ON GEOMETRIC AND PHOTOMETRIC REGISTRATION OF IMAGES

Murat Gevrekci and Bahadir K. Gunturk

Louisiana State University
Department of Electrical and Computer Engineering
Baton Rouge, LA 70803
lgevre1@lsu.edu, bahadir@ece.lsu.edu

ABSTRACT

Finding geometric and photometric relation among images is crucial in many computer vision tasks such as panoramic imaging, high dynamic range imaging, stereo imaging, and change detection. Most photometric registration algorithms require accurate geometric registration of images. On the other hand, geometric registration may fail when images are not aligned photometrically. There are two contributions of this paper: (i) A contrast invariant feature detection algorithm is proposed. This would allow geometric registration of images without photometric registration. (ii) A photometric registration algorithm that can handle scene occlusions is presented.

Index Terms— Contrast invariant feature extraction, photometric registration

1. INTRODUCTION

Accurate geometric and photometric registration of images is necessary in a variety of applications. For example, high dynamic range (HDR) imaging aims to construct scene radiance using multiple pictures of the same scene. This requires estimation of the camera response function (CRF) and exposure ratios. If the images were not taken with a camera fixed on a tripod, geometric registration would be required.

Significant amount of work has been done about photometric registration. Debevec and Malik constructed HDR radiance in addition to the CRF with known exposure durations [1]. They defined an objective function to minimize radiance differences, and used the second derivative of the inverse camera response function for regularization. Tsin and Kanade modeled the imaging process using statistical calibration, and estimated both CRF and white balancing parameters jointly in an iterative manner [2]. Mitsunga and Nayar fit a polynomial function to CRF iteratively starting from a rough estimate of the exposure ratios of images [3].

All these photometric registration methods require geometric registration. On the other hand, geometric registration may fail when images are not aligned photometrically. There are three possible approaches to this problem: (i) Images are first geometrically registered using an algorithm that is insensitive to photometric changes. This is followed by photometric registration. (ii) Images are first photometrically registered using an algorithm that is insensitive to geometric misalignments. This is followed by geometric registration. (iii) Geometric and photometric registration parameters are estimated jointly.

There are few algorithms that can be utilized for these approaches. In [4], an exposure-insensitive motion estimation algorithm based on the Lucas-Kanade technique is proposed to estimate motion vectors at each pixel. Although this method can be used to estimate large and dense motion field, it has the downside that it requires pre-knowledge of the CRF. Another exposure-insensitive algorithm is proposed in [5]. It is based on bit-matching on binary images. Although it does not require knowing CRF in advance, the algorithm is limited to global translational motion. In [6], an IMF estimation algorithm that does not require geometric registration is proposed. It is based on the idea that histogram specification gives the intensity mapping between two images when there is no saturation or significant geometric misalignment. And finally in [7], a joint geometric and photometric registration algorithm was proposed. There, the problem is formulated as a global parameter estimation, where the parameters are the parameters of geometric transformation, exposure rate, and CRF. Two potential problems associated with this approach are (1) getting stuck at a local minima and (2) limitation of using parametric CRF.

In this paper, we first propose a contrast invariant feature detection algorithm. This will be used to geometrically register images without photometric registration. Major advantages of the feature based registration are (i) the ability to handle large geometric variations and (ii) robustness to outliers. Our algorithm, which we call contrast invariant feature transform (CIFT), is based on obtaining a contrast signature of a feature detector. And it can be used with any feature detector. We will show how CIFT improves the phase congruency corner detector presented in [9].

The second contribution of the paper is an improvement of [6]. In [6], Grossberg and Nayar proposed an algorithm to align images photometrically without doing accurate geometric registration. Their algorithm is based on the idea that two differently exposed images should have the same histogram

This work was supported by the National Science Foundation under Grant No 0528785.

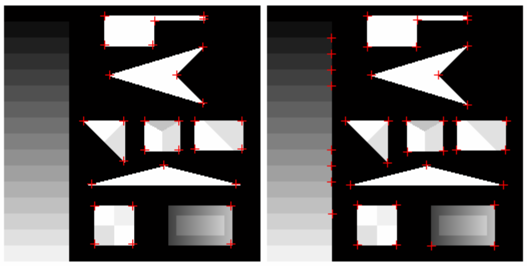


Fig. 1. Left: Result of the Harris corner detector. Right: Result of the Phase Congruency corner detector.

after histogram equalization. The negative aspect is that it requires all pixels to be visible. This is not very realistic since there might be occlusion due to moving objects and saturation due to limited dynamic range of camera. We improve this algorithm by adding occlusion robustness. Intensity mapping function (IMF) is found using the expectation maximization (EM) technique, starting with an unknown visibility map, which assigns binary values to pixels depending on being occluded or not. IMF and visibility map are updated iteratively leading to an intensity conversion with less error. Visibility map eliminates occluded pixels and improve the accuracy of IMF through iterations. It is also helpful in creating HDR images. Skipping occlusion elimination step may result in ghost effect in HDR imaging.

