A Multistage Approach for Image Registration

Francis Bowen, IEEE, Jianghai Hu, IEEE, and Eliza Du, Senior Member, IEEE,

Abstract—Successful image registration is an important step for object recognition, target detection, remote sensing, multi-modal content fusion, scene blending, as well as disaster assessment and management. The geometric and photometric variations between images adversely affect the ability for an algorithm to estimate the transformation parameters that relate the two images. Local deformations, lighting conditions, object obstructions, and perspective differences all contribute to the challenges faced by traditional registration techniques. In this work, a novel multistage registration approach is proposed that is resilient to view point differences, image content variations, and lighting conditions. Robust registration is realized through the utilization of a novel region descriptor which couples the spatial and texture characteristics of invariant feature points. The proposed region descriptor is exploited in a multistage approach. A multistage process allows the utilization of the graph-based descriptor in many scenarios thus allowing the algorithm to be applied to a broader set of images. Each successive stage of the registration technique is evaluated through an effective similarity metric which determines subsequent action. The registration of aerial and street view images from pre and post disaster provide strong evidence that the proposed method estimates more accurate global transformation parameters than traditional feature-based methods. Experimental results show the robustness and accuracy of the proposed multistage image registration methodology.

Index Terms—image registration, invariant feature point, region descriptors, SURF.

I. INTRODUCTION

For many image processing tasks, such as image fusion or stitching, image registration is often exploited as a preprocessing step [1-3]. Registration is a necessary task that estimates the transformation parameters relating two images. The purpose is to project one image such that both images contain some region which overlaps and may appear to be from the same perspective. This region of interest selects areas of the images that may share common details and features. Automated registration techniques are desirable [38].

Registration requirements may vary greatly amongst different applications. For instance, multi-modal medical registration will rely heavily upon shape context, while general object registration will necessitate the analysis of textural properties. Two matched images may vary geometrically through translation, rotation, affine, and perspective transformations, or photometric variations from object occlusion and pixel dissimilarities from lighting differences. Images of a scene before and after a disaster is particularly difficult to register due to many dissimilarities in geometric and photometric composition; however analysis of such images is an important goal for disaster management and risk planning [4,5].

Registration algorithms can be broadly classified as feature-based or intensity-based approaches. In the former, distinct image features such as corners [6], gradient edges [7], or shape descriptors [8] are used to define the geometrical mapping between images. Feature-based approaches rely on the detected feature’s invariance to affine, rotational, and translational transformations. With intensity-based methods, pixel intensities are compared directly for a sub-region of an image. In such a scenario, a similarity measure is coupled with an optimization algorithm in an attempt to identify the closest mapping of pixels [9].

Vast amounts of data captured before and after a disaster are often collected for future analysis. The ability to automate this process will lead to faster disaster management and response. Effective processing of such data may include image registration of two images with limited mutual information which is a challenging problem, as the images before and after a disaster often have serious local deformations.

The above processing approaches are not designed to deal with the large photometric and geometric variations that could simultaneously exist in an image set with scenes before and after a disaster. There are three challenges: 1) Large differences in pixel intensities or textural properties will adversely affect the ability to register two images as common features may be indistinguishable. Pixel intensities may vary due to differences in lighting conditions, as well as damage caused by natural disasters, such as from a fire or a flood. 2) Geometric variations are often present due to damage. Buildings that sustain large amounts of destruction may be unrecognizable after a disaster. Additionally, perspective differences when acquiring the images also add to the geometric differences between image pairs. And 3) aerial and satellite imagery is often acquired at different times and will therefore introduce rotational and scale differences.

Thomas et al. [10] proposed a registration approach that employed invariant feature points for disaster image alignment. However image content can vary tremendously where image gradients are skewed or completely different. Therefore a particular feature point may not be suitable for many scenarios. In [11] we proposed a graph-based region descriptor for accurately matching features across such image sets. It is the aim of this work to improve image registration in difficult situations through an innovative multi-stage approach that exploits a novel region descriptor as well as intensity-based analysis for a broad ability to register two images greatly affected by disasters. Such a method will be applicable to aerial and urban imaging where various levels of
transformations may invalidate other approaches. Many other applications could benefit from the proposed system where traditional registration methods have limitation when being applied to images with local geometric and photometric deformations.

