Placing Broadcast News Videos in their Social Media Context using Hashtags

Joseph G. Ellis, Svebor Karaman, Hongzhi Li, Hong Bin Shim and Shih-Fu Chang Columbia University {jge2105, svebor.karaman, hongzhi.li, h.shim, sc250}@columbia.edu

ABSTRACT

With the growth of social media platforms in recent years, social media is now a major source of information and news for many people around the world. In particular the rise of hashtags have helped to build communities of discussion around particular news, topics, opinions, and ideologies. However, television news programs still provide value and are used by a vast majority of the population to obtain their news, but these videos are not easily linked to broader discussion on social media. We have built a novel pipeline that allows television news to be placed in its relevant social media context, by leveraging hashtags. In this paper, we present a method for automatically collecting television news and social media content (Twitter) and discovering the hashtags that are relevant for a TV news video. Our algorithms incorporate both the visual and text information within social media and television content, and we show that by leveraging both modalities we can improve performance over single modality approaches.

Keywords

video analysis; video news; hash tagging; social media

1. INTRODUCTION

Social media platforms are a primary source of news information for many users. Twitter, in particular, is one of the most popular social media platform for news consumption and dissemination. Indeed, information about an event is often first posted on Twitter, oftentimes through first hand accounts of a news event. After this an engaging discussion is then added to and commented on by news agencies or other media outlets. Twitter users are often very active and post their opinions about news event by posting new tweets or reposting the existing popular tweets, enabling quick and wide spread dissemination of any information deemed valuable. To help build a conversation around a particular event, topic, or idea, Twitter users link their tweets to a subject by using hashtags. A hashtag, is a short word or acronym preceded by the "#" sign. Hashtags can then be searched easily within social media sites, making it simple to quickly obtain cogent information about a possibly very narrow topic or opinion in an efficient manner. This is the simplest way

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

MM '16, October 15 - 19, 2016, Amsterdam, Netherlands

© 2016 Copyright held by the owner/author(s). Publication rights licensed to ACM. ISBN 978-1-4503-3603-1/16/10...\$15.00

DOI: http://dx.doi.org/10.1145/2964284.2970929

Table 1: Examples of found hashtags for a set of news programs using our multimodal algorithm. Hashtags in red were found predominantly using our text pipeline, and those in blue were found using predominantly visual pipelines.

Frame from video	Found hashtags		
	<pre>#politics, #trump2016, #realdonaldtrump, #goptownhall, #votetrumpsc</pre>		
	<pre>#politics, #dumptrump, #realdonaldtrump, #gopdebate</pre>		
	#iphone, #apple, #cybersecurity		

to link related content into threads on social media platforms. As the discussion grows and evolve, different hashtags can be used in relation to a subject and the meaning of some hashtags may not be obvious given only the hashtag. Twitter users often like to use short hashtags due to the restriction on the number of characters in a tweet. Thus, to engage efficiently in the discussion requires some understanding of the hashtags related to a subject.

Although social media platforms have become major sources of news, especially for young people, traditional TV news still provides a unique value to customers. The news videos usually provide rich video footage and professional commentators' opinion which many not be covered by social media. Twitter attracts attentions from major news agencies and TV channels as it can provide an increased reach for their content. However, the content from TV broadcasts has to be properly integrated in the social media context in order to maximize this reach. By linking the broadcast news videos to hashtags, we let Twitter users have easy access to high quality content and can help TV news channels with a unique opportunity to reach a much wider audience.

The contributions of this paper are as follows:

- We present a pipeline and algorithms for automatically collecting social media and television news video and linking hashtags used on Twitter to the videos.
- We show improved performance for linking hashtags to television news when leveraging a multimodal approach over single visual or text modalities, demonstrating the importance of taking into account multiple modalities for this task.

2. RELATED WORKS

Twitter is one of the largest and fastest paced information channels on the Internet. For example, during the 2011 Tohoku earthquake information traveled throughout Twitter incredibly fast [19, 7]. Thanks to its rich amount and quick spreading of information, Twitter data attracts researchers from many different areas. Twitter data can be used to predict and respond to customers' complaints [10, 16], to predict election results, gather and spread breaking news, and even crime prediction [24].

