Hidden Markov Model For Event Photo Stream Segmentation

Jesse Prabawa Gozali, Min-Yen Kan — National University of Singapore
Hari Sundaram — Arizona State University

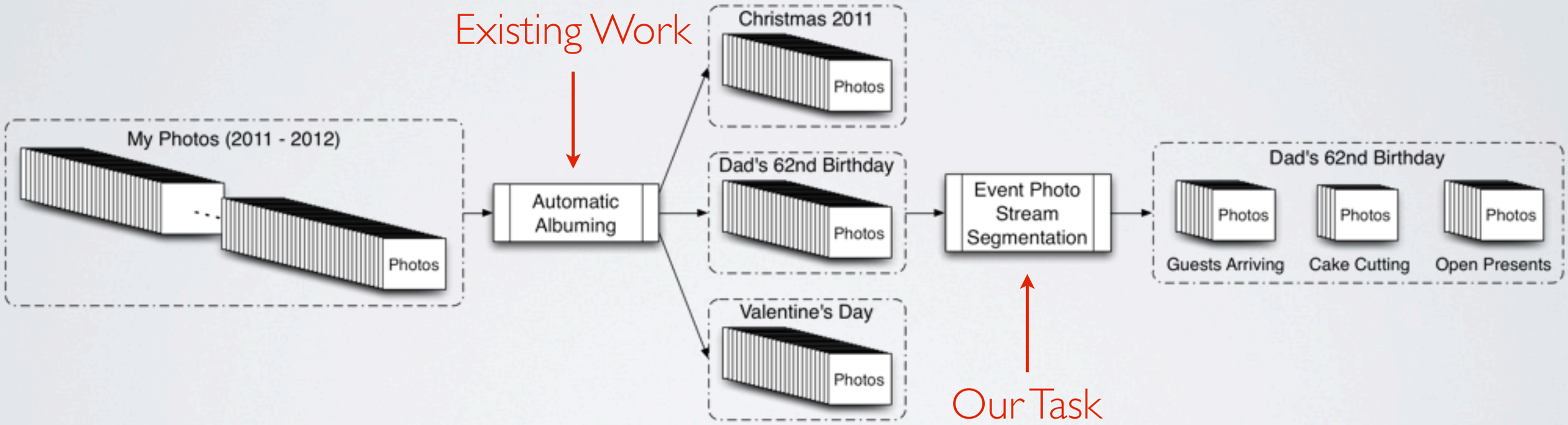
This work is supported by the NExT Research Center, funded by MDA, Singapore, under the research grant WBS: R-252-300- 001-490.

EVENT PHOTO STREAM SEGMENTATION

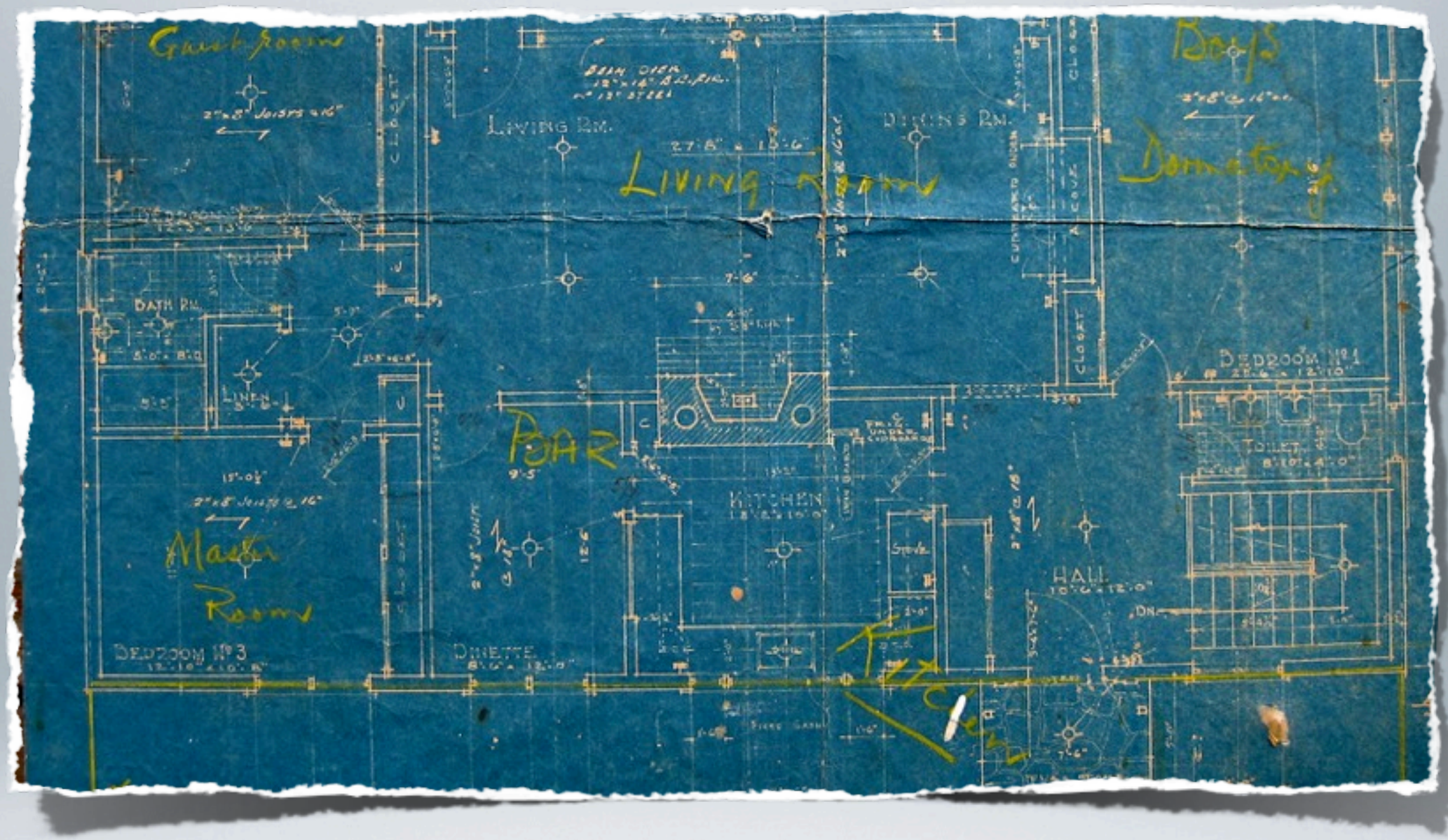
- An **event photo stream** is the chronological sequence of photos of a single event.
- **Event photo stream segmentation** is the process of finding contiguous groups of photos from an event photo stream, each corresponding to a photo-worthy moment in the event.

RELATED WORK

- **Automatic albuming**: Existing segmentation algorithms operate on large collection of photos (months/years) and produce segments (groups of photos) that correspond to events



- **Event photo stream segmentation** is different: data sparsity, indistinct time gaps, visual similarities

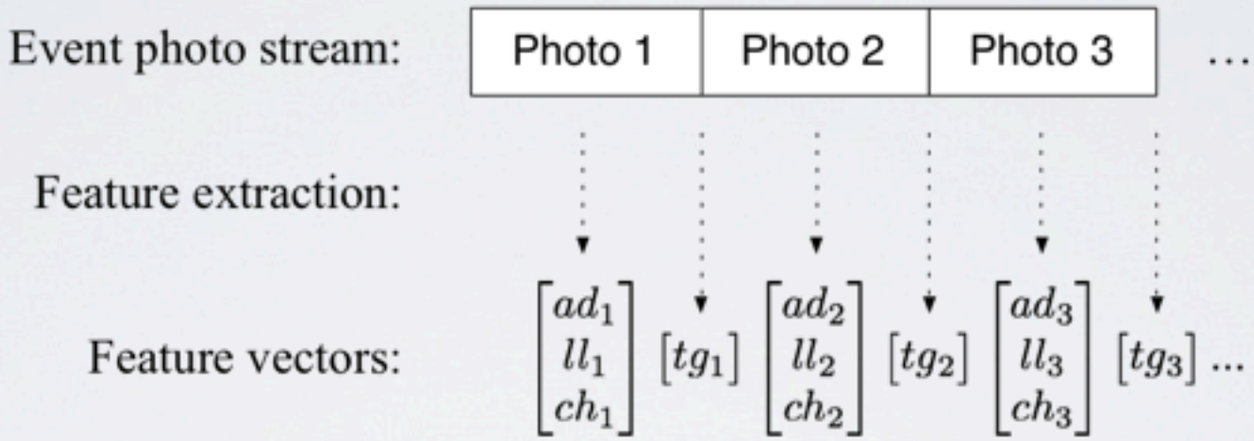


PROBLEM DEFINITION AND MODELING

http://www.flickr.com/photos/mburpee/4928429020/

MODELING AND PROBLEM DEFINITION

- An event photo stream is a sequence of **alternating feature vector types**

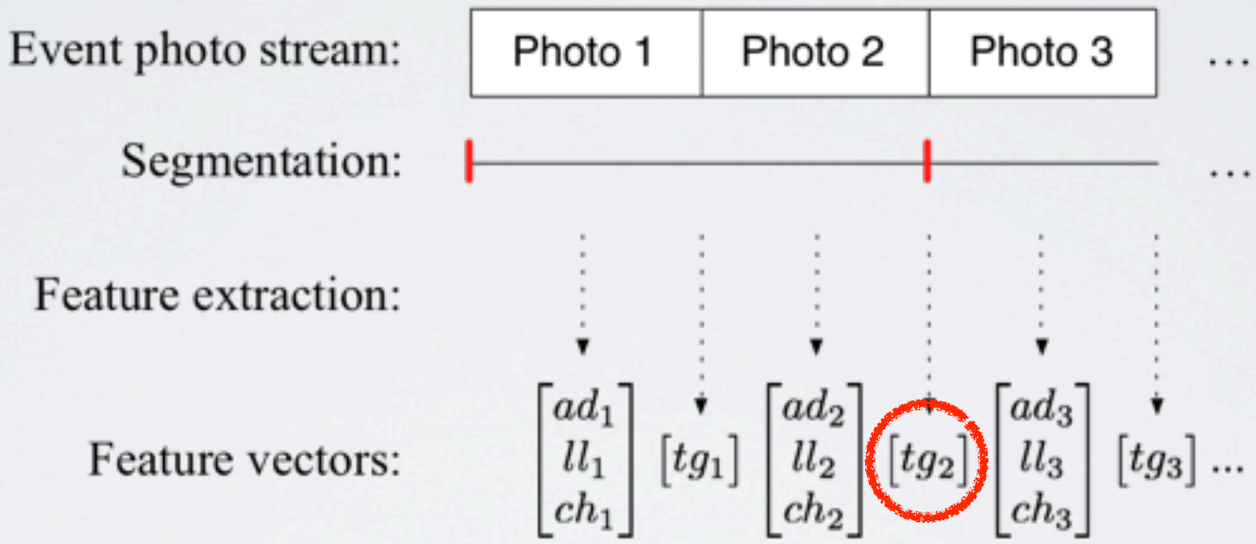


  - Photo feature (about the photo) — aperture diameter, LogLight, color histogram
  - Photo gap feature (about the gap between consecutive photos) — time gap

- Segmentation is the process of **identifying segment boundaries** amongst the gaps between consecutive photos, given the sequence of feature vectors.

