Motion Compensated Restoration of Colonoscopy Images

Nidhal Azawi and John Gauch

Department of Computer Science and Computer Engineering,
University of Arkansas, Fayetteville, Arkansas

ABSTRACT

Colonoscopy examinations are widely used for detecting colon cancer and many other colon abnormalities. Unfortunately, the resulting colon videos often have artifacts caused by camera motion and specular highlights caused by light reflections from the wet colon surface. To address these problems, we have developed a method for motion compensated colonoscopy image restoration. Our approach utilizes RANSAC-based image registration to align sequences of N consecutive images in the colonoscopy video and restores each frame of the video using information from these aligned images. We compare image alignment quality when N adjacent images are registered to each other versus registering images with larger step sizes between them. Three types of image pre processing were evaluated in our work. We found that the removal of non-informative images prior to image registration produced better alignment results and reduced processing time. We also evaluated the effects of image smoothing and resizing as a pre processing step for image registration.

KEYWORDS

RANSAC, Image alignment, Informative images, Non informative images, motion compensation, Colonoscopy.

1. INTRODUCTION

Image registration/alignment has been used in a wide range of image processing, computer vision and pattern recognition applications, including panorama creation, motion estimation, object recognition, and multi-sensor data fusion. For this reason, a lot of work has been done to develop fast and efficient image alignment methods.

The process of overlying two or more images by matching common features identified in the images using some methods is called image registration [1]. These images can be taken at different times may be taken in different angles or different camera/devices. Image registration-based feature matching involves feature detection and extraction, feature matching, transformation and fitting function, and image resampling and transformation.

Another image registration definition is illustrated by [2] who stated that image registration/aligning is the procedure to align two or more images after determining the optimal transformation that can fit or give the best transformation for a particular input image. Image registration also called image fusion, warping or matching. Registering two or more images helps

to combine information from multiple images. Image registration helps to integrate information for more than one image which are taken from different viewpoints, different angles, different times or different sensors. Therefore, it is very important step in image or video analysis

Image alignment methods can be classified into two broad categories based on how images are aligned with each other. The first category is area-based matching. Here, patches from one image are compared to patches in another image at different offsets to determine the (dx,dy) motion of the patch from frame to frame. A large number of methods have been devised for comparing patches, and for searching for (dx,dy) displacements with different accuracy/speed trade-offs. Recent examples of area-based approaches are described by [3], [4], [5] and [6].

The second category of image alignment methods is based on feature matching. Here, each image is examined to find feature points based on some search criteria, and the neighbourhood around each feature point is used to create a feature vector that can be matched against feature vectors from another image. Feature points are typically found by calculating geometric properties in an image, and detecting visually interesting points like corners, centres of bright/dark objects. Feature descriptors are chosen so they describe the local neighbourhood of feature points in a way that is robust to changes in position, orientation, scale, and illumination.

The scale invariant feature transform (SIFT) is one of the most widely used feature-based image alignment techniques [7]. Other recent examples of this approach include [8] and [9]. A hybrid approach using both area and feature matching was developed in [10].

Another proposed scheme for alignment of differently exposed images is by [11]. The proposed method consists of two stages. First, directional mapping to normalize images and to mitigate the effect of saturation has been implemented. Second, intensity invariant features have been represented using LBP (a non-parametric local binary pattern). The experimental results showed that their method achieved better accuracy than the state-of-the-art methods.

An efficient and robust method has been done in image alignment based on matching of relative gradient map. The match of the relative gradient feature from the training dataset has been used to find some candidate poses of the pattern from image. An iterative energy minimization approach is used to verify the candidate images. The authors show this approach is robust against non-uniform illumination [12].

Another approach is use viewpoint invariant patches (VIP) in the alignment of scenes and images especially if there are images that are seen or captured from different viewpoints [13]. VIP consists of features that are uniquely finds the matching transformation between 3D sciences. Features vector of VIP contains some invariant features such as 3D position, local gradient orientation in the patch plane, SIFT descriptors, and surface normal and the patch scale. The authors claim that their method able to distinguish between square and rectangle while affine invariant approaches could not recognize them. The proposed method rectified the image texture with respect to the geometry locality of the science. Ortho-texture (viewpoint independent of 3D science) can be seen using rectification.

Image registration in medical image analysis include applications of image registration to integrate information from computed tomography (CT), magnetic resonance imaging (MRI), single photon emission computed tomography (SPECT). Application areas include computer aided diagnosis, surgery simulation, intervention and treatment planning, radiation therapy,

anatomy segmentation, computational model building and image subtraction for contrast enhanced based approach, correction of scatter attenuation, partial volume corrections based on CT images, and assisted/guided surgery. Medical image registration has been applied on a wide range of body images such as brain [14], [15], [16], heart [17], breast, bones, wrist, entire body, liver, kidney, spine, knee, analysis of heart motion detection and many others [2].

In this paper, we address the problem of colonoscopy image registration. The proposed approach relies on three pre-processing steps, namely the removal of non-informative images, image resizing, median and mean filtering with or without image resizing. To the best of our knowledge we are the first researcher who tested these three pre-processing steps in image registration for colonoscopy images. By creating an image panorama from registered images, we are able to restore and enhance image details in colonoscopy images. The experimental set up shows that the removal of non-informative images allows enhancing the alignment results.

