Spatial Retargeting

Diego Gutierrez
Universidad de Zaragoza

(slides material also from Miki Rubinstein, Olga Sorkine, Arik Shamir and Susana Castillo)

The Retargeting Problem

Common solutions

- Homogeneous squeezing/stretching
- Cropping [DeCarlo and Santella 2002; Viola and Jones 2004…]
- Hybrid solution [modern TV sets]


original


squeeze


crop


hybrid

Visual Media Retargeting: Siggraph Asia Course 2009

Ariel Shamir

The Interdisciplinary Center, Herzliya

Olga Sorkine

New York Univeristy

Visual Media Retargeting: An Example

[Avidar & Shamir 2007]

Visual Media Retargeting: Scaling

Scaling

[Avidar & Shamir 2007]

Visual Media Retargeting: Seams

Insert & remove seams

Scaling

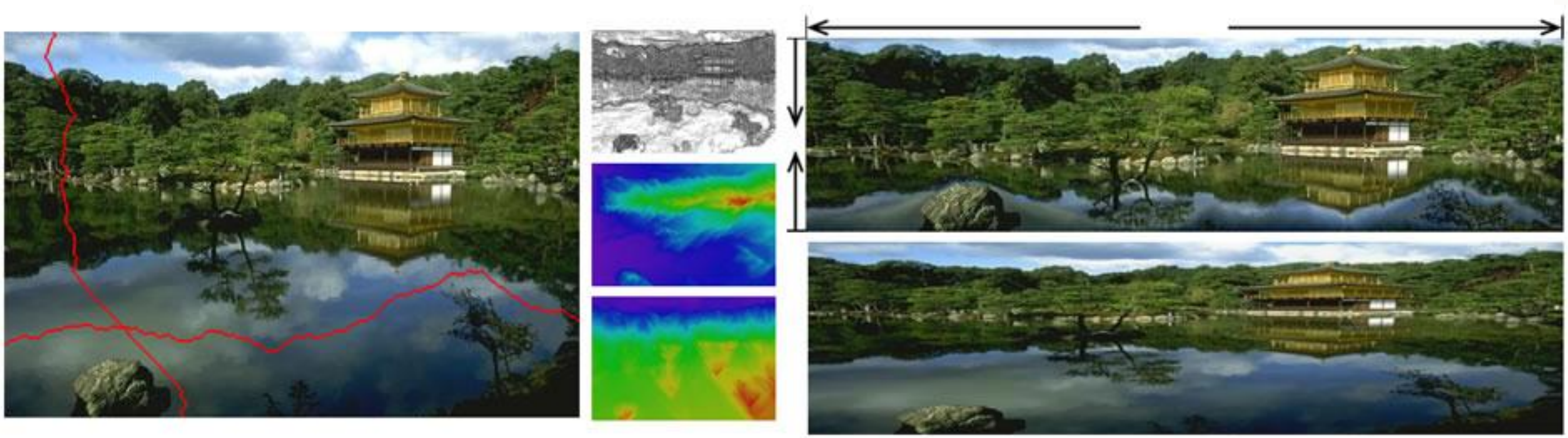
[Avidar & Shamir 2007]

Visual Media Retargeting: Energy Concept

1. Define an energy function **E**(**I**) (interest, importance, saliency…)

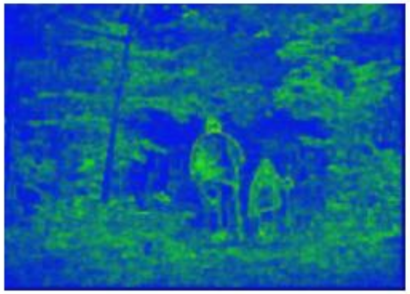
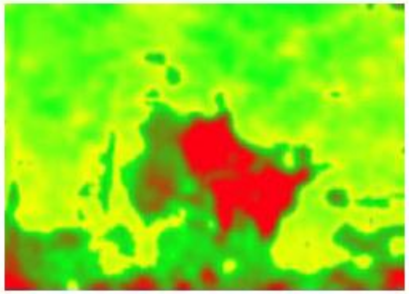
→

2. Use some operator(s) to change the image **I**



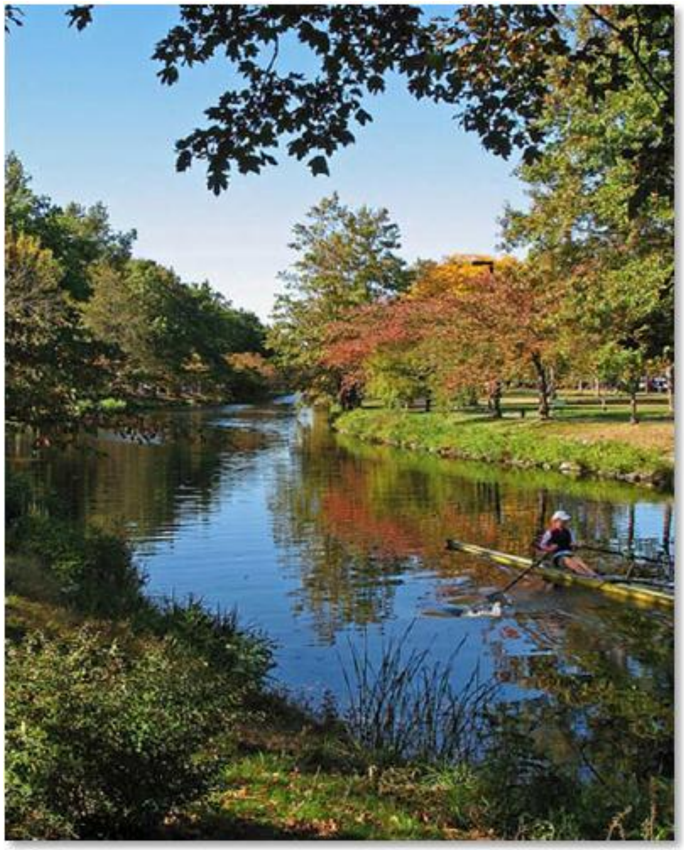
[Avidar & Shamir 2007]

- Magnitude of gradients (simple)
- Saliency (e.g. Itty's method) - multires

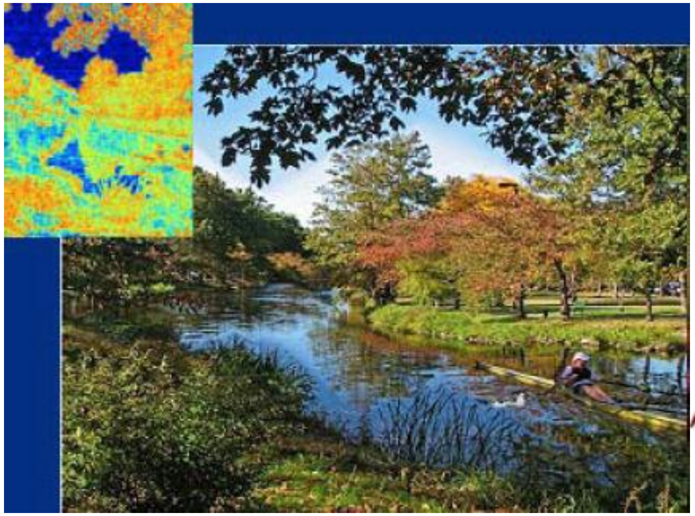




[Shamir and Sorkine 2009]

Different energy functions

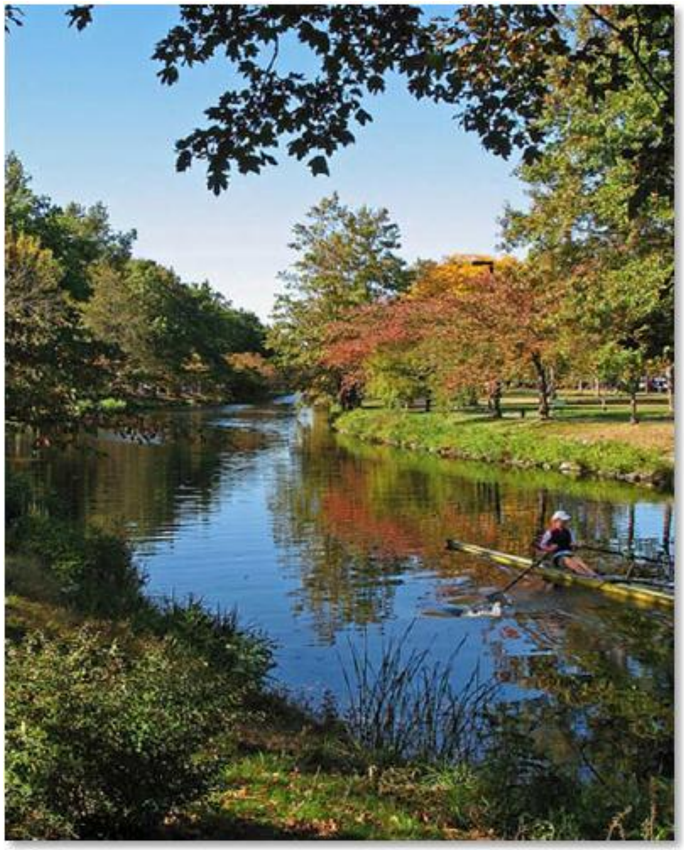


- •Histogram of Gradients
- •Entropy
- •E1
- •Mean Shift & E_1



[Shamir and Sorkine 2009]

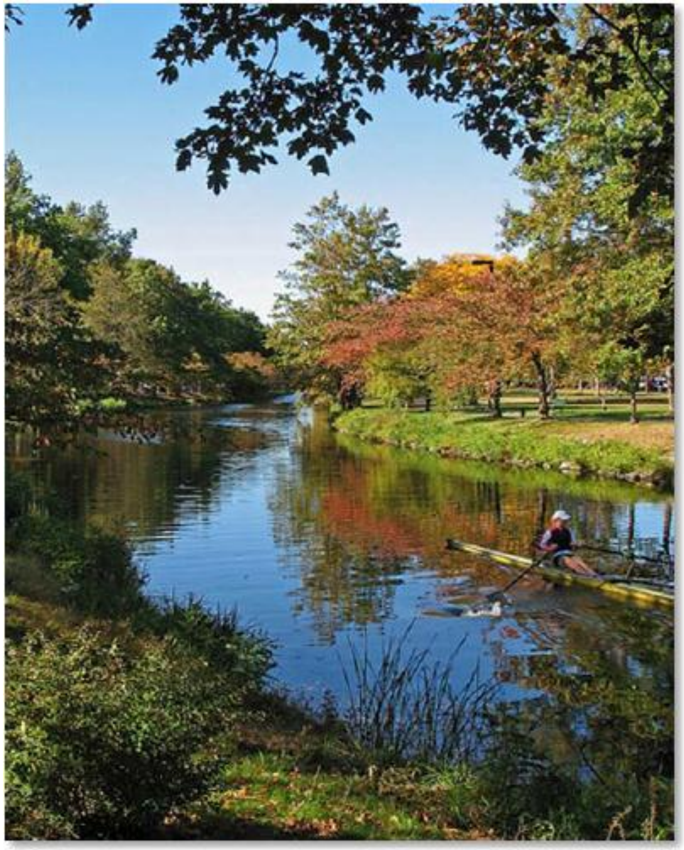
Different energy functions



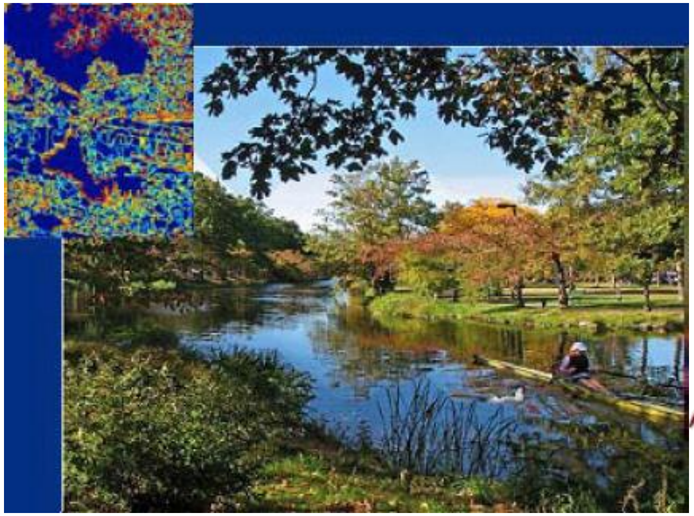
- Histogram of Gradients
- Entropy
- E1
- Mean Shift & E_1

[Shamir and Sorkine 2009]

Different energy functions

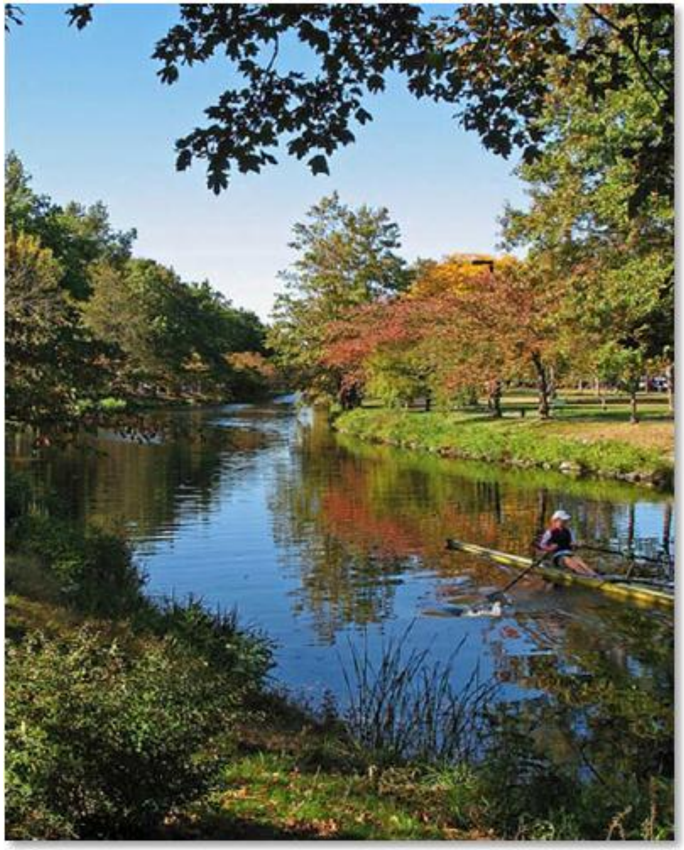


- Histogram of Gradients
- Entropy
- E1
- Mean Shift & E_1

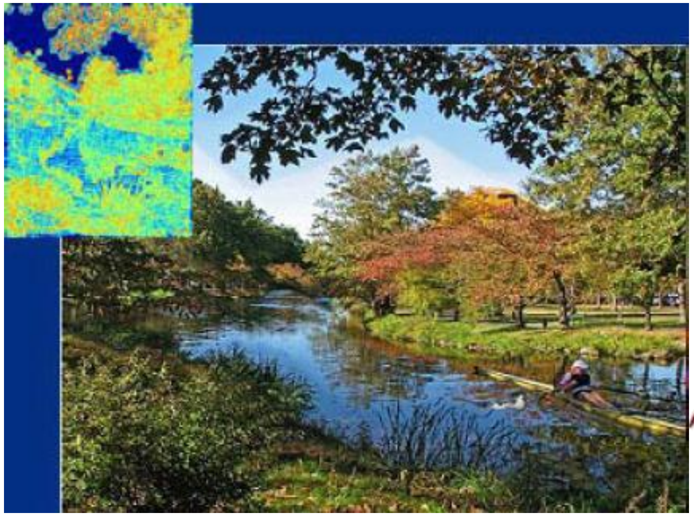


[Shamir and Sorkine 2009]