2. CONTRAST INVARIANT FEATURE TRANSFORM

Maybe the most widely used corner detector is the Harris corner detector [8]. The Harris corner detector is based on the image gradient, and is highly sensitive to contrast and illumination changes. An important feature detection algorithm that is relatively insensitive to illumination changes is proposed in [9]. The idea is based on the *local energy model*, which postulates that the Fourier components are in phase at corners and edges. Another advantage of this approach is the feature localization. Most gradient based algorithms apply Gaussian smoothing to deal with noise; and smoothing can sometimes change the location of the features critically. Figure 1 shows results of these algorithms applied on a test image. It is clear that the Phase Congruency method [9] is working better than the Harris corner detector. But it is still missing corners.

Here, we propose contrast invariant feature transform (CIFT) to improve feature detection against contrast changes. The CIFT stretches image contrast as a function of intensity. Combined with a feature detector, a "signature" of a point (as a function of intensity) is obtained. This signature is then used to determine feature locations. Figure 2 shows contrast stretching functions at two center intensities. The intensity range of the input image is scaled to 0 to 1. The contrast stretching functions are obtained using a sigmoid function. The center intensity is changed from 0 to 1 with small discrete increments to obtain a set of images. The feature detector is

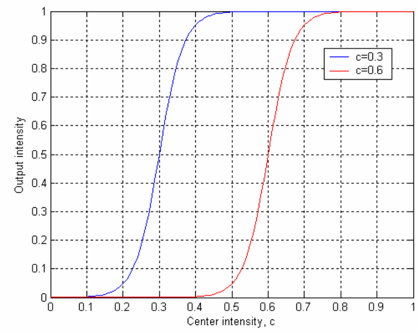


Fig. 2. Contrast stretching functions at two center intensities.

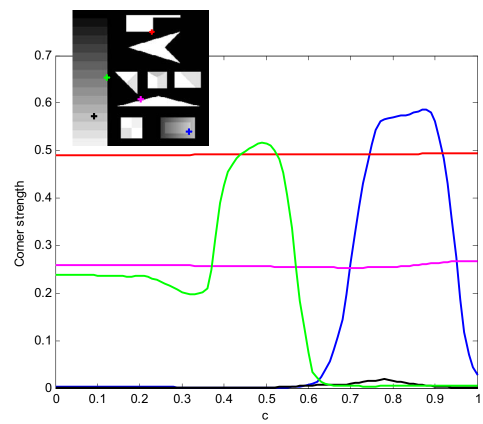


Fig. 3. Contrast signatures at different pixels for CIFT applied on Phase Congruency method.

applied on these images to obtain feature strength for each pixel as a function of center intensity.

Contrast signatures for several pixels are shown in Figure 3. Those pixels are selected from various features. For a strong (high-contrast) corner, the signature (red line) is always high. On the other hand for a smooth pixel, the signature (black line) is always low. For a strong edge, the signature (purple line) has a constant medium level response. For a low-contrast corner, the signature (blue line) peaks around a center contrast stretching intensity. For a T-junction corner, the signature (green line) shows edge-like behavior for some intensities and corner-like behavior for others.

After getting the signatures for each pixel, we can take the maximum response (corner strength) for each pixel. This corner strength can then be used to determine corners after thresholding and non-maximum suppression.

3. PHOTOMETRIC REGISTRATION

3.1. Estimating IMF Using Expectation Maximization Technique

The expectation maximization (EM) technique is suitable for finding maximum likelihood estimate iteratively in case model

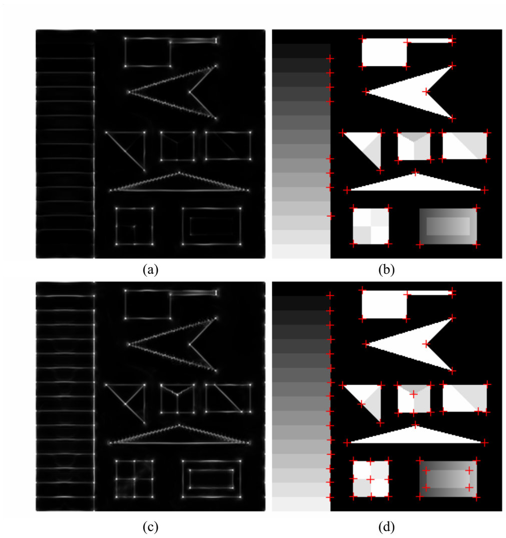


Fig. 4. Upper row: Corner strength and detected corners using the Phase Congruency method. Lower row: Corner strength and detected corners using CIFT + Phase Congruency.

