

Spot the Differences: From a Photograph Burst to the Single Best Picture

H. Emrah Tasli

Jan C. van Gemert

Theo Gevers

ISLA, Informatics Institute
University of Amsterdam

ABSTRACT

With the rise of the digital camera, people nowadays typically take several near-identical photos of the same scene to maximize the chances of a good shot. This paper proposes a user-friendly tool for exploring a personal photo gallery for selecting or even creating the best shot of a scene between its multiple alternatives. This functionality is realized through a graphical user interface where the best viewpoint can be selected from a generated panorama of the scene. Once the viewpoint is selected, the user is able to go explore possible alternatives coming from the other images. Using this tool, one can explore a photo gallery efficiently. Moreover, additional compositions from other images are also possible. With such additional compositions, one can go from a burst of photographs to the single best one. Even funny compositions of images, where you can duplicate a person in the same image, are possible with our proposed tool.

Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous;
H.1.2 [User/Machine Systems]: Human factors, Human information processing; I.2.10 [Vision and Scene Understanding]: Image Analysis

Keywords

Organizing Personal Photo Albums

1. INTRODUCTION

The wide availability of digital cameras and the abundance of storage space has altered the way that photos are taken since the analog days. In the pre-digital era, a photo was preferably taken only once to reduce film development costs. This stands in contrast to modern days, where one photo of a scene is never enough. To not miss the possibility of a better shot, nowadays people take several, near-identical, photos of the same scene.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

ACM Multimedia October 21 - 25 2013 Barcelona, Spain
Copyright 2013 ACM 978-1-4503-2404-5/13/10 ...\$15.00.

These near-identical photos represent the same semantic event or scene and thus the final choice typically remains a single image. The single best image could be the one where everybody has their eyes open, has the best composition, or where there is less unwanted artifact such as motion blur or over-exposure. Several of these traits can be analyzed quickly on the camera itself, however the final selection process is often done at home, at a larger screen to appreciate small detailed aesthetic differences that can make or break a photograph.

Selecting the best image from multiple near-identical photographs is now typically done manually, by comparing all image pairs. Hence, after a quick scan to remove the obvious lesser quality images, a naive approach may require $N(N + 1)/2$ operations to look at all image pairs, comparing minute differences to evaluate photo composition and image quality.

In this paper we present a system that helps the user in the photo selection process. We have two contributions. First, we help the user select the single best image by first aligning all images and visualizing the differences. Second, we allow the user to *create* completely new images by moving difference-regions from one photograph to another.

2. RELATED WORK

A related study in [4] proposes that people would like to make more than one photo of the same scene and hence, albums contain a considerable amount of near duplicates images. They propose to detect those near duplicates automatically. The study in [5] focuses on the activities of people with photos. They investigate the patterns of how people browse and search between photos. Another work on photo grouping [15] proposes a home photo management tool. An interface is also provided for easy browsing and searching. Another study [13] automatically groups photos from a single event into smaller selection of photos. The work in [14] focuses on a different aspect in organizing photos. They investigate the problem of matching images with different viewpoint, light and scale properties. The related work regarding selection of photos from a photo gallery mainly focuses on detecting similar or duplicate photos. In our proposed method, we also incorporate such a functionality but further extend it by providing an interface to actually create new compositions from the available images.

In order to merge different images, complete or remove some parts in a seamless fashion, we use the Poisson Image Editing Method [12]. This method requires a manually created foreground mask to merge the source and target image.

Other work addresses scene completion by using other images in an automatic manner [3]. The authors use manual selection of an image patch for automatic completion from a dataset of internet images. Training-set free inpainting methods focus on removing object patches; however they fill the removed region using the local information from the image itself [6, 17].

For the image alignment step we have used an affine transformation estimation between the images. This is a weak geometry transformation where plane models are well-preserved for viewpoint changes. For aligning images for harder constraints, we could also use more advanced methods as in [8, 9]. However computational complexity and performance of these methods makes them impractical in a responsive user interface.

3. PHOTO SELECTION

In order to create an interactive framework for selecting and/or creating a best composition of multiple shots of a scene, we developed a graphical user interface. In the design of the graphical interface, the main intention was to keep the interaction easy and intuitive.

In our photo selection tool, users start by selecting multiple photos of the same scene. Alternatively, the image folder can be input; and our system selects and groups the photos from the same scene. This is basically performed with a three channel color histogram comparison between the images in the folder and grouped by a similarity metric using a predefined threshold. In the image selection step, we have included an automatic image quality rating module. This module is useful for discarding the photos that are blurry or out of focus [7, 11]. One can also extend this functionality with more detailed aesthetic measures [1, 18]. However, some people might prefer to keep these shots on purpose. For such cases, image quality rating functionality can be manually deactivated.

