© 2018 IJSRST | Volume 4 | Issue 5 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Science and Technology

# A State-of-the art Review: A survey on Multimedia Tagging Techniques

**Kirubai Dhanaraj, Rajkumar Kannan**
Research Department of Computer Science Bishop Heber College, Tiruchirappalli, India

## ABSTRACT

Social media prevails in every walk of our life. Multimedia shared through social networks has tremendously increased the need for efficient retrieval methods and expects more accuracy in terms of annotating an existing image or video. Retrieval methods and annotation techniques are two sides in the development of an efficient multimedia retrieval system. Annotating the image and video is a challenging task. Collaborative user annotations can be incorporated into multimedia to increase the efficiency and accuracy in the retrieval methods. Collaborative user annotations are useful for two reasons: (i) Multi-label annotation for a multimedia is possible with less time-consuming even for large-scale image corpus (ii) Correlation between images and videos build a multi-class label propagation without much human effort and in reduced cost. There are many areas of research the collaborative annotations are incorporated with small modification in the existing machine learning algorithms. This survey paper presents the state-of-the-art annotation techniques for multimedia in the new era.

**Keywords:** Image annotation, video annotation, automatic approach, social annotation, collaborative annotation, crowdsourcing

## I. INTRODUCTION

Multimedia annotation is the process of describing the image or video in a textual description. The descriptions may be in the form of tag, label, and concepts for an image or video. Manual approaches are used to annotate the images but they are labour-intensive and time consuming. Automatic annotations approaches exist from the last decade, they find the concept similarity between the images and tags. Many information retrieval techniques are used by finding the similarity in the image concepts like sky, person, and car to label the images. An image can have multiple concepts (single-class, multiple-class labels) and it can be labelled using the training sample, and a training sample is unique for each concept. Automatic annotations are not suitable for efficient retrieval technique because of the semantic gap between the image concept and the labels that automatically generated using the training samples. Automatic annotations need more accurate training set but they cannot be generated by machine learning because the visual perception is the basic for annotation. Humans are more accurate than machines in visual perceptions and identifying concepts in real-world images. This is the feasible solution to train the sample for an automatic annotation, thus the research for collaborative annotation emerges in past few years. Social media and online multimedia sharing websites has huge amount of rapidly increasing images and videos along with the user descriptions and interlinked data. By leveraging the social user generated annotations the large-scale multimedia can be annotated, analyzed and can be retrieve efficiently.

## II. OUTLINE OF THE SURVEY

This paper presents a survey of the approaches, techniques used for multimedia annotation in the past

decade. The multimedia annotation techniques are broadly categorized into automatic and collaborative annotation. This survey is made with the automatic and collaborative approaches with relevant machine learning techniques and elaborates the future direction of research in multimedia retrieval.

The automatic annotation approaches finds the concept similarity [1, 2, 3, 4] between the image and labels or tags. The concept similarity is obtained by finding the visual similarity by low-level and semantic features of the image or video using the training samples. They cluster the similar concept images and propagate to new images of visually similar images [1, 2, 3, 4]. Label propagation through graph construction [5, 6], neighbourhood propagation [4, 7, 8] for labelling the nearest neighbour, random walks [6, 7, 8] to find the neighbourhood through hierarchically [8, 9] dividing partitioning the graph. On the other hand computer vision techniques [10] like object recognition, face recognition [10, 11], standard feature-based [11, 12] multimedia annotation (not discussed in this survey) are also increasing but they depend on semantic annotated accurate training samples.

The labels or tags that are generated by the automatic annotation are not most relevant when comparing to the tags that user generates for the same image in the internet. Social tags are better than the automatic annotations for training the large-scale images samples [6, 7, 8]. Collaborative annotation approaches [1, 2, 4, 10] leverages these user tags to find the semantic relationship between the image content and the tags.

Collaborative annotation approaches uses implicit user generated social tags to construct web-scale image graphs [5, 17, 26] that represents semantically similar images, finding the tag relevance [13, 15, 35] using semantic tag similarity and to improve the tag quality approaches like tag recommendation [14, 16 ], tag refinement [4, 15, 29, 33], tag filtering [17, 25] are used. Social tags can be explicitly collected as crowdsourced annotations like online games [18, 19, 20], paid tools for annotation like MTurk [21, 22], ESP [23, 24], LabelME and reCAPTCHA. The figure 2.1 shows the collaborative annotation models for multimedia retrievals.

