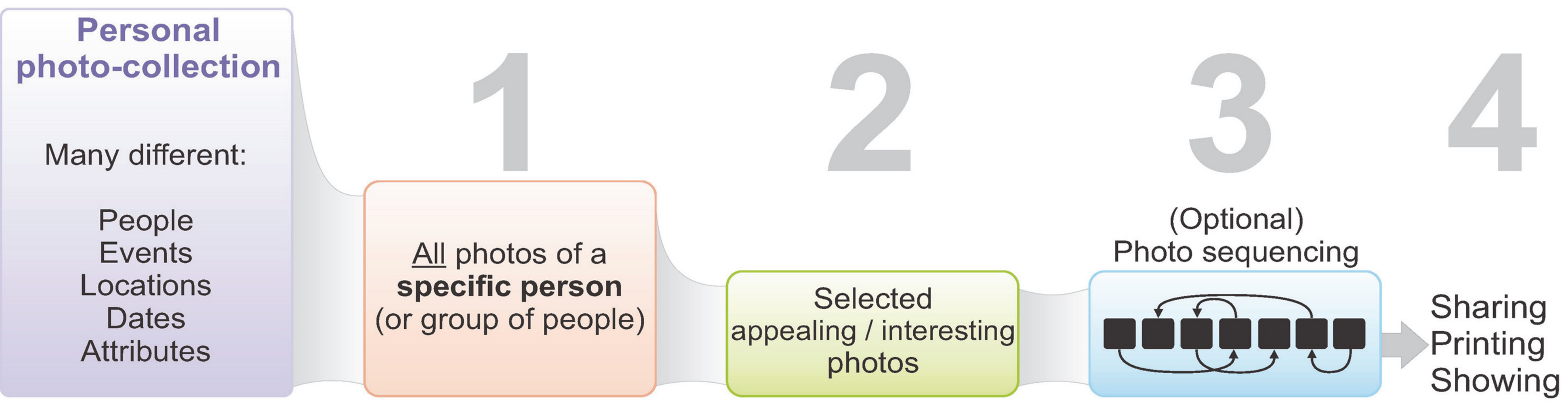
PROBABILISTIC APPROACH TO PEOPLE-CENTRIC PHOTO SELECTION AND SEQUENCING

Vassilios Vonikakis, Ramanathan Subramanian, Jonas T. Arnfred, and Stefan Winkler
Advanced Digital Sciences Center (ADSC), University of Illinois at Urbana-Champaign, Singapore

Motivation

- Personal photo-libraries keep growing in size
- Traditional approaches to organizing photo collections (event-, location-, time-, or image attribute-based) are not ideal for people-centric collections, which are inherently different
- How to automate the photo selection and sequencing process in order to create a pleasing people-centric slideshow?



Contributions

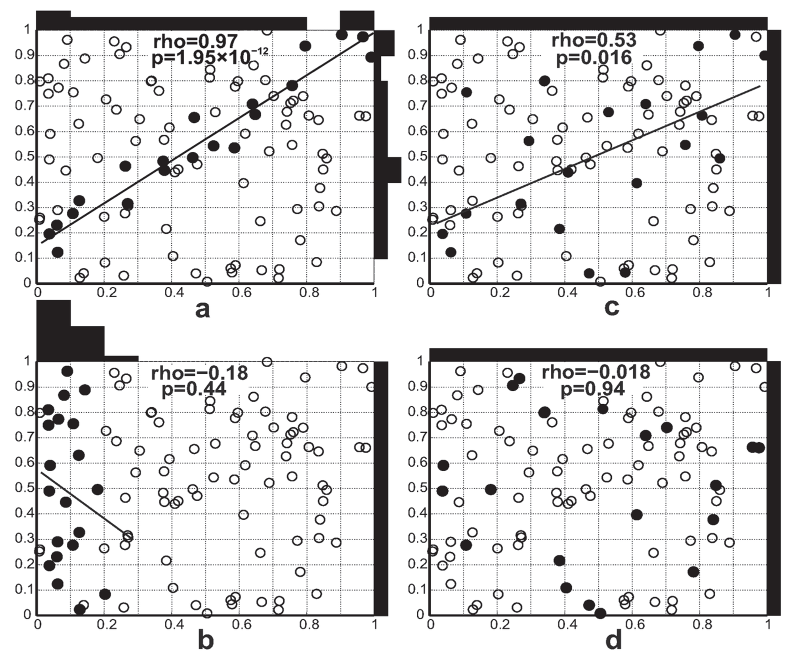
- Large-scale crowdsourcing study for discovering user preferences on people-centric photo summaries (first to include affect)
- Mixed Integer Linear Programming (MILP) dataset shaping technique to create compact, balanced, and representative datasets
- ILP-based technique to select and arrange images into an appealing slideshow, using probabilistic knowledge from crowdsourcing