An interesting and challenging problem using Twitter data is event detection and tracking. Topic Detection and Tracking (TDT) is a traditional research topic, which focus on finding events and linking documents to the related event in broadcast news streams [26, 25, 1]. However, TDT on twitter data faces new challenges due to the limited number of characters used in twitter messages and also the large number of noise and meaningless messages on twitter [9]. Several advanced methods have been developed recently to address the event detection and tracking problem on twitter data, to detect unspecified types of events [20, 18, 17], or discover predefined types of events from twitter messages [19, 4, 15].

Hashtags are used by social media users to convey an idea, event or feeling that is associated with a post. Recently, researchers have worked on automatically tagging a user's images with hashtags based on their user specific information [6, 2, 3]. In this application, we use hashtags [22, 5, 21, 23, 8] to organize tweets as topics and link them to broadcasting TV news events.

3. DATA COLLECTION

In this section we will describe the collection of data used for this work. We have collected a large scale multimodal dataset from twitter and broadcast news videos. The data that we have collected consists of content from the dates of 02/15/2016 - 02/24/2016.

3.1 Video news

To collect our set of video news programs we have built a recording infrastructure that automatically processes 100 hours of broadcast television news per day [13]. We begin by segmenting each of the full television programs into smaller semantically coherent segments that tend to be between 1-8 minutes long, and consist of one distinct topic. This segmentation is done using multimodal inputs, such as audio silence, anchor scenes, and text cues [11]. We have collected 3823 videos from this time window that should be used for matching to social media hashtags. Each of the videos include the closed caption transcripts that are associated with the video, and these are used for text matching.

3.2 Tweets

We collect tweets from the same 02/15/2016-02/24/2016 timeframe. To perform our tweet collection we have set up an infrastructure for downloading and crawling the freely available twitter public stream ¹. The Twitter public stream is a random sampling of the full twitter dataset, and therefore should be a reasonable representation of all of the information that flows through Twitter. The information from the public stream is returned in json formatted strings, and each tweet representation can consist of but is not limited to the following categories: images, videos, text, hashtags, usernames, geotags, timepoints, retweet status. Any tweets that did not contain hashtags were not used in our dataset. The collection statistics of our full dataset can be seen in Table 2. We can see that our dataset has over 2 million tweets, and more than 1 million images.

Table 2: Twitter Dataset Statistics

Hashtags	355,403	
Tweets	2,778,052	
Word Vocabulary Size	327,442	
Images	1,000,941	
Users	422,947	

4. VIDEO HASHTAGGING

In this section we will detail the algorithms and pipeline that we have used for processing the videos. Our goal is to link hashtags collected from Twitter to our video dataset. When attempting to process information on Twitter and in other microblogging formats, the information contained in one piece of information (tweet, post, etc.) tends to be very sparse, due to the small nature of each particular instance compared to the entire search space of the whole dataset. To solve this issue we have developed a framework that helps to densify the representation of social media data. Hashtags represent the idea, content, and feeling of a particular post and are therefore important for posting and searching social media. We thus organize all our Twitter content into hashtag representations.

We develop the following framework and representation for a hashtag. The i^{th} hashtag in our dataset is represented by a vector of the words (H_i^w) that are contained in tweets that have been marked with this hashtag. The images that are from tweets with this hashtag are represented as the set, (H_i^v) . We read over each of the tweets in our dataset, and augment the representation for each (H_i^w) using the words from the tweet if a hashtag appears in that particular tweet. Through this approach, we are able to build a dense representation for each hashtag based on all of the tweets that contain the hashtag. This allows for a more complete and robust representation of a topic than standard tweet-only approaches. We discard any hashtags that do not contain *at least* 300 tweets or *at least* 50 images as we are looking for popular hashtags. We represent the video news segments with a set of subsampled frames appearing in the segment (V_i^v) , and the full closed caption transcript (V_i^w) .

4.1 Text modalities

In this section we will describe how we link hashtags to broadcast news using text content. In our text matching frameworks we use unigram text features. In this work we have implemented two separate but related pipelines for performing text mapping. First, we have implemented a classic bag of words mapping pipeline that leverages the words that appear in each hashtag and the transcript for each video. This method, which we refer to as "bag-of-words", tends to work well for hashtags that have densely sampled our word vocabulary. However, for hashtags that have fewer tweets the bag of words representation can be quite sparse, and this can cause spurious matches to video transcripts.

