceedings*, (pp. 768-773) Los Alamitos, CA: IEEE Computer Society. <https://doi.org/10.1109/CVPR.2000.854952> (Published)

Detection of gradual transitions through temporal slice analysis, by NGO, C. W.; PONG, T. C.; CHIN, R. T.. (1999.0). *Proceedings of the 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'99), Fort Collins, Colorado, June 23-25*, (pp. 36-41) Fort Collins, Colorado: IEEE. (Published)

Camera break detection by partitioning of 2D spatio-temporal images in MPEG domain, by NGO, Chong-Wah. (1999.0). *Proceedings of the 1999 International Conference on Multimedia Computing and Systems, Florence, Italy, June 7-11*, (pp. 750-755) Los Alamitos, CA, United States: IEEE. (Published)

A high speed distributed file system for multimedia communications, by CHEE, C. L.; ERDOGAN, S. S.; NGO, C. W.; WONG, C. K. . (1995.0). *Proceedings of TENCON'94 - 1994 IEEE Region 10's 9th Annual International Conference on: 'Frontiers of Computer Technology', Singapore, August 22-26*, (pp. 254-258) Piscataway, NJ, United States: IEEE. (Published)

## Conference Papers

GroundNLQ @ Ego4D natural language queries challenge 2023, by HOU, Zhijian; JI, Lei; GAO, Difei; ZHONG, Wanjun; YAN, Kun; LI, Chao; CHAN, Wing-Kwong; NGO, Chong-Wah; DUAN, Nan; SHOU, Mike Zheng. (2023.0). *3rd International Ego4D workshop, Vancouver, 2023 June 19*, Vancouver. (Presented)

VIREO @ TRECVID 2021 ad-hoc video search, by WU, Jiaxin; NGUYEN, Phuong Anh; NGO, Chong-Wah. (2020.0). *TRECVID 2021, Gaithersburg, December 7-10*, Gaithersburg. (Published)

Vireo-TNO @ TRecviD 2014 zero-shot event detection and recounting, by LU, Yi-Jie; ZHANG, Hao; NGO, Chong-Wah; DE BOER, Maaike; SCHAVEMAKER, John; SCHUTTE, Klamer; KRAAIJ, Wessel. (2020.0). *2014 TREC Video Retrieval Evaluation, TRECVID 2014, November 10-14*, Orlando. (Published)

VireoJD-MM @ TRECVID 2019: Activities in extended video (ACTEV), by HOU, Zhijian; PAN, Ying-Wei; YAO, Ting; NGO, Chong-Wah. (2020.0). *TRECVID 2019 TREC Video Retrieval Evaluation, Gaithersburg, November 12-13, Gaithersburg*. (Published)

EURECOM at TRECVID AVS 2019, by FRANCIS, Danny; NGUYEN, Phuong Anh; HUET, Benoit; NGO, Chong-Wah. (2019.0). *2019 TREC Video Retrieval Evaluation, TRECVID 2019, Gaithersburg, November 12-13, Gaithersburg*. (Published)

Vireo @ TRecViD 2017: Video-to-text, ad-hoc video search and video hyperlinking, by NGUYEN, Phuong Anh; LI, Qing; CHENG, Zhi-Qi; LU, Yi-Jie; ZHANG, Hao; WU, Xiao; NGO, Chong-Wah. (2017.0). *2017 TREC Video Retrieval Evaluation, TRECVID 2017, Gaithersburg, November 13-15, Gaithersburg*. (Published)

Minimizing risk in video hyperlinking, by NGO, Chong-Wah; CHENG, Zhi-Qi; WU, Xiao. (2017.0). *2017 TREC Video Retrieval Evaluation, TRECVID 2017, Gaithersburg, November 13-15, Gaithersburg*. (Published)

VIREO@TRECVID 2016: Multimedia event detection, ad-hoc video search, video to text description, by ZHANG, Hao; PANG, Lei; LU, Yi-Jie; NGO, Chong-Wah. (2016.0). *2016 TREC Video Retrieval Evaluation, TRECVID 2016, Gaithersburg, USA, November 14-16, Gaithersburg*. (Published)

Multimedia event detection: Strong by integration, by ZHANG, Hao; DE BOER, Maaike; LU, Yi-Jie; SCHUTTE, Klamer; KRAAIJ, Wessel; NGO, Chong-Wah. (2015.0). *2015 TREC Video Retrieval Evaluation, TRECVID 2015, Gaithersburg, November 16-18, Gaithersburg*. (Published)

