

## Collection Flow

Ira Kemelmacher-Shlizerman  
University of Washington  
kemelmi@cs.washington.edu

Steven M. Seitz  
University of Washington and Google Inc.  
seitz@cs.washington.edu



Figure 1. Given a pair of images (first and last in the sequence) the in-between photos are automatically synthesized using our flow estimation method. Note the significant variation in lighting and facial expression between the two input photos.

### Abstract

*Computing optical flow between any pair of Internet face photos is challenging for most current state of the art flow estimation methods due to differences in illumination, pose, and geometry. We show that flow estimation can be dramatically improved by leveraging a large photo collection of the same (or similar) object. In particular, consider the case of photos of a celebrity from Google Image Search. Any two such photos may have different facial expression, lighting and face orientation. The key idea is that instead of computing flow directly between the input pair  $(I, J)$ , we compute versions of the images  $(I', J')$  in which facial expressions and pose are normalized while lighting is preserved. This is achieved by iteratively projecting each photo onto an appearance subspace formed from the full photo collection. The desired flow is obtained through concatenation of flows  $(I \rightarrow I') \circ (J' \rightarrow J)$ . Our approach can be used with any two-frame optical flow algorithm, and significantly boosts the performance of the algorithm by providing invariance to lighting and shape changes.*

### 1. Introduction

Despite significant progress in optical flow research, most methods are based on an assumption of brightness constancy; hence performance significantly degrades under differences in shading, due to lighting variations or changes in surface normals. An extreme case is estimating flow between photos of George Clooney and George W. Bush above, in which pixel intensities vary dramatically between the two input photos (first and last photo in Fig. 1).

Rather than considering optical flow as a purely pairwise

correspondence problem, in this paper we propose to leverage a large *collection* of similar photos to enable flow computation with changes in lighting and shape. As such, we are motivated by the vast stores of imagery available on the Internet and in personal photo collections; for any photo, you can find many more just like it. The case of faces is particularly interesting—we have access to thousands of photos of any celebrity (through Internet search), and a similarly large number of friends and family members (through tools like iPhoto or Facebook). Such collections implicitly describe the “appearance space” of an individual by capturing the subject under many poses, lighting, and expressions. The challenge is to model and leverage this appearance space for optical flow estimation. While we focus our attention on face applications, the approach does not employ face-specific assumptions and may be applicable more broadly to other families of objects that can be aligned to a common reference.

Instead of inventing a new optical flow algorithm, we seek to boost the performance of existing algorithms by normalizing (removing) confounding factors. For example, suppose we were able to normalize illumination, i.e., re-render the second image with the illumination of the first—this would likely lead to better flow performance with existing algorithms. However, this re-rendering task is not at all straightforward, as it would seem to require estimating the 3D shape corresponding to the second image and the lighting in both images. And even if we were able to do this, note that matching illumination is not sufficient, as the surface normals may change with the facial expression, leading to shading differences even with the same illumination. Similarly, the albedo or image exposure may also vary between shots.

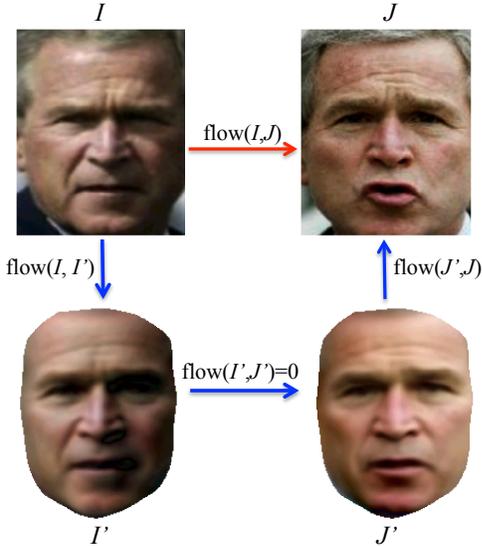


Figure 2. To compute flow from image  $I$  to  $J$ , we first project both images to a common neutral expression (brow relaxes in  $I'$ , mouth closes in  $J'$ ), then compute  $\text{flow}(I, I')$  and  $\text{flow}(J', J)$ . Observe how the neutral projections retain the shading, exposure, and color balance of the original; brightness constancy is much better satisfied between  $(I, I')$  and between  $(J, J')$  compared to the original image pair  $(I, J)$ . The desired flow is then obtained through concatenation of  $(I \rightarrow I') \circ (J' \rightarrow J)$ .

Rather than normalize shading, we propose to normalize *expression*. The key idea is to project each input photo onto a low-dimensional appearance subspace that retains the shading but converts the expression to neutral. Rather than computing flow between the image pair  $(I, J)$  directly, we instead compute flow between each input photo  $I$  and its normalized version  $I'$ , which yields the flow to a common expression (Figure 2). The flow between the input image pair is obtained through concatenation of  $(I \rightarrow I') \circ (J' \rightarrow J)$ .

