Transfiguring Portraits

Ira Kemelmacher-Shlizerman*
Computer Science and Engineering, University of Washington

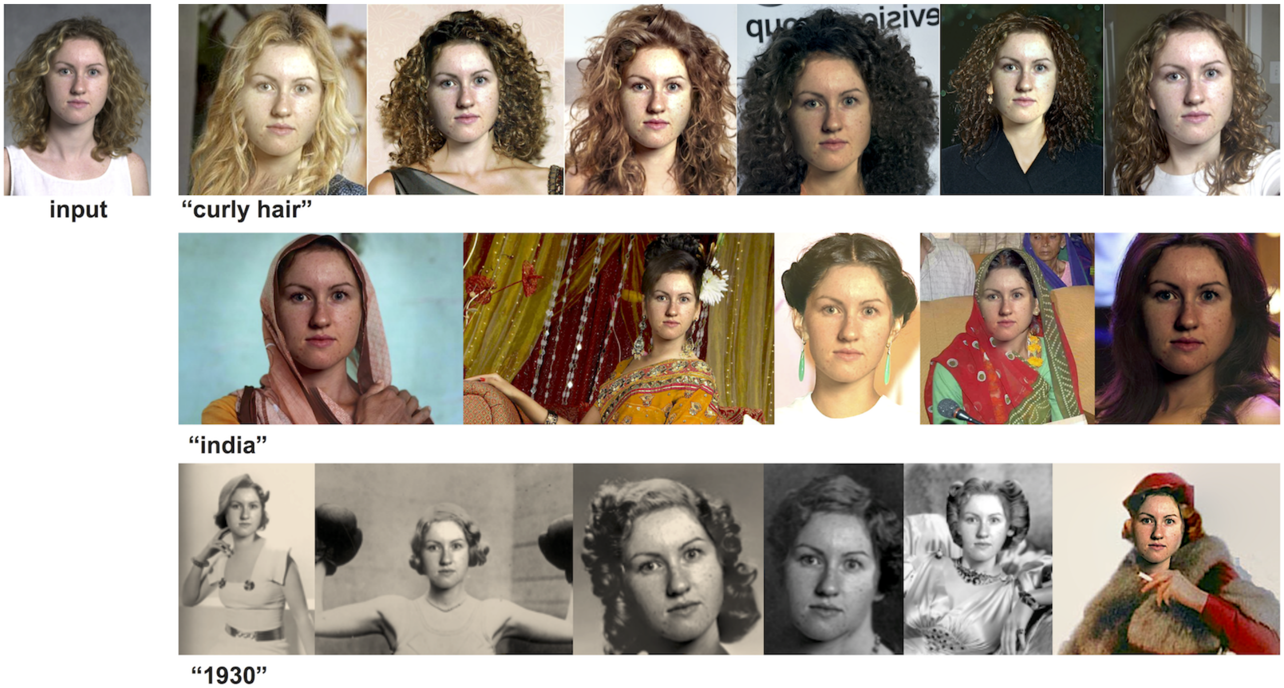

Figure 1: *Our system's goal is to let people imagine and explore how they may look like in a different country, era, hair style, hair color, age, and **anything** else that can be queried in an image search engine. The examples above show a single input photo (left) and automatically synthesized appearances of the input person with "curly hair" (top row), in "india" (2nd row), and at "1930" (3rd row).*

Abstract

People may look dramatically different by changing their hair color, hair style, when they grow older, in a different era style, or a different country or occupation. Some of those may transfigure appearance and inspire creative changes, some not, but how would we know without physically trying? We present a system that enables automatic synthesis of limitless numbers of appearances. A user inputs one or more photos (as many as they like) of his or her face, text queries an appearance of interest (just like they'd search an image search engine) and gets as output the input person in the queried appearance. Rather than fixing the number of queries or a dataset our system utilizes all the relevant and searchable images on the Internet, estimates a doppelgänger set for the inputs, and utilizes it to generate composites. We present a large number of examples on photos taken with completely unconstrained imaging conditions.

Keywords: face, big data, in the wild, appearance prediction, internet-based composites, portrait, imaginative.

Concepts: •**Computing methodologies** → **Image manipulation;** *Computational photography;*

*e-mail:kemelmi@cs.washington.edu

1 Introduction

In the show "The Americans" the Soviet KGB spies change identities just by changing hair style, color, and clothing style. When we go to a hair stylist we can browse a magazine with pictures of models and point to a photo of the kind of style we'd like to try. Occasionally we see celebrities and wonder if their style will fit us. Missing people are often times disguised by changing their hair. Actors make changes in their appearance to fit a particular role, for example Cate Blanchett depicted Bob Dylan's persona in "I'm Not There" . But how can we predict if some appearance change fits or not without physically trying? Or how can we explore a big variety of appearances, e.g., for people that are missing? This paper is about automatic synthesis of such changes.