Therefore, to overcome this issue we have developed a novel pipeline for matching hashtag text to video transcript. Our second method, which we will call "hashtag-probing" relies on first finding the most important words for each particular hashtag, and then using these words as a "probe" to find which videos contain information associated with this hashtag.

4.1.1 Bag-of-Words

For the "bag-of-words" representation of hashtags, we first filter all stopwords and words from our vocabulary that do not have at least 30 occurrences within our dataset. This filtering aims at removing words that appear only sparsely in social media. After this filtering our vocabulary size is 30, 845.

¹https://dev.twitter.com/streaming/public

Let our word count vector for the *i*'th hashtag be defined as:

$$H_i^c = \{c_{i1}, c_{i2}, ..., c_{in}\},\$$

where *n* is the number or words in our vocabulary, and c_{ik} represents the frequency of the *k*'th term of the vocabulary w_{ik} in the *i*'th hashtag. We build similarly the representation for the *j*'th video transcript, represented by V_j^c , using the same vocabulary as generated from our hashtags. We compute the *idf* weight of each word over all of the hashtags, and then apply a *tf-idf* weighting scheme on both the transcript representations and hashtags representations. Finally, we define the bag of words scoring function as the cosine similarity between the two *tf-idf* transformed vectors.

$$S_{BoW}(H_i^c, V_j^c) = \cos(tfidf(H_i^c), tfidf(V_j^c)),$$
(1)

4.1.2 Hashtag-Probing

We begin by finding the most relevant k unigram terms from all words associated with a hashtag, H_i^w . To accomplish this, we use *tf-idf* with logarithmic weighting over each term in the vector to represent each hashtag, which can be seen below,

$$log-tfidf(w_{ik}) = \frac{\log(c_{ik})}{(1 + \log(df_{ht}(w_{ik})))},$$
(2)

where $df_{ht}(w_{ik})$ is the number of hashtags this word appears in. We then take the top-k magnitude terms to represent each hashtag. In our experiments we set k to 40, and we represent the 40 probeterms in used for this hashtag as H_i^{pr} .

Finally, we have developed the following hashtag scoring function representing how closely a hashtag relates to a given transcript:

$$S_{pr}(H_i^{pr}, V_j^w) = \sum_{t_x \in V_i^w} \mathcal{I}_{H_i^{pr}}(t_x) \frac{1}{1 + \log(df_{ht}(t_x))}, \quad (3)$$

where V_j^w is the set of words that appear in the j'th video transcript, t_x is a word in the transcript, and $\mathcal{I}_{H_i^{pr}}$ is an indicator function that denotes if t_x appears in H_i^{pr} . We will compare this methodology to the "bag-of-words" approach in Sec. 5.

4.2 Visual modalities

In addition to the use of text analysis detailed above, we are looking into finding similarities between the broadcast news videos and the visual content posted on social media. We rely on three different types of representation: full images, faces and mid-level patterns. We detail in this section how each representation is computed and used to obtain a similarity score between videos and hashtags.

4.2.1 Full image matching

We first are looking for video frames and images posted on social media that may be depicting the same scene. We extract a global descriptor from each image in our dataset, either video frames or social media images. We decided to use the pre-trained *hybrid*-*CNN* model [30], which follows Caffe reference network architecture and has been trained on both the ImageNet and Places datasets, as it is a generic global feature able to cope with the variations of scene that can be depicted in news video and in images posted on social media. The *hybridCNN* feature of an image is extracted from the *fc7* layer and lies in \mathbb{R}^{4096} . The *hybridCNN* features are *l*2-normalized and compared with Euclidean distances.



Figure 1: Example of face matching. Left column: frames of news videos, right column: images from Twitter. Detected matching faces are overlaid as green rectangles.

