A Survey on Entropy Optimized Feature-based Bag-of-Words Representation for Information Retrieval

Swaroop Kale[1], H. A. Hingoliwala[2]

[1]Department of Computer Engineering, Jaywantrao Sawant College of Engineering, Pune, Maharashtra, India

[2]Professor, HOD, Department of Computer Engineering, Jaywantrao Sawant College of Engineering, Pune, Maharashtra, India

Abstract: ***In this paper, we present a supervised dictionary learning method for improving the component based Bag-of-Words (BoW) representation towards Information Retrieval. Taking after the bunch theory, which expresses that focuses in a similar group are probably going to satisfy a similar data require, we propose the utilization of an entropy-based enhancement basis that is more qualified for recovery of order. We show the capacity of the proposed strategy, curtailed as EO-BoW, to enhance the recovery execution by giving broad analyses on two multi-class picture datasets. The BoW model can be connected to different spaces too, so we additionally assess our approach utilizing a gathering of 45 time-arrangement datasets, a content dataset and a video dataset. The increases are three-crease since the EO-BoW can enhance the mean Average Precision, while decreasing the encoding time and the database stockpiling necessities. At long last, we give prove that the EO-BoW keeps up its representation capacity notwithstanding when used to recover objects from classes that were not seen amid the preparation.***

Keywords: Entropy optimized, Bag-of-words, Information Retrieval

1. Introduction

Information retrieval (IR) is the undertaking of recovering articles, e.g., pictures, from a database given the client's data require [28]. Early research concentrated generally on content recovery [20], [28], however then immediately extended to different territories, such as picture recovery [7], and video recovery [15], since the accessibility of computerized innovation prompted to an awesome increment of interactive media information. A comparative development in the accessibility of time arrangement information, i.e., information that are made out of a succession of estimations, for example, medicinal observing information [2], too brought the enthusiasm up in the recovery of time arrangement [5], [18].

We can all the more formally characterize IR as takes after: Given a gathering of items D = fd1; d2; :::; dng and an inquiry question q, rank the articles in D as per their importance to q and select a subset of the most pertinent items. As we as of now specified, the gathering can contain any kind of articles, for example, pictures, recordings, time-arrangement or content archives. Be that as it may, we concentrate primarily on picture recovery since it is the most examined and testing part of sight and sound data recovery. This is without loss of sweeping statement as the proposed strategy can be connected to a few different sorts of information, for example, video, sound and time-arrangement, with minor alterations.

The proposed strategy can't be straightforwardly utilized for content recovery, be that as it may it can use word-installing strategies, for example, [30], to make advanced representations for content reports.

A standout amongst the most critical difficulties in picture recovery is the supposed semantic hole [40], between the low-level representation of a picture and its larger amount ideas. In different words, the semantic hole depicts the wonder in which pictures with comparable semantic substance have extremely diverse low-level representations (e.g., shading data) furthermore, the other way around. The high dimensionality of pictures decreases the recovery execution significantly more, both as far as inquiry time and accuracy.

A few methodologies have been proposed to separate important low-dimensional components from pictures. The reason of highlight extraction is to bring down the dimensionality of the pictures, which lessens the capacity prerequisites and the recovery time, and to connect the semantic crevice, which increments the recovery accuracy. Maybe the most generally utilized what's more, fruitful strategy for this errand is the element based Sack of-Words model [39], otherwise called Bag-of-Features (BoF) or Bag-of-Visual Words (BoVW). The element based BoW approaches, portrayed in detail in Section 3.1, ought to not be mistaken for the standard printed BoW approaches [28], that are talked about later on in this Section. In the rest of the composition we truncate the component based Bag-of-Words models as BoW. The BoW demonstrate treats every picture as an archive that contains various diverse "visual" words. At that point a picture is spoken to as a histogram over a set of agent words, known as lexicon or codebook.

