INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY **DELHI**

#VisualHashtags
Visual Summarization of Social Media Events using Mid-Level Visual Elements

Sonal Goel (IIIT-Delhi), Sarthak Ahuja (IBM Research, India), A.V. Subramanyam (IIIT-Delhi), Ponnurangam Kumaraguru (IIIT-Delhi)

ACM Multimedia 2017

Introduction

- The data generated on social media sites grows at an increasing rate with more than **36% of tweets containing images**
- We aim to discover **#VisualHashtags**, i.e., meaningful patches that can become the visual analog of a regular text hashtag that Twitter generates
- These the entities which are both **representative and discriminative** to the target dataset besides being human-readable and hence more informative
- Our core novelty –
    - a novel pipeline to summarize images from social media events instead of the conventional method of only identifying key-images to represent the event.
    - Our approach includes a multi-stage filtering process which when coupled with the basic methodology to discover mid-level visual elements leads to an improvement in coverage over existing methods.

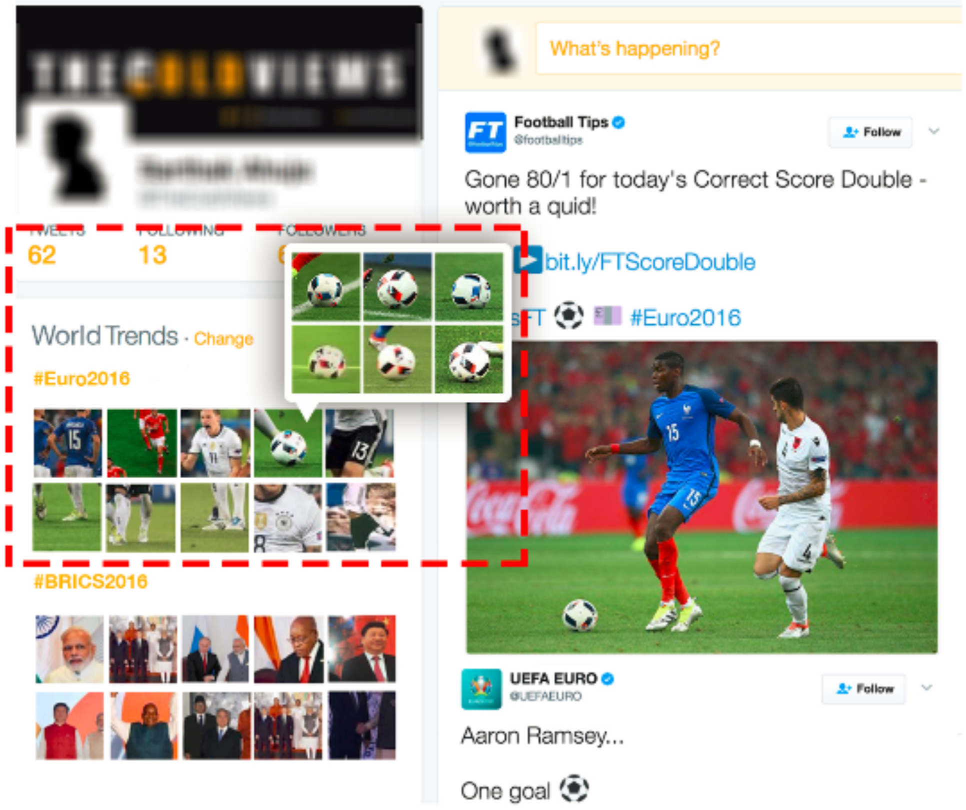


Figure 1: Highlighted in the image is a curated #VisualHashtag for the Euro2016 textual hashtag.

Prior Art

- Work on Social Media Image Dataset Summarization
  - Schinas et al. use both tweets and images to summarize an event. My reveal topics from a set of tweets as highly connected messages in a graph, whose nodes encode messages and whose edges encode their similarities. Finally, the images that best represent the topic are selected based on their relevance and diversity.
  - Cavalin et al. propose a social media imagery analytics system that processes and organize the images in more manageable way by removing duplicate, near-duplicate images and clustering images having similar content
- Work on Image Dataset Summarization at Patch Level
  - **Doersch et al.** collect data from Google Street View of different cities, and aim to automatically and the visual patches like windows, balconies, and street signs, that are most distinctive for a certain geo-spatial area.
  - Rematas et al. propose data-mining approach for exploring image collections by interesting patterns that use discriminative patches and further show the results on Pascal VOC and Microsoft COCO datasets.