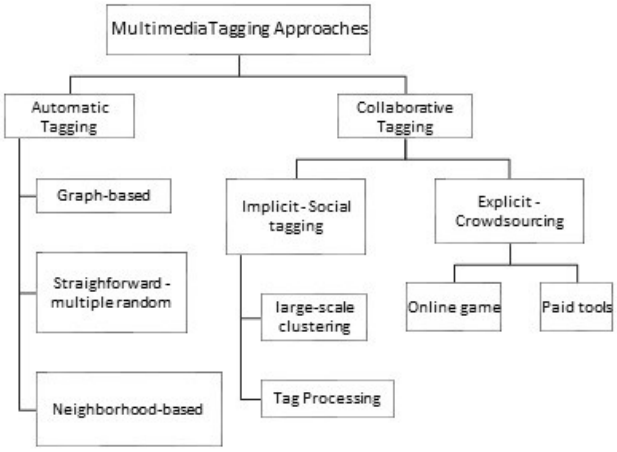


**Figure 1.** Multimedia tagging approaches

## III. AUTOMATIC ANNOTATION APPROACHES

Automatic annotation approach finds the visual and semantic similarity between the training samples and unlabeled images. Label propagation is achieved by constructing the probabilistic graphs [5, 6, 17, 26], label propagation through nearest-neighbours [1, 2, 4, 10, 11], Straight-forward random-walk hierarchical-based propagations [27, 28] and codebook generation [30, 31, 36,38].

Label propagation through graph by automatically constructing the graph for semantically similar images by forming tag as node, the value of the node are ranked, the edges represents the semantic similarity of the tag[2, 3, 4, 5]. The probabilistic transition was made between two similar nodes. Video annotation through search and graph reinforcement [2, 3, 4, 5, 34] generates a stable graph between visually similar keyframes from the video and tag similarity. It focuses on individual tagging collection on community tags in the social network. Jianping Fan et al [27, 37] presents a multi-level annotation of natural scenes using the salient features and relevant semantic concepts. Image

annotation by graph-based Inference with Integrated Multiple / single instance Representation [36, 39] is a unified framework. It combines the multiple-instances and single-instances representation of image. To obtain accurate region-level image annotation Jinhui Yuan et al[37, 38] presents a grid-structured graphical model that characterizes the spatial dependencies.

Graph-based label propagations consider each tag independently when handling multi-label propagation problem, the labels are also not in rank and construction of graph is time consuming. Large-scale label annotation algorithms meet the need of single-label case, and they are unclear when scaling to multiple labels. Stefan Siersdorfer et.al presents a neighbourhood-based tag propagation approach for automatically obtaining richer video annotation using content redundancy [25, 40]. This approach automatically analyses the dataset to find the near duplicates to extract additional information about the content. Tag ranks can also be provided depending on the additional label information while propagating the links between videos.

Multiple Random divide-and-conquer is a straightforward approach by hierarchically portioning the neighbourhood graph. Jianping Fan et al. [27] and Xiangyang Xue, et al. [28, 29] developed an automatic structured Max-margin learning algorithm to incorporate the inter-concept visual similarity of images and multiple base kernels for diverse visual similarity contexts between images. To obtain the inter-concept visual similarity relationships the high-dimensional multi-modal visual features for an image are extracted. They are partitioned into multiple feature subsets and each of it represents a specific image property. The mixture-of-kernels are used to obtain the diverse visual similarity between multiple feature subsets of an image. Structured Max-margin learning [28, 39] task predicts and estimate the inter-related classifiers more accurately. Multi-modal hierarchical image object annotation [39] is an automatic learner of image content without specifically labelling the individual objects. Image annotation Refinement using Random Walk with Restarts [13, 29] is a relevance model to decide the candidate annotations.

Concept categorization is an automatic approach for large-scale video indexing by comparing visual-based compact codebook [30, 31, 36, 38]. The vocabulary in the codebook model determines the quality of video annotation and indexing, the increased size of vocabulary leads to clarity but it increases the model complexity. This work solves the problem by incorporates discrete visual codeword for image features using unsupervised clustering approach. Automatic Annotation of Video sequences using Multimedia Ontology [34] automatically annotates the video clips with high level concepts by finding their similarity with the visual concepts of the ontology. Automatic annotation and semantic retrieval of video clips are performed by properly associating the similarity of the video clip to the high-level concept presented in the ontology to derive and perform complex queries to the Multimedia ontology.