The idea is to let one transfigure their appearance from images without making any physical changes (Fig. 1). Instead of addressing the challenging problem of artificially rendering realistic hair color and hair style in any input photo, e.g., from precaptured 3D models of hair styles, our key idea is that the billions of photos of people on the Internet have already captured all possible hair styles, colors, and more. For example, a search of "curly blond hair" returns thousands of variations of that hair style. Inspired by [Hays and Efros 2007] that leveraged the Internet for scene completion, we have created a tool that leverages the Internet to let people explore and imagine themselves in new styles and scenarios.

SIGGRAPH 2016 Technical Papers, July 24-28, 2016, Anaheim, CA
ISBN: 978-1-4503-ABCD-E/16/07
DOI: http://doi.acm.org/10.1145/9999997.9999999

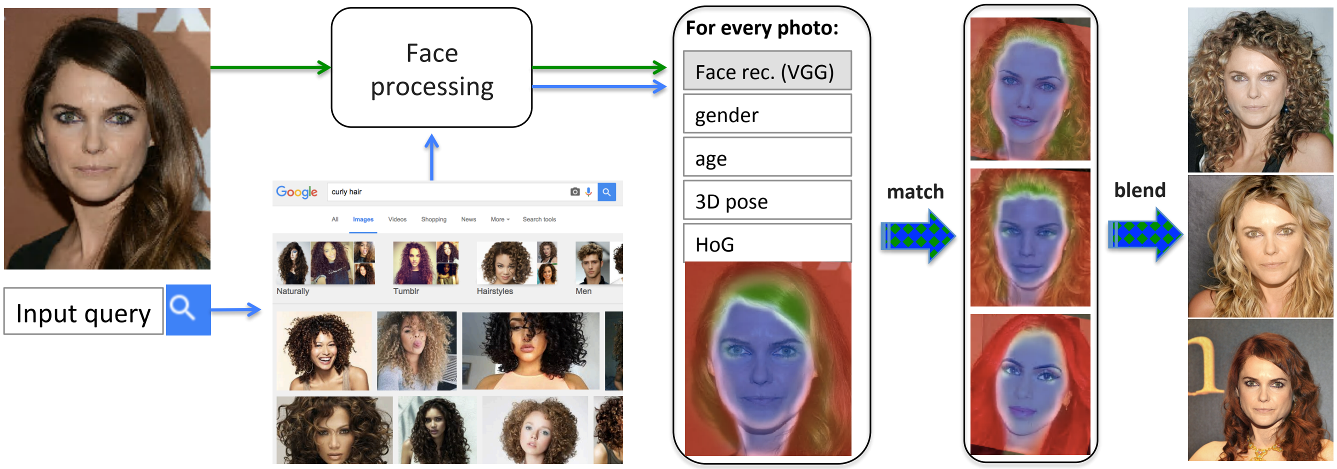

Figure 2: *Illustration of our system. The system gets as input a photo and a text query. The text query is used to search a web image engine. The retrieved photos are processed to compute a variety of face features and skin and hair masks, and ranked based on how well they match to the input photo. Finally, the input face is blended into the highest ranked candidates.*

Our system gets as input one or more photos of a person's face, lets the user search for **any** description they wish, automatically processes the target photo set, and renders the input face in the queried description. Specifically it has three key capabilities:

1. **Free-form query.** The user is allowed to search for anything, and is not restricted to a prescribed set of styles. This opens up the possibilities to explore how they may look like in, e.g., a different country, an era, or even as an astronaut.
2. **Doppelgänger sets.** People's faces vary significantly due to differences in gender, ethnicity, and facial expression. Compositing two arbitrary faces will likely fail due to difference in face shape and silhouette. However, given a large number of images per query that capture many different people and poses, allows us to incorporate face, age, and gender recognition algorithms and find people whose face is most similar to the input person in their face and head shape. We call such similar people, the doppelgänger set. We may then replace the input's hair with hair styles from the doppelgänger set and imagine the input person in a new style.
3. **In the wild.** We target a completely unconstrained and uncalibrated setup that allows to input any photo(s) from a personal photo collection. Naturally, Internet photos include vast variability in imaging conditions (lightings, poses, cameras) which we leverage in the matching and compositing stage for best candidates selection.

The key technical components of our system are face processing (face and fiducial points detection, and alignment), attributes recognition (age and gender), skin and hair mask computation, doppelgänger set estimation via face and attributes recognition as well as matching of imaging conditions, and blending using skin and hair masks. The full system (illustrated in Fig. 2) is completely automatic, and designed with simplicity in mind.

Furthermore, to allow age-related queries, and specifically ones that require to predict how a child may look like in the future, we use a facial age progression method [Kemelmacher-Shlizerman et al. 2014] to modify the face, and then treat it as an input to our matching system. By this we are able to predict a full head aging rather than just the face. The aging process is also automatic.

The novelty of our system is not in a particular technical component (the individual components are drawn from a broad literature) but in the application and their unique combination. We demonstrate that the system enables people to imagine themselves in an unlimited number of styles with a click of a button.

2 Related Work

Modifying appearances can involve complex physical changes, e.g., wearing wigs, coloring hair, putting on makeup, hats, or growing beards. Commercial image manipulation software such as Adobe Photoshop provides variety of tools for replacement, blending, and adjustment of image parts. Such tools require skilled manual processing of each individual photo. In this section, we describe related work in automatic synthesis of people's appearances and styles.