Dataset Shaping (for Crowdsourcing)

Goal: Narrow down image sets for workers to manageable size, subject to the following constraints:

1. Maximize diversity in attributes
2. Minimize selection biases
3. Minimize correlations between attributes

Method: Mixed Integer Linear Programming (MILP) (details in paper)



Toy example for dataset shaping (selecting 20 out of 100 datapoints):
(a) Strong correlation between dimensions
(b) Underrepresented x-axis
(c) Enforcing uniform distribution
(d) Above + minimizing cross-correlations

Matlab source code available at http://tiny.cc/dshape

Crowdsourcing Study

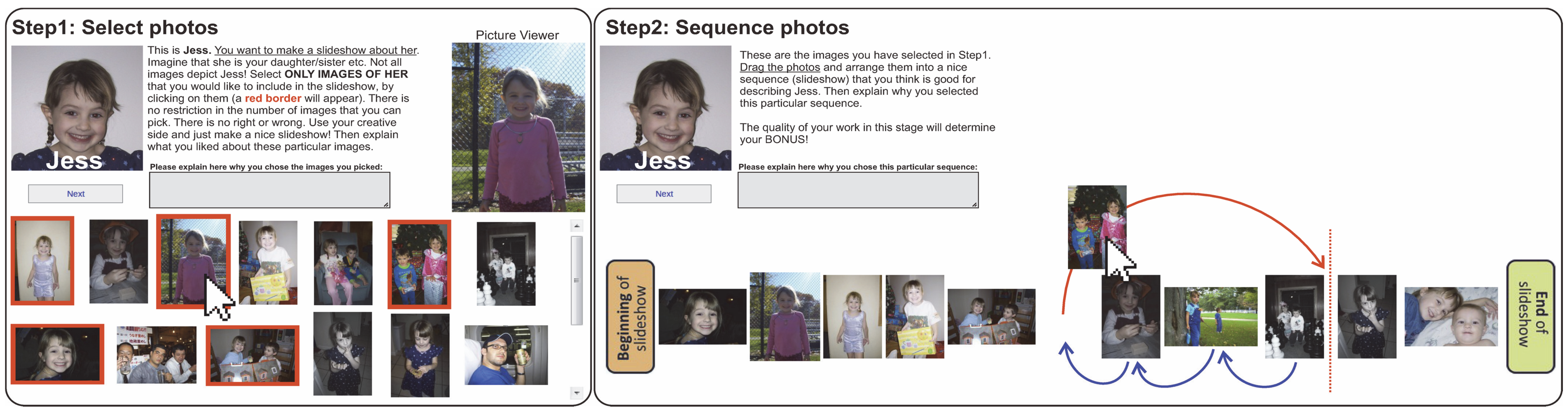
Datasets: 5 people-centric photo albums

1. Adult male (316 photos)
2. Adult female (381 photos)
3. Couple (281 photos)
4. Girl (278 photos), from Gallagher dataset [21]
5. Baby (177 photos), from Gallagher dataset [21]

Dataset shaping to obtain 5x3 balanced subsets with 40, 60 or 80 images each

Workers: 465 English-speaking workers

1. Selection Task
2. Sequencing Task



User interface for crowdsourcing study

VISUAL ATTRIBUTES USED IN THE CS EXPERIMENT.