depends on hidden variables. EM has been used in computer vision applications extensively. Since an intensity mapping function between two images is valid for only unoccluded pixels, we need to use a visibility map to define the the visible (unoccluded) pixels.

Let V be the visibility map between two images. The visibility map takes a value of 1 (for unoccluded pixels) or 0 (for occluded pixels). The maximum likelihood estimate can be found by maximizing the log-likelihood of the conditional probability $p(I_1, I_2|g)$:

$$\hat{g} = \arg\max_g \log p(I_1, I_2|g). \tag{1}$$

This requires estimation of the visibility map V, which can be incorporated into the picture using the following marginalization:

$$p\left(I_1, I_2|g\right) = \int p\left(I_1, I_2, V|g\right)\, dV. \tag{2}$$

Since V is binary for a pixel, equation (1) can be written as

$$\hat{g} = \arg\max_g \log \sum_V p(I_1, I_2, V|g), \tag{3}$$

where the summation is over the pixels for which the visibility map is 1.

EM aims to find the maximum likelihood estimate of both the visibility map and IMF. At the start of algorithm, both g and V are unknown; and the algorithm alternates between expectation (E) and maximization (M) steps until convergence. At the kth iteration, the steps are as follows.

- *Step E*: We want to find g which maximizes likelihood in equation (3), but V is unknown initially. So, instead, we use the expected value V given the images and the current estimate of g. The following Q function is to be maximized:

$$Q(g, g^{[k]}) \equiv \sum_V p(V|I_1, I_2, g^{[k]}) \log p(I_1, I_2, V|g). \tag{4}$$

- *Step M*: The new IMF is found as follows:

$$g^{[k+1]} = \arg\max_g Q(g; g^{[k]}). \tag{5}$$

3.2. Initial Estimation of IMF

EM algorithm requires an initial estimate for IMF. Histogram-based approach in [6] is a good candidate for initializing IMF since it does not require geometric registration. Let $H_1(\cdot)$ and $H_2(\cdot)$ be the histogram equalization functions calculated from I_1 and I_2, respectively. When there is no saturation or occlusion, histogram-equalized images must be identical: $H_2(I_2) = H_1(I_1)$, from which we can write $I_2 = {H_2}^{-1}(H_1(I_1))$. As a result, the initial estimate for IMF is given as

$$g^{[0]}(\cdot) = {H_2}^{-1}(H_1(\cdot)). \tag{6}$$

3.3. Estimating Visibility Map

The visibility map is a binary, and the probability $p(V|I_1, I_2, g^{[k]})$ will be either 1 or 0 based on the difference $|g^{[k]}(I_1(x)) - I_2(x)|$. Using a predetermined threshold T_k, and denoting x as the spatial location, visibility map value at location x at the kth iteration is found by

$$V(x,k) = \begin{cases} 0 & \text{, if } |g^{[k]}(I_1(x)) - I_2(x)| > T_k \\ 1 & \text{, else} \end{cases} \tag{7}$$

The threshold T_k could in general be a function of iteration. Intuitively, we may start with a large threshold, therefore taking a lot pixels as visible. As we progress with the iterations, we may reduce the threshold value to have a more and more accurate estimation of visible/invisible pixels. As the visibility map becomes more accurate, the IMF will also become more accurate.

3.4. Estimating IMF

To complete the algorithm we also need to define $p\left(I_1, I_2, V|g\right)$. Let $H_{i,V}$ be the histogram equalization function (of the ith image) found from the visible pixels only. We then use a Gaussian model to define $p\left(I_1, I_2, V|g\right)$:

$$p\left(I_1, I_2, V|g\right) \propto \exp\left(-\|(g(I_1) - H_{2,V}^{-1} H_{1,V}(I_1))\|^2 / 2\sigma^2\right), \tag{8}$$

where σ^2 is the noise variance. Assuming independence among different spatial locations, the IMF estimate at kth iteration is