MODELING AND PROBLEM DEFINITION

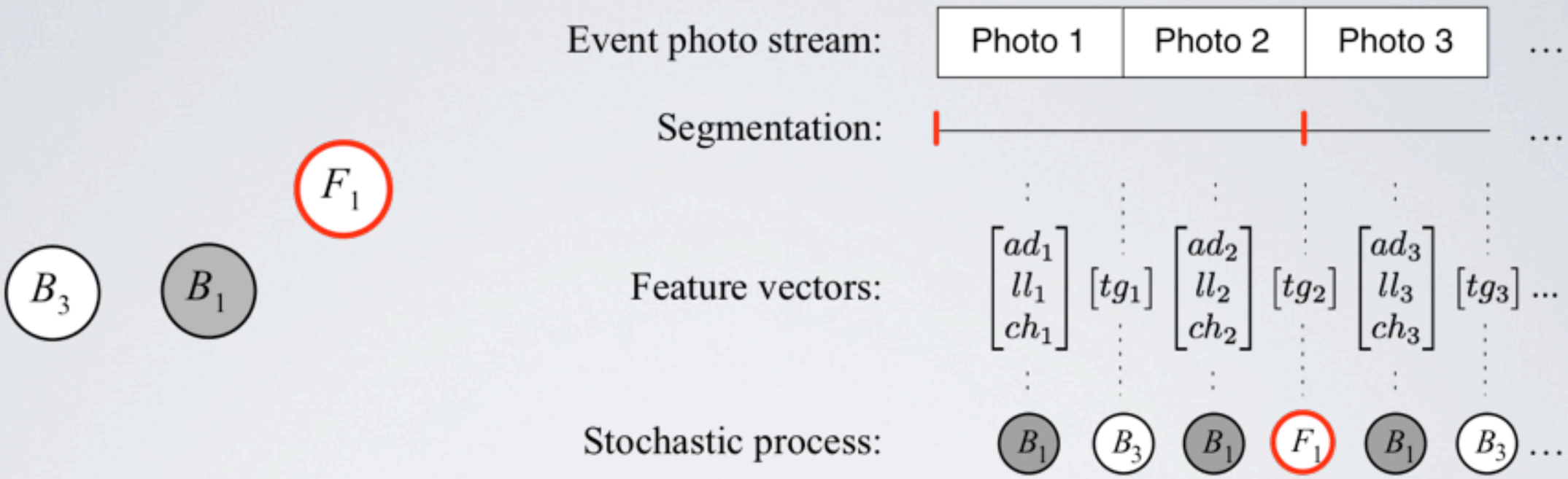
- An event photo stream is a sequence of **alternating feature vector types**



  - Photo feature (about the photo) — aperture diameter, LogLight, color histogram
  - Photo gap feature (about the gap between consecutive photos) — time gap

- Segmentation is the process of **identifying segment boundaries** amongst the gaps between consecutive photos, given the sequence of feature vectors.

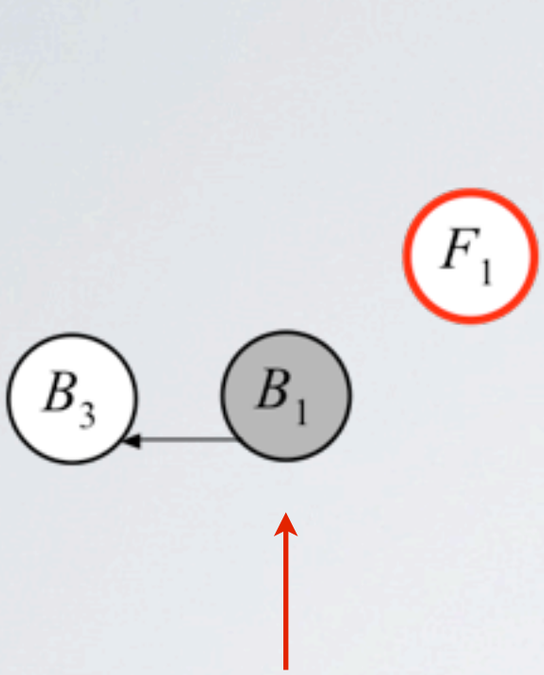
A GENERATIVE MODEL

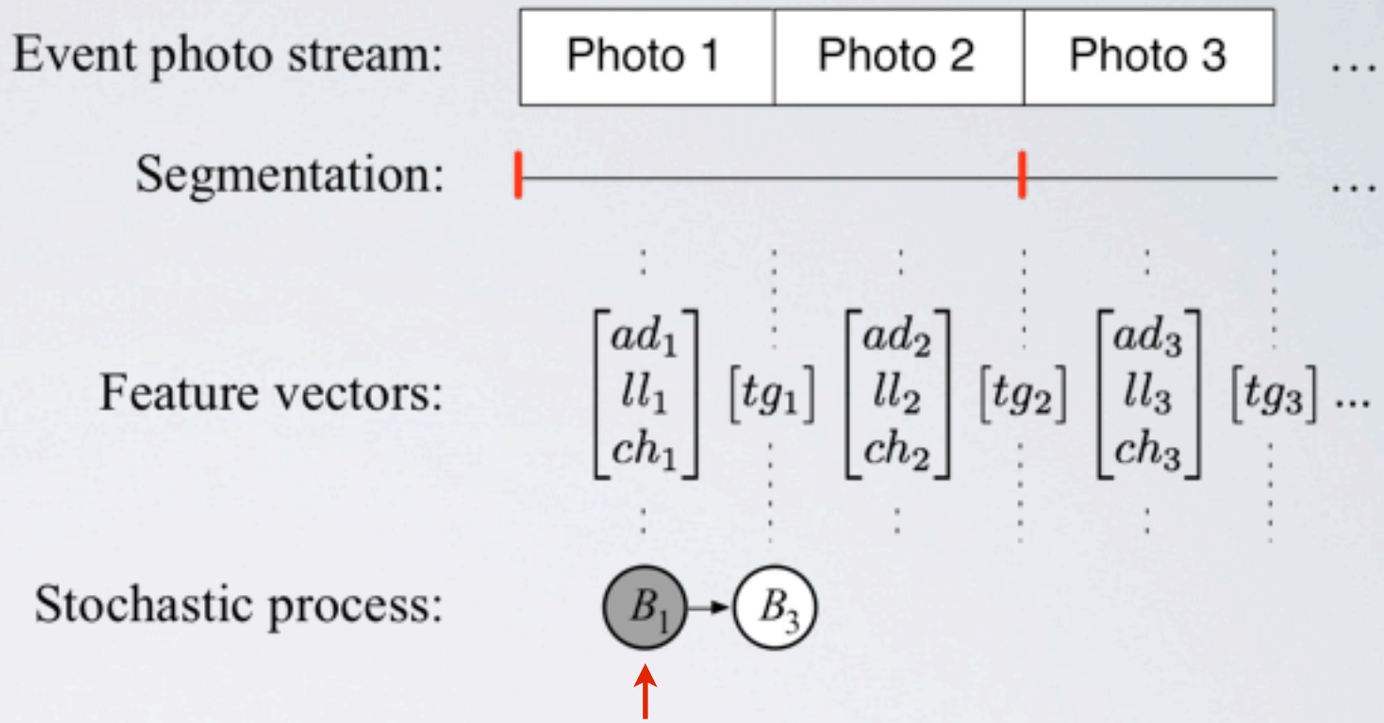


- Our method is based on a generative model:
  - Foreground states produce feature vectors corresponding to segment boundaries
  - Background states produce the surrounding feature vectors

- The simplest model has 3 states: foreground state, background state (photo feature), background state (photo gap feature)