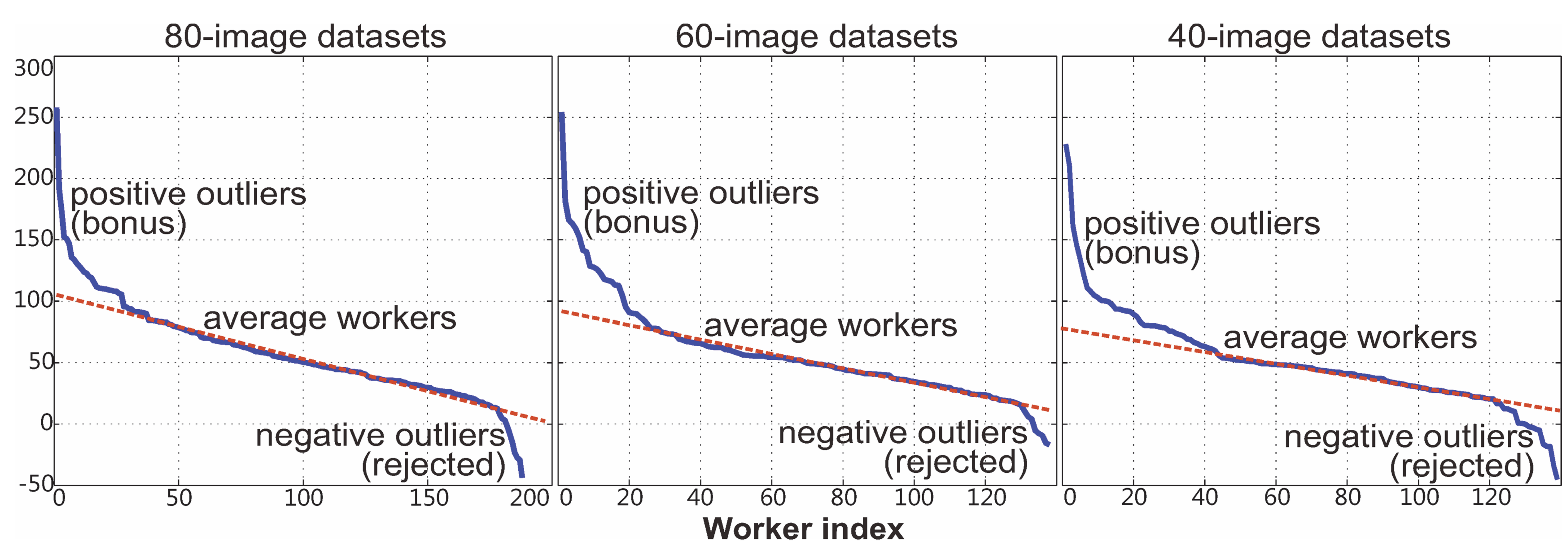
	Attribute	Indicates	Implementation
People-centric	Face count	Group or individual	Viola-Jones frontal face detector along with skin detection to eliminate false-positives [50]
	Face size	Type of shot: close-up, full-body, long distance	Ratio of face bounding box size to image size
	Face exposure	Face exposure	Mean luminance within bounding box
	Face composition	Aesthetic impression	Minimum distance from the face center to the 5 power-points in the image (4 for the *rule of thirds* and 1 for the center of the image) [42]
	Face yaw	Frontal / profile	Yaw of the head pose as estimated by the IntraFace library [51]
	Face valence	Pleasant / unpleasant facial expression [-1,1]	Regressor trained on Radboud images with VA annotations [47] with geometric features extracted on 49 facial points detected by IntraFace [51]
	Face intensity	Neutral / apex [0,1]	Regressor trained on Radboud images with IN annotations [47] with the same geometric features as for VA
Image-centric	Capturing Period	Primacy/recency, age of person	EXIF timestamps in conjunction with character's age (only for *male*, *female*, *couple* albums)
	Scene type	Indoorness/outdoorness	Ranking function based on *Relative Attributes* [52], using gist and color histograms
	Sharpness	Level of fine details	Computed similar to [53]
	Exposure	Overall exposure	Mean value of the luminance component
	Contrast	Overall contrast	Coefficient of variation (σ/μ) for the luminance channel
	Colorfulness	Color vividness	Computed similar to [54]