In this research, we propose a comprehensive three-phase image registration method that takes advantage of feature point detection but imposes a strict method for identifying optimal interest points for the estimation of the homography matrix. Invariant feature points and their spatial relationships are leveraged to identify the suitable control points for the estimation of the transformation parameters required for accurate registration. A k-nearest neighbor graph constructed from a collection of Speeded-up Robust Feature (SURF) feature points is described along with the matching criteria for the region descriptor and subsequent control point identification. Our previous works in [11,16] demonstrate the effectiveness of the proposed region descriptor for image feature matching while this work extrapolates the core concepts to image registration. Finally, the discrete linear transform (DLT) approach is utilized for the transformation parameter estimation.

The multistage methodology exploits the invariant features in a variety of approaches to successfully register disaster images that differ by a multitude of transforms. The hierarchical process first attempts registration using the most computationally simple approach then evaluates the result. Each successive stage is attempted if the result of the previous stage is unsatisfactory. Moreover, the computational complexity increases with each successive stage, but successive stages offer greater resilience to transformations.

The analysis of pre and post disaster image sets is a challenging proposition due to the multitude of extreme transformations. Structural and textural variations coupled with perspective differences provide a challenge for image registration algorithms; therefore, the effectiveness of the proposed registration method is illustrated through a diverse assortment of disaster image scenarios from aerial perspectives as well as urban landscapes.

II. RELATED WORK

Image registration approaches can be classified as feature, intensity or Fourier-based. Feature-based techniques attempt to identify matching features across images in order to estimate an appropriate set of transformation parameters. Suitable features may be structural elements of a building [12], hard corners for general objects [6], or general shape characteristics of the image content [8]. Intensity-based methods compare pixel intensity distributions and underlying relationships. These techniques often rely upon an optimization algorithm such as the Particle Swarm Theory [13-14] or gradient descent [15]. Fourier-based approaches exploit the shift properties of a signal in the frequency domain. Translation, rotation and scale can be estimated by analyzing the normalized cross-correlation matrix that is easily computed from the Fourier transform [17-18]. Although feature and intensity-based categories are the most common, the third category, Fourier-based methods, offer additional information that is extracted from frequency-domain analysis. The Fourier approaches can be utilized to aid in feature or intensity-based algorithms.

A. Feature-based Methods

Invariant feature points such as the Scale Invariant Feature Transform (SIFT) and Speeded Up Robust Features (SURF) are common approaches for identifying registration control points [19-21]. The ubiquitous SURF approach has been used extensively to identify control points for homography estimation due to the feature point’s invariance to translation, rotation and scale.

In [22], Bay proposed the SURF descriptor as an alternative to the SIFT descriptor which aims to compute multi-scale feature points that are invariant to scale, rotation and translational deformations. The core operations of SURF that rely upon the integral image, such as convolution, are calculated in a fixed amount of time, greatly improving the computational efficiency. Interest points are discovered using a multi-scale Hessian detector where second order Gaussians are estimated using box filters. A descriptor is computed for each detected interest point utilizing the sum of the Haar wavelets in the x and y directions. Each vector is composed of components $\Sigma d_x$, $\Sigma d_y$, $\Sigma |d_x|$, and $\Sigma |d_y|$, where $d_x$ and $d_y$ are the responses in the x and y direction, respectively.

The effectiveness of feature-based approaches relies on the characteristics of the selected invariant feature points which include the discriminative nature of the feature points as well as the matching process. Common feature-based approaches such as SURF and SIFT-based methods are invariant to scale, translation, and rotation variations, but may not be discriminative enough to provide a one-to-one mapping for all feature points. This scenario is evident in situations with many redundant textural patterns.