A GENERATIVE MODEL

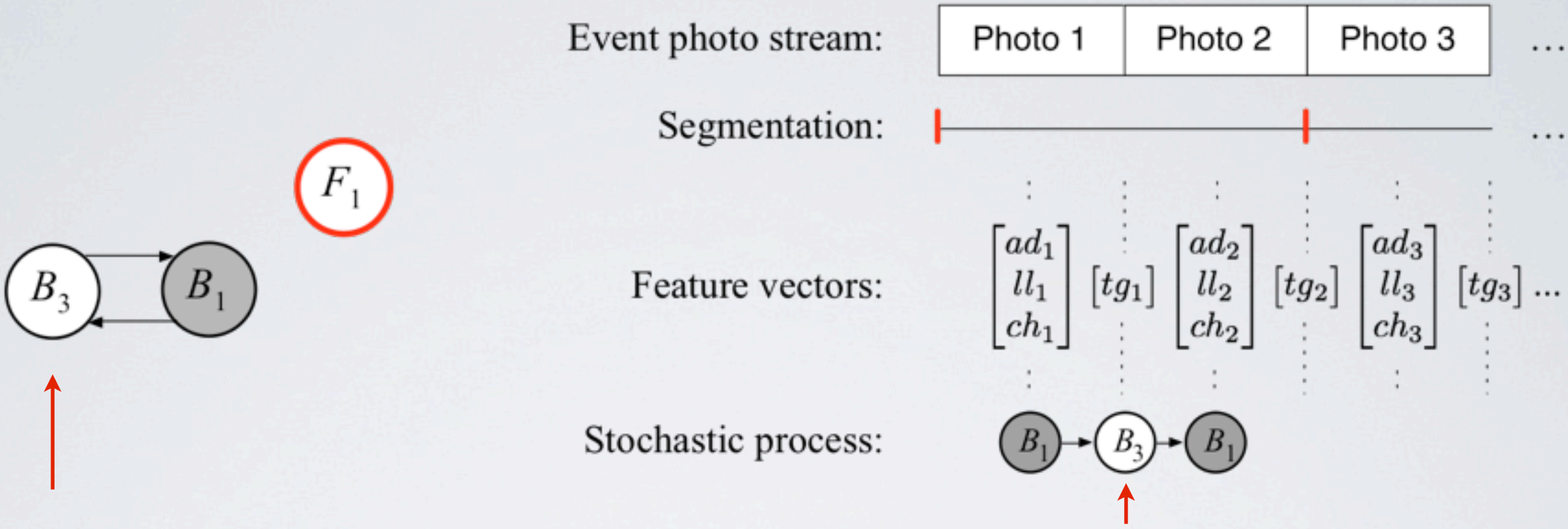




- Our method is based on a generative model:
  - Foreground states produce feature vectors corresponding to segment boundaries
  - Background states produce the surrounding feature vectors

- The simplest model has 3 states: foreground state, background state (photo feature), background state (photo gap feature)

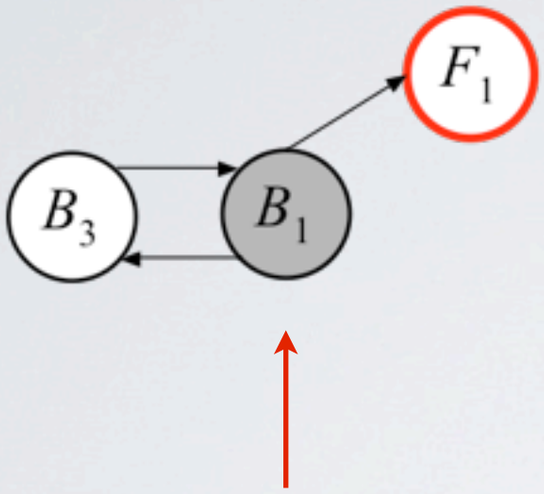
A GENERATIVE MODEL

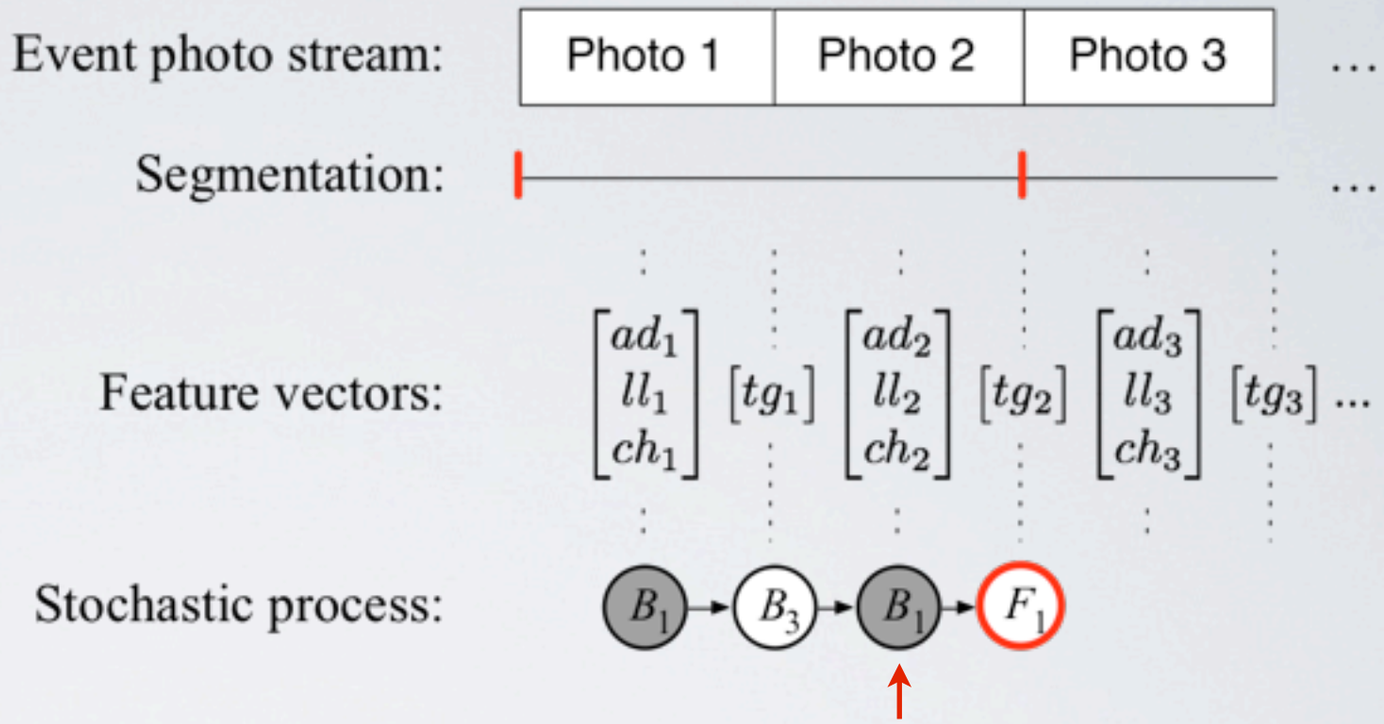


- Our method is based on a generative model:
  - Foreground states produce feature vectors corresponding to segment boundaries
  - Background states produce the surrounding feature vectors

- The simplest model has 3 states: foreground state, background state (photo feature), background state (photo gap feature)

A GENERATIVE MODEL

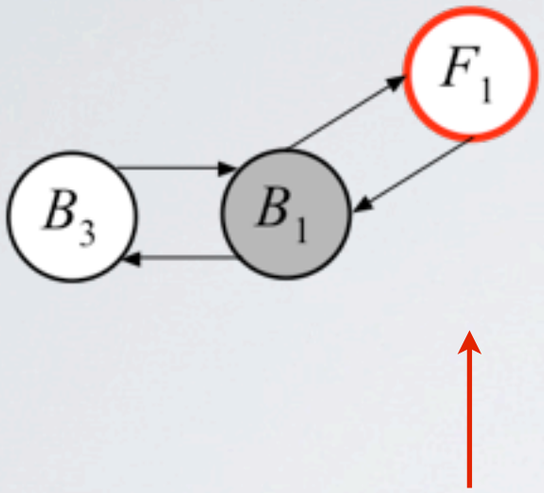


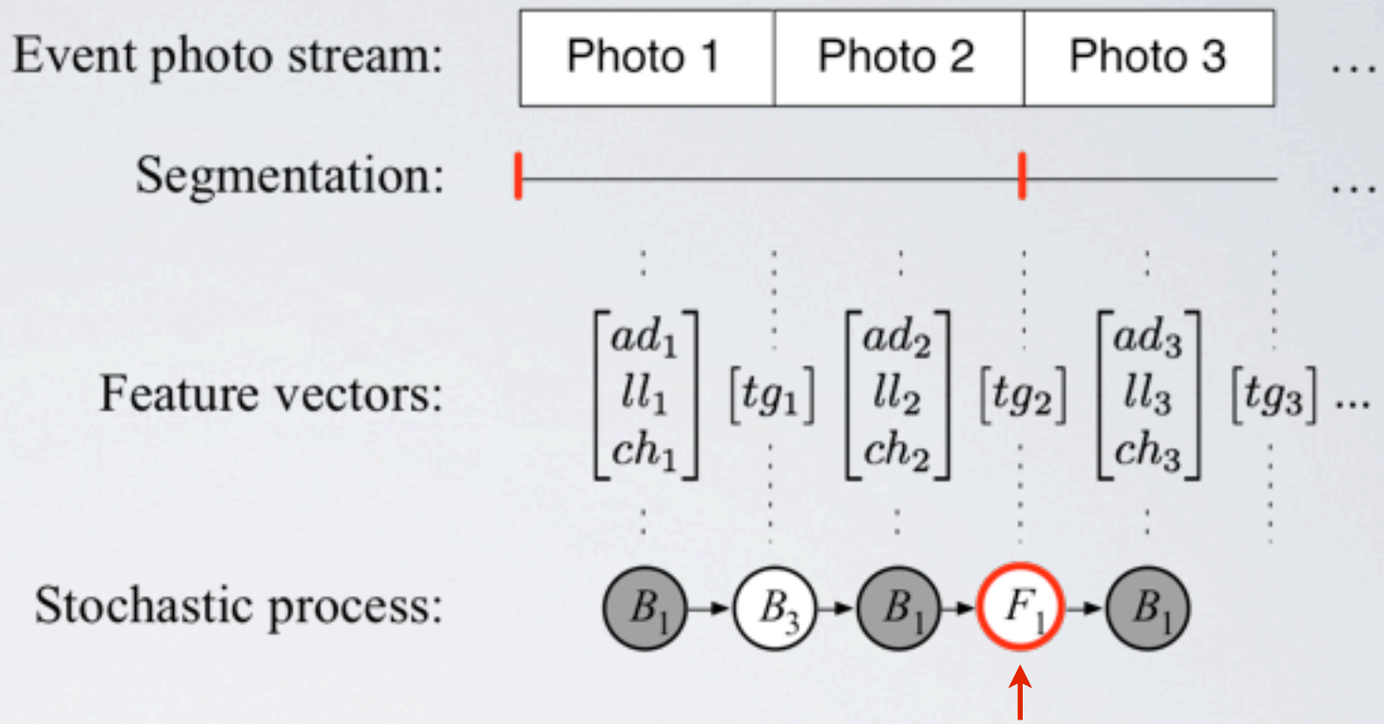


- Our method is based on a generative model:
  - Foreground states produce feature vectors corresponding to segment boundaries
  - Background states produce the surrounding feature vectors