3.1 Image Alignment

Once the images of the same scene are selected, they are aligned using an affine transformation between the images. To align two images, we have used the standard procedure as follows. 1) SIFT keypoints in the images [10] are found. 2) SIFT descriptor for each detected keypoint is computed. 3) The keypoints between two images are matched. 4) The affine transformation model is estimated using the standard RANSAC [2] procedure. The VLFeat implementation has been widely used in this part of the work [19].

3.2 Canvas Mode

When the images are aligned, images are shown in the "canvas" mode where one can see the panorama of different perspectives of individual shots from the same scene. This panorama is similar to a stitched image mosaic, with all original images visible. See Figure 1 for the canvas mode view. We believe that, this type of visualization is native when we want to select between multiple photos. We first go through the images and try to mark the individual differences; and select the one that best covers the salient regions in the scene (Buildings, statues, stationary or moving objects/people of interest).

3.3 Reference Viewpoint Selection

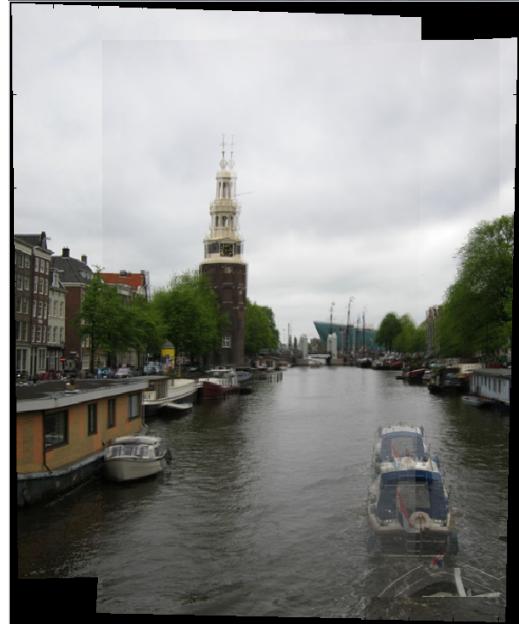


Figure 1: Canvas visualization mode

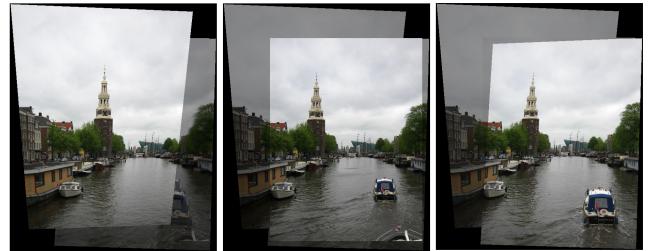


Figure 2: Selecting between views using the slider.

In the "canvas" visualization mode, user can interact with the GUI using a vertical sliding bar to switch between alternative viewpoints. At this point one can still observe which part of the whole canvas the selected single view corresponds to. Figure 2 shows the individual viewpoint selection for a three image setup. When one of the views is selected, others are visualized with a lower (60%) transparency value.

3.4 Spotting the Differences

The selection of the viewpoint defines the reference view; as seen in Figure 3(a). All the other views are aligned with respect to the reference view as in Figure 3(b). At this point, the differences between reference view and all the other images are computed. This step is performed by a window based histogram intersection. It satisfies the goal of detecting dissimilarity regions (blobs) between the reference and corresponding images. The dissimilarity is computed pixel-wise,

$$D_{r,q}(p) = \sum_{q=1:N} \min(h_i(p_r^W), h_i(p_q^W)), \quad (1)$$

where $D_{r,q}(p)$ corresponds to the dissimilarity of a pixel p between the reference r and other image q . $h_i(r)$ is the i^{th} bin in an N bin histogram for a selected window size of W .