2. OUR APPROACH

Image alignment is an important component of our research. This is a very challenging task because we are dealing with colonoscopy images taken with a moving camera with significant changes in illumination and a number of images artefacts. It should be possible to align sequences of colonoscopy images with gradual changes in viewpoint, but it may not be possible to align very long sequences of images or sequences with rapid motion to each other. Our work will try to overcome these difficulties by preprocessing the colonoscopy video to identify and remove non-informative images from the input prior to registration (see figure1). The method we use to find and remove bad images is based on feature-based image classification described in our earlier paper [18].

To register sequences of N colonoscopy images to each other we used RANSAC (random sample consensus) to solve for the projective transformation that produces the best image alignment. RANSAC is a widely used for fitting models to some data in the presence of outliers. As the name suggest, this approach uses trial and error approach to find model parameters that best fit the data. The RASAC algorithm works as follows [19].

Let X represent the set of experimental data points we wish to model, we choose S1 points from X at random and build the parametric model through these points. To evaluate this model, we check the error tolerance for other points in X to find the subset that are less than distance D from the model. We call this the consensus set S*. The goal of RASAC is to find the parameters with the largest size consensus set S*, so we repeat this process until we find S* with more than V members or until a pre-determined number of random trials T has been performed. The speed and accuracy of RANSAC is controlled by three parameters, the distance threshold D, the target consensus set size V, and the maximum number of trials T.

They key to effective image registration is finding corresponding points in adjacent images. We do this by extracting a collection of feature points from each image and match their corresponding feature vectors to identify potential point correspondences. We performed RANSAC based image registration with four different types of image features (SURF, BRISK, FAST, and HARRIS) and our experiments show that SURF provides the best registration accuracy for our colonoscopy images [18].
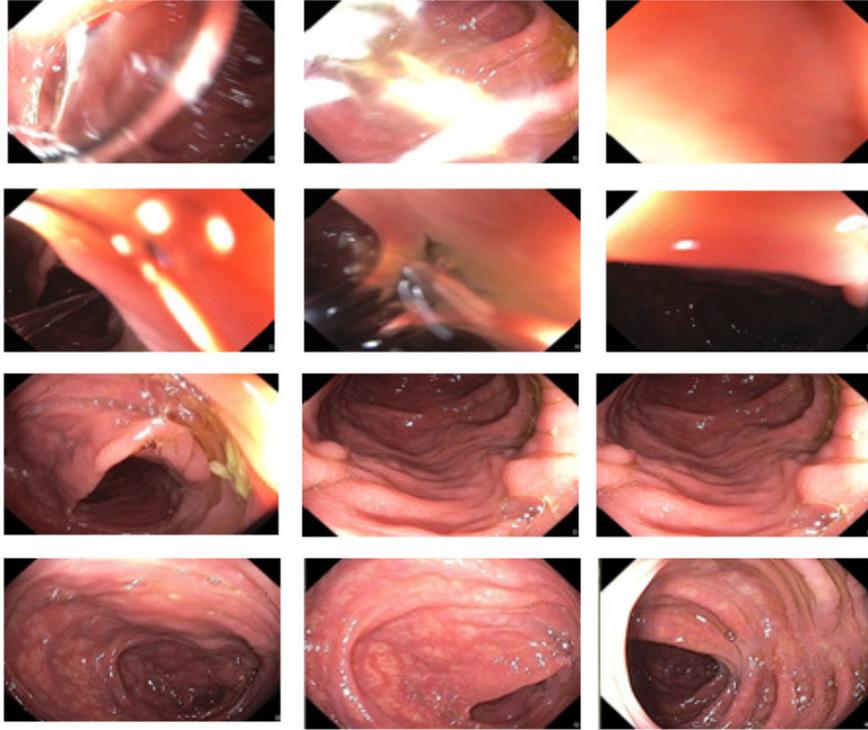
Figure 1. An illustration typical colonoscopy images. The top six images are non-informative because they are very blurred or have large specular highlights. The bottom six images are informative views of the colon, with small specular highlights and little blurring.

We evaluated RANSAC based image registration with affine and projective transformations. Affine transformations capture translation, rotation, scaling and sheer between consecutive images, while projective transformations also capture changes due to changes in viewpoint. Affine transformations preserve parallelism while projective transformations do not. Affine transformation can be defined in terms of the motion of vertices of a triangle while projective transformation is defined by the transformation of quadrangle vertices. The affine transform equation is $\overline{A \cdot H_A = C}$ where

$$
A = \begin{pmatrix} x1 & y1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x1 & y1 & 1 \\ x2 & y2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x2 & y2 & 1 \\ x3 & y3 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x3 & y3 & 1 \end{pmatrix} \quad H_A = \begin{pmatrix} hf1 \\ hf2 \\ hf3 \\ hf4 \\ hf5 \\ hf6 \end{pmatrix} \quad C = \begin{pmatrix} X1 \\ Y1 \\ X2 \\ Y2 \\ X3 \\ Y3 \end{pmatrix}
$$

The affine homography matrix is represented as a vector called H_A that contains six degrees of freedom (DOFs). Hence, the minimum number of points needed to solve for homography is three

matching points (x1,y1), (x2,y2) and (x3,y3). These three matching points are combined into matrix A

