Registration of the Prokudin-Gorskii Colour Photographs Using a Multiresolution SSIM Algorithm

Michael Wirth and Bruce Bobier

Dept. Computing and Information Science,
University of Guelph, Guelph, Ontario, Canada
mwirth@uoguelph.ca
http://www.uoguelph.ca/~mwirth

Abstract. Russian photographer Prokudin-Gorskii was one of the pioneers of colour photography, creating projected colour composites through the acquisition of images using red, green, and blue colour filters. This paper explores the use of a multiresolution algorithm incorporating a structural similarity index to align each of the three-frame photographs. The algorithm is applied to medium and high resolution photographs in the Prokudin-Gorskii collection.

Keywords: image registration, SSIM, colour photographs, Prokudin-Gorskii, multiresolution.

1 Introduction

Sergey Mikhaylovich Prokudin-Gorskii (1863-1944) was a Russian-born chemist, inventor and pioneering colour photographer who captured cultural aspects of the Russian Empire through photographic surveys of its geographic regions. Between 1905 and 1915, Prokudin-Gorskii travelled throughout Central Asia, the Marinskii Canal and Murmansk Railway, the Ural Mountains, Volga River and Napoleonic War regions, capturing thousands of photographs. The photographs capture historically significant scenes of traditionally garbed ethnic populations, agricultural and farming practices, the ornate interiors of churches and monasteries, religious artifacts, and scenic landscapes, as well as infrastructure such as factories, railways, and canals [2]. Amongst his photographs were glass plates composed of three frames, representing monochromatic images acquired using blue, green and red filters. The glass positives are Ilford "red label" plates chemically rendered by Prokudin-Gorskii's own process. Each plate is 9×24 cm with each frame 8.5cm wide and a variable 7.5-8cm in height. After processing, Prokudin-Gorskii used a triple-lens projector to combine the plates into a composite colour image. An example of the three colour frames is shown in Figure 1.

In 1918, Prokudin-Gorskii left Russia with a portion of his photographs and settled in France. Prokudin-Gorskii's collection was acquired by the U.S. Library of Congress (LOC) [1] in 1948, and comprises some 1902 three-frame glass plates

M. Kamel and A. Campilho (Eds.): ICIAR 2009, LNCS 5627, pp. 917–926, 2009.

Fig. 1. Example of three colour frames (Blue-Green-Red) from a Prokudin-Gorskii plate (Image 00152)

Fig. 2. Example of misregistration of the three colour frames from Fig.1 (Image 00152)

and 2,433 sepia-tone prints. In 2000, the LOC contracted with JJT Inc. to scan all 1902 triple-frame glass positives. The scans were performed in 16-bit grayscale, with a resolution of 1010ppi. The LOC had the plates manually aligned to generate composite colour images by photographer Walter Frankhauser in 2001, a process he dubbed *digichromatography* [1]. In 2004 the LOC had Blaise Agüera y Arcas develop an algorithm to register the images [3]. An example of a composite image produced *without* registration is shown in Figure 2. This paper explores the use of a multiresolution algorithm incorporating a structural similarity index to register each of the three-frame photographs.

2 Image Registration Process

2.1 Preprocessing

Prior to registration, several preprocessing steps are performed. With high-resolution images that have 16-bits of greyscale information per pixel, the images are first resampled to 8-bits per pixel, as the additional information provides little perceptible difference in image quality, and resampling greatly improves the computation time. For both image resolutions, the three-frame positive is divided into three equally sized images representing the blue, green and red color channels. A number of pixels are then trimmed from the sides of each image (17 pixels for medium resolution, and 200 pixels for high resolution images) to remove the border surrounding each frame. In some high resolution images (e.g. 00033u.tif), dividing the original image and trimming each channel results in the green channel having one fewer rows of pixels due to the original plate scan having a number of rows that is not divisible by three. Thus, for each high-resolution image, a quick comparison of dimensions is performed between the blue and green channels, and if the heights are different, the green channel is zero-padded with one row along the bottom.