Different energy functions



- Histogram of Gradients
- Entropy
- E1
- Mean Shift & E_1



[Shamir and Sorkine 2009]

Simple operators: cropping

- Crop s.t. important (salient) parts remain
- Use domain-specific tools, such as face detector, gaze estimation… [DeCarlo and Santella 2002; Viola and Jones 2004]

original

crop

[Shamir and Sorkine 2009]

Simple operators: scaling

- Cam combine with cropping techniques (done on modern TV sets – center remains, peripheral data is scaled)
- Distorts content but is perfectly temporally coherent (video)

original squeeze hybrid

[Shamir and Sorkine 2009]

Discrete vs continuous

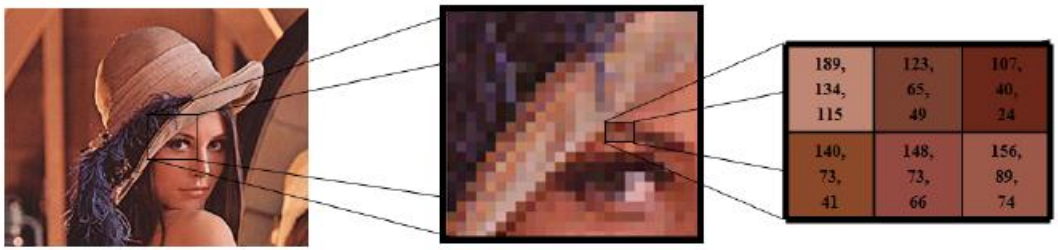


Figure 2: *A digital image as a 2D discrete grid of pixels. In this case the cells contain 3 values of RGB color.*

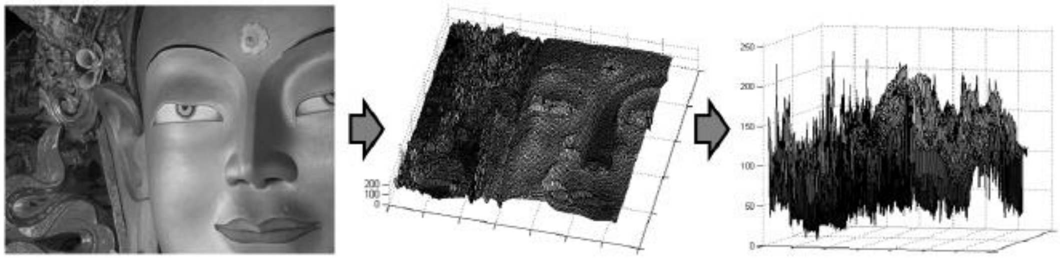


Figure 3: *A digital image as a sampling of a continuous function.*

[Shamir and Sorkine 2009]

Problem statement

- Given an image **I** of size (*n x m*), we want to produce an image **I*** of size (*n* x m**) which is a good representative of image **I**

- **But what is a “good representative”?** No definitions exist

- Goals of a retargeting algorithm:
  - 1. The important *content* of **I** should be preserved in **I***.
  - 2. The important *structure* of **I** should be preserved in **I***.
  - 3. **I*** should be *artifact*-free

Discrete approaches

- **Seam carving for content aware image resizing**
  SIGGRAPH 2007
  S. Avidan and A. Shamir
- **Improved seam carving for video retargeting**
  SIGGRAPH 2008
  M. Rubinstein, A. Shamir and S. Avidan
- **Seam carving for Media Retargeting**
  Trans. Of the ACM
  S. Avidan and A. Shamir
- **Multi-Operator Media Retargeting**
  SIGGRAPH 2009
  M. Rubinstein, A. Shamir and S. Avidan

[Shamir and Sorkine 2009]

Continuous approaches

- **Feature-aware textureing**

  EGSR 2006

  R. Gal, O. Sorkine and D. Cohen-Or
- **Non-homogeneous content-drive video retargeting**

  ICCV 2007

  L. Wolf, M Guttmann and D. Cohen-Or
- **Optimized scale-and-stretch for image resizing**

  SIGGRAPH ASIA 2008

  Y. Wang, C. Tai, O. Sorkine and T. Lee
- **Shrinkability maps for content-aware video resizing**

  Pacific Graphics 2008

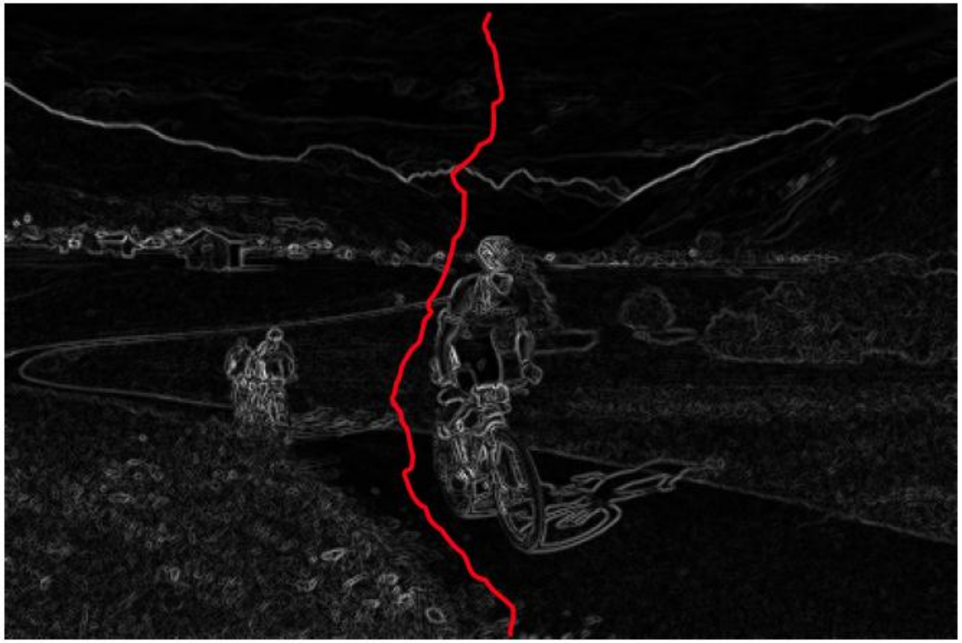
  Y. Zhang, S. Hu and R. Martin

[Shamir and Sorkine 2009]

Discrete example: Seam carving

[Rubinstein, Avidan and Shamir 2007]

Seam carving



[Rubinstein, Avidan and Shamir 2007]

Seam carving

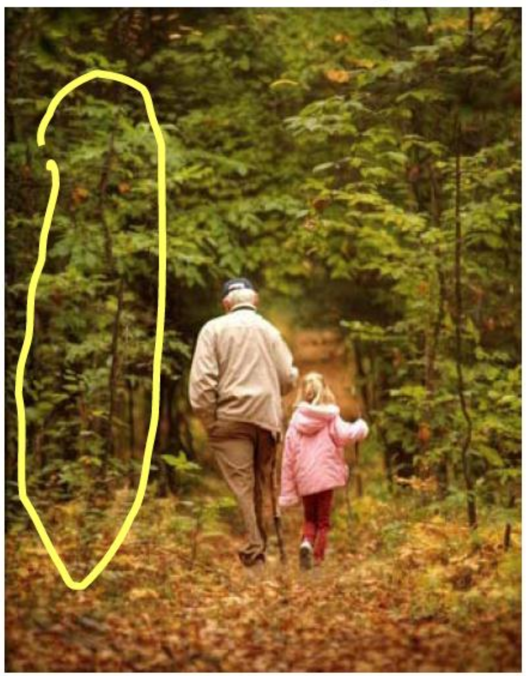
[Rubinstein, Avidan and Shamir 2007]

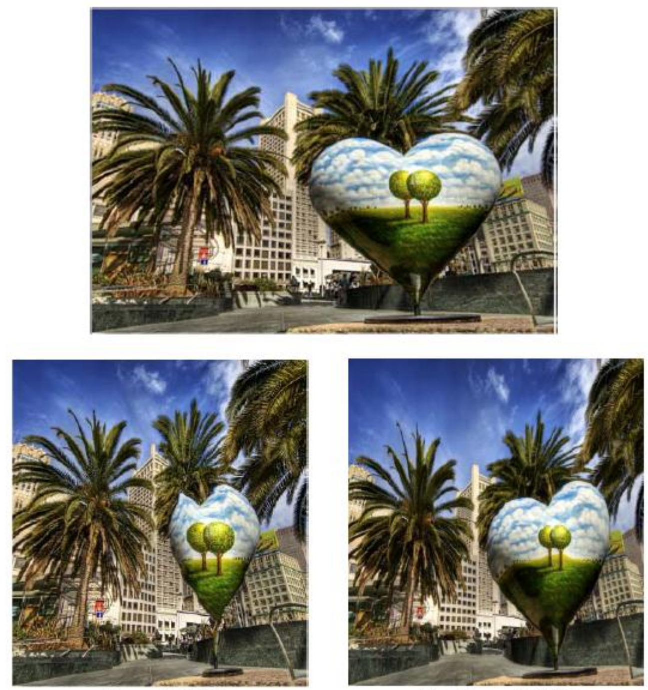
Seam carving

[Rubinstein, Avidan and Shamir 2007]

Seam carving: problems

- Discrete and greedy – may break structures
- Can run out of good seams in one direction





[Rubinstein, Avidan and Shamir 2007]

Continuous example: Warping

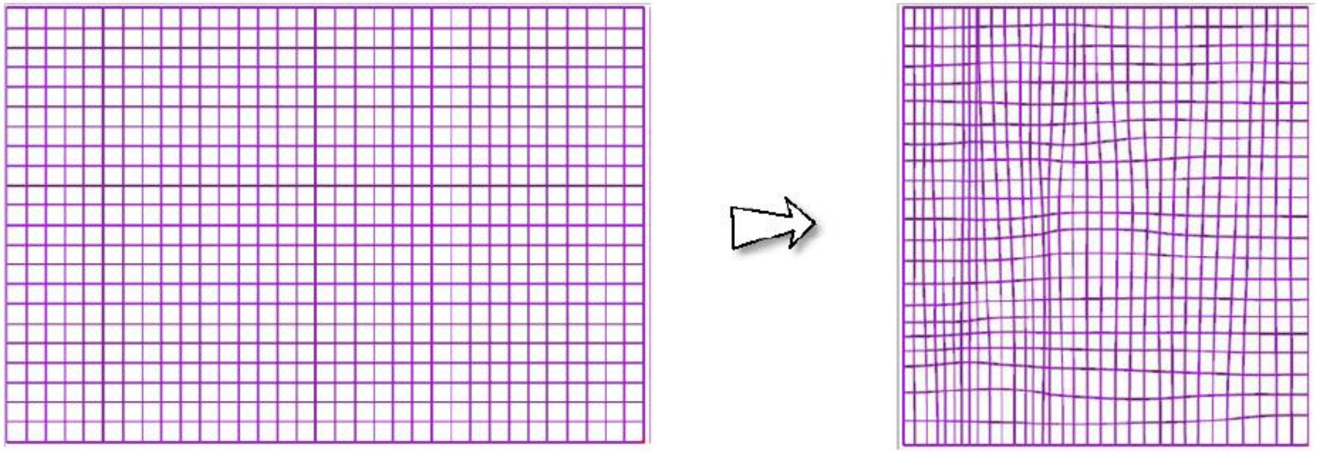
- Allow important regions to **uniformly scale**
- Find **optimal** local scaling factors by global optimization
- Result: preserve the **shape** of important regions, distort non-important ones

importance map

[Wang, Tai, Sorkine and Lee 2008]

Continuous example: Warping

- Grid mesh, preserve the shape of the important quads

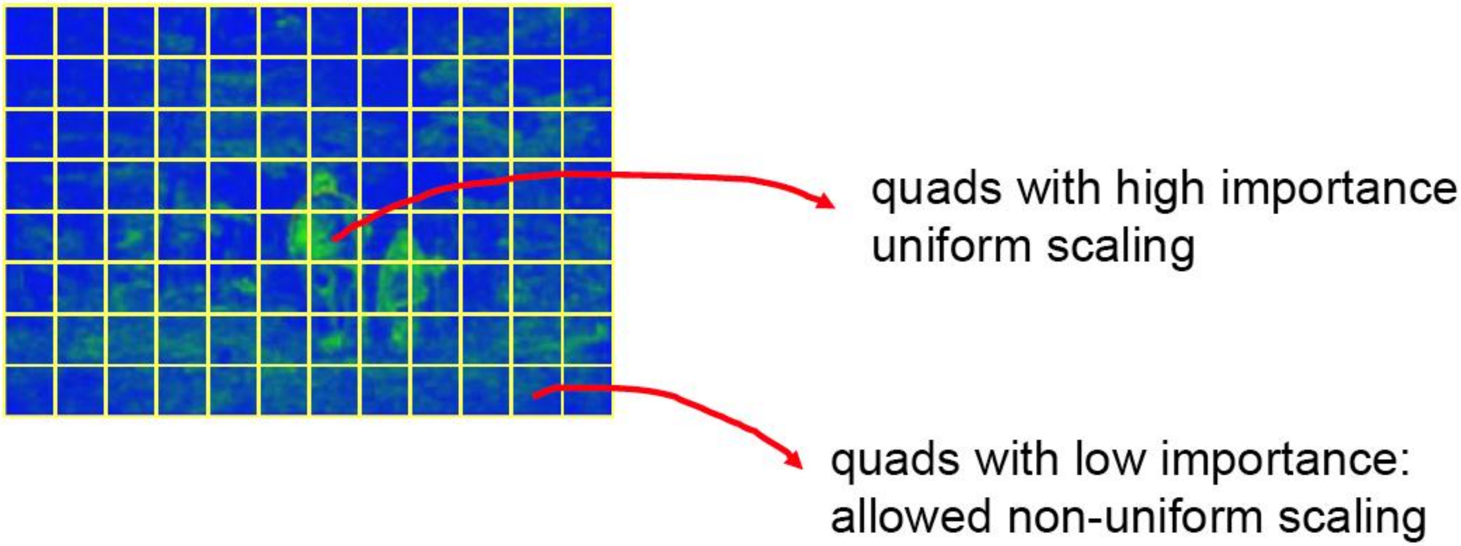


- Optimize the location of mesh vertices, interpolate image

[Wang, Tai, Sorkine and Lee 2008]

Continuous example: Warping

- Grid mesh, preserve the shape of the important quads



- Optimize the location of mesh vertices, interpolate image

[Wang, Tai, Sorkine and Lee 2008]

Video?