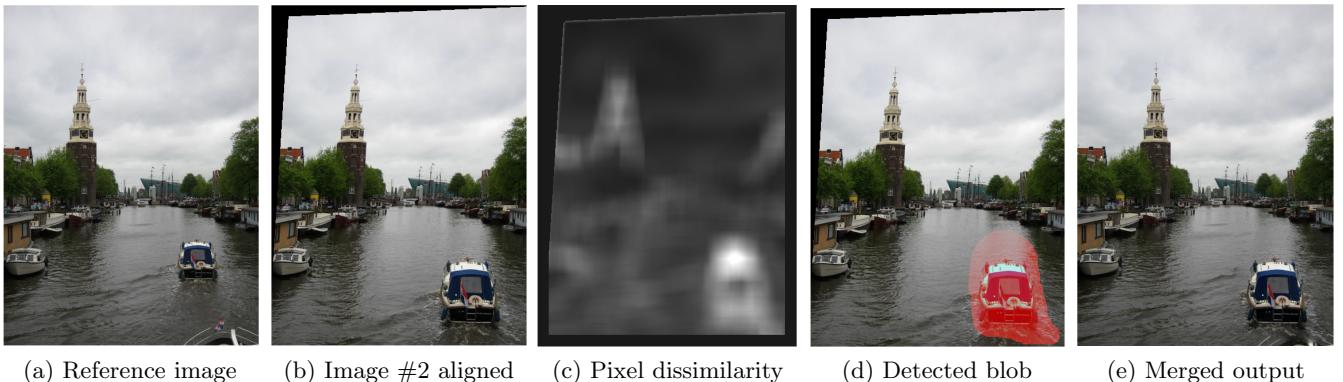


Figure 3: The reference image (a) is selected in the “canvas” mode, other views are aligned with respect to the reference. Only one of the other views is shown here on (b). Estimated dissimilarity cost between reference (a) and source (b) is used to detect blobs. Detected blob is shown in a rectangle for user selection in (a). As an algorithmic output detected blob is visualized in (d) in red color. Blob shown in (d) is taken out from the source image(b) and painted on top of the reference image (a). The final output is shown on (e). Notice the missing flag on the boat at the bottom of the photo

The histogram intersection cost is computed for each pixel in the image.

The estimated pixel level dissimilarity (Figure 3(c)) is used to detect the regions where possible interesting differences between the aligned images are present. A multi-level thresholding on the dissimilarity image defines different blobs with varying sizes. The blobs that are defined in the highest threshold are selected as candidate regions and the final assignment is done using a watershed segmentation starting from these candidate region seeds. The detected blob is shown on the image in the red channel in Figure 3(d). Users can click inside this region to change between alternative paintings.

3.5 Painting the Photo

The detected dissimilarity blob on the image is used to modify the selected view with the detected region from the alternative view. When the blob is selected with the mouse, the corresponding patch from the alternative view is merged on the reference view. For a seamless modification on the reference image, the Poisson Image Editing tool [12] is used. The detected blob from the source image is painted on the target image. The resulting image is shown in Figure 3(e).

4. RESULTS

Our proposed tool offers user friendly interaction for selecting or creating images from a photo album. Visual results and ease of interaction supports this argument.

A careful look at the reference and painted images in Figure 3(a) and Figure 3(e), one can observe that the waves of the boat in the water are still present in the final painted image. This is a logical anomaly; however, not a visual show-stopper for such an application. Moreover, one can observe the disappearance of the flag of the bigger boat at the bottom of the final painted image. These artifacts are mostly due on the blob region segmentation and could be fixed with further user interaction [16].

As a proof of concept, Figure 4 shows possible modifications and capabilities of the proposed tool. Figure 4(a)

shows the reference image selected in the canvas mode. Possible changes on top of the reference images are shown in Figure 4(b)-4(e). One can remove the foreground object like in Figure 4(b) as soon as the background is available in the other images. One can also add additional objects Figure 4(c) and Figure 4(d) (same in this case) from the other images.

Additional visual results are provided in a supplementary video. The graphical user interface, canvas mode, image selection using the slider, blob detection and image matting results can be seen in this video.

We have made a limited evaluation of our tool in terms of usability and output performance. More detailed analysis with regards to an evaluation metric might propose extensions for our future directions.

A main weakness of the proposed method is that possible misalignment of the images might cause erroneous blob regions. Such cases produce dissimilar area candidates; however this would cause no harm since user is supposed to select between dissimilar areas. In order to overcome such cases, we have tried more powerful image matching tools like [9] and [8]. However, these methods have been discarded due to performance issues. [9] can create major artifacts in aligned images and [8] requires very long (20 minutes for a 1200x900 resolution image) computation times; measured with a powerful Intel Xeon 3.4 GHz processor. Therefore the affine model is utilized for matching images in our photo gallery.

5. CONCLUSION

Our proposed method supplies an easy interface for detecting the differences in multiple images of the same scene. In contrast to previous methods, this tool enables not only the best image selection but also a composition of a new image using patches from different photos. With the canvas view module, an efficient viewpoint selection is followed by the dissimilar region identification. With the proposed composition capability, one can process commonly used bursts of photographs to create a single best image.