2.2 Transformation Model

Image registration is an amalgam of various processes. Central to any registration algorithm is the transformation model used to relate the target and reference images. The simplest form of transformation is a linear, or *rigid*, transformation, a global mapping which takes into account spatial changes such as translation and rotation. The spatial dissimilarity between frames is primarily in the form of horizontal and vertical translations due to the mechanics of photograph acquisition.

2.3 Structural Similarity Index

The registration algorithm described here is based on the notion of matching image intensity values using some measure of correspondence. Such correlation, or area-based methods use corresponding windows of a predefined size, from within the target and reference images and calculate a numerical measure based on their homogeneity. From a spatial point of view, area-based methods only allow rigid transformations. Commonly used similarity measures include Normalized Cross-Correlation (NCC) and Sum of Squares Difference (SSD), however the caveat with such measures is that they match image intensities directly, without any structural analysis, resulting in a sensitivity towards intensity changes. The measure used in this study, *structural similarity index* (SSIM) [6], is a more recent measure that determines the loss of structural information between the target and reference images in an attempt to provide an image similarity measure that is more consistent with human perception than is provided by traditional metrics. The structural information in an image is comprised of those attributes that represent the structure of the objects, independent of average luminance and contrast [6]. The SSIM separates the task of similarity measurement into comparisons of structure, luminance and contrast.

The SSIM index produces a value in the range [0,1], where 0 indicates zero correlation, i.e. the reference image is entirely different than the target, and 1 indicates that they are identical.

2.4 Matching Criteria

The registration algorithm attempts to align the reference channel to the target channel by selecting a square window from each channel and using this window to measure similarity. Using the entire channel for comparison requires a substantial amount of computation to be performed at each comparison, and thus, only a subsection of the image is used here to decrease computation time. The algorithm uses four main parameters for registration: *windowSize*, which refers to the dimensions of the window used for comparing the two images; *numMoves*, which represents the number of pixels by which the window is slid horizontally and vertically across the reference image and; *xOffset* and *yOffset*, which specify where the window begins in terms of the number of pixels from the channels top left corner. Here, *windowSize* is set to 38% of the channels width, such that the window for similarity comparison has an area equal to 15% the size of the channel (Figure 3).

Fig. 3. Example of a window used in calculating SSIM (Image 00147)

3 Multiresolution Registration

One of the problems with the use of area-based methods is the extreme computational load. To speed up searching, it is often beneficial to use a multiresolution pyramidal image representation. The Laplacian pyramid [4] is an approach that is commonly used in registration algorithms, and works by downsampling the image at each successively lower level of the pyramid such that it has half the resolution of the level above it. For example, if the bottom level (original image) has 128x128 bits of information, the second level would have 64x64, the third/top level would have 32x32, etc. Figure 4 shows an example of a four-level Laplacian pyramid, where the images have been resized to show the effects of downsampling.

Fig. 4. Example of a four-level Laplacian pyramid for image 00859. Left-to-right: Bottom level of the pyramid (original image), Downsampled to $\frac{1}{2}$ the resolution; Downsampled to $\frac{1}{4}$ the resolution and Downsampled to $\frac{1}{8}$ the resolution(resized to show detail).