- Naïve… every frame by itself

[Shamir and Sorkine 2010]

Video?

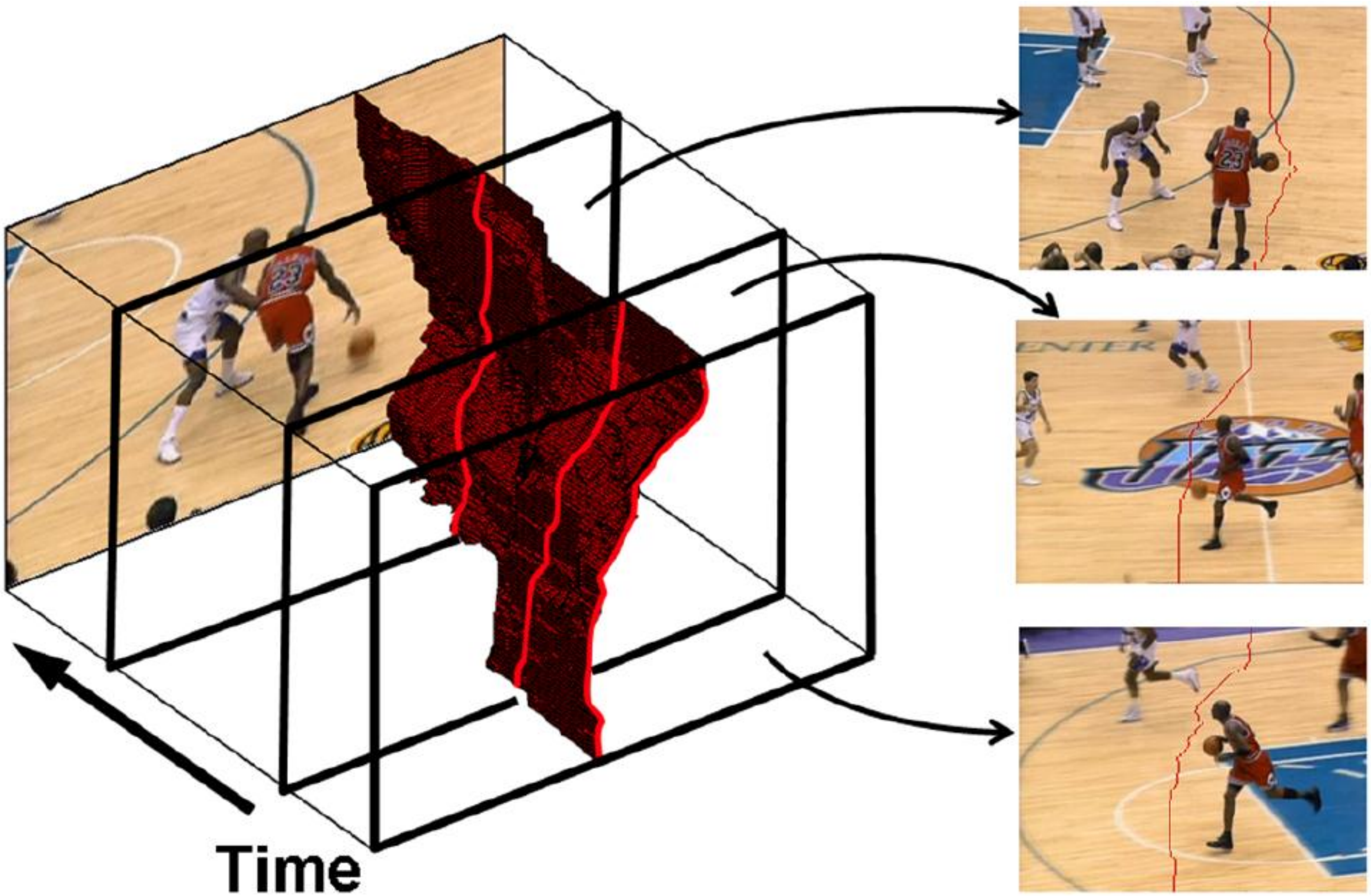
- Camera movement
- Object movement
- Seams should adapt and change through time!

- ➔ Global Solution (video cube)

[Shamir and Sorkine 2010]

SIGGRAPHASIA2011 HONG KONG

Video?



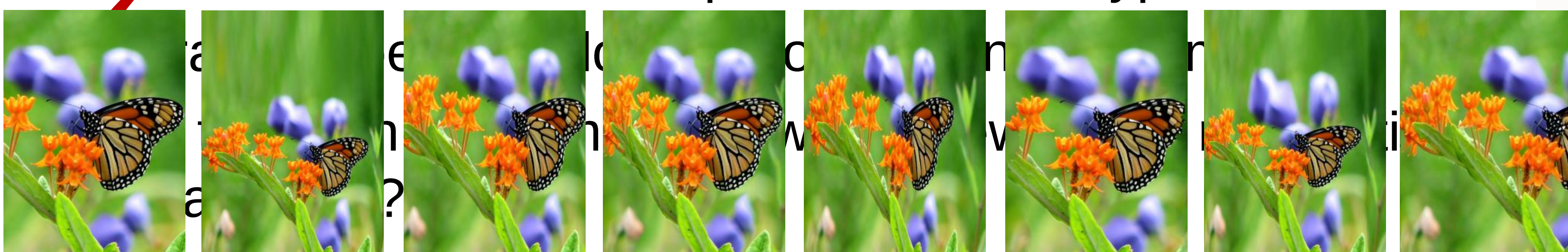
[Shamir and Sorkine 2010]

Current State of Retargeting Research

- No clear evaluation methodology!
  - Mostly visual comparison
  - Small subset of previous techniques

Source

- Relation between the operator and the type of content?



- Computational retargeting measure?

[Rubinstein, Gutierrez, Sorkine and Shamir 2010]

A Comparative Study of Image Retargeting

Miki Rubinstein, Diego Gutierrez, Olga Sorkine and Arik Shamir

ACM Transactions on Graphics, Vol. 29(5) (SIGGRAPH Asia 2010)

- Benchmark and evaluation methodology for image retargeting

RetargetMe

http://people.csail.mit.edu/mrub/retargetme/

- Comprehensive perceptual study and analysis of image retargeting

[Rubinstein, Gutierrez, Sorkine and Shamir 2010]

Goals

- What is the “correct” way to retarget this image?

[Rubinstein, Gutierrez, Sorkine and Shamir 2010]

Goals

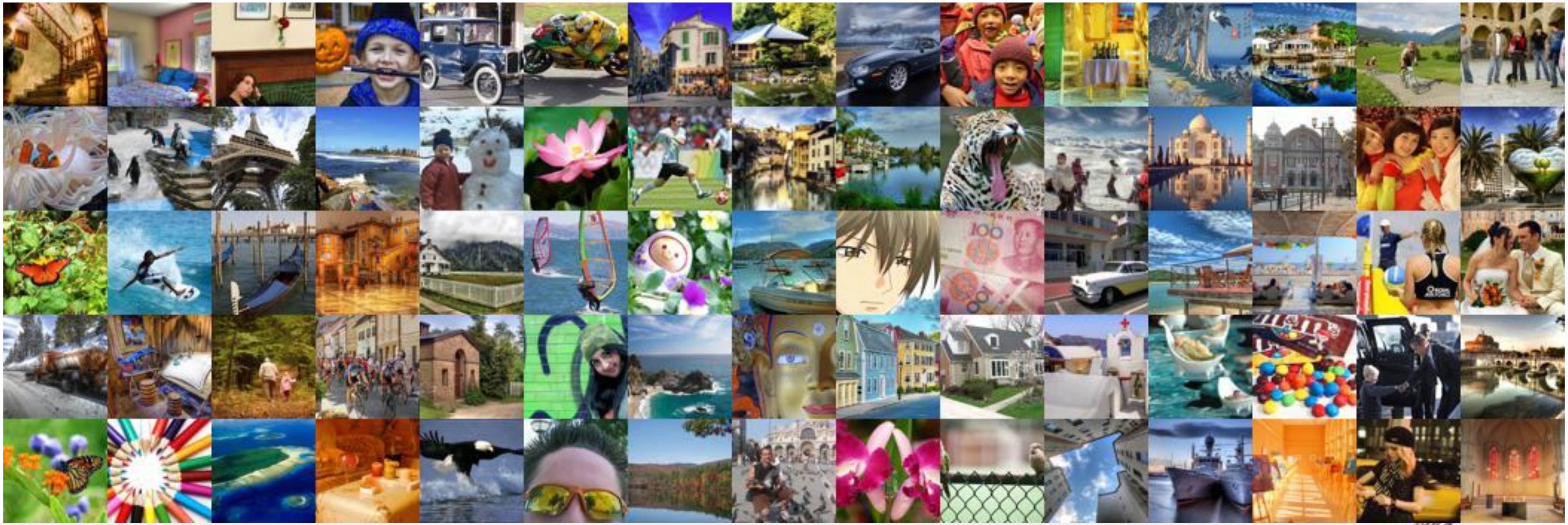
- The dataset and user study
- User response (subjective) analysis
  - Is there consensus between viewers?
  - When is one method better than another?
- Computational (objective) analysis
  - Can an image distance measure predict retargeting quality?

[Rubinstein, Gutierrez, Sorkine and Shamir 2010]

SIGGRAPHASIA2011 HONG KONG

Constructing the Dataset

- Image Retargeting objectives:
  1. Preserve the important ***content*** and ***structures***
  2. Limit ***artifacts***



[Rubinstein, Gutierrez, Sorkine and Shamir 2010]

Retargeting Operators

Seam Carving **[SC]**	[Rubinstein et al. 2008]	Discrete
Shift Map **[SM]**	[Pritch et al. 2009]	Discrete
Multi-Operator **[MULTIOP]**	[Rubinstein et al. 2009]	Discrete
Warping **[WARP]**	[Wolf et al. 2007]	Continuous
Streaming Video **[SV]**	[Krähenbühl et al. 2009]	Continuous
Scale-and-Stretch **[SNS]**	[Wang et al. 2008]	Continuous
Cropping **[CR]**	[Manual]	Reference
Scaling **[SCL]**	[Cubic interpolation]	Reference

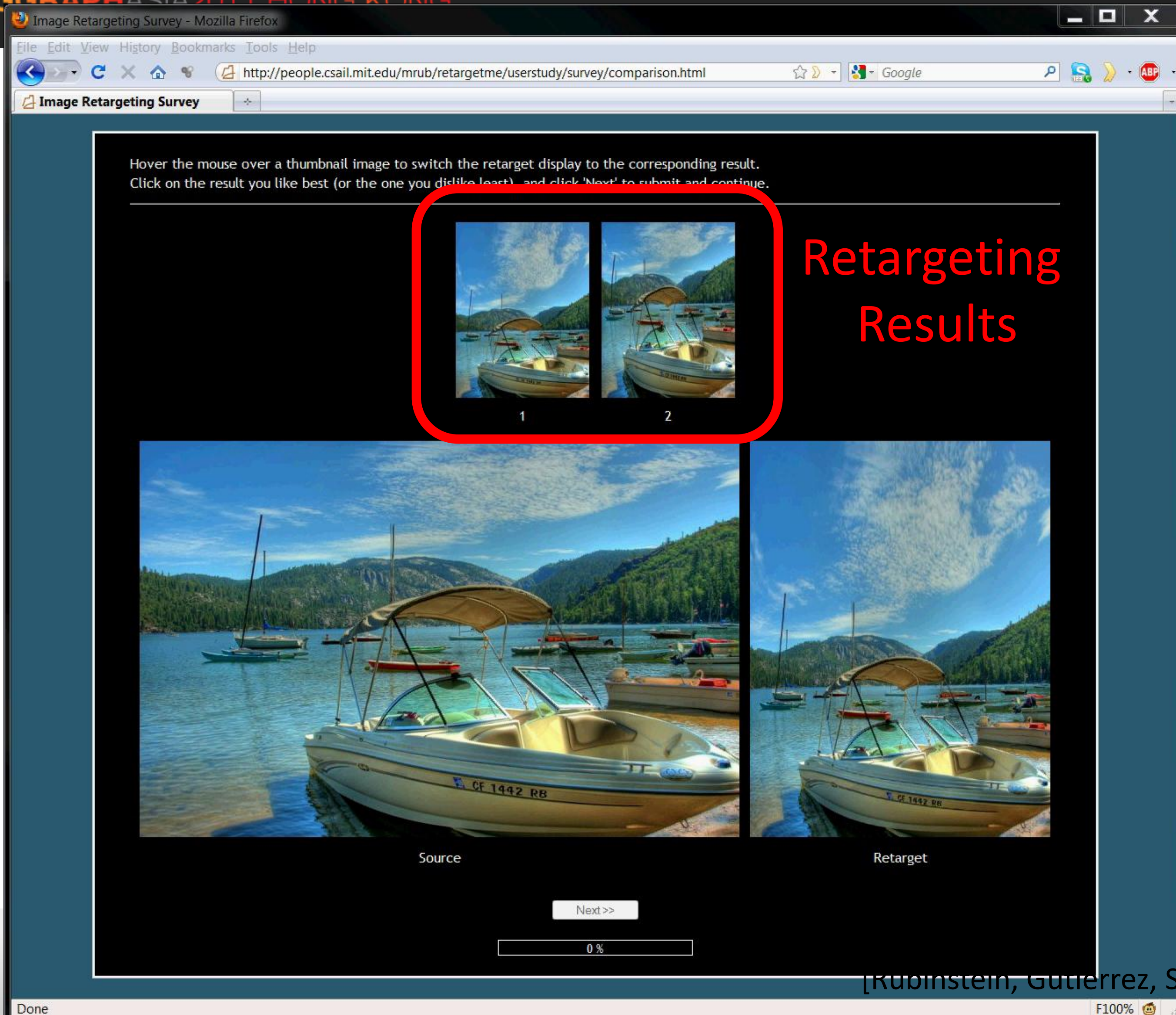
[Rubinstein, Gutierrez, Sorkine and Shamir 2010]