## IV. IV COLLABORATIVE ANNOTATION APPROACHES

### *A. Implicit approach*

Social annotations for the online multimedia are better than automatic annotations. Automatic annotation approaches finds the similarity between label training images and unlabeled online images. They lack to find the similarity between the label or tag and the image content, so the tags generated automatically are not relevant and not in the top positions in ranking the tags [33, 41]. Collaborative annotation leverages the social annotations generated by the user to be takes as initial training sample to promote the tag for large clusters that are semantically similar with other image contents. Visually similar images are also ranked closely by graph-propagation [2, 4, 5], Re-ranking [33, 41] and relevance scores [4, 5,

9]. The collaborative tags are generated social by different user and they are noisy, ambiguous, incomplete and irrelevant. Tag processing approaches improves the quality of tags by tag recommendation [14, 15, 31], and tag refinement [2, 4, 5, 41] and etc.

Tat-Seng Chua et.al presents a large-scale multi-label propagation approach [1, 3, 4] using minimized Kullback-Leibler divergence for single image labels. Locality Sensitive Hashing is used for candidate selection of similar neighbours for an image and to speed up the scaling (large-scale) process by constructing the *$l_1$-graph.* The author [50] proves the efficiency and accuracy by testing the algorithm in NUS-WIDE dataset, among 269,648 images a part of 161,789 unlabeled images are indicated with 81D label vector for 81 distinct concepts using multi-label propagation. This problem can also be done directly as tag ranking task instead of probabilistic label propagation.

A graph-based semi-supervised learning approach [17, 18, 20] is presented to annotate a large-scale image corpus by label-propagation over nosily-tagged web images. By using the greatly available online user annotations these training samples can be annotated using machine learning techniques to eliminate few human errors like incorrect tags and incomplete tags. It is an efficient method to annotate large-scale images from with perfectly build training samples. Training label refinement strategy [1, 41] was developed previously to define Semantic model using sparse graph construction for noisily annotated tags. Jinhui Tang et.al [26, 39] work improves the efficiency of noisily annotated training samples by incorporating Locality Sensitive Hashing. The training images are segmented to semantic regions depending on their labels and they specify different semantic clusters. Image annotation using noisy tags [17, 20, 39] in addition to that semantic video indexing framework was presented in [42, 43] by incorporating the noisy user tags, for image-to-video indexing approach rather than text-to-video. A probabilistic approach is employed to estimate the relevance score by indicating the probabilities of correctly tagged images.

Labelling an image for bi-concepts is not possible, but the need for searching bi-concepts is a challenge in multimedia retrieval systems. Cees. G. M. Snoek [36, 38, 39] presents a multimedia search engine by harvesting social images to define bi-concepts using the co-occurrence of two distinct visual concepts. It collects de-noised positive and informative negative training examples from social web. It creates a codebook for bi-concept detector by estimating the relevance of bi-concept with respect to an image using k-means clustering. The multimedia search engine achieves bi-concept image search by artificially combining individual single-concept detectors.

Collaborative tag depends on social user interest and their use the vocabulary by their choice. These user generated tags may not properly describe the content of the multimedia and sometimes they are irrelevant, negatively annotate, and have noisy tags. To refine the social annotations and to enhance the quality tag processing during tagging such as tag recommendation [14, 15] or after tagging such as tag refinement [4, 33, 35] and re-ranking are the state-of-art approaches in social multimedia annotation and retrieval.

Graph reinforcement technique is an inductive learning process [5, 6, 7] that creates a strong prediction for weakly annotated set of each similar video. Graph is created through correlative near neighbours to extract better annotation or to create a new annotation of one of its similar documents. Elaheh Momeni et al. [23] generated an automated support to increase the quality of tag by tags to descriptive annotations. Descriptive annotation consists of supplementing features based on text and linguistics, semantic and topical, author and social features classifiers are used to classify the usefulness as positive class and not usefulness as negative class. It

improves the quality and efficiency of the user generated social tags.