Worker Evaluation & Screening

- Microtask: filter distractor images (facial recognition)
- Reliability score relative to other workers

a_j	Description	Sign of w	Indicates a worker...
a_1	Ratio of selected distractor photos over selected album photos	w_1: –	paid attention to the task (engagement)
a_2	Comments regarding selection (number of characters)	w_2: +	had a specific reason for his/her selections
a_3	Length of comments – sequencing (number of characters)	w_3: +	had a specific reason for his/her sequencing
a_4	Number of selected images	w_4: +	was engaged in the task
a_5	Time spent selecting images (s)	w_5: +	was not rushing
a_6	Time spent arranging images (s)	w_6: +	was not rushing
a_7	Time spent on instructions (s)	w_7: +	was not rushing
a_8	Time spent on questions (s)	w_8: +	was not rushing
a_9	Number of clicks in sequencing	w_9: +	was diligent in this task
a_{10}	Not owning any cameras (binary)	w_{10}: –	had no photo-experience
a_{11}	Number of wrong answers (binary)	w_{11}: –	paid attention to questions
a_{12}	Contradictory answers (binary)	w_{12}: –	paid attention to questions

Metadata for evaluating workers, ordered by weight



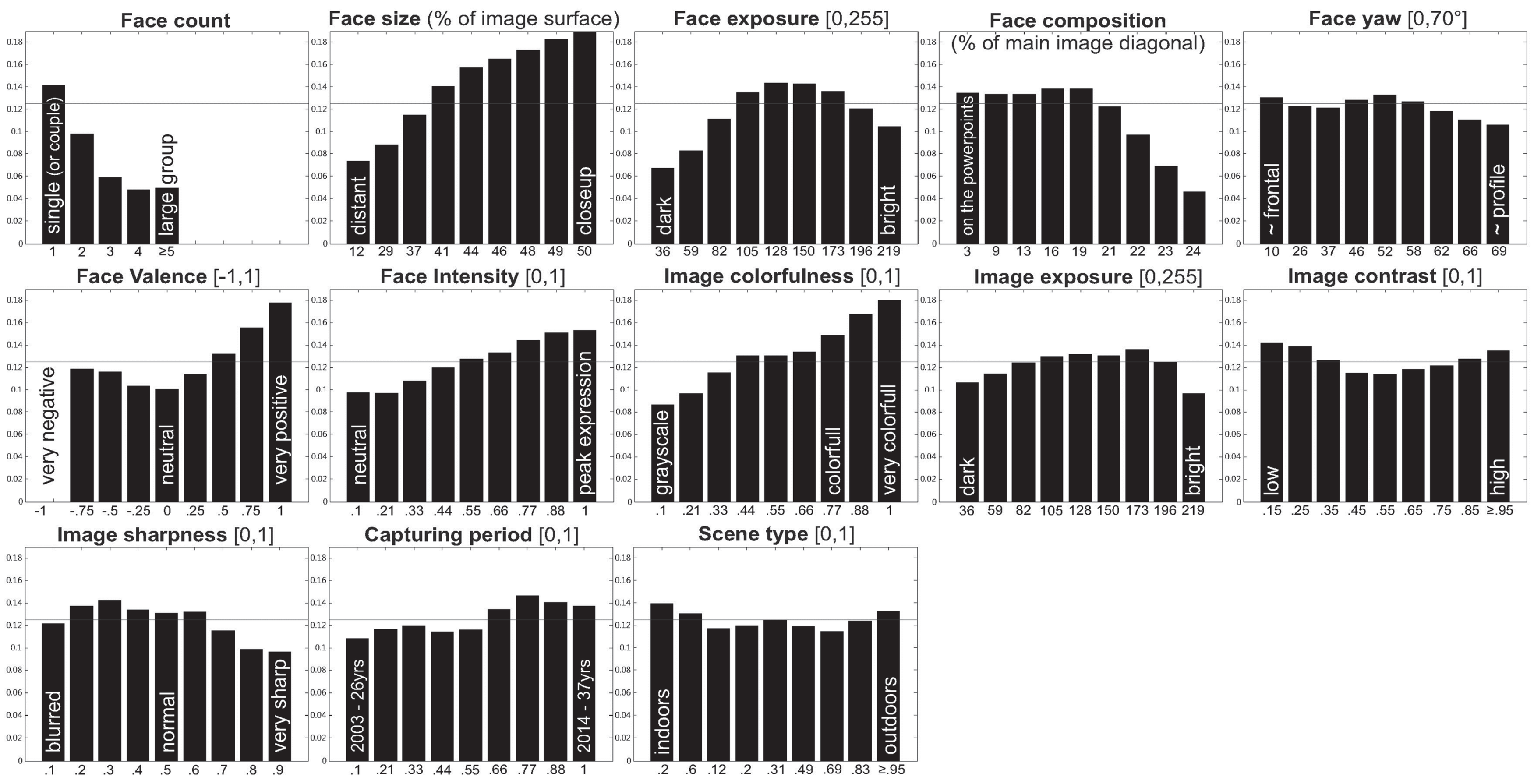
Worker evaluation based on relative reliability (~10% of workers were rejected)

