All Smiles: Automatic Photo Enhancement by Facial Expression Analysis

https://sites.google.com/site/allsmilespaper

Rajvi Shah[1,2]
rajvi.shah@research.iiit.ac.in
[1]CVIT, IIIT Hyderabad, India

Vivek Kwatra[2]
kwatra@google.com
[2]Google Research, Mountain View, CA, USA

ABSTRACT

We propose a framework for automatic enhancement of group photographs by facial expression analysis. We are motivated by the observation that group photographs are seldom perfect. Subjects may have inadvertently closed their eyes, may be looking away, or may not be smiling at that moment. Given a set of photographs of the same group of people, our algorithm uses facial analysis to determine a *goodness score* for each face instance in those photos. This scoring function is based on classifiers for facial expressions such as smiles and eye-closure, trained over a large set of annotated photos. Given these scores, a best composite for the set is synthesized by (a) selecting the photo with the best overall score, and (b) replacing any low-scoring faces in that photo with high-scoring faces of the same person from other photos, using alignment and seamless composition.

Categories and Subject Descriptors

I.4.9 [**Image Processing and Computer Vision**]: Applications;

Keywords

Image Composition, Image Enhancement, Face Enhancement, Facial Analysis

1. INTRODUCTION

A photograph shot to capture a perfect moment can often turn out to be unsatisfactory. Group photographs are especially susceptible to problems. Subjects may have inadvertently closed their eyes, may be looking away, or may have a sullen expression on their faces instead of a pretty smile. Having everyone in the group look just right at the same moment in time can be a challenging task, especially when kids are involved. Modern digital cameras have utilities like *click on smile*, but those features may not work well in a group photograph scenario. Multiple clicks and *burst-mode* images improve the chances of capturing a good photo but do not ensure it.

In this paper, we tackle the challenging problem of *automatically* synthesizing the perfect composite from a given set of group photographs. The key novelty of our approach is that it brings together ideas from two disparate research areas: *image compositing* and *facial analysis*. General image compositing can be ill-defined in the absence of constraints. Therefore, most compositing algorithms either operate under simplifying assumptions or rely on user input. On the other hand, facial analysis research (face detection, recognition, matching *etc.*) is driven towards automation. It is also highly mature, as evident by face detectors which are commonplace in consumer cameras and social networks. In our work, we employ the power of facial analysis to automate image compositing in a specific but prevalent problem setting.

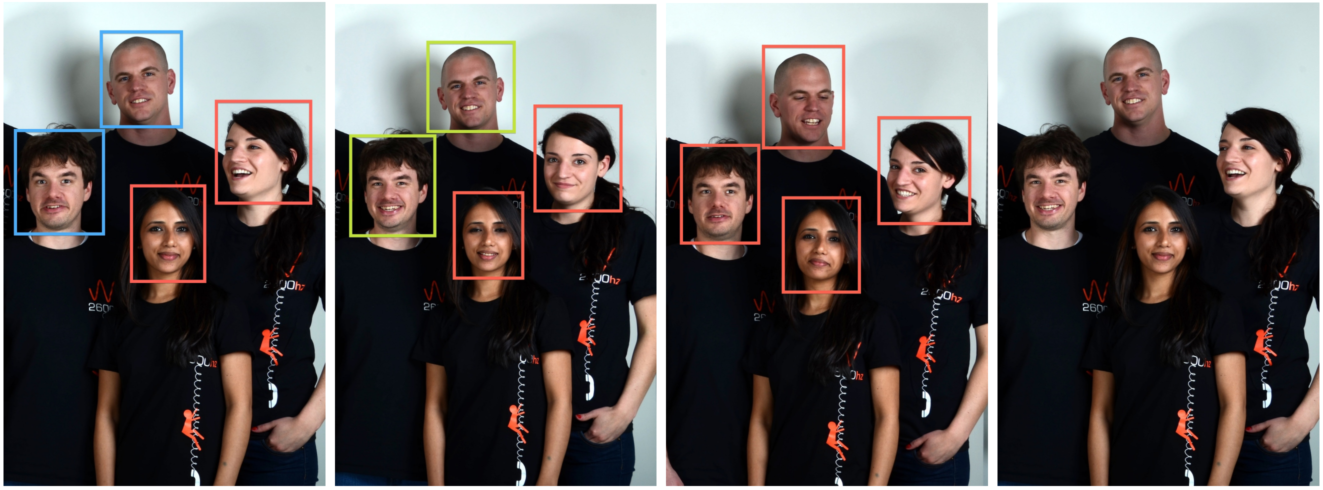
Furthermore, our technical contributions include a scoring function for evaluating the *goodness* of a face, based on classifiers that learn attributes such as *smile/no-smile* and *open/closed eyes*. Figure 1 demonstrates various stages of our pipeline. Firstly, given the set of input photos, our framework detects all faces, and then groups together faces of the same person. Secondly, it assigns a goodness score to each face instance. Finally, it selects a target photo based on the overall scores, and replaces any inferior (low/negative-scoring) faces in the target with superior (high/positive-scoring) faces from other photos.

2. RELATED WORK

Efforts have been made previously for automatic detection of smile [20, 17] and eye-closure [18, 13] events. Albuquerque et al. [2, 3] proposed a framework for selecting an attractive portrait of a person from a video sequence based on smile and eye-closure attributes. Their framework detects eye and mouth regions in a face using the AdaBoost algorithm [19] and trains an SVM classifier over PCA-based features for labeling these regions as good or bad. Our goodness evaluation mechanism shares some attributes with Albuquerque et al. [2]. However, our work differs from theirs in several respects. For example, we use a combination of pyramidal histogram-of-gradient features, rectangular features and color features for describing eye-closure and smile events, which are more detailed and discriminative. Combining these features gives us a relatively high dimensional feature vector. We, therefore, use AdaBoost with decision stumps over these features as weak classifiers, which also serves as a feature selection mechanism.

This distinction is important given that the portrait selection system of [2] is trained and tested mainly on images taken in a semi-controlled environment with a small number of subjects. On the other hand, our system is trained on a large database of real-world face images with small or no overlap in subjects, and performs well on a variety of faces. We show the results of the goodness scoring and composition on real-world albums. Also, [2, 7] demonstrate results only on frames within a video, while our system performs well even when the analyzed images are farther apart in time.