The Wisdom of Social Multimedia [42, 43] is leveraged to predict and forecast by implicit vote for an image or video whenever it is viewed. Jose San Pedro et al [33, 44] leverages user comments for aesthetic aware image search re-ranking, while Taku Kuribayashi et al. [40] ranks the classical music using content descriptions. A relevance-based ranking scheme for social image search [9, 35] to automatically rank the image according to their relevance to the query tag. Cees G. M. Snoek [26, 36] develops a learning classifier for visual concepts categorization using relevant negatives.

### *B. Gaming and Crowdsourcing*

Collaborative tagging explicitly gathers user annotations through online gaming like LabelMe, ESP, reCAPTCHA and paid online user annotation tools like MechanicalTurk. In this section this survey discusses in detail of various crowdsourcing approaches.

Interactive image tagging framework is a hashing-based image tagging to enable quick clustering of image regions and dynamic multi-scale clustering label for a large group of similar region tagging [10, 45]. Set of visually close images does not satisfy with specific labels, this work segments the images into multiple regions and lets the user to annotate it and a dynamic multi-scale clustering using locality sensitive hashing is applied to cluster these manual labels.

Robert Di Salvo et al. Presents a collaborative web-based platform to enhance the label from video ground truth annotation [45, 46]. It presents a platform with annotating windows where the videos are explored with their ground-truth annotation, that the user selects the best of existing annotation or generate new annotation for the video. The existing annotation can be best by marking the object as it is already annotated using object tracking algorithm in computer vision. It enhances the clarity of labelling and increases the accuracy when compared with existing ground truth annotation, but fails to link the videos of similar annotations.

A collaborative Design Assistant developed to ensure the cross platform user interface that dynamically updates the web applications depending on different user interfaces [24, 42, 43]. It is an expert system to provide dynamic changes made by the user web applications even in different browser platforms. Online users relay on the user generated tags and reviews for web items sold through online sites, this creates platform for designers to attract desirable tags when published [47]. Vangelis Hristidis et al [48] presents an optimization task that designs a new item expected based on the maximum number of desirable tags.

Carl Vondrick et.al [18, 49] presents an experiment with people annotating the real-world videos with some computer assistance. The user studies show that by extracting pixel-based features from manually labelled key frames are able to leverage more sophisticated interpolation strategies to maximize performance. Video processing algorithm [21] is capable of predicting boredom videos of internet are used to improve multimedia retrieval and recommendation.

A novel crowdsourcing workflow presented by Joho Kim et.al [46] extracts step-by-step annotation for How-to Videos. It annotates with procedural steps with timing, textual descriptors, before and after thumbnail images. It is similar to [18, 19] and complements in computer vision algorithms for clustering in timing, and uses Manhattan distance metric to measure the similarity between two images. The results are compared with ground-truth annotations and show better optimal solutions.

The previous crowdsourced video annotation can be complemented by the C.G.M. Snoek et.al multimedia search engine for semantic access to archival of rock n' roll concert videos [18, 36, 39]. It explores a novel crowdsourcing mechanism for multimedia retrieval of rock n' roll concert videos by user feedback to improve and extend automated content analysis results and shares video fragments using timeline based video player. Unlike the contemporary video annotation methods, they collect user feedback by a graphical overlay which specifies the pre-defined labels for the video fragments, that are asked to correct by user and on demand the user can create new labels. The video is fragmented not like shot or frames and they are specified with time-lines to mark the interesting fragments according to the automatically pre-defined labels. This increases the interestingness for users to give feedback for existing label and to create new labels for the videos.

