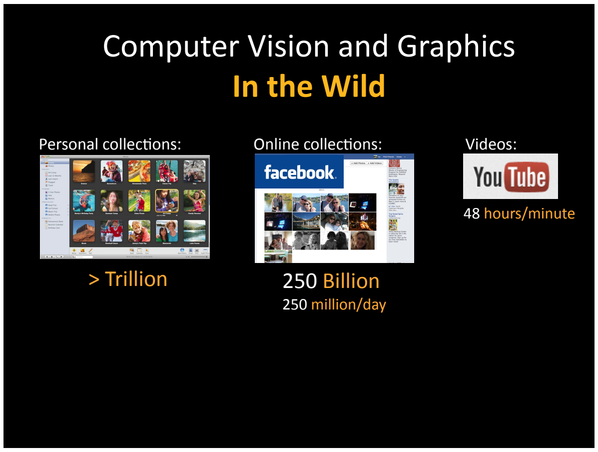
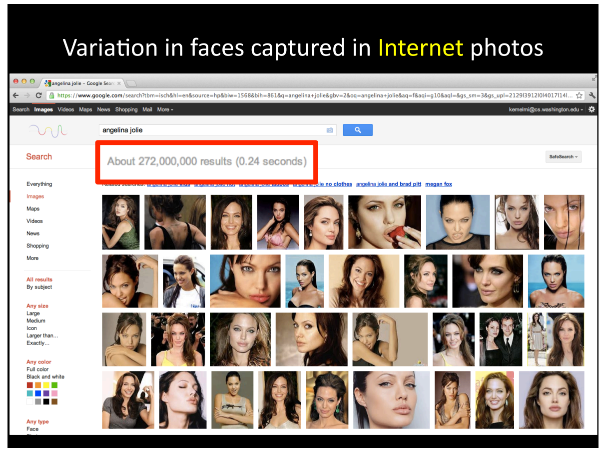
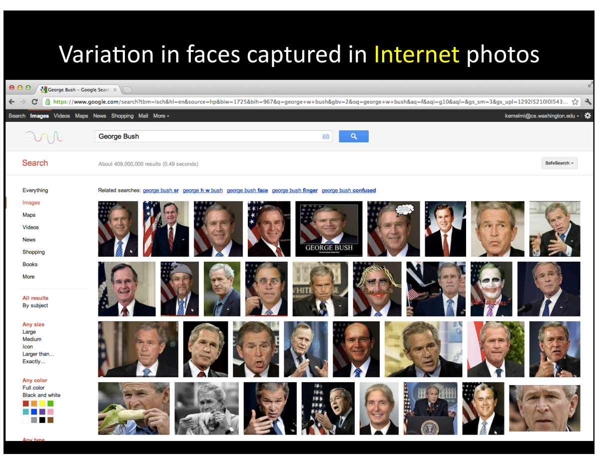
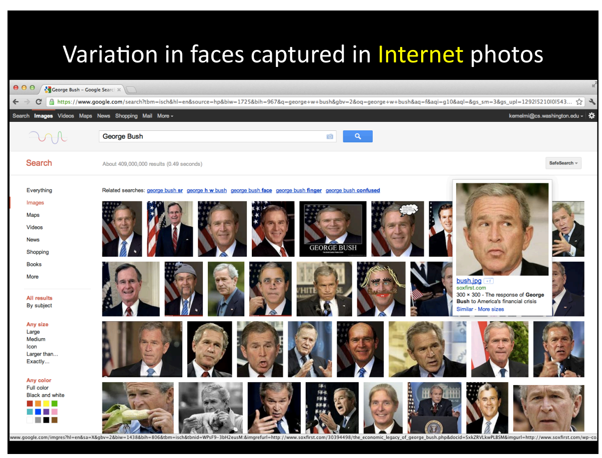
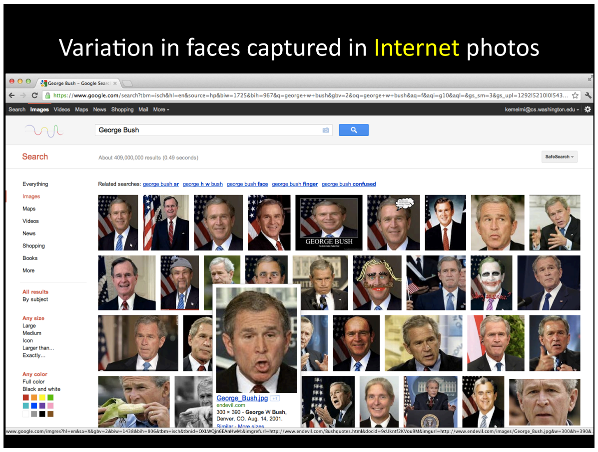

Computer Vision and Graphics
In the Wild
Personal collections:
> Trillion
Online collections:
facebook
250 Billion
250 million/day
Videos:
YouTube
48 hours/minute

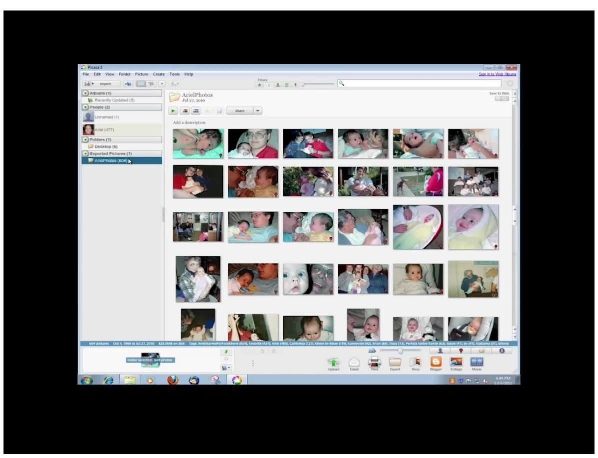

Variation in faces captured in Internet photos
angelina jolie
About 272,000,000 results (0.24 seconds)