Probabilistic Approach to People-Centric Photo Selection and Sequencing

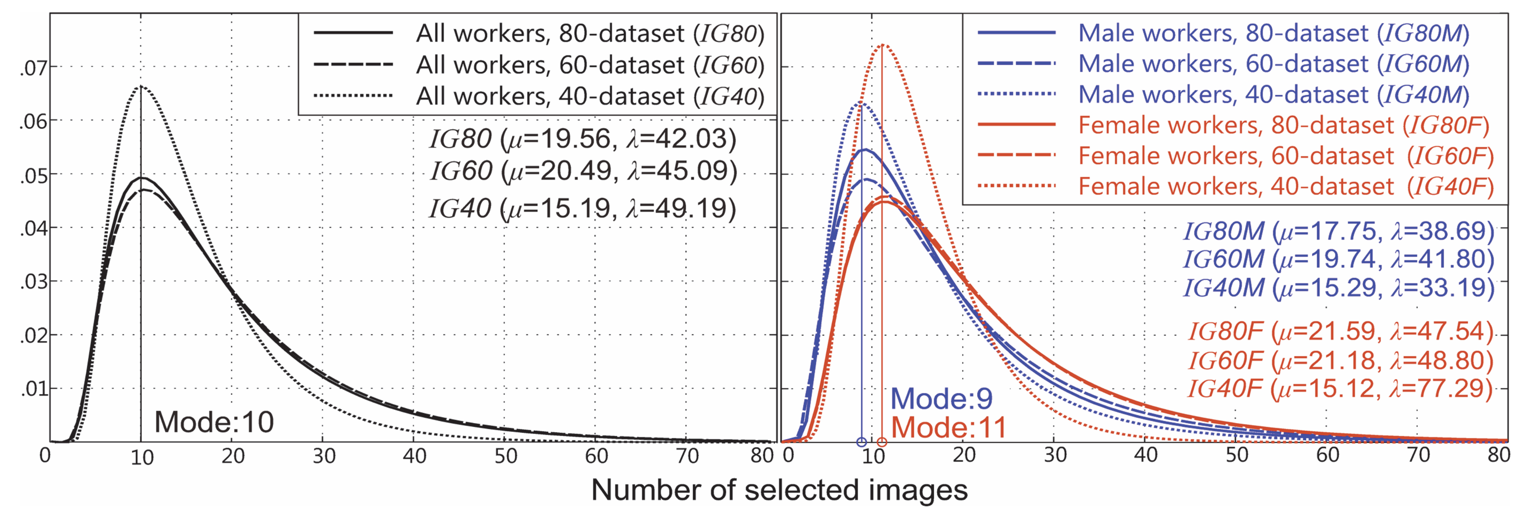
Vassilios Vonikakis, Ramanathan Subramanian, Jonas T. Arnfred, and Stefan Winkler
Advanced Digital Sciences Center (ADSC), University of Illinois at Urbana-Champaign, Singapore