More than a decade ago, the Interactive PhotoMontage paper [Agarwala et al. 2004] showed that it's possible to combine a sequence of photographs using graph-cut and gradient domain blending techniques to create a realistically looking single composite. Although requiring manual user interaction and sequences of photos captured with similar imaging setup, their method produced exciting results. Our system is completely automatic (no user interaction) since it requires processing of a much larger number of photos.

Leveraging large face datasets for graphics applications was shown to be useful by [Bitouk et al. 2008]. They created a large dataset of faces, computed attributes for each face and used the data for the de-identification task. Any input photo in which a person would like to hide their identity would be matched against the database, the highest ranked database face would be then used to replace the face of the input person. In this paper, we consider the reverse task: create many different appearances but preserve the identity of the input person. This requires accurate face and hair segmentation and recognition, e.g., changing the chin of the person is allowed in the de-identification task, however will change the identity in our recasting task.

Face reenactment approaches are related since they blend a source person's expression into the target [Garrido et al. 2014; Thies et al. 2015]. They focus on video tracking and aim to preserve the identity of the target person, while we are interested in preserving the identity of the source. They also do not perform hair replacement and style exploration.

[Nguyen et al. 2008] showed that by analyzing photos of people with beard and without beard, they can create a PCA space of beard and non beard photos and erase beards from new input photographs. [Kemelmacher-Shlizerman et al. 2014] leveraged the Internet for prediction how a child may look like in older age. They focused on the face area only. In this paper, we incorporated that method with our head matching and replacement approach to show full head predictions.

Related to our doppelganger set selection is the recent work of [Crowley et al. 2015]. They applied the VGG face recongition model of [Parkhi et al. 2015] for the task of matching people to paintings. The model proved to be robust across various painting and photographic styles. We therefore use it in our matching process. Additionally, we extend the face similarity metric used in [Kemelmacher-Shlizerman et al. 2011] and [Berthouzoz et al. 2012].

[HaCohen et al. 2011] explored how to use a personal photo collection to enhance colors and quality of a photograph. Their idea is to find corresponding parts within the photo collection and transfer colors across the corresponding parts. More generally, to enable style transfer across photographs: [Shih et al. 2014] transfered photographic style learned from a set of photos to a new input, [Liu et al. 2014] transfered style from a collection of photos returned by a web search, and a number of deep learning methods, e.g., [Gatys et al. 2015] learn a style of a particular artist, e.g., Van Gogh, and apply it on new photographs. Those projects share the goal of adjusting the style of a photo. This paper is about modifying the content and style.

Finally, [Hu et al. 2015] leveraged a large dataset of 3D hair models to generate new hair styles in 3D. Our application is similar in a sense of creating hair styles from large data, and different in the data itself: we target 2D photo-realistic appearances, and allow any style that can be learned from images on the Internet.

3 The Method

The input to our system is 1) one or more photos of a person's face, let's denote it by "source set", and 2) a text query. The query is used to search an image search engine, let's denote the set of retrieved photos by the engine a "target set". The two sets of photos (source and target) are processed in parallel, and matched to compute a target doppelgänger set per source photo. Next, each source photo and its corresponding doppelgänger set are composited to take the face from source and head from each of the targets. The results are ranked and top outputs are provided back to the user. Below we describe those steps in detail.

3.1 Face processing

Each photo (target or source) is processed independently. First, the face and facial landmarks are detected. We use the face detection algorithm of [Mathias et al. 2014] since it is robust to large out of plane rotations. Given the box around the face, we run the IntraFace landmark detector [Xiong and De la Torre 2013] to estimate corners of eyes, nose, and mouth. We follow the frontalization pipeline of [Kemelmacher-Shlizerman et al. 2011] to align and warp each photo to a frontal pose using a template 3D model. Via this process we also estimate the 3D pose (roll, pitch, and yaw angles) for each photo.

Next, we estimate a number of features per photo: age, gender, HoG, and face recognition features VGG. For age and gender estimation we follow the method reported in [Levi and Hassner 2015] which uses a deep network model trained on the Adience dataset. We augment that dataset with photos from [Kemelmacher-Shlizerman et al. 2014] and create classifiers for age and gender using Caffe [Jia et al. 2014]. For comparison of facial expressions, we compute Histograms of Gradients [Dalal and Triggs 2005] per photo and estimate them on the warped photos. Finally, we compute face recognition features using the leading face identification method by [Parkhi et al. 2015] that uses the VGG 16 layer model to train a face recognition model. We apply it on the detected and aligned photos.

3.2 Doppelgänger set estimation

Given the computed features per photo we match each source photo to each target photo. The source set may include one or more photos, the target set typically includes about 1000 photos. These photos are taken in the wild with various lightings, poses, and capturing different people. Our final goal is to create a composite using source person's face and target head, with the goal of preserving the identity of the source person. We found that photos of people with very similar face shapes are more amenable to effective composition (Fig. 4, the forehead and face sides silhouettes are very similar). We, therefore, rank the target photos based on their similarity to the input person, and estimate the source's doppelgänger set. The similarity is computed as the L2 norm between the VGG's face recognition features. This is done for each source photo and in the end (source,target) pairs are ordered based on their rank.

