International Journal of Engineering & Technology, 7 (2.7) (2018) 257-261

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research Paper

Joint graph regularization based semantic analysis for cross-media retrieval: a systematic review

Monelli Ayyavaraiah [1, 2] *, Dr. Bondu Venkateswarlu [3]

[1] *Research Scholar, Department of Computer Science and Engineering, School of Engineering*
Dayananda Sagar University, Bangalore, India
[2] *Assistant Professor, MGIT, Hyderabad, India*
[3] *Assistant Professor, Department of Computer Science and Engineering, School of Engineering*
Dayananda Sagar University, Bangalore, India
**Corresponding author E-mail: ayyavaraiah50@gmail.com*

Abstract

The large number of heterogeneous data are rapidly increasing in the internet and most data consist of audio, video, text and images. The searching of the required data from the large database is difficult and time taking process. The single media retrieval is used to get the needed data from the large dataset and it has the drawback, it can only retrieve the single media only. If the query is given as the text and acquired result are present in text. The users demand the cross-media retrieval for their queries and it is very consistent in providing the result. This helps the users to get more information regarding to their queries. Finding the similarities between the heterogeneous data is very complex. Many research is done on the cross-media retrieval with different methods and provide the different result. The aim is to analysis the different cross-media retrieval with the joint graph regularization (JGR) to understand the various technique. The most of researches are using the parameter of MAP, precision and recall for their research.

Keywords: *Cross-Media Retrieval; Joint Graph Regularization; MAP; Heterogeneous Data; Single Media Retrieval.*

1. Introduction

The major part of the big data consists of heterogeneous multimedia content such as text, image, audio, video, these are all increasing rapidly [1]. Many has been conducted in the content based media retrieval and the prevailing methods are single-media retrieval and multi-modal retrieval [2]. For former one, the user query and retrieved result are present in the same media and later, multi-modal retrieval is used like user queries are present in form of one media and retrieved result are presented in the different format [3]. This method is used for the better result of the system, such as image and text as a results. The existing method focus on the single media retrieval and ignore the relationships of the different modalities, which is important for better understanding of multi-media contents [4]. These method does not support that the query is in the image and the result is text, image, audio and video. The cross-media retrieval is required by the user for the various types of result for their submitted queries of single media as search [5]. Although user can get the various media as the result of their queries, which is more comprehensive than the traditional retrieval methods.

For example, user visited the Golden Gate Bridge, by using the photo of the place to search and cross-media retrieval provide user to get the result in the form of text, video and audio description about the place [6]. The user can give the query in many format like image or text, which is convenient. For instance, sound of unfamiliar bird given as query and result is get in the textual description, without recording of similar sound [7]. Content-based cross media retrieval is an interesting and yet this is a hard to provide the required results [8]. The finding the similarity between the homogeneous data is the difficult task and finding similarities between the heterogeneous is much more challenging [9]. In the cross-media retrieval heterogeneous metric is learned with the help of joint graph regularized technique, which is used to compare the different Ease of Use media to exploit the better structure information [10]. In this paper, the review has been done on a JGR researches for the better understanding of the method and analyses the advantages and disadvantageous for each individual algorithm. In Xie, et al. [39] research, used the self-taught learning in cross-modal, which provides the improved retrieval result. The heterogeneous metric is measured across various media with Joint graph regularized is used in Zhai, et al. [32] research, which helps to learn the high level sematic metric.

2. Literature review

Yu-xin, et al. [11], proposed the model for the cross-media uniform representation and also done the cross-media correlation with understanding and deep mining. This provides the cross media description and generation of the data and the retrieval with intelligent engines for intelligent application. The construction of graph with the cross media knowledge, learning methodologies was made and also done the reasoning and evolution. The iTrend is not efficient in collecting the cross-media data and under-utilization of cross media data.

Sun, et al. [12], used the active learning SVM algorithm with regularization path, which fit each solution path of SVM for every value of model parameters. From this method, candidate model parameters are traced from every solution path of this algorithm and the unlabeled samples are used to select the best model parameter. The experiment is tested on the real time dataset, which shows the effectiveness of the proposed system. There are some samples are not

considering in this process and that is also needed to take in classification for optimization of the algorithm.

