Connecting the Dots Using Contextual Information Hidden in Text and Images

Md Abdul Kader,[1] Sheikh Motahar Naim,[1] Arnold P. Boedihardjo,[2] M. Shahriar Hossain[1]
[1]The University of Texas at El Paso, El Paso, TX 79968, Phone: (915) 747-5000
[2]U. S. Army Corps of Engineers, Alexandria, VA 22315
{mkader, snaim}@miners.utep.edu, arnold.p.boedihardjo@usace.army.mil, mhossain@utep.edu

1 Introduction

The publicly available data feeds are increasing exponentially providing a massive source of intelligence, ironically this plethora of information is what makes succinct details hidden during the analysis of an event of interest. Traditional search engines help narrow down the scope of analysis to a specific event but it is not an easy task for an analyst to study massive amount of relevant documents and get a complete understanding of how each sub-event compose a complex set of interactions between the actors involved in a major event. Here, we describe a framework called *Storyboarding* that provides summarizations of an event as chains of coherent documents and relevant image objects using publicly available data. Summarization of an event involves the task of identifying relevant entities (e.g., person and location), discovering non-obvious connections between the entities to construct a coherent story – a task which sometimes is referred in the literature as "connecting the dots", and providing relevant information (e.g., images) to explain the connections better. Our approach provides a mechanism to build a story between two news articles published at two different times using a sequence of intermediate published articles.

The storyboarding framework builds a knowledge base first, which helps in providing an image context for each story. A straightforward way to incorporating images during summarization of events is to generate the story using only a similarity network of news articles, as done in Storytelling (Hossain et al. 2012), and then attach images to enrich the presentation of the story. This straightforward approach does not take contents of the images into account whereas the images are likely to carry vital information for a story-building process. In our Storyboarding framework, we leverage image objects (e.g., faces) as added pieces of information to the text content. As an example of our Storyboarding output, Figure 1 and Section 3.3 explain a sub-event of the Boston Marathon Bombing tragedy.

The main contributions of the storyboarding framework are (1) a **knowledge base** that connects textual information with facial features using a probabilistic technique, (2) a **Frontalization** mechanism that enhances face feature extraction by complementing conventionally used techniques, and (3) **summarization** of events that leverages text and images to explain events as chains of documents over time giving an idea about how the story evolved.

2 Methodology

The Storyboarding framework has three operational stages that are described in the following subsections.

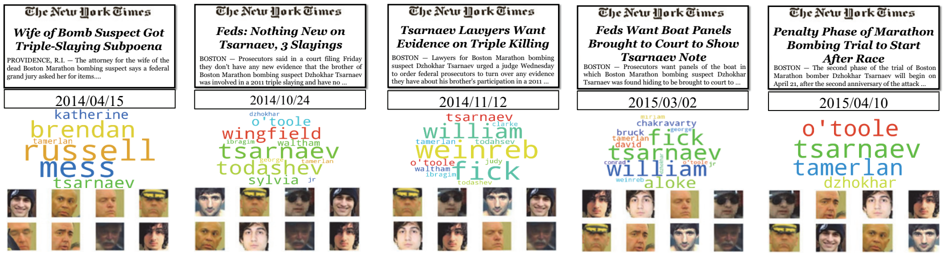
2.1 Knowledge Base, κ

The knowledge base $\kappa = \{\mathcal{D}, \mathcal{E}^\kappa, \mathcal{F}\}$ is constructed using images and textual context of Wikipedia pages that are related to politics and terrorism. $\mathcal{D}$, $\mathcal{E}^\kappa$, and $\mathcal{F}$ are the set of text content of pages in Wikipedia, the set of person entities in $\mathcal{D}$, and the set of faces extracted from the images of Wikipedia pages respectively. We represent the person entities using vector space modeling: the weight $W(e, d)$ of entity $e \in \mathcal{E}^\kappa$ in the document $d \in \mathcal{D}$ is a variant of tf-idf modeling with cosine normalization. To detect faces from images we used Viola-Jones algorithm (Viola and Jones 2001) with *trained cascade classification* model. Since many of the faces are side-facing, we use a frontalization method to be able to capture positions of some key facial points in a projected plane where a side-faced photo represents a corresponding front-posing face. We detect five facial key points – two eye centers, nose and two mouth corners – using a pretrained convolutional neural network (Sun, Wang, and Tang 2013). The detected facial key points are then frontalized. To extract facial features of length L for each face, we combine Eigenface with all possible angles between every three of five frontalized points and relative distances between every pair of frontalized points.

2.2 Context Modeling

The context of a face $f \in \mathcal{F}$ is a ranked list of person entities ordered by relevance to f, and expressed as a probability distribution over the set of entities $\mathcal{E}^\kappa$. The contexts of the faces are leveraged to generate chains of contextual documents in Section 2.3. The probability of an entity $e \in \mathcal{E}^\kappa$ for the given face f is calculated as

$$\log(P(e|f)) \propto \log(P(e)) - \sum_{l=1}^{L} \left(f^l \times \log\left(\sum_{d \in D^e} W(e,d) \right) \right) + \sum_{l=1}^{L} \left(f^l \times \log\left(\sum_{d \in D^e} P(f^l|d) \times W(e,d) \right) \right) \quad (1)$$



The New York Times

Wife of Bomb Suspect Got Triple-Slaying Subpoena

PROVIDENCE, R.I. — The attorney for the wife of the dead Boston Marathon bombing suspect says a federal grand jury asked her for items....