The projective transform is given by $\overline{A \cdot H_p = C}$ where

$$A = \begin{pmatrix} x1 & y1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x1 & y1 & 1 \\ x2 & y2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x2 & y2 & 1 \\ x3 & y3 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x3 & y3 & 1 \\ x4 & y4 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x4 & y4 & 1 \end{pmatrix} \quad H_P = \begin{pmatrix} hp1 \\ hp2 \\ hp3 \\ hp4 \\ hp5 \\ hp6 \\ hp7 \\ hp8 \end{pmatrix} \quad C = \begin{pmatrix} X1 \\ Y2 \\ X3 \\ Y4 \\ X5 \\ Y6 \\ X7 \\ Y8 \end{pmatrix}$$

The projective homography matrix is represented as a vector is called HP which has eight DOF and for that reason the minimum number of points required to solve for homography is four points. These four matching points (x1,y1), (x2,y2), (x3,y3) and (x4,y4) are gathered into tow dimensional matrix A. The coordinate points (Xi,Yi) for affine or projective transformation can be calculated by multiplying the matching points A by the corresponding homography matrix [20] [21].

Using RANSAC to align colonoscopy images with affine transformations yields a large number of non singular transformation matrices, which means there is no viable affine transformation that can successfully align these two images. Hence affine transformations are not a good choice for image registration. Since the camera capturing colonoscopy video is changing position during the procedure, we will use RANSAC to calculate the best projective transformation that aligns all pairs of images within a moving 10 frame window of the colonoscopy image. The algorithm we use to register, and process colonoscopy images has the following steps:

- Loop over all sets of 10 consecutive images in the colonoscopy video.
- Detect and extract features points for all 10 images in the sequence.
- Find the matching feature points for all pairs of images im1 and im2.
- Determine the best projective transformation using RANSAC algorithm.
- Exclude all transforms that fail any condition below:
    o The number of inlier points less than 5 points.
    o The determinant of the transform less than 0.5.
    o The image difference after alignment is less than before alignment.
- Save aligned images in an output directory and create image panorama.

3. EXPERIMENTAL RESULTS

To evaluate the effectiveness of this image registration algorithm on colonoscopy images, we performed a number of experiments using a collection of 1000 typical colonoscopy images. These images have been automatically classified as being informative or non-informative using feature-based image classification [18].

3.1 Evaluation Metrics

- In each of our experiments, we considered the following four evaluation metrics. Three of these measures are objective, while one is subjective and depends on the viewer's requirements.
- The alignment error is calculated as the average RMSE between pairs of images after alignment for all images that are successfully aligned.

$$RMSE = \sqrt{\sqrt{\frac{\sum_{i=1}^{H}\sum_{j=1}^{W}[im1(i,j) - T(im2(i,j)]^2}{H \cdot W}}}$$

where T(im2(i,j)) is image im2(i,j) after it has been transformed by the optimal projective transformation T to align with image im1(x,y).
- The percentage aligned is the percentage of image pairs with valid projective transformations out of the total number of images in the sequence.
- The average computation time for aligning images in the colonoscopy video.
- The visual quality of the panorama image generated from the aligned images compared to the original images in the colonoscopy video. Panorama images that have specular highlights removed and/or have improved image detail would be considered high quality, and panoramas that are highly distorted would be considered low quality.

3.2 Parameter Selection

We performed RANSAC based image registration with four different types of image features (SURF, BRISK, FAST, and HARRIS) and our experiments show that SURF provides the best registration accuracy for our colonoscopy images [18]. This image registration algorithm has several parameters that control the accuracy and speed of colonoscopy image alignment.

The identification of SURF feature points is controlled by a metric threshold. As this threshold is decreased more SURF feature points are detected. We experimented with a range of metric thresholds between [0..1000] and had the best alignment results with a metric threshold of 100. The number random trials used by RANSAC to find the optimal transformation effects the speed and accuracy of the results. As the number of trials increases, the quality of the alignment improves, but the computational cost increases. We experimented with a range of values between [400..3000] and selected 2500 to align images in reasonable time.

After choosing the metric threshold and the number of trials, we conducted experiments to evaluate two pre processing operations, median filtering and image resizing.

To smooth these images to remove noise, we performed median filtering with a 10x10 mask. The root mean square image alignment error (RMSE) when median filtering was used was 3.52 for the 150 images we aligned. The percentage aligned after median filtering was 30% for this group of images. The RSME without median filtering was slightly lower at 3.44 and the alignment percentage increased to 43%. Median filtering reduces the number of matching points which in turn reduces the number of frames that can be successfully aligned.

The images we extracted from our colonoscopy video were 1347x540 pixels. We experimented with image resizing prior to image alignment with a scale factor of 0.5 (673x270) and with a scale factor of 0.25 (336x135). Unfortunately, these resized images were too small for our algorithm to find enough matching points to successfully align any images. Median filtering before or after image resizing did not improve these results, so we will use our original images in our subsequent image alignment experiments.

3.3 Pairwise Image Alignment

Our first experiment performed pairwise image alignment with 1000 adjacent colonoscopy images. For each pair of images, we calculated the projective transformation using RANSAC that provided the best image alignment. Our experiments show that the average RMSE for the 1000 images was equal to 8.85. This alignment error was reduced to 7.8 when the non-informative images were removed from the input sequence prior to alignment. Similarly, the percentage successfully aligned for the full video sequence was 61.5%. This was increased substantially to 80.6% when non-informative images were omitted from the input sequence. These improvements in alignment error and percentage aligned are to be expected because the non-informative images are so highly distorted [18].