- The simplest model has 3 states: foreground state, background state (photo feature), background state (photo gap feature)

A GENERATIVE MODEL

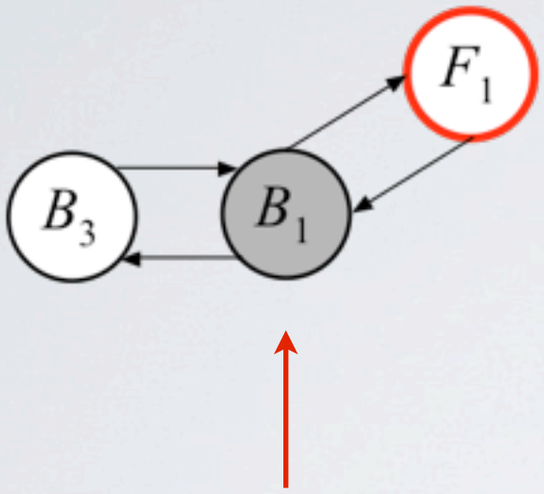


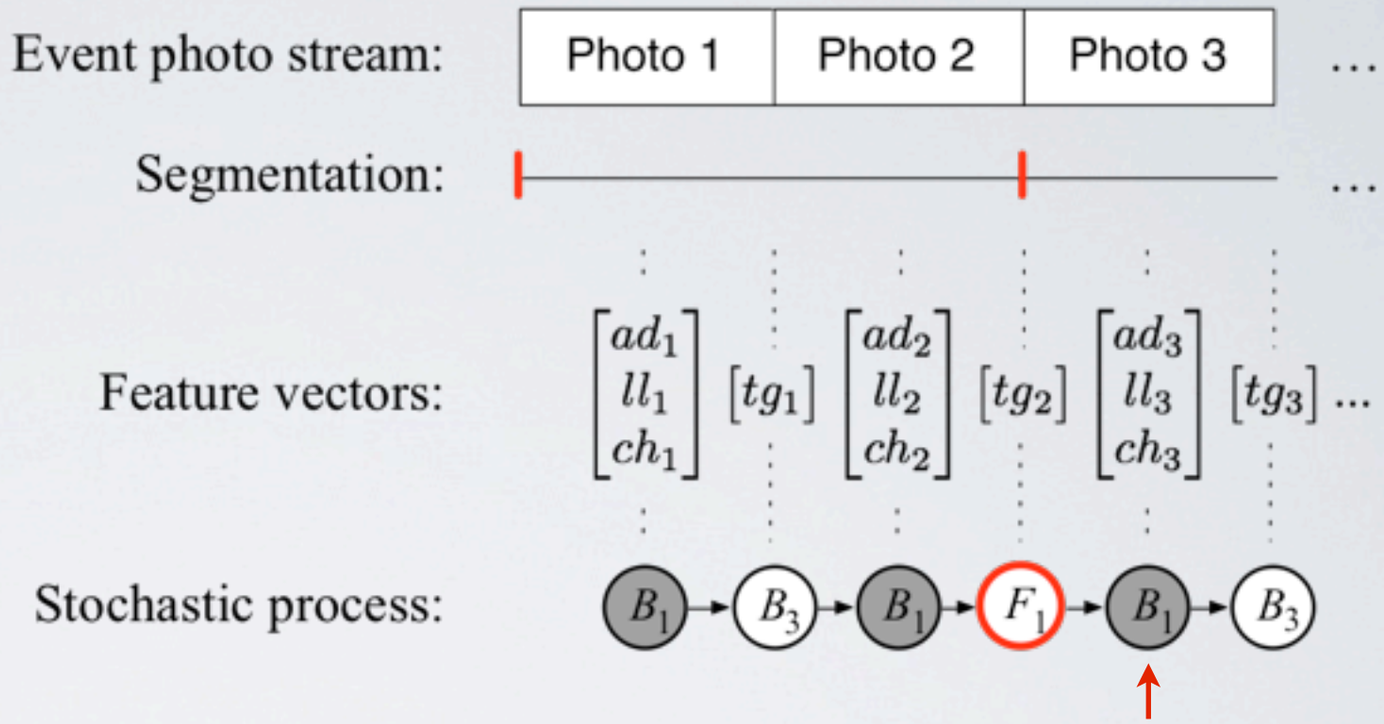


- Our method is based on a generative model:
  - Foreground states produce feature vectors corresponding to segment boundaries
  - Background states produce the surrounding feature vectors

- The simplest model has 3 states: foreground state, background state (photo feature), background state (photo gap feature)

A GENERATIVE MODEL

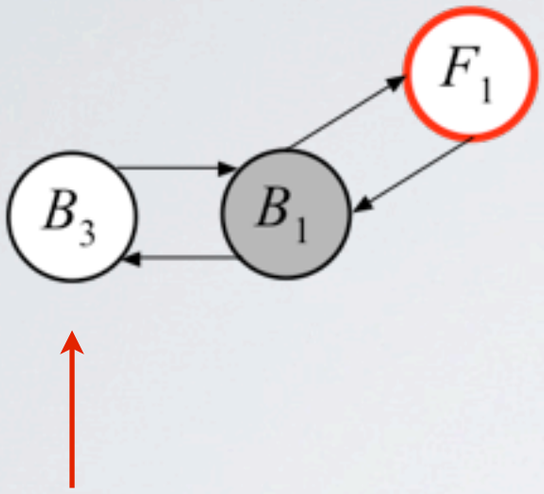


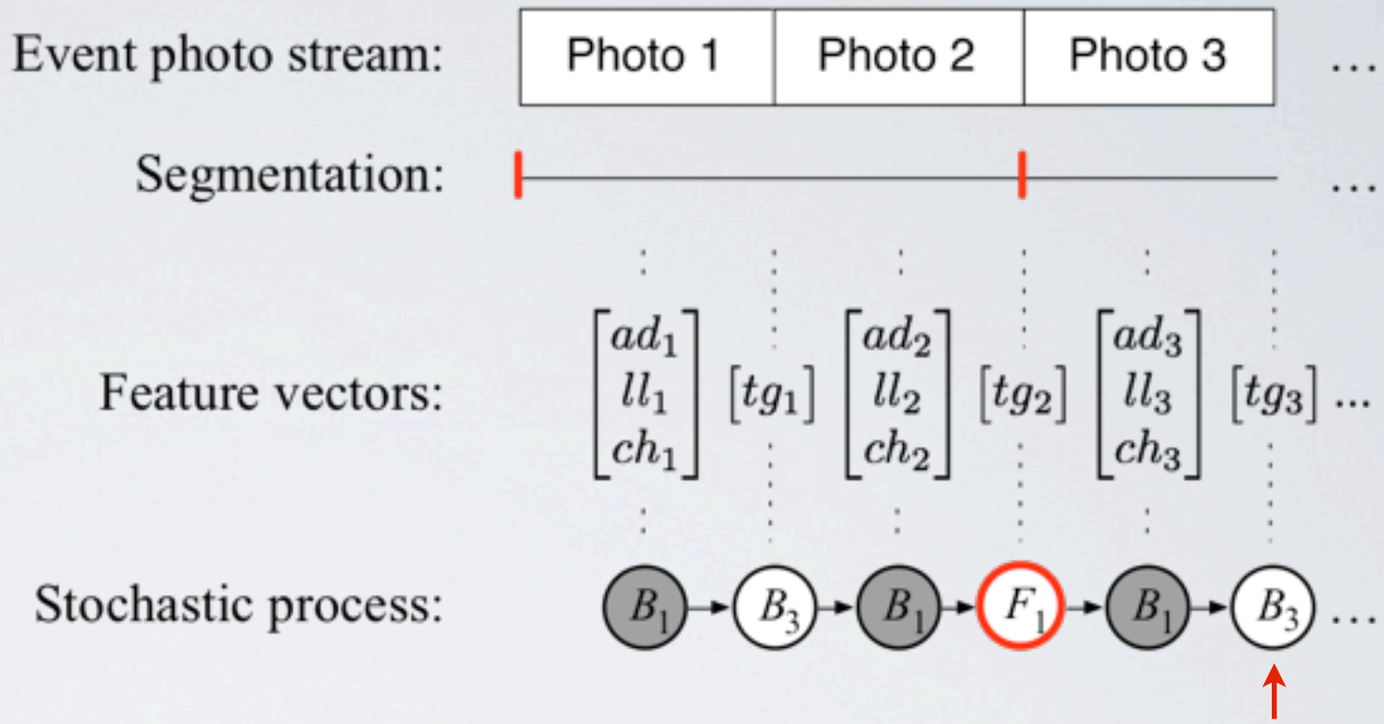


- Our method is based on a generative model:
  - Foreground states produce feature vectors corresponding to segment boundaries
  - Background states produce the surrounding feature vectors

- The simplest model has 3 states: foreground state, background state (photo feature), background state (photo gap feature)