These histograms portray the relating pictures and they can be utilized for the resulting recovery undertakings. The BoW pipeline can be condensed as takes after:

1) Include extraction, in which numerous elements, for example, Filter descriptors [26], are separated from every picture. That way, the component space is shaped where each picture is spoken to as an arrangement of components.
2) Word reference learning, in which the removed components are utilized to take in a word reference of delegate highlights (likewise called words or codewords),
3) Include quantization and encoding, in which each highlight is spoken to utilizing a codeword from the learned lexicon and a histogram is removed for every

picture. That way, the histogram space is framed where every picture is spoken to by a consistent dimensionality histogram vector.

Not at all like printed words, visual words are not predefined what's more, the nature of the removed representation basically depends on how these words are picked. Early element based BoW approaches (e.g., [23], [39]) utilized unsupervised bunching calculations, for example, k-means, to group the arrangement of components what's more, realize a word reference. These unsupervised methodologies accomplished promising outcomes and created codebooks that were non sufficiently specific to be utilized for any undertaking. In any case, taking in a discriminative word reference custom-made to a particular issue is relied upon to perform essentially better.

In fact, managed word reference learning, is appeared to accomplish prevalent execution regarding grouping exactness [13], [24], [25]. These techniques create discriminative word references that are helpful for the given grouping issue. Despite the fact that an exceedingly discriminative representation is sought for characterization undertakings, it is not generally ideal for recovery since it may seriously mutilate the similitude between pictures keeping in mind the end goal to pick up segregation capacity. This can be better comprehended by an illustration. Assume that we need to take in a word reference that recognizes apples from oranges utilizing just two code-words and a great discriminative word reference learning calculation exists. After the preparation procedure, every apple is spoken to by a vector (x; y) 2 N(0; 0:1) N(1; 0:1) and every orange by a vector (x; y) 2 N(1; 0:1) N(0; 0:1), where N(; 2) is a typical appropriation with mean and change 2. This codebook would be astounding for characterization and recovery of oranges also, apples, as it figures out how to directly isolate the two classes with an extensive edge. Presently, what might happen on the off chance that we utilize this representation to recover an another natural product, for example, pears on the other hand bananas? Does this word reference picked up its separation capacity to the detriment of its representation capacity and would really perform more regrettable (than an unsupervised word reference) on this assignment? Our analyses (Section 4.2.2) affirm this speculation, since we found that a profoundly discriminative representation exceeds expectations at its educated space, yet its discriminative capacity outside this area is extremely constrained.

The intrigued peruser is alluded to [34], where the issues of exchange learning and area adjustment are discussed.We moreover expect significance criticism strategies, i.e., techniques that are used to better distinguish the client's data require [10], [11], [45], not to work effectively outside the preparation area, since the representation capacity is now lost. Preferably, the previously mentioned issue would be illuminated in the event that we could enhance the educated representation for each conceivable data require. Since this is somewhat infeasible, we can take in a representation utilizing a huge scale database with commented on pictures, for example, ImageNet [37], that covers a wide scope of data needs. Be that as it may, it is not generally conceivable to procure a huge arrangement of explained preparing information, for the most part in light of the high cost of explanation. In this manner, a decent representation for recovery ought to an) enhance the recovery measurements inside its preparation area and b) have the capacity to "exchange" the educated information to other comparative spaces.