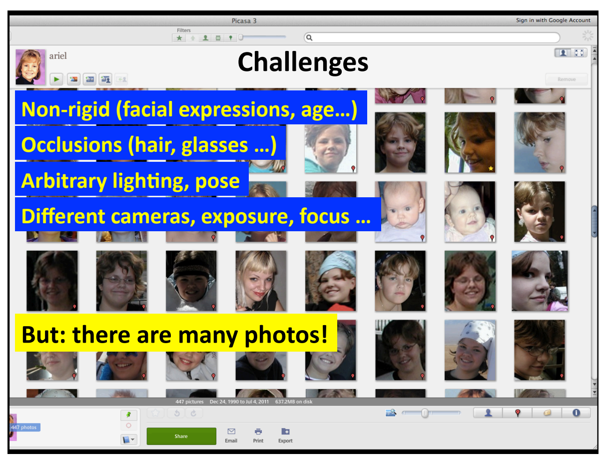

Variation in faces captured in Internet photos
George Bush


Variation in faces captured in Internet photos
George Bush


Variation in faces captured in Internet photos
George Bush
Search


© Photos: Ariel McClendon




Challenges
Non-rigid (facial expressions, age…)
Occlusions (hair, glasses …)
Arbitrary lighting, pose
Different cameras, exposure, focus …
But: there are many photos!

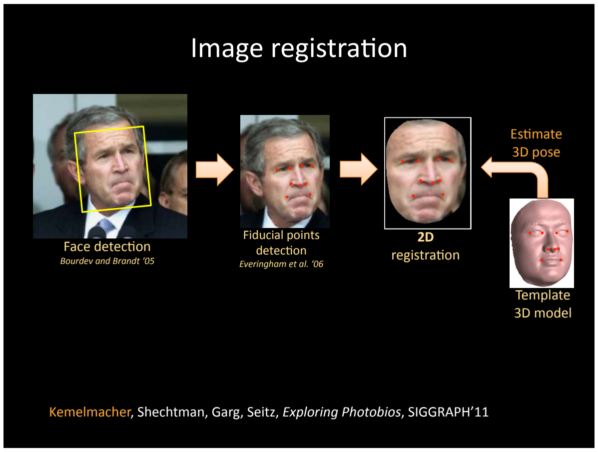


Image registration
Face detection
Bourdev and Brandt '05
Fiducial points
detection
Everingham et al. '06
2D
registration
Estimate
3D pose
Template
3D model
Kemelmacher, Shechtman, Garg, Seitz, Exploring Photobios, SIGGRAPH'11

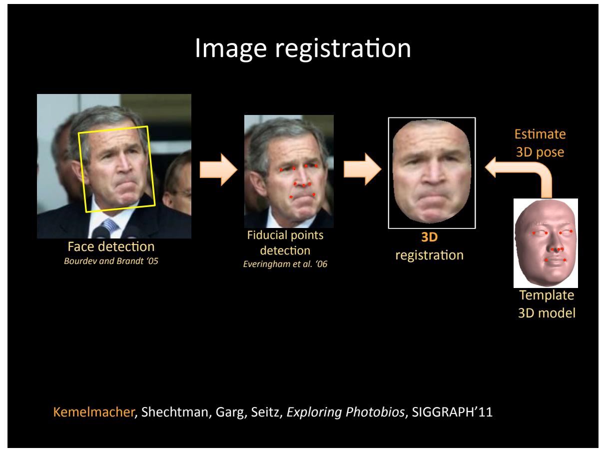
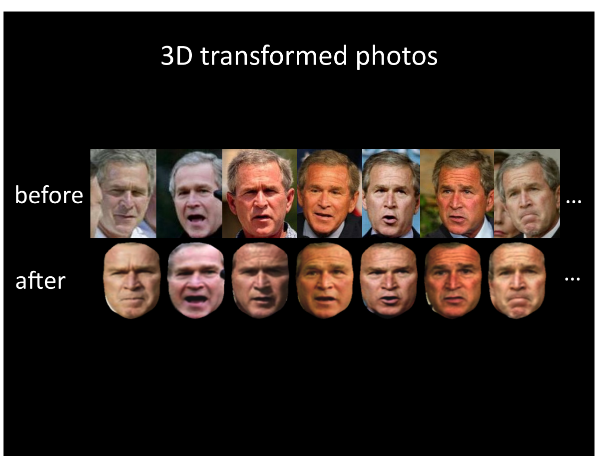
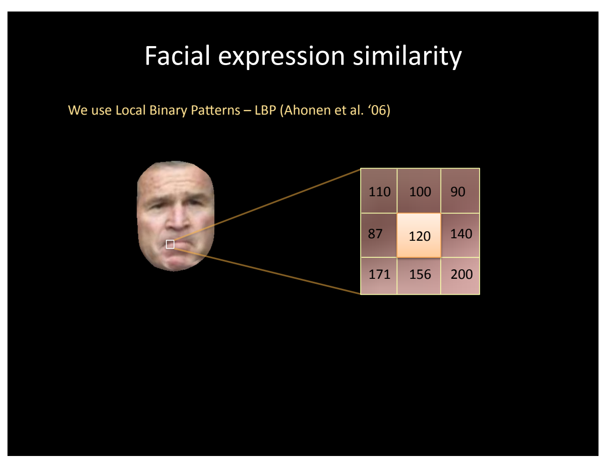
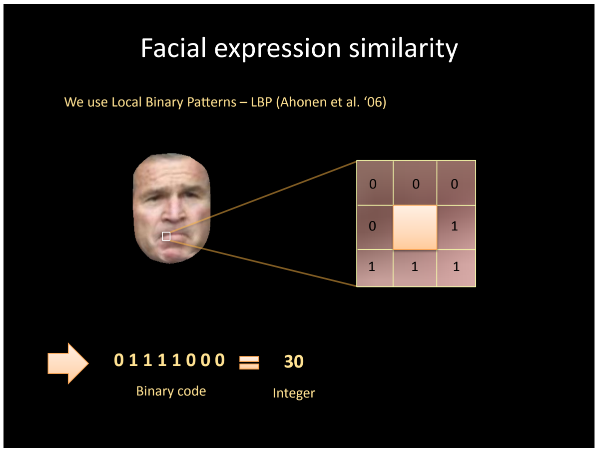
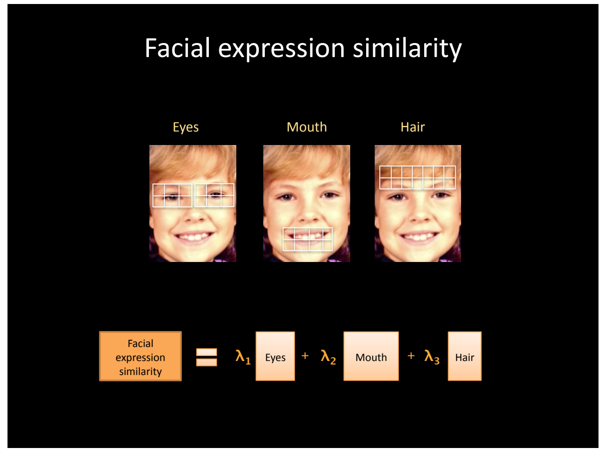


