A Comparative Study of Image Retargeting

Michael Rubinstein, Diego Gutierrez, Olga Sorkine, Ariel Shamir

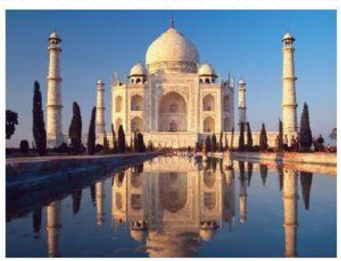
SIGGRAPH ASIA 2010

Presentation by Moeka Takagi

Goal

- provide common ground for comparison between existing and future retargeting methods
- Take into account subjective and objective results

Brick House (L, T, G)



Taj Mahal (L, G, S)

butterfly (F, G)

Fatem (L, P, T, G)

boat (L, F)

Collecting Data

- Collecting pure retargeting data is challenging
- Manual retargeting requires proficient artist
  - takes too long to resize an image
  - limits size of resulting image
- Artists may insert bias in replicating retargeting technique
- Concentrate on existing retargeting methods

Evaluating Data

- Difficult
- Results depend on media content itself; certain methods can work better for certain content
- Evaluation is subjective
- Is there even a consensus?

Three Main Objectives

- Preserving the important content of the image
- Limiting visual artifacts in resulting media
- Preserving internal structures of original media

Creating the Benchmark Set

- chose a set of attributes that could be mapped to the three main objectives
- people and faces, lines and/or clear edges, evident foreground objects, texture elements or repeating patterns, specific geometric structures, symmetry
- gathered 80 images from various retargeting papers having one or more of attributes

Retargeting Methods

- Discrete: removes or inserts pixels/patches to preserve content

  Seam Carving (SC), Shift-maps (SM), Cropping (CR)
- Continuous: optimize a mapping from source media size to target size

  Nonhomogeneous warping (WARP), Scale and Stretch (SNS), Streaming Video (SV), Scaling (SCL)
- Multi-operator(MULTIOP) combination of SC, SCL, and CR

Quick Recap of Methods

- Seam Carving: Removes or duplicates chains of pixels with least importance in image
- Shift-maps: removes entire objects, not seams
- Nonhomogenous warp: amount of deformation is proportional to importance, uses face detectors
- Scale-and-stretch: important regions uniformly scale and preserve shape
- Multi-operator: uses SC, SCL, CR all together
- Streaming Video: warping method using line detection, user markings of lines and objects

Retargeted Images

- restricted changes to either width or height of the image
- reduced considerable amount, 25% or 50%
- authors of retargeting papers retargeted the images
- 37 images used for study with various attributes

Subjective Analysis

- paired comparisons technique

  participants shown two retargeted images, side by side, and are asked to choose one they like better

- web-based interface allowed user to switch between various retargeted results or original image

Subjective Analysis

- Total number of comparisons too large, 1036 comparisons
- Followed linked-paired comparison design
- Each pair is compared by same number k of participants
- Within pairs compared by each participant, each stimulus appears an equal number of times, β
- Given any two participants, there are exactly λ pairs compared by both of them
- Used $k = 3$, $\beta = 3$, $\lambda = 4$

Subjective Analysis

- Each participant assign 12 out of 28 possible paired comparisons per image
- Each image had 21 participants, total of 252 votes
- Total study had 210 participants

  40% female, 60% males

  average age, 30

Subjective Analysis

- Also conducted *no reference image test* where original image was not shown to 210 new participants
- Sometimes, participants asked to choose out of set of reasons one for not choosing a result

Subjective Analysis

- Complete agreement means everyone voted same way
- High disagreement means people tend not to agree
- Coefficient of agreement:

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

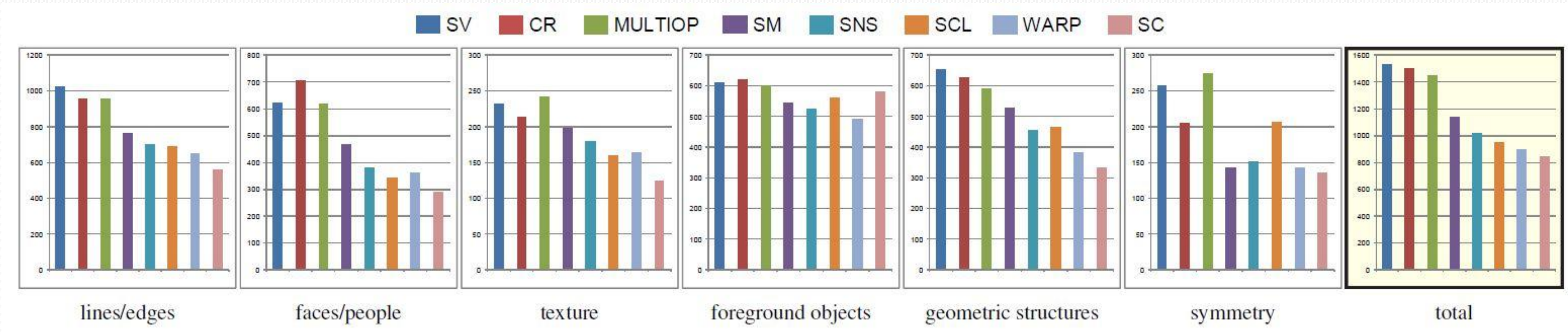
- a_{ij} = number of times method i was chosen over method j, m = the number of participants, t = number of retargeting methods tested
- u=1 means complete agreement
- u= -1/m means even distribution of answers