B. Intensity-based Methods

In contrast to feature-based approaches, intensity-based image registration directly exploits the pixel intensities and their distributions. The same optimization techniques may still apply, such as the particle swarm theory; however, a new set of similarity measures must be employed. The functions attempt to relate the statistical models of the pixels.

Recent work in [23-25] provides examples of the statistical analysis required for medical image intensity-based registration. Entropy functions such as mutual information provide the basis for such approaches; however entropy-based metrics are adversely affected by noise and differences in image content.

Due to the statistical nature of intensity-based approaches, optimization algorithms are often coupled with entropy metrics. In [13] Wang et al. propose the use of the Particle Swarm Theory, while Shen et al. in [15] outline an intensity-based registration technique based on the gradient descent optimization algorithm and motion estimation. Similarly in [39], Qin et al. outline a method which matches image patches using statistical analysis and optimization techniques. Such scenarios pose the problem of incorrect convergence to local minima or maxima. Moreover, the computational latency of
such methods may limit the applications in which the intensity-based algorithm can be applied.

Intensity-based registration approaches are advantageous in natural scenes where lighting variations are minimal; however, these approaches are adversely affected by noise. The effectiveness of intensity-based methods is determined by the similarity metric used as well as the search window size. Furthermore, since typical approaches utilize a rectangular search window, intensity-based registration techniques may have difficulty in datasets that exhibit non-rigid transformations [37].

C. Fourier-based Methods

Fourier-based methods are a popular registration method where sub-pixel registration is not required [18,26]. Moreover, it is a common approach for use as an initial step prior to a more accurate registration method [17].

The Fourier Shift Theorem offers a valuable tool for determining the translation, rotation, and scale parameters that relate two images. The registration parameters may be recovered by identifying the peak within the cross-correlation matrix, which is estimated by computing the inverse Fourier Transform of the cross power spectrum of the two images.

Determination of the translation parameters that relate two images is often accomplished using the original reference and query images; however calculation of the rotation and scale parameters requires transferring the images to the log-polar domain. The Fourier-Mellin approach applies the log-polar transform to the frequency domain images while the approach outlined in [27-28] employs the log-polar transform in the spatial domain.

Fourier-based registration approaches suffer from similar drawbacks as intensity-based approaches where differences in image content or noise greatly reduce the method’s accuracy. Moreover, recovery of rotation parameters is not accurate for large angle differences.

Each of the three common registration techniques offers advantages under specific, yet different, scenarios; thus it is the intention of the proposed multistage approach to realize these advantages while exploiting a novel graph-based region descriptor. The aim of the hierarchal methodology is to successfully estimate registration parameters for a broad range of transformations in disaster images. Scenes involving disasters often entail a multitude of variations including photometric and geometric differences that pose a challenge to image registration techniques. A multistage approach will attempt to address as many such variations with a hybrid method involving feature, intensity and Fourier-based registration procedures.

III. PROPOSED GRAPH-BASED REGION DESCRIPTOR

In this research, we propose a novel region descriptor that couples spatial and textural characteristics of invariant feature points. This is accomplished by representing image features with a directed graph of clustered SURF feature points. The feature points are grouped by their textural and spatial characteristics which are then compared to each other through a fast graph comparison process. Feature point clustering will group similar feature points while a sub-graph matching scheme will identify similar graphs across images. The centroids of matched clusters are regarded as the control points which are utilized in image registration.

A. Invariant Feature Point Clustering

K-means clustering has long been exploited for many applications where clustering may be necessary [29-30]. In this work, k-means clustering is utilized to group feature points based on their spatial relationships. This will result in grouping together feature points into dense groups. For this work, k-means was chosen due to its simplicity and fast execution time. The clustering required for this application does not require better accuracy than what k-means can provide at the expense of system resources and runtime. The original method proposed by MacQueen [31] involves a three step process: initial cluster assignment, calculation of cluster center, and cluster reassignment. Steps 2 and 3 are repeated until the method converges to the point where clusters remain unaltered.