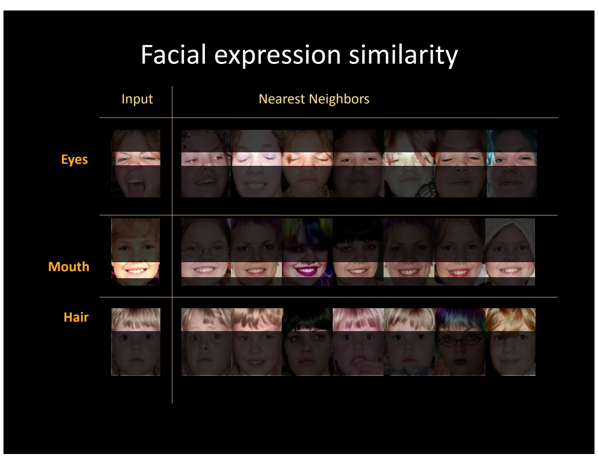
Image registration
Face detection
Bourdev and Brandt '05
Fiducial points
detection
Everingham et al. '06
3D
registration
Estimate
3D pose
Template
3D model
Kemelmacher, Shechtman, Garg, Seitz, Exploring Photobios, SIGGRAPH'11

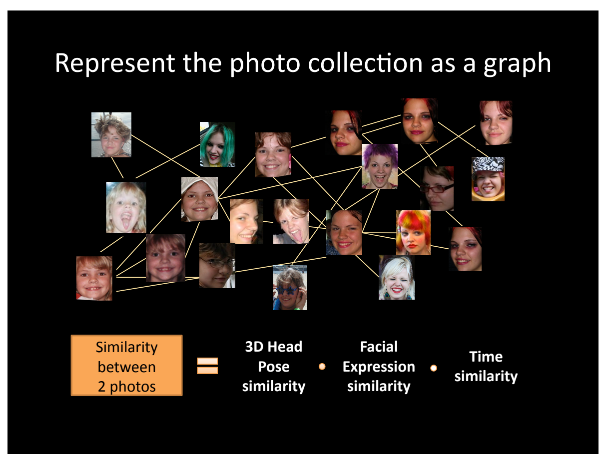

3D transformed photos
before
…
after
…

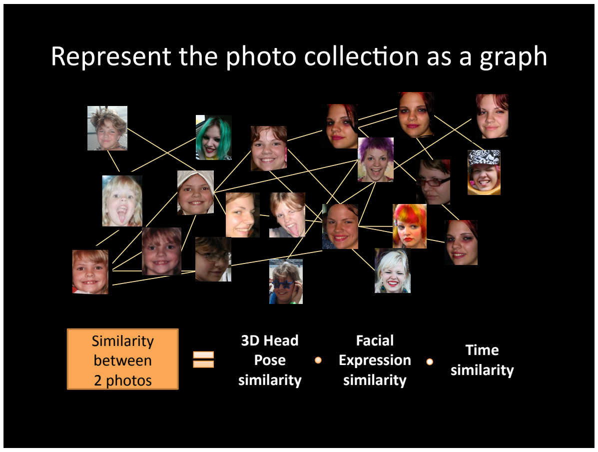

Facial expression similarity
We use Local Binary Patterns – LBP (Ahonen et al. '06)
110
100
90
87
120
140
171
156
200