Our approach is based on the well-known observation [6, 21, 19, 1, 7, 2, 8] that in image collections with lighting variations, the first few eigenfaces (PCA components) tend to capture shading effects very well. While these prior results apply only to rigid scenes, in this paper we observe that the first few PCA components of large image collections of faces with expressions (non-rigid shape variations) and lighting variations capture mostly the shading, i.e. shading changes dominate expression changes. Hence, projecting onto a low-rank subspace has the effect of *removing* most expression differences among the photos. In practice, however, the low-rank projection tends to smooth out fine details (which are important for optical flow). We therefore introduce an iterative approach that computes flow and warps each image to its low rank projection, re-estimates the low-rank subspace, and repeats until convergence. The

resulting subspace does a much better job of matching the illumination, shading changes, albedo, and imaging conditions, e.g., non linear camera response and white balance, while still reducing the expression to neutral.

Another advantage of our approach is that it requires only  $O(n)$  flow computations to derive pairwise flows across a collection of  $n$  images, instead of running optical flow for all  $O(n^2)$  image pairs. This performance improvement is significant for large collections, and is achieved by computing flow to a neutral reference and deriving the pairwise flows via concatenation.

The paper is organized as follows. Section 2 summarizes related work in optical flow. Section 3 introduces the idea of expression normalization and analyzes its properties. Section 4 introduces the collection flow algorithm, and Section 5 presents results.

## 2. Related work

Classical work on optical flow is based on an assumption of brightness constancy. While most modern optical flow methods also employ this constraint, there are a number of notable exceptions. In particular, several researchers have explored ways to generalize the optical flow constraint equation to handle certain types of intensity changes ranging from bias-gain variations [13], physically-based radiometric changes [12], and other parametric changes in the intensity field [17]. HaCohen et al. [10] solve for a global parametric color change in concert with solving for optical flow. All of these methods operate by introducing additional parameters to solve for, and thus require more reliance on smoothness to regularize flow. Another avenue for coping with illumination changes is to incorporate more robust matching techniques, e.g., SIFT flow [16].

Beginning with Pentland [18], several authors [25, 22] have explored the special case of optical flow generated by a *rigid* scene moving under fixed or variable illumination. In these cases, the lighting and/or object motion is usually assumed to be known, and the problem reduces to reconstructing the scene geometry.

More related to our work, Hager and Belhumeur demonstrated illumination-invariant tracking via linear combination of a set of template images of an object under different lighting [11]. While they limited their scope to simple parametric motions (e.g., rotation, translation, scale), our approach is inspired in part by their insights.

For the specific case of faces, there is a large literature on tracking and alignment techniques. However, few of these techniques provide dense optical flow fields, are fully automated, and work robustly in the presence of illumination changes. [26] compute nonrigid facial motion under “moderate” illumination changes, by introducing an outlier model and allowing for local bias-gain variations.

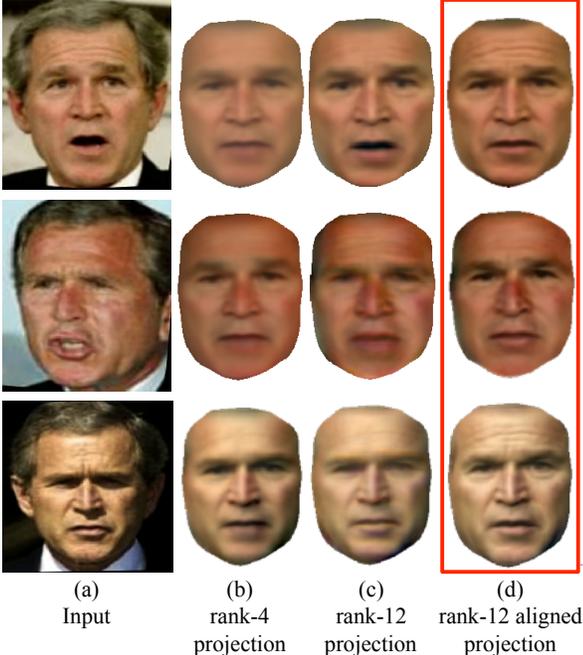


Figure 3. (a) Selected input photos, (b) rank-4 projection: expression is neutralized, (c) rank-12 projection: better match to input photo (e.g. sharper), but expression no longer neutral, (d) rank-12 warped projection using our flow algorithm: much better matching to input photo while expression is neutralized. All input images were warped to frontal prior to rank projection.

### 3. Expression normalization

Suppose we are given a collection of photos of a person’s face captured “in the wild”, i.e., in unconstrained conditions. Suppose, for the moment, that the faces are all frontal (we will relax this later), but that face expression, lighting, albedo, image exposure, age, and other factors may change from one image to the next. Figure 3(a) shows a sample of such images downloaded from Google Image Search for George W. Bush. Our objective is to compute optical flow between any pair of such images.

Now try the following experiment: put all  $n$  photos into a matrix  $M$  where each column corresponds to the pixels in one image laid out in a vector. Compute the best rank-4 approximation  $M_4$  of  $M$ , by forming the singular value decomposition and setting singular values  $4 - n$  to zero. Now display each column of  $M_4$  as a 2D image and compare with the original photo (corresponding column of  $M$ ), as shown in Figure 3(b). The resulting images capture most of the original lighting and shading, but the expression has been changed to neutral!