SIGGRAPH ASIA 2011 HONG KONG

Comparative Analysis

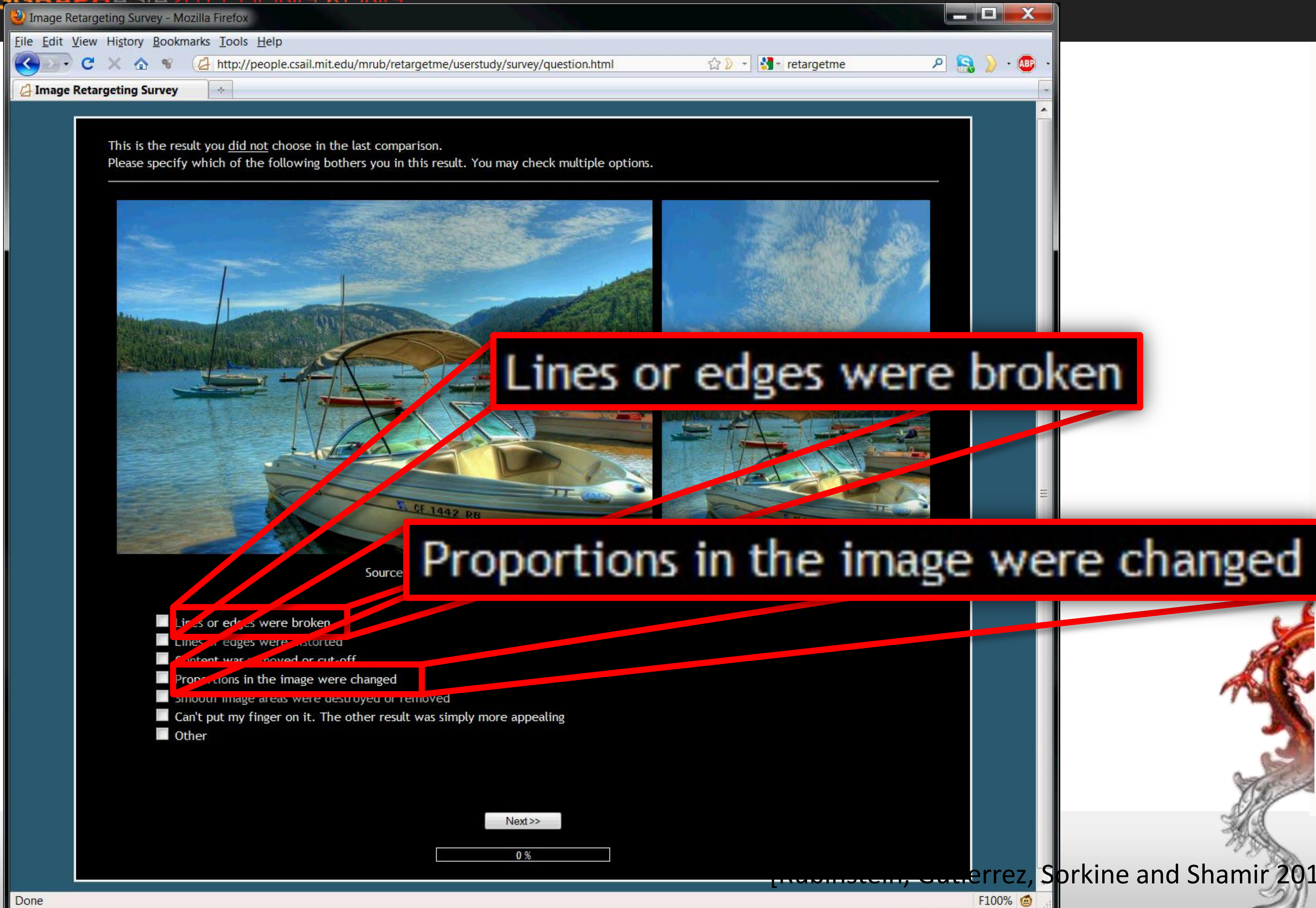
[Rubinstein, Gutierrez, Sorkine and Shamir 2010]

The Survey Interface



[Rubinstein, Gutierrez, Sorkine and Shamir 2010]

Additional Questions



[Rubinstein, Gutierrez, Sorkine and Shamir 2010]

User Statistics

- Each participant performs **12** comparisons over **5** images

- **210** participants; **252** votes per image
  - Half amazon mechanical turk Artificial Artificial Intelligence
  - Half (25 cents per completed survey)

- Average time to complete: 20 minutes

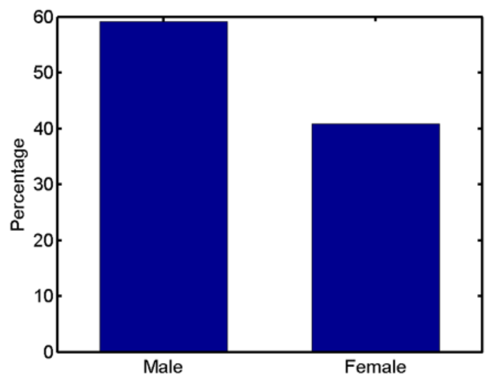
  “It was a very interesting survey. Very nice experience”

  “i need your more survey so that i can help u a lot”

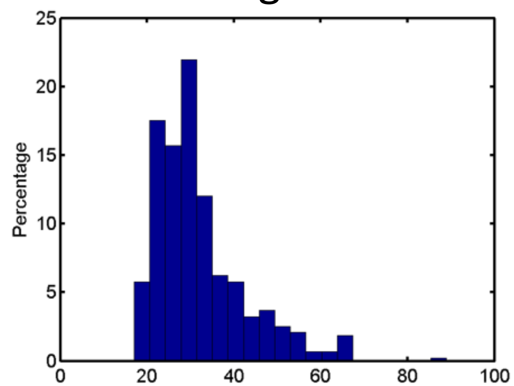
[Rubinstein, Gutierrez, Sorkine and Shamir 2010]

User Statistics

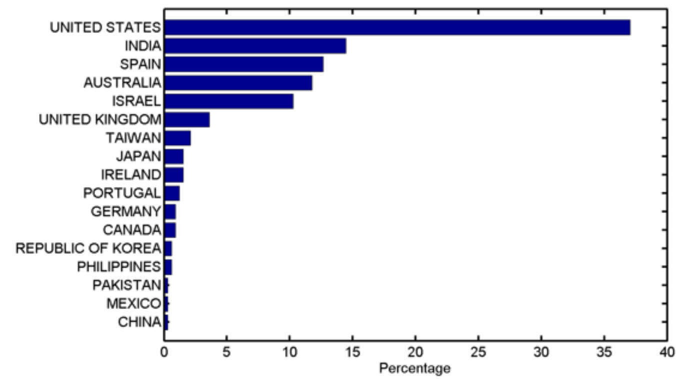
Gender



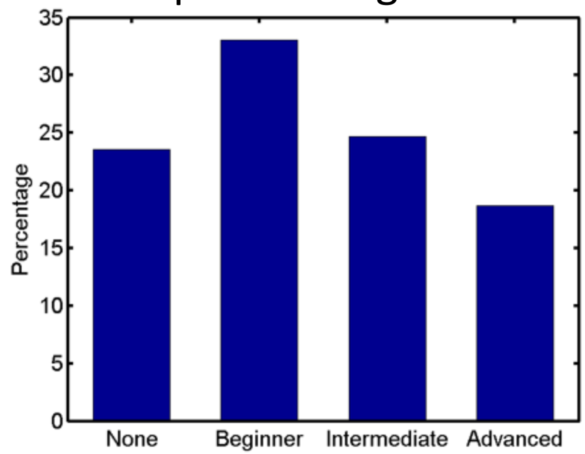
Age



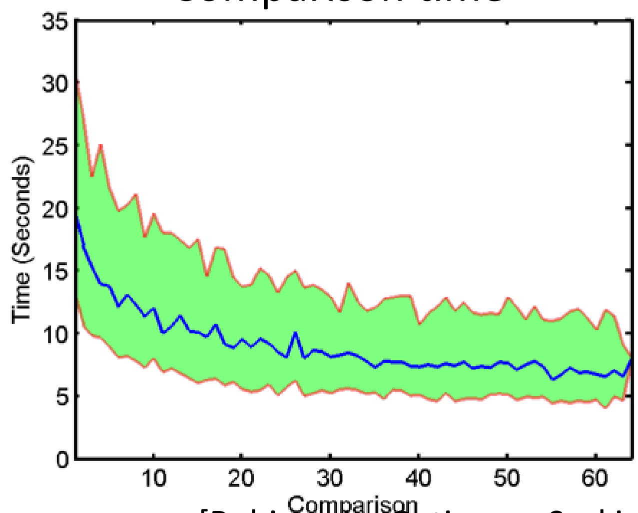
Country



Graphics Background



Comparison time



[Rubinstein, Gutierrez, Sorkine and Shamir 2010]

User Agreement

- Similarity of votes = consensus on "good" retargeting

- *Coefficient of Agreement* [Kendall 1940]

$$u = \frac{2\Sigma}{\binom{m}{2}\binom{t}{2}} - 1, \qquad \Sigma = \sum_{i=1}^{t}\sum_{j=1}^{t}\binom{a_{ij}}{2}$$

- a_{ij} = # times method i chosen over method j
- m = # participants
- t = 8 (# retargeting operators)
- $u \in [-\frac{1}{m}, 1]$

[Rubinstein, Gutierrez, Sorkine and Shamir 2010]

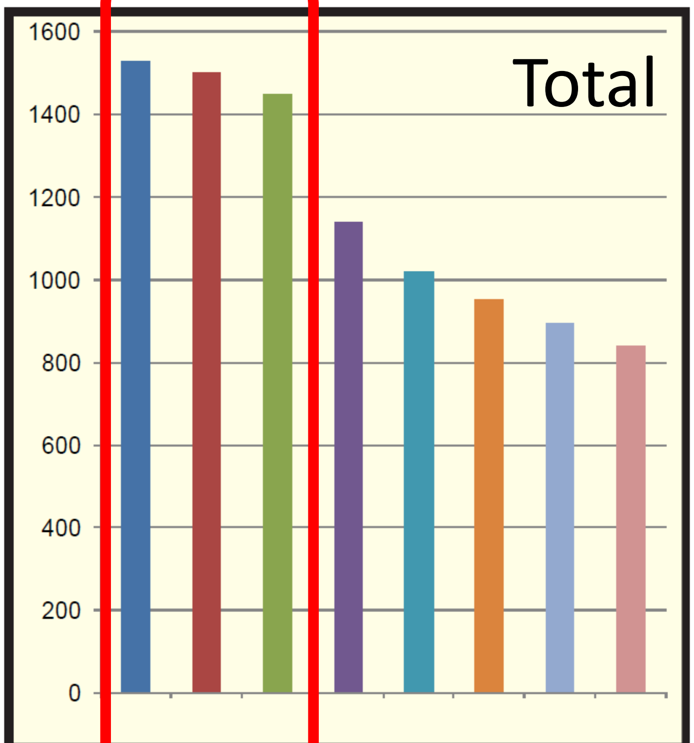
User Agreement

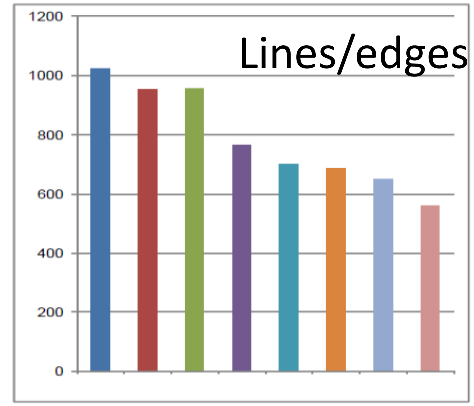
	lines/ edges	faces/ people	Textur e	foregroun d objects	Geometri c Structure s	Symmetr y	Total
u	0.073	**0.166**	0.070	**0.146**	0.084	**0.132**	**0.095**

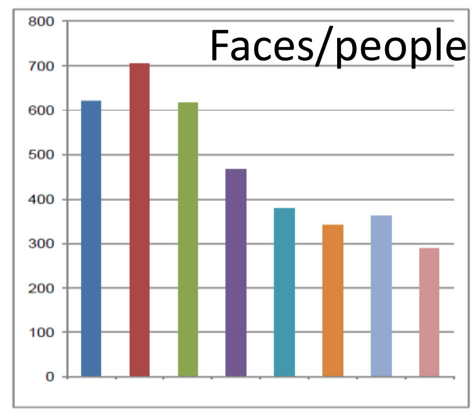
- Low agreement in general
- Greater agreement on images containing faces/people, evident foreground objects and symmetry.