CVMP '12, December 5 – 6, 2012, London, United Kingdom
Copyright 2012 ACM 978-1-4503-1311-7/12/11 ...$15.00.



(a) Input A (selected target)

(b) Input B (source faces)

(c) Input C

(d) Composite Result

Figure 1: Overview of our pipeline. (a) - (c) Input photographs: *A,B,C*. Faces of all subjects are detected, scored and matched across photos. *A* is selected as the target based on overall score. Two faces in *A* (blue boxes) have better scoring counterparts in *B* (green boxes). (d) The composited result obtained by replacing the two target faces in *A* with source faces from *B*.

Recently, Fiss et al. [7] also proposed a novel approach for selecting candid portrait frames from a high-resolution video. The motivation of their work is identifying photo-journalistic style portraits from a continuous video, but not necessarily enhancement of group photographs. They conduct a psychological experiment for portrait selection and demonstrate that the frames selected as candid portraits are highly correlated with the most expressive or communicative frames in the video. They capture this knowledge by computing optical-flow along face landmarks and identifying video frames with apex facial expressions suitable as portrait images.

As mentioned earlier, our approach is not specific to video frames but applies to general still image sets. Hence, we employ image based features to model specific facial expressions. Also, unlike above approaches, our framework is not limited to selection alone. We combine goodness based source selection with composition to create a completely automatic photo enhancement framework.

Kwatra et al. [10] proposed a graph-cut [6] based formulation for seamless composition of multiple images. This formulation poses image composition as a *labeling problem*, where the label of each pixel indicates its source image. Agarwala et al. [1] employed a similar framework for compositing a perfect group photograph from a set of problematic photographs taken in bursts. Their framework allows a user to select desired source regions by marking strokes on input images. Pixels marked by user strokes are constrained to be copied from the source images, while labels (source images) for the unmarked pixels are computed using graph cuts.

Yang et al. [21] introduced a technique called *Expression Flow* for 3*D* aware replacements of face components. This technique allows a user to interactively transfer expressions between two faces of the same person, even when the faces are rotated out-of-plane. Contrarily, in our work, we replace the entire face instead of only touching the problematic components, and restrict the replacements to frontally-oriented faces. While our replacement strategy may not be able to reconstruct all possible expressions of a person, our conservative approach minimizes the possibility of synthesizing distorted or unrealistic faces. Expression Flow computation uses reconstruction of 3D face structure, which can be error-prone and also requires user interaction. By restricting our algorithm to conservative 2D compositing, we have an approach which is completely automatic, highly robust, and simple to implement.

Bitouk et al. [4] proposed an automatic framework for swapping faces of two *different* people from a large database for face de-identification applications. In their framework, a composite face is created by replacing the interior face region (containing facial features) of the target face by that of the source face. To ensure a seamless replacement, a minimum error boundary is computed using dynamic programming. Since, this framework automatically swaps faces of two different people, the source face for swapping is selected based on a the similarity measure which determines quality of the final composite along the replacement boundary. On the contrary, we emphasis on facial feature quality for source selection. They show an example of using their method for swapping faces between the same person for photo enhancement purposes, but it requires manual source selection and is limited to burst-mode photographs. [9, 14] are other examples of consumer applications for compositing group photographs.

In our work, we replace user interaction with a learning based selection mechanism, thereby automating the whole pipeline.

3. FACE DETECTION AND GROUPING

We use off-the-shelf tools for face detection and grouping. Face detection, landmarks and pose identification [16, 19, 22] and recognition [12] are well studied problems and not the contribution of this paper. Hence, we do not go into their internal details. Instead, we assume the availability of following modules, which are derived from, or similar to, the above references.

Detector and Landmarker: This module detects all the faces in a given image and extracts landmark locations in those faces. A total of 12 meaningful landmarks are extracted: two eye centers, nose tip and nose root, four eyebrow corners and four lip corners. These locations are highlighted by green markers in Figure 3 (top-right).

Pose Estimator: This module estimates the face pose in terms of yaw, pitch and roll, based on landmark locations.

Template matcher: This module extracts templates from faces for matching, and computes match scores for face template pairs.

Given the set of input photos, we detect all faces in these photos, and compute landmarks for them, which are used later during goodness evaluation and face replacement. We then compute templates for each face and group them together such that a unique person-identity can be assigned to each group. The template matching is fairly robust to illumination variations, allowing successful face matching under different capture settings. We outline our algorithm for assigning person identities in Algorithm 1. This algorithm builds the person-identities incrementally. At each step, a new face is either assigned to an existing identity, or creates a new identity of its own based on the matching score.

4. GOODNESS EVALUATION

Given a person's face image, we need to automatically decide whether that face is a candidate for replacement. We train boosted classifiers for two common face attributes: *open vs. closed eyes* and *smiling vs. not smiling*. The classifiers are calibrated using cross-validation to return a continuous score between 0 and 1. We combine these scores with face pose information to build a joint scoring function for overall goodness evaluation. In the following sub-sections, we provide a step-by-step explanation our experiments.

4.1 Dataset Collection

To train classifiers for identifying problematic facial attributes, we created a dataset of 14000 face images from the web using *Google Image Search*. We used queries like 'family photo', 'friends outing', 'convocation photos' *etc*. to obtain images corresponding to a group photograph setting. To improve the number of negatives in the data, we also used specific queries like 'not smiling', 'frown', 'closed eyes' *etc*. Our classifiers are trained using frontal faces only. Hence, we discarded profile-view faces from the collected dataset and rotate tilted face images to frontal-view prior to training. The remaining face images were manually annotated by multiple (at least three) operators for the following attributes:

1. Person is smiling or not smiling.
2. Person has eyes open or closed.
3. Photo looks good or should be retaken.

The table in Table 1 shows the number of annotations eventually used for each attribute. The totals are different in each case because we used multiple operators per sample and only kept the samples that received a majority vote for a given attribute. A few example faces from the training set are shown in Figure 2 with good (superior) and not-so-good (inferior) photos.

4.2 Training Sub-Classifiers

To build the goodness classifier, we first train sub-classifiers for detecting smiles and eye-closure. Here, we discuss our choice of features and learning mechanism for these two classification tasks.

Pyramidal Histogram of Oriented Gradients (P-HOG).