Wang, et al. [13], established the method that combines the contest and semantic similarities, which combines the low-level and high-level similarities through the Markov-chain and also uses the heterogeneous similarity measure between the uni-media. The list of query is ranked by finding the better path across the chain and contest similarity is measured from the internal structure of the modality and the semantic similarity is calculated from the sematic correlation between the different modalities. The result shows that the proposed multi-mode method provides the better result while compared to the previous method on image retrieval function. This method is works only on the media that measured in content and sematic similarities and it not suitable for media like video, audio etc.

Arulmozhi and Abirami [14], consider the importance of the ANN and their general classification; the different categories for learning the hash were analyzed along with their advantageous and disadvantageous. The different bit assignment types were also learned and investigates the various methods to reduces quantization errors along with its pros and cons. The performance of Hash function has to be improved for the better classification.

Qi, et al. [15], provides the method of Cross-media Similarity Metric (UNCSM) that is Unified Network, which associate the cross-media shared representation learning with distance metric. The deep network is built with two-pathway, pre-trained is done with contrastive loss and fine-turning is done by applied the double triplet similarity loss to learn the shared representation for each media type. The result shows that the proposed system gives the high performance when compared with the 8 state-of-art method with the 4 datasets. The classification has to be made for the different types of media like audio, video etc.

3. Proposed methodology

The cross-media retrieval identifies the similarities among the different media to give the different result to the user. The many techniques are present in the cross-media retrieval and one of the most important technique is JGR. The overview of the JGR is shown in figure 1.

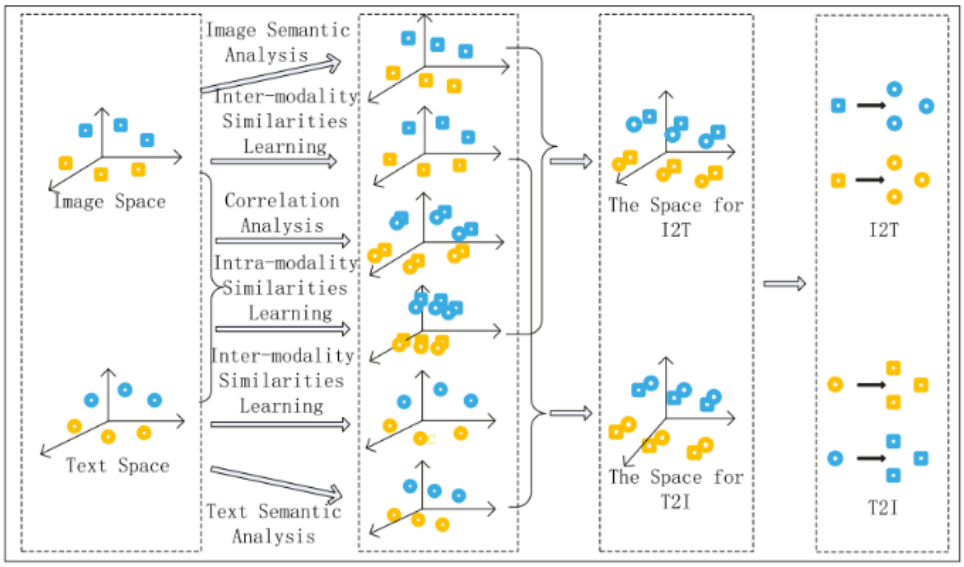

Fig. 1: The Overview of the JGR Technique.

3.1. Heterogeneous metric learning

The two sets of heterogeneous pair wise constraints among the heterogeneous media objects are obtained, shown in equation (1):

$$S = \left\{(x_i, y_j)\middle| l_i^x = l_j^y\right\}$$

$$D = \{(x_i, y_j)| l_i^x \neq l_j^y\} \quad (1)$$

Where S denoted as the set of similarity constrain and D is denoted as set of dissimilarity constrains. Each pairwise constraint (x_i, y_j) indicates the two heterogeneous media objects x_i and y_j are relevant from the investigation with help of category label. The S and D on the dataset $\mathbb{D}$ with a single matrix $Z = \{z_{ij}\}_{m\times n}$:

$$z_{ij} = \begin{cases} 1, & (x_i, y_j) \in S; \\ -1, & (x_i, y_j) \in D. \end{cases} \quad (2)$$