The last is particularly essential as it guarantees that we can learn utilizing a little and agent preparing set. To this end, we propose a managed word reference learning strategy that produces recovery arranged codebooks by holding fast to the group theory. Bunch speculation states that focuses in a similar bunch are probably going to satisfy the same data require [43]. We select an arrangement of centroids in the histogram space and we take in a codebook that minimizes the entropy of every bunch. Every centroid can be considered as a delegate question and the improvement plans to augment the significant data around it. The entropy target goes about as a mellow segregation paradigm, which attempt to make the bunches as immaculate as would be prudent. To comprehend why this rule varies from other more discriminative criteria, for example, the Fisher's proportion [13], or max-edge goals [24], [25], take note of that we don't push bunches or focuses far from each other. Rather, the bunches are fitted to the current information circulation and we just attempt to move unimportant focuses to the closest applicable bunch. This permits us to enhance the codebook without over-fitting the representation over the preparing area. We approve our cases by illustrating the capacity of the proposed technique to effectively recover pictures that have a place with classes that are not seen amid the preparing process. We ought to specify that our technique is definitely not constrained to picture recovery. It is sufficiently general to be connected to any undertaking that includes BoW representations, for example, video [15], sound [36], and time arrangement recovery [12].

The term Bag-of-Words is likewise used to allude to the common dialect preparing strategies that handle a content archive as accumulation of its words [28]. A standout amongst the most broadly utilized such techniques utilizes a term recurrence (tf) histogram, which tallies the appearances of every word, to speak to a report as a vector. In these methodologies the expressions of the word reference are predefined and can't be modified. Lexicon learning for this representation points chiefly to prune the word reference by selecting the most helpful elements [27], of changing the words and removing another representation. In spite of the fact that the proposed strategy can't be connected when the printed BoW representation is utilized, with the coming of the word-inserting models, e.g., [30], is conceivable to outline word to a "significant" low-dimensional persistent vector.

2. BOF Image Representation

A Bag of Features strategy is one that speaks to pictures as order less accumulations of nearby elements. The name originates from the Bag of Words representation utilized as a part of printed data recovery. This segment gives a clarification of the Bag of Features picture representation, concentrating on the abnormal state prepare free of the application. More advanced varieties and other usage points of interest are talked about later in this report.

There are two normal viewpoints for clarifying the BoF picture representation. The to begin with is by similarity to the Bag of Words representation. With Bag of Words, one speaks to a record as a standardized histogram of word tallies. Generally, one numbers every one of the words from a lexicon that show up in the report. This word reference may prohibit certain non informative words, for example, articles (like "the"), and it might have a solitary term to speak to an arrangement of equivalent words. The term vector that speaks to the archive is a meager vector where every component is a term in the word reference and the estimation of that component is the quantity of times the term shows up in the archive partitioned by the aggregate number of lexicon words in the archive (and accordingly, it is likewise a standardized histogram over the terms). The term vector is the Bag of Words archive representation – called a "sack" since all requesting of the words in the archive have been lost.

The Bag of Features picture representation is closely resembling. A visual vocabulary is built to speak to the word reference by bunching highlights extricated from an arrangement of preparing pictures. The picture highlights speak to neighborhoods the picture, similarly as words are nearby elements of a record. Grouping is required so that a discrete vocabulary can be produced from millions (or billions) of neighborhood elements examined from the preparation information. Every element group is a visual word. Given a novel picture, elements are distinguished and allotted to their closest coordinating terms (bunch focuses) from the visual vocabulary. The term vector is at that point just the standardized histogram of the quantized elements recognized in the picture.

The second approach to clarify the BoF picture representation is from a codebook viewpoint. Components are extricated from preparing pictures and vector quantized to build up a visual codebook. A novel picture's components are doled out the closest code in the codebook. The picture is lessened to the arrangement of codes it contains, spoke to as a histogram. The standardized histogram of codes is precisely the same as the standardized histogram of visual words, however is inspired from an alternate perspective. Both "codebook" and "visual vocabulary" wording is available in the overviewed writing.

The BoF expression vector is a minimized representation of a picture which disposes of large-scale spatial data and the relative areas, scales, and introductions of the elements. A contemporary huge scale BoF-based picture recovery framework may have a lexicon of 100,000 visual words and 5,000 components separated per picture. Hence in a picture where there are no copy visual words (bizarre), the term vector will have 95% of its components as zeros. The solid sparsity of term vectors considers proficient ordering plans and other execution enhancements, as examined in later areas.