A GENERATIVE MODEL

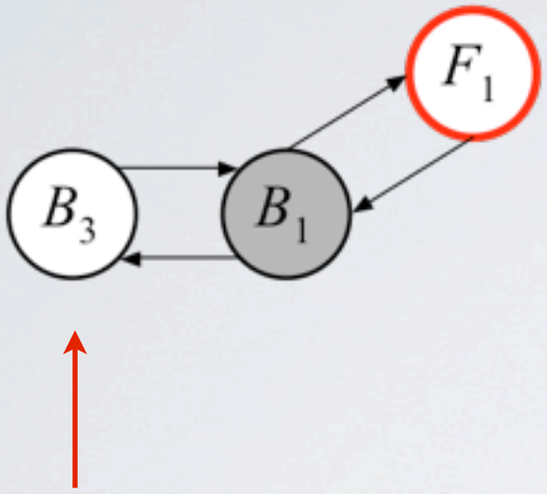


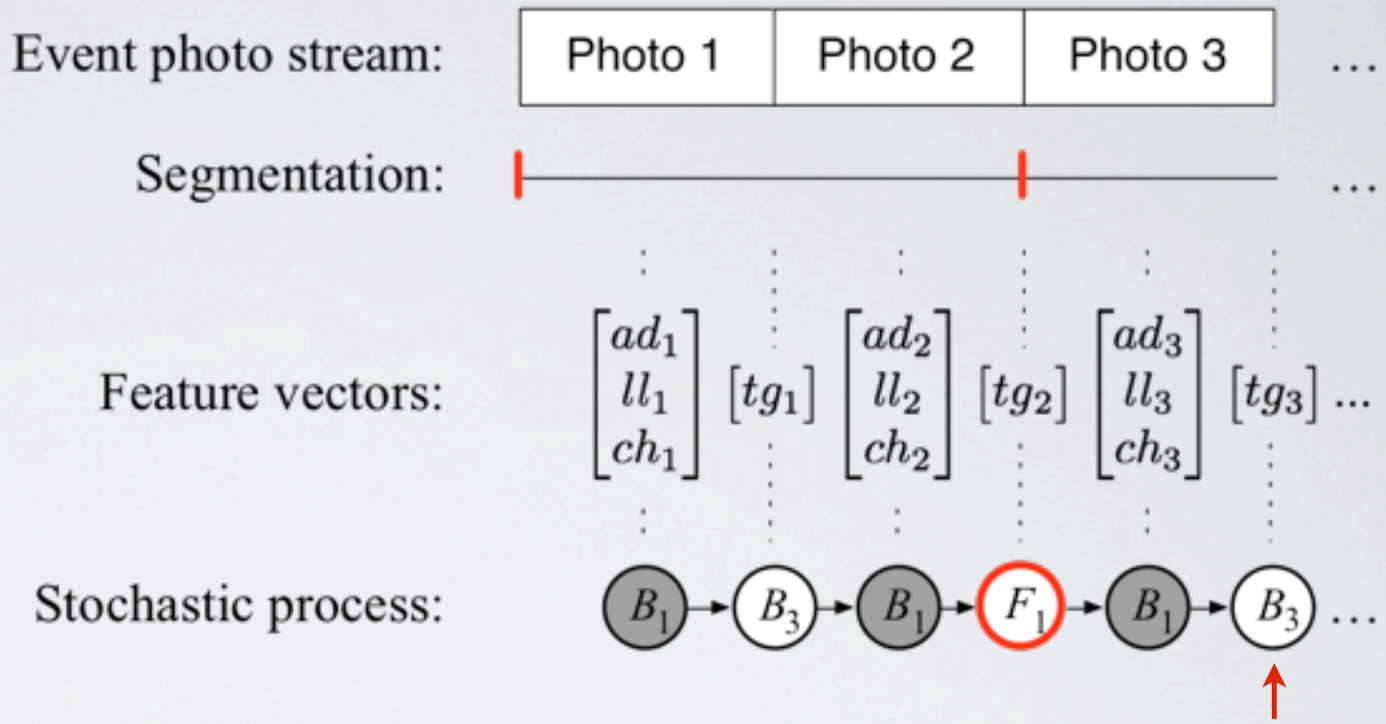


- Our method is based on a generative model:
  - Foreground states produce feature vectors corresponding to segment boundaries
  - Background states produce the surrounding feature vectors

- The simplest model has 3 states: foreground state, background state (photo feature), background state (photo gap feature)

A GENERATIVE MODEL



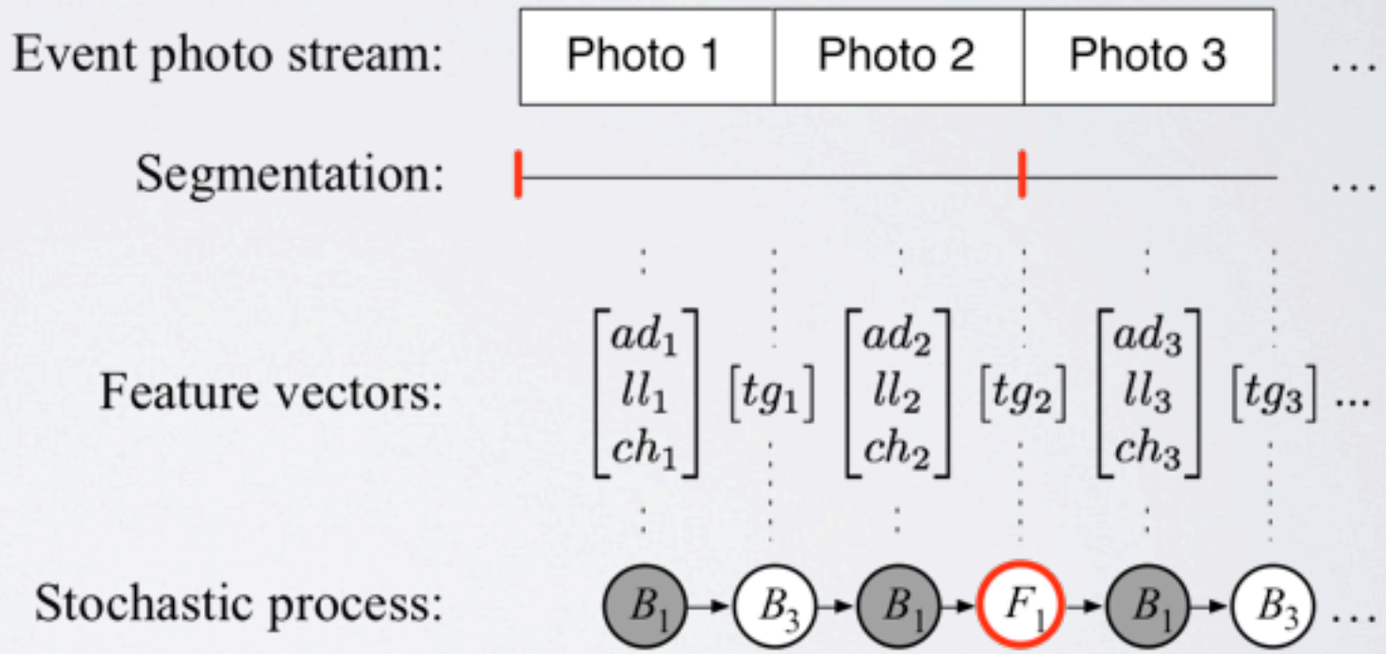


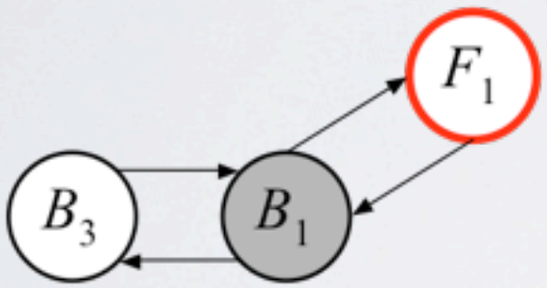
- Our method is based on a generative model:
  - Foreground states produce feature vectors corresponding to segment boundaries
  - Background states produce the surrounding feature vectors

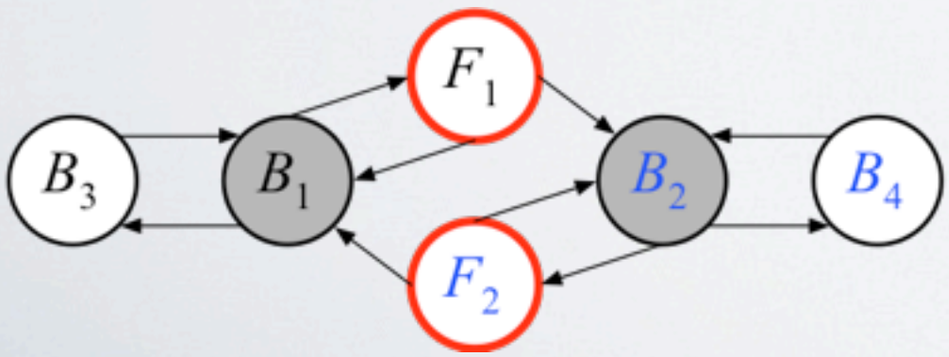
- The simplest model has 3 states: foreground state, background state (photo feature), background state (photo gap feature)

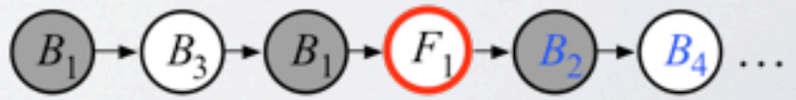
A GENERATIVE MODEL

- **High-level meaning**: Features in each segment (group of photos) follow the output distribution of the background states (B_1 and B_3).
- What if the segments follow more than one output distribution?
  We build larger models using the 3-state model as a basic building block







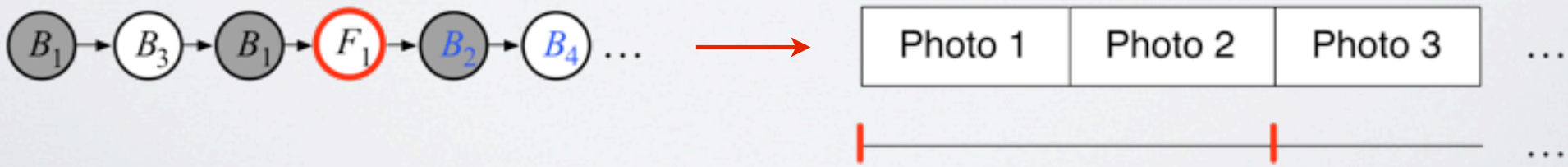


HIDDEN MARKOV MODEL

- The stochastic process of our model can be described by a **Hidden Markov Model** (HMM)
  - Computations are very efficient
  - Successfully applied in other domains (speech, text segmentation, topic detection, information extraction)

- With a trained HMM, given the sequence of feature vectors, we can find the most likely state sequence taken by the HMM to produce the sequence of feature vectors (Viterbi algorithm).