For any two given heterogeneous media objects x_i and y_j, let $d(x_i, y_j)$ denotes the heterogeneous distance between the objects. The similarities of the heterogeneous objects are not present in the metrics; hence it doesn't fall in the standard frameworks of metric learning [16], [17], [18], [19]. To learn the similarities between the heterogeneous media, the method is needed to find the heterogeneous similarities [20], [21]. The learning is carried out for the multiple linear transformation U and V of the input, which maps the same output, instead of the single learn transformation used in the previous methods. Let $U \in R^{d^x * x}, V \in R^{d^y * c}$ be the distance parameter matrices for $X \in R^{d^x * m}$ and $Y \in R^{d^y * n}$ respectively, here d^x,d^y are the dimensions of the original media types, c is the dimension of the mapped space, m and n are the count of media objects of the media X and media Y respectively [22], [23]. The heterogeneous distance measure technique is given as follows:

$$d(x_i, y_j) = \sqrt{(U^T x_i - V^T y_j)^T (U^T x_i - V^T y_j)} \quad (3)$$

The two parameter matrices $U_{d^x * c}, V_{d^y * c}$ are learn from the training heterogeneous multimedia dataset $\{X_{d^x * m}, Y_{d^y * n}\}$. Frequently used notations and their respective description is given in the table 1.

Table 1: Notations and Description of Symbols

Notation	Description
m, n	Number of training examples of media object x,y
p, q	Number of testing examples of media object x,y
d^x	Features dimension of media object x
d^y	Feature dimension of media object y
λ, ω	Regularization parameters
X	$d^x \times m$ data matrix of media object x
Y	$d^y \times n$ data matrix of media object y
Z	$m \times n$matrix with heterogeneous constrains
U	$d^y \times c$ transformation matrix for media x
V	$d^y \times c$ transformation matrix for media y
O	$c \times (m+n)$ data matrix for all media objects.

3.2. Objective function

The general regularization framework for the heterogeneous distance metrics learning is given in the equation (4) as follows:

$$\underset{U,V}{argmin}\ f(U,V) + \omega g(U,V) + \lambda r(U,V) \tag{4}$$

Where $f(U,V)$ is the loss function defined on the set of similarities and dissimilarity constrain $S\ and\ \mathcal{D}$, $g(U,V)$ and $r(U,V)$ are regularizer defined on the target parameter matrices $U,V.\ \lambda > 0, \omega > 0$ are the balancing parameters [24].

3.3. Loss function

The loss function $f(U,V)$ is defined as the minimizing the loss function results in minimizing the distance between the media objects with similarity constrains. The sum of squared distance expression for defining the loss functions in terms of its effectiveness and efficiency is shown below:

$$f(U,V) = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n} z_{ij}\left|\left|U^T x_i - V^T y_j\right|\right|^2 \tag{5}$$

Where z_{ij} is given in the equation (2) and to balance the influence of the similarity and dissimilarity, normalization is done on the element of Z in the column wise to check that the addition of each column value is zero.

3.4. Scale regularization

We define the regularization item $r(U,V)$ is given in equation (6):

$$r(U,V) = \frac{1}{2}||U||_F^2 + \frac{1}{2}||V||_F^2 \tag{6}$$

Where $||U||_F^2$ and $||V||_F^2$ are used to control the scale parameter matrices and reduce over fitting.

3.5. Joint graph regularization

The JGR item $g(U,V)$ implemented to identifying the similarities from the heterogeneous data. The similarity constraints in both modalities are helpful for metric learning and attempt to make the learned transformation consistent with the similarity constraints in both modalities [25], [26].