[Rubinstein, Gutierrez, Sorkine and Shamir 2010]

Operator Ranking

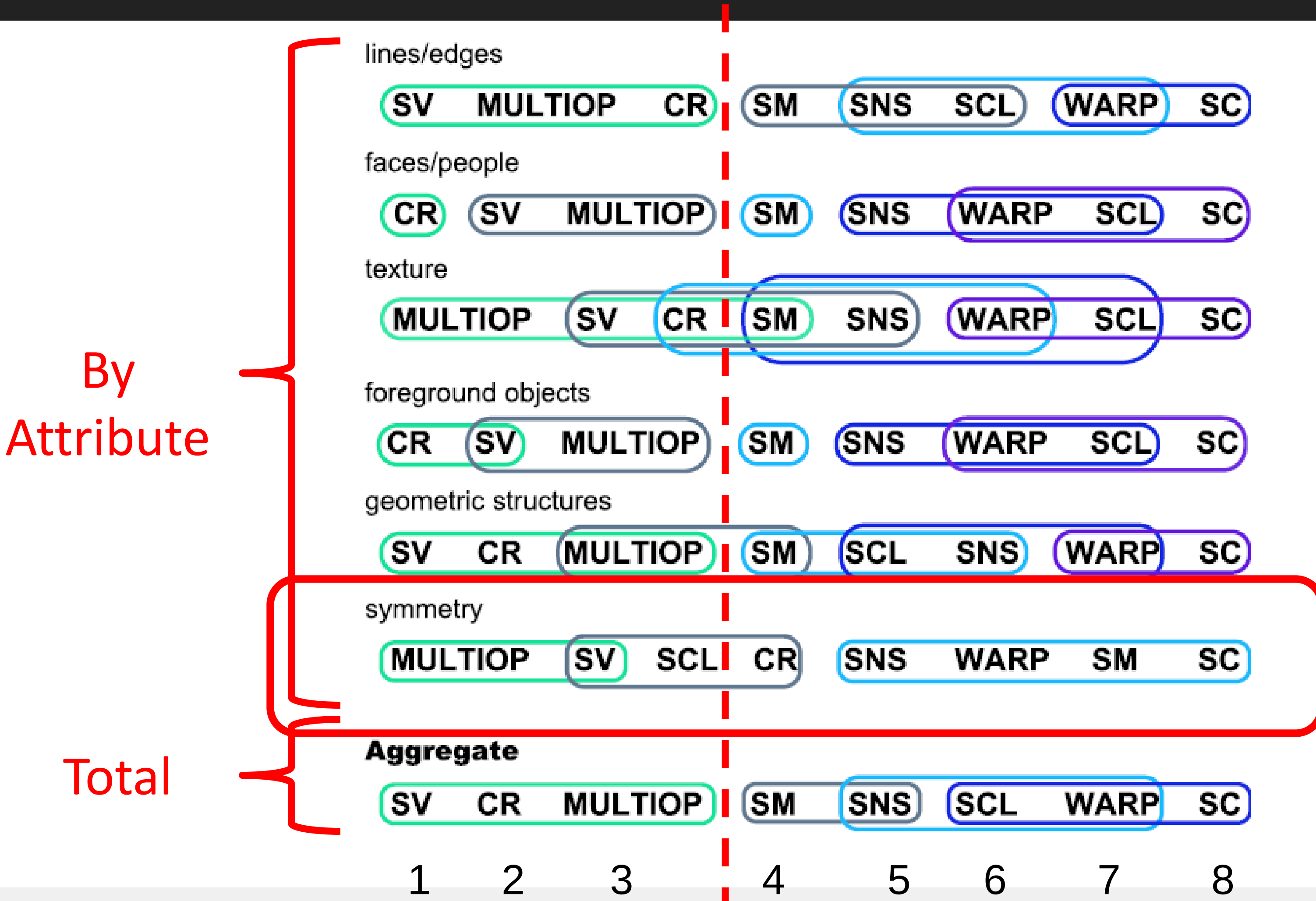






SV	CR	MULTIOP	SM	SNS	SCL	WARP	SC
Streaming Video	Cropping	Multi-operator	Shift-maps	Scale & Stretch	Scaling	Nonhomo. Warping	Seam Carving

Operator Ranking

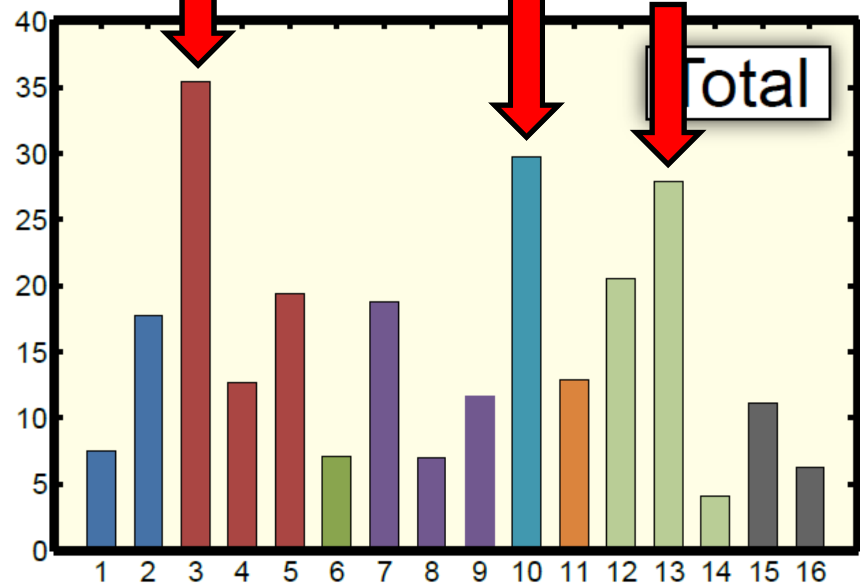


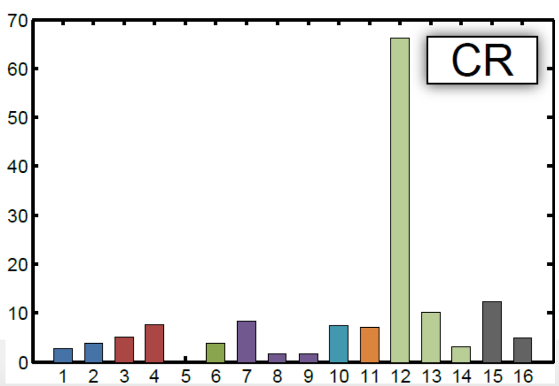
Rank

[Rubinstein, Gutierrez, Sorkine and Shamir 2010]

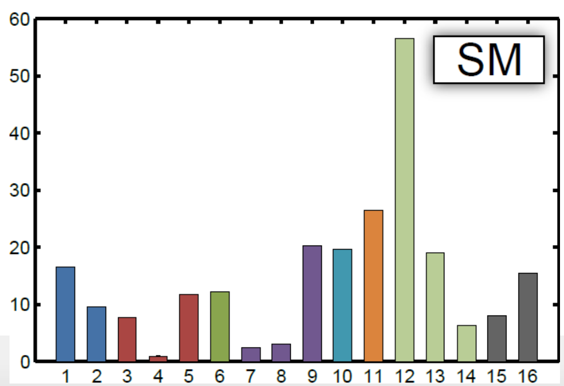
Additional Questions

Attribute	Reason	ID
lines/edges	Lines or edges were broken	1
lines/edges	Lines or edges were distorted	2
faces/people	People or faces were squeezed	3
faces/people	People or faces were stretched	4
faces/people	People or faces were deformed	5
texture	Textures were distorted	6
foreground objects	Foreground objects were squeezed	7
foreground objects	Foreground objects were stretched	8
foreground objects	Foreground objects were deformed	9
geometric structures	Geometric structures were distorted	10
symmetry	Symmetry was violated	11
Common	Content was removed or cut-off	12
Common	Proportions in the image were changed	13
Common	Smooth image areas were destroyed or removed	14
Common	Can't put my finger on it. The other result was simply more appealing	15
Common	Other	16

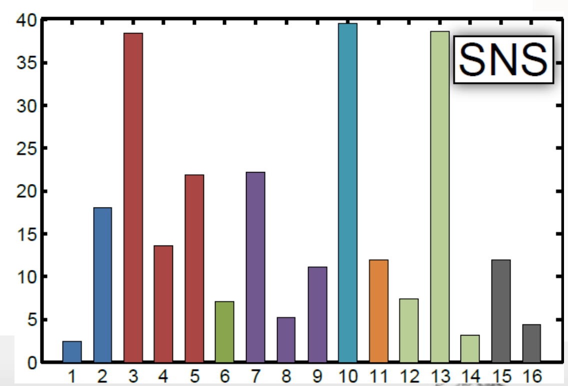




Cropping



Shift-maps



Scale & Stretch

Partial Conclusion

(At least for our retargeted setup)

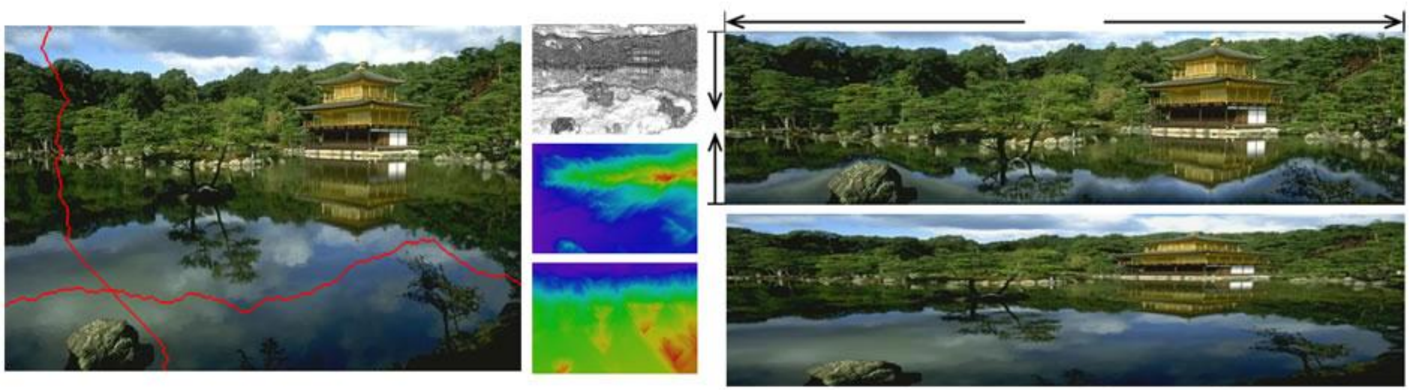
SUBJECTIVE:

Clear *and consistent* division in groups

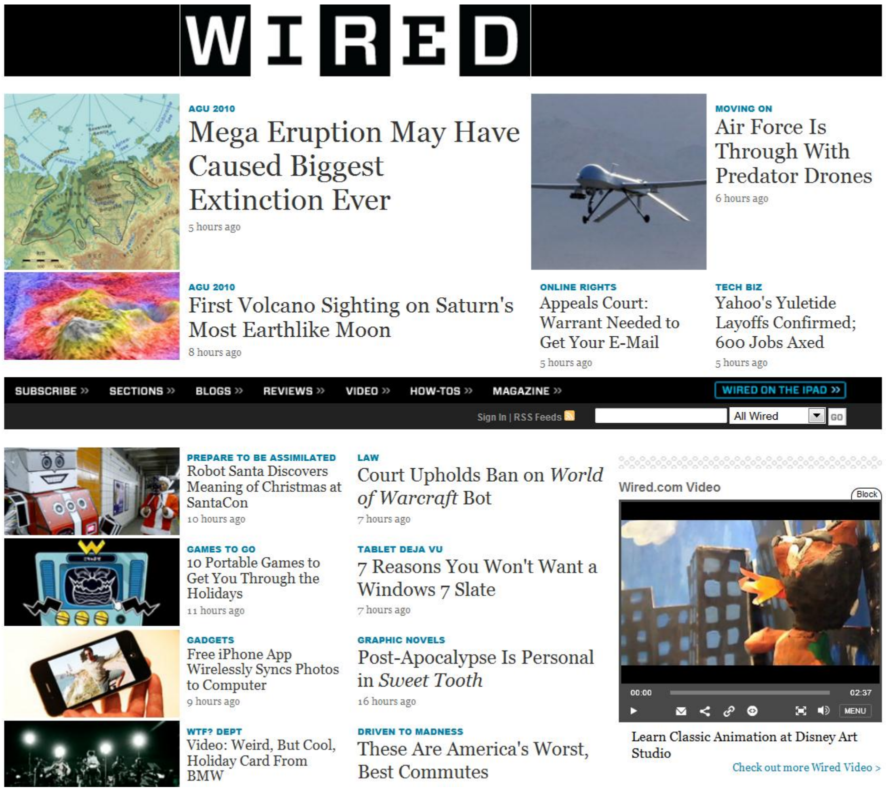
- CR, SV, MULTIOP: good!
- SCL, SC, WARP: not so good

Greater agreement for *faces/people* and *foreground objects:*

- Saliency at object level?



Source is Usually Unknown!