Once pairwise alignments have been calculated, it is possible to partition the 1000 images into aligned sequences by connecting adjacent images that are successfully aligned to each other. In our case, this resulted in 22 sequences of images that varied in length from [2..385] images. Once we have calculated this partition of the colonoscopy video into separate sequences, we can focus our image restoration and display efforts on these sequences. For example, we can recreate 22 video clips that contain only the informative images, or in some cases we can create a panorama image using these images (see figure 2).
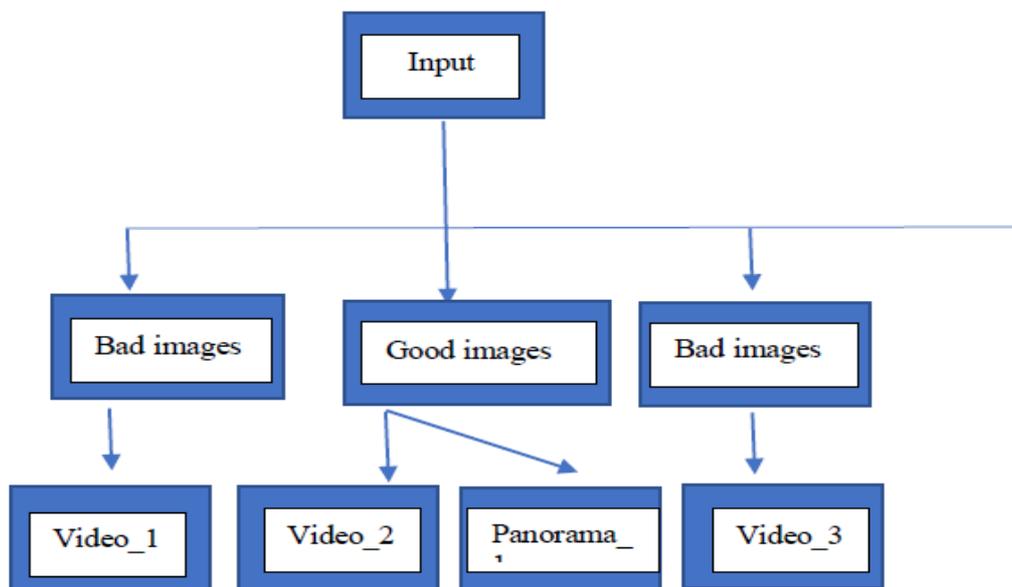
Figure 2. Visualization structure for pair image alignment. Once the input video has been classified into informative (good) images and non-informative (bad) images, we have the option of creating video clips or panoramas from the good images for each aligned sequence of images.

3.4 **Sequence Image Alignment**

Our second experiment, we performed image sequence alignment as a pre processing step to image panorama creation. For each frame in the colonoscopy video, we calculated the best alignment with the 10 subsequent images. Figure 3 shows that in some cases only a subset of the 10 subsequent images were able to be successfully aligned with the starting image.
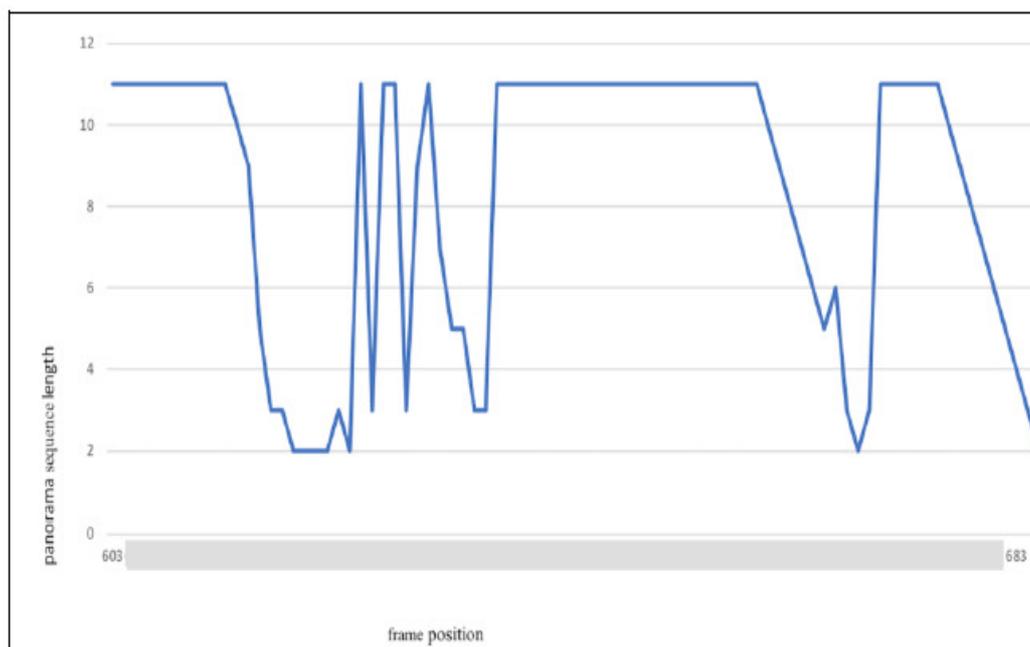

Figure 3. Plot of panorama sequence length for colonoscopy frames 603 to 683. Notice that the sequence length ranges from 11 frames down to only 2 frames. This is because some portions of the colonoscopy video have high motion or contain non-informative images.