	Yes	No
Smiling	8691	1033
Open Eyes	11494	321
Good Photo	7442	760

Table 1: Number of labeled samples for each attribute.

Algorithm 1 Identity Assignment

```
Let $\mathbf{Id}(i, f)$ be identity of face $f$ in image $i$.
Let $T_k = \{t_k^n\}$ be the list of all face templates for person $k$.
Let $S = \{T_k\}$ be the set of all template lists $T_k$, one per person.
BEGIN: S = { }
for each image $i$ in $album$ do
  for each face $f$ in image $i$ do
    $t_{if} = \mathbf{GetTemplate}(i, f)$
    $bestScore = 0,\ matchIndex = -1$
    for $k := 1 \rightarrow size(S)$ do
      for $n := 1 \rightarrow size(T_k)$ do
        $score = \mathbf{GetPairwiseMatchScore}(t_{if}, t_k^n)$
        if $score > bestScore$ then
          $bestScore = score$
          $matchIndex = k$
        end if
      end for
    end for
    if $bestScore \geq matchThreshold$ then
      $\mathbf{Id}(i, f) = p = matchIndex$
      $T_p = T_p \cup t_{if}$
    else {create new person identity}
      $\mathbf{Id}(i, f) = p = size(S) + 1$
      $T_p = \{t_{if}\},\ \ S = S \cup T_p$
    end if
  end for
end for
END
```

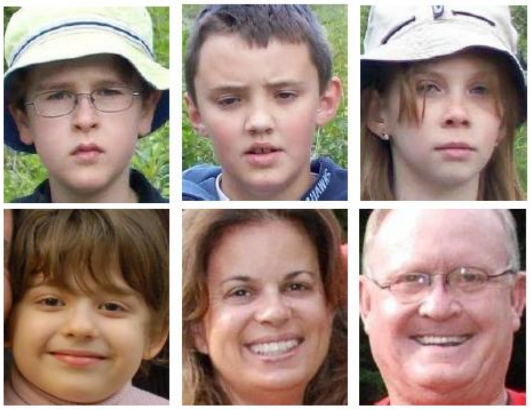


Figure 2: Examples of inferior (top) and superior (bottom) faces.

These features encode local shape and spatial layout of the shape at various scales [5]. Local shape is captured by the histogram of orientation gradients within a spatial window, while spatial layout is captured by gridding the image into regions at multiple resolutions. The final feature vector is a concatenation of orientation histograms for all spatial windows at all grid resolutions, as shown in Figure 3.

Since eye-closure is a localized event, it is sufficient to limit the feature extraction only to the eye region. This region-of-interest is shown by the outer green bounding box in Figure 3 (top-left). On the contrary, even though smile is an action mainly localized to the mouth region, it causes subtle changes in cheek and eye muscles as well. Hence, for smile detection, we extract features from both the entire face as well as the mouth region. These regions are shown in Figure 3 by the blue box and yellow grid, respectively.

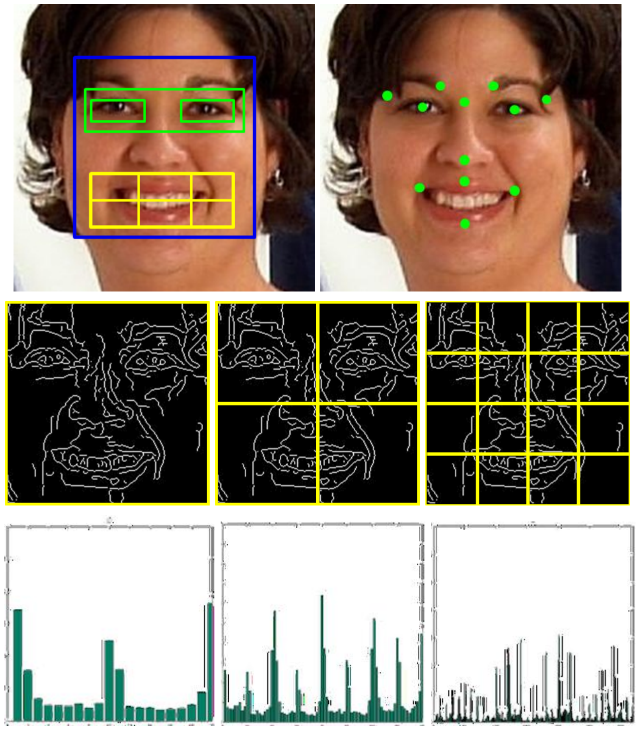


Figure 3: Row 1: Regions of interest (left) and face landmarks (right). Shown at various grid levels (row 2) for the blue bounding box are pyramidal histograms of oriented gradients (row 3).

We quantize orientation angles $\in [0, 180]$ into 20 bins for histogram computation, which gives us a 20 dimensional feature vector for a single spatial window. For eye-closure, we extract features for two pyramid levels leading to a $(20 \times 2^0 \times 2^0) + (20 \times 2^1 \times 2^1) = 100$ dimensional feature vector per color channel. For smile detection, we extract features for three pyramid levels, producing a 420 dimensional feature vector per channel. In our experiments, we found that extracting features from all three R,G,B color channels improves accuracy over using only the luminance channel.

Rectangular Features.

These features [19, 11] encode average intensity difference of adjacent rectangular regions. They are also known as Haar-like features due to their analogy to Haar-wavelets. [11] proposed an extended set of 14 rectangular filters which encode edges, lines and center-surround differences. These filters are shown in Figure 4. For eye-closure classification, these features are extracted from three regions as shown by the green boxes in Figure 3, generating a $14 \times 3 = 42$ dimensional feature vector per channel. For smile classification, these features are extracted from 6 regions as shown by the yellow grid in Figure 3, producing an 84 dimensional feature vector per channel.

Pyramidal Histogram of Color Features.

These features encode spatial color distribution, and are helpful for both smile and eye-closure events, since they have strong color associations caused by the difference in teeth/lips color and iris/skin color respectively. We extract these features only from the regions localized to the eyes and the mouth.

