Photographically Guided Alignment for HDR Images

Ahmet Oğuz Akyüz†

Middle East Technical University

Abstract

This paper presents an automatic image alignment algorithm that alleviates the need to keep the camera still during the capture of a bracketed sequence to obtain an HDR image. Our algorithm assumes that the misalignment between the two consecutive exposures is translational. Using a photographically guided random search, our algorithm first finds properly exposed high contrast regions. The shift amount is then found by analyzing and matching the pixel correlations inside these regions.

Categories and Subject Descriptors: Image Processing and Computer Vision [I.4.3]: Enhancement—Registration

1. Introduction

Alignment for the purpose of generating HDR images is a relatively new topic. Ward observed that conventional image alignment methods often fail when used for aligning differently exposed photographs [War03]. He proposed an alignment approach based on median threshold bitmaps (MTB), which are largely invariant with respect to changing exposure. By comparing the MTBs of two adjacent exposures one can more easily detect the translational shifts between the two exposures. Ward's method has recently been extended to accommodate more types of image artifacts such as rotation [Gro06, JLW08, LHW*09].

Cerman and Hlaváč [CH06] estimated the translation and rotation amounts by computing the correlation of the images in the Fourier domain. Tomaszewska and Mantiuk [TM07] proposed to use standard computer vision techniques to achieve alignment. They employed a modified SIFT to extract local features in the images to be aligned. These features were fed into RANSAC to reduce the number of key-point pairs. The refined set of key-points was then used to compute a homography between the images.

In addition to methods developed for static images, several algorithms have been devised to register differently exposed video frames [KUWS03, ST04]. These methods have typically used optical flow for motion detection. A common feature of the existing methods is that they exploit invariants

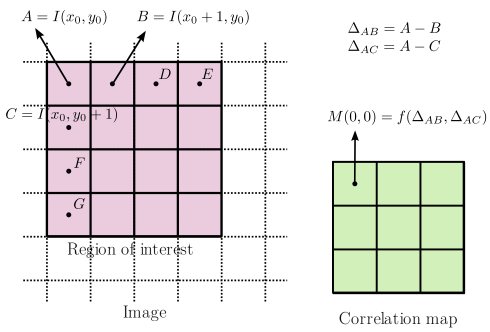

Figure 1: *The computation of correlation map (see Table 1).*

that tend to stay constant across exposures. This, too, is one of the key features of our algorithm. However, we use a different type of invariant that allows for a simpler and more efficient implementation than the previous methods.

2. Algorithm

The key observation of our algorithm is that although absolute pixel values change from one exposure to another, correlations between pixels tend to remain constant. For example, if a pixel's value is larger than its right neighbor and smaller than its left neighbor in a given exposure, this relationship will also hold in the next exposure unless the pixels become under- or over-exposed, or affected by noise. By forming a series of such correlations in carefully selected areas of an image, we can construct *correlation maps* that are largely invariant between two adjacent exposures. This simplifies the problem of image alignment to searching for the most

† akyuz@ceng.metu.edu.tr

Figure 2: *The constituent exposures of the middle image are aligned by using our algorithm with a correlation map size of* 32 *pixels and a search diameter of* 80 *pixels in each dimension. The hand-held bracketed sequence contained* 5 *exposures.*

$f(s,t)$	$s > \varepsilon$	$s < -\varepsilon$	$\lvert s \rvert \leq \varepsilon$
$t > \varepsilon$	1	2	3
$t < -\varepsilon$	4	5	6
$\lvert t \rvert \leq \varepsilon$	7	8	9

Table 1: f *maps pixel correlations to ordinal values.*

similar correlation maps. Our approach is akin to the ordinal measures for image correspondence [BN98].