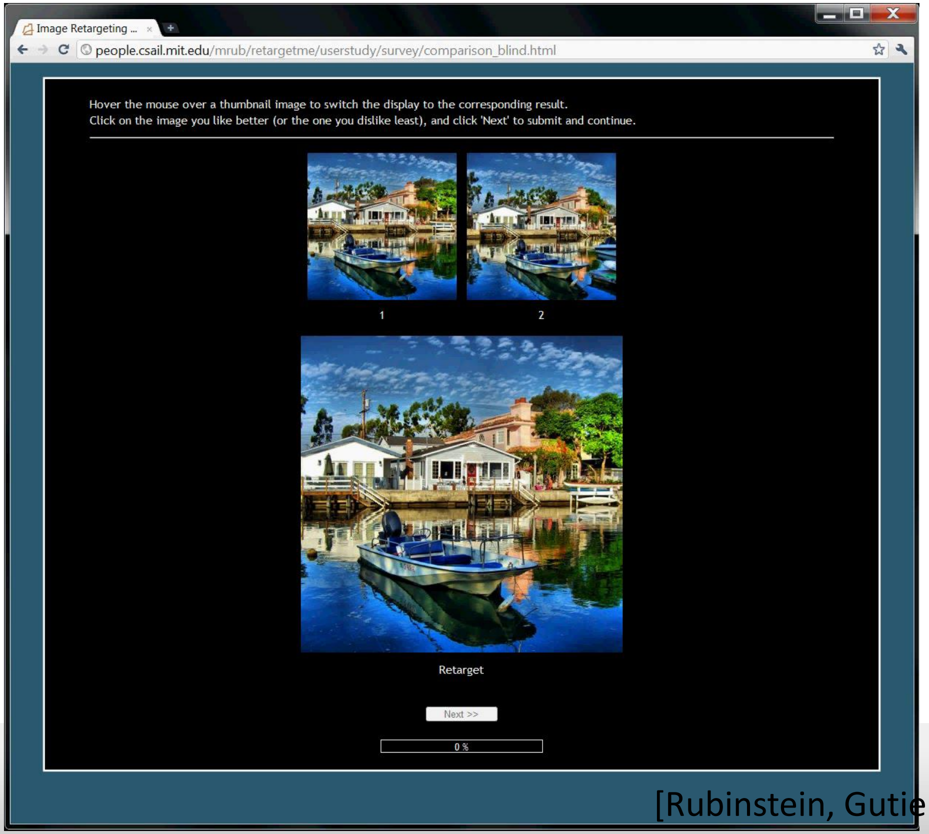


[Rubinstein, Gutierrez, Sorkine and Shamir 2010]

“No Reference” Experiment Results

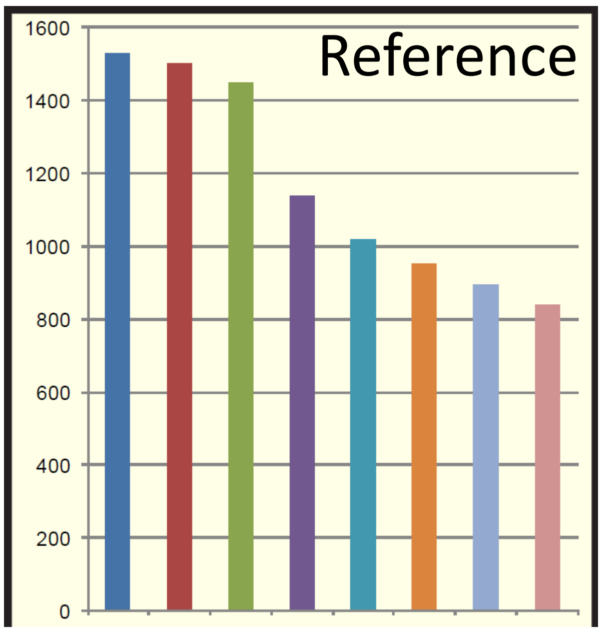
SIGGRAPHASIA2011 HONG KONG

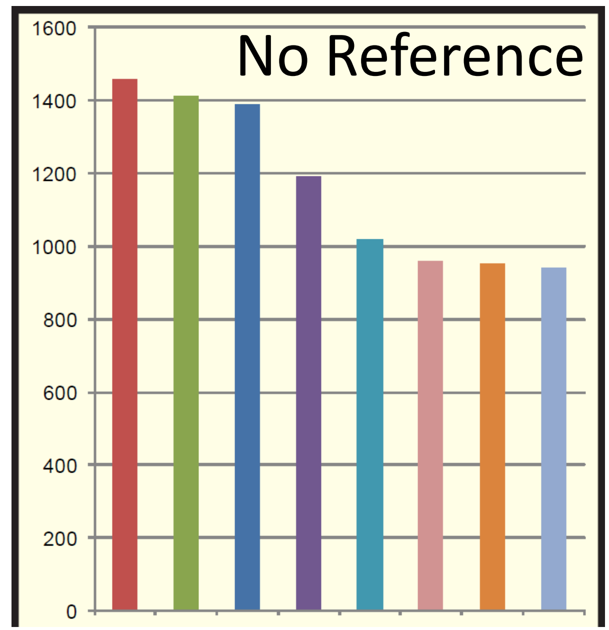
- Similar setup, source image **not shown**
- **New** set of 210 participants



[Rubinstein, Gutierrez, Sorkine and Shamir 2010]

"No Reference" Experiment Results





SV Streaming Video | CR Cropping | MULTIOP Multi-operator | SM Shift-maps | SNS Scale & Stretch | SCL Scaling | WARP Nonhomo. Warping | SC Seam Carving

lines/ edges	faces/ people	texture	foreground objects	geometric structures	symmetry	Aggregate	Rank product
0.964	0.988	0.946	0.737	0.950	0.957	0.978	0.985

Analysis of the users' responses: significance test

[Rubinstein, Gutierrez, Sorkine and Shamir 2010]

Computational Retargeting Measures

- Goal: can computational image distance measures predict human retargeting preferences?
  - Can be used to evaluate new operators
  - Can be used to develop new operators – [Simakov et al. 2008], [Rubinstein et al. 2009]

[Rubinstein, Gutierrez, Sorkine and Shamir 2010]

(Non-blind) Retargeting Measures

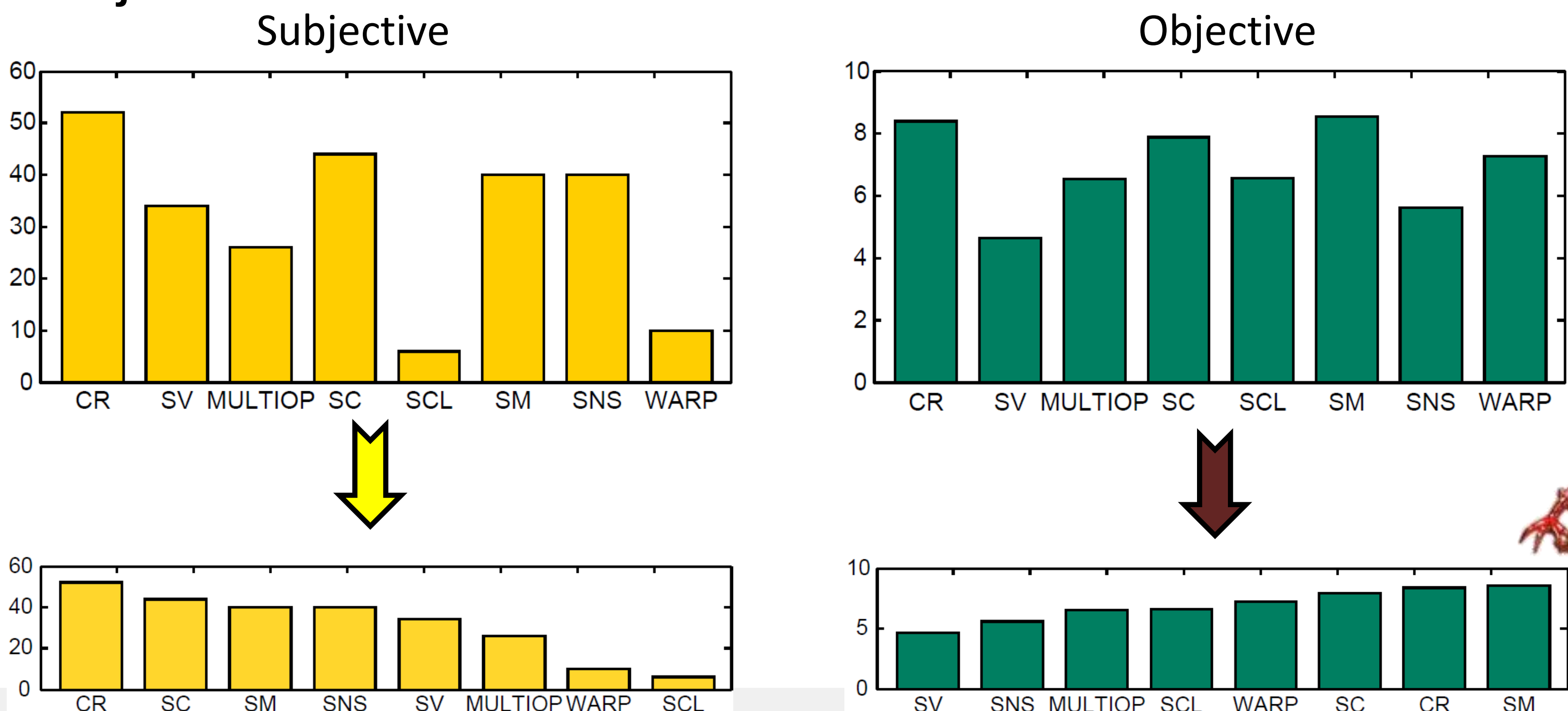
Objective Measures

- High level semantics:
  - Bidirectional Similarity [**BDS**] - Simakov et al. 2008
  - Bidirectional Warping [**BDW**] - Rubinstein et al. 2009
  - SIFT Flow [**SIFTflow**] – Liu et al. 2008
  - Earth Mover's Distance [**EMD**] - Pele and Werman 2009
- Low level features
  - Edge Histogram [**EH**] – Menjunath et al. 2001
  - Color Layout [**CL**] – Kasutani and Yamada 2001
- See dataset website and supplemental material for more details

[Rubinstein, Gutierrez, Sorkine and Shamir 2010]

How to Evaluate an Objective Measure?

- Define rate of agreement as the *correlation between rankings* induced by the user responses, and the objective measure



Objective Analysis Results

Metric	lines/ edges	faces/ people	texture	Foreground objects	geometric structures	symmetry	total
BDS	0.04	0.19	0.06	0.17	0.00	-0.01	0.08
BDW	0.03	0.05	-0.05	0.06	0.00	0.12	0.05
EH	0.04	-0.08	-0.06	-0.08	0.10	0.30	0.00
CL	-0.02	-0.18	-0.07	-0.18	-0.01	0.21	-0.07
SIFTflow	0.10	0.25	**0.12**	0.22	0.08	0.07	0.14
EMD	**0.22**	**0.26**	0.11	**0.23**	**0.24**	**0.50**	**0.25**

- The results were spectacular(ly poor!)
- We tried something else:
  - SIFT-flow [Liu et al. 2008]: SIFT
  - Earth mover’s distance [Pele & Werman 2009]: EMD
- Somewhat better ☺

[Rubinstein, Gutierrez, Sorkine and Shamir 2010]

Can computational image distance metrics predict human retargeting perception?

Metric	Attribute						Total		
	Lines/Edges	Faces/People	Texture	Foreground Objects	Geometric Structures	Symmetry	Mean	std	p-value
BDS	0.040	0.190	0.060	0.167	-0.004	-0.012	0.083	0.268	0.017
BDW	0.031	0.048	-0.048	0.060	0.004	0.119	0.046	0.181	0.869
EH	0.043	-0.076	-0.060	-0.079	0.103	0.298	0.004	0.334	0.641
CL	-0.023	-0.181	-0.071	-0.183	-0.009	0.214	-0.068	0.301	0.384
RAND	-0.046	-0.014	0.048	-0.032	-0.040	0.143	-0.031	0.284	0.693
SIFTflow	0.097	0.252	**0.119**	0.218	0.085	0.071	0.145	0.262	0.031
EMD	**0.220**	**0.262**	0.107	**0.226**	**0.237**	**0.500**	**0.251**	0.272	1e-5

(a) Complete rank correlation ($k = \infty$)

Metric	Attribute						Total		
	Lines/Edges	Faces/People	Texture	Foreground Objects	Geometric Structures	Symmetry	Mean	std	p-value
BDS	0.062	0.280	0.134	0.249	-0.025	-0.247	0.108	0.532	0.005
BDW	0.213	0.141	0.123	0.115	0.212	0.439	0.200	0.395	0.002
EH	-0.036	-0.207	-0.331	-0.177	0.111	0.294	-0.071	0.593	0.013
CL	-0.307	-0.336	-0.433	-0.519	-0.366	0.088	-0.320	0.543	1e-6
SIFTflow	0.241	**0.428**	**0.312**	**0.442**	**0.303**	0.002	0.298	0.483	1e-6
EMD	**0.301**	0.416	0.216	0.295	0.226	**0.534**	**0.326**	0.496	1e-6

(b) Rank correlation with respect to the three highest rank results ($k = 3$).

Table 6: *Correlation of objective and subjective measures for the complete rank (top) and for the three highest ranked results (bottom). In each column the mean τ correlation coefficient is shown ($-1 \leq \tau \leq 1$), calculated over all images in the dataset with the corresponding attribute. The last three columns show the mean score, standard deviation, and respective p-value over all image types. Highest score in each column appears in bold.*

[Rubinstein, Gutierrez, Sorkine and Shamir 2010]

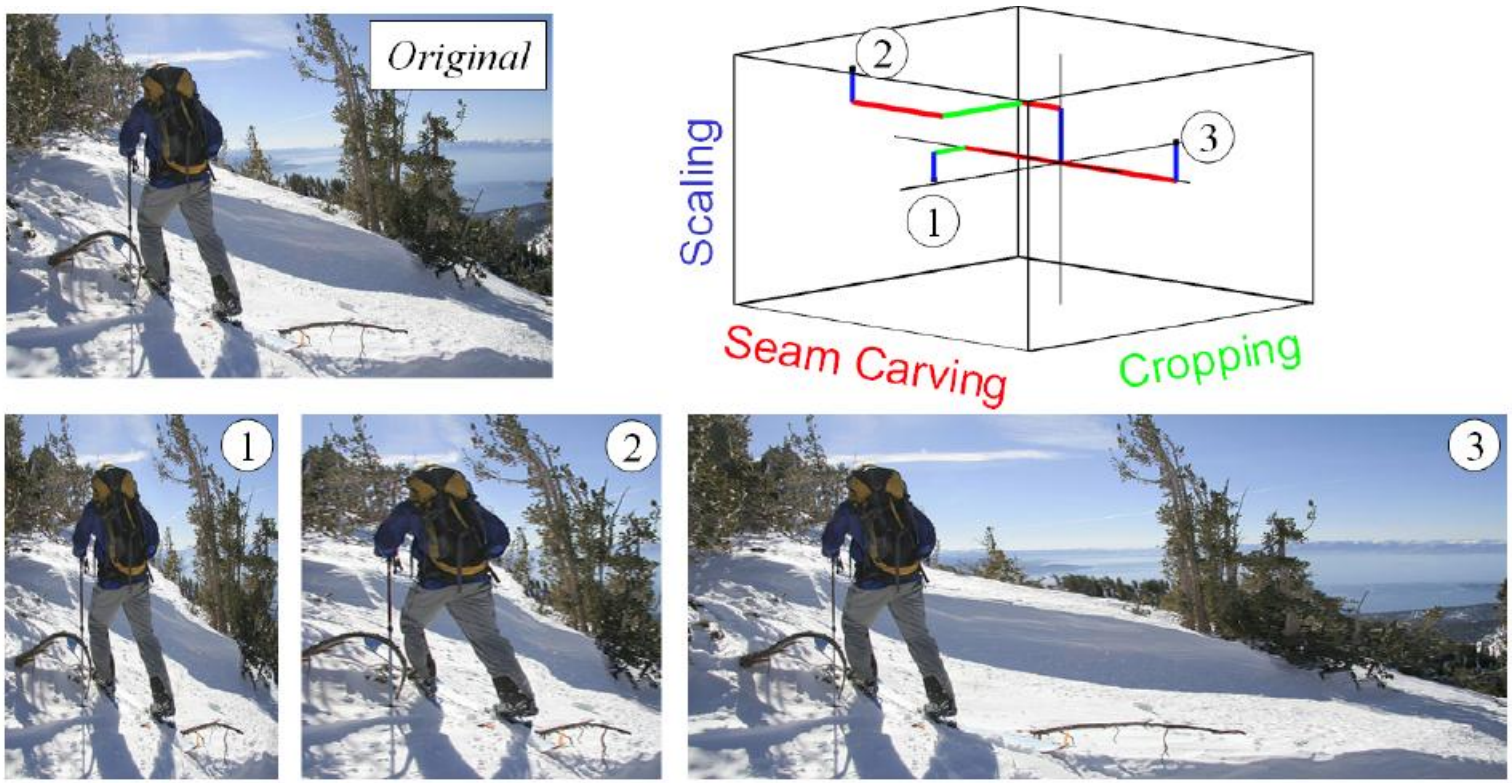
Conclusions

SUBJECTIVE:
More recent algorithms **do** outperform their predecessors in a (surprisingly) consistent way

Cropping is the simplest and one of the best:
loss of info OK
distortion **NOT** OK
bring it back!