4.3 Overall Classifier and Scoring Function

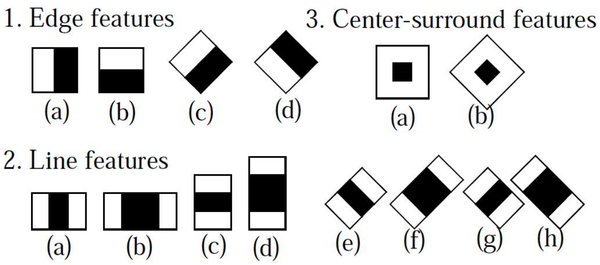


Figure 4: Extended Set of Haar-like Features (Image credit:Lienhart and Maydt [11]).

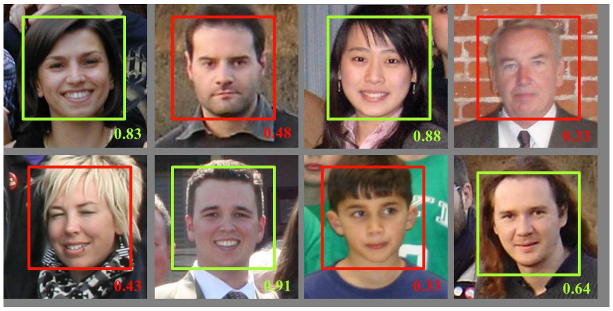


Figure 5: Faces evaluated for overall goodness. Superior faces (score ≥ 0.5) marked green, inferior faces marked red.

We combine all of the three features described above, which results in a relatively high dimensional feature vector. We therefore use the AdaBoost [8] learning algorithm for training our eye-closure and smile detection classifiers. This algorithm uses decision stumps on individual features as weak classifiers, which also results in automatic feature selection. The trained classifiers return a score that can be thresholded for the classification task. However, we need a continuous score to rank the various faces of a person. The raw scores need to be converted into membership probabilities in order to combine various attributes. We carry out this process, known as calibration, by performing logistic regression over the raw scores [15].

The overall goodness score y may be expressed as a function of the individual sub-classifiers scores x_i: $y = f(x) = \sum_i w_i x_i$. We tried learning the weights w_i using the *good photo vs. should be retaken* annotations, resulting in a two-level hyper-classifier. However, we found that simply combining the scores with uniform weights was sufficient. We also take the pose of the face into account in the goodness score. Frontal faces are preferred over profile faces. Instead of learning a classifier in this case, we simply treat faces with both yaw and pitch $< 30^\circ$ as frontal. Non-frontal faces can either be assigned a large penalty, or simply not used as targets for face replacements, even if they contribute to the overall score of the photo. Figure 5 shows the goodness scores for some faces using the overall scoring function.

5. FACE REPLACEMENT

To create a composite from a set of group photos, we first select the best target photo based on the total goodness score across all faces in that photo. Then, each face in the target photo with a low goodness score is replaced by the best scored face of the same person in other photos. Face replacement is a three step process. Firstly, source (superior) and target (inferior) faces are

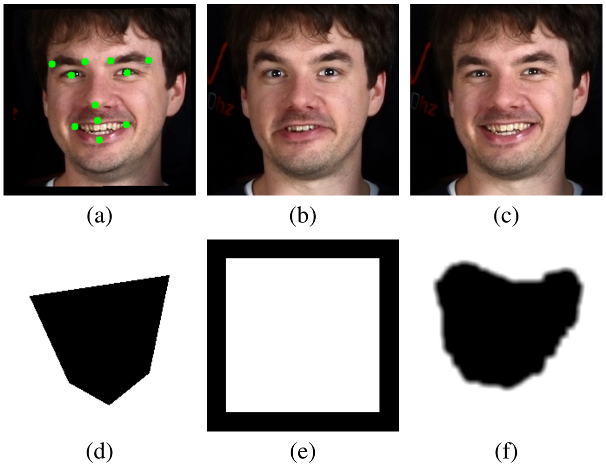


Figure 6: (a) Source face after alignment and color correction w.r.t. target face. Green dots represent face landmarks. (b) Target face. (c) Final composite. (d) Source and (e) target opacity masks supplied to graph cuts for stitching. Black regions in these masks represent constrained pixels, which are directly copied from source or target. Note that the source mask encloses all landmarks shown in (a). (f) Blending mask used to create the final composite. Black region corresponds to source pixels, with soft weighting across the transition boundary.

aligned using feature based registration. Secondly, the source face is color corrected to match the target photo's illumination. Finally, the aligned and color corrected source face is pasted on the target photo in a seamless manner using graph-cut optimization and alpha blending.

Face Alignment.

We use feature based registration to align source and target faces. This alignment requires corresponding points in both source and target images. We use the face landmarks extracted using the Landmarker module as corresponding points between the two face images. To reduce the registration error, we also extract and match Harris corner points in both images. We then fit a parametric transformation to these feature correspondences. This transformation is restricted to a similarity transform (translation, rotation, scale) to avoid distortions, and we solve for its four parameters using least squares. If a similarity transform cannot be computed for a given pair of source and target faces, we do not perform the replacement.

Color Correction.

To compensate for illumination variation between the source and target images, we color correct the source image as:

$$I_s^c \leftarrow I_s^c + \bar{I}_t^c - \bar{I}_s^c,$$

where $\bar{I}^c$ refers to the mean value of color channel c in image I; subscripts s and t correspond to source and target, respectively. Figure 6a and Figure 6b show a pair of source and target images after alignment and color correction.

Seamless Replacement.

If the source face is simply copied to the target image after alignment, it can lead to artifacts along the face boundary. Hence, we use graph-cut optimization [6, 10] for seamless face replacement. During replacement, the inner pixels must come from the (superior) source face, whereas the outer pixels must come from the (inferior) target face image. A convex polygon is fit to contain all the landmark points, and pixels inside this polygon are constrained to come from the source face. The outer border of the face is constrained to come from the target image. Figure 6d and Figure 6e show the opacity masks for corresponding source and target faces. Constrained pixels are copied as-it-is from the respective images. Graph-cut optimization finds the optimal seam passing through the unconstrained pixels by minimizing the total transition cost from source to target pixels. We use the same quadratic formulation for this cost as [10]:

$$C_{pq}(s,t)|_{s \neq t} = |I_s(p) - I_t(p)|^2 + |I_s(q) - I_t(q)|^2,$$

where $C_{pq}(s,t)$ represents the cost of transitioning from the source image at pixel p to the target image at pixel q. Once the optimal seam is computed, we blend the source and target faces along the seam using alpha blending to obtain the final composite:

$$I_c = \alpha \cdot I_s + (1-\alpha) \cdot I_t,$$

where α is obtained by blurring the binary mask corresponding to the source region computed by graph cuts (Figure 6f).