Crowdsourcing Results

- Estimates of image selection probability and relative image position
- Workers typically chose 10 photos for a slideshow
- People-centric (facial) attributes have large influence
- Differences between male and female workers for certain attributes



Crowdsourcing results: marginal selection likelihood across all 5 albums



Predicting selection probability (image appeal) for 3 albums from Gallagher dataset

Automatic Image Selection & Sequencing

Probability of selecting image j for position i in slideshow:

$$e_{ij} = \prod_{m=1}^{M} P_j\left(sel = 1 \mid a_{jm} = k, t = d(i)\right) = \prod_{m=1}^{M} p^{m}_{kd(i)}$$

Maximize $\sum_{i=1}^{N} e_{i\hat{s}(i)}$

Combinatorial problem, solve via Integer Linear Programming (ILP)

Evaluation via Benchmarking

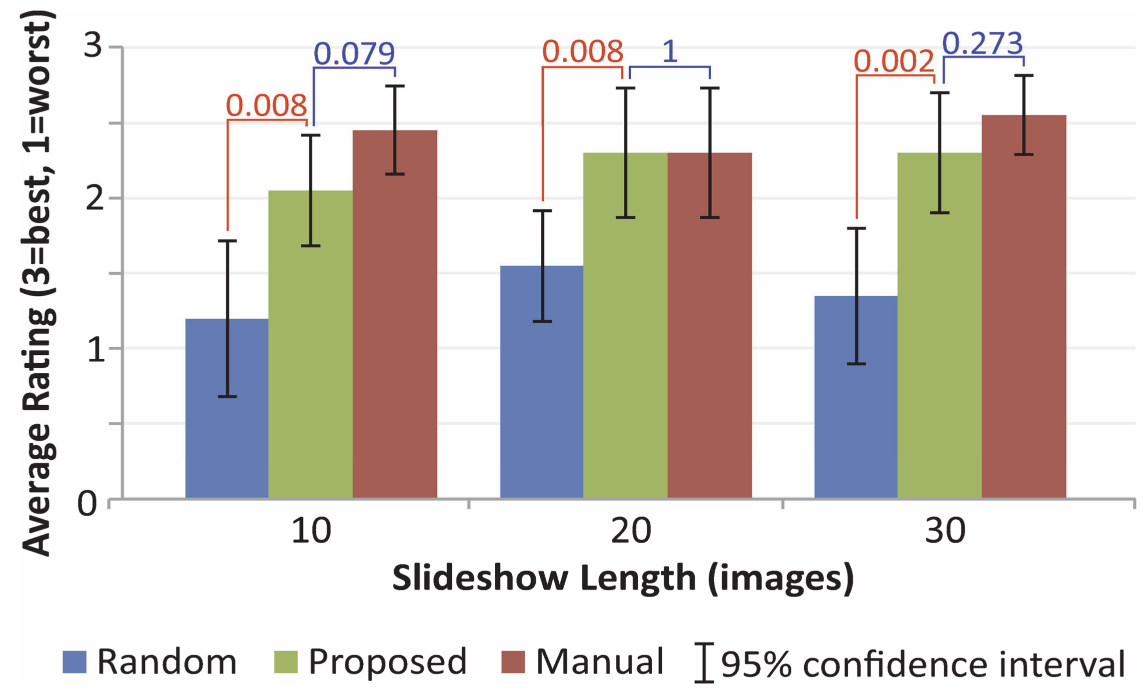
- Album-based cross-validation for selection and sequencing
- Comparison against greedy algorithm and regression approach
- Kendall's τ (rank correlation coefficient)
- Image sequencing is more challenging than selection
- People-centric attributes are critical