Given a set of n observations, \( S = \{x_d \mid 1 \leq d \leq n\} \), the aim of the algorithm is to produce \( k \) clusters, \( S = \{S_1 \cup ... \cup S_k\} \), where each cluster has an associated mean, \( \bar{m}_i, \ldots, \bar{m}_k \). Observations are then assigned to each cluster contingent upon the distance between the observation and the cluster means. Set \( S_j \) at the next iteration of \( m \) is defined as,

\[
S_j^{i+1} = \{x_d \mid \|x_d - \bar{m}_j\| < \|x_d - \bar{m}_p\| \} \forall p \neq j
\]  

(1)

The \( j \)-index represents the \( j \)-th cluster, while \( i \) denotes the iteration number, \( x_d \) is a sample from the population where \( 1 \leq d \leq n \), \( \bar{m}_j \) is the mean for the \( j \)-th cluster and \( \bar{m}_p \) is the mean from another cluster besides \( j \). The algorithm has converged when the following condition is reached.

\[
S_j^{i+1} = S_j^{i} \forall j
\]  

(2)

For the proposed region descriptor, the cluster centers are initially selected at random. For a set of feature points, \( S_{FP} \), the number of clusters is defined as,

\[
k = \left\lfloor \frac{|S_{FP}|}{5} \right\rfloor
\]  

(3)

The denominator is chosen such that the average number of nodes per graph is 5. The total number of feature points is divided by 5 to provide the initial number of clusters for the k-means algorithm. Convergence within a certain number of iterations is not guaranteed, however a limit of 100 iterations was imposed to prevent unnecessary runtime of the clustering method. This value allows the region descriptor to be unique but not so large that it would inhibit descriptor matching with additional latency. Initial cluster centers are represented by the spatial coordinates of the randomly selected feature points.

The objective function provides the scoring mechanism for the reassignment phase of the clustering. For this work, a 2-dimensional Gaussian function is exploited during the update phase of the clustering algorithm.
Figure 1. Example clustered feature points using the k-means approach with a 2D Gaussian. Feature points indicated in the same color belong to the same cluster.

B. Graph Creation

The proposed region descriptor outlined in [11,16] creates a connected graph utilizing Dijkstra’s method. The shortest-path algorithm identifies a suitable ordering of feature points based on descriptor and spatial distances. This process is improved upon by employing a filtering scheme and 1-nearest neighbor algorithm.

As an alternative to Dijkstra’s method, we propose the use of a greedy algorithm that is coupled by a feature point filtering scheme for reducing graph sizes. For multi-scale descriptors, such as SIFT and SURF, the descriptor is determined by a sampling window that is related to the detected scale. In the proposed method, feature points with overlapping sampling windows are discarded.

The initial feature point of a cluster is selected where the node is chosen based on the descriptor distance and overall average descriptor value for that particular cluster. Given the initial feature point, \( FP_0 \), all feature points within the neighborhood of \( 6s_0 \) pixels are discarded, where \( s_0 \) denotes the scale of the initial node. The initial node is then regarded as the current node. The distance between two descriptors is computed using the Euclidean distance of the feature point’s components.

The descriptor distance is computed from the current node to all other nodes. The next node chosen is the node with the least distance to the current node. Subsequent iterations are started with feature point filtering. This is repeated until all feature points in a cluster are assigned to the graph or discarded. Lastly, the resulting graph is discarded if it contains less than four nodes. Figure 2 illustrates example graphs created using feature point filtering and the greedy selection process.

Figure 2. Example graphs created from clustered SURF keypoints.

C. Graph Matching

If we are comparing two graphs, \( G_1 \) and \( G_2 \), where \( n = |V_1|, \ m = |V_2|, \) and \( n \geq m \), we must make \( n - m + 1 \) comparison to test \( G_2 \) against all sub-graphs of \( G_1 \). In the following figure, the vertices of \( G_1 \) are given as the set \( \{V_1, V_3\} \) and the vertices of \( G_2 \) as \( \{V_6, V_8\} \).