- With the state sequence, we can find the feature vectors that correspond to the foreground states (segment boundaries) to produce the segmentation.



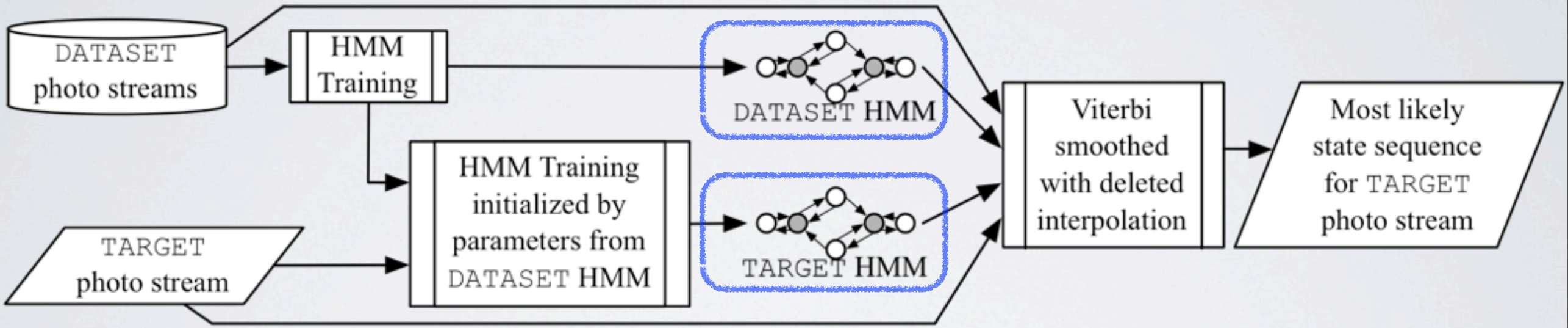
TRAINING

http://www.flickr.com/photos/tomas_sobek/7412838894/

TRAINING

- Given a sequence of feature vectors we want to segment (TARGET)
- We need to train the HMM to produce TARGET — or in other words, find the best model parameters to produce TARGET with the highest probability
- Thus, we train the HMM with
  - TARGET
  - and other sequences of feature vectors (DATASET) — to alleviate data sparsity
- The model parameters learnt from the two data sources are smoothed using deleted interpolation (Jelinek & Mercer, 1980)

WORKFLOW


DATASET
photo streams
HMM
Training
DATASET HMM
HMM Training
initialized by
parameters from
DATASET HMM
TARGET HMM
TARGET
photo stream
Viterbi
smoothed
with deleted
interpolation
Most likely
state sequence
for TARGET
photo stream

EVALUATION AND RESULTS

http://www.flickr.com/photos/lokesh/6767422267/

EVALUATION

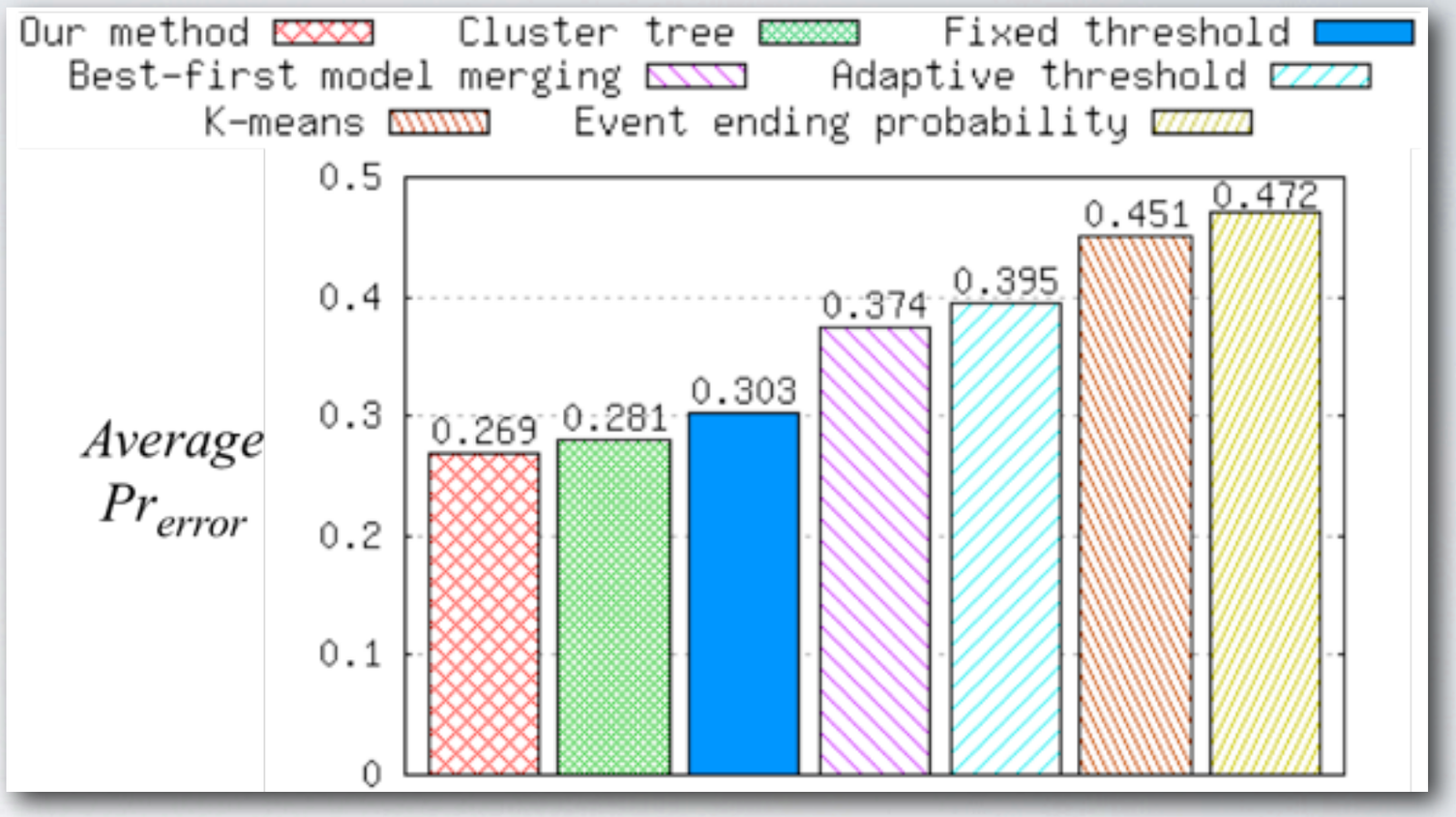
- Dataset
  - 28 event photo streams of various event types (four from Flickr, 24 from seven volunteers)
  - Ground truth segmentation provided by the volunteers (except Flickr ones done by first author)

- Six Baselines
  - Cluster Tree (Graham et. al., 2002) — state-of-the-art
  - Fixed threshold (Platt et. al., 2000), Best-first model merging (Platt et. al., 2003), Adaptive threshold (Platt et. al., 2003), K-means (Loui & Savakis, 2003), Event ending probability (Zhao et. al., 2006)

RESULTS

Metric: Pr_{Error}
(Georgescul et. al., 2006)