Indeed, this phenomenon has been observed previously in the face recognition literature, as the first few *eigenfaces* are often dominated by shading effects (Figure 4). The same effect has also been observed recently in the context of pho-

tometric stereo [14]. However, this phenomenon is still not understood. In this section, we analyze the reasons for this behavior, bringing together known results in this area and contributing new observations. Furthermore, we identify limitations of this approach for expression normalization and propose more powerful expression normalization techniques. In Section 4, we present a method that leverages expression normalization for optical flow estimation.

### 3.1. Low-rank projection

Why does the expression get normalized under low-rank projection? For a rigid Lambertian scene under directional lighting,  $M$  is known to be low rank; rank-3 with no shadows [21], rank-9 with attached shadows [19, 1]. In particular, more than 90% of the image energy is in the first 4 basis images [7]. Similar theoretical [2, 8] and empirical [6] results have been shown for non-Lambertian scenes as well.

As these results apply only to rigid scenes, we now turn our attention to the *non-rigid* case. Our main observation is that the change in image intensities caused by non-rigid face motion is typically small compared to the effect of changing the illumination. The intuition is that face motion due to expression change (not head rotation) has three components: 1) changes in intensity caused by optical flow, and 2) shading changes caused by shape deformation (changing surface normals), and 3) changes in visibility (e.g., open mouth). The first component is significant only at edges, the second component is significant only at wrinkles and dimples, and the third is most pronounced only in the mouth and eyes—all effects are *sparse* in the image. These effects are dominated by the intensity changes induced by moving the light source, which affect *all* pixels and can be very large.

To formalize this argument, let’s assume *the lighting is fixed*, but the facial expression (geometry) changes between images  $I$  and  $I'$ . To facilitate analysis, we ignore occlusions and assume the motion is small enough that we can approximate the images as consecutive in time  $t$ . Assuming Lambertian reflectance and directional illumination, the image intensity at each image point  $(x, y)$  is given by:

$$\begin{aligned}
 I(x, y) &= \rho(x, y)(l_1 + \mathbf{l}^T \mathbf{n}(x, y)) \\
 I'(x + u, y + v) &= \rho(x, y)(l_1 + \mathbf{l}^T \mathbf{n}'(x + u, y + v))
 \end{aligned}$$

where  $\mathbf{u} = (u, v)$  is the flow,  $\rho(x, y)$  is the albedo (scalar for each point on the object),  $l_1$  is ambient and  $\mathbf{l}$  are directional lighting coefficients and  $\mathbf{n}(x, y)$  is the surface normal vector at each point on the surface. Linearizing  $I'$  leads to the following optical flow equation:

$$\frac{dI}{dt} = -\nabla I \mathbf{u} - \rho \mathbf{l}^T \frac{d}{dt} \mathbf{n}. \quad (1)$$

The left hand side of this equation is the pixel intensity differences between the images, and the right hand side explains these differences in terms of two components: a term

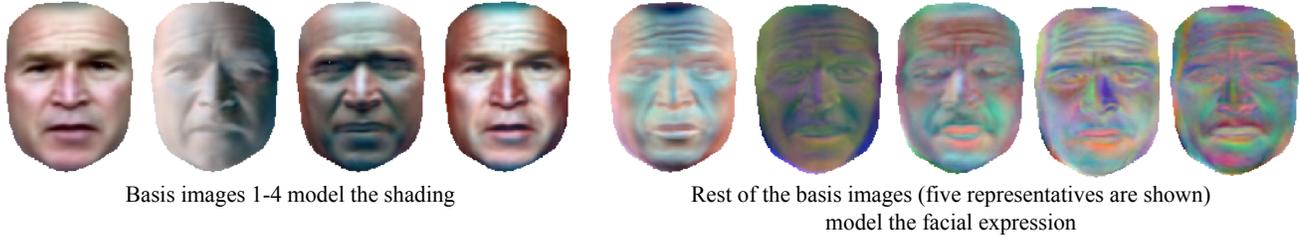


Figure 4. Left singular vectors for Internet images of George Bush, with magnitude decreasing left to right. Observe that the first 4 images span the shading variation of a neutral face, whereas subsequent images capture facial expression and other effects.

depending on image gradients, and a second term depending on changes in surface normal. Let's examine each of these terms. The first term  $\nabla I \mathbf{u}$  is significant only where the image gradient is large. However, large gradients are sparse in natural images and even more so for faces [24] which are dominated by smooth regions. Hence, this term will have a limited effect overall. The second term  $\rho \mathbf{l}^T \frac{d}{dt} \mathbf{n}$  captures the change in shading due to changing surface normals. Note, however, that changes in local surface orientation are somewhat limited due to the elastic tension of skin and constrained bone/muscle movement. For example, no matter how much you deform your face, most points on the right side of your face will have normals pointing to the right. Contrast this to lighting changes, which can create arbitrary changes in  $\mathbf{l}$  and affect nearly every pixel in the image. Therefore, both terms, and hence  $\frac{dI}{dt}$  are small relative to the intensity changes caused by large illumination changes.