When we evaluated our image alignment results for the entire 1000 frame sequence, we had a RMSE of 4.16 and a percentage aligned of 37%. When we ran the experiment again, excluding the non-informative images, the RMSE increased slightly to 4.38 and the percentage aligned increased significantly to 48%.

When we compare our image sequence alignment results to our pairwise image alignment results we can see that the percentage aligned is much lower for sequence alignment. This is because we are attempting to align images that are more than one frame apart from each other in time, so there has been more motion, and it becomes more difficult to find and match image features. Consequently, it is harder to successfully align images as the sequence length increases.

The average CPU time for image alignment and panorama creation was also significantly reduced from 15.7 seconds for the original video down to 5.1 seconds when non-informative images were excluded. This large change in CPU time can be explained by looking at the RANSAC image alignment process. When two images can be successfully aligned, the algorithm converges before the maximum number of iterations is reached. When two images can not be aligned, RANSAC

will attempt the maximum number of iterations before failing. Therefore, attempting to align non-informative images to other images wastes a lot of CPU time.

The results from our image sequence alignment and panorama creation are illustrated in the figures below. In figure 4 and figure 5, we show how some subsequences produce very good panoramas. In figure 6 and figure 7, we show how non-informative subsequences produce very poor panoramas that has no useful information.
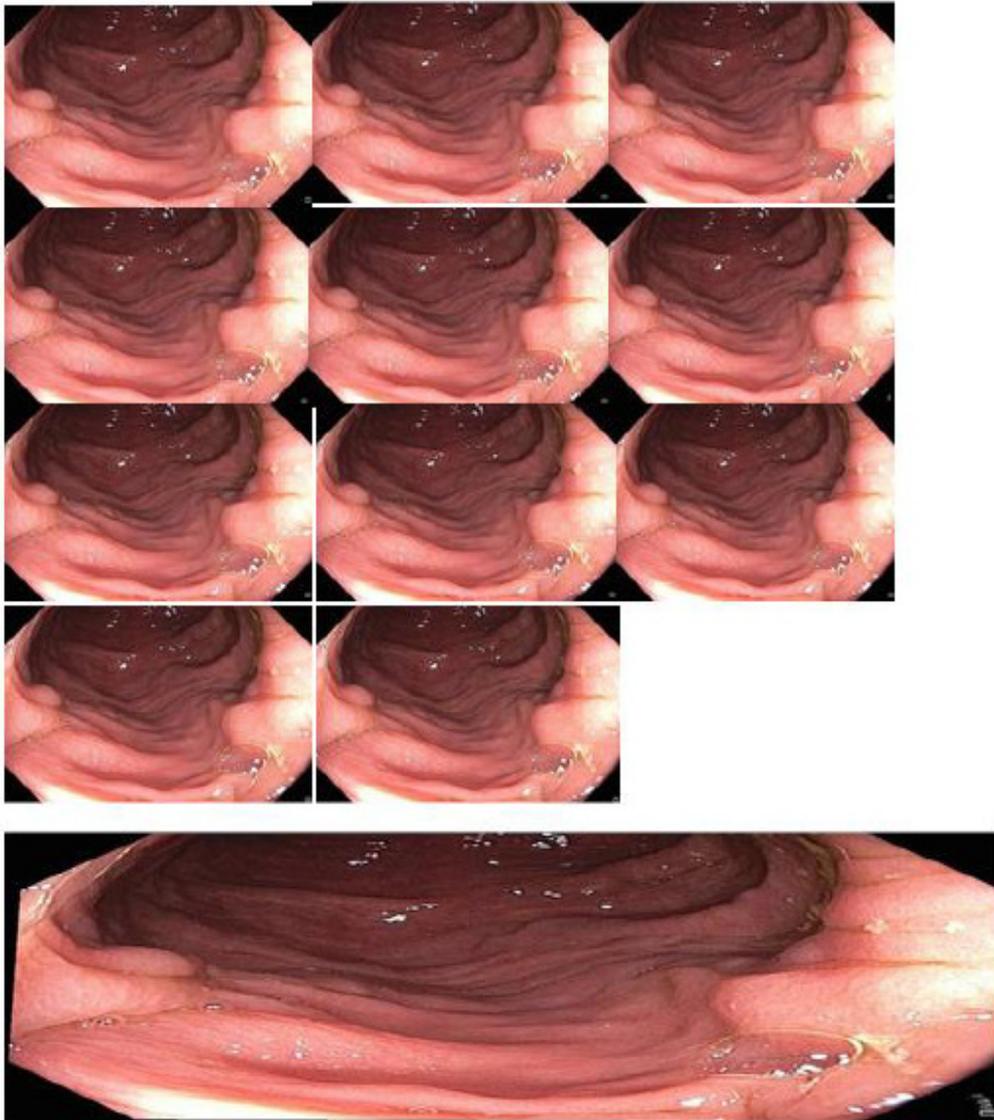