Evaluation via User Study

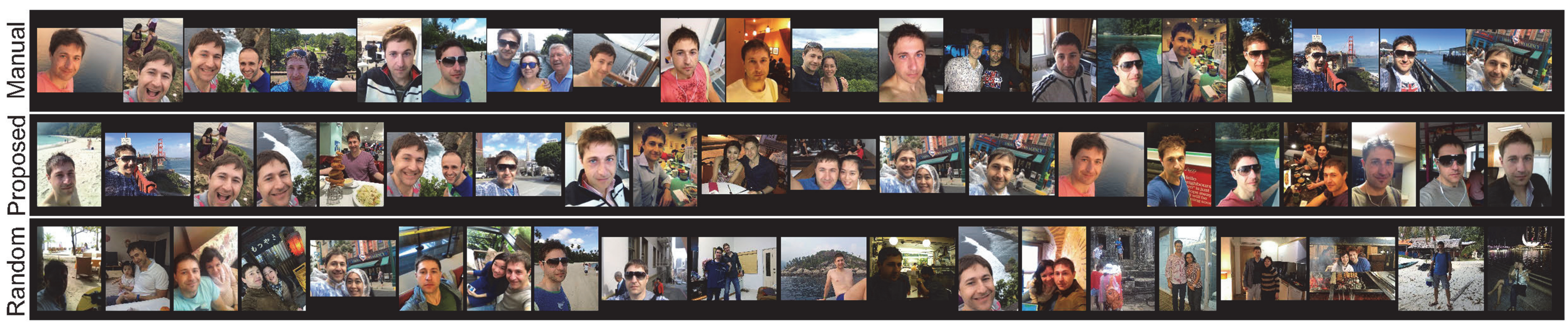
- User study on photolibrary with 715 images
- Slideshows created with 10/20/30 images
- Using manual (expert), random, and proposed method
- 15 users rated quality of 9 slideshows

Performance comparison for image sequencing (rank correlation mean and std)

Methods		Model tested on: Baby	Girl	Adult male	Adult female	Couple	Mean
Image attributes (6)	Regression	0.0866 (0.2175)	-0.0097 (0.1937)	0.0130 (0.3009)	-0.0015 (0.2422)	-0.0553 (0.2832)	0.0066 (0.2475)
	Greedy	0.0135 (0.2519)	0.0006 (0.2974)	-0.0500 (0.1942)	-0.0152 (0.2101)	-0.0341 (0.2272)	-0.0171 (0.2362)
	Proposed	0.1079 (0.2675)	0.0358 (0.2325)	0.0154 (0.2325)	0.0032 (0.2525)	-0.0217 (0.2477)	0.0282 (0.2466)
Image + people attributes (13)	Regression	0.0094 (0.2691)	-0.0265 (0.2389)	0.0125 (0.2837)	-0.0349 (0.2429)	-0.0996 (0.2656)	-0.0278 (0.2601)
	Greedy	0.0294 (0.2589)	0.0039 (0.2765)	0.0514 (0.2102)	0.0737 (0.2182)	0.0228 (0.2933)	0.0362 (0.2514)
	Proposed	**0.3602 (0.1927)**	**0.3046 (0.1549)**	**0.3230 (0.1797)**	**0.3813 (0.1890)**	**0.3345 (0.1620)**	**0.3407 (0.1757)**



User study results



Comparing the output of our proposed ILP-based slideshow creation method (middle row) to manual (top row) and random (bottom row) selection & sequencing

Conclusions

- Extensive crowdsourcing study on image selection and sequencing for people-centric slideshows from personal photolibraries
- ILP-based slideshow creation method outperforms others and creates slideshows similar to human experts
- People-centric attributes play a critical role in estimating selection and sequencing preferences
- Probabilities can also be learned for individual users, resulting in personalized slideshows

Image selection probabilities, Matlab source code for dataset shaping and slideshow creation available at http://tiny.cc/dshape