At an abnormal state, the method for creating a Bag of Features picture representation is appeared in Figure 1 and abridged as takes after:

1) Build Vocabulary: Extract highlights from all pictures in a preparation set. Vector quantize, alternately group, these components into a "visual vocabulary," where every bunch speaks to a "visual word" or "term." In a few works, the vocabulary is known as the "visual codebook." Terms in the vocabulary are the codes in the codebook.
2) Assign Terms: Extract highlights from a novel picture. Utilize Nearest Neighbors or a related procedure to relegate the elements to the nearest terms in the vocabulary.
3) Generate Term Vector: Record the checks of every term that shows up in the picture to make a standardized histogram speaking to a "term vector." This term vector is the Bag of Features representation of the picture. Term vectors may likewise be spoken to in courses other than basic term recurrence, as talked about later.

There are various plan decisions required at every progression in the BoF representation. One key choice includes the decision of highlight recognition and representation. Many utilize an intrigue point administrator, for example, the Harris-Affine finder or the Maximally Stable Extremal Regions (MSER) At each intrigue point, regularly a couple of thousand for every picture, a high-dimensional component vector is utilized to portray the neighborhood picture fix. Lowe's 128-measurement SIFT descriptor is a prominent decision.

3. Feature Detection and Representation

3.1 Feature Detection

Feature detection is the way toward choosing where and at what scale to test a picture. The yield of highlight identification is an arrangement of key points that determine areas in the picture with comparing scales and introductions. These key points are particular from highlight descriptors, which encode data from the pixels in the area of the key points. In this way, highlight recognition is a different procedure from highlight representation in BoF approaches. There is a significant collection of writing that spotlights on distinguishing the area and degree of good elements from no less than two distinctive sub-fields of PC vision. The first created from the objective of discovering key points helpful for picture enrollment that are steady under minor relative and photometric changes. These element discovery techniques are alluded to as Interest Point Operators. The second gathering distinguishes highlights in view of computational models of the human visual consideration framework. These strategies are worried with discovering areas in pictures that are outwardly notable. For this situation, wellness is regularly measured by how well the computational techniques anticipate human eye obsessions recorded by an eye tracker. In the accompanying two subsections, we examine both ways to deal with highlight discovery.

At long last, there is research that recommends producing key points by inspecting the pictures utilizing a framework or pyramid structure, or even by irregular testing.

3.2 Feature Descriptors

Notwithstanding figuring out where and to what degree a component exists in a picture, there is a different collection of research to decide how to speak to the neighborhood of pixels close to a limited district, called the component descriptor. The easiest approach is to just utilize the pixel power values, scaled for the extent of the district, or an eigenspace representation thereof. Standardized pixel representations, in any case, have performed more awful than numerous more modern representations .Nowak et al , among others) and have generally been surrendered by the BoF investigate group.

The most prevalent component descriptor in the BoF writing is the SIFT (Scale Invariant Highlight Transform) descriptor (Lowe, 2004). In a nutshell, the 128 dimensional SIFT descriptor is a histogram of reactions to arranged slope channels. The reactions to 8 angle introductions at each of 16 cells of a 4x4 matrix produce the 128 parts of the vector. The histograms in every cell are piece shrewd standardized. At scale 1, the cells are frequently 3x3 pixels. A contrasting option to the SIFT descriptor that has increased expanding prominence is SURF . The SURF calculation comprises of both include location and representation perspectives. It is intended to create highlights much the same as those created by a SIFT descriptor on Hessian-Laplace intrigue focuses, however utilizing proficient approximations. Reported outcomes show that SURF gives a noteworthy accelerate while coordinating or enhancing execution. Different descriptors which have been proposed incorporate Gabor channel banks, picture minutes, what's more, others. A review by Mikolajczyk and Schmid thinks about a few element descriptors, what's more, demonstrates that SIFT-like descriptors have a tendency to beat the others much of the time . The descriptors that were assessed, in any case, need shading data. This is as opposed to the bio-mimetic vision group which normally incorporates a shading opponency angle to highlight representations. There is confirmation that including shading data in highlight recognition and depiction may enhance BoF picture recovery execution. A late paper by van de Sande et al. presents an assessment of shading highlight descriptors. Reported outcomes demonstrate a mix of shading descriptors beats SIFT on a picture grouping undertaking and that, of the shading descriptors, OpponentSIFT is most for the most part valuable.