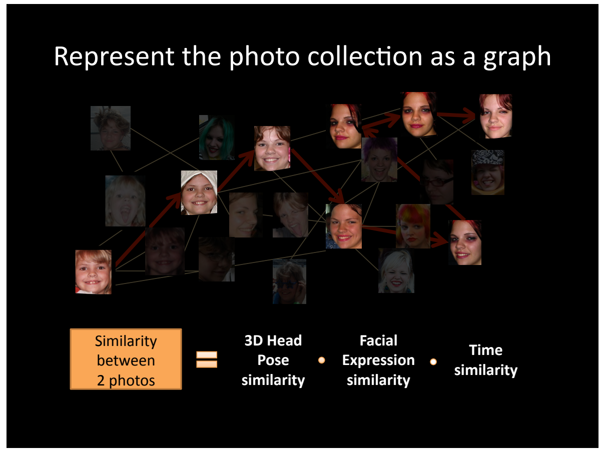

Facial expression similarity
We use Local Binary Patterns – LBP (Ahonen et al. '06)
0 0 0
0 1
1 1 1
0 1 1 1 1 0 0 0
Binary code
30
Integer

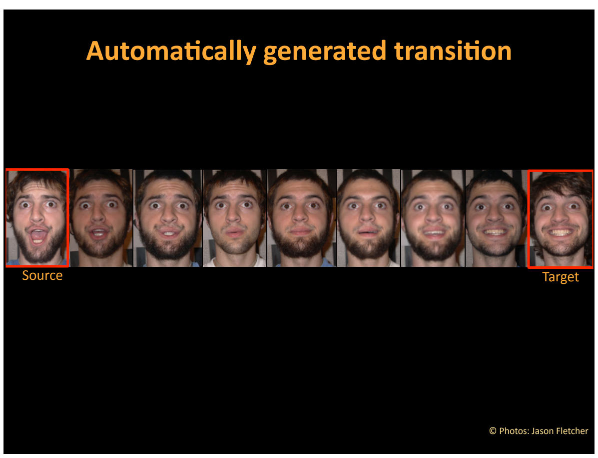

Facial expression similarity
Eyes
Mouth
Hair
Facial expression similarity
λ1 Eyes + λ2 Mouth + λ3 Hair

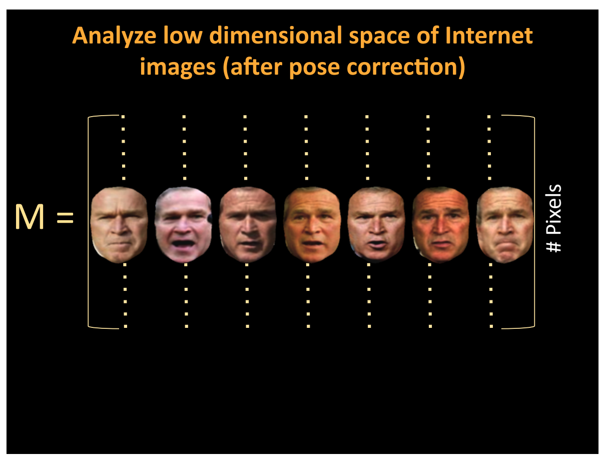

Facial expression similarity
Input
Nearest Neighbors
Eyes
Mouth
Hair


Represent the photo collection as a graph
Similarity between 2 photos
3D Head Pose similarity
Facial Expression similarity
Time similarity

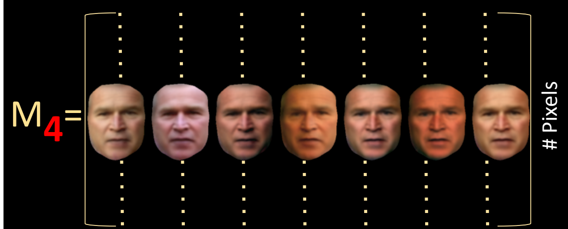

Represent the photo collection as a graph
Similarity between 2 photos
3D Head Pose similarity
Facial Expression similarity
Time similarity


Represent the photo collection as a graph
Similarity between 2 photos
3D Head Pose similarity
Facial Expression similarity
Time similarity


Automatically generated transition
Source
Target
© Photos: Jason Fletcher


Analyze low dimensional space of Internet images (after pose correction)
M =
Pixels

Analyze low dimensional space of Internet images (after pose correction)



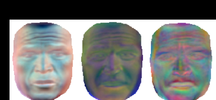
Expression gets normalized Shading captured

Why expression normalized by rank=4?

Lighting dominates facial expression

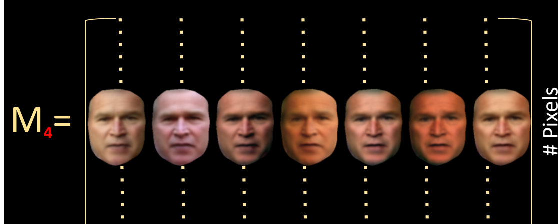
1) Rank 4 for rigid has >90% energy
   Ramamoorthi & Hanrahan'01, Basri & Jacobs'01
2) Lighting affects all pixels (dense)
3) Expression (non-rigid) affects mostly edge pixels (sparse) & captured by rank=5,6,…