Subjective Results

	lines/ edges	faces/ people	texture	foreground objects	geometric structures	symmetry	Aggregate
u (with ref.)	0.073	0.166	0.070	0.146	0.084	0.132	0.095
u (no ref.)	0.047	0.086	0.027	0.075	0.059	0.054	0.059
R'	107	83	53	91	85	53	129

- More agreement for faces/people, foreground objects, and symmetry sets
- Agreement drops significantly without a reference image

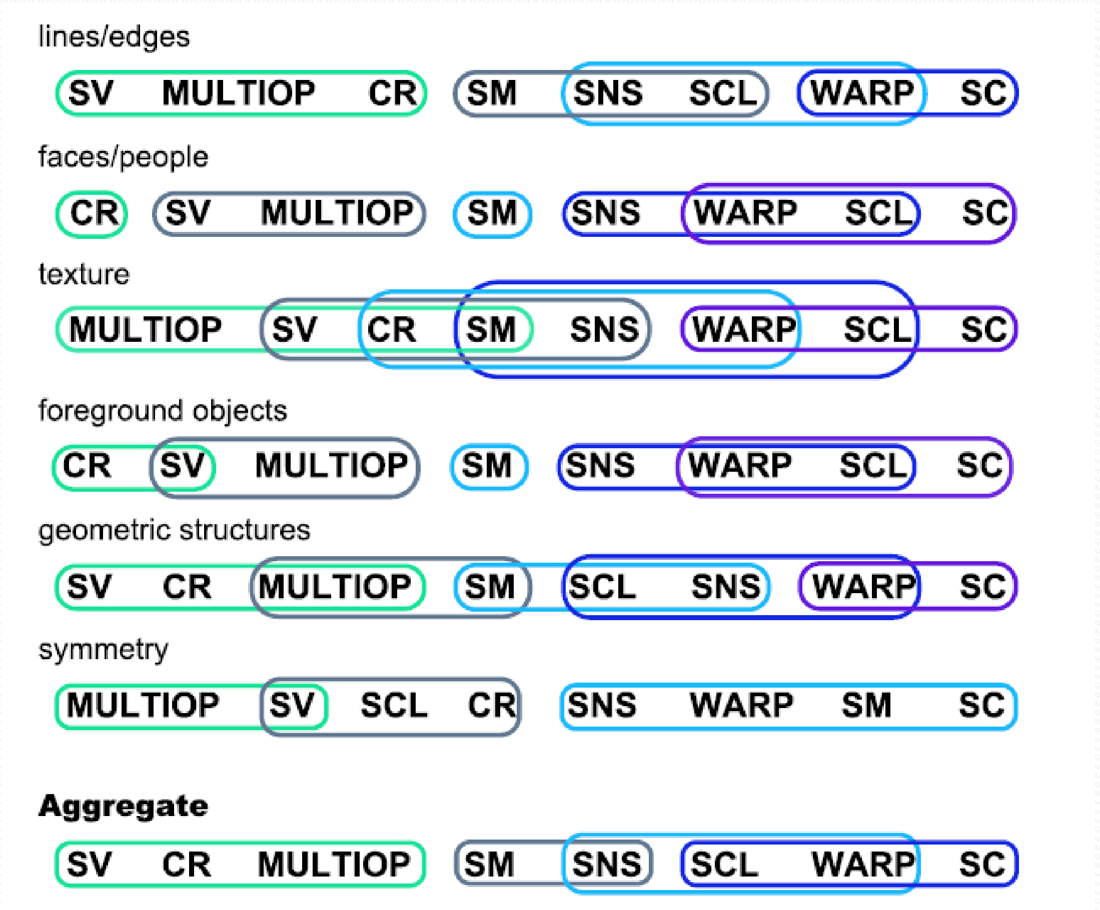
Subjective Results



$$\Psi(\mathcal{O}) = \left(\prod_i r_{\mathcal{O},i}\right)^{1/b}$$

- ro,i = specific ranking for retargeting method o
- i = category of attributes