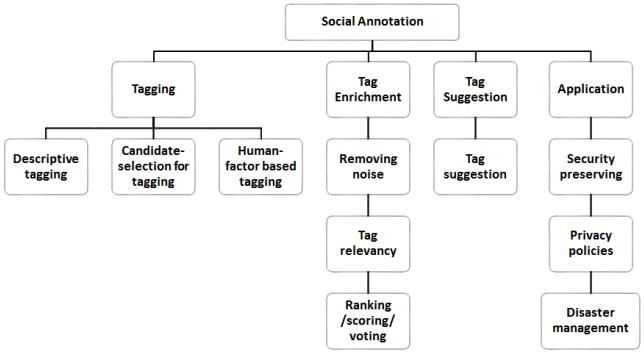


**Figure 2.** Social Annotation

Another crowdsourcing on-line game was developed by DiSalvo et.al [19, 20, 46] for annotating objects in the videos. Big noisy annotations are collected through an on-line flash game, where the user takes photos of object appearing throughout the game levels. After collecting the big noisy annotations, the machine learning algorithms are applied on the results to cluster most clicked area using k-means clustering, to identify the objects in image using region growing where initial seeds are giving and to perform image segmentation by means of a probabilistic approach. The drawback in this on-line gaming is to pre-define initial seed positions or that may lead to inaccurate region selection and it cannot identify the objects having similar texture and color for the background and object.

Mackay's EVA is an earliest system annotated the video using the mouse movement but now social media has introduced many user interactions in the form of tags, tweets, microposts. The new culture among the attendees of the academics conference in the last few years has generated a huge collection of microposts that exploits the descriptive nature of the videos in the conferences. It serves as the metadata for video analysis and annotation, and also can be used as browsing aids. Polemic tweets [42, 43] annotate the video sequences by crowdsourcing the videos and synchronizing the tweets with the videos. Timestamp in the video and the tweet annotation is an issue to be solved when these tweets are incorporated in the videos.

Table 1: Lists the Tagging techniques and purposes

| Task | Tagging Approaches | Purpose |
|---|---|---|
| **Tagging** | | |
| Descriptive Tagging | Logistic Regression [23,24], Naive Bayes Classifiers, INFORMATIVENESS, SUBJECTIVITY TONE[23, 32, 35] | Usefulness classifier |
| | | Predicts usefulness using only one feature |
| | | Predicts usefulness using only particular semantic class. |
| Candidate | PCA[11], Region-level [22], MOM-LDA [39], 2D | Face Recognition |

| | | |
|---|---|---|
| Selection | Hidden Markov Model [19, 20, 37], Discriminative random fields [37], Locality Sensitivity Hashing [39] | Accurate region-level annotation |
| | | Multimodal component |
| | | Segmentation |
| Human Factor | Two-tier top-k algorithm, PTAS [47], User feedback, video fragments [35, 36, 44, 49], clustering, segmentation Manhattan distance [44,45,46] | Tag maximization |
| | | To correct predefined labels |
| | | Depends on user interestingness labels (not as shot/ keyframes) |
| | | Timeline to obtain similar object |
| **Tag Processing** | | |
| Tag Relevance | Probabilistic approach [42, 43], Classical Kernel density estimation (KDE), Gaussian Kernel - Visual similarity based [1, 2 ,3 ,4, 5], k-means clustering [37, 38], 2D Hidden Markov Model [26, 37], K-means, Region growing, Grab-cut [19,20] | To estimate relevance for image concept |
| | | To combine individual concepts |
| | | To measure semantic divergence between two tags based on their co-occurrence frequency |
| | | For accurate region level annotation |
| | | For inter-concept visual similarity relationship between images |
| Tag Refinement | SURF Feature, Nearest-Neighbors [4, 5, 6] Sparse graph construction [32], Weighted edges [32], Directed and Weighted graph [27], KL-D, KNN, LSH [9, 10, 11], User feedback [38, 48, 49] | To find visually similar images |
| | | To refine training labels |
| | | Overlays between neighbor-based |
| | | Image-wise multilabel |
| | | To correct / refine pre-defined labels |
| Tag Suggestion | Label propagation through Regularization framework, Zhou's Regularization framework, Iterative EM algorithm [5, 22] , Uniform Histogram Binning [30, 35, 36], Radius-based clustering[35, 36] | Single-graph, Multi-graph reinforcement |
| | | For discretizing a continuous feature space |
| | | For optimizing the convergence solution |
| | | Assigns features with fixed radius of similarity for one cluster |
| Ranking / Scoring / Voting | TagRank – Overlap graph [25, 30], Normalized Google distance [24], Visual Concept Networks *HDMVFS, Mixture-of-kernels, Markov Networks [27, 28] , Manhattan Distance [46], Rank+ algorithm [4, 11] | Content based tag propagation in video graph |
| | | To measure semantic divergence between two tags based on their co-occurrence frequency |
| | | For inter-concept visual similarity relationship between images |
| | | To find the difference between two images |
| Noise Removal | Semantic Modelability[24,39], Label propagation with K-NN Sparse graph[17, 39 ] Affinity graph -Undirected with weighted edges, Overlap graph – Directed and weighted graph [25, 30 ], LSH [39] | Concept space construction |
| | | Removes semantically unrelated links |
| | | Overlaps graph - between neighbor-based tagging videos |
| | | To specifically assign labels to semantic regions of an image |

