

TRECVID 2018 Linking Task: Social Media Video Storytelling

João Magalhães
NOVALINCS

Universidade NOVA Lisboa, Portugal
jm.magalhaes@fct.unl.pt

David Semedo
NOVALINCS

Universidade NOVA Lisboa, Portugal
df.semedo@campus.fct.unl.pt

Saverio Blasi
BBC R&D

London, Wood Lane, UK
saverio.blasi@bbc.co.uk

ABSTRACT

The social-media video storytelling linking task seeks to advance the area of visual summarization, with collaborative videos, images and texts available from professional media and social-media users. The goal is to illustrate a news story with social-media content. Starting from a news story topic and a stream of social-media video and images, the goal is to link a story-segment to image and video material, while also preserving a good flow of the whole visual story.

1 INTRODUCTION

When an event occurs in real-life, there may be little information available at first. However, as media and textual descriptions about the event are published in social-media, they form an evolving story, which is of critical value for the affected or interested persons. This evolving story corresponds to a semantic topic that can be tracked over time with both textual and visual information from multiple social-media sources (i.e., end-users or online services). Moreover, as social-media sources continue to publish information about the story, it becomes critical to select the most relevant information. Thus, based on all collaborative audio-visual and textual information, one can create summaries or stories of those real-world events.

The timeline of an event, e.g. a music festival, a sport tournament or a natural disaster [2], contains visual and textual pieces of information that are strongly correlated. There are several ways of presenting the same event, by covering specific *storylines*, thus offering different perspectives about that same event. These storylines, illustrated in Figure 1, refer to a story topic and related subtopics, and are structured into *story segments* that should describe narrow occurrences over the course of the event. More formally, we define a *Visual Storyline* as a sequence of segments, referring to an event topic, with each segment being defined by a textual description and comprising an image or a video.

2 TASK

The goal is to illustrate a news story with social-media visual content. Starting from a news story topic and a stream of social-media video and images, the goal is to link a story-segment to image and video material, while also preserving a good flow of the whole visual story. A news story topic is an actual news narrative and the news segments correspond to particular sentences of the news, that a journalist may wish to illustrate. For each story segment (a sentence query with a strong visual component), systems should propose the *single video or image* that satisfy the two requirements:

- Best illustrates the news segment;
- Makes the best transition from the previous video/image illustration.

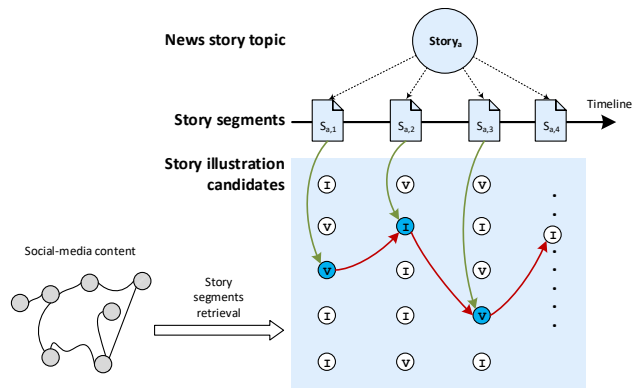


Figure 1: Computation of social visual stories is greatly influenced by story transitions.

In this task, a visual storyline is composed of a set of images/videos organised in a sequence to provide a cohesive narrative. Tackling the task of illustrating a storyline means taking into account not only the relevance of the individual pieces of content, but also the way they transition from one to the other, Figure 1. As such, assuring the quality and meaningfulness of these transitions is an important component of the editing process.

2.1 Data

To enable social media visual storyline illustration, a data collection strategy was designed to create a suitable corpora, limiting the number of retrieved documents to those posted during the span of the event. Events adequate for storytelling were selected, namely those with strong social-dynamics in terms of temporal variations with respect to their semantics (textual vocabulary and visual content). In other words, the unfolding of the event stories is encoded in each collection. Events that span over multiple days like music festivals, sports competitions, etc., are examples of good candidates of storylines. Taking the aforementioned aspects into account, the data for the following events was crawled (Table 2):

The Edinburgh Festival (EdFest) consists of a celebration of the performing arts, gathering dance, opera, music and theatre performers from all over the world. The event takes place in Edinburgh, Scotland and has a duration of 3 weeks in August.

Le Tour de France (TDF) is one of the main road cycling race competitions. The event takes place in France (16 days), Spain (1 day), Andorra (3 days) and Switzerland (3 days).