In short, the reason why rank-4 projection normalizes expression is that 1) lighting changes dominate the variance in image pixels, hence the top singular vectors will model illumination effects, not expression changes, and 2) a rank-4 projection captures 90% of the shading effects due to illumination. Hence, a rank-4 projection will generally have the effect of normalizing the expression and roughly matching the lighting. We note this analysis applies only when the lighting variation in the image collection is large. If the light source is constant or moves less than the normals on the face, expression changes will dominate.

### 3.2. Higher rank projections

This expression normalization effect with rank-4 projection is very compelling, however it has a number of limitations. First, the rank-4 basis captures an *average* face, with fine details smoothed out (Figure 3(b)). Second, the illumination of the rank-4 projection will only roughly match that of the input image due to expression changes. Third, the rank-4 projection is not sufficient to capture the changes in surface shape due to the expression change, i.e., the surface normals are not precisely matched so brightness constancy will be violated to some extent. Finally, higher-rank projections may be needed to get a more accurate match to the

input image to account for effects like shaving a beard or getting a suntan which may cause very significant intensity changes over a large region of the face.

Figure 3(c) shows the result of a rank-12 projection instead of rank-4. Indeed, increasing the size of the basis results in a more faithful fit to the original photo. However, the expression normalization property (observed with rank-4) is lost with rank-12. In the rest of the paper, we will show how to capture higher order effects (most importantly to capture the intensity change due to surface shape variations) while retaining the normalization property.

### 3.3. Warped projections

Suppose we had precise pixel-to-pixel correspondence and could map all of the input photos onto a single reference expression. Ignoring occlusions, this allows us to remove optical flow effects, and explain the appearance changes purely in terms of geometry and reflectance changes. In particular, let's represent the key expressions using a set of  $k$  basis shapes, each with a set of surface normals  $\mathbf{n}_i(x, y)$ , and albedos  $\rho_i(x, y)$  for  $i = 1 \dots k$ . By combining these basis expressions, we can represent any face in their linear span<sup>1</sup>. If the scene is Lambertian, we can thus capture this space of expressions with a rank  $4k$  basis. Note that this representation allows capturing not just changes in shape, but also changes in albedo, e.g., due to growing a beard, getting a suntan, or applying makeup. Similar arguments apply for modeling exposure changes or nonlinear camera response curves (approximated as linear combinations, as in [9]).

Hence, low-rank approximation is an even more powerful tool when the input photos can be aligned. Figure 3(d) shows the result of a rank-12 projection on an *warped* image set generated using the method in Section 4. Note how both lighting and fine details (e.g., red mark on his nose in center image) are much more accurately match between the aligned result and the original input images, while still maintaining the expression normalization property. In the next section, we introduce an iterative approach for con-

<sup>1</sup>With the caveat that the normals must be integrable, or will be projected onto the closest integrable set.

structuring a warped face space and solving for flow in tandem.

#### 4. Flow estimation algorithm

We seek to compute optical flow between any pair of  $n$  photos from a large collection of a person’s face. Because lighting changes degrade optical flow performance, we propose to leverage expression normalization as shown in Figure 2. I.e., given a pair of images  $(I, J)$ , we first compute expression normalized versions  $(I', J')$  and compute flow from  $I$  to  $I'$  and  $J'$  to  $J$ , the composition of which yields the desired flow field from  $I$  to  $J$ . Hence, solving the pairwise flow problem reduces to computing the flow  $(I, I')$  between each photo and its expression normalized version. This reduction also enables calculating all  $n^2$  pairwise flows with only  $O(n)$  runs of an optical flow algorithm.

We begin by computing  $I'$  using rank-4 projection, as described in Section 3.1, and estimate flow between  $I$  and  $I'$ . We could stop here. However, due to the limitations with rank-4 projection, as discussed in Section 3.2, better results can be obtained by producing a *warped projection*, as described in Section 3.3. We accomplish this by warping each input photo to its normalized expression, using the recovered flow. We iterate these steps (project, compute flow, warp) until convergence while increasing the projection rank gradually in each iteration, enabling progressively more accurate image reconstructions. More details are provided below.

Note that *any* flow algorithm can be used to compute these intermediate flow fields  $(I, I')$ —we are not inventing a new flow algorithm, rather adjusting the input images to fit the operation range of any state of the art optical flow estimation method by leveraging large photo collections.