## V. CONCLUSION

Collaborative tagging bridges the semantic gap by leveraging the social annotations to semantically label and propagate to the visually similar image or video content. It can be regarded as a combination of manual labelling, model-based annotations and data-driven tag processing approach. The future direction in the area of multimedia tagging are estimating and evaluating label quality, to find the label inference, to trace anti-spam or cheating in online labels. Research in this area also needs strong data mining techniques. The collaborative annotations can also be leveraged to annotate large-scale multimedia, to annotate real-world videos, to annotate for multi-class of objects and to annotate the cultural heritages.

## VI. REFERENCES


[1]. L. S. Kennedy, S.-F. Chang, I. V. Kozintsev, To search or to label? Predicting the performance of search-based automatic image classifiers, in:Proc. of ACM MIR, Santa Barbara, CA, USA, 2006, pp. 249–258.

[2]. B. Sigurbj¨ornsson, R. van Zwol, Flickr tag recommendation based on collective knowledge, in: Proc. Of WWW, Beijing, China, 2008, pp. 327–336.

[3]. X. Li, T. Uricchio, L. Ballan, M. Bertini, C. G. M. Snoek, A. Del Bimbo, Socializing the semantic gap: A comparative survey on image tag assignment, refinement and retrieval, arXiv preprint arXiv:1503.08248 (2015).

[4]. Kirubai Dhanaraj, Rajkumar Kannan, Harnessing the Social Annotations for Tag Refinement in Cultural Multimedia, IJSRCEIT, 2018, pp. 1802-1808.

[5]. Emily Moxley, TaoMei, B. S. Manjunath, Video Annotation Through Search and Graph Reinforcement Mining, Published in IEEE Transaction on Multimedia Vol.12, No.3 April 2010 pp 184 – 193.

[6]. L. Ballan, M. Bertini, T. Uricchio, A. Del Bimbo, Data-driven approaches for social image and video tagging, Multimedia Tools and Applications 74 (2015) 1443–1468.

[7]. Y. Yang, Y. Yang, Z. Huang, H. T. Shen, Tag localization with spatial correlations and joint group sparsity, in: Proc. of CVPR, Providence, RI, USA, 2011, pp. 881–888.

[8]. X. Cao, X.Wei, Y. Han, Y. Yang, N. Sebe, A. Hauptmann, Unified dictionary learning and region tagging with hierarchical sparse representation, Computer Vision and Image Understanding 117 (2013) 934–946.

[9]. Xiangyu Chen, Yadong Mu, Shuicheng Yan, and Tat-Seng Chua, Efficient large-scale Image Annotation by Probabilistic Collaborative Multi-Label Propagation, ACM MM 2010

[10]. R. Kannan, G. Ghinea, S. Swaninathan, Salient region detection using patch level and region level image abstractions, 2015, IEEE, Signal Processing Letters 22(6), pp 686-690.

[11]. G. Ghinea, R. Kannan, S. Kannaiyan, Gradient-Oriented based PCA subspace for novel face recognition, IEEE Access 2014, pp 914-920.

[12]. H. Li, L. Yi, Y. Guan, H. Zhang, DUT-WEBV: A benchmark dataset for performance evaluation of tag localization for web video, in: Proc. Of MMM, Huangshan, China, 2013, pp. 305–315.

[13]. J. Song, Y. Yang, Z. Huang, H. T. Shen, J. Luo, Effective multiple feature hashing for large-scale near-duplicate video retrieval, IEEE Transactions on Multimedia 15 (2013) 1997–2008.

[14]. Sophia Swamiraj, Rajkumar Kannan, Twitter based stock recommendations using SVM and Ant Colony Optimization Methods. Advances in Natural and Applied Sciences, 2017, 11(9): pp 306-313.

[15]. H. Li, L. Yi, B. Liu, Y. Wang, Localizing relevant frames in web videos using topic model and relevance filtering, Machine Vision and Applications 25 (2014) 1661–1670.