For heterogeneous data with multiple representations, a joint undirected graph, $G = (V,W)$ on the dataset. Each element w_{ij} of the similarity matrix $W = \{w_{ij}\}_{(m+n)*(m+n)}$ means the similarity between the $i-th$ media and $j-th$ media object [27], [28], [29], [30]. Note that all of the heterogeneous media objects $o_i \in \mathbb{D}, i = 1, \dots\dots, m+n$ are incorporated into joint graph. Here, the adoption of the label information to construct the similarity matrix:

$$w_{ij} = \begin{cases} 1, & l_i == l_j\ \Lambda\ i \neq j; \\ 0, & otherwise. \end{cases} \tag{7}$$

Where $l_i = l_i^x$ for $0 < i \leq m$ and $l_i = l_{(i-m)}^y$ for $m < i \leq m+n$ and set $w_{ii} = 0$ for $1 \leq i \leq m+n$ to avoid self-reinforcement. The normalized graph Laplacian L is defined as:

$$\bar{L} = I - D^{-1/2} W D^{-1/2} \tag{8}$$

Where I is an $(m+n) \times (m+n)$ identity matrix and D is an $(m+n) \times (m+n)$ diagonal matrix with $d_{ii} = \Sigma_j w_{ij}$. It should be noted $\bar{L}$ is symmetric and positive semidefinite, with eigenvalue in the interval. It is defined as

$$O = (U^T X \quad V^T Y)$$

$$\bar{L} = \begin{pmatrix} \overline{L^x} & \overline{L^{xy}} \\ \overline{L^{yx}} & \overline{L^y} \end{pmatrix} \tag{9}$$

Where O represents for all of the media objects in the learned metric space, $\bar{L}$ denotes the normalized graph Laplacian. The formulation of regularization is based on this is given in equation (10)

$$g(U,V) = \frac{1}{4}\sum_{i,j=1}^{m+n} || \frac{o_i}{\sqrt{d_{ii}}} - \frac{o_j}{\sqrt{d_{jj}}} ||^2 w_{ij}$$

$$= \frac{1}{2} tr(O\bar{L}O^T)$$

$$= \frac{1}{2} tr(U^T X \overline{L^x} X^T U) + \frac{1}{2} tr(U^T X \overline{L^{xy}} Y^T V)$$

$$+ \frac{1}{2} tr(V^T Y \overline{L^{yx}} X^T U) + \frac{1}{2} tr(V^T Y \overline{L^y} Y^T V) \tag{10}$$

Where $tr(X)$ is the measure of trace of a matrix X. The regularization $g(U,V)$ penalizes large changes of the mapping function (U,V)between two nodes linked with a large weight. In other words, minimizing $g(U,V)$ encourages the smoothness of a mapping label information.

4. Comparative table

Several latest research papers are taken for the analysis in this section. Some of the important parameters are taken for the evaluation.

