AUTOMONTAGE: PHOTO SESSIONS MADE EASY

Nithya Manickam & Sharat Chandran

Dept. of Computer Science, IIT Bombay

ABSTRACT

Nostalgia apart, group photo sessions are tedious; it is difficult to get acceptable expressions from all people at the same time. The larger the group size, the harder it gets. Ironically, we miss many expressions in the scene while the group assembles, or reassembles in the taking of the photographs.

A solution to the problem is using a video of the scene, and automatically extracting an acceptable, possibly stretched, photo montage. In this work, we automate the process. We extract faces, assess the quality, and paste them back appropriately at the correct position to create a pleasing memory.

Index Terms— Automatic Photo Montage, Easy Group Photo Session, Photo Mosaicing

1. INTRODUCTION

Its no more that only professional photographers who take pictures. Almost anyone has a good camera, and often takes a lot of photographs. Group photographs as a photo session in reunions, conferences, weddings, and so on are de rigueur. It is difficult, however, for novice photographers to capture good expressions at the right time, and realize a consolidated acceptable picture.

A video shoot of the same scene ensures that expressions are not missed. Sharing the video, however, may not be the best solution. Besides the obvious bulk in the video, poor expressions (”false positives”) are also willy nilly captured and might prove embarassing. A good compromise is to produce a mosaiced photograph assembling good expressions, and discarding poor ones. This can be achieved by a cumbersome manual editing; in this paper, we provide an automated solution, illustrated in Figure 1. The photo shown has been created from a random youtube video excerpt and the photo shown does not exist in any frame of the original video.

1.1. Technical Contributions

The technical contribution in this work includes:

- A frame analyzer that detects camera panning motion to generate candidate frames
- An expression analyser from detected faces
- A photo patcher that enables seamless placement of faces in group photos

Details of these steps appear in Sec. 2.

Fig. 1. An automatic photo montage created by assembling “good” expressions from a video (randomly picked from Youtube). This photo did not exist in any of the input frames.

1.2. Related Work

Research in measuring the quality of face expressions has appeared elsewhere, in applications such as medical patient expression detection to sense pain [1], measurement of children’s facial expression during problem solving [2], and analysing empathetic interactions in group meeting [3]. Our work focuses on generating an acceptable photo montage from photo-session video and is oriented towards a targeted goal of *discarding* painful expressions, and recognizing memorable expressions.

In regard to photo editing researchers have come up with many interesting applications like organizing photos based on the person present in the photos [4][5][6][7], correcting an image with closed eye to open eye [8], and morphing photos [9]. Many of these methods either require manual intervention, or involves a different and enlarged problem scope resulting in more complicated algorithms rendering them in applicable to our problem.

The closest work to ours is presented in [10], where a user selects the part they want from each photo. These parts are then merged using graph cut to create a final photo. Our work differs from [10] in a few ways. Faces are selected automatically by determine pleasing face expressions (based on an offline machine learning strategy). Video input are allowed enabling a larger corpus of acceptable faces, and the complete process is automated, thereby making it easier for end-users.

2. METHODOLOGY

In this section, we present a high level overview of our method followed by the details. The steps involved in photo montage creation are

1. Base Photo Selection
2. Facial Expression Measurement
3. Montage Creation

The first step in this problem is to identify plausible frames which can be merged to create mosaic. Frames in a video "far away" in time, or unrelated frames cannot me merged. Next, we measure the facial expression of all detected faces, and select a manageable subset. In the last phase, selected faces are substituted in the mosaiced image using the technique of graph cut and blending. Figure 2 illustrates these steps.

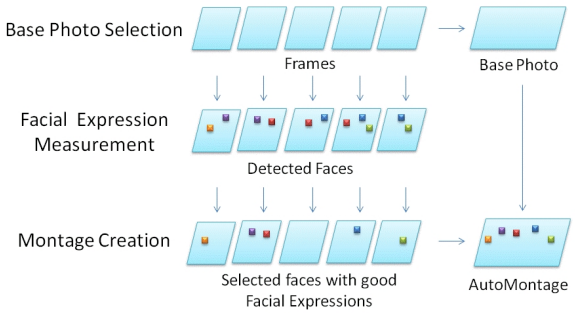

Fig. 2. A schematic of our method. In the first step frames are analyzed to detect a base photo frame which can either be a mosaic of frames, or a single frame from the video. In the next step, detected faces are tracked and grouped together. In the last step, faces with good expression are patched to the base photo to form the required photo montage.