We call a set of pairwise pixel relations in a rectangular region of an image a correlation map. Assume that $I(x,y)$ represents the intensity of the pixel at coordinates (x,y) such that $A = I(x_0,y_0)$, $B = I(x_0+1,y_0)$ and $C = I(x_0,y_0+1)$ (Figure 1). To find the corresponding value in the correlation map, we first compute the horizontal and vertical pixel differences, $\Delta_{AB} = A - B$ and $\Delta_{AC} = A - C$. The correlation map value is then computed as $M(i,j) = f(\Delta_{AB},\Delta_{AC})$, where f is defined in Table 1. Here, ε is a small tolerance used to prevent the influence of random intensity fluctuations. We set $\varepsilon = 1$, but it can also be set based on the noise level.

It is important to use a suitable image region for computing a correlation map. First, this region should preferably correspond to the important elements of the captured scene. Second, the region should contain ample intensity variations in both horizontal and vertical directions. To achieve these, we perform randomized rejection sampling that favors high contrast areas in the photographically important parts of the captured scene. Although it is an ill-posed problem to accurately detect which parts of the scene are important, we attempt to estimate it by following the rule of thirds of photography which suggests that the most important elements of the composition should be placed at the intersection of the imaginary lines that divide the scene into three equal parts in either direction. Thus, we randomly select a center pixel in proximity of these intersections and compute the mean horizontal (H) and vertical (V) edge magnitudes within a small region around it. If both H and V are greater than a mean edge threshold τ we accept this region, otherwise we generate a new random point. We initially set $\tau = 50$ and reduce its value if no region with sufficient contrast is found.

Once a correlation map is constructed, we search the second image for the most similar correlation map around the same neighborhood. We use the *Hamming distance*, d_h, as a measure of similarity between the correlation maps:

$$d_h(M_1,M_2) = \sum_{x,y}(|\,\mathrm{sgn}(M_1(x,y) - M_2(x,y))|) \qquad (1)$$

where $\mathrm{sgn}(x) = x/|x|$. We choose M_2 that minimizes d_h. In our method we only align consecutive images; offsets are accumulated to determine the shifts between disjunct pairs. A sample result is illustrated in Figure 2.

The biggest limitation of our method is the lack of support for rotation. A potential extension that would also support rotational alignment can be made by transforming the images into the log-polar domain prior to alignment where horizontal and vertical shifts can be used to simultaneously solve for the rotation angle and the translation offsets.

References

[BN98] Bhat D., Nayar S.: Ordinal measures for image correspondence. *IEEE PAMI 20*, 4 (1998), 415–423.

[CH06] Cerman L., Hlaváč V.: Exposure time estimation for high dynamic range imaging with hand held camera. In *Computer Vision Winter Workshop, Czech Republic* (2006).

[Gro06] Grosch T.: Fast and robust high dynamic range image generation with camera and object movement. In *Proc. of Vision Modeling and Visualization* (2006), pp. 277–284.

[JLW08] Jacobs K., Loscos C., Ward G.: Automatic high-dynamic range image generation for dynamic scenes. *IEEE Computer Graphics and Applications 28*, 2 (2008), 84–93.

[KUWS03] Kang S. B., Uyttendaele M., Winder S., Szeliski R.: High dynamic range video. *ACM Trans. Graph. 22*, 3 (2003), 319–325.

[LHW*09] Lu P.-Y., Huang T.-H., Wu M.-S., Cheng Y.-T., Chuang Y.-Y.: High dynamic range image reconstruction from hand-held cameras. In *CVPR* (2009), pp. 509–516.

[ST04] Sand P., Teller S.: Video matching. *ACM Trans. Graph. 23*, 3 (2004), 592–599.

[TM07] Tomaszewska A., Mantiuk R.: Image registration for multi-exposure high dynamic range image acquisition. In *WSCG: Proc. of the 15th Intl. Conf. in Central Europe on Computer Graphics, Visualization and Computer Vision* (2007).

[War03] Ward G.: Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures. *Journal of Graphics Tools 8*, 2 (2003), 17–30.