Author	Methodology Employed	Dataset	Advantage	Limitation	Performance measure
Peng, et al. [31]	Semantic correlations	Wikipedia, XMedia and NUS-WIDE datasets	This has the high efficiency than the existing method in the three database.	The weight value of the data is not used, which helps to improves the performance of the system.	cross modal factor analysis(CFA), Canonical Correlation analysis(CCA) and Joint representation learning
Zhai, et al. [32]	Joint Graph Regularized Heterogeneous Metric Learning (JGRHML)	Wikipedia and XMedia dataset	This helps to learn the high level sematic metric through label propagation.	The jointly modeling multiple modality is not used in this method, that helps in used for many applications.	CCA,CFA, CCA+SMN and JGRHML.
Xie, et al. [33]	Multi-graph Cross-modal Hashing (MGCMH)	Wikipedia, NUS-WIDE dataset	Nystrom method approximation is used to construct the effective graph and it shows the better performance when we compared to two unsupervised cross-modal hashing method.	The design the effective graph with the increasing complexity in graph construction, whose time may be linear is more difficult.	Composite Hashing with Multiple Information Sources (CHMIS), Cross-view Hashing (CVH), Inter-Media Hashing (IMH).
Wang, et al. [34]	Semi-Supervised Semantic Factorization Hashing (S3FH)	Wikipedia, NUS-WIDE dataset	S3FH changes the semantic label into the hash codes that stores the well preserve semantic value. This experimental result shows that it has high efficiency than the existing method.	The decomposition of the value causes in eigenvalue, while constrain of hash code orthogonally.	MAP score of cross-modal hash value, Precision and recall.
Deng, et al. [35]	Bayesian personalized ranking based heterogeneous metric learning (BPRHML)	Wikipedia dataset and the corel5k image dataset	They integrate the homogeneous and heterogeneous graph regularization into objective value and produce the better performance.	This method is only done on the images and need to test on the other media like audio, text and video.	MAP, Precision and Recall.
Cao, et al. [36]	Correlation among attributes.	Shoes dataset	This methods takes the advantageous of the structural sparsity of queries and experimental results shows that it shows the advance performance when compared it to the existing method of image retrieval technique with similar queries.	The performance of the image retrieval has to be increased and only image is retrieved in this method.	Ranking score (NDCG from 10 to 100 with increment value of 10 level).
Lei Huang and Yuxin Peng , [37]	Semantic Entity Projection (SEP)	Wikipedia dataset	The proposed method uses the SEP technique and experimental result shows that the system achieves the better result when we compared to the other existing method.	The correlation between the entity levels were not fully exploited and also delivers poor result for lower amount of data.	MAP score.
J. Qi , et al. [38]	Deep Multimodal Learning Method (DML)	Wikipedia, NUS Wide 10k dataset.	The experimental result shows that the proposed system is effectively when compared with other methods.	The technique has to be implement to utilize the unlabeled data in the system to improve the efficiency.	MAP score.
L. Xie, et al. [39]	A Cross-Modal Self-Taught Learning (CMSTL)	Wikipedia articles and NUS-WIDE	The hierarchical generation was used to obtain the multi-modal topic and it gives the better performance in cross-modal retrieval.	The technique has to extended to the image hashing and also on the other media.	MAP, precision and recall.
K.I. Kim, et al. [40]	Link Confidence Measure	Local dataset.	This method exploits the intermediates match output and predicts potential similarities and this helps the maximum connectivity of the graph for the limited computational resources.	The evaluation is not provided due to it is difficult to construct objective criteria.	Hit ratio.
X. Zhai, et al. [41]	Semantic Information	Wikipedia and XMedia dataset	The iterative optimization algorithm solves the problem of corresponding optimization and shows the better result in both single-media retrieval and cross-media retrieval with the state-of-art method.	The more kind of correlation among different media can be used to improve the performance.	MAP score, precision and recall.

5. Conclusion

The more number of heterogeneous data are increasing in the internet and it is increasing difficult to get the required data in the large datasets. The single media retrieval is used to give the same media result as the queries. The user requires the cross-media retrieval for the consistent result, which helps to give more information. The finding the similarities across the different media is very difficult and makes it hard task for cross-media retrieval. The lot of researches is done on cross media retrieval and one of the important feature used in the cross-media retrieval is JGR method. The investigation of the methods used in the JGR technique in the cross-media retrieval on the several research is made in this paper. The many research is used the Wikipedia dataset to experiment their method. The iterative optimization algorithm is used in the Zhai, et al. [40] research, which helps to solve the problem of corresponding optimization and in Wang, et al. [34], semi-supervised sematic factorization hashing is used that turns the semantic label into hash value, provides the better result when it compared to the existing methods of similar methods.

References

[1] O. Allani, H.B. Zghal, N. Mellouli, and H. Akdag, "Pattern graph-based image retrieval system combining semantic and visual features", Multimedia Tools and Applications, pp.1-30, 2017.

[2] Y. Yang, F. Nie, D. Xu, J. Luo, Y. Zhuang, and Y. Pan, "A multimedia retrieval framework based on semi-supervised ranking and relevance feedback", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 4, pp.723-742, 2012.

[3] Y. Yang, Y.T. Zhuang, F. Wu, and Y.H. Pan, "Harmonizing hierarchical manifolds for multimedia document semantics understanding and cross-media retrieval", IEEE Transactions on Multimedia, vol. 10 no. 3, pp.437-446, 2008.

[4] Y.T. Zhuang, Y. Yang, and F. Wu, "Mining semantic correlation of heterogeneous multimedia data for cross-media retrieval", IEEE Transactions on Multimedia, vol. 10, no. 2, pp.221-229, 2008.

[5] M. Li, L. Li, and F. Nie, "Ranking with adaptive neighbors", Tsinghua Science and Technology, vol. 22, no. 6, pp.733-738, 2017.