While face recognition features find very similar faces, they are relatively invariant to pose, age, and facial expression. Blending two faces with a different pose produces artifacts, especially when the pose differences are significant and a full 3D model (which is not available) is required to correct for pose. The selected ranked pairs of (source,target) photos are then re-ranked based on the following similarity function:

$$D(s,t) = ||P_s - P_t||^2 + ||\text{Age}_s - \text{Age}_t||^2 + \chi^2(H_s, H_t) \quad (1)$$

where P_i is a 3-vector that includes the roll, yaw, and pitch angles, Age_i is the estimated age of target, and source, and H_i are the HoG features.

3.3 Synthesis

To create the appearance change we compose the face of the source to heads of its target doppelgänger set. The composition involves two steps:

1. Skin and hair masking to blend the source face into the target head.
2. Transforming the face to fit the query, e.g., in case of searching for an older age the source face is age progressed to the target age.

Prior work on face transformations and modifications in "in the wild" photos assumed fixed mask that separated the face from the head, e.g., [Bitouk et al. 2008; Kemelmacher-Shlizerman et al. 2014]. Recently, [Liu et al. 2015] trained a deep network for skin and hair segmentation by labeling photos in the LFW dataset [Huang et al. 2008]. We modify their method to fit out aligned photo collections and estimate masks for source and target photos. Our final filtering of source and target pairs is based on comparison of masks. We calculate an edge map from each mask and compute the L2 norm between the two masks. Pairs with smaller difference are ranked higher. Fig. 3 illustrates the masks.

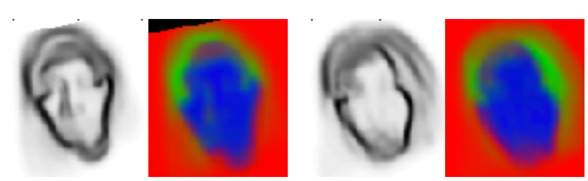

Figure 3: *Example edge maps of a corresponding pair. Original photos can be found at https://goo.gl/BdsEmT and http://favimages.com/wp-content/uploads/2013/01/singer-joan-baez-photos-folk-music.jpg)*

In case the query included an older age, we incorporated the method of [Kemelmacher-Shlizerman et al. 2014] to age progress the face.

Figure 4: *Comparison of with and without face recognition matching (Doppelganger vs. Non-Doppelganger), and with fixed (e.g., Bitouk et al.), or varying mask (our method). Non-Doppelganger photos do not blend well with the input, and fixed mask may change the identity of the person. Both are undesirable in our application. More example can be found at http://grail.cs.washington.edu/tportraits/. Original non-doppelganger and doppelganger pairs: row 1 https://goo.gl/rcjO2s and http://goo.gl/M4bAps, row 2: http://www.departments.bucknell.edu/WRC/history/1926to1930/1926kennedy.gif and https://s-media-cache-ak0.pinimg.com/236x/25/1a/ee/251aee3f2a8be9a310f3b430aa8d22cc.jpg*

Input:	MM	MM	KR	KR	KR	IK	IK	IK	IK	GC	CB
Query:	curly man	sikh	curly hair	1930	black hair	1930	india	curly	black hair	curly man	BD
#images retrieved:	991	147	1512	429	727	429	883	1512	727	991	736
#images post-filters:	84	23	300	56	87	89	42	76	49	108	36
#great:	56	2	275	30	64	30	24	30	28	67	17
#slight artifact:	23	4	22	20	18	36	8	32	11	21	6
#significant artifact:	5	17	3	6	5	23	10	14	10	20	13

Table 1: *Rows 3-4: number of images retrieved from the web and filtered out by the algorithm per input-query combination. Rows 5-7: number of high quality outputs, number of results with artifacts, and number of unacceptable composites. Abbreviations in rows 1-2: MM=Matthew McConaughey, KR=Keri Russell, IK=Ira Kemelmacher, GC=George Clooney, CB=Cate Blanchett, BD=Bob Dylan.*

Specifically, we identified that age progression is required if one of the words is ”age” or ”year old”. We estimated the difference between the searched age and estimated input age, and if the difference is more than 1 year we ran age progression. The query was used to identify the target age space and the input's age was used as query for data that created the source space. Both sets were processed and illumination spaces were created as in [Kemelmacher-Shlizerman et al. 2014].

Once face was transformed and masks were estimated we blend the two photos given the two masks, similarly to [Levin et al. 2004]. Prior to the final blending the source photo is aligned using the fiducial points to the target photo. To ensure preservation of identity, the blending mask is defined as the source mask (Fig. 3). Fig. 4 shows examples of how blending may go wrong in each step of the method, as well as, compares our results to [Bitouk et al. 2008].

4 Results

Our system is implemented as a web interface (Fig. 6), and is easy to experiment with. The system runs on a dual six-core Intel X5680 (3.33GHz) linux machine (Fedora 22) with NVIDIA GeForce GTX 750ti, and 24GB RAM. The code is written in MATLAB and uses Caffe [Jia et al. 2014]. It currently takes 1 minute to get the results back per single input and text query, we plan to make it more efficient in the future and make the system widely available.