6. EXPERIMENTS AND RESULTS

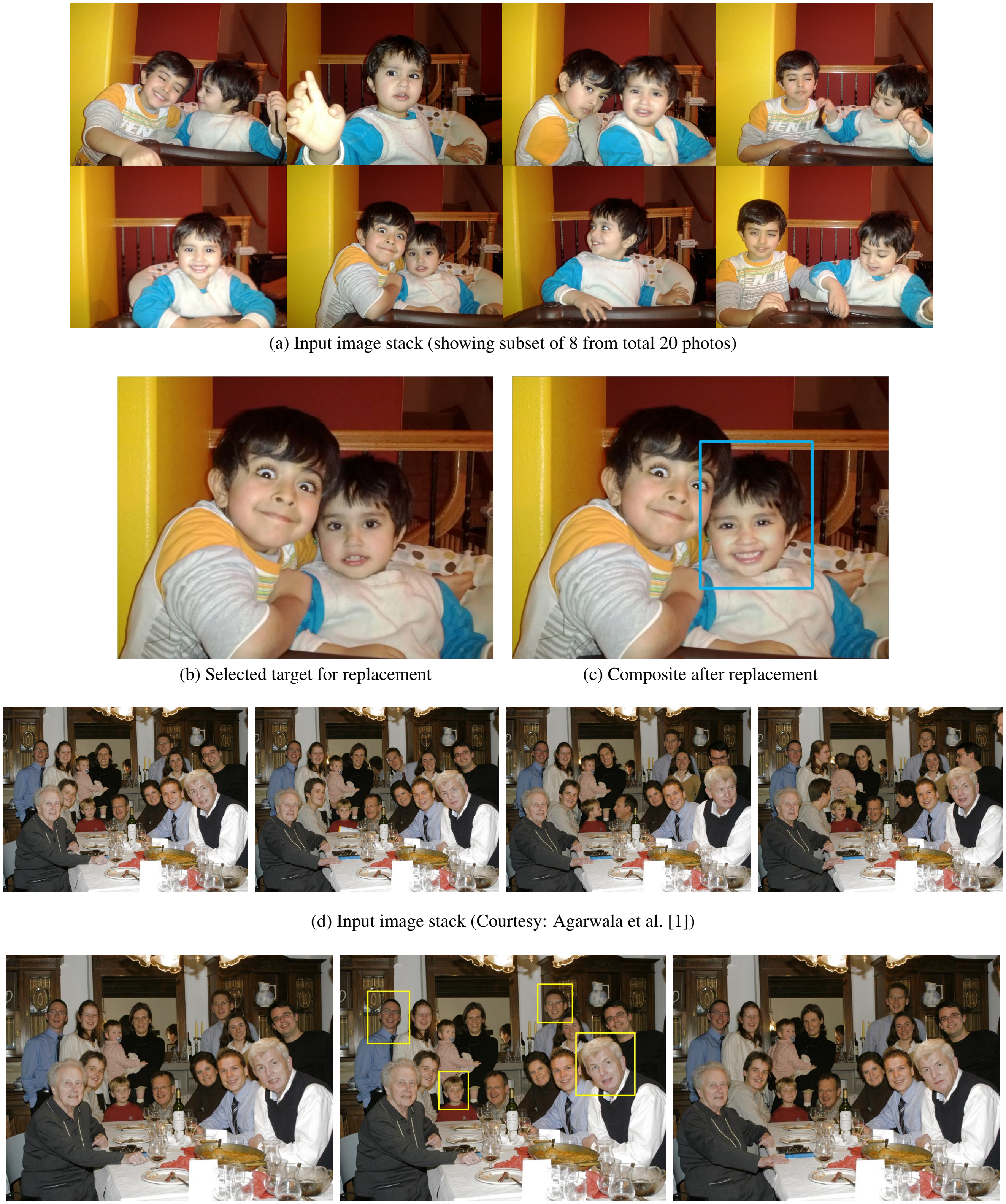
We have applied our technique to photos taken in a variety of settings. In the following results, we highlight face replacements through bounding boxes drawn over the composited result. Figure 7 shows two examples with input stacks, selected targets and composited results. The top example consists of 20 input photos with a variety of facial expressions and poses. Our algorithm was able to select a good target (Figure 7b) containing only frontal faces and with only one face needing replacement. The composition results in both kids having smiles. In the bottom example, we show a comparison with [1]. Their method requires a user to specify regions of interest from different photos via a brushing tool, while we achieve a qualitatively similar result fully automatically.

Figure 8 demonstrates that while our method is great for adding smiles to faces, it also acts as an automatic technique for selecting the best photo from a set. In this example, the selected target (Figure 8c) has the highest score for all faces, *i.e.* no face needs replacement. This can be used as-it-is as a representative of the set. However, the user may, of course, manually choose a different photo as target (Figure 8b) and use our algorithm to create a composite with face replacement (Figure 8d).

Photos taken in *burst mode*, where all photos are taken within a few seconds, form good candidates for our technique. However, we can also robustly handle cases where photos are taken relatively far apart in time. Figure 9 shows a specific example where the source and target photos were clearly not taken one-after-another (child and mom vs. child and dad), but still results in a successful composite.

Figure 11 shows a before (target) and after (composited) result, along with a plot of *smile* vs. *open eyes* scores for the source faces of one subject. The plot shows three faces, only one of which has open eyes and a wide smile, and consequently gets selected as the best face. Figure 12 is an example before and after result created from photos of a large group of people. Several faces get replaced in this example. Figure 10 shows more examples of before and after replacement results.