Vireo-Tno @ TRECVID 2015: Multimedia event detection, by ZHANG, Hao; LU, Yi-Jie; DE BOER, Maaike; TER HAAR, Frank; QIU, Zhaofan; SCHUTTE, Klamer; KRAAIJ, Wessel; NGO, Chong-Wah. (2015.0). *2015 TREC Video Retrieval Evaluation, TRECVID 2015, Gaithersburg, November 16-18, Gaithersburg*. (Published)

Direct or indirect match? Selecting right concepts for zero-example case, by LU, Yi-Jie; DE BOER, Maaike; ZHANG, Hao; SCHUTTE, Klamer; KRAAIJ, Wessel; NGO, Chong-Wah. (2015.0). *2015 TREC Video Retrieval Evaluation, TRECVID 2015, Gaithersburg, November 16-18, Gaithersburg*. (Published)

Vireo-Tno @ TRECVID 2014: Multimedia event detection and recounting (MED and MER), by NGO, Chong-Wah; LU, Yi-Jie; ZHANG, Hao; YAO, Ting; TAN, Chun-Chet; PANG, Lei; DE BOER, Maaike; SCHAVEMAKER, John; SCHUTTE, Klamer; KRAAIJ, Wessel. (2020.0). *2014 TREC Video Retrieval Evaluation, TRECVID 2014, Orlando, November 10-12, Orlando*. (Published)

Vireo @ TRecViD 2014: Instance search and semantic indexing, by ZHANG, Wei; ZHANG, Hao; YAO, Ting; LU, Yijie; CHEN, Jingjing; NGO, Chong-Wah. (2020.0). *2014 TREC Video Retrieval Evaluation, TRECVID 2014, Orlando, November 10-12, Orlando*. (Published)

Search of small objects by topology matching, context modeling, and pattern mining, by ZHANG, Wei; NGO, Chong-Wah. (2013.0). *TRECVID 2013 TREC Video Retrieval Evaluation, Gaithersburg, November 20-22, Gaithersburg*. (Published)

Vireo/ECNU @ TRECVID 2013: A video dance of detection, recounting and search with motion relativity and concept learning from wild, by NGO, Chong-Wah; WANG, Feng; ZHANG, Wei; TAN, Chun-Chet; SUN, Zhanhu; ZHU, Shi-Ai; YAO, Ting. (2013.0). *TRECVID 2013 TREC Video Retrieval Evaluation, Gaithersburg, November 20-22, Gaithersburg*. (Published)

The vireo team at MediaEval 2013: Violent Scenes Detection by mid-level concepts learnt from youtube, by TAN, Chun Chet; NGO, Chong-Wah. (2013.0). *Multimedia Benchmark Workshop, MediaEval 2013, Barcelona, Spain, October 18-19, Barcelona*. (Published)

VIREO@TRECVID 2012: Searching with topology, recounting will small concepts, learning with free examples, by ZHANG, Wei; TAN, Chun-Chet; ZHU, Shi-Ai; YAO, Ting; PANG, Lei; NGO, Chong-Wah. (2012.0). *TREC Video Retrieval Evaluation, TRECVID 2012, Gaithersburg, November 26-28, Gaithersburg*. (Published)

The shanghai-hongkong team at mediaeval2012: Violent scene detection using trajectory-based features, by JIANG, Yu-Gang; DAI, Qi; TAN, Chun Chet; XUE, Xiangyang; NGO, Chong-Wah. (2012.0). *Multimedia Benchmark Workshop, MediaEval 2012, Pisa, Italy, October 4-5, Pisa, Italy*. (Published)

VIREO@TRECVID 2011: Instance search, semantic indexing, multimedia event detection and known-item search, by NGO, Chong-Wah; ZHU, Shi-Ai; ZHANG, Wei; TAN, Chun-Chet; YAO, Ting; PANG, Lei; TAN, Hung-Khoon. (2011.0). *TREC Video Retrieval Evaluation, TRECVID 2011, Gaithersburg, December 5-7, Gaithersburg*. (Published)

VIREO at TRECVID 2010: Semantic indexing, known-item search, and content-based copy detection, by

NGO, Chong-Wah; ZHU, Shi-Ai; TAN, Hung-Khoon; ZHAO, Wan-Lei. (2010.0). *TREC Video Retrieval Evaluation, TRECVID 2010, Gaithersburg, November 15-17*, Gaithersburg. (Published)