##### 4.1. Iterative alignment

Given a set of frontal or pose-corrected images (details on pose correction in Section 5), we apply the following algorithm:

1.  $k = 4$ , initialize flow fields  $F_i$  to identity, stack input images as columns of matrix  $M$ .
2. compute rank- $k$  singular value decomposition  $M_k$  of  $M$ , extract projected images  $I'_i$  from columns of  $M_k$
3. compute flow  $F_i$  from  $I'_i$  to  $I_i$
4. inverse warp  $I_i$  to  $I'_i$  using flow  $F_i$
5.  $k = k + 1$
6. repeat step 2 until flow converges

In every iteration, we both improve the alignment, and increase the rank of the projection, allowing more accurate modeling of fine details. It is important that the rank be small initially and increase slowly, to avoid capturing expression changes in the basis (we seek a projection that

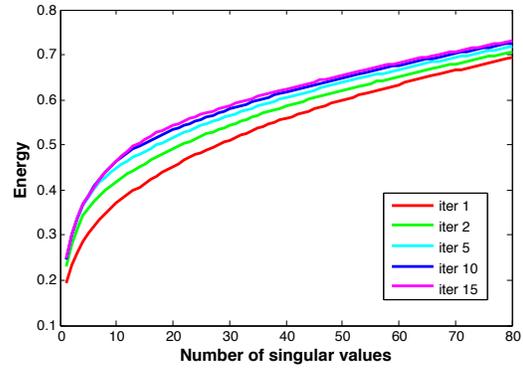


Figure 5. Plot of % total energy (y-axis) captured by singular values 1-x as a function of collection flow iteration. E.g., after 15 iterations, the first 15 singular values capture 50 % of the energy (up from 40% initially). This shows that after the iterative procedure, significantly more of the energy is captured in the first few singular values. I.e., the aligned images are better fit by the linear model than the original set of images.

normalizes expression). In early iterations, the low-rank projection strongly regularizes the alignment, compensating for imperfect flow. Then, as more basis images are added, the projection quality and flow improves, thus improving the alignment. The additional basis images add more degrees of freedom in the surface normals (effectively adding additional basis “shapes”) and albedos, as discussed in Section 3.3, enabling the projection to fit not just the lighting, but also changes in intensity due to shape and reflectance of the input.

Specifically, by increasing the projection rank and re-warping the images with each iteration, the projection accounts better for the surface normal difference term  $\rho \mathbf{l}^T \frac{d}{dt} \mathbf{n}$  in Eq. (1). This term is usually ignored in most optical flow methods [23]. Figure 5 plots the improvement in alignment quality over iterations.

So far, we have assumed that the lighting variations are large (which is typical in Internet collections). If this is not the case, choosing a smaller initial value for  $k$  could make sense. If the lighting is constant, for example in the case of a photos in a high school yearbook,  $k = 1$  may give the best results; in this case, each face will initially be registered to the average face (with linear intensity scaling).

## 5. Experiments

In this section, we discuss results of our algorithm on several image collections of celebrities downloaded from Google Image Search and a personal photo collection. We first describe how we pre-process the downloaded collections and correct for rigid pose variation. We further discuss details related to the flow estimation approach and running times, show our expression warping results and compare

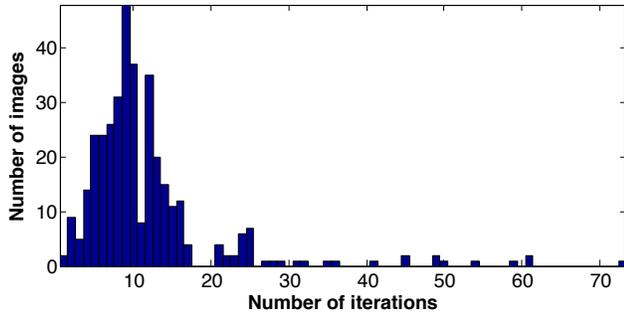


Figure 6. Histogram of number of images (of George Bush) vs. number of iterations needed to converge. E.g., most images required fewer than 15 iterations.

with state-of-the-art optical flow algorithms.

**Preprocessing and rigid pose correction:** We downloaded about 800 photos per person and used Picasa to identify and recognize faces. We used the preprocessing pipeline from [14] that includes face and fiducials detection, pose estimation and warping to frontal pose, and masking the face region, which successfully registers about 500 photos per person. We eliminate photos with extreme poses (filtered by pitch and yaw, i.e. we kept photos within the range of  $\pm 25$  degrees yaw and  $\pm 5$  degrees pitch angles), leaving about 400 photos per person. Fig. 7 (a) shows example input images, and (b) the same images warped to frontal position by this procedure. Flow is then computed on pose corrected and masked images.

**Flow estimation details and running times:** We estimated flow between pairs of photos of the same person and also between different people. The approach operates robustly for a wide range of variations, e.g., pose, expression, lighting, age, and identity. For collections containing a single person we used  $N = 400$  photos. For collections containing two people, we used  $N = 600$  (300 for each person). The algorithm we chose for flow estimation between each image and its low rank projection is Ce Liu’s [15] implementation of Brox et al. [3] combined with Bruhn et al. [4]. We use the following parameters in their implementation:  $\alpha = 0.01$ ,  $\text{ratio}=0.75$ ,  $\text{minWidth}=60$ ,  $\text{nOuterFPIterations}=5$ ,  $\text{nInnerFPIterations}=1$ ,  $\text{nCGIterations}=50$ . The running time of each flow estimation is around 4 sec.