The *keyword-based* approach, consists of querying the social media APIs with a set of keyword terms. Thus, a curated list of keywords was manually selected for each event. Furthermore, hashtags in social media play the essential role of grouping similar content

Table 1: Development data covers the 2016 editions (relevance judgments available).

| Event | Stories | Docs | Docs w/images | Docs w/videos | Crawling span | Crawling seeds |
|------------|---------|--------|-----------------|----------------|---------------------------------------|---|
| EdFest2016 | 20 | 34,297 | Twitter: 29,558 | Twitter: 4,739 | From: 2016-07-01 Until: 2017-01-01 | Terms Hashtags Edinburgh Festival, Edfest, Edinburgh Festival 2016, Edfest 2016 #edfest, #edfringe, #EdinburghFestival, #edinburghfest |
| TDF2016 | 20 | 75,385 | Twitter: 67,032 | Twitter: 8,353 | From: 2016-06-01 Until: 2017-01-01 | Terms Hashtags le tour de france, le tour de france 2016, tour de france #TDF2016, #TDF |

Table 2: Test data covers the 2017 event editions (no relevance judgments available).

| Event | Stories | Docs | Docs w/images | Docs w/videos | Crawling span | Crawling seeds |
|------------|---------|--------|-----------------|----------------|---------------------------------------|---|
| EdFest2017 | 15 | 39,022 | Twitter: 34,302 | Twitter: 4,720 | From: 2017-07-01 Until: 2017-10-19 | Terms Hashtags Edinburgh Festival, Edfest, Edinburgh Festival 2017, Edfest 2017 #edfest, #edfringe, #EdinburghFestival, #edinburghfest, #BBCedfest, #EdinburghFringe, #edinburghfringefestival |
| TDF2017 | 15 | 69,089 | Twitter: 59,534 | Twitter: 9,555 | From: 2017-07-01 Until: 2017-10-19 | Terms Hashtags le tour de france, le tour de france 2017, tour de france #TDF2017, #TDF, #TourdeFrance |

(e.g. content belonging to the same event) [1]. Therefore a set of relevant hashtags grouping content of the same topic was also manually defined. The data collected is detailed in Table 2.

2.1.1 Development data. The development data covers the 2016 editions of the above events and for each event there’s 20 stories. Several stories were generated with simple baselines and evaluated with crowd-sourcing. Three annotators were presented with each story title, and asked to rate each segment illustration as relevant or non-relevant, as well as rate the transitions between each of the segments. Finally, using the subjective assessment of the annotators, the score proposed in Section 2.3 was calculated for each story.

For each visual storyline, annotators were asked to rate the transitions between each sequential pair of images with a score of 0 (“bad”), 1 (“acceptable”) or 2 (“good”); they were also asked to rate the story quality on 1 to 5 scale.

2.1.2 Test data. The test data covers the 2017 editions of the above events and for each event there’s 15 stories. The topics are available for download, but the ground truth will only be available after submissions.

2.2 Story topics

For the identification of event storylines, along with a focused crawling of social-media data about particular events, a set of professional news¹ stories covering these same events was also collected. Two requirements were established regarding the identified storylines: general interestingness, i.e. news worthy and/or informative storylines, and availability of enough relevant supporting documents and media elements on the collected data.

In this context, three types of storylines were identified:

Published News. To identify reliable storylines we collected news articles from well-reputed news websites (e.g. BBC News, Reuters, etc.). Specific sentences were marked as story segments that need to be illustrated by an image or a video.

Investigative topics. When journalists edit news stories, the common practice is to investigate online media as well as additional information that supports a story. These storylines were constructed taking this workflow into account.

Review topics. Often, media professionals (e.g. journalists) need to cover several independent events, e.g. as music festivals spanning several days, and edit a news video that provides a summary of the highlights. Hence, these storylines correspond to the coverage of an event through a summary of the highlights.

Due to their interestingness and informational value, stories in these categories are naturally present in the data sources, hence they are typically well supported with enough relevant information.

2.3 Evaluation metric

Media editors are constantly judging the quality of news material to decide if it deserves being published. The task is highly skillful and deriving a methodology from such process is not straightforward. The task of identifying visual material suitable to describe each story segment is, from the perspective of media professionals, highly subjective. The motivation for why some content may be used to illustrate specific segments can derive from a variety of factors. While subjective preference obviously plays a part in this process (which cannot be replicated by an automated process), other factors are also important which come from common practice and general guidelines, and which can be mimicked by objective quality assessment metrics.