The *full* similarity $S_F(H_i^F, V_j^F)$ between a hashtag *i* and a video *j* is defined as:

$$S_F(H_i^F, V_j^F) = \sum_{v_j^k \in V_j^F} \sum_{h_i^l \in H_i^F} \mathbf{I}(d(v_j^k, h_i^l) < t_F) \frac{1}{d(v_j^k, h_i^l)},$$
(4)

where t_F is a threshold on the distance set to 0.6 in our experiments, v_j^k is the feature of keyframe *i* in video *j*, h_i^l is the feature from image *l* of a hashtag *i*, V_j^F is the set of features for the keyframes of video *j* and H_i^F is the set of features from images in a hashtag H_i .

4.2.2 Face matching

A topic or event is often heavily related to a small set of key persons. When analyzing the visual content, the most reliable information to identify people are their faces. Face bounding boxes are obtained through the face detector NPD [14] run on all the keyframes of the videos and on all images of hashtags selected for visual matching. Each bounding box is then described by a binary code of 512 bits, enabling fast face matching computation using the hamming distance. We trained a deep network for face classification on the CASIA Webface dataset [27]. We use the architecture defined in [27], with one additional fully connected layer of 512 nodes with sigmoid activation, namely the "hash layer" between the pool5 and fc6 layers. Given a cropped face image using the NPD bounding box, we extract the activations of the hash layer and binarize it by thresholding the output at 0.5. An example of face match obtained is given in Figure 1. The face similarity $S_{fa}(H_i^{fa}, V_i^{fa})$ between a hashtag *i* and a video *j* is defined similarly as Eq. (4)but restricted to the single best match using the hamming distance and setting the *face* acceptance threshold (t_f) to 0.2.

4.2.3 Image Pattern Matching

Visual patterns are discriminative and representative image patches in a given image set. Visual patterns are often used as mid level image feature representation in image retrieval and classification tasks [28, 29]. We follow the pipeline proposed in [12] to discover visual



Figure 2: Example of matched image pairs using the pattern bank. Left column: frames of news videos, right column: images from Twitter. Some faces in the second row were blurred.

patterns in the images associated with each hashtag. Visual patterns are aggregated to form a *pattern bank*. We then detect visual patterns from all images in twitter and all frames in news videos. Each image or frame is represented by the set of visual patterns which are detected in it. The similarity of a twitter hashtag and video is defined as the Jaccard similarity between their detected pattern sets, where V_j^{pa} is the patterns extracted from the j^{th} video, and H_k^{pa} are the patterns extracted from the k^{th} hashtag, see below. Only hashtags with similarity values higher than 0.25 are considered.

$$S_{pa}(V_i^{pa}, H_k^{pa}) = \frac{P_{V_i^{pa}} \cap P_{H_k^{pa}}}{P_{V_i^{pa}} \cup P_{H_k^{pa}}}.$$
(5)

4.3 Multimodal Fusion

Each of our particular matching methods have strength and weaknesses to link hashtags to the video content. Some hashtags may have iconic images but few tweets, hence having sparse text representations that are hard to match, but may match well using the visual modalities. Leveraging a combination of each methods should provide a performance gain over one single approach. We include a normalizing constant for each component so that the each of the scoring functions produce results of similar magnitudes for their best matches. Therefore, we propose the following linear combination of our 5 scoring functions, to obtain the final multi-modal score $S_{mm}(H_k, V_i)$:

$$S_{mm}(H_k, V_i) = \alpha_{BoW} S_{BoW}(H_k^w, V_i^w) + \alpha_{pr} S_{pr}(H_k^{pr}, V_i^w) + \alpha_F S_F(H_k^F, V_i^F) + \alpha_{fa} S_{fa}(H_k^{fa}, V_i^{fa}) + \alpha_{pa} S_{pa}(H_k^{pa}, V_i^{pa}),$$
(6)

where the α_* are normalization weights estimated from all hashtags as $\alpha_{pr} = \frac{1}{\max(S_{pr}(\cdot, V_i^w))}, \ \alpha_{BoW} = \frac{1}{\max(S_{BoW}(\cdot, V_i^w))}, \ \alpha_{pa} = \frac{1}{\max(S_{pa}(\cdot, V_i^{pa}))}, \ \alpha_{fa} = \frac{1}{\max(S_{fa}(\cdot, V_i^{fa}))}, \ \text{and} \ \alpha_F = \frac{0.5}{\max(S_F(\cdot, V_i^F))}$ as the full image matching is the noisiest modality. Note that these weights are video specific.

Table 3: Evaluation results.