VIREO/DVMM at TRECVID 2009: High-level feature extraction, automatic video search, and content-based copy detection, by NGO, Chong-Wah; JIANG, Yu-Gang; WEI, Xiao-Yong; ZHAO, Wanlei; LIU, Yang; WANG, Jun; ZHU, Shuai; CHANG, Shih-Fu. (2009.0). *TREC Video Retrieval Evaluation, TRECVID 2009, Gaithersburg, November 16-17*, Gaithersburg. (Published)

Beyond semantic search: What you observe may not be what you think, by NGO, Chong-Wah; JIANG, Yu-Gang; WEI, Xiaoyong; ZHAO, Wanlei; WANG, Feng; WU, Xiao; TAN, Hung-Khoon. (2008.0). *TREC Video Retrieval Evaluation, TRECVID 2008, Gaithersburg, November 17-18*, Gaithersburg. (Published)

Columbia University/VIREO-cityU/IRIT TRECVID2008 high-level feature extraction and interactive video search, by CHANG, Shih-Fu; HE, Junfeng; JIANG, Yu-Gang; EL KHOURY, Elie; NGO, Chong-Wah; YANAGAWA, Akira; ZAVESKY, Eric. (2008.0). *TREC Video Retrieval Evaluation, TRECVID 2008, Columbia University, October 26*, Columbia University. (Published)

Concept detection: Convergence to local features and opportunities beyond, by CHANG, Shih-Fu; HE, Junfeng; JIANG, Yu-Gang; EL KHOURY, Elie; NGO, Chong-Wah; YANAGAWA, Akira; ZAVESKY, Eric. (2008.0). *TREC Video Retrieval Evaluation, TRECVID 2008, Columbia University, October 26*, Columbia University. (Published)

Experimenting VIREO-374: Bag-of-visual-words and visual-based ontology for semantic video indexing and search, by NGO, Chong-Wah; JIANG, Yu-Gang; WEI, Xiaoyong; WANG, Feng; ZHAO, Wanlei; TAN, Hung-Khoon; WU, Xiao. (2007.0). *TREC Video Retrieval Evaluation, TRECVID 2007, Gaithersburg, November 5-6*, Gaithersburg. (Published)

Modeling local interest points for semantic detection and video search at TRECVID 2006, by JIANG, Yu-Gang; WEI, Xiaoyong; NGO, Chong-Wah; TAN, Hung-Khoon; ZHAO, Wanlei; WU, Xiao. (2006.0). *TREC Video Retrieval Evaluation, TRECVID 2006, Gaithersburg, November 13-14*, Gaithersburg. (Published)

Motion-based approach for BBC rushes structuring and characterization, by NGO, Chong-Wah; PAN, Zailiang. (2005.0). *TREC Video Retrieval Evaluation, TRECVID 2005, Gaithersburg, November 14-15*, Gaithersburg. (Published)

## Research Grants

### Singapore Management University

AI-InterRAI: AI-Assisted InterRAI Assessment and Evaluation for Person-Centred Care Planning and Healthy Ageing, Singapore-New Zealand Joint Grant Call on Leveraging AI for Healthy Ageing, AI Singapore, National University of Singapore (NUS) , PI (Project Level): DAI Bing Tian , Co-PI (Project Level): TAN Kar Way, NGO Chong Wah, 2025, S\$1,329,998.8

Food Recognition: Causality-driven Cross-modal Cross-lingual Domain Adaptation, Academic Research Fund (AcRF) Tier 2, Ministry of Education (MOE) , PI (Project Level): NGO Chong Wah , Co-PI (Project Level): Rajesh Krishna BALAN, 2023, S\$749,177

Unleashing the Power of Pre-trained Models for VisualQA: A Skill-based Framework, Academic Research Fund (AcRF) Tier 2, Ministry of Education (MOE) , PI (Project Level): Jing JIANG , Co-PI (Project Level): NGO Chong Wah, 2023, S\$656,253

Interpretable Embedding for Interactive Video Search, SMU Internal Grant, Ministry of Education (MOE) Tier 1 , PI (Project Level): NGO Chong Wah, 2020, S\$98,950

### Other Institutions

Automated CWD Detection with Machine Learning for Vessel-based Line Transect Survey, Marine Ecology Enhancement Fund, Hong Kong Airport Authority Co-PI (Project Level): NGO Chong Wah, 2020

Sport Video Analysis and Retrieval, Innovation and Technology Support Programme, Innovation and Technology Fund, Hong Kong PI (Project Level): NGO Chong Wah, 2019

Query-aware Video Recounting, Attention, Captioning, Indexing, General Program, National Natural Science of Foundation of China (NSFC) Co-PI (Project Level): NGO Chong Wah, 2019