Kemelmacher and Seitz, *Collection Flow*, CVPR'12

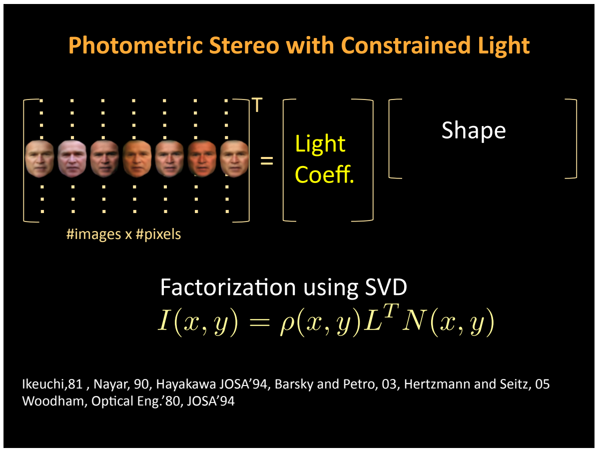


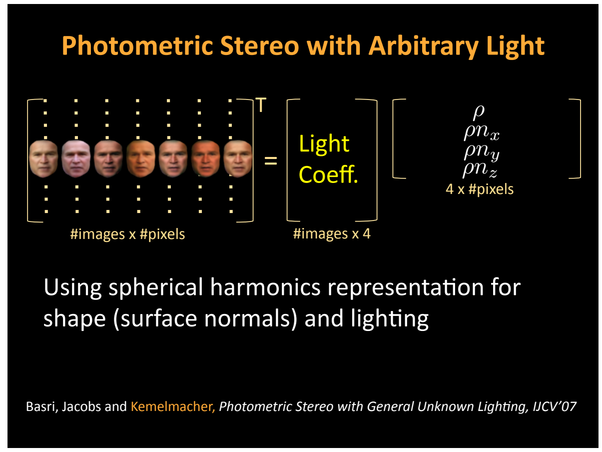
…

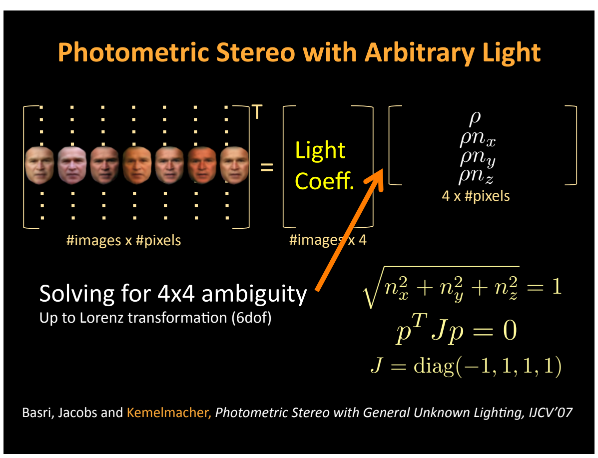
This is actually a photometric stereo setup

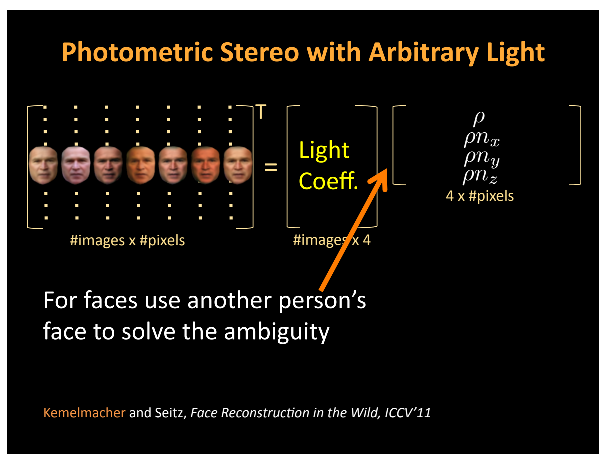


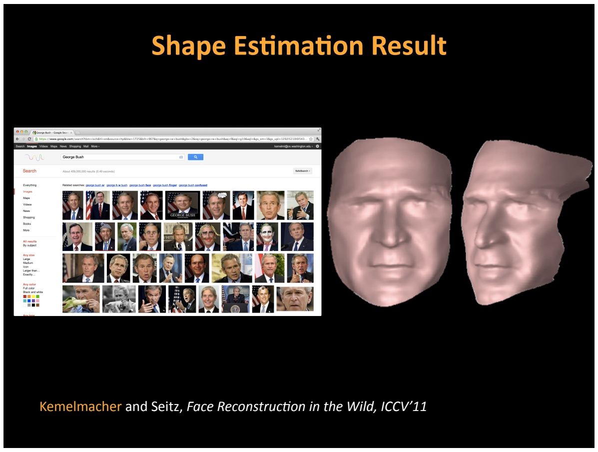
rigid object, fixed pose, varying lighting

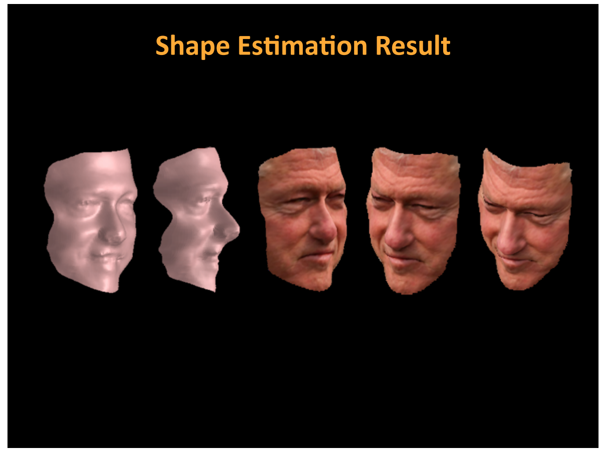

Photometric Stereo with Constrained Light
T
=
Light
Coeff.
Shape
#images x #pixels
Factorization using SVD
$I(x,y) = \rho(x,y)L^T N(x,y)$
Ikeuchi,81 , Nayar, 90, Hayakawa JOSA'94, Barsky and Petro, 03, Hertzmann and Seitz, 05
Woodham, Optical Eng.'80, JOSA'94