Methodology

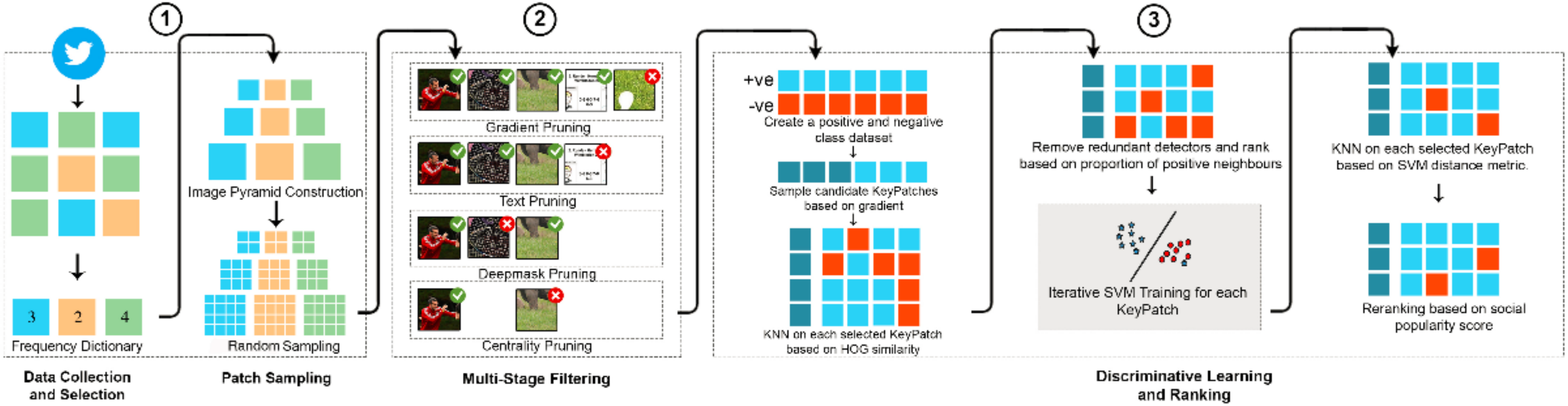


Figure 2: Overall flow of the approach. (1) First collect images from Twitter and remove duplicate images, use these images for patch sampling; (2) Apply multi-stage filtering to prune noisy and non-informative patches; (3) Apply discriminative learning to discover mid-level visual elements and rank them using social popularity score to finally present #VisualHashtag.

Algorithm

Algorithm 1 Discriminative Learning

$D1, D_2, D_3 \ldots D_n$ ▷ Set of n Detectors
$K_1, K_2, K_3 \ldots K_n$ ▷ Clusters of nearest neighbors
$\mathbf{P} = P_1, P_2, P_3 \ldots P_n$ ▷ Positive dataset divided into l parts
$\mathbf{N} = N_1, N_2, N_3 \ldots N_l$ ▷ Negative dataset divided into l parts
for $i = 1$ to n **do**
    K_i = HogBasedKNN(D_i, P_1, k)
end for
for $i = 1$ to l **do**
    P^* = ChooseWithoutReplacement($\mathbf{P}$)
    N^* = ChooseWithoutReplacement($\mathbf{N}$)
    for $j = 1$ to n **do**
        SVM_j = trainSVM(D_j, K_j, N^*)
        K_j = [K_j, topSVMDetections(SVM_j, P^*, k)]
    end for
end for
for $j = 1$ to n **do**
    $Score_j$ = score(D_j,K_j)
end for

$$score_i = \sum_{j=1}^{n} (-1)^c S_j \frac{n-j+1}{n}$$

- **score$_i$** is the score of the detector
- **n** is the number of nearest neighbors,
- **c** is the class of the nearest neighbor (1 for negative, 0 for positive),
- **S$_j$** is the frequency (number of duplicates) of the image to which the patch belongs.