Figure 3 Partial graph matching

For the illustrated example of Figure 3, a matching score is generated for the comparison of sets \( \{V_1, V_3\} \) and \( \{V_6, V_8\} \), \( \{V_2, V_3, V_5\} \) and \( \{V_6, V_7, V_8\} \), and finally between \( \{V_3, V_4, V_5\} \) and \( \{V_6, V_7, V_8\} \). The minimal score represent the similarity between \( G_1 \) and \( G_2 \).

In this work, the similarity metric is designed to account for the angle similarities between successive graph nodes and the descriptor characteristics. Moreover, efficient angle comparisons are realized through the use of Binary Angle Measurement (BAM) and the Gray code encoding scheme.

Provided a directed graph with \( n \) vertices, \( V = \{v_0, v_1, ..., v_{n-1}\} \), and \( n - 1 \) edges, \( E = \{e_{01}, e_{12}, ..., e_{(n-3)(n-2)}\} \), where \( e_{ij} \) denotes the edge between vertices \( v_i \) and \( v_j \), we define the \( n - 2 \) angles as \( A = \{a_{01}, a_{12}, ... , a_{(n-4)(n-3)}\} \), where \( a_{ij} \) represents the angle between edges \( e_i \) and \( e_j \). As a result of the graph’s directed edges, the angle order is critical. For a given graph, two angle descriptors can be constructed to represent the structure of the graph. First, a descriptor for a graph with \( n \) nodes will be structured with \( n - 2 \) elements where \( n \geq 4 \). Let \( d_{\theta_1} \) be a graph descriptor composed of the ordered angles,

\[
d_{\theta_1} = \begin{bmatrix} \theta_{0,1} \\ \theta_{1,2} \\ \vdots \\ \theta_{(n-3),(n-2)} \end{bmatrix}
\]  

(4)

The angles utilized in the computation of \( d_{\theta_1} \) are derived from consecutive vectors of a directed graph, however the angle between non-consecutive vectors offer an additional structural characteristic of the graph-based region descriptor. When coupled with \( d_{\theta_1} \), the descriptor would provide a rotational and scale invariant representation of a graph’s
structure. The angle between any two vectors, $e_k$ and $e_m$ is defined from the dot product, $\cos^{-1} \frac{e_k \cdot e_m}{||e_k|| ||e_m||}$. Using this notation, the non-consecutive vector angle descriptor is stated in (5).

$$d_{\theta_2} = \begin{bmatrix} \theta_{0,2} \\ \theta_{1,3} \\ \vdots \\ \theta_{(n-3),(n-1)} \end{bmatrix}$$ (5)

The angles from $d_{\theta_1}$ and $d_{\theta_2}$ are typically represented with floating point numbers in the range $0^\circ$ to $359^\circ$, where a detected angle of $360^\circ$ is changed to $0^\circ$. In such a scenario, we may use the Euclidean distance for comparing the descriptors; however, we propose encoding the angles using the Gray Code encoding scheme while using the Hamming distance as a similarity measure between two descriptors. The binary encoding process starts by first converting the angle in degrees to a binary equivalent using the binary angle measurement approach. This binary string is then converted to a Gray Code string for later comparison using the Hamming distance. A Gray Code encoding with Hamming distance comparison offers a convenient form for processing on computers or dedicated hardware such as GPUs and reconfiguration hardware platforms.

The comparison of two graphs using the angle and descriptor information is completed where a sub-graph of a larger graph is compared to the entire smaller graph. For each comparison, a score is generated while the smallest of the resulting scores is regarded as the similarity score for the two region descriptors. If both graphs contain the same number of nodes, a single score is generated and assigned.