$$g^{[k]}(I_1) = H_{2,V}^{-1} H_{1,V}(I_1). \tag{9}$$

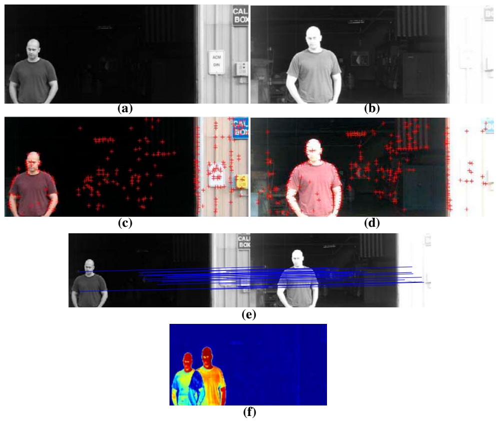


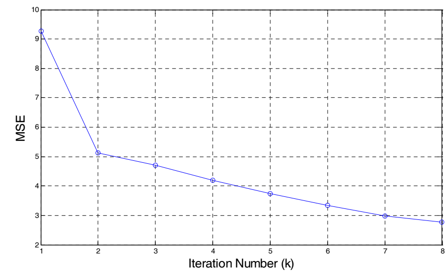
Fig. 5. (a)-(b) Input images. (c) Features extracted using CIFT + Phase Congruency. (d) Inliers found after RANSAC. (e) Residual after geometric and photometric registration.

4. EXPERIMENTAL RESULTS

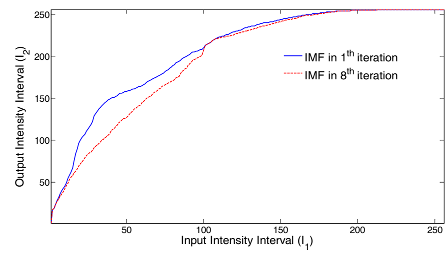
Here, we provide an example to show how iterative refinement of IMF improves the accuracy of intensity conversion between differently exposed images. We captured two images with exposure times of 1/200 seconds and 1/60 seconds. These images are shown in Figure 5. There is also a person moving in the scene. We first applied feature extraction to these images using the CIFT + Phase Congruency method. This is followed by feature matching and outlier rejection with RANSAC. The extracted features are shown in figures 5(c) and (d). The inliers after the RANSAC iterations are shown in Figure 5(e). Figure 5(f) shows the residual image after geometric and photometric registration. In Figure 6(a), the change in the mean square error of the residual (outside the estimated occlusion map) is shown as a function of iterations. Figure 6(b) shows the initial and final intensity mapping functions.

5. CONCLUSION

In this paper we presented a contrast invariant feature detection algorithm. The algorithm can be used with any feature detection technique to improve robustness against illumination changes. This would allow geometric registration of images without photometric registration. We also presented a photometric registration algorithm that can handle scene occlusions. The algorithm is based on the expectation maximization technique. We provided examples to demonstrate these algorithms. In the future, we plan to combine the CIFT algorithm with affine invariant region descriptors to achieve contrast and affine invariance together.



(a)



(b)

Fig. 6. (a) Iterative change in mean square error. (b) IMF in the 1*st* and 8*th* iterations.

6. REFERENCES

[1] P. E. Debevec and J. Malik, "Modeling and rendering architecture from photographs," in *Proc. of the ACM SIGGRAPH*, 1997, pp. 369–378.

[2] Y. Tsin, V. Ramesh, and T. Kanade, "Statistical calibration of the ccd imaging process," *Proc. Int. Conf. Computer Vision*, vol. 1, pp. 480–487, 2001.

[3] T. Mitsunaga and S.K. Nayar, "Radiometric self calibration," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, June 1999, vol. 1, pp. 374–380.

[4] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High dynamic range video," *ACM Transactions on Graphics*, vol. 23, no. 3, 2003.

[5] G. Ward, "Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures," *Journal of Graphics Tools*, vol. 8, no. 2, pp. 17–30, 2003.

[6] M. D. Grossberg and S. K. Nayar, "Determining the camera response from images: what is knowable?," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 11, pp. 1455–1467, November 2003.

[7] F. M. Candocia, "Joint registering images in domain and range by piecewise linear comparametric analysis," *IEEE Trans. Image Processing*, vol. 12, no. 4, pp. 409–419, 2003.

[8] C. G. Harris and M. Stephens, "A combined corner and edge detector," in *Fourth Alvey Vision Conference*, 1988, pp. 147–151.

[9] P. Kovesi, "Image features from phase congruency," *Videre: A Journal of Computer Vision Research*, vol. 1, no. 3, 1999.