The total running time of our algorithm is therefore  $\text{Niter} * (\text{SinglePairFlow} * N + \text{pcaTime})$  where  $\text{SinglePairFlow}$  is the time takes for flow estimation between one pair of photos,  $N$  number of photos in the collection and  $\text{pcaTime}$  is the time takes to compute the low rank projection images at each iteration. The number of iterations is plotted in Fig. 6. We observed that images that are similar to many others in the collection typically need fewer iterations (e.g., less than 10 iterations), whereas rare poses, expressions, or illuminations require more. We stop computing flow for

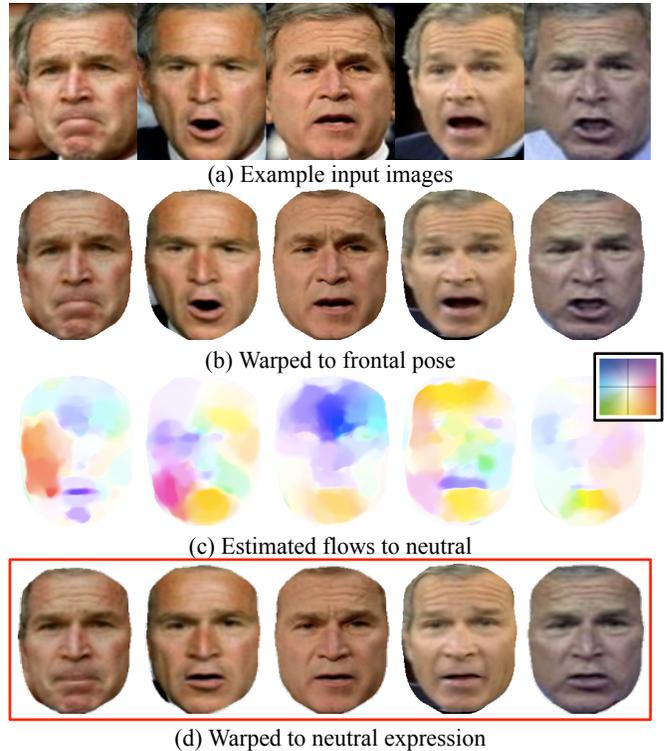


Figure 7. Pose normalization pipeline. (a) A few example images, (b) rigid pose correction, (c) estimated flow to neutral (color-code in the upper right), and (d) images warped using the flow.

an image when the L2 norm of the difference between the current estimated flow and the one in previous iteration is below a fixed threshold=20. To estimate the low rank projections we use the randomized PCA algorithm of Rokhlin et al. [20] that typically takes 0.8 sec on a matrix produced from 400 images of size  $200 \times 150$ . Running optical flow on all pairs with 400 photos would take 177 hours. Using our collection flow method with 15 iterations per image requires 7 hours (this is the  $O(n)$  vs.  $O(n^2)$  savings).

**Facial expression normalization:** Fig. 7 (c) shows several estimated flows and (d) shows the faces warped using the flows to neutral expression.

**Morphing:** We apply a standard morph effect by warping each input photo to the desired in-between (linearly interpolating the flow) and cross-fading the results. We further augment the effect by linearly interpolating the *poses* of the two input images and applying pose correction. Fig. 8 shows several results illustrating changes in expression, pose, lighting, age, and identity (input images at far left and right). **See supplementary material for videos of these transitions.** The fact that the in-between photos look sharp and lack ghosting artifacts is an indication of high quality flow.