[16]. Krassimira Ivanova, Peter Stanchev, Evgeniya Velikova, Keon Vanhoof, Benoi Depaire, Rajkumar Kannan, Iliya Mitov, Features for Art Painting

Classification Based on Vector Quantization of MPEG-7 Descriptors, published in Springer-Verlag, 2011, ICDEM 2010, LNCS 6411, pp 146-153.

[17]. Jinhui Tang, Richang Hong, Shuicheng Yan, Tat-Seng Chua, Guo-Jun Qi, Ramesh Jain, Image Annotation by kNN-Sparse Graph-based label propagation over Noisily-tagged web images, published in ACM Transaction on Intelligent Systems and Technology, Vol. 1, No. 1, September 2010, pp-111-126

[18]. Carl Vondrick, Donald Patterson, Deva Ramana, Efficiently Scaling up CrowdSourced Video Annotation, Springer, International Journal on Computer Vision, September 2012

[19]. R. Di Salvo, D. Giordano, I. Kavasidis, A Crowdsourcing Approach to support Video Annotation, ACM VIGTA Conference, July 2013.

[20]. Kyra Schlining, SusanVon Thun, Linda Kuhnz, Brian Schlining, Lonny Lundsten, Nancy Jacob senstout, Lori Chaney, Judith Connor, Debris in the deep: Using a 22-year video annotation database to survey marine litter in monterey, published in Elsevier, Journal on SciVerse ScienceDirect on January 2013.

[21]. Mohammad Soleymani, Martha Larson, Crowdsourcing for affective annotation of video: Development of a viewer-reported boredom corpus, published in the proceeding of SIGIR 2010 workshop on Crowdsourcing for Search Evaluation, July 2010.

[22]. Jinhui Tang, Qiang Chen, Meng Wang, Shuicheng Yan, Tat-Seng Chua, Ramesh Jain, Towards Optimizing Human Labeling for Interactive Image Tagging, Published in ACM Transaction on Multimedia Computing Communication and Applications Vol.9, No.4, August 2013.

[23]. Elaheh Momeni, Clarie Cardie, Myle Ott, Properties, Predictions, and Prevalence of Useful User-generated Comments for Descriptive Annotation of Social Media Objects, published in Association for the Advancement of Artificial Intelligence, 2013.

[24]. Vivian Genaro Motti, Dave Raggett, Quill: A Collaborative Design Assistant for Cross Platform Web Application User Interfaces, Published in ACM WWW 2013 Companion, May 2013, pp 3-5.

[25]. Stefan Siersdorfer, Jose San Pedro, Mark Sanderson, Automatic Video Tagging Using Content Redundancy, Published in ACM SIGIR July 2013.

[26]. Jinhui Tang, Haojie Li, Guo-Jun, Tat-Seng Chua, Image Annotation by Graph-based Inference with Integrated Multiple / Single Instance Representations, published in IEEE Transaction on Multimedia Vol.12, No.2, February 2011, pp 131-141

[27]. Jianping Fan, Yi Shen, Chunlei Yang, Ning Zhou, Structured Max-margin Learning for Inter-related Classifier Training and Multilabel Image Annotation, published in IEEE Transaction on Image Processing Vol.20, No.3, March 2011 pp 837- 854.

[28]. Xiangyang Xue, Hangzai Luo, Jianping Fan, Structured Max-margin Learning for Multi-label Image Annotation, published in ACM CIVR July 2010, pp 82-88

[29]. Changhu Wang, Feng Jing, Lei Zhang, Hong-Jiang Zhang, Image Annotation Refinement using Random Walk with Restarts, performed at Microsoft Research Asia, published in ACM MM October 2006

[30]. Jan C. Van Gemert, Cees G. M. Snoek, Cor J. Veenam, Arnold W. M. Smeulders, Jan-Mark Geusebroek, Comparing Compact codebooks for Visual Categorization, Published in Elsevier Journal on Computer Vision and Image Understanding 114, 2010, pp 450 – 462.