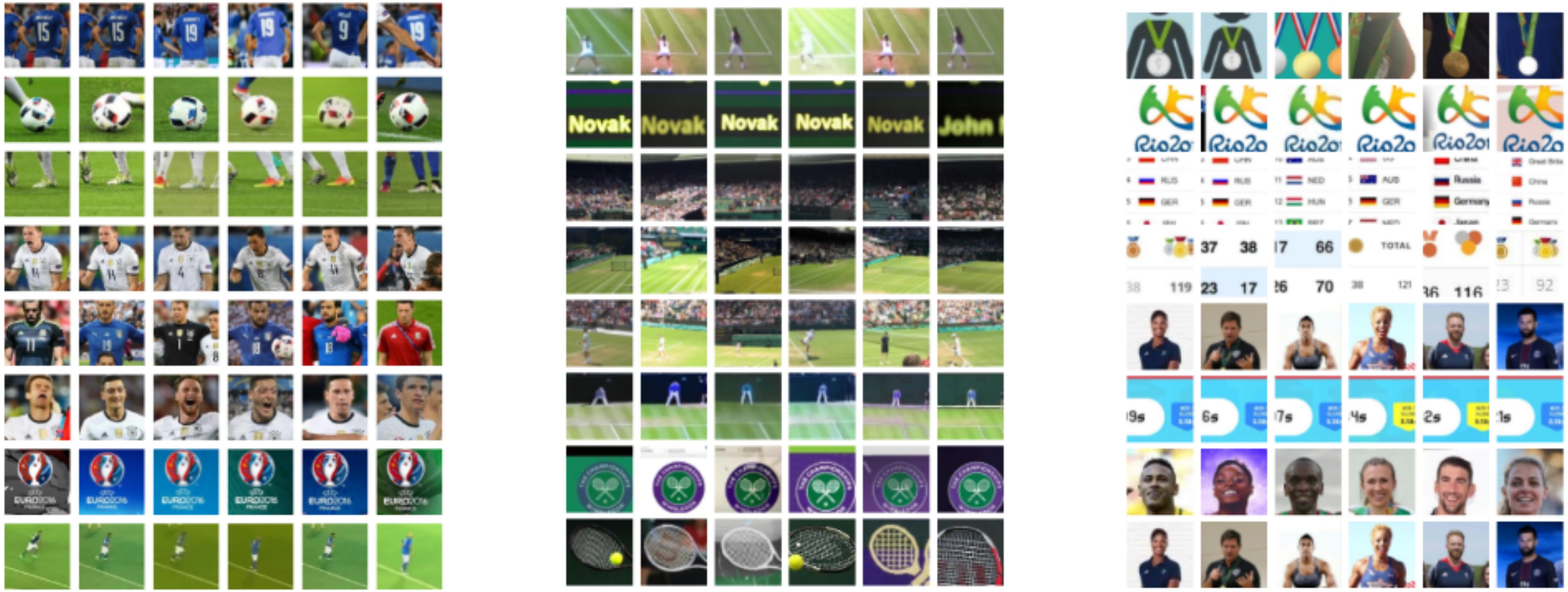
Dataset

Table 1: Details of Data Collection.

Event	Total Images	Unique Images	Category
EuroCup	3,489	827	Sports
Wimbeldon	3,229	1,327	Sports
Olympics	2,264	1,968	Sports
BREXIT	3,728	1212	Politics
BRICS	4,618	1,102	Politics
UNGA	2,756	1,572	Politics
US-Elections	98,813	5,000	Before-Election
US-Elections	218,289	5,000	After-Election

- The data generated on social media sites grows at an increasing rate with more than **36% of tweets containing images**
- We aim to discover **#VisualHashtags**, i.e., meaningful patches that can become the visual analog of a regular text hashtag that Twitter generates
- These the entities which are both **representative and discriminative** to the target dataset besides being human-readable and hence more informative