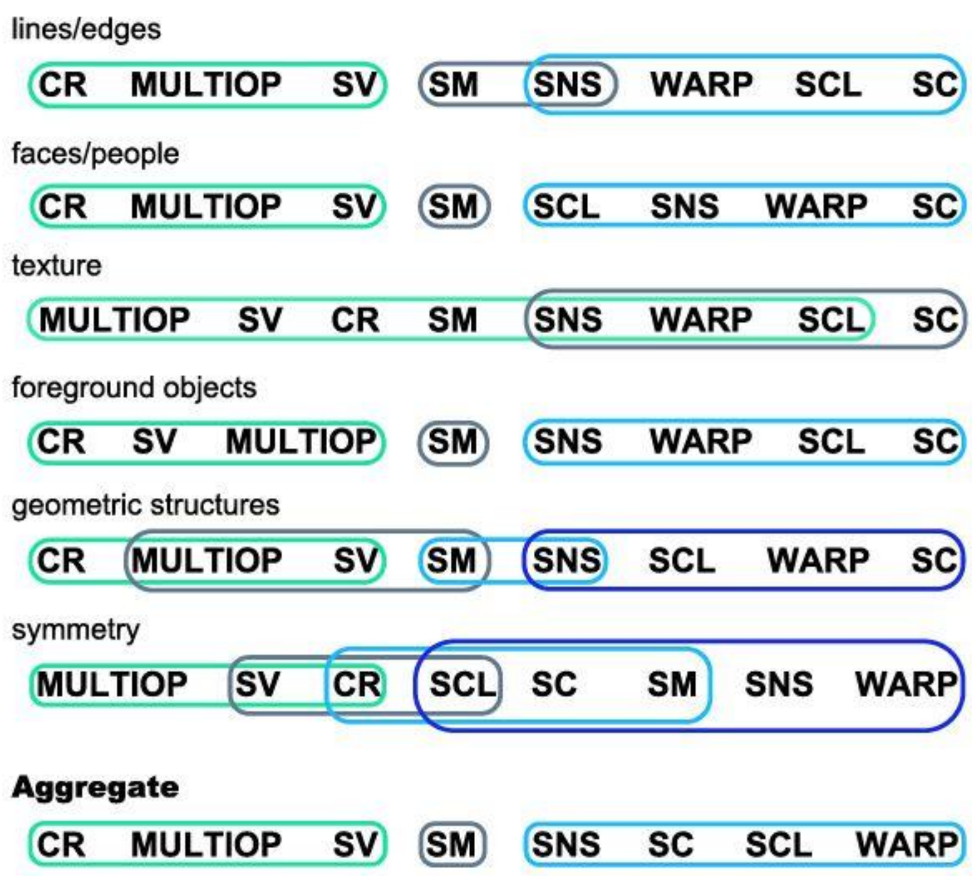
Subjective Results



Subjective Observations

- In general, CR, SV, MULTIOP were ranked highest, while SCL, SC, WARP were ranked lowest
- SV and MULTIOP are content-aware methods, and CR doesn't create any artifacts
- Loss of content is preferred over deformation artifacts

No-reference Results

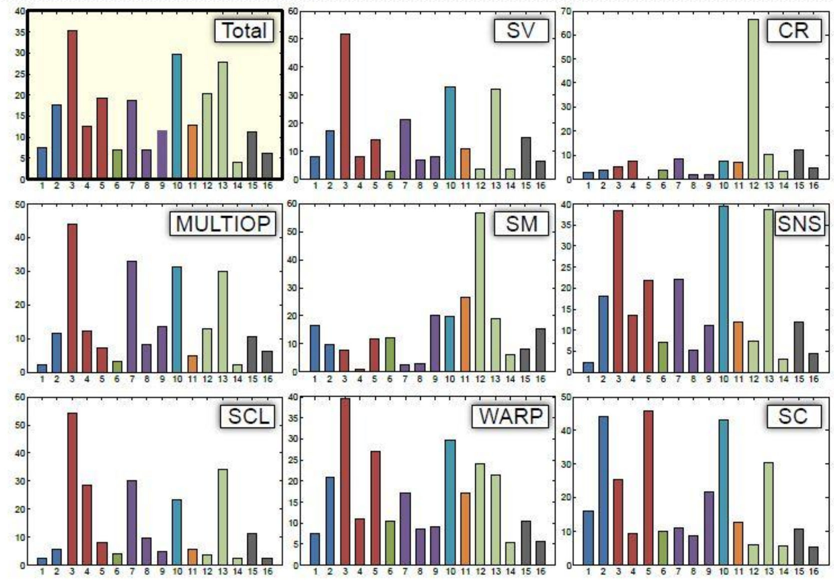


No-reference Observations

- Results show similar pattern as the test with reference image
- CR, SV, MULTIOP still ranked high
- Main difference: CR almost always preferred choice
- No reference image to show the loss of content

