Motion Compensated Automatic Image Compositing for GoPro Videos

Ryan Lustig Balu Adsumilli David Newman
GoPro *
Carlsbad, CA

Abstract

Image composition for GoPro videos captured in the presence of significant camera motion is a manual and time consuming process. Existing techniques typically fail to automate this process due to the wide-capture field of view and high camera motion of such videos. The proposed method seeks to solve these problems by developing an image registration algorithm for fisheye images without expensive pixel warping or loss of field of view. Background subtraction is performed to extract moving foreground objects, which are noise corrected and then layered on a reference image to build the final composite. The results show marked improvements in accuracy and efficiency for automating image composition.

Keywords: image registration, background subtraction, motion compensation, fisheye, composite

Concepts: •**Computing methodologies** → **Motion processing;** *Computational photography;*

1 Introduction

Image compositing is a creative process of combining image elements from multiples sources to produce a final image for visual effect. So called "action shot composites" are a popular use of this technique where multiple foreground objects are merged or mosaicked onto a single reference image as in Fig. 1. This is a very time consuming task, especially in the presence of camera motion, requiring the user to manually extract foreground objects and carefully layer each one onto a final image.

Existing techniques of automating this process only work using images in rectilinear space [Matt 2013]. The usual process is to first align each image to a common reference image then extract the moving foreground object. For wide field of view fisheye images, such techniques fail for a few reasons. First, when camera motion is present, the severe lens distortion affects the image registration quality because straight lines curve and objects change size across the scene. Second, high frequency noise in images (swaying trees, etc.) results in poor foreground segmentation due to the low quality background segmentation algorithms used. To solve this problem, an improved image registration algorithm was developed and a high quality background subtraction was applied.

Image registration is extensively researched with a variety of feature-based and pixel-based methods [Szeliski 2006]. Usually, these techniques focus on the registration of rectilinear images with low to medium field of view. With the growing popularity of action cameras with wide-angle (fisheye) lenses, so does the importance

Figure 1: *Automated action shot composite using a fisheye lens.*

for a registration algorithm for fisheye lenses. The above registration techniques can be naïvely applied to fisheye images, but the images must first be de-warped, cropped, registered, then warped back into fisheye space. This process is computationally expensive and the resulting image is cropped, thereby losing field of view. [Xiong and Turkowski 1997] proposes to use multi-level gradient-based registration to register fisheye images while self-calibrating the distortion and field of view, but this is based on a equi-distance lens camera model and does not work well for GoPro videos.

Background subtraction is also a well studied topic. In the last decade alone, many algorithms have been proposed to solve this problem [Bouwmans et al. 2008; Sobral and Vacavant 2014]. Such algorithms range from multiple Gaussian, frame-based, feature-based, and non-parametric methods.

A method is proposed to perform automatic compositing of videos in the presence of fisheye distortion. This method robustly compensates for camera motion and high frequency noise without expensive pixel warping or loss of field of view. Applying background subtraction to each registered image extracts the moving foreground objects which are then blended on to a final composite.

2 Technical Approach

Consider two fisheye images (I_{ref}, I_{src}) that contain camera and subject motion (Figs. 2a and 2b). A fisheye imaging model is a mapping from 3D space points to 2D fisheye image points. Thus, 2D image features between the pair of images are first detected and tracked (in general, any modern feature detection and extraction algorithm works well). Then, given a calibrated or known fisheye lens distortion model, the 2D image space feature coordinates are transformed into 3D spherical coordinates, where the radius is enforced to always lie on the unit sphere.

Two methods are used to eliminate incorrectly matched features. The first method clusters the feature tracks since such tracks on the background result from camera motion and are typically more consistent than foreground tracks. This clustering leads to candidate regions from the background. The clusters that are small in size or have large variances are removed. The remaining feature tracks are refined using RANSAC (Fig. 2c). The model used in RANSAC

*email: {rlustig, badsumilli, dnewman}@gopro.com

SIGGRAPH '16, July 24-28, 2016, Anaheim, CA,
ISBN: 978-1-4503-4371-8/16/07
DOI: http://dx.doi.org/10.1145/2945078.2945090

Figure 2: Motion Compensation Process: (a) *Reference I_{ref} and* (b) *source I_{src} fisheye images, depicting camera motion between each image.* (c) *Clustering and RANSAC on tracked features (on suppressed background) to detect foreground inliers (green) and outliers (red).* (d) *Registered source image obtained by re-pointing in spherical image space.* (e) *Binary mask of moving foreground object after noise removal.* (f) *Final action shot composite with extracted foreground objects layered on the reference image.*

comes from [Umeyama 1991], which determines the optimal rotation and translation between corresponding sets of 3D points.

Next, the source fisheye image I_{src} is mapped into spherical space (using a known or calibrated lens distortion model) and then re-pointed by applying the computed rotation above. Finally, this image is transformed back into the fisheye image space, thereby aligning it with the reference image I_{ref}. This step is computationally efficient and it preserves the image's field of view since no cropping is performed (Fig. 2d). A high quality background subtraction algorithm processes each registered image and then noise removal is performed to clean up the extracted foreground objects (Fig. 2e). Finally, each foreground object is layered onto I_{ref}. This operation is repeated for multiple frames producing the final action shot composite (Fig. 2f).

3 Discussion

Currently, the proposed automatic compositing algorithm robustly handles camera motion with rotations up to 20 degrees and for small translations (e.g., shaky handheld camera footage following a moving foreground subject, x-y planar camera motions). The robustness decreases as the camera translation increases. Additionally, since the proposed idea is tested on GoPro videos, the lens distortion is known/calibrated. Further investigations include exploring structure from motion to accurately model the 3D scene for improved registration as well as extending the lens distortion model to the general case. Lastly, automatically choosing video segments best suited for image composition requires additional investigation.

References

BOUWMANS, T., EL BAF, F., AND VACHON, B. 2008. Background modeling using mixture of gaussians for foreground detection-a survey. *Recent Patents on Computer Science 1*, 3, 219–237.

MATT, 2013. The galaxy s4: A life companion. http://www.samsung.com/uk/discover/mobile/galaxy-s4-a-life-companion/, June. Accessed on March, 1 2016.

SOBRAL, A., AND VACAVANT, A. 2014. A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. *Computer Vision and Image Understanding 122*, 4–21.

SUNKAVALLI, K., JOSHI, N., KANG, S. B., COHEN, M. F., AND PFISTER, H. 2012. Video snapshots: Creating high-quality images from video clips. *Visualization and Computer Graphics, IEEE Transactions on 18*, 11, 1868–1879.

SZELISKI, R. 2006. Image alignment and stitching: A tutorial. *Foundations and Trends® in Computer Graphics and Vision 2*, 1, 1–104.

UMEYAMA, S. 1991. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 4, 376–380.

XIONG, Y., AND TURKOWSKI, K. 1997. Creating image-based vr using a self-calibrating fisheye lens. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, IEEE, 237–243.