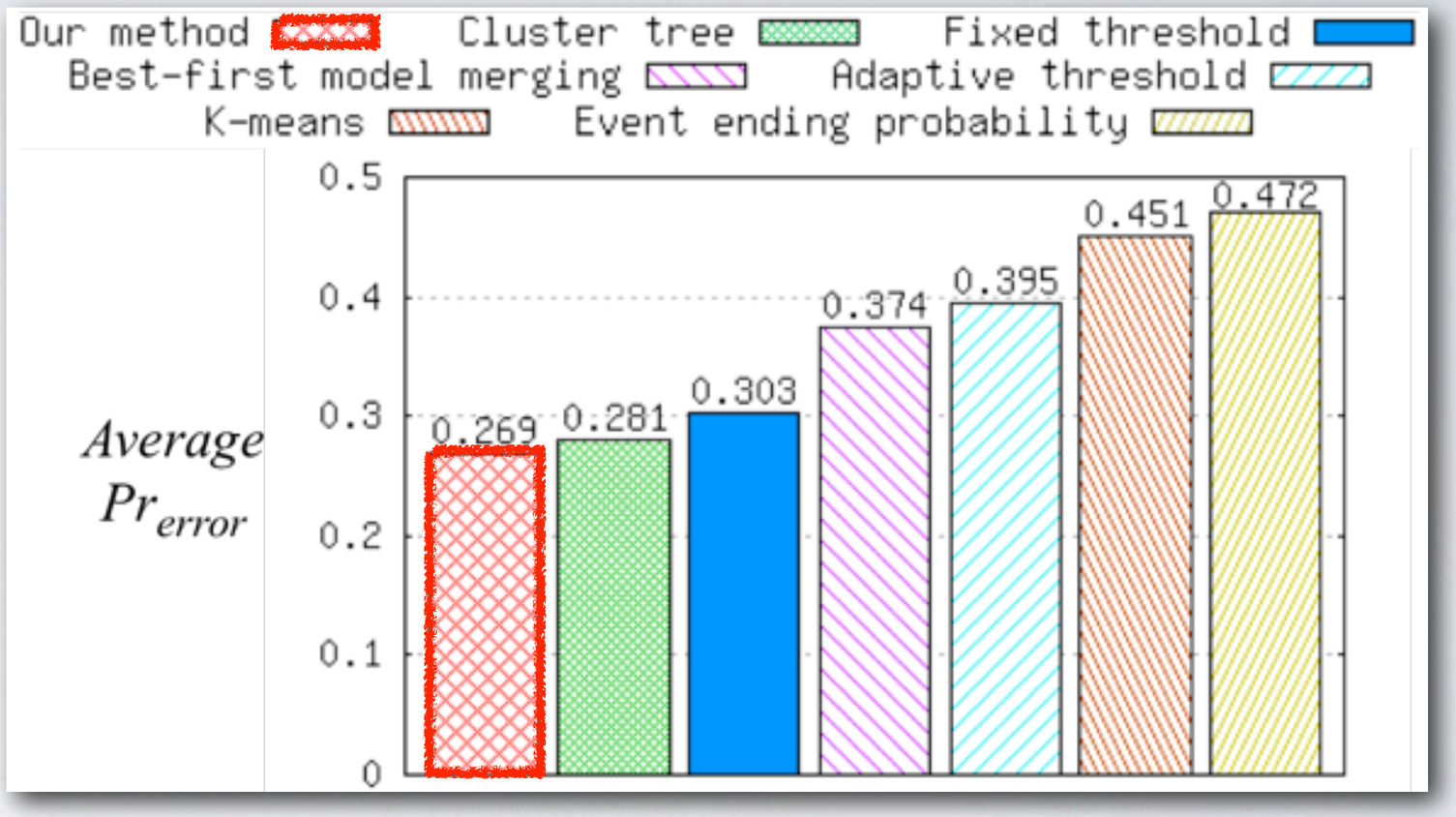
- Error rate against the ground truth segmentation
- Smaller is better
- A method that proposes no segment boundaries or all segment boundaries will receive an error rate of about 0.5



RESULTS

Metric: Pr_{Error}
(Georgescul et. al., 2006)

- Error rate against the ground truth segmentation
- Smaller is better
- A method that proposes no segment boundaries or all segment boundaries will receive an error rate of about 0.5

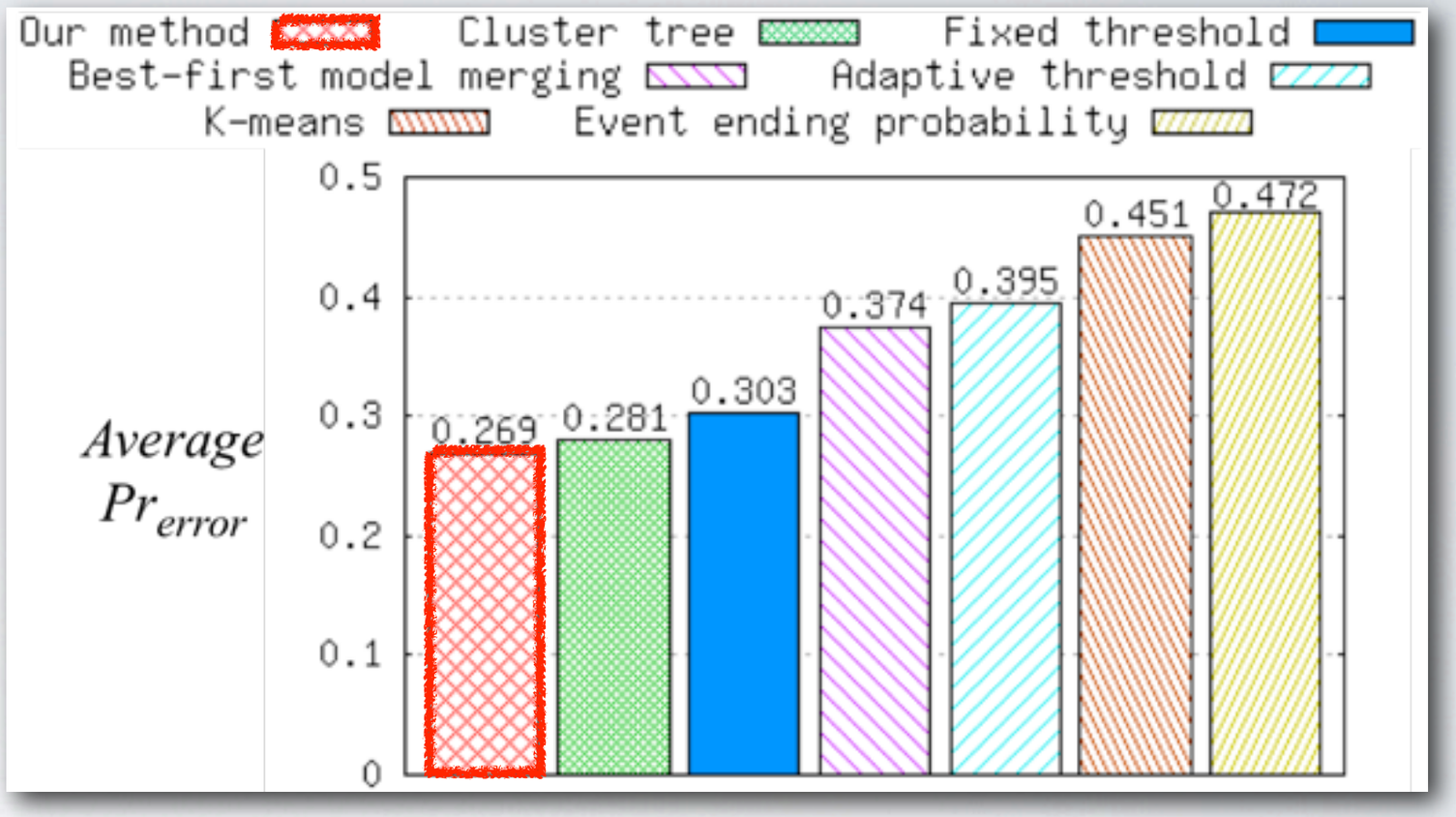


We perform better than state-of-the-art ($p < 0.1$) and other baselines ($p < 0.005$)

RESULTS

Metric: Pr_{Error}
(Georgescul et. al., 2006)

- Error rate against the ground truth segmentation
- Smaller is better
- A method that proposes no segment boundaries or all segment boundaries will receive an error rate of about 0.5

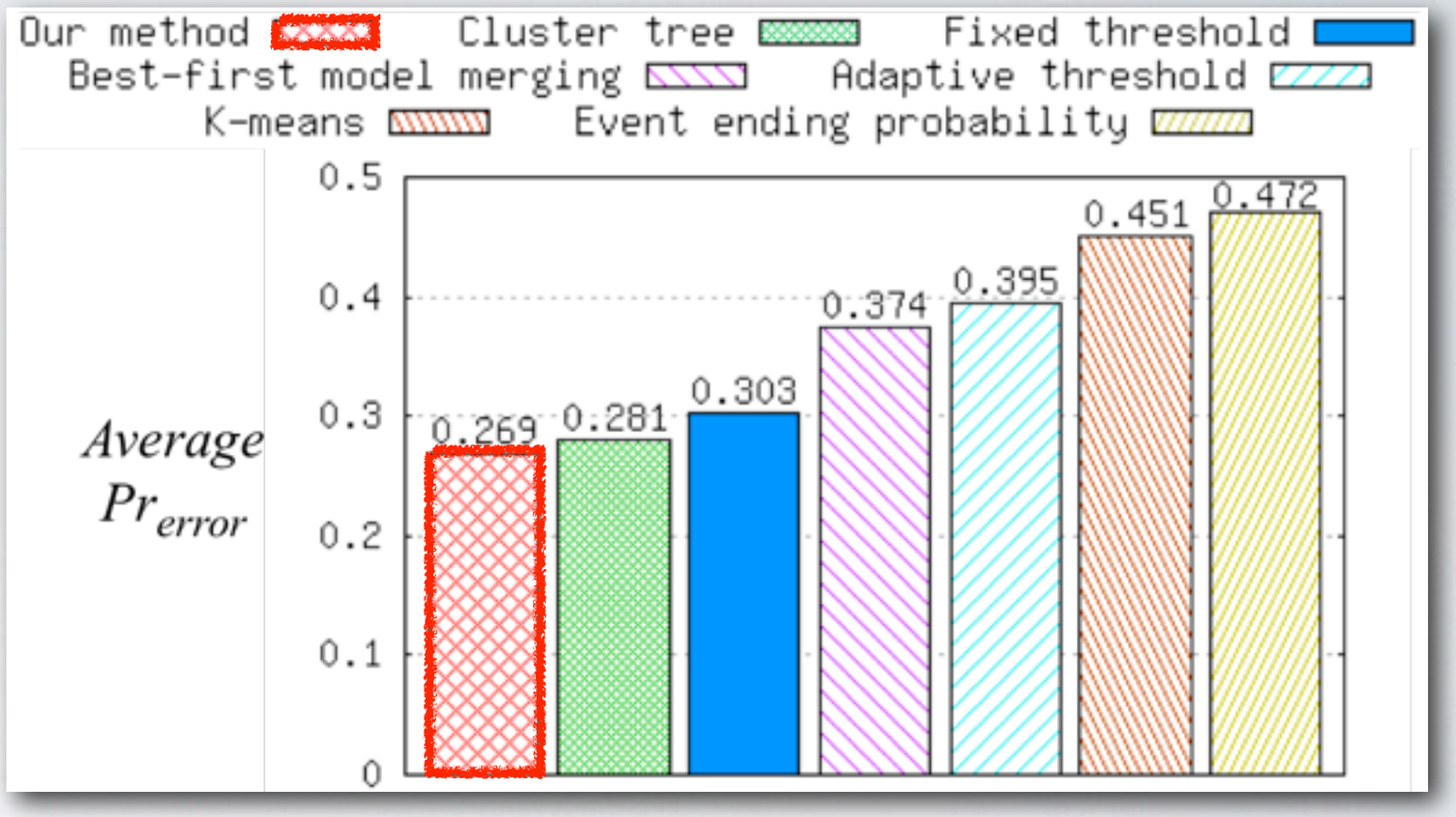


The event photo stream segmentation baseline performed better than the automatic albuming baselines