Figure 4. An illustration of image sequence alignment showing eleven consecutive input images and the resulting image panorama. Notice that the panorama includes additional information on the left and right sides of the first image in the sequence.
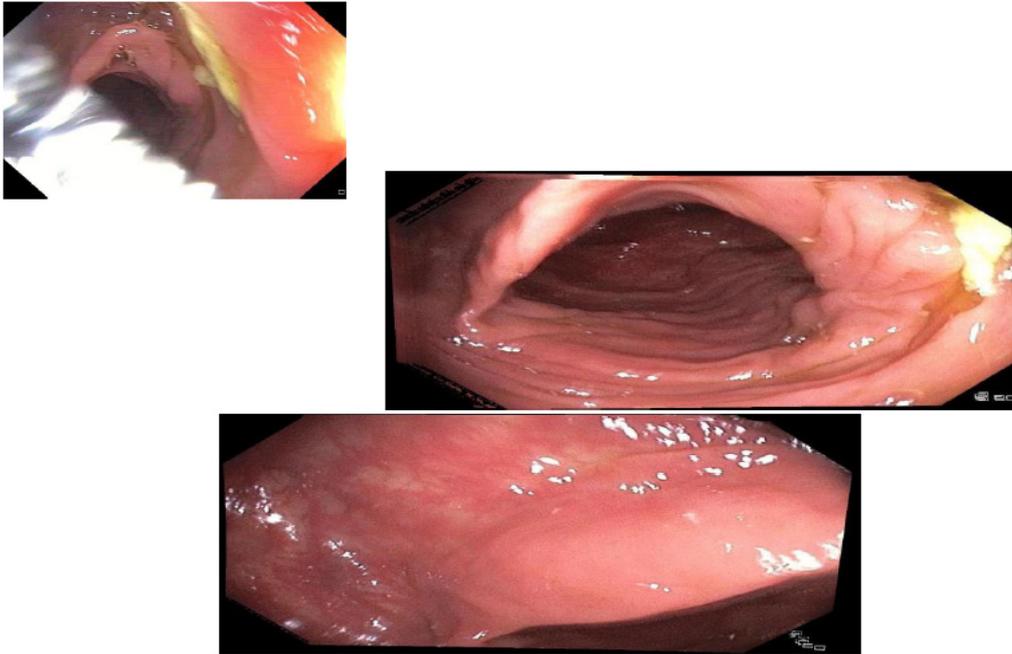
Figure 5. An illustration of two additional image panoramas. In both cases, the width of the panorama is larger than the first frame in the image sequence and include more information about adjacent colon features.
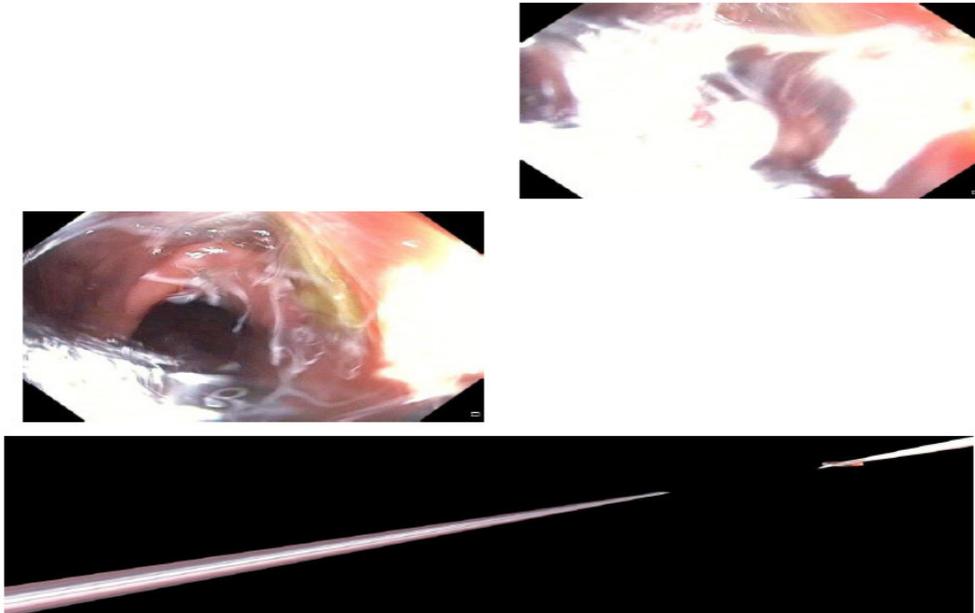