Runtime Performance: The performance of our method depends on the number and resolution of photos and faces in the input stack. Maximum time is spent in face detection and feature extraction; scoring and replacement is quick. Most of our examples took



(a) Input image stack (showing subset of 8 from total 20 photos)

(b) Selected target for replacement

(c) Composite after replacement

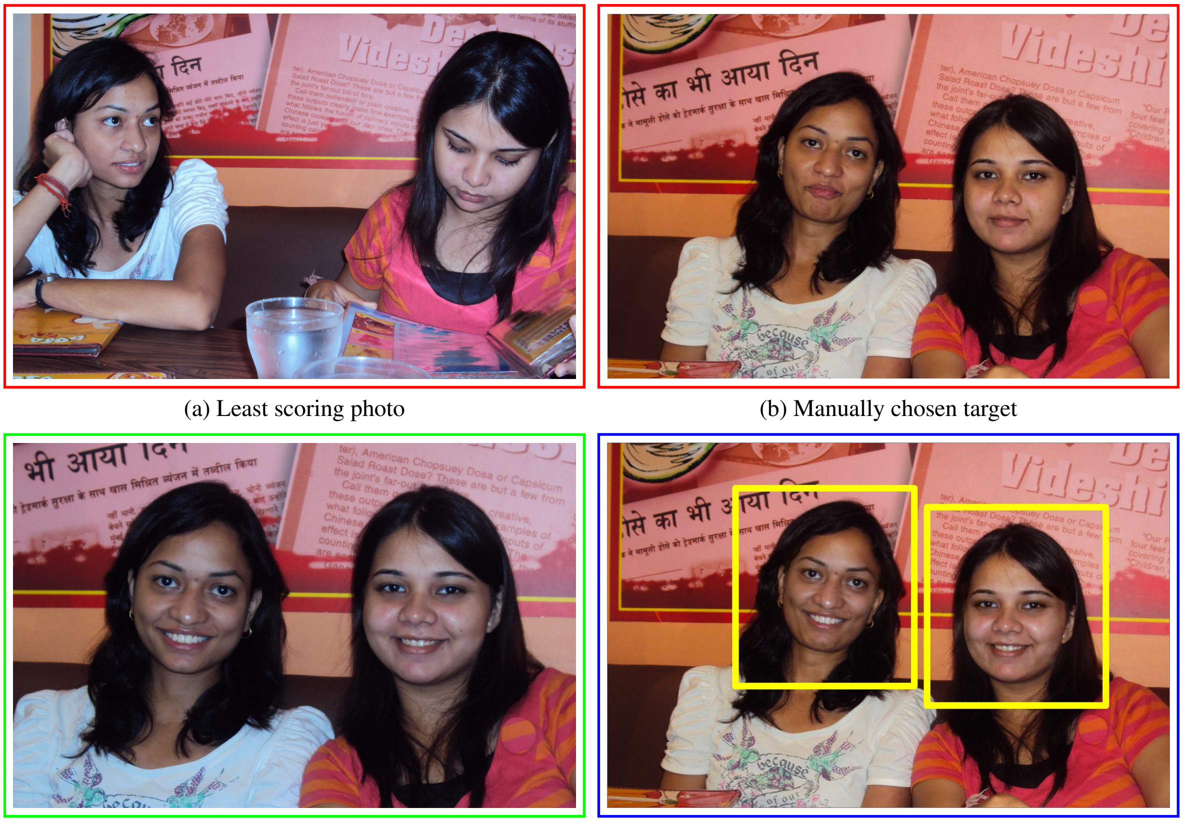
(d) Input image stack (Courtesy: Agarwala et al. [1])

(e) Selected target for replacement

(f) Composite after replacement

(g) Agarwala et al. [1]'s result

Figure 7: (a-c) Our pipeline and composition result for a stack of 20 input photos (only 8 shown), with a variety of facial expressions and poses. (d-f) Result on image stack from Agarwala et al. [1]. (g) Their method requires user interaction, while ours is fully automatic.



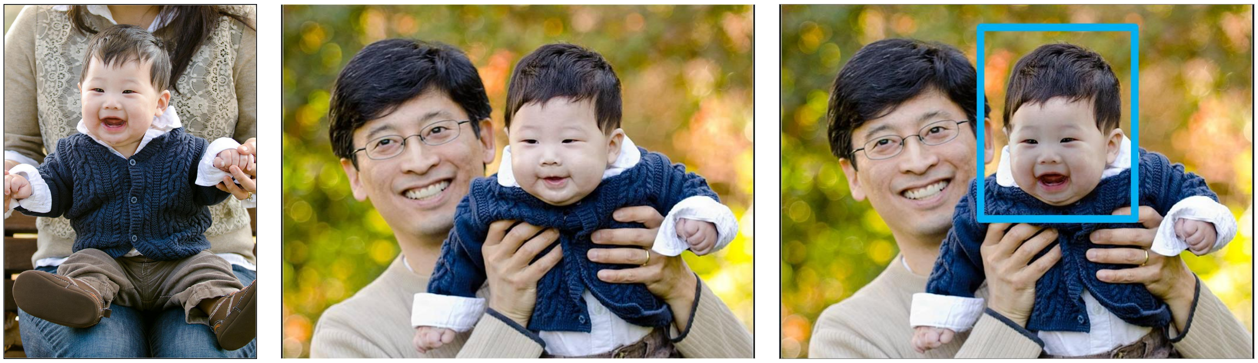
(a) Least scoring photo

(b) Manually chosen target

(c) Automatic best selection

(d) After Composition over (b)

Figure 8: (a-c): Input stack. In (c), our algorithm provides an automatic best photo selection even when no replacement is required. (d) User manually chooses (b) as the target for compositing producing the shown composite.



(a)

(b)

(c)

Figure 9: Example demonstrating robustness of our technique. Source (a) and target (b) photos were taken farther apart in time than a typical "burst". (c) Composition.

between $1 - 15$ seconds on a 3.5GHz, 6-core, 12GB RAM Intel Xeon workstation. The most expensive example, Figure 7 (top) with 20 photos took 60 seconds. We believe that our technique is amenable for interactive applications, and especially suited to processing of photo bursts.