At each pyramid level, the multiresolution algorithm essentially implements the rigid matching algorithm. Firstly, *numMoves* is set to 10 and 24 for medium and high resolution images respectively, with $numMoves^2$ comparisons being made at each level. For the medium and high resolution images, three- and four-level pyramids are used respectively, resulting in a total of $10^2 \times 3 = 300$ and $24^2 \times 4 = 2304$ comparisons being made for the medium and high resolution images. Secondly, the pyramid algorithm uses a window that is centered horizontally, but is shifted down by setting the value of *xOffset* to 40% of the channels width. This modification is used here as a result of the empirical observation that the content many of the images is bottom heavy, with large areas of homogeneous texture or sky in their upper section, and more static information such as landscapes in the lower section. By setting *xOffset* to 40%, it is believed that greater amounts of pertinent visual information will be present in the window, while still allowing part of the window to cover the image center for images where the main content is centered in the image (Figure 3). The location of the reference window is initially offset horizontally and vertically from the targets location by $-\frac{1}{2} * numMoves$ pixels, and the window is slid down and to the right in a nested iterative manner. For example, with medium resolution images, when the algorithm begins, the reference window is displaced 5 pixels above and to the left of the target window. At each pyramid level starting with the coarsest, a rigid matching algorithm is used, such that a window is slid across the reference channel and compared to the target channels window. With each iteration, the SSIM value that is calculated for the two windows is compared against the best alignment previously seen, and if the current location produces a better similarity value, this value is saved for future comparisons. Additionally, the x and y values of the location where the target and reference most closely match are saved.

After all iterations are completed for a given pyramid level, the original resolution reference image is shifted to the position where the maximal similarity value was achieved, multiplied by a scaling factor representing the current pyramid level. For example, if at the third pyramid level, the maximal similarity was achieved by shifting the reference image by $x =$ -3 pixels horizontally and $y =$ 1 pixels vertically, the full scale reference image would be recreated from the original three-frame image by shifting its position by $x \times 2^{pyramidLevel-1} =$ -12 pixels horizontally, and 4 pixels vertically (recalling that levels two and three have a resolution equal to $\frac{1}{2}$ and $\frac{1}{4}$ of the original image respectively). Some of the high resolutions require a channel to be displaced 150 pixels or more, and the pyramid algorithm enables this with significantly fewer comparisons by allowing a channel to be shifted by a total of $\sum_p \frac{1}{2} \times numMoves \times 2^{pyramidLevel-1}$ where p is the number of pyramid levels. Thus, a channel can be shifted a maximum of 620 pixels in high resolution images and 70 pixels in medium resolution images. After the blue (reference) and green (target) channels have been registered, the above algorithm is similarly applied to register the green (reference) and red (target) channels. The benefit of registering the red channel to the green is that some three-frame positives are laterally skewed, and therefore the horizontal displacement between the green and red frames is less than between the blue and red. Additionally, there is typically less contrast difference between the green and red channels than between the blue and red. Finally, a new RGB composite image is created from the shifted blue, green and red channels.

4 Experimental Work

4.1 Data Set

The data set used in this paper consists of 1918 images in total, consisting of all 1902 medium resolution images, and 16 high resolution images. The medium resolution images are approximately 400×1024 8-bit images (0.5Mp), whilst the higher resolution images are approximately 3800×9800, 16-bit images (37Mp). The high resolution images were chosen based on their use in a similar project at Carnegie-Mellon [6], and the potential difficulty for registration due to lighting changes and large areas of homogeneous texture. The high resolution photographs used are: 00033u, 00066u, 00120u, 00139u, 00147u, 00152u, 00153u, 00225u, 00237u, 00252u, 00264u, 00271u, 00794u, 00797u, 01443u, and 01754u.

4.2 Results

Using the algorithms described, all 1902 medium resolution images and 16 high resolution images were registered. Two colour composites are shown in Figures 5 and 6. Evaluation of registration accuracy is achieved through visual inspection, as misregistration manifests itself in the form of an *anaglyph* image, with a 3D effect. The vast majority of the medium resolution images were correctly registered using the SSIM metric, with only 23 images (1.2% of the data set) being grossly misaligned. These misaligned images are listed in Table 1, along with reasons for the incorrect alignments. The primary cause of the misaligned images is subject movement or homogeneous textures in the registration window. For images where water, sky or vegetation in contained in the comparison window, there is often a substantial difference between exposures, due to changes in the environment. Prokudin-Gorskiis original glass plates had much less silver in the emulsion than

Fig. 5. Registration example (Image 00152): Alim Khan, Emir of Bukhara, 1911

Fig. 6. Registration example: Mills in Ialutorovsk district of Tobolsk Province, 1912 (Image 00859)

would be used today and thus the photographs needed to be exposed for a significant period, during which the scene may change due to trees, water, smoke, people or clouds moving in the wind. In some images, this effect is quite pronounced, such as the clouds in image 01275, which is otherwise correctly aligned (Figure 7). Some other examples showed only minor misalignments, in regions representing less than 1% of the images area. An example is shown in Figure 8, where a boat is moving across the waters surface during successive plate exposure. This misalignment represents only 0.05% of the images total area.