4. Entropy Optimized BoW Model

The objective of the EO-BoW strategy is to take in a codebook that minimizes the entropy in the histogram space by utilizing a preparation set of pictures, where the i-th picture is commented on by a name li ∈ {1, ..., NC } and NC is the quantity of preparing classes. Instinctively, the entropy in the histogram space is minimized when the picture histograms are accumulated in immaculate groups, i.e., every bunch contains pictures of a similar class. So as to gauge the entropy in the histogram space, the si vectors are grouped into NT bunches. The centroid of the k-th group is indicated by ck (k = 1...NT). At that point, the entropy of the k-th group can be characterized as: where pjk is the likelihood that a picture of the k-th group has a place with the class j. This likelihood is evaluated as pjk = hjk/nk, where nk is the quantity of picture histograms in group k and hjk is the quantity of picture histograms in group k that have a place with class j. As it was at that point said, every centroid can be considered as a delegate question for which we need to improve the codebook. As indicated by the bunch speculation low entropy bunches, i.e., groups that contain for the most part vectors from pictures of a similar class, are best for recovery undertakings to high-entropy bunches, i.e., groups that contain vectors from pictures that have a place with a few unique classes. Hence we mean to take in a codebook that minimizes the add up to entropy of a bunch design, which is characterized as: where rk = nk/N is the extent of pictures in bunch k.

5. Conclusions

In this paper we proposed a regulated word reference learning technique, the EO-BoW, which streamlines a recovery arranged target work. We exhibited the capacity of the proposed technique to enhance the recovery execution utilizing two picture datasets, an accumulation of time-arrangement datasets, a content dataset and a video dataset. Initially, for a given word reference measure, it can enhance the mAP over the pattern techniques and other cutting edge representations. Second, these changes permit us to utilize littler representations which promptly means bring down capacity necessities and quicker recovery. In return for these, our strategy requires a little arrangement of explained preparing information. In spite of the fact that the picked up execution is associated to the size and the nature of the preparing dataset, we demonstrated that the proposed strategy does not lose its representation capacity notwithstanding when a little preparing dataset is utilized. At last, we exhibited that the EOBoW enhances the recovery execution utilizing two distinctive similitude measurements, the Euclidean and the chi-square separation. Hence, it can be consolidated with any inexact closest neighbor procedure that works with these comparability measurements to further build the recovery speed.

References

[1] A. Andoni and P. Indyk, ―Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions," in 47th Annual IEEE Symposium on Foundations of Computer Science, 2006, pp. 459– 468.

[2] M. M. Baig, H. Gholamhosseini, and M. J. Connolly, ―Acomprehensive survey of wearable and wireless ECG monitoring systems for older adults," Medical & Biological Engineering & Computing, vol. 51, no. 5, pp. 485–495, 2013.

[3] W. Bian and D. Tao. (2009) The COREL database for content based image retrieval. [Online]. Available: https://sites.google.com/ site/dctresearch/Home/content-based-image-retrieval

[4] Y.-L. Boureau, F. Bach, Y. LeCun, and J. Ponce, ―Learning mid-level features for recognition," in IEEE

Conference on Computer Vision and Pattern Recognition, 2010, pp. 2559–2566.

[5] F. K.-P. Chan, A. W.-C. Fu, and C. Yu, ―Haar wavelets for efficient similarity search of time-series: with and without time warping," IEEE Transactions on Knowledge and Data Engineering, vol. 15, no. 3, pp. 686–705, 2003.