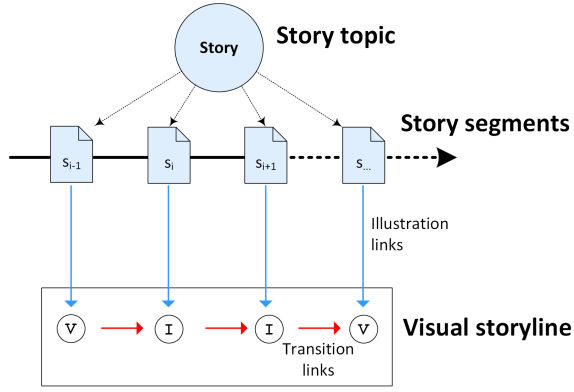
The first step towards the quantification of visual storyline quality concerns the human-judgement of these different dimensions. This is achieved in a sound manner by judging specific objective characteristics of the story – Figure 2 illustrates the visual storyline quality assessment framework. In particular, storyline illustrations are assessed in terms of *relevance of illustrations* (blue links in Figure 2) and *coherence of transitions* (red links in Figure 2). Once a visual storyline is generated, annotators will judge the relevance of the illustration to the story segment as:

- $s_i=0$: the image/video is not relevant to the story segment;
- $s_i=1$: the image/video is relevant to the story segment;
- $s_i=2$: the image/video is highly relevant to the story segment.

Similarly with respect to the *coherence* of a visual storyline, each story transition is judged by annotators as the degree of affinity between pairs of story segment illustrations:

- $t_i=0$: there is no relation between the segment illustrations;

¹We collected news from BBC, Guardian and Reuters.



(a) Visual story editing assessment framework.

$$\begin{array}{c}
 s_{i-1} \downarrow \quad s_i \downarrow \\
 \textcircled{v} \xrightarrow{t_i} \textcircled{i} \quad \text{pairwise}Q(i) = \underbrace{\beta (s_{i-1} + s_i)}_{\text{segments illustration}} + \underbrace{(1 - \beta) (s_{i-1} s_i + t_i)}_{\text{transitions}}
 \end{array}$$

(b) Visual story quality assessment metric.

Figure 2: Methodology for evaluating visual storyline illustration.

- $t_i=1$: there is a relation between the two segments;
- $t_i=2$: there is an appealing semantic and visual coherence between the two segment illustrations.

These two dimensions can be used to obtain an overall expression of the "quality" of a given illustration for a story of N segments. This is formalised by the expression:

$$\text{Quality} = \alpha \cdot s_1 + \frac{(1 - \alpha)}{2(N - 1)} \sum_{i=2}^N \text{pairwise}Q(i) \quad (1)$$

The function $\text{pairwise}Q(i)$ defines quantitatively the perceived quality of two neighbouring segment illustrations based on their relevance and transition:

$$\text{pairwise}Q(i) = \underbrace{\beta \cdot (s_i + s_{i-1})}_{\text{segments illustration}} + \underbrace{(1 - \beta) \cdot (s_{i-1} \cdot s_i + t_{i-1})}_{\text{transition}} \quad (2)$$

where α weights the importance of the first segment, and β weights the trade-off between *relevance of segment illustrations* and *coherence of transitions* towards the overall quality of the story.

Given the underlying subjectivity of the task, the values of α or β that optimally represents the human perception of visual stories, are in fact average values. Nevertheless, we posit the following two reasonable criteria: (i) illustrating with non-relevant elements ($s_i = 0$) completely breaks the story perception and should be penalised. Thus, we consider values of $\beta > 0.5$; and (ii) the first image/video perceived is assumed to be more important, as it should grab the attention towards consuming the rest of the story. Thus, α is a boost to the first story segment s_1 . It was empirically found that $\alpha = 0.1$ and $\beta = 0.6$ adequately represent human perception of visual stories editing.

3 SUBMISSIONS

The following figure illustrates what is supposed to be submitted: the video segment links and the video transition links are the key elements that make the story summary. From a submission point of view, teams will have to consider the following:

- There will be a total of 30 story topics, organised into 3-5 visual story segments each.
- Each segment is story segment is either (1) a text description provided by the media editor to be used as query or (2) media + text that is fixed by the media editor. In both situations transitions are an important part.
- Each run is composed of a sequence of videos/images intended to illustrate the sequence of story segments.
- Teams may submit up to 5 runs. Hence, each team can create up to 5 alternative visual summaries for each story. This totals 150 visual stories.

Runs submission:

Format - Submissions format is detailed in file `info_queries+evaluation.txt` (available in the Github repository²);

Evaluation - We provide a script for validating and evaluating submissions on the task repository.

Submissions - For each event dataset, a run submission file should be created. In other words, assignments of images/videos to segments of the an event, should be placed in the same submission file.

REFERENCES

- [1] David Laniado and Peter Mika. 2010. Making Sense of Twitter. In *Proceedings of the 9th International Semantic Web Conference on The Semantic Web - Volume Part I (ISWC'10)*. Springer-Verlag, Berlin, Heidelberg, 470–485.
- [2] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. 2010. Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors. In *Proceedings of the 19th International Conference on World Wide Web (WWW '10)*. 10.

²<https://github.com/davidfsemedo/TRECVID-VisualStories>