S3: A Multi-perspective Video Search Paradigm, France/Hong Kong Joint Research Scheme, Research Grants Council (RGC) of Hong Kong PI (Project Level): NGO Chong Wah, 2018

Chinese Food Scanner: Modeling Ingredient Composition from Multimedia Perspective, General Research Fund, Research Grants Council (RGC) of Hong Kong PI (Project Level): NGO Chong Wah, 2017

Multimedia Hyperlinking: From Similarity Ranking to Entity Linking, General Research Fund, Research Grants Council (RGC) of Hong Kong PI (Project Level): NGO Chong Wah, 2016

Zero-Shot Multimedia Event Detection and Evidence Localization, Microsoft Research Asia Collaborative Research Grant, Microsoft Research Asia PI (Project Level): NGO Chong Wah, 2015

A Joint Learning/MOOC Platform for Hong Kong's Tertiary Education Sector, Teaching and Learning Fund, University Grants Council (UGC) of Hong Kong Co-PI (Project Level): NGO Chong Wah, 2015

Beyond Relevancy: Multimedia Analytics for Exploratory Video Search, General Research Fund, Research Grants Council (RGC) of Hong Kong PI (Project Level): NGO Chong Wah, 2014

Big data pattern discovery based on multi-modal feature fusion, National High-tech R&D Program (863 program), Ministry of Science and Technology, China Co-PI (Project Level): NGO Chong Wah, 2014

Beyond Concept Annotation: Multimedia Event Detection and Recounting, General Research Fund, Research Grants Council (RGC) of Hong Kong PI (Project Level): NGO Chong Wah, 2013

Mobile Media Retrieval: From Instance Search to Question Answering, General Research Fund, Research Grants Council (RGC) of Hong Kong PI (Project Level): NGO Chong Wah, 2012

Large-Scale Exploratory Video Search by Social Networking, Microsoft Research Asia Windows Phone Academic Program, Microsoft Research Asia PI (Project Level): NGO Chong Wah, 2012

On the Fusion of Heterogeneous Modalities for Large-Scale Multimedia Search, General Program, National Natural Science of Foundation of China (NSFC) PI (Project Level): NGO Chong Wah, 2012

Scalable Partial Near-Duplicate Search and Its Applications to Topical Video Browsing, General Research Fund, Research Grants Council (RGC) of Hong Kong PI (Project Level): NGO Chong Wah, 2010

Web-Scale Multimedia Search: From Semantic Detection to Cross-Media Reasoning, General Research Fund, Research Grants Council (RGC) of Hong Kong PI (Project Level): NGO Chong Wah, 2009

Multimedia-based Hyperlink Construction and Modeling for Web Video Search, General Research Fund, Research Grants Council (RGC) of Hong Kong PI (Project Level): NGO Chong Wah, 2008

Exploiting Interaction and Co-occurrence of Multi-modal Features for Early Fusion in Retrieval, General Research Fund, Research Grants Council (RGC) of Hong Kong PI (Project Level): NGO Chong Wah, 2006

Common Pattern Discovery in Images and Videos, General Research Fund, Research Grants Council (RGC) of Hong Kong PI (Project Level): NGO Chong Wah, 2005

A Spatio-temporal Framework for Video Representation and Retrieval, General Research Fund, Research Grants Council (RGC) of Hong Kong PI (Project Level): NGO Chong Wah, 2002

## TEACHING

---

### Courses Taught

Singapore Management University

Undergraduate Programmes :

Data Management

## Operating System Concepts with Android Operating Systems

### Postgraduate Professional Programmes :

Capstone Project - Data Science and Engineering  
Introduction to Artificial Intelligence

### Postgraduate Research Programmes :

Empirical Research Project 1  
Empirical Research Project 2  
Empirical Research Project 3  
Empirical Research Project I  
Empirical Research Project II  
Empirical Research Project III

## THESES AND DISSERTATIONS

---

### Theses and Dissertations Supervised

#### Other Institutions

Supervisor, "Towards Learning Compatible, self-explainable, transferable features for cross-modal recipe retrieval", Thesis by Bin Zhu , PhD, City University of Hong Kong, 2021

Supervisor, "Interactive Video Retrieval", Thesis by Phuong-Anh Nguyen , PhD, City University of Hong Kong, 2020

Supervisor, "Cross-Modal Cooking Recipe Retrieval", Thesis by Jing Jing Chen , PhD, City University of Hong Kong , 2018