2.1. Base Photo

In our work, faces from video are detected using the method in [11]. We then detect camera movement direction by tracking the position and sizes of the faces. For each shot, we accumulate frames until the camera direction changes. The resulting clusters are used to serve as the base photo. In case there is little or no camera movement, the scene is considered static and the frame having a maximum number of faces is selected as the base photo.

2.2. Facial Expression Measurement

Measuring facial expression is, as expected, critical Facial expression is measured as deviation from neutral expressions as illustrated in Figure 3. We have manually collected around one hundred neutral expression faces from various wedding videos for training our system.

Fig. 3. Facial expression measured as deviation from neutral expression

2.2.1. *Offline Alignment*

Intuitively, alignment is achieved using the position of the eyes as the reference. Neutral faces are aligned using the following steps:

1. Color space conversion: The face is converted from RGB to TSL (Tint, Saturation, and Luminance) color space.
2. Skin regions with values I_s are detected.
3. In the non skin region where $I_s = 0$, regions in the top half of the face are examined for two symmetrical and almost spherical regions which represents the eyes.
4. Non-skin regions in the bottom half of the face are examined for the occurrence of the mouth. When a region is horizontally in between the eyes is found, it is detected as the mouth part. Rectangular regions extracted are measured relative to the positions of eyes and mouth to achieve alignment.

2.2.2. *Neutral expression*

A sparse quantification of neutral expression is achieved using dimensionality reduction techniques. In brief,

1. Faces are contrast enhanced
2. Mean images are computed and subtracted from neutral faces.

3. The actual dimensionality reduction using SVD factorization

For any test face x_t, the facial expression measure is computed as deviation from the stored principal component vectors. Similar to training phase, the face is first aligned, enhanced, and then mean centered vector is projected onto the neutral face vector eigen space to obtain a value p.

The Euclidean distance between the projection p and mean of projection of neutral eigen face P is used to measure the quality of an expression.

$$\delta_1 = p - \frac{1}{N}\sum_{i=1}^{N} P_{ki}$$

We also compute the minimum Euclidean distance between the projection p and each neutral eigenface from P.

$$\delta_2 = \min_{i=1}^{N} p - P_i$$

The facial expression measure is computed as the arithmetic mean of δ_1 and δ_2.

$$\delta = \frac{\delta_1 + \delta_2}{2}$$

2.3. Montage

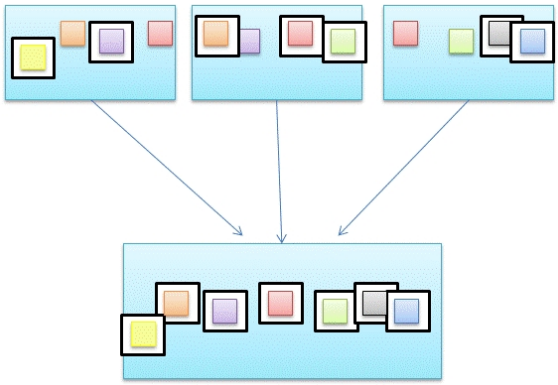

Fig. 4. Illustration of the montage creation process. Best expressions from various frames previously selected are patched to create a new frame.

The montage is illustrated in Figure 4 and created as follows:

1. As a rough indicator of the desired position, the detected face's coordinates are mapped to the corresponding coordinates in the mosaiced image using the computed parameters from the mosaicing algorithm.
2. Simultaneously multiple faces which have similar coordinates are grouped, and the face with the maximum facial expression measure (δ) is selected.
3. A broad alignment with the body position is also made
4. Given these tentative positions, Graph cut [13] is used to find the accurate boundary of the inserted face. In brief, the base boundaries are tied to source node and assigned a high weight. In-between nodes are assigned the absolute difference of gradient level.
5. Around the graph-cut segmentation, image blending is done between the base photo and the selected face.