Methods	MAP@1	MAP@2	MAP@3	MAP@5
BoW (text)	0.64	0.66	0.63	0.56
Probing (text)	0.80	0.67	0.62	0.50
Full (image)	0.24	0.20	0.16	0.13
Face (image)	0.51	0.39	0.31	0.25
Patterns (image)	0.49	0.39	0.37	0.32
Multimodal	0.89	0.82	0.72	0.64

5. EXPERIMENTS

5.1 Experimental setting

To evaluate the performance of our hashtagging algorithms we have annotated part of our dataset with relevant hashtags for each video. To obtain a very high recall of relevant hashtags for each video, we return all hashtags that are possible candidates according to any of our modalities. For example, we return all hashtags for which a word in the transcript can be found in the set of words of our hashtag probing algorithm. We obtain a large set of possible hashtags for each video to be pruned through manual annotation.

A hashtag is considered "relevant" to a video if the tweets and concepts behind the hashtag are closely related to the video content. We have developed a website for annotation in which the annotators can search tweets and photos from a hashtag, and then can decide if the hashtag is relevant for the given video. We annotated 47 news segments across a variety of different topics, and over 30 different news programs. We use mAP@K as our evaluation metric for the performance of our hashtagging algorithms. We present results for K equal to 1, 2, 3, and 5 as we are only looking for a small set of relevant hashtags.

5.2 Results

We can see the results from our methods in Table. 3. The text based methods have the highest performance for this task, which makes sense as a large portion of news is text-based. Our probing method performs better than the standard BoW approach for the top results in our retrieved hashtags. Whereas, the BoW method begins performing better when comparing hashtags we are not as confident about i.e. that are lower in the retrieved hashtag results. The visual content can also do a reasonable job of finding and predicting hashtags for videos, especially the face matching approach that gives the best visual mAP@1 performance and the pattern based approach which can find important local image patches within hashtags and then find these patches in the news frames.

However, our multimodal method performs the best across each of the different levels of mAP evaluation. The visual and text information are indeed complementary in nature for this task, which leads to improved performance, examples are shown in Table 1. With the high precision of our method we could enable a TV news channel to link their video content to social media discussions.

6. CONCLUSIONS

We have presented a new research challenge of linking television video content to its proper context within social media using hashtags. Our approach demonstrates the value of leveraging the visual and text modalities within both television and social media as they bring disparate but complimentary information for this task. We believe that automatically leveraging hashtags to place traditional media in its social media context is a useful and important challenge. We plan to extend the annotation and release our dataset in the future to foster research on this problem.

7. ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-11-44155 and the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA R&D Contract No. 2014-14071600012. Any opinion, views, findings, and conclusions expressed in this material are those of the authors and should not be interpreted as necessarily representing the views, official policies or endorsements, either expressed or implied, of the National Science Foundation, the ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