Supervisor, "Zero-Example Multimedia Search Using Large-scale Knowledge Bank", Thesis by Yijie Lu , PhD, City University of Hong Kong , 2018

Supervisor, "Deep Fine-grained Video Representation for Large-scale Multimedia Retrieval and Recounting", Thesis by Hao Zhang , PhD, City University of Hong Kong, 2017

Supervisor, "Multimedia Content Understanding by Analyzing Perceived Signals: Face, Sentiment and Emotion", Thesis by Lei Pang , PhD, City University of Hong Kong , 2015

Supervisor, "Searching and Mining Visual Instances in Large-Scale Video Collection", Thesis by Wei Zhang , PhD, City University of Hong Kong , 2015

Supervisor, "Multimedia Search by Self, External and Crowdsourcing Knowledge, ACM SIGMM Outstanding Thesis Award", Thesis by Ting Yao , PhD, City University of Hong Kong, 2014

Supervisor, "Video Content Recounting by Commonsense Knowledge", Thesis by Chun Chet Tan , PhD, City University of Hong Kong , 2014

Supervisor, "Sampling Web Images for Large-scale Concept Learning", Thesis by Shuai Zhu , PhD, City University of Hong Kong , 2013

Supervisor, "Topical Video Browsing by Cross-Media and Cross-Source Linking", Thesis by Song Tan , PhD, City University of Hong Kong , 2013

Supervisor, "Algorithm and Techniques for Large-scale Near-duplicate Video Mining", Thesis by Wanlei Zhao , PhD, City University of Hong Kong , 2010

Supervisor, "Concept-based Video Search by Semantic and Context Reasining", Thesis by Xiao-Yong Wei , PhD, City University of Hong Kong , 2010

Supervisor, "Video Hyperlinking for Multimedia Search", Thesis by Hung Khoon Tan , PhD, City University of Hong Kong , 2010

Supervisor, "Large-scale Semantic Concept Detection, Fusion and Selection for Domain Adaptive Video Search", Thesis by Yu-Gang Jiang , PhD, City University of Hong Kong , 2009

Supervisor, "On Clustering, Detection, and Threading of Topics for Large-scale Videos with Multiple Modalities", Thesis by Xiao Wu , PhD, City University of Hong Kong, 2008

Supervisor, "Common Pattern Discovery and Multiple Appearance Modeling for Visual Content Analysis", Thesis by Tan Hong Khoon , MPhil, City University of Hong Kong , 2007

Supervisor, "Content Parsing of Home Videos by Motion Analysis", Thesis by Zailiang Pan , MPhil, City University of Hong Kong , 2005

Supervisor, "Polyphonic Song Analysis and Representation for Query-by-singing Systems", Thesis by Tan-Wan Leung , MPhil, City University of Hong Kong , 2005

## Theses and Dissertations Assessed

### Other Institutions

Co Supervisor, "Video Content Analysis and Its Applications for Multimedia Authoring of Presentations", Thesis by Feng Wang , PhD, Hong Kong University of Science and Technology , 2006

Co Supervisor, "Modeling with Panoramic Image Network for Image-based Walkthroughs", Thesis by Sau Kuen Wan , MPhil, City University of Hong Kong , 2005

Co Supervisor, "Matching and Retrieval of 3D deformable models", Thesis by Kwok Leung Tam , MPhil, City University of Hong Kong , 2004

## EXTERNAL SERVICE – PROFESSIONAL

---

Core member, Hong Kong Computer Society, Artificial Intelligence Specialist Group , 2018 - Present

Steering Committee Member, ACM Multimedia Asia, 2017 - Present

Steering Committee Member, IEEE ICME 10K Award Committee , 2016 - 2018

Steering Committee Member, TRECVID, NIST USA , 2015 - Present

Editor Associate Editor, ITE Trans. on Media Technology and Applications , 2015 - Present

Editor Associate Editor, IET Computer Vision , 2015 - Present

Steering Committee Member, ACM Conf. on Multimedia Retrieval , 2015 - Present

Guest Editor, International Journal of Multimedia Tools and Applications , 2014

Guest Editor, Multimedia Systems Journal , 2013

Guest Editor, IEEE Multimedia Magazine , 2013

Editor Associate Editor, IEEE Trans. on Multimedia , 2011 - 2014

Editor Associate Editor, Journal of Multimedia Data Engineering and Management , 2010 - Present

Editor Associate Editor, Journal of Advances in Multimedia , 2010 - 2017

Chairperson, ACM Hong Kong Chapter , 2008 - 2009