**Comparison to leading flow algorithms:** We compare collection flow to other state-of-the-art flow methods: 1)

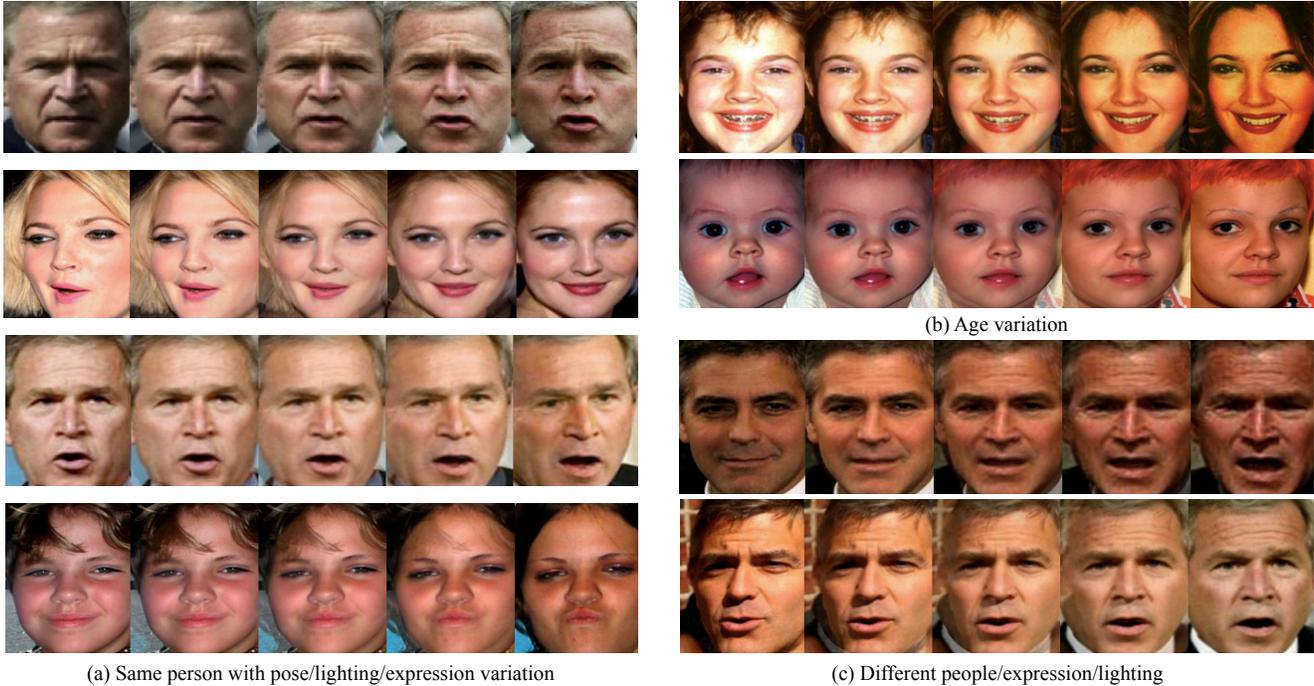


Figure 8. Morphing sequences: far left and right images are given as input and all the in-between views are *automatically synthesized* by our method.

Liu’s [15] implementation of Brox et al. [3] combined with Bruhn et al. [4], 2) SIFT flow [16] and 3) dense duality based TV- $L^1$  optical flow by Chambolle and Pock [5]. Fig. 9 presents the results. We ran all flow algorithms on pose-corrected and masked images (results are worse on the original photos).  $I, J$  are the input images, (a) shows  $J$  warped to  $I$ , (b) vice versa, and (c) show the morphed image (at the midpoint of the transition). These input images are particularly challenging, due to the dramatic illumination and shape differences (brightness constancy is strongly violated), and collection flow produces significantly better results. For image pairs with less variations, the performance difference between algorithms is less significant. Please see the supplemental material for other comparisons and video versions of the morphs.

## 6. Summary

In this paper, we presented a method for optical flow estimation between a pair of images allowing variations due to lighting, non-rigid surface shape changes, and pose. Our key idea is to estimate flow between the input images by leveraging large photo collections. Traditional optical flow estimation methods assume brightness constancy and resort to smoothing to account for its violations (e.g., when the input images have different lighting). In contrast we have shown that lighting and shape variations can be accounted for by projecting the input images to a reduced appearance

space constructed from photos of the same person. This reduction dramatically improves flow computation in unstructured photo collections. We have also analyzed the low dimensional representation of a person’s photos in the presence of both lighting and non rigid shape variations. While we focused on faces in this paper our approach maybe applicable more generally.

## Acknowledgements

This work was supported in part by National Science Foundation grant IIS-0811878, the University of Washington Animation Research Labs, Adobe, Google, and Microsoft.

## References

- [1] R. Basri and D. W. Jacobs. Lambertian reflectance and linear subspaces. *PAMI*, 25(2):218–233, 2003. 2, 3
- [2] P. N. Belhumeur and D. Kriegman. What is the set of images of an object under all possible lighting conditions? *IJCV*, 28(3):245–260, 1998. 2, 3
- [3] T. Brox, A. Bruhn, N. Papenbergh, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *ECCV*, pages 25–36, 2004. 6, 7
- [4] A. Bruhn, J. Weickert, and C. Schnrr. Lucas/kanade meets horn/schunck: Combining local and global optic flow methods. *IJCV*, 61:211–231, 2005. 6, 7

