# Joint Photo Stream and Blog Post Summarization and Exploration

Gunhee Kim[1], Seungwhan Moon[2] Leonid Sigal[3]
[1]Seoul National University. [2]Carnegie Mellon University. [3]Disney Research Pittsburgh.

**Research Objective**. We propose to take advantage of large collections of photo streams and blog posts in a *mutually-beneficial* way for story-based summarization and exploration (see Fig.1 for an example). Blogs usually consist of sequences of images and associated text; they form a *storytelling* narrative, by digesting key events and expressing them with concise sentences and representative images. Thus, blog posts can help achieve a *story-based* semantic summarization of large-scale and ever-growing sets of photo streams that are often unstructured and associated with missing or inaccurate semantic labels. In the reverse direction, each blog benefits from a large set of photo streams, which can interpolate various photo paths between consecutive images in the blog. Each blog is written based on a single person's experience with a small number of selected images. The photo-path interpolation, achieved with a collection of photo streams, allows blog authors to explore alternative paths by other visitors who follow a similar itinerary.

To implement joint summarization and exploration, we first collect a large set of photo streams and blog posts for an event of interest. We then jointly perform the two base tasks so that they help each other: (i) *alignment* between the images of blogs and photo streams, and (ii) *summarization* of photo streams. The alignment task discovers correspondences between the blog photos and the images in the photo streams. The summarization task selects the most important images from photo streams with maximal coverage and minimal redundancy. Since blog photos selected by users are more semantically meaningful, we encourage photo stream summaries to have the images that are closer matches to blog photos. In the reverse, summaries of photo streams can make the alignment task faster and more focused.

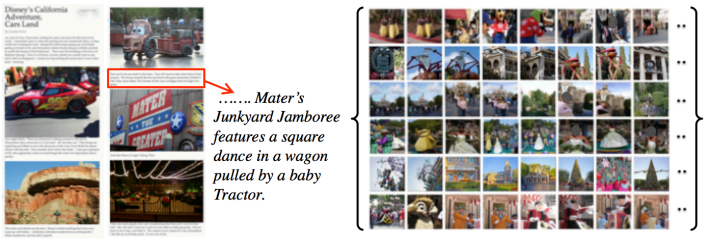
**Approach**. The input of our approach is two-fold: a set of photo streams $\mathcal{P} = \{P^1, \cdots, P^L\}$ and a set of blog posts $\mathcal{B} = \{B^1, \cdots, B^N\}$ for a topic of interest (*e.g.* trips to Disneyland). Each photo stream is a set of images taken in sequence by a single user in a single day, while each blog post comprises a sequence of pairs of images and text blocks.

For joint exploration between blogs and photo streams, we solve the two subproblems: (i) alignment from blog images to photo streams, and (ii) summarization of photo streams. The alignment is achieved by building a bipartite image graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the vertex set consists of images of blogs and photo streams (*i.e.* $\mathcal{V} = \mathcal{I} \cup \mathcal{P}$), and the edge set $\mathcal{E}$ represents correspondences between them. We denote the adjacency matrix by $\mathbf{W} \in \mathbb{R}^{|\mathcal{I}| \times |\mathcal{P}|}$ where $|\mathcal{P}|$ is the number of images in all photo streams. Thus, the goal of alignment reduces to an estimate of $\mathbf{W}$. On the other hand, summarization aims to predict the best subset $S^l \subset P^l$ for each photo stream $P^l \in \mathcal{P}$. We use $\mathcal{S}$ to denote a set of summaries $\mathcal{S} = \{S^1, \cdots, S^L\}$.

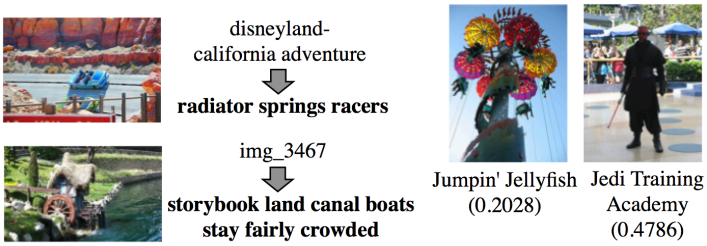
Once we formalize the objectives and constraints for alignment and summarization, we formulate the two base tasks as two sets of ranking SVM problems with latent variables (*e.g.*, [1]). We alternate between solving the two coupled latent SVM problems, by first fixing the summarization and solving for alignment and vice versa. The major strength of our formulation lies in its flexibility; one can easily add additional constraints, while using the exact same formulation and optimization, which can be efficiently solved via stochastic gradient descent.

**Experiments**. We crawl the *Disneyland* dataset, consisting of about 540K images of 6K photo streams from FLICKR and 10K blog posts with 120K associated images from BLOGGER and WORDPRESS. Although we mainly discuss the proposed approach in the context of *Disneyland*, our approach can be extended to any problem domain, with little modification, because our NLP pre-processing (*e.g.* keyword extraction) is unsupervised.

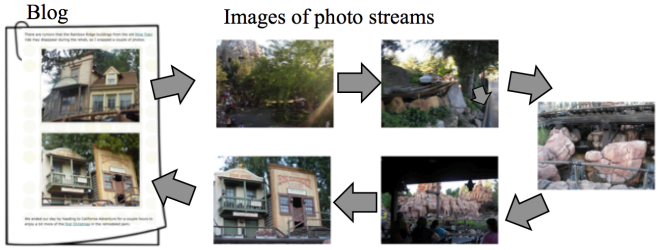
In the experiments, we focus on showing that blog posts and photo streams are indeed mutually beneficial. First, we demonstrate the usefulness of blogs toward photo streams with two tasks of semantic knowledge transfer. We show that blog posts improve the image localization accuracy




(a) Input: large sets of blog posts and photo streams




(b) Blogs transfer semantic knowledge to photo streams (title, location)




(c) Photo streams help interpolation between blog images

Figure 1: Motivation for joint summarization and exploration between large collections of photo streams and blog posts. (a) The input is two-fold: a set of photo streams and blog posts from *Disneyland*, which are captured by multiple users and at different times. (b) Blogs benefit photo stream summarization by transferring semantic knowledge. Examples show *automatic image titling* and attraction-based *image localization*. (c) Photo streams can enhance blog posts by allowing interpolation between consecutive blog images. Two blog images of an attraction entrance used as a query, result in an illustration of what happens inside an attraction.

(*i.e.* finding where photos were taken), and automatic image titling (*i.e.* creating descriptive titles for images). Second, we show that a large set of photo streams leads to better path interpolation between consecutive blog images. We quantitatively evaluate the performance of our approach for path interpolation via crowdsourcing, using Amazon Mechanical Turk.

**Contributions**. (1) To the best of our knowledge, our work is unique in jointly leveraging large sets of blog posts and photo streams for mutually-beneficial summarization and exploration. We show that blogs are useful for story-based summary of photo streams along with semantic knowledge transfer. At the same time, a large set of photo streams help interpolate plausible image paths between any two consecutive blog photos.

(2) We propose an approach for jointly solving alignment and summarization tasks in a unified ranking SVM framework. We alternatingly solve one of the two problems while conditioning on the solution of the other.

(3) For evaluation, we collect *Disneyland* dataset, consisting of 10K blog posts with 120K associated images, and 6K photo streams of 540K images. We demonstrate that blog posts and photo streams indeed help each other for summarization and exploration.

[1] Thorsten Joachims, Thomas Finley, and Chun-Nam John Yu. Cutting-Plane Training of Structural SVMs. *Mach Learn*, 77:27–59, 2009.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.