Figure 1 presents typical outputs of our system on a single input photo (left) with queries ”curly hair”, ”1930”, and ”india”. Figure 5 shows results of applying the method on five input photos. Adding more input photos is easy and allows to create composites with non-frontal faces. The method automatically ranks any provided (source,target) pair. Figure 7 shows photos of Keri Russell transformed to ”1930”, ”curly hair”, and ”black hair”. The system processed a 100 photos of Keri Russell (the top photos that were retrieved in an image search engine) and matched each photo to the target set of ”curly”, ”1930”, and ”black hair”. The results presented in the figure use six different inputs (can identify based on facial expression and makeup). It is interesting to compare the two different people (Figs. 5 and 7) and their ”curly” synthesized photos. Although the query is the same the resulted top ranked outputs are different. This is due to different face shape of the input people. All the presented photos are automatically generated and in each case we show the top ranked results. Figure 8 demonstrates hair style change for George Clooney.

Figure 5: *Same person synthesized in "india woman" (top), "black hair woman" (bottom left), and "curly hair woman" (bottom right). Total of five input photos was used. The composites in the bottom row use a single photo for all results. The top row ones use four different photos.*

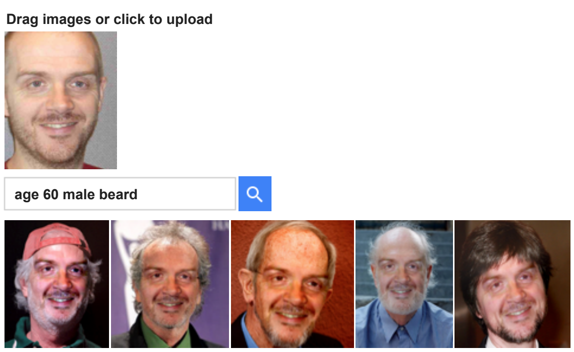

Figure 6: *Screen shot of our web interface. A user inputs one of more photos, searches for a text query and gets variety of renderings of the input person in the queried appearance.*

Figure 9 shows a full head prediction (top ranked result) of a 1 year old future appearance at age 5, 10, 15, 30, 40, and 60. The outputs were produced by age progressing the face and searching for the most similar person in the target sets. Target sets were retrieved by searching for, e.g., "age 5 boy". While age progressing the face part is a big leap towards automatic generation of photos, for important tasks like the missing children search, creating a full head model is essential since the ability of humans to recognize a person is highly correlated with hair style [Kumar et al. 2009]. Guaranteeing that this is indeed how the baby hair will look like in older ages is beyond the scope of this paper. However, our system is the first to allow head replacement combined with age progression and automatically searches hundreds of queries (currently missing children foundations manually match heads from several dozen templates). One way to constrain the space is by incorporating photos of parents, which will be straight forward in our system. Additionally, having the ability to explore hair styles that are not genetically probable is still of great benefit, e.g., children trafficking often involves coloring the child's hair.

Figure 4 compares our results to [Bitouk et al. 2008]. The two key differences between our methods is that 1) we incorporate face recognition while their method focuses mostly on pose and lighting matching, and 2) we use person-specific mask for skin and hair while they (and other previous works) use a fixed mask. We show that using a fixed mask typically changes the identity of the person (note how the chin of the two results changed) which is a positive effect for de-identification task, however cannot be used for style transfer. We also show that by matching similar people we can achieve higher accuracy results.

Figure 10 shows example results of Cate Blanchett in the role of Bob Dylan. (a)-(h) are automatically generated by inputing photos of Cate Blanchett (first 100 that come up in image search) and query "Bob Dylan". (i) shows the actual Cate Blanchett physically transformed to Bob Dylan to play the role in the movie "I'm Not There". We envision that actors or directors may use our system to easily predict how a person may look like playing someone else, or in different era and style as in Fig. 7 (bottom row).

In Table 1 we quantify the quality of the results. We present numbers of images retrieved and filtered out by the algorithm, and the quality of the resulted composites for query/input combinations that appear in the paper. We have also generated many other results from many different queries and inputs and found that on average 75% are same quality as presented in the results' Figures in this paper, 25% are same quality as in Fig 11. I.e., the results presented in the paper are typical. In the next section we discuss limitations and possible solutions.

5 Discussion

We have presented a system for synthesizing possible appearances of a person from a single or more photographs. We called this process transfiguring appearance. One of our main goals was to build a system that will be extremely intuitive to use. The only thing we require from the user is to enter a text query just like they'd do in a search engine. Another goal was not to restrict the user to a particular set of styles, instead we encourage creativity and let people explore any queries they wish and automatically process all retrieved photos.