When each node of the graph is represented by a SURF feature point, the similarity between two nodes is computed using the Euclidean distance between the two descriptors, $S_{FP}$, whereas the angle similarity measure for two graphs, $\alpha$ and $\rho$, is defined as,

$$S_1 = \sum_{k=1}^{\lceil d^a_{\theta_1} \rceil} \left\lfloor \frac{d_H \left( d^{a}_{\theta_1}(k), d^{p}_{\theta_1}(k) \right) }{2^n} \right\rfloor$$

$$S_2 = \sum_{k=1}^{\lceil d^a_{\theta_2} \rceil} \left\lfloor \frac{d_H \left( d^{a}_{\theta_2}(k), d^{p}_{\theta_2}(k) \right) }{2^n} \right\rfloor$$

$$S_\theta = S_1 + S_2.$$ (6)

where $d^{a}_{\theta_1}(k)$ and $d^{a}_{\theta_2}(k)$ are the $k$-th angle in the $d_{\theta_1}$ and $d_{\theta_2}$ descriptors, respectively, for graph $\chi$. The overall score for comparing two graphs is then defined as,

$$S = \frac{\sum_{\chi} S_{FP}}{|V|} + S_\theta.$$ (7)

The graph-based region descriptor will form the basis of the detected control points for the proposed image registration method. Fast matching and the feature point’s highly discriminative nature make the descriptor an ideal candidate for the registration of disaster images. The region descriptor is defined by the area spanned by the feature points within a graph. After two region descriptors are matched, the centroid of each graph defines the control points for the subsequent parameter estimation and registration.

IV. PROPOSED REGISTRATION METHODOLOGY

Each traditional approach to image registration offers a unique set of advantages and disadvantages. The following proposed methodology provides a comprehensive approach that couples each traditional method with the proposed graph-based region descriptor. This hybrid approach allows for a broad application of the proposed technique.

![Figure 4: Overview of proposed registration method.](image)

Initially, the reference and query image are coarsely registered utilizing the cross-correlation matrix in the frequency domain where the rotation and translation parameters are obtained. The effectiveness of the coarse registration is evaluated using color pixel distributions. If the initial registration is evaluated to be accurate, a limited search domain is coupled with the proposed region descriptor to identify registration control points. In the situation where the coarse registration is detected to fail or the limited window region descriptor method is identified to be ineffective, a comprehensive approach is attempted where an intensity-based method is employed to match similar image patches.
while image features are extracted to provide the control points. This hybrid registration technique is evaluated where an unrestricted region descriptor search is leveraged if the hybrid is determined to be inaccurate. If the comprehensive registration stage is invalid, an unrestricted search stage is executed where an exhaustive search is performed utilizing the previously computed region descriptor. Figure 4 provides a visual summary of the overall steps and decision process for the proposed registration methodology while the following sections discuss each stage in detail.

A. Coarse Registration

Given two images in the frequency domain, \( I_1(u, v) \) and \( I_2(u, v) \), the frequency domain can be utilized to determine translation parameters according to the following,

\[
(x_p, y_p) = \max_{x,y} \left[ \mathcal{F}^{-1} \left\{ \frac{I_1(u, v) I_2^*(u, v)}{||I_1(u, v)||} \right\} \right] \tag{8}
\]

The point, \((x_p, y_p)\), represents the peak of the cross-correlation matrix which is estimated as the inverse Fourier transform of the normalized product of \( I_1(u, v) \) and the complex conjugate of \( I_2(u, v) \). The peak represents the translation parameters relating the original images. In order to recover the rotation parameters, the process is repeated after first converting the input images into the log-polar domain.

The Fourier Shift theorem can be exploited to determine the rotation and translation parameters that relate two images. To recover the rotation parameters, the Canny edge map for each image is transformed into the log-polar domain, while the original Canny edge maps are used directly for translation estimation. As a preprocessing step, morphological operators are applied to the edge maps prior to mapping into the log-polar domain.

The Canny edge detector, originally proposed in [34], is a popular two-pass method for classifying pixels as either an edge or background pixel. During the second pass, two thresholds are exploited to determine the state of the pixel. In this work the two thresholds are determined for each test case according to (9) and (10),

\[
\tau_{\text{low}} = \frac{2}{3} * \sum_x \sum_y I(x, y) \frac{1}{|I(x,y)|} \tag{9}
\]

\[
\tau_{\text{high}} = \frac{4}{3} * \sum_x \sum_y I(x, y) \frac{1}{|I(x,y)|} \tag{10}
\]

Mathematical morphology operators in image processing are used to describe portions of the image using shapes instead of individual pixels. The operators are applied using set theory and can all be decomposed into two basic operations, erosion and dilation [35]. While many applications apply morphology to binary images, their uses have since been extrapolated to grayscale and color imaging.

