Supplementary Material for
Video Co-summaries: Video Summarization by Visual Co-occurrence

Anonymous CVPR submission
Paper ID 865

#1. Closed-form updates of Algorithm 1

Suppose \( C \) is the co-occurrence matrix; \( u \) and \( v \) are the two binary vectors we are looking for (see notation\(^1\)). Recall that the sparse maximal biclique problem is written as:

\[
\begin{align*}
\max_{u,v} & \quad \sum_{ij} C_{ij} u_i v_j - \lambda_u \| u \|_1 - \lambda_v \| v \|_1 \\
\text{subject to} & \quad u_i + v_j \leq 1 + 1(C_{ij} \geq \epsilon), \forall i,j \\
& \quad u \in [0,1]^m, v \in [0,1]^n.
\end{align*}
\]

Suppose we solve for \( u \) with \( v \) fixed, (1) becomes:

\[
\begin{align*}
\max_{u\in[0,1]^m} & \quad \sum_i (C_{i:} v - \lambda_u) u_i \\
\text{subject to} & \quad u_1 \leq 1 + 1(C_j \geq \epsilon) - v_j, \forall j, \\
& \quad \vdots \\
& \quad u_m \leq 1 + 1(C_m \geq \epsilon) - v_j, \forall j.
\end{align*}
\]

As \( u_i \) is upper-bounded by \( n \) constraints, \( \forall i \), \( u_i \leq 1 + 1(C_{ij} \geq \epsilon) - v_j, \forall j \), for notational simplicity we denote the upper bound as \( \hat{u}_i = \min_i \{1(C_{ij} \geq \epsilon) - v_j\}_i \). Because \( u_i \) are independent of each other, we can derive the update

\[
\begin{align*}
\tilde{u}_i &= \begin{cases} 
0, & \text{if } C_{i:} v < \lambda_u, \\
\min(1 + \hat{u}_i, 1 + \min(0, \hat{u}_i)), & \text{otherwise.}
\end{cases} 
\end{align*}
\]

Introducing a non-positive operator \( (x)_- = \min(0, x) \), we can rewrite the second case in (3) as \( u_i = 1 + (\hat{u}_i)_- \), and rewrite the update (3) in a more compact way:

\[
\tilde{u}_i = \min(1(C_{i:} v \geq \lambda_u), 1 + (\hat{u}_i)_-).
\]

Similarly, we can write the upper bound for \( v_j \) as \( \hat{v}_j = \min_j \{1(C_{ij} \geq \epsilon) - u_i\}_j \), and its update in compact way:

\[
v_j = \min(1(u^\top C_{ij} \geq \lambda_u), 1 + (\hat{v}_j)_-).
\]

Since each element in \( u \) and \( v \) can be computed independently, these closed-form updates are parallelizable, implying our proposed algorithm can be extended larger scale dataset.

#2. User study GUI for concept visualization

For the concept visualization experiment in Sec. 4.3 of the main manuscript, we designed a user study GUI. Figs. 1 and 2 illustrate the snapshots of the evaluation webpage. The goal is to identify how human perceive video summaries gives the videos sharing one topic, e.g., Surfing. Before the study begins, we provide the introductory paragraph as follows:

"Imagine a video search website that shows you a quick summary of each video as a preview. The quality of video summaries (e.g., relevance to the query/title/etc.) is crucial for better user experience.

We want to understand how people perceive the quality of video summaries according to a specific query. This questionnaire will gather data to help answer this question."
Figure 1. Entrance page of our AMT-like webpage for user study on concept visualization in Sec. 4.3 of the main manuscript. Part 1 collects the user demography. Part 2 gives instructions about what the users will evaluate. Part 3 provides examples about Good, Neutral and Bad summaries using the query \textit{Surfing}. 
Figure 2. Evaluation page of our AWT-like webpage for user study on concept visualization in Sec. 4.3 of the main manuscript. The progress, i.e., 1 out of 10 categories, is shown on top of the webpage. The query string Surfing is highlighted with bright yellow color. A quadruplet of summaries (each generated from different methods) is shown at a time. The order of the summaries is completely random. Totally 5 sets of quadruplets are shown to a user. A user is required to select at least one good summary and one bad summary to continue to the next set of videos. To assist the users, we used green, blue and red colors to indicate Good, Neutral and Bad summaries.

To perform a fair comparison, we provided a quadruplet of 4 summaries at a time (as described in Sec. 4.3 in the manuscript). Each summary was generated from one method, i.e., $k$-means (baseline approach), LiveLight [2] (state-of-the-art unsupervised video summarization), co-clustering [1] (originally for word-document classification and our first attempt), and MBF (the proposed maximal biclique finding algorithm). The order of the summaries were randomized in a way that the viewers cannot tell by which method each summary was generated, and have to select at least one good summary and one bad summary to continue to the next quadruplet. This study was carried out across 20 subjects (14 males and 6 females ranging from 23 to 33 years old). The entire evaluation took about 30 minutes for a user.

3. More results on concept visualization

This section shows the qualitative results of concept visualization that could not be fitted into the main manuscript due to space limitation. Figs. 3~12 show the results of different methods on visualizing the concept of each video category. Each row shows the results of $k$-means (baseline approach), LiveLight [2] (state-of-the-art unsupervised video summarization), co-clustering [1] (originally...
for word-document classification and our first attempt), and MBF (the proposed maximal biclique finding algorithm), respectively. In each row, the summaries are ranked from left to right according to the quality scores computed in Sec. 3.3 in the manuscript. Each summary is represented as a snapshot that was taken as the central frame of a summary.

Figure 3. Illustration of different methods on visualizing the concept of Base jump.

Figure 4. Illustration of different methods on visualizing the concept of Bike polo.

References


Figure 5. Illustration of different methods on visualizing the concept of *Eiffel Tower*.

Figure 6. Illustration of different methods on visualizing the concept of *Excavators river Crossing*.
Figure 7. Illustration of different methods on visualizing the concept of *Kids playing in leaves*.

Figure 8. Illustration of different methods on visualizing the concept of *MLB*.
Figure 9. Illustration of different methods on visualizing the concept of NFL.

Figure 10. Illustration of different methods on visualizing the concept of Notre Dame Cathedral.
Figure 11. Illustration of different methods on visualizing the concept of Statue of Liberty.

Figure 12. Illustration of different methods on visualizing the concept of Surfing.