While some of our results look near flawless, some exhibit artifacts which we discuss below. For the application of style exploration, however, we focused on providing the user with a very broad range of good results, rather than a narrow range of perfect results. Our idea is that the user will benefit from seeing many different styles that came from many different photos (even if they have slight imperfections). E.g., if user is searching for "blond hair" the system

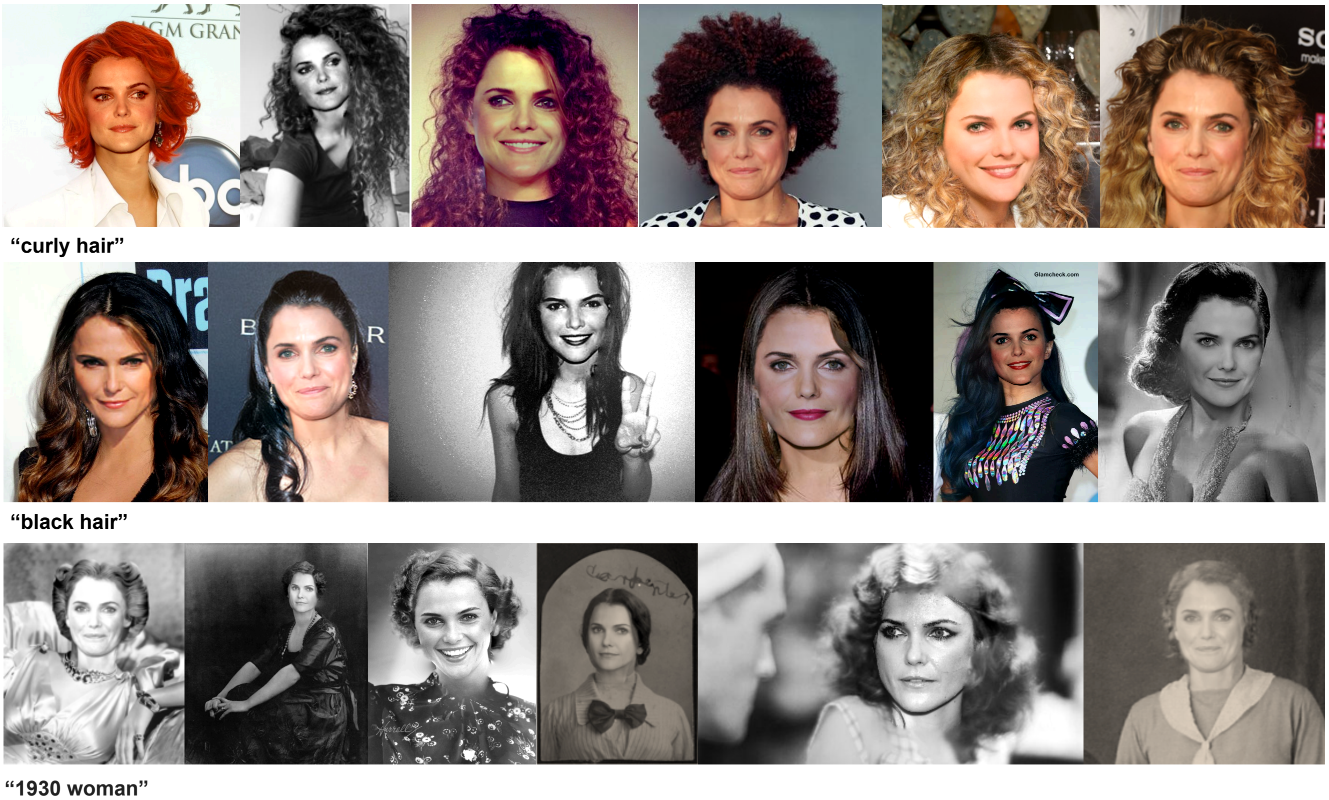

Figure 7: *Automatically generated photos of Keri Russell with various queries. The presented photos are synthetic. See source photos at https://goo.gl/TYTpfJ, and source and target photos on the paper website http://grail.cs.washington.edu/tportraits/.*

Figure 8: *Automatically generated photos of George Clooney with "curly male hair". See the target photos used in these examples on website http://grail.cs.washington.edu/tportraits/.*

will show curly blond, straight blond, blond hair with different outfits, and many more. In future work, it would be great to solve the remaining artifacts and below we discuss possible solutions.

Limitations Currently, the two images (source and selected target) are adjusted using histogram equalization prior to blending which may create color leaks and artifacts. To improve color, contrast and other image adjustments during the blending stage, we can potentially combine recent style transfer methods in our system. Specifically, it will be exciting to incorporate the deep learning based neural art method of [Gatys et al. 2015], transferring photographic styles for portraits by [Shih et al. 2014], and harmonization method by [Sunkavalli et al. 2010] to our appearance exploration application.

The automatic mask estimation algorithm currently provides hair and skin regions. Typical failure is when background (or clothing) is considered part of hair. It would be interesting to retrain the method and add labels for clothing, beards, and glasses. It will be straight forward to incorporate the extended masks in our system. Similarly segmentation of neck and hands would be useful, e.g., using the semantic segmentation of [Zheng et al. 2015].

Figure 11 illustrates typical limitations of the current system.