8. REFERENCES

- J. Allan, J. G. Carbonell, G. Doddington, J. Yamron, and Y. Yang. Topic detection and tracking pilot study final report. 1998.
- [2] S. Amer-Yahia, S. Anjum, A. Ghenai, A. Siddique, S. Abbar, S. Madden, A. Marcus, and M. El-Haddad. Maqsa: a system for social analytics on news.
- [3] L. Ballan, M. Bertini, T. Uricchio, and A. Del Bimbo. Data-driven approaches for social image and video tagging. *Multimedia Tools and Applications*, 74(4):1443–1468, 2015.
- [4] H. Becker, M. Naaman, and L. Gravano. Beyond trending topics: Real-world event identification on twitter. 2011.
- [5] H.-C. Chang. A new perspective on twitter hashtag use: Diffusion of innovation theory. *Proc. of the American Society* for Information Science and Technology, 47(1):1–4, 2010.
- [6] E. Denton, J. Weston, M. Paluri, L. Bourdev, and R. Fergus. User conditional hashtag prediction for images.
- [7] S. Doan, B.-K. H. Vo, and N. Collier. An analysis of twitter messages in the 2011 tohoku earthquake. In *Electronic Healthcare*, pages 58–66. Springer, 2011.
- [8] F. Godin, V. Slavkovikj, W. De Neve, B. Schrauwen, and R. Van de Walle. Using topic models for twitter hashtag recommendation. In *Proc. of the International Conference on World Wide Web*, pages 593–596, 2013.
- [9] J. Hurlock and M. L. Wilson. Searching twitter: Separating the tweet from the chaff. In *ICWSM*, pages 161–168, 2011.
- [10] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao. Target-dependent twitter sentiment classification.
- [11] B. Jou, H. Li, J. G. Ellis, D. Morozoff-Abegauz, and S.-F. Chang. Structured exploration of who, what, when, and where in heterogeneous multimedia news sources. In *Proc.* of the 21st ACM international conference on Multimedia, pages 357–360. ACM, 2013.
- [12] H. Li, J. G. Ellis, H. Ji, and S.-F. Chang. Event specific multimodal pattern mining for knowledge base construction. In *Proceedings of the 24th ACM international conference on Multimedia*. ACM, 2016.
- [13] H. Li, B. Jou, J. G. Ellis, D. Morozoff, and S.-F. Chang. News rover: Exploring topical structures and serendipity in heterogeneous multimedia news. In *Proc. of the 21st ACM international conference on Multimedia*, pages 449–450. ACM, 2013.
- [14] S. Liao, A. Jain, and S. Li. A fast and accurate unconstrained face detector. *Transactions on Pattern Analysis and Machine Intelligence*, 2014.
- [15] K. Massoudi, M. Tsagkias, M. De Rijke, and W. Weerkamp. Incorporating query expansion and quality indicators in

searching microblog posts. In *Advances in information retrieval*, pages 362–367. Springer, 2011.

- [16] A. Pak and P. Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. In *LREc*, volume 10, pages 1320–1326, 2010.
- [17] S. Petrović, M. Osborne, and V. Lavrenko. Streaming first story detection with application to twitter. In *Human* Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pages 181–189, 2010.
- [18] S. Phuvipadawat and T. Murata. Breaking news detection and tracking in twitter. In Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on, volume 3, pages 120–123. IEEE, 2010.
- [19] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proc. of the 19th international conference on World wide* web, pages 851–860. ACM, 2010.
- [20] J. Sankaranarayanan, H. Samet, B. E. Teitler, M. D. Lieberman, and J. Sperling. Twitterstand: news in tweets. In Proc. of the 17th acm sigspatial international conference on advances in geographic information systems, pages 42–51. ACM, 2009.
- [21] T. A. Small. What the hashtag? a content analysis of canadian politics on twitter. *Information, Communication & Society*, 14(6):872–895, 2011.
- [22] O. Tsur and A. Rappoport. What's in a hashtag?: content based prediction of the spread of ideas in microblogging communities. In *Proc. of the fifth ACM international conference on Web search and data mining*, pages 643–652. ACM, 2012.
- [23] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe. Predicting elections with twitter: What 140 characters reveal about political sentiment. *ICWSM*, 10:178–185, 2010.
- [24] X. Wang, M. S. Gerber, and D. E. Brown. Automatic crime prediction using events extracted from twitter posts. In *Social Computing, Behavioral-Cultural Modeling and Prediction*, pages 231–238. Springer, 2012.
- [25] Y. Yang, J. G. Carbonell, R. D. Brown, T. Pierce, B. T. Archibald, and X. Liu. Learning approaches for detecting and tracking news events. *IEEE Intelligent Systems*, (4):32–43, 1999.
- [26] Y. Yang, T. Pierce, and J. Carbonell. A study of retrospective and on-line event detection. In *Proc. of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 28–36. ACM, 1998.
- [27] D. Yi, Z. Lei, S. Liao, and S. Z. Li. Learning face representation from scratch. arXiv preprint arXiv:1411.7923, 2014.
- [28] J. Yuan, Y. Wu, and M. Yang. Discovery of collocation patterns: from visual words to visual phrases. In *Computer Vision and Pattern Recognition*, 2007. CVPR'07. IEEE Conference on, pages 1–8. IEEE, 2007.
- [29] J. Yuan, M. Yang, and Y. Wu. Mining discriminative co-occurrence patterns for visual recognition. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 2777–2784. IEEE, 2011.
- [30] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva. Learning deep features for scene recognition using places database. In *Advances in neural information processing* systems, pages 487–495, 2014.