Not chosen because…

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



Objective Analysis-Methods

- Bidirectional Similarity

  For every patch in image, looks for well-matched patch in other image

  Distance between images is defined as mean distance in color space between corresponding patches
- Bidirectional Warping

  Result mapping will maintain order of patches in image

  Distance is taken to be the mean or maximal distance between corresponding patches in color space

Objective Analysis-Methods

- Standard edge histogram

  partitions pictures into smaller blocks and calculates edge type (vertical, horizontal, diagonal, non-directional, no edge) for each block
- Standard color layout

  partitions pictures into smaller blocks and computes a representative color for each block

Objective Analysis-Methods

- SIFT-flow

  robustly captures structural properties
- Earth-Mover's Distance

  uses “ground distance,” cost of transforming a unit of mass between distributions

Evaluation

- Create subjective similarity vector
- $s = \langle s_1, \ldots, s_n \rangle$ for 8 methods, s_i is number of times the retargeting result T_i was favored
- higher s_i = better method i

- Create objective distance vector
- $o = \langle o_1, \ldots, o_n \rangle$ For given image I, compare with targeted image by given objective measure D
- $o_i = D(I, T_i)$, lower o_i = better method i

Compare

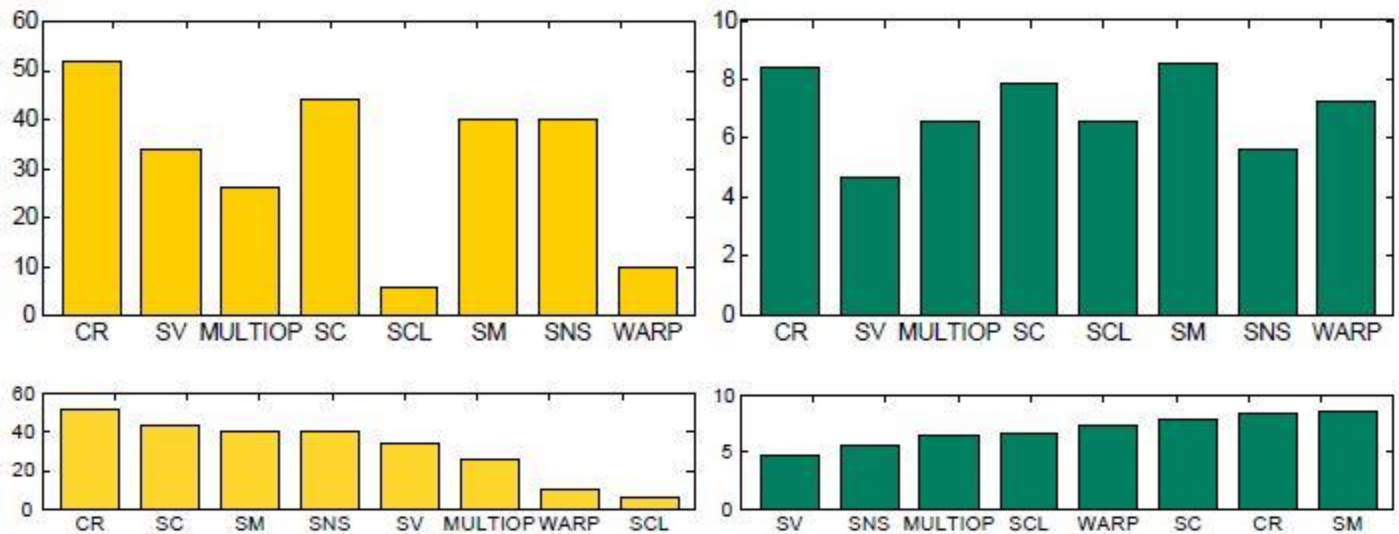
- Sort s vector in descending order, o vector in ascending order and determine rank of s and o
- Use Kendall t distance to measure correlation between rankings

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)}$$

- n = length of rankings, n_c = number of agreeing pairs, n_d = number of disagreeing pairs
- -1<= t <= 1, increasing t indicated increasing agreement

Correlation Results

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



Conclusions

- SV and MULTIOP performed well
- Cropping, although naive, still favored as well
- Still a long way to imitating human perception
- SIFTflow and EMD, measures not used before for retargeting, generally agree better with user preferences

References

- http://people.csail.mit.edu/mrub/papers/retBenchmark.pdf
- http://people.csail.mit.edu/mrub/retargetme/supplemental.pdf
- http://en.wikipedia.org/wiki/Rank_product