Currently the results are evaluated visually. While this is enough for some applications, e.g., an actor may judge if they like the appearance or not, for other applications it would be exciting to train

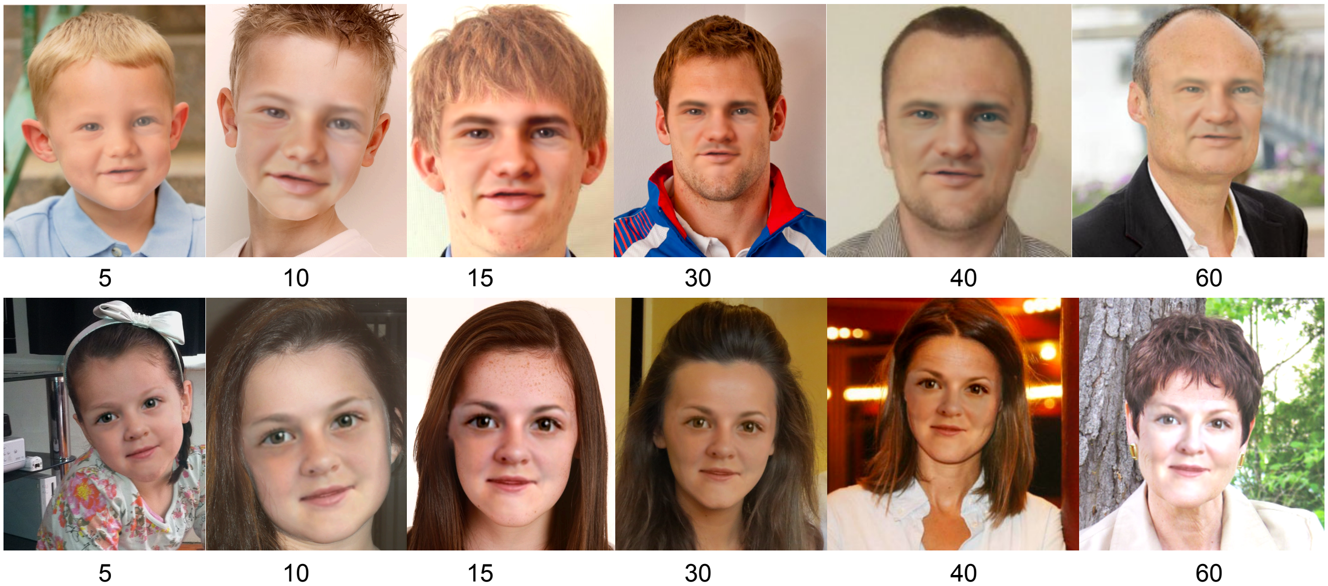

Figure 9: *Two example aging sequence produced by our system. The input is a single photo of a child and the output is a sequence of full head synthesis in older ages. Inputs are 1 year old boy (top) and 4 year old girl (bottom). The query for each age is, e.g., "age 5 boy" for top example, and "age 5 girl" for bottom. Ages are specified below each synthesized result.*

Figure 10: *Actors often change their appearance to fit a new role. Our system could help predict how they may look like. We present predicted appearances of Cate Blanchett acting Bob Dylan created by our system by searching for "Bob Dylan". The actual photo of Cate Blanchett acting Bob Dylan in "I'm Not There" can be found at http://cms.positively-bobdylan.com/wp-content/uploads/2007/08/ blanchett-imnotthere2.jpg*

a classifier that will rank the realism of the results. We have experimented with the pre-trained realism model of [Zhu et al. 2015] and concluded that the model needs to be retrained on face photos, rather than places. Since our system can produce a large number of possible appearances it is easy to create large number of positive and negative examples adequate for deep network training ([Zhu et al. 2015] used 22K photos for training).

Content creation for person identification research. An open problem in person recognition research, is whether a person can be identified across ages. Even more challenging is to enable disguise-invariant recognition, e.g., if the person changes hair style and color, will they still be identified as same or not? One of the biggest issues is lack of appropriate data to enable such research. Our system may be used to easily provide large quantities of data.

Additional examples. More examples and all source and target photos can be found at the paper website http://grail.cs.washington.edu/tportraits/.

Acknowledgements

Input photo in Figure 1 is from personal photo collection of the author. Input photo in Figure 2 is used with permission by Getty Images. Figure 4: row 1 left is Creative Commons https://commons.wikimedia.org/wiki/File:Keri_Russell_2010_2.jpg, row 2 left used with permission from Getty Images. Figure 6: photo used with permission from Bernard Deconinck.

References

AGARWALA, A., DONTCHEVA, M., AGRAWALA, M., DRUCKER, S., COLBURN, A., CURLESS, B., SALESIN, D., AND COHEN, M. 2004. Interactive digital photomontage. *ACM Transactions on Graphics (TOG) 23*, 3, 294–302.

BERTHOUZOZ, F., LI, W., AND AGRAWALA, M. 2012. Tools for placing cuts and transitions in interview video. *ACM Transactions on Graphics (TOG) 31*, 4, 67.