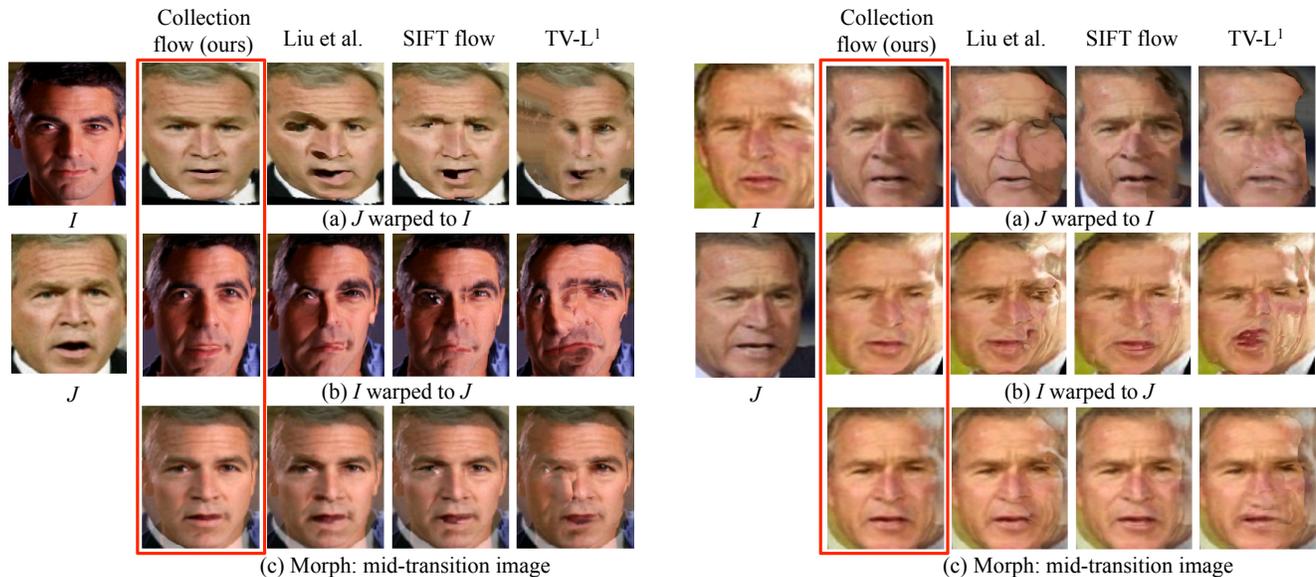


Figure 9. Comparison between our results and leading optical flow methods. Note the distortions that appear in images warped using other methods—these are due to significant lighting variations which traditional optical flow methods are not designed to handle.

- [5] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Mathematical Imaging and Vision*, 40(1):120–145, 2011. 7
- [6] R. Epstein, P. Hallinan, and A. Yuille. 5 plus or minus two eigenimages suffice: An empirical investigation of low-dimensional lighting models. In *Proc. IEEE Workshop on Physics-based Modeling in Computer Vision*, pages 108–116, 1995. 2, 3
- [7] D. Frolova, D. Simakov, and R. Basri. Accuracy of spherical harmonic approximations for images of lambertian objects under far and near lighting. In *ECCV*, pages 574–587, 2004. 2, 3
- [8] R. Garg, H. Du, S. M. Seitz, and N. Snavely. The dimensionality of scene appearance. In *ICCV*, pages 1917–1924, 2009. 2, 3
- [9] M. Grossberg and S. Nayar. Modeling the Space of Camera Response Functions. *PAMI*, 26(10):1272–1282, Oct 2004. 4
- [10] Y. HaCohen, E. Shechtman, D. B. Goldman, and D. Lischinski. Non-rigid dense correspondence with applications for image enhancement. *ACM Trans. Graph.*, 30(4):70, 2011. 2
- [11] G. D. Hager and P. N. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *PAMI*, 20:1025–1039, 1998. 2
- [12] H. W. Haussecker and D. J. Fleet. Computing optical flow with physical models of brightness variation. *PAMI*, 23(6):661–673, 2001. 2
- [13] H. Jin, P. Favaro, and S. Soatto. Real-time feature tracking and outlier rejection with changes in illumination. In *ICCV*, pages 684–689, 2001. 2
- [14] I. Kemelmacher-Shlizerman and S. M. Seitz. Face reconstruction in the wild. In *ICCV*, 2011. 3, 6
- [15] C. Liu. *Beyond Pixels: Exploring New Representations and Applications for Motion Analysis*. PhD thesis, MIT, 2009. 6, 7
- [16] C. Liu, J. Yuen, and A. Torralba. Sift flow: Dense correspondence across scenes and its applications. *PAMI*, 33(5):978–994, 2011. 2, 7
- [17] S. Negahdaripour. Revised definition of optical flow: Integration of radiometric and geometric cues for dynamic scene analysis. *PAMI*, 20(9):961–979, 1998. 2
- [18] A. Pentland. Photometric motion. *PAMI*, 13:879–890, 1991. 2
- [19] R. Ramamoorthi and P. Hanrahan. A signal-processing framework for inverse rendering. In *SIGGRAPH*, pages 117–128, 2001. 2, 3
- [20] V. Rokhlin, A. Szelam, and M. Tygert. A randomized algorithm for principal component analysis. *SIAM J. Matrix Analysis Applications*, 31(3):1100–1124, 2009. 6
- [21] A. Shashua. Geometry and photometry in 3d visual recognition. Technical report, PhD thesis, M.I.T Artificial Intelligence Laboratory, 1992. 2, 3
- [22] D. Simakov, D. Frolova, and R. Basri. Dense shape reconstruction of a moving object under arbitrary, unknown lighting. In *ICCV*, pages 1202–1209, 2003. 2
- [23] S. Vedula, S. Baker, P. Rander, R. Collins, and T. Kanade. 3-d scene flow. *PAMI*, 27(3):475–480, 2005. 5
- [24] Y. Weiss. Deriving intrinsic images from image sequences. In *ICCV*, pages 68–75, 2001. 4
- [25] L. Zhang, B. Curless, A. Hertzmann, and S. M. Seitz. Shape and motion under varying illumination: Unifying structure from motion, photometric stereo, and multi-view stereo. In *ICCV*, pages 618–625, 2003. 2
- [26] J. Zhu, S. C. Hoi, and L. V. Gool. Unsupervised face alignment by robust nonrigid mapping. In *ICCV*, 2009. 2