[6] Y. Peng, X. Zhai, Y. Zhao, and X. Huang, "Semi-supervised cross-media feature learning with unified patch graph regularization", IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. (3), pp.583-596, 2016.

[7] Z. Pan, W. Chen, M. Zhang, J. Liu, and G. Wu, "Virtual reality in the digital Olympic museum", IEEE Computer Graphics and Applications, vol. 29, no. 5, 2009.

[8] R. Ren, and J. Collomosse, "Visual sentences for pose retrieval over low-resolution cross-media dance collections", IEEE Transactions on Multimedia, vol. 14, no. 6, pp.1652-1661, 2012.

[9] X. Zhai, Y. Peng, and J. Xiao, "Learning cross-media joint representation with sparse and semisupervised regularization", IEEE Transactions on Circuits and Systems for Video Technology, vol. 24, no. 6, pp.965-978, 2014.

[10] F. Wu, X. Jiang, X. Li, S. Tang, W. Lu, Z. Zhang, and Y. Zhuang, "Cross-modal learning to rank via latent joint representation", IEEE Transactions on Image Processing, vol. 24, no. 5, pp.1497-1509, 2015.

[11] Y.X. Peng, W.W. Zhu, Y. Zhao, C.S. Xu, Q.M. Huang, H.Q. Lu, Q.H. Zheng, T.J. Huang, and W. Gao, "Cross-media analysis and reasoning: advances and directions", Frontiers of Information Technology & Electronic Engineering, vol. 18, no. 1, pp.44-57, 2017.

[12] F. Sun, Y. Xu, and J. Zhou, "Active learning SVM with regularization path for image classification", Multimedia Tools and Applications, vol. 75, no. 3, pp.1427-1442, 2016.

[13] S. Wang, P. Pan, Y. Lu, and L. Xie, "Improving cross-modal and multi-modal retrieval combining content and semantics similarities with probabilistic model", Multimedia Tools and Applications, vol. 74, no. (6), pp.2009-2032, 2015.

[14] P. Arulmozhi, and S. Abirami, "A comparative study of hash based approximate nearest neighbor learning and its application in image retrieval", Artificial Intelligence Review, pp.1-33, 2017.

[15] J. Qi, X. Huang, and Y. Peng, "Cross-media similarity metric learning with unified deep networks", Multimedia Tools and Applications, pp.1-19, 2017.

[16] Zhai, X., Peng, Y. and Xiao, J., 2012. Effective heterogeneous similarity measure with nearest neighbors for cross-media retrieval. Advances in Multimedia Modeling, pp.312-322.

[17] X. Zhai, Y. Peng, and J. Xiao, "Cross-media retrieval by intra-media and inter-media correlation mining. Multimedia systems, vol. 19, no. 5, pp.395-406, 2013.

[18] B. Lu, G.R. Wang, and Y. Yuan, "A novel approach towards large scale cross-media retrieval", Journal of Computer Science and Technology, vol. 27, no. 6, pp.1140-1149, 2012.

[19] B. Lu, G. Wang, and Y. Yuan, "Towards large scale cross-media retrieval via modeling heterogeneous information and exploring an efficient indexing scheme", In Computational Visual Media pp. 202-209, Springer, Berlin, Heidelberg, 2012.

[20] H. Zhang, Y.Y. Wang, H. Pan, and F. Wu, "Understanding visual-auditory correlation from heterogeneous features for cross-media retrieval", Journal of Zhejiang University SCIENCE A, vol. 9, no. 2, pp.241-249, 2008.

[21] H. Zhang, and J. Weng, "Measuring multi-modality similarities via subspace learning for cross-media retrieval", In Pacific-Rim Conference on Multimedia, pp. 979-988, Springer, Berlin, Heidelberg, November 2006

[22] K. Liu, S. Wei, Y. Zhao, Z. Zhu, Y. Wei, and C. Xu, "Accumulated reconstruction error vector (AREV): a semantic representation for cross-media retrieval", Multimedia Tools and Applications, vol. 74, no. 2, pp.561-576, 2015.

[23] P. Bellini, D. Cenni, and P. Nesi, "Optimization of information retrieval for cross media contents in a best practice network", International Journal of Multimedia Information Retrieval, vol. 3, no. 3, pp.147-159, 2014.