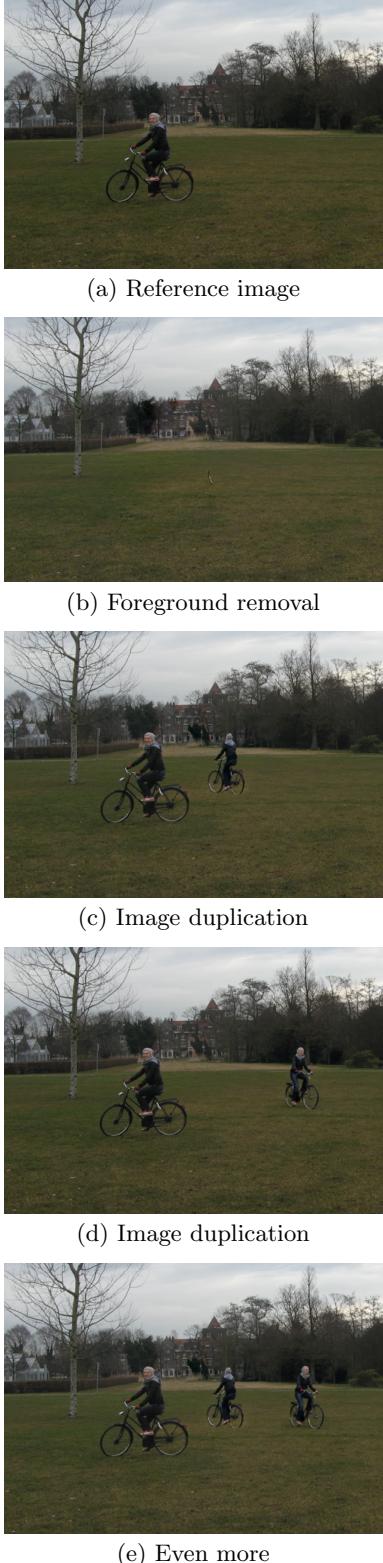


Figure 4: (a) Reference image selected in the "canvas" mode (b) Foreground object can be removed from the image. (c)-(e) One can also add multiple objects from other images

6. REFERENCES

- [1] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Studying aesthetics in photographic images using a computational approach. In *ECCV*, 2006.
- [2] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 1981.
- [3] J. Hays and A. A. Efros. Scene completion using millions of photographs. *ACM Trans. Graph.*
- [4] A. Jaimes, S. Chang, and A. C. Loui. Detection of non identical duplicate consumer photographs. In *Information Communications Signal Processing*, 2003.
- [5] D. Kirk, A. Sellen, C. Rother, and K. Wood. Understanding photowork. In *SIGCHI conference on Human Factors in computing systems*, 2006.
- [6] N. Komodakis and G. Tziritas. Image completion using efficient belief propagation via priority scheduling and dynamic pruning. *Transactions on Image Processing*, 2007.
- [7] A. Levin. Blind motion deblurring using image statistics. In *NIPS*, 2006.
- [8] W.-Y. Lin, L. Liu, Y. Matsushita, K.-L. Low, and S. Liu. Aligning images in the wild. In *Computer Vision and Pattern Recognition*, 2012.
- [9] C. Liu, J. Yuen, A. Torralba, J. Sivic, and W. T. Freeman. Sift flow: Dense correspondence across different scenes. In *European Conference on Computer Vision*, 2008.
- [10] D. Lowe. Object recognition from local scale-invariant features. *International Conference on Computer Vision*, 1999.
- [11] Y. Luo and X. Tang. Photo and video quality evaluation: Focusing on the subject. In *ECCV*, 2008.
- [12] P. Pérez, M. Gangnet, and A. Blake. Poisson image editing. In *ACM SIGGRAPH 2003 Papers*, 2003.
- [13] K. Rodden. How do people organise their photographs. In *Proceedings of the BCS IRSG Colloquium*, 1999.
- [14] F. Schaffalitzky and A. Zisserman. Multi-view matching for unordered image sets, or Should we organize my holiday snaps? In *ECCV*, 2002.
- [15] Y. Sun, H. Zhang, L. Zhang, and M. Li. Myphotos: a system for home photo management and processing. In *Proceedings of the tenth ACM international conference on Multimedia*. ACM, 2002.
- [16] H. E. Tasli and A. A. Alatan. Interactive object segmentation for mono and stereo applications : Geodesic prior induced graph cut energy minimization. *ICCV Workshop on Human Interaction on Computer Vision*, 2011.
- [17] C. E. Toyama, A. Criminisi, P. Pérez, and K. Toyama. Object removal by exemplar-based inpainting. In *Computer Vision and Pattern Recognition*, 2003.
- [18] J. C. van Gemert. Exploiting photographic style for category-level image classification by generalizing the spatial pyramid. In *International Conference on Multimedia Retrieval*, 2011.
- [19] A. Vevaldi and B. Fulkerson. Vlfeat an open and portable library of computer vision algorithms. In *Proceedings of ACM International Conference on Multimedia*, 2010.