Discussion and Limitations: Our method works well on a variety of cases, and we strive hard to avoid failure cases. A common failure case for face replacement would be when the two faces have different poses. To alleviate that, we discard faces which are not sufficiently frontal or cannot be aligned using a similarity transform. However, if facial appearance changes significantly, then artifacts can still occur. An example is shown in Figure 13, where the hair of the subject are positioned differently in the source and the target, causing artifacts. Another notable issue here is that the source was lower resolution than the target, resulting in loss of resolution. We can also potentially replace a superior face with an

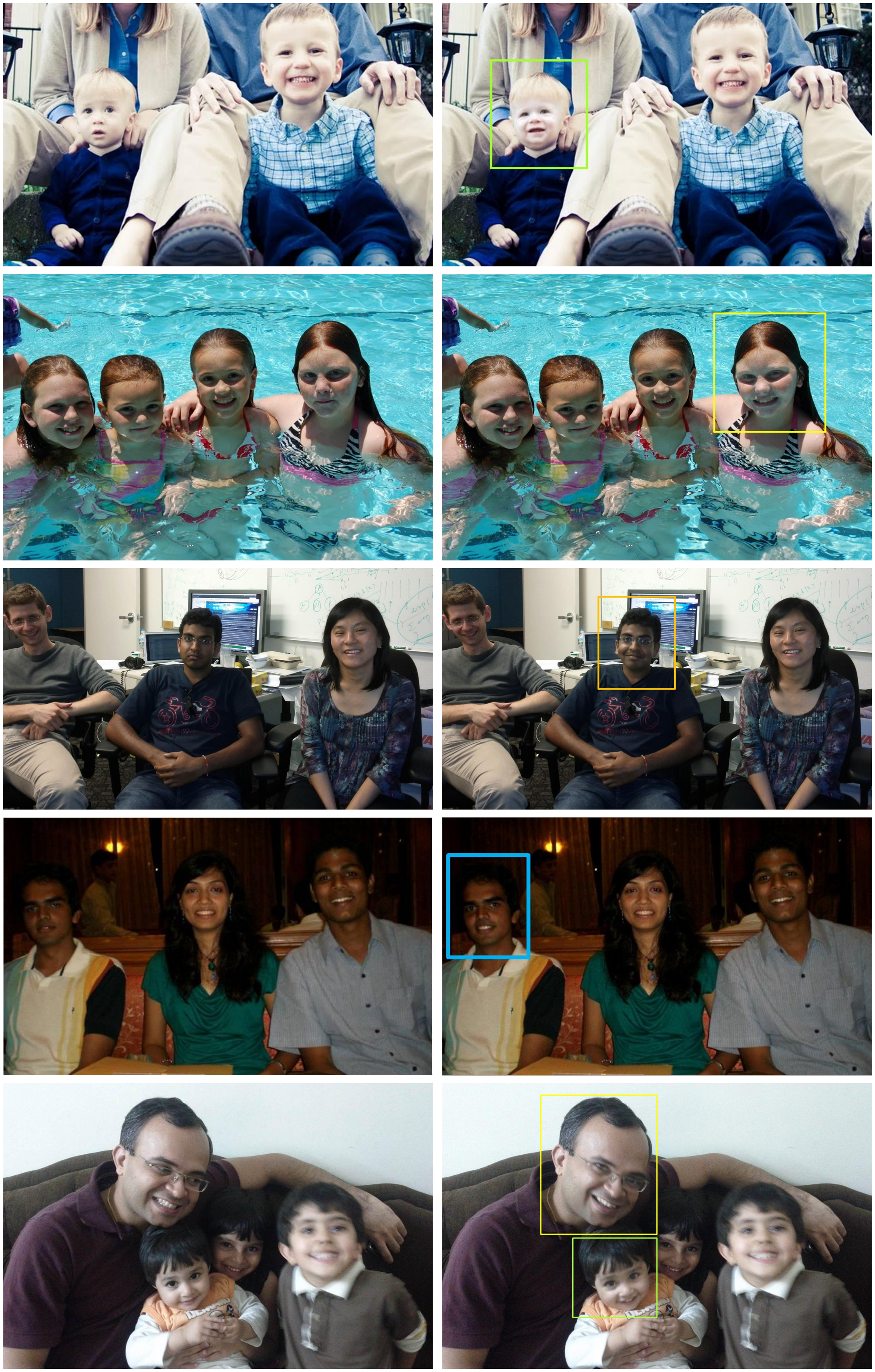


Figure 10: Examples of before (left) and after replacements (right).

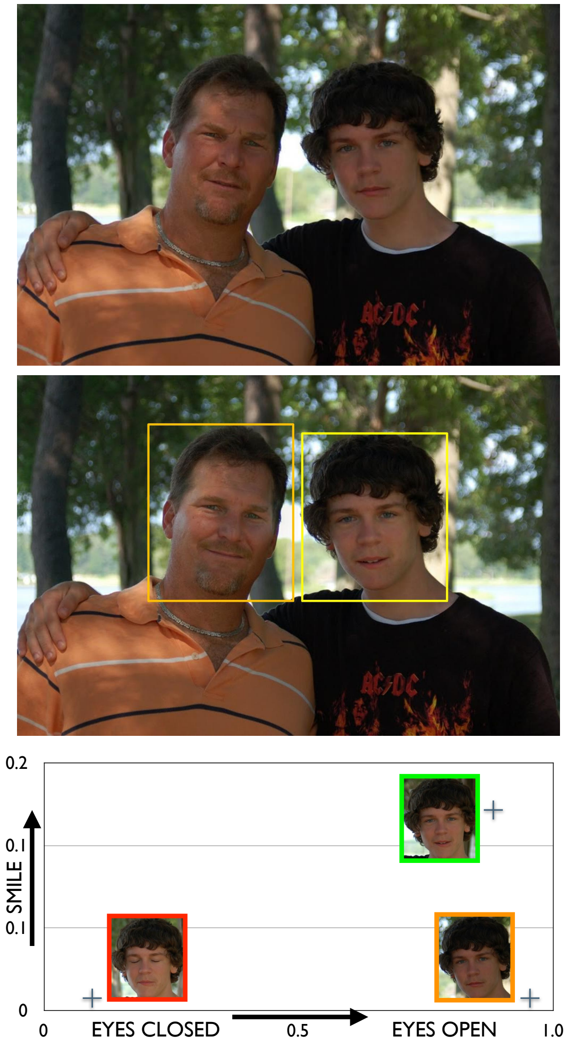


Figure 11: Top: Before and after result. Bottom: Plot showing "open eyes" and "smile" scores for one subject's faces. The face on top-right (green box) scores best and is chosen as source.

inferior face if the classifier ranked faces incorrectly. This usually happens when the faces have similar scores. Our solution here is to only replace a face if the scores differ by more than a threshold.