[24] D. Damm, C. Fremerey, V. Thomas, M. Clausen, F. Kurth, and M. Müller, "A digital library framework for heterogeneous music collections: from document acquisition to cross-modal interaction", International Journal on Digital Libraries, pp.1-19, 2012.

[25] J. Yan, H. Zhang, J. Sun, Q. Wang, P. Guo, L. Meng, W. Wan, and X. Dong, "Joint graph regularization based modality-dependent cross-media retrieval", Multimedia Tools and Applications, pp.1-19, 2017.

[26] Y. Zhuang, Q. Li, and L. Chen, "A Unified Indexing Structure for Efficient Cross-Media Retrieval", In DASFAA pp. 677-692, 2009, April.

[27] Y. Zhuang, and Y. Yang, "Boosting cross-media retrieval by learning with positive and negative examples", In International Conference on Multimedia Modeling pp. 165-174, Springer, Berlin, Heidelberg. 2007, January.

[28] K. Song, Y. Tian, and T. Huang, "Improving the image retrieval results via topic coverage graph", Advances in Multimedia Information Processing-PCM 2006, pp.193-200, 2006.

[29] H. Zhang, X. Gao, P. Wu, and X. Xu, "A cross-media distance metric learning framework based on multi-view correlation mining and matching", World Wide Web, vol. 19, no. 2, pp.181-197, 2016.

[30] A.K. Smith, K.H. Cheung, K.Y. Yip, M. Schultz, and M.B. Gerstein, "LinkHub: a Semantic Web system that facilitates cross-database queries and information retrieval in proteomics", BMC bioinformatics, vol. 8, no. 3, p.S5, 2007.

[31] Y. Peng, X. Zhai, Y. Zhao, and X. Huang, "Semi-supervised cross-media feature learning with unified patch graph regularization", IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 3, pp.583-596, 2016.

[32] X. Zhai, Y. Peng, and J. Xiao, "Heterogeneous Metric Learning with Joint Graph Regularization for Cross-Media Retrieval", In AAAI. 2013, June.

[33] L. Xie, L. Zhu, and G. Chen, "Unsupervised multi-graph cross-modal hashing for large-scale multimedia retrieval", Multimedia Tools and Applications, vol. 75, no. 15, pp.9185-9204, 2016.

[34] J. Deng, L. Du, and Y.D. Shen, "Heterogeneous Metric Learning for Cross-Modal Multimedia Retrieval" In WISE (1) pp. 43-56, 2013, October.

[35] J. Wang, G. Li, P. Pan, and X. Zhao, "Semi-supervised semantic factorization hashing for fast cross-modal retrieval", Multimedia Tools and Applications, pp.1-19, 2017.

[36] X. Cao, H. Zhang, X. Guo, S. Liu, and X. Chen, "Image retrieval and ranking via consistently reconstructing multi-attribute queries", In European Conference on Computer Vision pp. 569-583, Springer, Cham. 2014, September.

[37] L. Huang, and Y. Peng, "Cross-Media Retrieval via Semantic Entity Projection", In International Conference on Multimedia Modeling pp. 276-288, Springer, Cham. 2016, January.

[38] J. Qi, X. Huang, and Y. Peng, "Cross-Media Retrieval by Multimodal Representation Fusion with Deep Networks", In International Forum of Digital TV and Wireless Multimedia Communication pp. 218-227, Springer, Singapore, 2016, November.

[39] L. Xie, P. Pan, Y. Lu, and S. Jiang, "Cross-modal self-taught learning for image retrieval", In International Conference on Multimedia Modeling pp. 257-268, Springer, Cham. 2015, January.

[40] K.I. Kim, J. Tompkin, M. Theobald, J. Kautz, and C. Theobalt, "Match graph construction for large image databases", In European Conference on Computer Vision pp. 272-285, Springer, Berlin, Heidelberg. 2012, October.

[41] X. Zhai, Y. Peng, and J. Xiao, "Learning cross-media joint representation with sparse and semisupervised regularization", IEEE Transactions on Circuits and Systems for Video Technology, vol. 24 no. 6, pp.965-978, 2014.