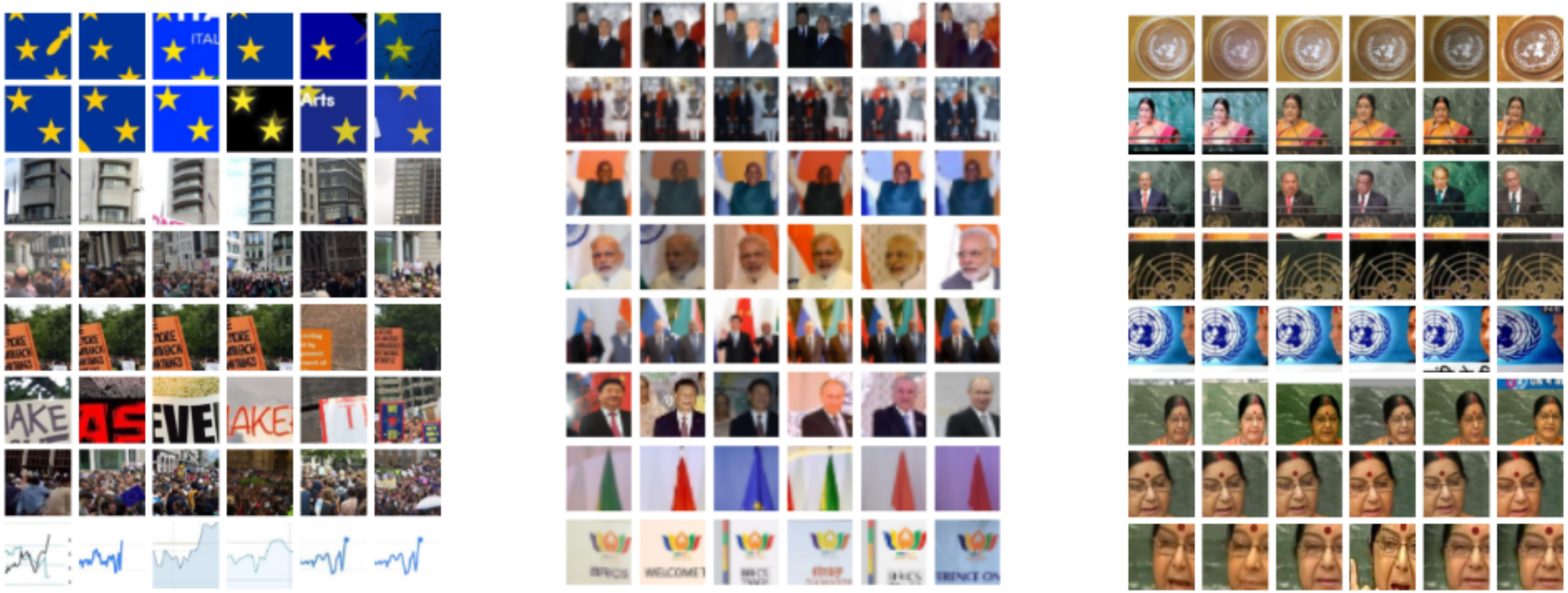
Results – 1 (Sports)



(a) #VisualHashtags for EuroCup (b) #VisualHashtags for Wimbledon (c) #VisualHashtags for Olympics

Figure 3: Summarizing sports events for (a) EuroCup (b) Wimbledon (c) Olympics.

Results – 2 (Politics)



(a) #VisualHashtags for BREXIT (b) #VisualHashtags for BRICS (c) #VisualHashtags for UNGA

Figure 4: Summarizing politics events for (a) BREXIT (b) BRICS (c) UNGA.

Results – 3 (Temporal Analysis)

(a) Summarizing US Election2016 before the election day

(b) Summarizing US Election2016 after the election day

Figure 5: Analysing the visual elements dominating the dataset before and after elections.

Results – 3 (Pattern Mining)