For the medium resolution images, the pyramid algorithm had a mean time of 3.57061 seconds. For the large images, the pyramid algorithm had a mean time of 442.626 seconds.

Table 1. List of images that were misaligned by the multiresolution SSIM algorithm

Cause	Image Number
Broken slide	00060 01706
Moving water	00097 00335 00336 00793 00882 01098 01423
Moving sky	00243 00302 01275
Moving vegetation	00965 01089 01206 01798
Homogeneous texture	00536 00537 01895
Unknown	00190 01803 01873

Fig. 7. Example of clouds moving between exposures (Image 01275)

Fig. 8. Example of a boat moving between exposures (Image 01098)

Fig. 9. The image is correctly aligned in the center and lower right, but is not correctly registered in the upper left due to the glass plate shifting during exposure (Image 00066)

5 Conclusion

This article has introduced an algorithm to register the three-frame positives from the Prokudin-Gorskii collection using a multiresolution SSIM algorithm. All 1902 medium resolution images were registered, along with 16 high resolution images. The majority of the resulting full color composites are correctly aligned. Although the algorithm presented here was shown to generate well registered images, there are several directions that future research may take. First, the SSIM metric is computationally expensive, and needs to be optimized in order for the metric to be more usable for this problem. One example, might be the use of SSIM in the complex wavelet domain [7]. An open problem is to evaluate the effectiveness and speed of other algorithms for registering the Prokudin-Gorskii images.

Second, future work may approach registration using a local search algorithm to find the optimal shift more quickly than the algorithm presented here. The search domain for registering these images can be viewed in three dimensions, where the x and y axes represent the horizontal and vertical displacement of the reference image, and the z axis represents the similarity value between the reference and target channels. During experimentation, it was observed that the search space has many local maxima when using the SSIM metric, and thus local search techniques that attempt to escape local maxima such as stochastic hill climbing, simulated annealing and random walks may be more efficient. Additionally, these approaches might be used within a Laplacian pyramid to further improve computation time.

Third, when the original photographs were taken, the glass plate often physically shifted in one or more of the three dimensions, causing the scene to be skewed or distorted between frames. The algorithm presented in this article only performs rigid alignment, which results in composite images that are aligned in one area, but may be misaligned in another due to shifts of the glass plate during exposure (e.g. Figure 9). A solution to this problem might employ an algorithm that uses non-rigid transformations to allow local warping of the image features to compensate for local deformations.

References

1. Library of Congress Prokudin Gorskii Collection, http://www.loc.gov/exhibits/empire/
2. Adamson, J., Zinkham, H.: The Prokudin-Gorskii legacy: Color photographs of the Russian Empire 1905-1915. Comma 3, 107–144 (2002)
3. Agüera y Arcas, B. Reconstructing Prokudin-Gorskii's Color Photography in Software, http://memory.loc.gov/pp/prokhtml/prokcompos.html
4. Burt, P.J., Adelson, E.H.: The Laplacian pyramid as a compact image code. IEEE Transactions on Communications 31(4), 532–540 (1983)
5. Dellaert, F.: Prokudin-Gorskii Images (2001), http://www.cs.cmu.edu/~dellaert/aligned/
6. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing 13(4), 600–612 (2004)
7. Wang, Z., Simoncelli, E.P.: Translation insensitive image similarity in complex wavelet domain. In: IEEE Int. Conference on Acoustics, Speech, and Signal Processing, pp. 573–576 (2005)