Figure 6. An illustration of unsuccessful panorama creation with non-informative images. The three colonoscopy images above were incorrectly aligned to each other by our RANSAC method. This produces singular transformations, and a highly distorted image panorama.
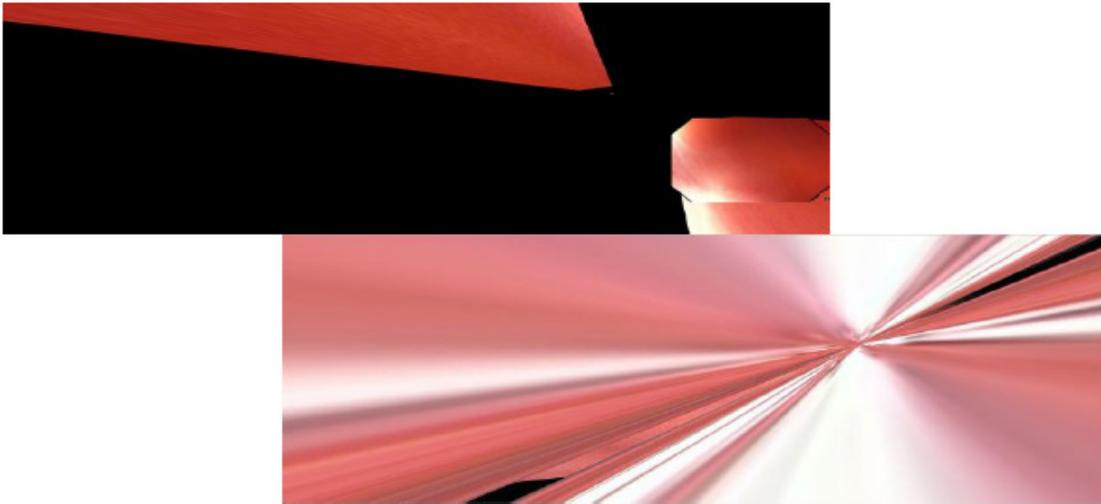
Figure 7. More examples of unsuccessful panorama creation. The two panoramas above were produced by aligning sequences that contained one or more non-informative images, which yielded singular transformations, and highly distorted image panoramas.
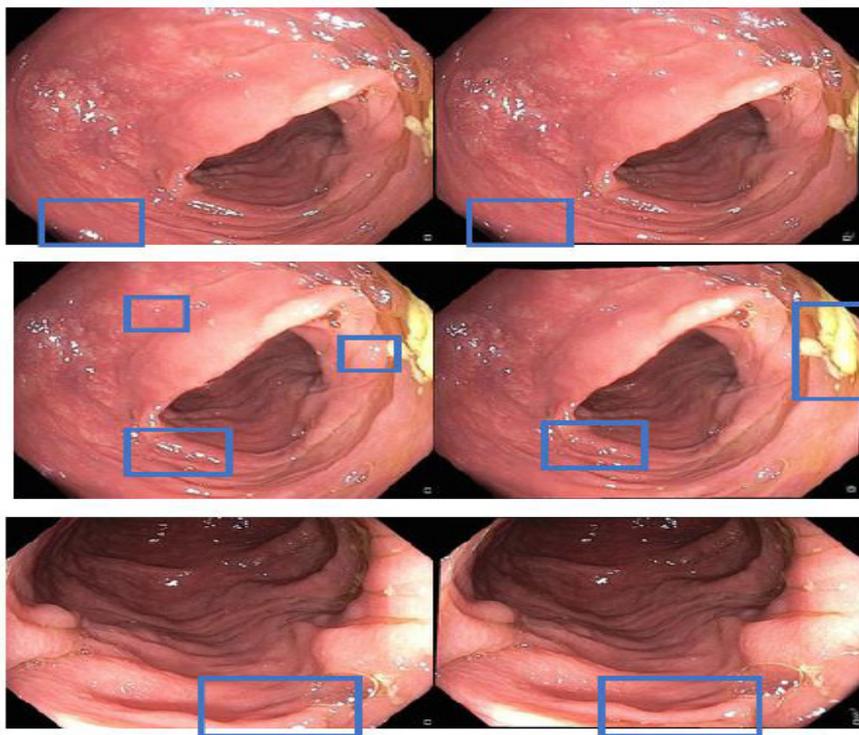

Figure 8. Three examples of panoramas that were created with image sequences that had zooming out motions. The images on the left are the original colonoscopy images, and the images on the right are the corresponding panoramas. Blue boxes indicate areas of where specular highlights have been removed and where more image detail is visible.

One benefit of our image sequence alignment is that it restores some important details in our colonoscopy images. This can be seen in figure 8 which shows some original images and enhanced versions that have been generated using our approach. Notice that some specular noise has been removed, also some image details have been added that are indicated using blue rectangles.

4. CONCLUSIONS AND FUTURE WORK

In this paper we described our method for motion compensated colonoscopy image restoration. As a first step we perform RANSAC-based image registration to align sequences of N consecutive images in the colonoscopy video. We then use this sequence to construct panorama images that improve image quality. We have demonstrated that this approach successfully removes unwanted specular highlights from colonoscopy images, and in many cases adds details that are not present in the original image.

Our experiments verify that the removal of non-informative images prior to image registration reduces the CPU time necessary for image alignment. This is because the RANSAC algorithm executes the maximum number of iterations without finding a good alignment transform for non-informative images. We also experimented with different sequence lengths and found that sequences of 11 consecutive images provided a good trade-off between CPU time and panorama quality.

For future work, we will focus on improving the quality of image alignment using different image registration techniques. In addition, we will explore methods to reduce the CPU time needed to perform this image restoration. By combining image alignment transformations from frame, A to B and from frame B to C, we should be able to get better estimates of the transformation from A to C. Since we are performing many independent image alignment operations, CPU time can also be reduced using parallel programming on a cluster or using GPUs.

REFERENCES

[1] L. Dung, C. Huang, and Y. Wu, "Implementation of RANSAC Algorithm for Feature-Based Image Registration," Journal of Computer and Communications, pp. 46–50, 2013.

[2] F.P.M. Oliveira, J.M.R.S. Tavares. Medical Image Registration: a Review. Computer Methods in Biomechanics and Biomedical Engineering 17(2):73-93, 2014.