RESULTS

Metric: Pr_{Error}
(Georgescul et. al., 2006)

- Error rate against the ground truth segmentation
- Smaller is better
- A method that proposes no segment boundaries or all segment boundaries will receive an error rate of about 0.5

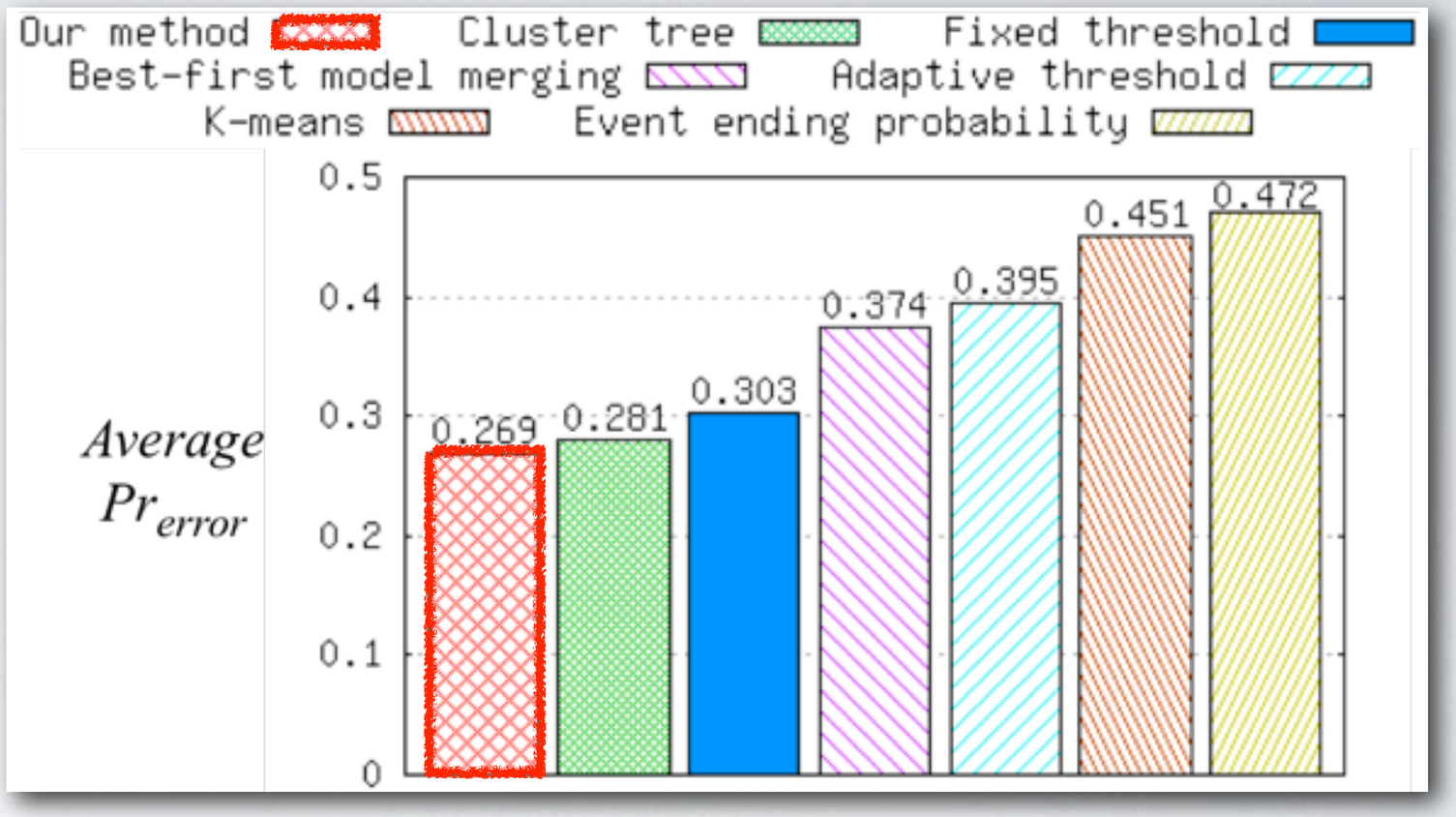


Among the automatic albuming methods, the simple fixed threshold method works the best

RESULTS

Metric: Pr_{Error}
(Georgescul et. al., 2006)

- Error rate against the ground truth segmentation
- Smaller is better
- A method that proposes no segment boundaries or all segment boundaries will receive an error rate of about 0.5

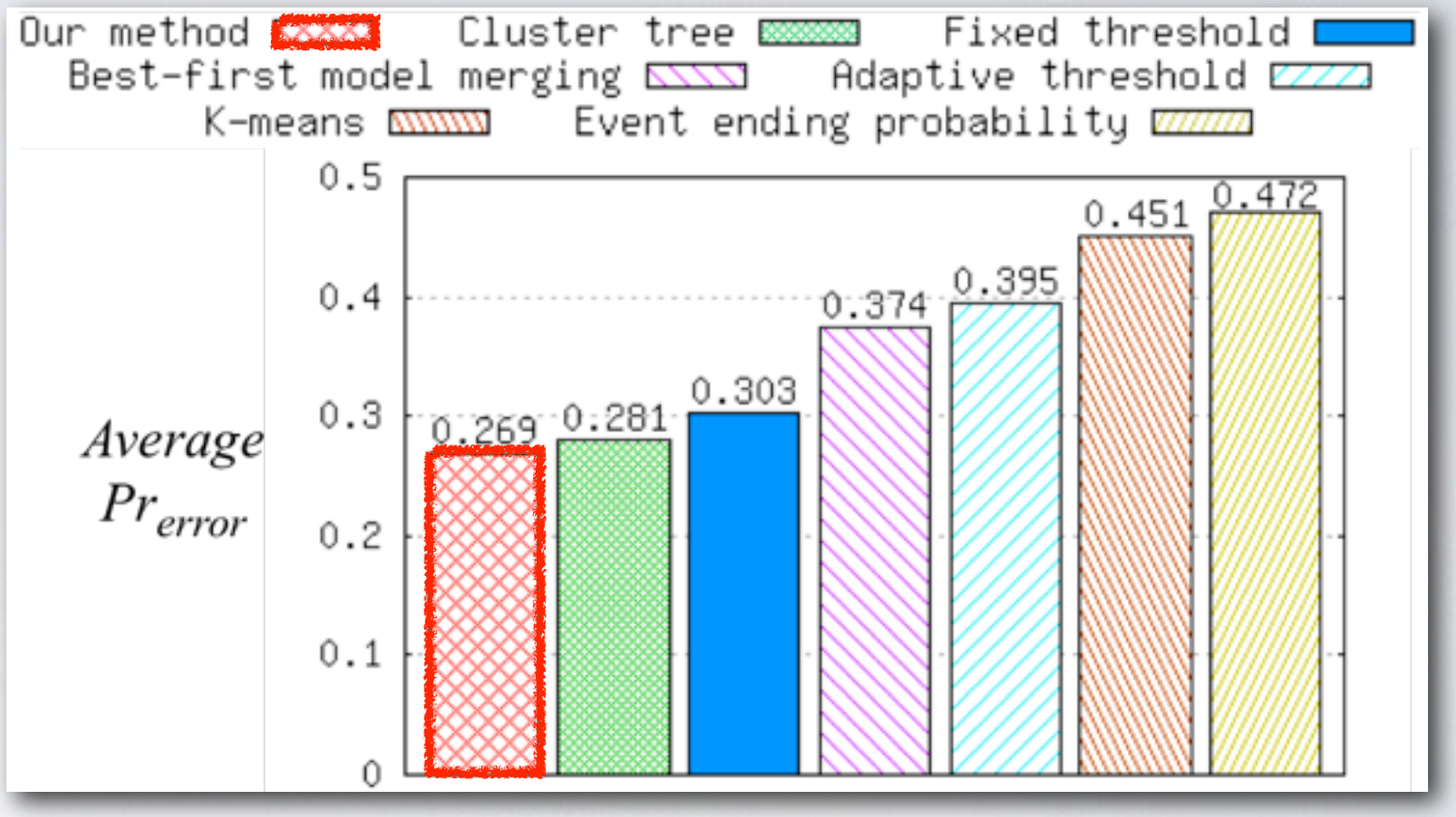


These methods performed better than the fixed threshold method for automatic albuming

RESULTS

Metric: Pr_{Error}
(Georgescul et. al., 2006)

- Error rate against the ground truth segmentation
- Smaller is better
- A method that proposes no segment boundaries or all segment boundaries will receive an error rate of about 0.5



The baseline methods that rely heavily on heuristics performed the worst

CONCLUSION

- Proposed an automatic algorithm to segment event photo streams (chronological sequence of photos from a single event)
  - Based on observation of alternating feature vector types
  - Trained using unsegmented, unlabelled event photo streams
  - Using deleted interpolation smoothing to alleviate data sparsity
  - Using only simple features (metadata and color histogram)
  - Outperform all baselines with statistical significance
- Complements existing photo organization methods that operate on events, faces, geolocations

FUTURE WORK

- Per-state HMM training from segmented, labelled training data
- Visual features
- User Study
  - JCDL 2012: "How Do People Organize Photos In Each Event and How Does It Affect Storytelling, Searching and Interpretation Tasks?"
  - End-user application: "*Chaptrs* photo browser"

Thank You!

Jesse P. Gozali
jprabawa@comp.nus.edu.sg
http://wing.comp.nus.edu.sg