BITOUK, D., KUMAR, N., DHILLON, S., BELHUMEUR, P., AND NAYAR, S. K. 2008. Face swapping: automatically replacing

(a) (b) (c) (d) (e)

Figure 11: *Typical current limitations of our system: (a) beard was not captured by the mask in this case, (b) turban is overlaid using the face due to use of source mask, (c)-(d) contrast of the images doesn't fit, and clothing is not part of the mask. Red arrows point to the issues. (e) uses the same target as (d) but with a different source photo and since source in (e) has a beard, it blends better (blue arrow).*

faces in photographs. *ACM Transactions on Graphics (TOG) 27*, 3, 39.

CROWLEY, E. J., PARKHI, O. M., AND ZISSERMAN, A. 2015. Face painting: querying art with photos. In *British Machine Vision Conference*.

DALAL, N., AND TRIGGS, B. 2005. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, IEEE, 886–893.

GARRIDO, P., VALGAERTS, L., REHMSEN, O., THORMAEHLEN, T., PEREZ, P., AND THEOBALT, C. 2014. Automatic face reenactment. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, IEEE, 4217–4224.

GATYS, L. A., ECKER, A. S., AND BETHGE, M. 2015. A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576*.

HACOHEN, Y., SHECHTMAN, E., GOLDMAN, D. B., AND LISCHINSKI, D. 2011. Non-rigid dense correspondence with applications for image enhancement. *ACM transactions on graphics (TOG) 30*, 4, 70.

HAYS, J., AND EFROS, A. A. 2007. Scene completion using millions of photographs. *ACM Transactions on Graphics (TOG) 26*, 3, 4.

HU, L., MA, C., LUO, L., AND LI, H. 2015. Single-view hair modeling using a hairstyle database. *ACM Transactions on Graphics (TOG) 34*, 4, 125.

HUANG, G. B., MATTAR, M., BERG, T., AND LEARNED-MILLER, E. 2008. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*.

JIA, Y., SHELHAMER, E., DONAHUE, J., KARAYEV, S., LONG, J., GIRSHICK, R., GUADARRAMA, S., AND DARRELL, T. 2014. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*.

KEMELMACHER-SHLIZERMAN, I., SHECHTMAN, E., GARG, R., AND SEITZ, S. M. 2011. Exploring photobios. In *ACM Transactions on Graphics (TOG)*, vol. 30, ACM, 61.

KEMELMACHER-SHLIZERMAN, I., SUWAJANAKORN, S., AND SEITZ, S. M. 2014. Illumination-aware age progression. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, IEEE, 3334–3341.

KUMAR, N., BERG, A. C., BELHUMEUR, P. N., AND NAYAR, S. K. 2009. Attribute and simile classifiers for face verification. In *ICCV*, IEEE, 365–372.

LEVI, G., AND HASSNER, T. 2015. Age and gender classification using convolutional neural networks.

LEVIN, A., ZOMET, A., PELEG, S., AND WEISS, Y. 2004. Seamless image stitching in the gradient domain. In *Computer Vision-ECCV 2004*. Springer, 377–389.

LIU, Y., COHEN, M., UYTTENDAELE, M., AND RUSINKIEWICZ, S. 2014. Autostyle: automatic style transfer from image collections to users' images. In *Computer Graphics Forum*, vol. 33, Wiley Online Library, 21–31.

LIU, S., YANG, J., HUANG, C., AND YANG, M.-H. 2015. Multi-objective convolutional learning for face labeling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3451–3459.

MATHIAS, M., BENENSON, R., PEDERSOLI, M., AND VAN GOOL, L. 2014. Face detection without bells and whistles. In *Computer Vision–ECCV 2014*. Springer, 720–735.

NGUYEN, M. H., LALONDE, J.-F., EFROS, A. A., AND DE LA TORRE, F. 2008. Image-based shaving. *Robotics Institute*, 141.

PARKHI, O. M., VEDALDI, A., AND ZISSERMAN, A. 2015. Deep face recognition. *Proceedings of the British Machine Vision 1*, 3, 6.

SHIH, Y., PARIS, S., BARNES, C., FREEMAN, W. T., AND DURAND, F. 2014. Style transfer for headshot portraits. *ACM Transactions on Graphics (TOG) 33*, 4, 148.

SUNKAVALLI, K., JOHNSON, M. K., MATUSIK, W., AND PFISTER, H. 2010. Multi-scale image harmonization. In *ACM Transactions on Graphics (TOG)*, vol. 29, ACM, 125.

THIES, J., ZOLLHÖFER, M., NIESSNER, M., VALGAERTS, L., STAMMINGER, M., AND THEOBALT, C. 2015. Real-time expression transfer for facial reenactment. *ACM Transactions on Graphics (TOG) 34*, 6, 183.

XIONG, X., AND DE LA TORRE, F. 2013. Supervised descent method and its applications to face alignment. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, IEEE, 532–539.

ZHENG, S., JAYASUMANA, S., ROMERA-PAREDES, B., VINEET, V., SU, Z., DU, D., HUANG, C., AND TORR, P. 2015. Conditional random fields as recurrent neural networks. *arXiv preprint arXiv:1502.03240*.

ZHU, J.-Y., KRAHENBUHL, P., SHECHTMAN, E., AND EFROS, A. A. 2015. Learning a discriminative model for the perception of realism in composite images. In *Proceedings of the IEEE International Conference on Computer Vision*, 3943–3951.