[3] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High dynamic range video," ACM Trans. Graph., vol. 23, no. 3, pp. 319–325, 2003.

[4] F. M. Candocia, "On the Featureless Registration of Differently Exposed Images," in Proc. Int. Conf. Imaging Science, Systems & Technology, Las Vegas, NV, USA, Jun. 2003, vol. I, pp. 163–169.

[5] Hossain and B. K. Gunturk, "High Dynamic Range Imaging of Non-Static Scenes," in Proc. SPIE Digital Photography VII, 2011, vol. vol. 7876.

[6] H. Q. Luong, B. Goossens, A. Pizurica, and W. Philips, "Joint photometric and geometric image registration in the total least square sense," Pattern. Recognition. Lett., vol. 32, no. 15, pp. 2061–2067, 2011.

[7]   D. G. Lowe, "Object recognition from local scale-invariant features," Proc. of the Int. Conf. on Computer Vision, pp. 1150–1157, 1999.

[8]   B. Zitova and J. Flusser, "Image registration methods: A survey," Image Vis. Comput., vol. 21, pp. 977–1000, 2003.

[9]   S. Oldridge, G. Miller, and S. Fels, "Mapping the problem space of image registration," in Proc. Can. Conf. Computer and Robot Vision, St. John's, NF, Canada, May 2011, pp. 309–315.

[10] M. Tico and K. Pulli, "Robust image registration for multi-frame mobile applications," in Proc. Asilomar Conf. Signals, Systems & Computers, Pacific Grove, CA, USA, 2010, pp. 860–864.

[11] S. Wu, Z. Li, J. Zheng, and Z. Zhu, "Exposure-robust alignment of differently exposed images," IEEE Signal Process. Lett., vol. 21, no. 7, pp. 885–889, 2014.

[12] S. Wei and S. Lai, "Robust and efficient image alignment based on relative gradient matching," IEEE Trans. image Process., vol. 15, no. 10, pp. 2936–43, 2006.

[13] C. Wu, B. Clipp, X. Li, J. M. Frahm, and M. Pollefeys, "3D model matching with viewpoint-invariant patches (VIP)," 26th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR, pp. 1–8, 2008.

[14] P. A. Freeborough and N. C. Fox, "Modelling Brain Deformations in Alzheimer Disease by Fluid Registration of Serial 3D MR Images", vol. 22. 1998.

[15] D. Leow, A. D. Klunder, C. R. Jack, A. W. Toga, A. M. Dale, M. A. Bernstein, P. J. Britson, J. L. Gunter, C. P. Ward, J. L. Whitwell, B. J. Borowski, A. S. Fleisher, N. C. Fox, D. Harvey, J. Kornak, N. Schuff, C. Studholme, G. E. Alexander, M. W. Weiner, and P. M. Thompson, "Longitudinal stability of MRI for mapping brain change using tensor-based morphometry," Neuroimage, vol. 31, no. 2, pp. 627–640, 2006.

[16] K. A. Ganser, H. Dickhaus, R. Metzner, and C. R. Wirtz, "A deformable digital brain atlas system according to Talaicrach and Tournoux," Med. Image Anal., vol. 8, no. 1, pp. 3–22, 2004.

[17] X. Huang, J. Ren, G. Guiraudon, D. Boughner and T. M. Peters, "Rapid Dynamic Image Registration of the Beating Heart for Diagnosis and Surgical Navigation," in IEEE Transactions on Medical Imaging, vol. 28, no. 11, pp. 1802-1814, Nov. 2009. doi: 10.1109/TMI.2009.2024684.

[18] N. Azawi, J. Gauch, "Automatic Method for Classification of Informative and Noninformative Images in Colonoscopy Video", Int. Conf. on Medical Image Processing and Analysis (ICMIPA), Vancouver, Canada, August 2018.

[19] M. A. Fischler and R. C. Bolles, "Paradigm for Model," vol. 24, no. 6, 1981.

[20] D. Wierzbicki, "Multi-Camera Imaging System for UAV Photogrammetry," Sensors (Basel, Switzerland) vol. 18,8 2433. 26 Jul. 2018, doi:10.3390/s18082433.

[21] R. Redzuwan, N. A. M. Radzi, N. M. Din, and I. S. Mustafa, "Affine versus projective transformation for SIFT and RANSAC image matching methods," 2015 IEEE Int. Conf. Signal Image Process. Appl., pp. 447–451, 2015.

AUTHORS

Nidhal K. Azawi finished her bachelor and master's degree in computer science / image processing at the University of Technology / department of computer science in Baghdad. She was an assistant teacher in the computer science and Engineering in the University of Al_Mustansiriya in Baghdad. She is now a PhD student researcher in computer science and engineering department in the University of Arkansas. Her research is in medical image processing. Her research interest seeks to combine machine learning with image processing and computer vision for advancements in the medical field pertaining to image enhancement, reconstruction, motion analysis, and visualization.

John M. Gauch joined the computer science and computer engineering department at the University of Arkansas as a professor in 2008. He held previous faculty positions at the University of Kansas and Northeastern University. His research interests include real-time digital video processing, content-based image and video retrieval, biomedical image enhancement, image segmentation, and motion tracking applications. This research has resulted in over sixty publications in these areas including one book and one patent.