Interestingly, scaling and seam carving do not do very well on their own… but are two of the three in MULTIOP:
combination of simple methods?

Conclusions



Conclusions

OBJECTIVE:
We are a long way from predicting human perception

Four similarity image metrics did not perform well at all

Two metrics not originally designed for that purpose did somewhat better

Optimize retargeting wrt those?

Further research is (badly!) needed

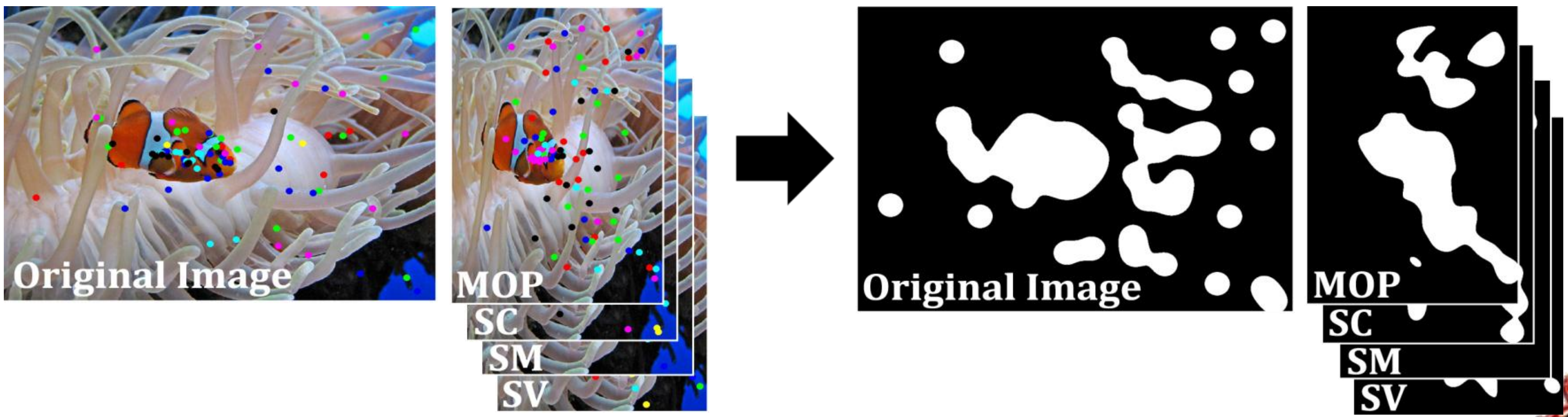
Conclusions

We need **video** analysis and experiments!

Image Retargeting Quality Assessment
Computer Graphics Forum, 2011, Vol. 30, No. 2, Eurographics 2011,
Yong-Jin Liu, Xi Luo, Yu-Ming Xuan, Wen-Feng Chen, Xiao-Lan Fu

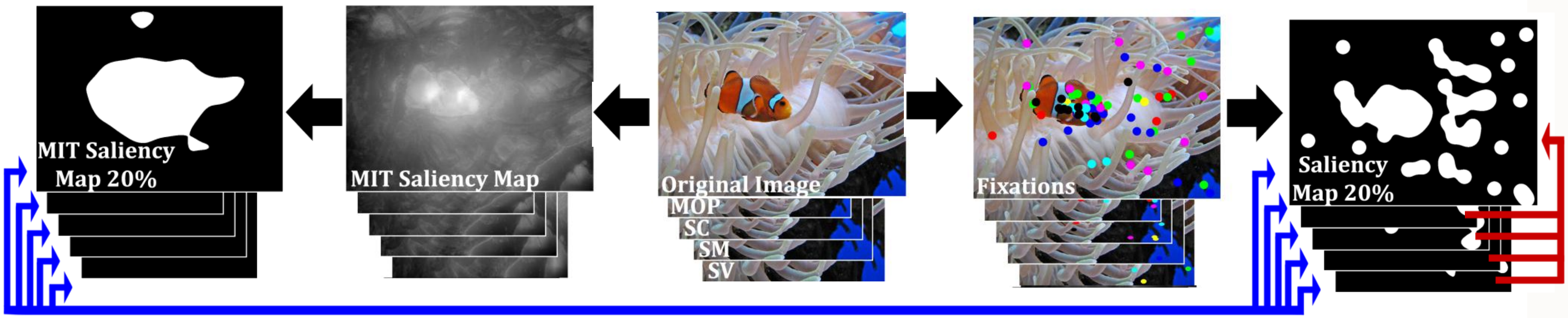
$$ColSim(C^0_{ori}, C^0_{ret}) = w_L SalSim(L^{*0}_{ori}, L^{*0}_{ret}) + w_a SalSim(a^{*0}_{ori}, a^{*0}_{ret}) + w_b SalSim(b^{*0}_{ori}, b^{*0}_{ret})$$

Using Eye-Tracking to Assess Different Image Retargeting Methods



[Castillo, Judd and Gutierrez 2011]

Overview

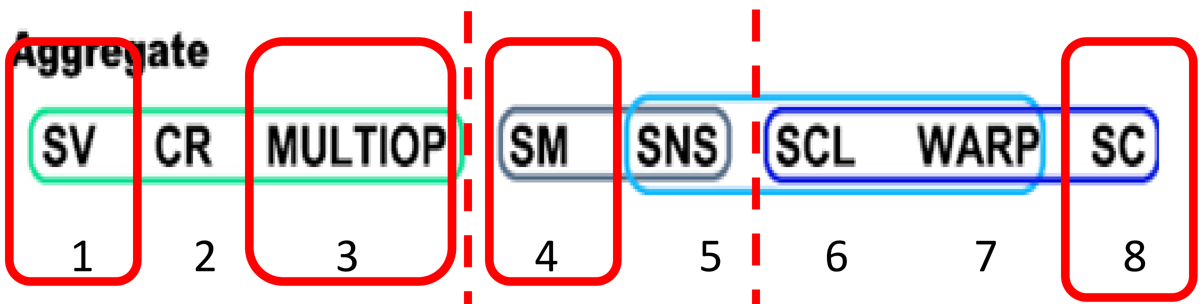


[Castillo, Judd and Gutierrez 2011]

Retargeting Operators

- Seam Carving **[SC]** [Rubinstein et al. 2008]
- Shift Maps **[SM]** [Pritch et al. 2009]

- Multi-Operator **[MULTIOP]** [Rubinstein et al. 2009]
- Streaming Video **[SV]** [Krähenbühl et al. 2009]

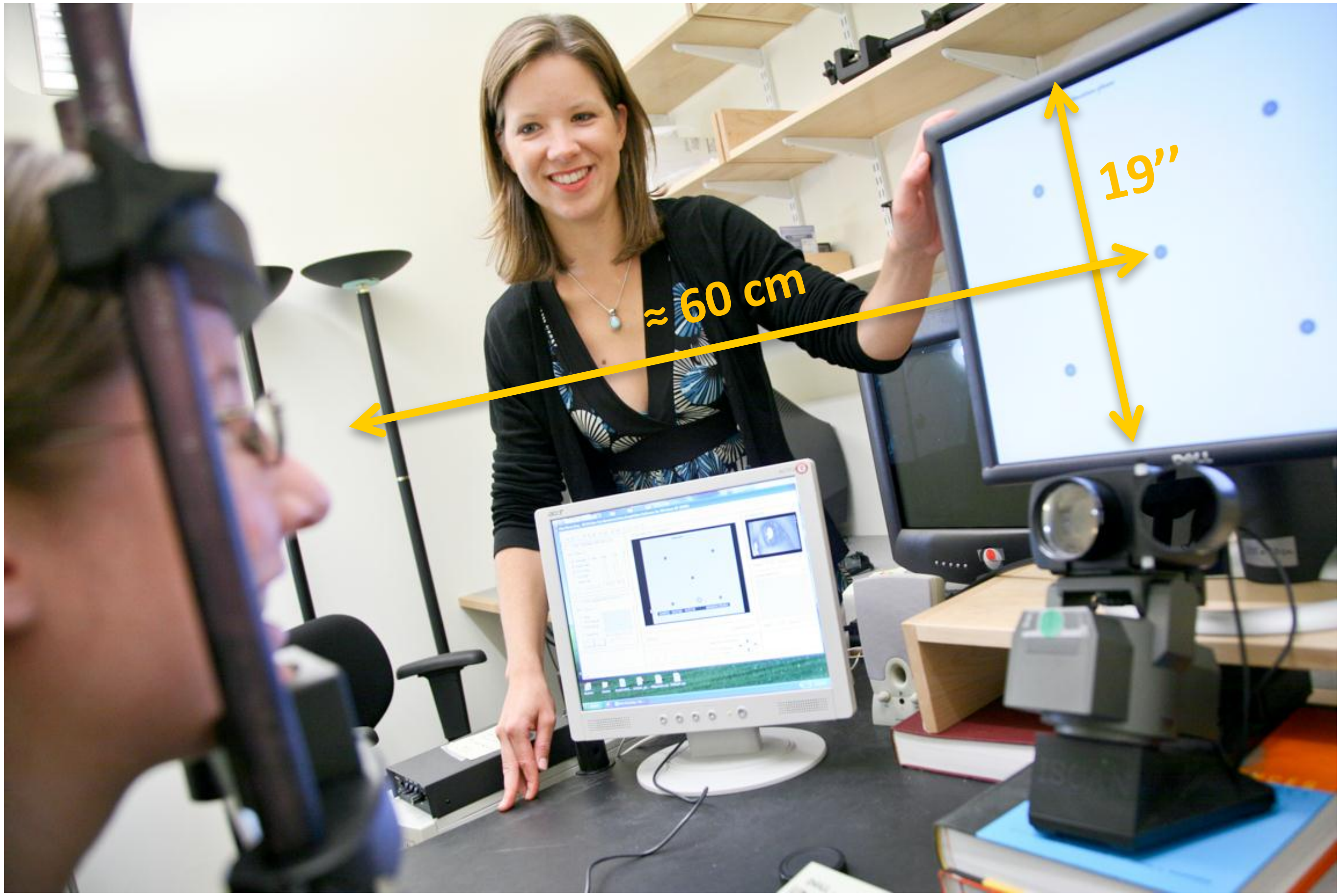


Rank **[Rubinstein et al. SIGAsia 2010]**

[Castillo, Judd and Gutierrez 2011]

SIGGRAPHASIA2011 HONG KONG

Collect eye tracking data



[Photo Credit: Jason Dorfman CSAIL website]

Screen resolution 1280x1024

Each image shown for 5 seconds

[Castillo, Judd and Gutierrez 2011]

Eye tracking data

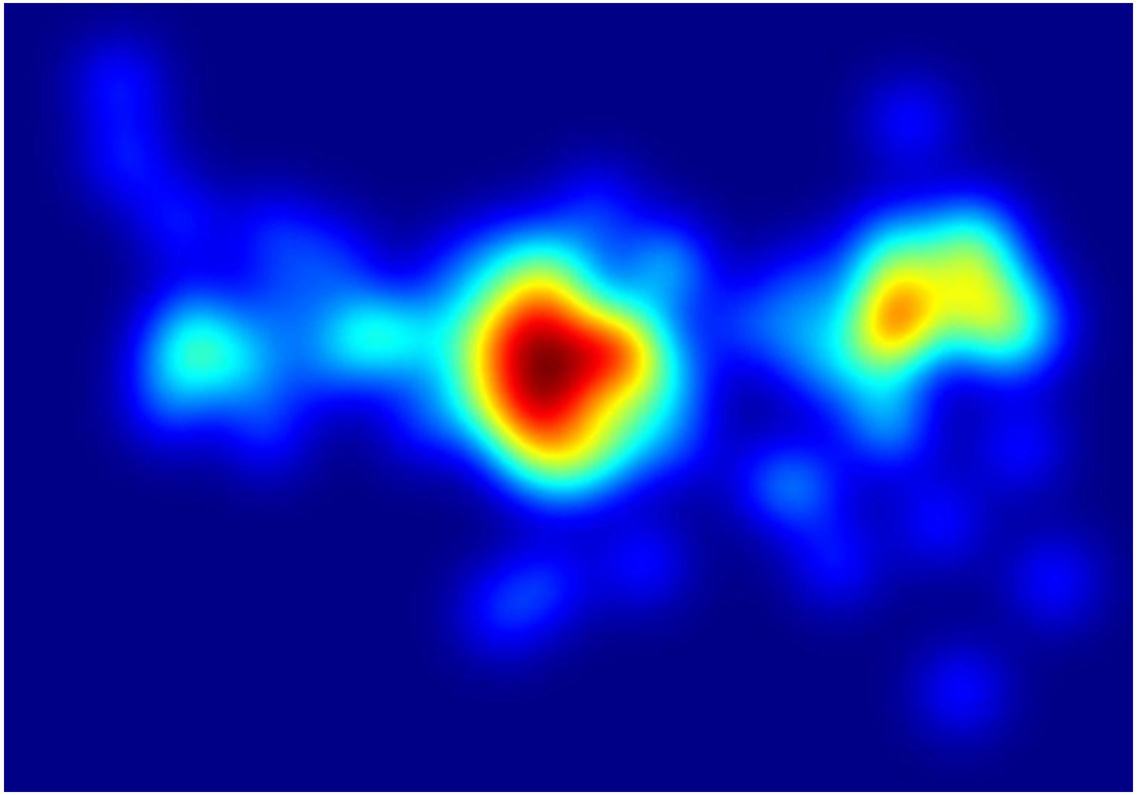
Contextual guidance of eye movements and attention in real-world scenes: The role of global features on object search **[Torralba et al. 2006]**