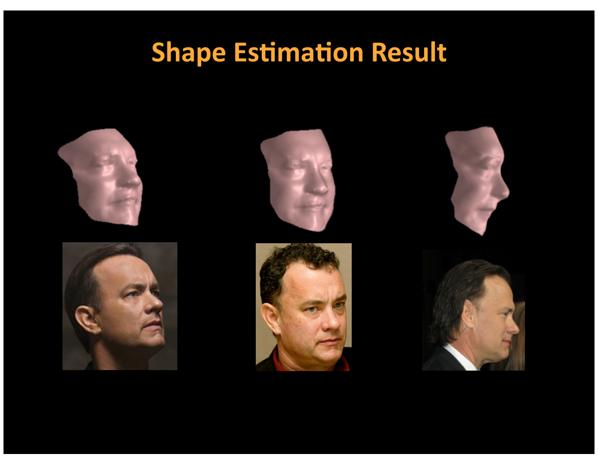

Photometric Stereo with Arbitrary Light
T
=
Light
Coeff.
ρ
ρn_x
ρn_y
ρn_z
4 x #pixels
#images x #pixels
#images x 4
Using spherical harmonics representation for shape (surface normals) and lighting
Basri, Jacobs and Kemelmacher, *Photometric Stereo with General Unknown Lighting, IJCV'07*

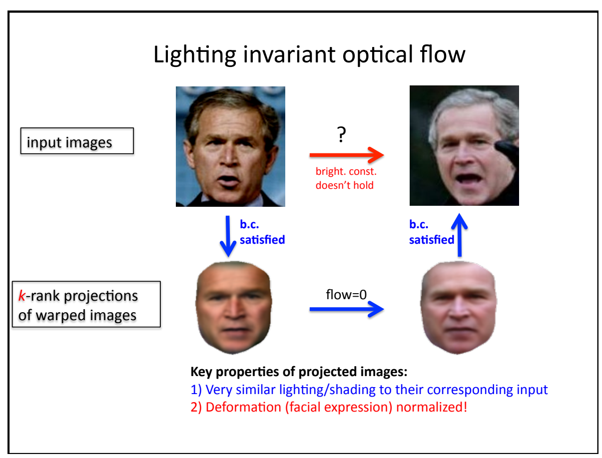

Photometric Stereo with Arbitrary Light
T
=
Light
Coeff.
ρ
ρn_x
ρn_y
ρn_z
4 x #pixels
#images x #pixels
#images x 4
Solving for 4x4 ambiguity
Up to Lorenz transformation (6dof)
$\sqrt{n_x^2 + n_y^2 + n_z^2} = 1$
$p^T J p = 0$
$J = \mathrm{diag}(-1, 1, 1, 1)$
Basri, Jacobs and Kemelmacher, *Photometric Stereo with General Unknown Lighting, IJCV'07*


Photometric Stereo with Arbitrary Light
T
=
Light
Coeff.
ρ
ρn_x
ρn_y
ρn_z
4 x #pixels
#images x #pixels
#images x 4
For faces use another person's face to solve the ambiguity
Kemelmacher and Seitz, *Face Reconstruction in the Wild, ICCV'11*


Shape Estimation Result
Kemelmacher and Seitz, *Face Reconstruction in the Wild, ICCV'11*


Shape Estimation Result


Shape Estimation Result

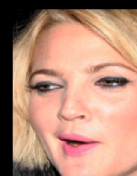

Lighting invariant optical flow
input images
?
bright. const. doesn't hold
b.c. satisfied
b.c. satisfied
k-rank projections of warped images
flow=0
Key properties of projected images:
1) Very similar lighting/shading to their corresponding input
2) Deformation (facial expression) normalized!

Result of flow estimation applied to morphing

Accounts for lighting
and shape changes

Challenging for state of the art

Note the distortions
due to violation of the
brightness constancy



C. Liu. PhD thesis, MIT, 2009
Brox et al. ECCV'04
Bruhn et al. IJCV'05

State of the art flow on unstructured photos

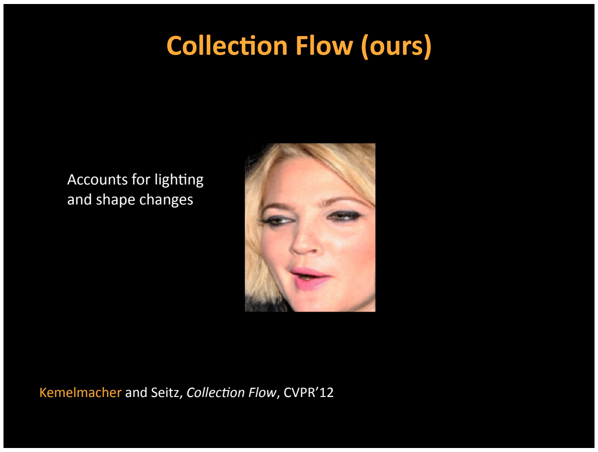
TV-L_1

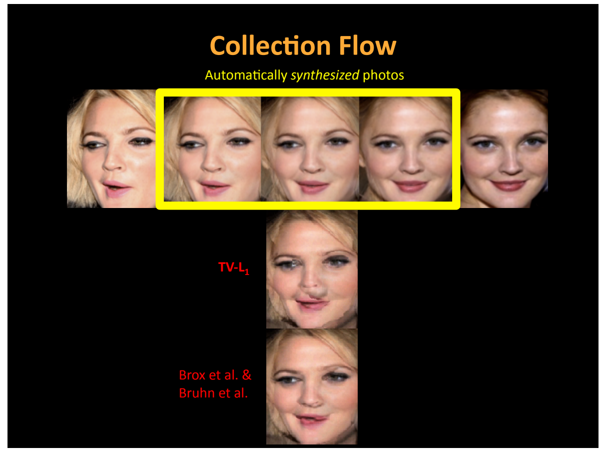
Chambolle & Pock, J. Mathematical Imaging and Vision 2011