7. CONCLUSION AND FUTURE WORK

In this paper, we have proposed an automatic framework for enhancement of group photographs using facial expression analysis and image composition techniques. Our framework automatically detects problematic or inferior faces in an image and replaces them with superior faces of the same person from other source images. Face detection and recognition techniques are used to automatically group faces of the same person from multiple photographs. We have employed a large dataset of face images from the web to train robust classifiers for two common facial attributes in photos, smiling vs. not smiling and open vs. closed eyes. We have introduced a novel goodness scoring function, which makes use of these classifier scores as well as face pose information to automate the source face selection procedure in our composition framework. We have demonstrated the effectiveness of our approach through a variety of examples that bring a smile on people's faces. In future, we would like to optimize our implementation for real-time performance, making it an attractive utility for computational cameras. We would also like to learn and incorporate more subtle facial attributes for goodness evaluation.

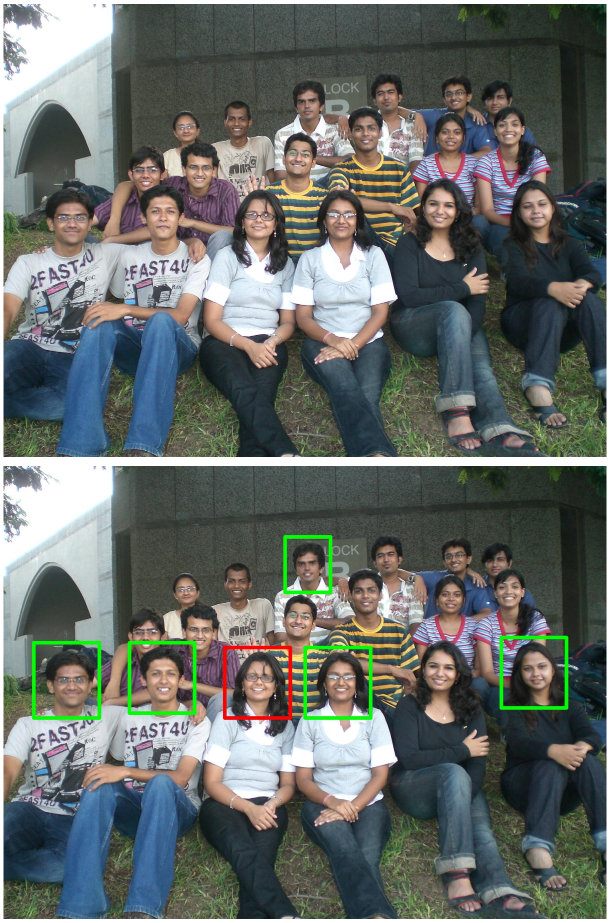


Figure 12: Example demonstrating multiple replacements in a group photograph with many subjects. The replacement in the red box exhibits a limitation (see Figure 13 for details).

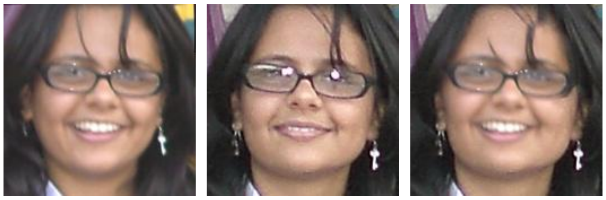


Figure 13: Failure case: Source (left) has hair positioned differently than the target (middle) and is lower resolution, resulting in artifacts and blurring in the composite (right).

References

[1] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen. Interactive digital photomontage. *ACM Trans. Graph.*, 23(3), 2004.

[2] G. Albuquerque, T. Stich, and M. Magnor. Qualitative portrait classification. In H. Lensch, B. Rosenhahn, H. Seidel,

P. Slusallek, and J. Weickert, editors, *Proc. Vision, Modeling and Visualization (VMV) 2007*.

[3] G. Albuquerque, T. Stich, A. Sellent, and M. Magnor. The good, the bad and the ugly: Attractive portraits from video sequences. In *Proc. European Conference on Visual Media Production (CVMP) 2008*, London, UK, 2008.

[4] D. Bitouk, N. Kumar, S. Dhillon, P. Belhumeur, and S. K. Nayar. Face swapping: automatically replacing faces in photographs. *ACM Trans. Graph.*, 27(3), 2008.

[5] A. Bosch, A. Zisserman, and X. Munoz. Representing shape with a spatial pyramid kernel. In *In proceedings of the ACM CIVR 2007*.

[6] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23, November 2001.

[7] J. Fiss, A. Agarwala, and B. Curless. Candid Portrait Selection From Video. *ACM Transactions on Graphics*, 30(6), 2011.

[8] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.*, 55, August 1997.

[9] GroupShot. `http://www.groupshot.com`.

[10] V. Kwatra, A. Schödl, I. Essa, G. Turk, and A. Bobick. Graphcut textures: Image and video synthesis using graph cuts. *ACM Transactions on Graphics, SIGGRAPH 2003*, 22(3), July 2003.

[11] R. Lienhart and J. Maydt. An extended set of haar-like features for rapid object detection. In *IEEE ICIP 2002*, 2002.

[12] C. Liu and H. Wechsler. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Trans. Image Processing*, 11, 2002.

[13] Z. Liu and H. Ai. Automatic eye state recognition and closed-eye photo correction. In *ICPR'08*, 2008.

[14] Photo Fuse. `http://explore.live.com/windows-live-essentials-photo-gallery-get-started`.

[15] J. C. Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In *ADVANCES IN LARGE MARGIN CLASSIFIERS*. MIT Press, 1999.

[16] H. A. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20, January 1998.

[17] C. Shan. An efficient approach to smile detection. In *FG*, 2011.

[18] R. Sun and Z. Ma. Robust and efficient eye location and its state detection. In *Proceedings of the ISICA 2009*. Springer-Verlag.

[19] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, 2001.

[20] J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett, and J. Movellan. Toward practical smile detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31, 2009.

[21] F. Yang, J. Wang, E. Shechtman, L. Bourdev, and D. Metaxas. Expression flow for 3d-aware face component transfer. *ACM Transactions on Graphics*, 30(4), 2011.

[22] C. Zhang and Z. Zhang. A survey of recent advances in face detection. *Microsoft Research Technical Report, MSR-TR-2010-66*, 2010.