[6] D. Chatzakou, N. Passalis, and A. Vakali, ―Multispot: Spotting sentiments with semantic aware multilevel cascaded analysis," in Big Data Analytics and Knowledge Discovery, 2015, pp. 337–350.

[7] R. Datta, D. Joshi, J. Li, and J. Z. Wang, ―Image retrieval: Ideas, influences, and trends of the new age," ACM Computing Surveys, vol. 40, no. 2, 2008.

[8] K. Eguchi and V. Lavrenko, ―Sentiment retrieval using generative models," in Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2006, pp. 345–354.

[9] D. Gorisse, M. Cord, and F. Precioso, ―Locality-sensitive hashing for chi2 distance," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 2, pp. 402–409, 2012.

[10] X. He, D. Cai, and J. Han, ―Learning a maximum margin subspace for image retrieval," IEEE Transactions on Knowledge and Data Engineering, vol. 20, no. 2, pp. 189–201, 2008.

[11] S. C. Hoi, M. R. Lyu, and R. Jin, ―A unified log-based relevance feedback scheme for image retrieval," IEEE Transactions on Knowledge and Data Engineering, vol. 18, no. 4, pp. 509–524, 2006.

[12] A. Iosifidis, A. Tefas, and I. Pitas, ―Multidimensional sequence classification based on fuzzy distances and discriminant analysis," IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 11, pp. 2564–2575, 2013.

[13] ——, ―Discriminant bag of words based representation for human action recognition," Pattern Recognition Letters, vol. 49, pp. 185–192, 2014.

[14] H. Jegou, M. Douze, C. Schmid, and P. P ´ erez, ―Aggregating ´ local descriptors into a compact image representation," in IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 3304–3311.

[15] Y.-G. Jiang, C.-W. Ngo, and J. Yang, ―Towards optimal bag-offeatures for object categorization and semantic video retrieval," in Proceedings of the 6th ACM international conference on Image and Video retrieval, 2007, pp. 494–501.

[16] I. T. Jolliffe, Principal Component Analysis, 2nd ed., ser. Springer Series in Statistics. New York: Springer-Verlag, 2002.

[17] E. Keogh, X. Xi, L. Wei, and C. A. Ratanamahatana. (2006) The UCR time series classification/clustering homepage. [Online]. Available: http://www. cs. ucr. edu/˜ eamonn/time series data

[18] E. J. Keogh and M. J. Pazzani, ―Relevance feedback retrieval of time series data," in Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, 1999, pp. 183–190.

[19] Y. Kuang, M. Byrod, and K. Astrom, ―Supervised feature quantization with entropy optimization," in IEEE International Conference on Computer Vision Workshops, 2011, pp. 1386–1393.

[20] W. Lam, M. Ruiz, and P. Srinivasan, ―Automatic text categorization and its application to text retrieval," IEEE Transactions on Knowledge and Data Engineering, vol. 11, no. 6, pp. 865–879, 1999.

[21] I. Laptev, M. Marszałek, C. Schmid, and B. Rozenfeld, ―Learning realistic human actions from movies," in IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.

[22] S. Lazebnik and M. Raginsky, ―Supervised learning of quantizer codebooks by information loss minimization," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 7, pp. 1294– 1309, 2009.

[23] S. Lazebnik, C. Schmid, and J. Ponce, ―Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, 2006, pp. 2169–2178.

[24] X.-C. Lian, Z. Li, B.-L. Lu, and L. Zhang, ―Max-margin dictionary learning for multiclass image categorization," in Proceedings of the 11th European Conference on Computer Vision, 2010, pp. 157–170.

[25] H. Lobel, R. Vidal, D. Mery, and A. Soto, ―Joint dictionary and classifier learning for categorization of images using a max-margin framework," in Proceedings of the 6th Pacific-Rim Symposium, 2014, pp. 87–98.