[31]. Lei Wu, Linjin Yang, Nenghai Yu, Xian-Sheng Hua, Learning to Tag, Published in ACM WWW April 2009, pp 361 -370

[32]. Jinhui Tang, Richang Hong, Shuicheng Yan, Tat-Seng Chua, Guo-Jun Qi, Published in ACM Conference on Multimedia, October 2009 pp- 223-232

[33]. Jose San Pedro, Tom Yeh, Nuria Oliver, Leveraging user comments for aesthetic aware

image search reranking, published in ACM WWW April 2012, pp 439 – 448.

[34]. Marco Bertini, Alberto Del Bimbo, Carlo Tornial, Automatic Annotation and Semantic Retrieval of Video Sequences using Multimedia Ontologies, published in ACM MM, October 2006, pp 679- 682

[35]. Dong Liu, Xian-Sheng Hua, Meng-Wang, Hong Jiang Zhang, Boost search relevance for tag-based social image retrieval, published in IEEE on published in ICME 2009, pp-1636-1639.

[36]. Xirong Li, Cees G. M. Snoek, Marcel Worring, Dennis Koelma, Arnold W. M. Smeulders, Bootstrapping Visual Categorization with Relevant Negatives, published in IEEE Transaction on Multimedia, Vol 15, No. 4, June 2013, pp 933 -945.

[37]. Jianping Fan, Yuli Gao, Hangzai Luo, Multi-level annotation of Natural Scenes using Dominant Image Components and Semantic Concepts, published in ACM MM 04, October 2004, pp 540-547

[38]. Xirong Li, Cees G. M. Snoek, Marcel Worring, Arnold W. M. Smeulders, Harvesting Social Images for Bi-concept Search, published in IEEE Transaction on Multimedia, Vol.14, August 2012, pp 1091- 1104.

[39]. Jinhui Tang, Shuicheng Yan, Chunxia Zhao, Tat-Seng Chua, Ramesh Jain, Label-specific training set construction from web resource for image annotation, published in Elsevier Journal on Signal Processing , 2012.

[40]. Tahu Kuribayashi, Yasuhito Asano, Masatoshi Yoshikawa, Ranking Method specialized for content description of classical music, published in ACM WWW 2013 companion, May 2013.

[41]. Guangyu Zhu, Shuicheng Yan, Yi Ma, Image Tag Refinement Towards Low-Rank, Content-Tag Prior and Error Sparsity, MM'10 October 25-29, 2010, pp: 461-470

[42]. Yang Yang, Zheng-Jun Zha, Heng Tao Shen, Tat-Seng Chua, Robust Semantic Video Indexing by Harvesting Web Images, Published in Springer-Verlag MMM 2013, Part-I, LNCS 7732, pp 70-80, 2013.

[43]. Samuel Huron, Petra Isenberg, Jean Daniel Fekete, PolemicTweet: Video Annotation and Analysis through Tagged Tweets, published in Proceedings of the IFIP TC13 conference on Human-computer Interaction (INTERACT), version 1, April 2013.

[44]. Xin Jin, Andrew Gallagher, Liangliang Cao, Jiebo Luo, Jiawei Han, The Wisdom of Social Multimedia: Using Flickr for Prediction and Forecast, published in ACM MM October 2010.

[45]. Isaak Kavasidis, Simone Palazzo, Robert Di Salvo, Daniela Giordana, Concetto Spampinato, An Innovative web-based collaborative platform for Video annotation, Published in Springer on Multimedia Tools and Applications, March 2013.

[46]. Joho Kim, Phu Nguyen, Sarah Weir, Philip J. Guo, Robert C. Miller, Krzysztof Z. Gajos, Crowdsourcing step-by-step Information Extraction to Enhance existing How-to Videos, submitted to ACM CHI 2014.

[47]. Mahashweta Das, Gautam Das, Vegelis Hristidis, Leveraging Collaborative Tagging for Web Item Design, Published in ACM KKD, August 2011.

[48]. A. Makadia, V. Pavlovic, S. Kumar, A new baseline for image annotation, in: Proc. of ECCV, Marseille, France, 2008, pp. 316-329.

[49]. Cees G. M. Snoek, Bauke Freiburg, Johan Oomen, Roeland Ordelman, Crowd sourcing Rock n' Roll Multimedia Retrieval, ACM Conference on Multimedia, october 2010.

[50]. Okasana Yakhnenko, Vasant Honavar, Annotating images and image objects using a hierarchical Dirichlet process model, published in ACM MDM / KDD August 2008.