3. EXPERIMENTS

To compare our method, we ran our experiments against the stack of images provided in [10]. As can be seen, the output photo generated had good expressions of most of the people. The result is presented in Figure 6. Note that we have automatically generated the photo montage without user input compared to the original method.

Fig. 6. AutoMontage created from family stack of image [10]. We are able to automatically generate the photo montage as opposed to the method in [10].

Our system has also been tested on other group photo sessions collected from youtube. Our algorithm successfully created photo montages from all these videos. Examples are presented in Figure 6 and Figure 7.

In Figure 6, though most of the faces looks good, there is an artifact in the third person from top left corner. This is introduced by the patching scheme when multiple face's best expression was substituted. This scenario demands more accurate detection of faces to avoid such artifacts.

Fig. 5. AutoMontage created by generating a mosaic of the Youtube video available in [12].

Fig. 7. Photo Montage example having acceptable expressions. Youtube Video is available in [14].

4. CONCLUSION

With the increased usage of camera by novices, tools to make photography sessions are becoming increasingly valuable. Our work successfully creates photo montage from photo session videos and combine the best expressions into a single photo.

5. REFERENCES

[1] P. Lucey, J.F. Cohn, K.M. Prkachin, P.E. Solomon, and I. Matthews, "Painful data: The unbc-mcmaster shoulder pain expression archive database," in *Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, March 2011, pp. 57 –64.

[2] G.C. Littlewort, M.S. Bartlett, L.P. Salamanca, and J. Reilly, "Automated measurement of children's facial expressions during problem solving tasks," in *Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, March 2011, pp. 30 –35.

[3] S. Kumano, K. Otsuka, D. Mikami, and J. Yamato, "Analyzing empathetic interactions based on the probabilistic modeling of the co-occurrence patterns of facial expressions in group meetings," in *Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, March 2011, pp. 43 –50.

[4] Sung-Ho Lee, Jong-Woo Han, Oh-Jung Kwon, Tae-Hyun Kim, and Sung-Jea Ko, "Novel face recognition method using trend vector for a multimedia album," in *Consumer Electronics (ICCE), 2012 IEEE International Conference on*, jan. 2012, pp. 490 –491.

[5] Jun Li, Joo Hwee Lim, and Qi Tian, "Automatic summarization for personal digital photos," in *Information, Communications and Signal Processing, 2003 and Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint Conference of the Fourth International Conference on*, dec. 2003, vol. 3, pp. 1536 – 1540 vol.3.

[6] Cheng-Hung Li, Chih-Yi Chiu, Chun-Rong Huang, Chu-Song Chen, and Lee-Feng Chien, "Image content clustering and summarization for photo collections," in *Multimedia and Expo, 2006 IEEE International Conference on*, july 2006, pp. 1033 –1036.

[7] M. Das and A.C. Loui, "Automatic face-based image grouping for albuming," in *Systems, Man and Cybernetics, 2003. IEEE International Conference on*, oct. 2003, vol. 4, pp. 3726 – 3731 vol.4.

[8] Zhaojie Liu and Haizhou Ai, "Automatic eye state recognition and closed-eye photo correction," in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, dec. 2008, pp. 1 –4.

[9] Suk Hwan Lim, Qian Lin, and A. Petruszka, "Automatic creation of face composite images for consumer

applications," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, march 2010, pp. 1642 –1645.

[10] Aseem Agarwala, Mira Dontcheva, Maneesh Agrawala, Steven Drucker, Alex Colburn, Brian Curless, David Salesin, and Michael Cohen, "Interactive digital photomontage," in *Proceedings of SIGGRAPH 2004*, 2004.

[11] P. Viola and M. Jones, "Robust real-time face detection," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, 2001, vol. 2, p. 747.

[12] "Youtube video jc class of 1976 reunion-group photo session," `http://www.youtube.com/watch?v=9FZTh_BkRD4`.

[13] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, 1999, vol. 1, pp. 377 –384 vol.1.

[14] "Youtube video 94rotary convention_youth-hub committee 17th. rotaractors," `http://www.youtube.com/watch?v=KQ73-P9HGiE`.