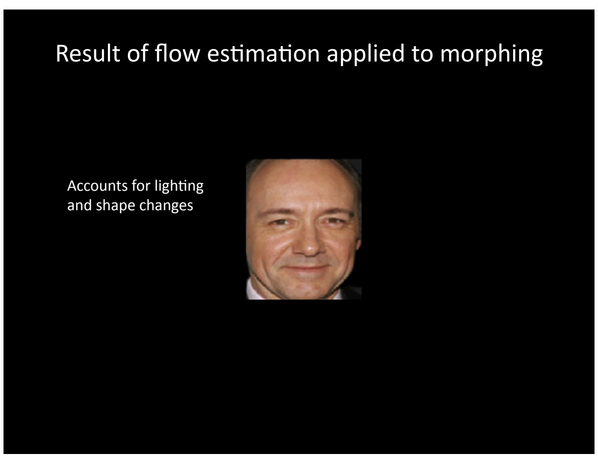
State of the art flow on unstructured photos

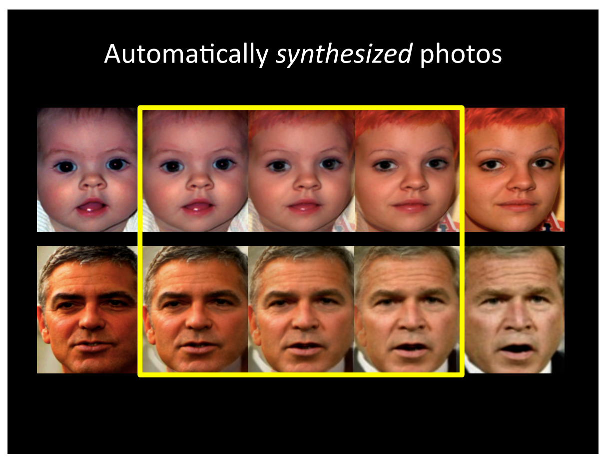
SIFT Flow

C. Liu, J. Yuen, and A. Torralba. Sift flow: Dense correspondence across scenes.. PAMI'11


Collection Flow (ours)
Accounts for lighting
and shape changes
Kemelmacher and Seitz, *Collection Flow*, CVPR'12

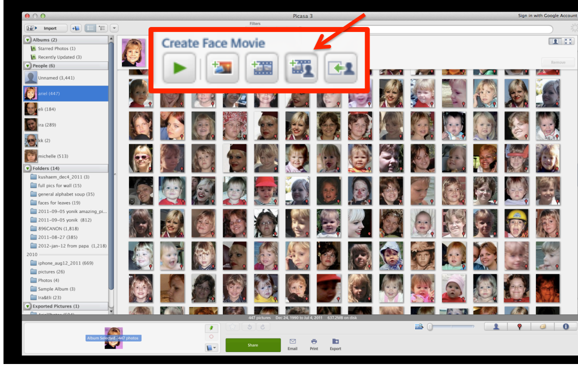

Collection Flow
Automatically *synthesized* photos
TV-L_1
Brox et al. &
Bruhn et al.


Result of flow estimation applied to morphing
Accounts for lighting
and shape changes


Automatically *synthesized* photos

Animation from Internet photos

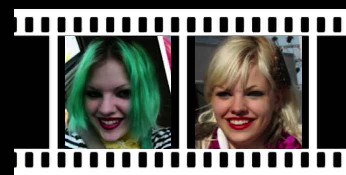
Face Movie feature in Google's Picasa



Being John Malkovich

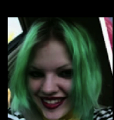
Kemelmacher, Sankar, Shechtman, Seitz
Being John Malkovich, ECCV'10

How to animate the resulted photo sequence?



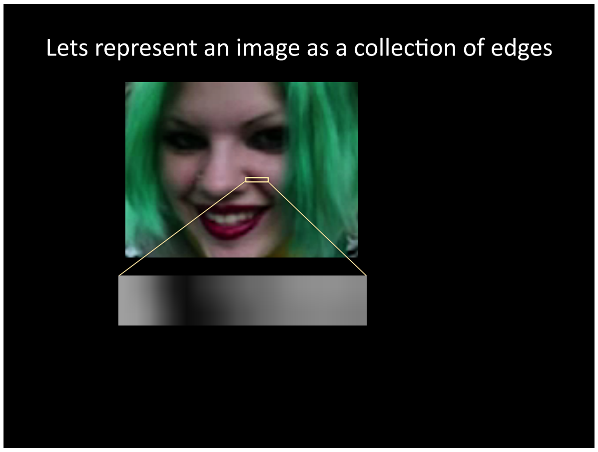
Linear interpolation (cross-dissolve)
between 2 images
can create physical motion