[26] D. G. Lowe, ―Object recognition from local scale-invariant features," in Proceedings of the 7th IEEE international conference on Computer vision, vol. 2, 1999, pp. 1150–1157.

[27] R. E. Madsen, S. Sigurdsson, L. K. Hansen, and J. Larsen, ―Pruning the vocabulary for better context recognition," in Proceedings of the 17th International Conference on Pattern Recognition, vol. 2, 2004, pp. 483–488.

[28] C. D. Manning, P. Raghavan, and H. Schutze, ¨ Introduction to Information Retrieval, 1st ed. Cambridge: Cambridge University Press, 2008.

[29] C. D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard, and D. McClosky, ―The Stanford CoreNLP natural language processing toolkit," in Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, 2014, pp. 55– 60.

[30] T. Mikolov, K. Chen, G. Corrado, and J. Dean, ―Efficient estimation of word representations in vector space," arXiv preprint arXiv:1301.3781, 2013.

[31] M. Muja and D. G. Lowe, ―Fast approximate nearest neighbors with automatic algorithm configuration," in International Conference on Computer Vision Theory and Application, 2009, pp. 331–340.

[32] R. Negrel, D. Picard, and P.-H. Gosselin, ―Web-scale image retrieval using compact tensor aggregation of visual descriptors," MultiMedia, IEEE, vol. 20, no. 3, pp. 24–33, 2013.

[33] J. Nocedal and S. Wright, Numerical Optimization, 2nd ed., ser. Springer Series in Operations Research and Financial Engineering. New York: Springer-Verlag, 2006.

[34] S. J. Pan and Q. Yang, ―A survey on transfer learning," IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345–1359, 2010.

[35] B. Pang and L. Lee, ―A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," in Proceedings of the 42nd annual meeting on Association for Computational Linguistics, 2004, p. 271.

[36] M. Riley, E. Heinen, and J. Ghosh, ―Atext retrieval approach to content-based audio retrieval," in Proceedings of the 9th International Conference on Music Information, 2008, pp. 295–300.

[37] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein et al., ―ImageNet large scale visual recognition challenge," International Journal of Computer Vision, pp. 1–42, 2014.

[38] C. Schuldt, I. Laptev, and B. Caputo, ―Recognizing human actions: ¨ a local svm approach," in Proceedings of the 17th International Conference on Pattern Recognition, vol. 3, 2004, pp. 32–36.

[39] J. Sivic and A. Zisserman, ―Video google: A text retrieval approach to object matching in videos," in Proceedings of the 9th IEEE International Conference on Computer Vision, 2003, pp. 1470–1477.

[40] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, ―Content-based image retrieval at the end of the early years," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 12, pp. 1349–1380, 2000.

[41] R. Socher, A. Perelygin, J. Y. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts, ―Recursive deep models for semantic compositionality over a sentiment treebank," in Proceedings of the Conference on Empirical Methods in Natural Language Processing, vol. 1631, 2013, p. 1642.

[42] D. Tao, X. Tang, X. Li, and Y. Rui, ―Direct kernel biased discriminant analysis: a new content-based image retrieval relevance feedback algorithm," IEEE Transactions on Multimedia, vol. 8, no. 4, pp. 716–727, 2006.

[43] E. M. Voorhees, ―The cluster hypothesis revisited," in Proceedings of the 8th annual international ACM SIGIR conference on Research and development in information retrieval, 1985, pp. 188–196.

[44] J. Yang, Y.-G. Jiang, A. G. Hauptmann, and C.-W. Ngo, ―Evaluating bag-of-visual-words representations in scene classification," in Proceedings of the international workshop on Workshop on multimedia information retrieval, 2007, pp. 197–206.

[45] X. S. Zhou and T. S. Huang, ―Relevance feedback in image retrieval: A comprehensive review," Multimedia systems, vol. 8, no. 6, pp. 536–544, 2003.