The New York Times

Feds: Nothing New on Tsarnaev, 3 Slayings

BOSTON — Prosecutors said in a court filing Friday they don't have any new evidence that the brother of Boston Marathon bombing suspect Dzhokhar Tsarnaev was involved in a 2011 triple slaying and have no ...

The New York Times

Tsarnaev Lawyers Want Evidence on Triple Killing

BOSTON — Lawyers for Boston Marathon bombing suspect Dzhokhar Tsarnaev urged a judge Wednesday to order federal prosecutors to turn over any evidence they have about his brother's participation in a 2011 ...

The New York Times

Feds Want Boat Panels Brought to Court to Show Tsarnaev Note

BOSTON — Prosecutors want panels of the boat in which Boston Marathon bombing suspect Dzhokhar Tsarnaev was found hiding to be brought to court to ...

The New York Times

Penalty Phase of Marathon Bombing Trial to Start After Race

BOSTON — The second phase of the trial of Boston Marathon bomber Dzhokhar Tsarnaev will begin on April 21, after the second anniversary of the attack ...

Figure 1: The story contains five news articles associating Boston bombers' involvement in the Waltham triple murder.

In practice, we keep record of maximum of L_C of entities with the highest probabilities as the context of a face.

2.3 Automatic Summarization as Stories

The core storyboarding task focuses on forming a chain of $\{$article, face-set$\}$ pairs using a subset of news articles from a news corpus and face contexts generated from a knowledge base κ. That is, we have two kinds of information associated with each news article: the weighted list of entities extracted from the text of the article and the most relevant faces found in the knowledge base. Both the types are leveraged in a heuristic search algorithm to build a path of $\{$article, face-set$\}$ pairs between two articles $\{a_s, a_t\} \in \mathcal{A}$. Our heuristic algorithm maintains three properties during exploration: (1) any two consecutive articles during the search must maintain a maximum allowable distance θ, (2) face contexts of one article, as combined from certain number of most relevant faces, cannot be more than τ distant from the face contexts of a neighboring article, and (3) the search must have a progression over time, i.e., $Timestamp(a_i) \leq Timestamp(a_{i+1})$ for two consecutive articles a_i and a_{i+1}.

3 Experimental Results

The three specific questions we seek to answer in the following three subsections are: (Section 3.1) Does frontalization of key points result in better face recognition accuracy? (Section 3.2) How good are the contexts generated for the face images? (Section 3.3) Do the stories generated by the framework provide meaningful summarization of events?

3.1 Impact of Frontalization on Recognition

Although Storyboarding targets a different problem than face recognition, we evaluate the proposed frontalization technique using face recognition and a dataset of 293 labeled faces of 9 persons. In addition to the original face, we included mirror image of each face to make sure that we have at least two poses of the same face. Figure 2(left) compares face recognition accuracies at different training and test rations with different combinations of techniques. The figure shows that Eigenface with frontalization and mirror faces gives better accuracy than any other combination. The standard deviation of the accuracies was 2%-5%.

3.2 Quality of Face-Contexts

To evaluate the quality of the generated contexts against ground truth information, we sample 21 faces from $\mathcal{F}$, and manually attach most appropriate person entities to each of them with the help of human experts. We then compare these ground truth contexts with our automatically generated context. Figure 2(right) shows that Eigenface combined with both frontalization and mirroring provided the best average accuracy with different context sizes.

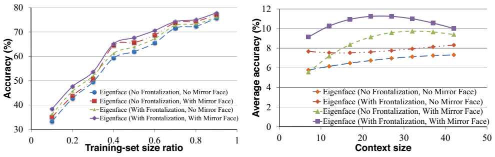


Figure 2: Comparison of face recognition accuracies (left) and context generation accuracies for human annotated test data (right).

3.3 Examples of Generated Stories using Boston Marathon Bombing Data

New York Times returns 1028 articles with the query "Boston Marathon Bombing". The storyboarding framework discovers a number of sub-events, each of which provides a concise mental model of a branch of the Boston Marathon Bombing tragedy and events afterwards (see Section 2.3). The first three articles of the story in Figure 1 describe the connection of the Boston bombers to the Waltham murders. The story moves forward till the penalty phase of the imprisoned bomber Dzhokhar Tsarnaev. The term cloud of the storyboard highlights the Boston Bombers and a friend of the Boston Bombers, *Todashev*, who is connected to the Waltham murders. Each related face list automatically selected from the knowledge base captures faces of people connected to the event. For example, the first face of the second article of the story belongs to Todashev. This face is found repeatedly in the consecutive articles. Rest of the faces in the story are of the bombers, police officers, and rescue crews.

References

Hossain, M. S.; Butler, P.; Boedihardjo, A. P.; and Ramakrishnan, N. 2012. Storytelling in entity networks to support intelligence analysts. In *KDD '12*, 1375–1383.

Sun, Y.; Wang, X.; and Tang, X. 2013. Deep convolutional network cascade for facial point detection. In *CVPR '13*, 3476–3483.

Viola, P., and Jones, M. 2001. Rapid object detection using a boosted cascade of simple features. In *CVPR '01*, volume 1 I–511.