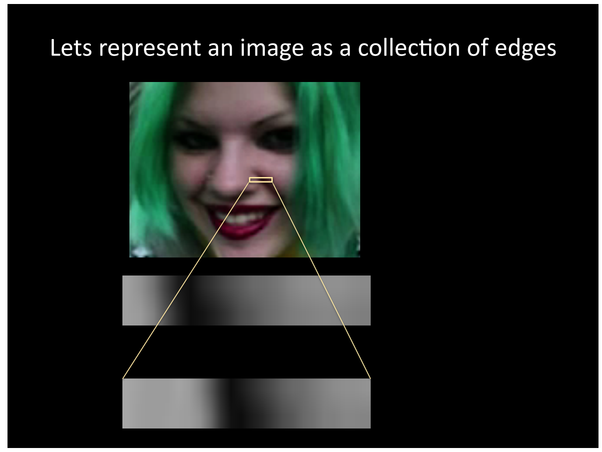
What is Cross-Dissolve?

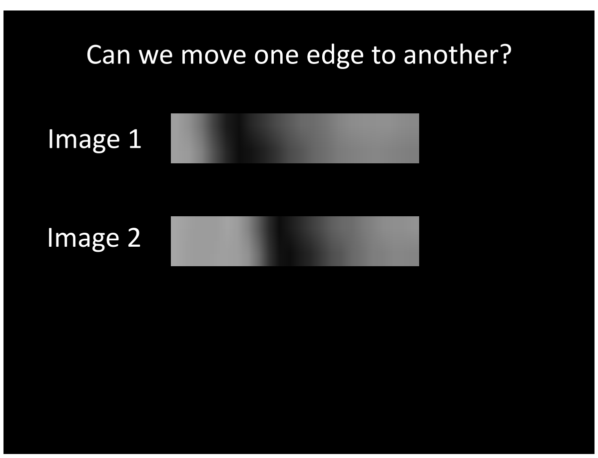
FADE OUT  + FADE IN

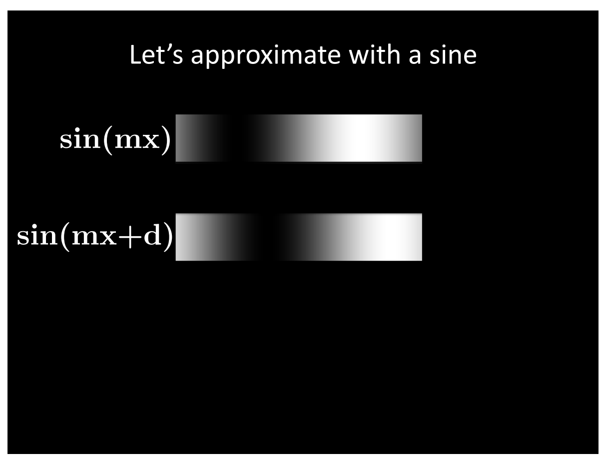
Cross-Dissolve produces motion

Why?


Lets represent an image as a collection of edges


Lets represent an image as a collection of edges


Can we move one edge to another?
Image 1
Image 2

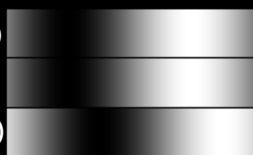

Let's approximate with a sine
sin(mx)
sin(mx+d)

While doing cross-dissolve we see motion!

$\sin(\mathrm{mx})$

Cross dissolve

$\sin(\mathrm{mx+d})$

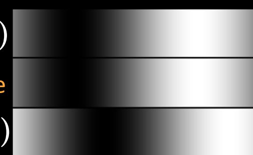


It is actually a physical motion!

$\sin(\mathrm{mx})$

Cross dissolve

$\sin(\mathrm{mx+d})$



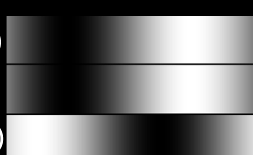
$\mathrm{c}\ \sin(\mathrm{mx+k})$

Motion effect with larger phase shift

$\sin(\mathrm{mx})$

Cross dissolve

$\sin(\mathrm{mx+d})$

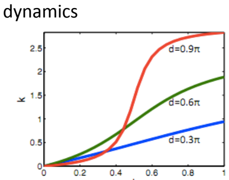


$\mathrm{c}\ \sin(\mathrm{mx+k})$

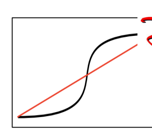
The motion is non linear

$$k = \arctan \frac{t \sin d}{(1-t) + t \cos d}$$

There is slow-in slow-out dynamics



Rule of Thumb in animation



Walt Disney Animation Principles, 1930
J. Lasseter, SIGGRAPH'87

Kemelmacher et al. SIGGRAPH'11

When motion is produced?

- Approximately same frequency edges
- The distance edges can move depends on the frequency

Cross-dissolve produces edge rotation

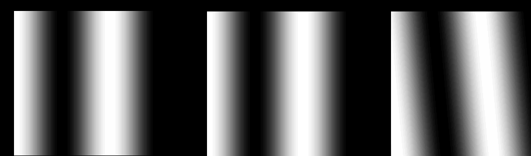


Image 1 | Cross dissolve = edge rotation (proof in the paper) | Image 2

Cross-dissolve produces light motion

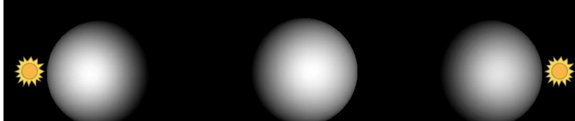


Image 1 | Cross dissolve = light motion (proof in the paper) | Image 2