$$N = C(\,Similarity\,(I_{xi}, I_{yj})\,), \quad \forall\,(i,j) \qquad (2)$$

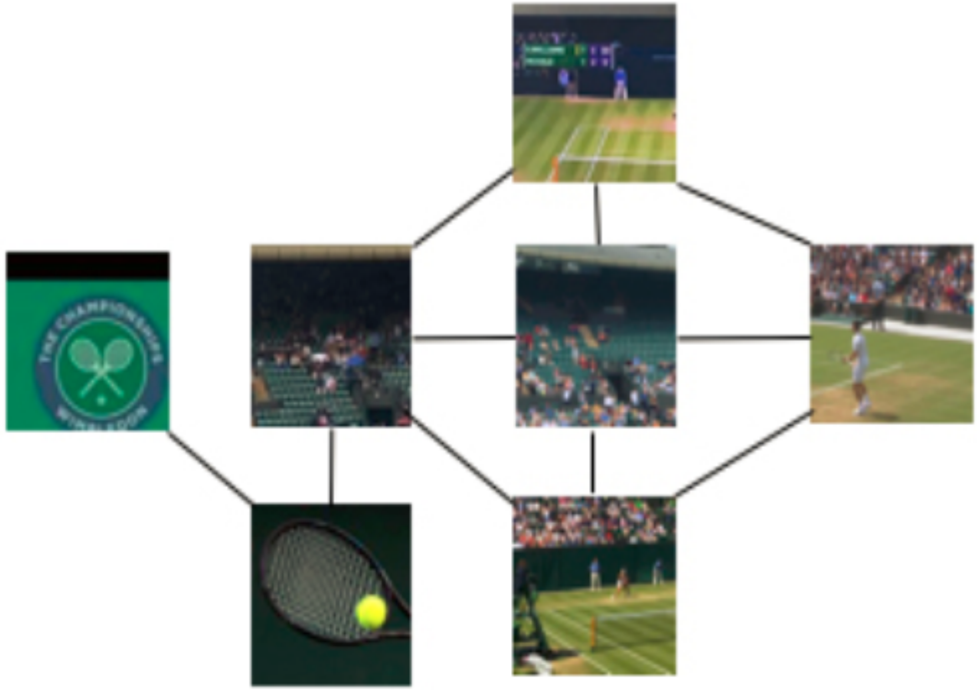


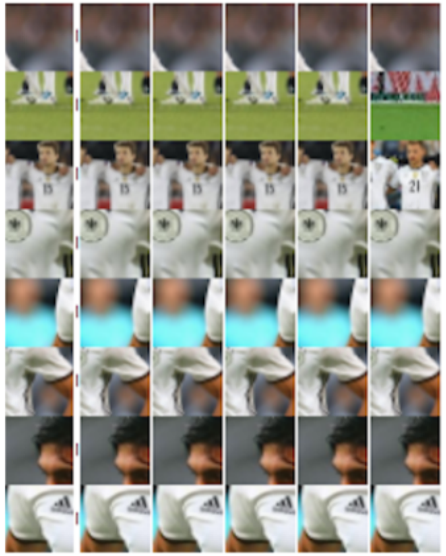
Figure 6: Graph showing connections between different patches signifying their co-occurrence in images.

Evaluation - Quantitative

Table 2: Number of patches selected at each filtering stage, and percentage of noisy patches pruned at the end.

Event	Initial Patches	Grad. Pruned	Text Pruned	Deepmask Pruned	Centr. Pruned	Reduce %
Euro	20,635	19,428	19,237	14,607	11,917	42%
Wimb	33,151	31,222	30,604	21,983	17,140	48%
Olymp	49,176	46,312	43,669	32,727	26,211	47%
BREXIT	30,264	28,828	27,677	17,544	13,937	54%
BRICS	27,526	26,363	25,124	18,431	14,821	46%
UNGA	39,276	38,108	36,229	24,021	19,849	49%

(a) Summarizing EuroCup using (FILT_DISC)



(b) Summarizing EuroCup using DISC

Table 5: Percentage of purity and coverage of the results with different nearest-neighbors for DISC and FILT_DISC.

Approach	P@5	C@5	P@10	C@10	P@15	C@15	P@20	C@20	P@25	C@25
DISC	100.0	0.77	100.0	1.11	94.3	1.37	85.5	2.03	79.0	2.57
FILT_DISC	93.0	10.17	88.0	20.80	77.0	22.72	68.0	24.41	67.0	25.36

Evaluation - Qualitative

User Study – 21 People

- Given the set of patches below, choose the most appropriate event which it summarizes.
- Select all the patches that can be distinctly linked with the event chosen above.
- Select all the patches that are Meaningful ,i.e. covering a meaningful part of an image
- How many of the below rows demonstrate strong correlation (containing similar elements like faces/buildings, etc) among their elements?

Metrics

(1) **Precision (Pr@N):** percentage of patches among the top N that are relevant/meaningful to the corresponding event, averaged
among all events.
(2) **Success (S@N@D):** percentage of responses, where there exist at least D relevant/meaningful patches amongst the top N.
(3) **Mean Reciprocal Rank (MRR):** Computed as 1/r, where r is the rank of the first relevant/meaningful patch returned, averaged over all events.

Table 4: Precision, Success, MRR, Mean and Std. Dev., based on the qualitative analysis of the summarized events.

Quality-measures	Pr@1	Pr@5	Pr@10	S@10@1	S@10@3	S@10@5	MRR	Mean	S.Dev	Category
Relevance (Rel.)	0.94	0.57	0.45	1.0	0.75	0.48	0.95	4.5	2.4	Sports
Meaningfulness (Mea.)	0.95	0.58	0.44	1.0	0.79	0.43	0.97	4.3	2.0	Sports
Intersection (Rel.+ Mea.)	0.90	0.46	0.32	0.95	0.60	0.27	0.50	3.3	1.9	Sports
Relevance (Rel.)	0.79	0.54	0.41	1.0	0.57	0.37	0.87	4.1	2.6	Politics
Meaningfulness (Mea.)	0.79	0.58	0.41	1.0	0.71	0.38	0.87	4.1	2.1	Politics
Intersection (Rel.+ Mea.)	0.75	0.44	0.29	0.97	0.46	0.17	0.52	2.9	1.9	Politics

Future Work

- Currently, our approach is centered around events that contain images with relative stylistic coherence and uniqueness, and thus #VisualHashtags generated also focus on concrete entities.
- As future work, the technique can be modified to summarize more abstract phenomenon like violence, summer, etc.
- Mid-level patches obtained as a summary of a particular viral event, can be further generalized to pave way for finding higher-level image features that can cover the essence of an event.
- While the current approach needs to be re-run to generate #VisualHashtags at different time instances, dynamic re-summarization would be an interesting direction to explore, making it a more real-time system.