Fixations for 7 users

[Castillo, Judd and Gutierrez 2011]

Eye tracking data

Learning to predict where humans look **[Judd et al. 2009]**

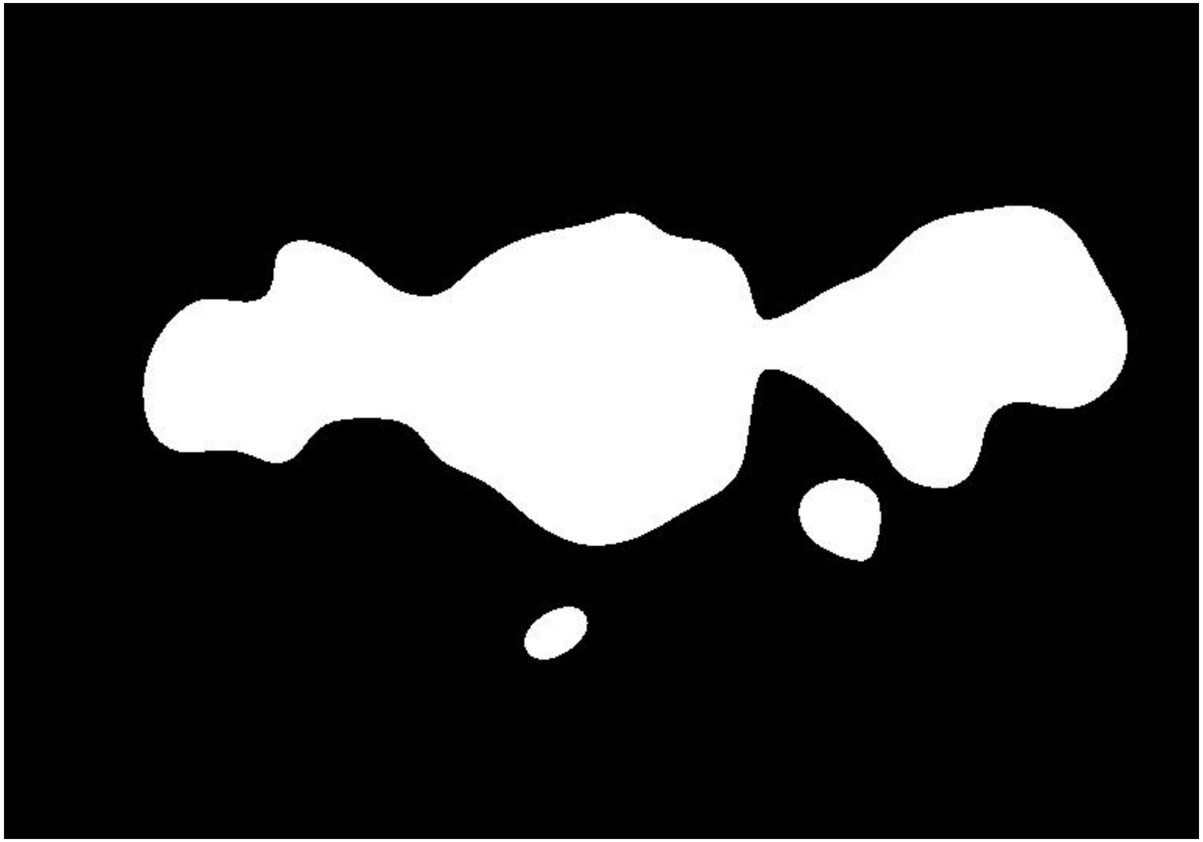


Average fixation locations / continuous saliency map

[Castillo, Judd and Gutierrez 2011]

Eye tracking data

Learning to predict where humans look **[Judd et al. 2009]**

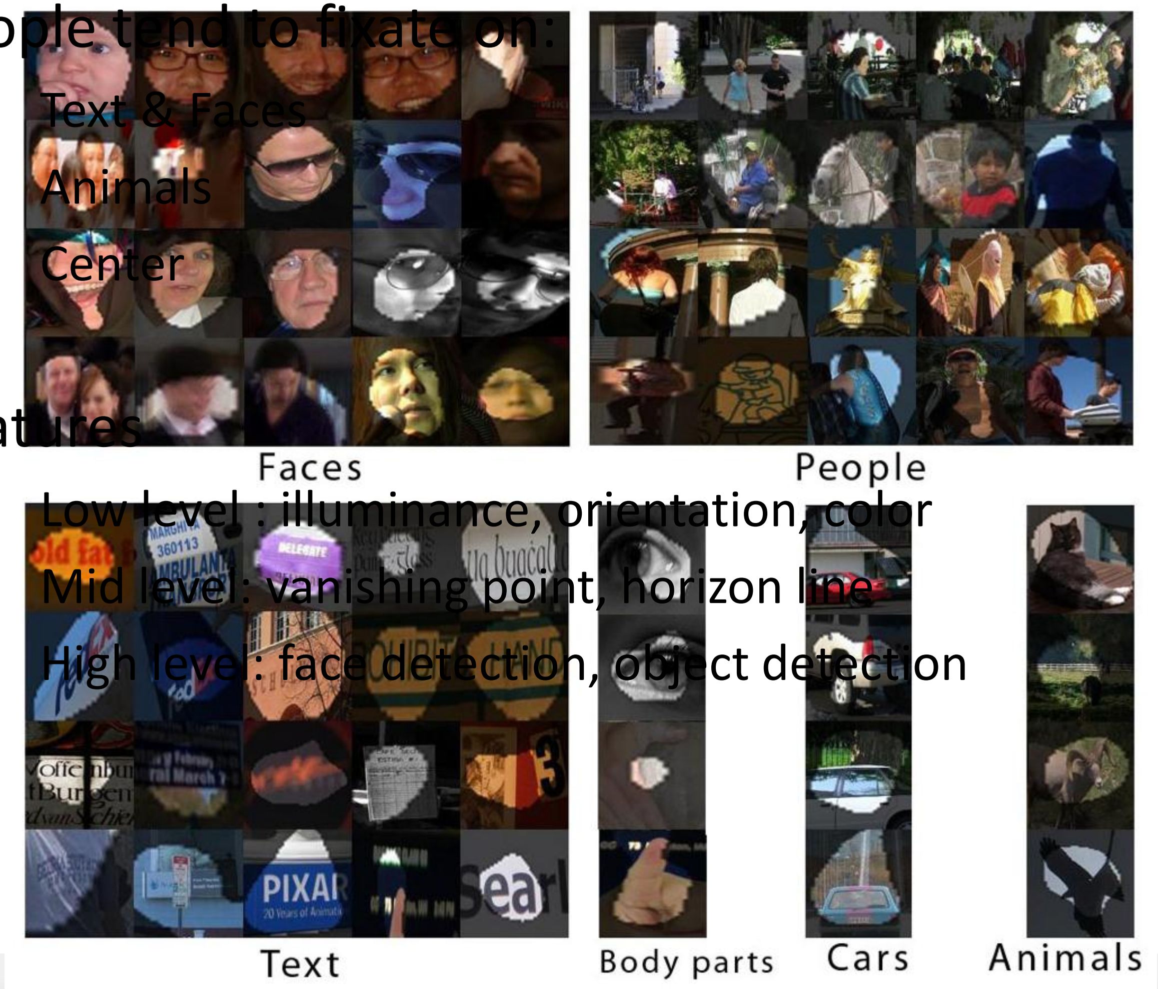


Top 20% salient locations

[Castillo, Judd and Gutierrez 2011]

MIT Predictive Model of Saliency

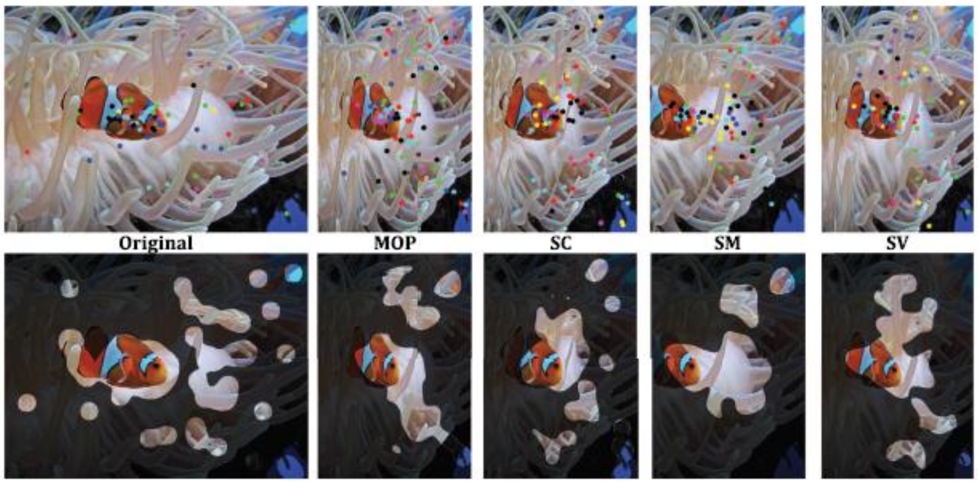
- People tend to fixate on:
  1. Text & Faces
  2. Animals
  3. Center
- Features
  - Low level : illuminance, orientation, color
  - Mid level: vanishing point, horizon line
  - High level: face detection, object detection



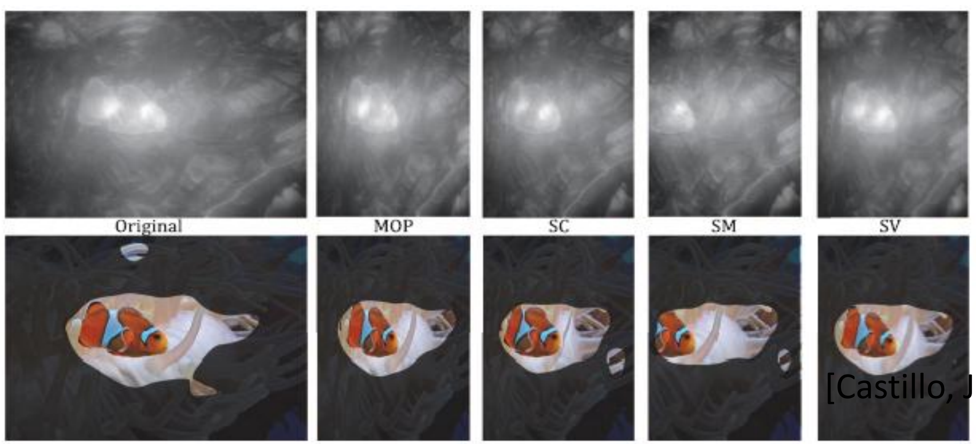
[Judd et al. 2009]

MIT Predictive Model of Saliency

Saliency Maps from eye-tracking data

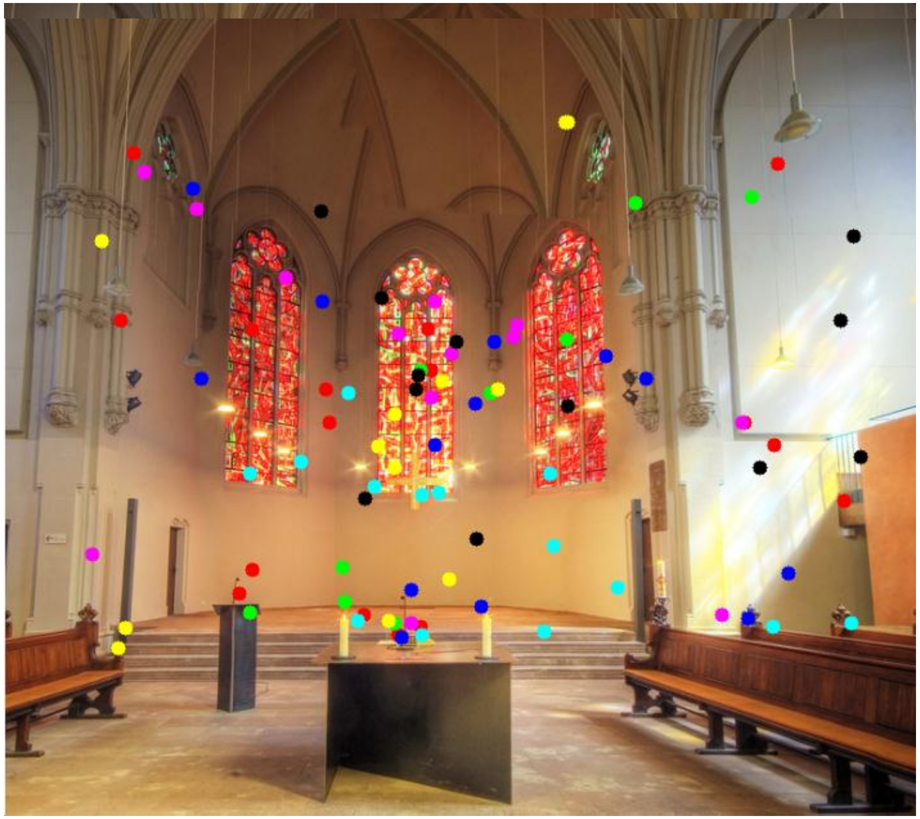


Saliency Maps predicted by the model from Judd et al. [2009]



[Castillo, Judd and Gutierrez 2011]

Examples and Discussion



[Castillo, Judd and Gutierrez 2011]

Conclusions

- Lots of methods in the past few years, in top-notch places
- Relatively small impact in industry

RetargetMe

http://people.csail.mit.edu/mrub/retargetme/
or Google: "retargetme"

- We need more (and better!) metrics
- Does video retargeting *really* work?

Conclusions

- Eye-tracking data framework
- The model of saliency from Judd et al. [2009] can be an useful tool in a retargeting context when using an eye tracker is not feasible
- Analysis of 4 retargeting operators with 6 image distance measures
  - Using eye-tracking data can improve the predicting capabilities of these measures
- Alteration of the image *semantics.*
  - Content removal alters RoIs although the results can be aesthetically pleasing
- **Attentional tension** between RoIs and artifacts
  - Large artifacts can remain unnoticed when not in a RoI *(At least for our 5 second task)*