For the proposed coarse registration phase, morphological operations are applied to the image edge maps that are used for rotational estimation. They are used such that major features of an image, such as the side of a building or a major roadway, are preserved while smaller features that may negatively affect the registration outcome are removed. The aim of the operators is to reduce the image to a few large distinct features instead of many small features.

The fundamental morphological operators, erosion, \( \mathcal{E}_m \), and dilation, \( \mathcal{D}_m \), can be used to define other useful operators. The equation (11) below will preserve the main objects in an image while eliminating disjoint internal regions and external nubs. An example canny edge map is given in Figure 5a along with the resulting binary image from applying the \( M(I, \mathcal{S}_D, \mathcal{S}_E) \) operator in Figure 5c.

\[
M(I, \mathcal{S}_D, \mathcal{S}_E) = (I \oplus \mathcal{S}_D) \ominus \mathcal{S}_E \oplus \mathcal{S}_D. \tag{11}
\]

![Figure 5](image-url)

<table>
<thead>
<tr>
<th>a.</th>
<th>b.</th>
<th>c.</th>
</tr>
</thead>
</table>
| Original canny edge map | Canny edge map after closing operation | Binary image after dilating closed edge image according to \( M(I, \mathcal{S}_D, \mathcal{S}_E) \).

B. Registration Evaluation

After the coarse registration stage and subsequent registration attempts, the registered query image is compared to the original query image in order to determine the effectiveness of that particular registration approach. The hypothesis is that an image’s color histograms should be similar before and after registration. For images in the RGB color space, there are three histograms associated with the image. Each color channel’s histogram consists of 256 bins while each distribution is normalized.

Evaluating two color images in the RGB color space requires the calculation of three similarity scores; therefore the proposed registration verification method imposes a threshold that is the sum of the three histogram scores, \( \tau_h = \mathcal{S}_h^r + \mathcal{S}_h^g + \mathcal{S}_h^b \). The similarity scores are calculated using the Euclidean distance between the color channel histograms. Through experimentation, it has been determined that a value of 0.15 for \( \tau_h \) offers a fair trade-off between registration false acceptance and rejection.

C. Limited Window Search

The proposed system uses an initial registration step to improve feature point matching accuracy and subsequent control point identification. If the coarse registration step is surmised to be valid, a registration attempt is performed involving the proposed graph-based region descriptor within a localized neighborhood.

Invariant feature points are extracted from the original reference image and the FFT registered query image. For this work, k-means clustering is exploited to identify graph nodes from their spatial relationships and the graph is constructed by
using the k-nearest neighbor shortest path algorithm, with $k = 1$.

In the limited neighborhood registration approach, feature points are matched utilizing a small search space of the query image that is determined from the centroid of the graph-based region descriptor in the reference image. Provided a graph in the reference image with centroid at $C_r = (x_r, y_r)$, the subset of query graphs, $\overline{G_Q}$, that are tested against the reference graph is defined as,

$$\overline{G_Q} = \{ G_i | \|C_i - C_r\| \leq r_w \}, \overline{G_Q} \subset G_Q$$  \hspace{1cm} (12)

Figure 6. Limited search neighborhood where the red square represents the centroid of the reference graph while green dots denote centroids of graphs in the query image that lie within the search window.

The reference graph is then compared to all graphs in the set $\overline{G_Q}$, where the graph pair with smallest score is chosen as a match. For each matched pair, the centroids are acknowledged as the control points for subsequent transformation parameter estimation using the direct linear transform.

D. Comprehensive Intensity and Feature-based Search

The second registration approach couples intensity and feature-based approaches to accurately determine registration control points. This method is attempted if either the initial coarse registration or the limited neighborhood registration fails the histogram comparison.

Our work in [36] provided strong evidence that a hybrid method which couples feature and intensity-based analysis for image registration can be provide accurate image registration for situations involving scenes before and after a disaster. A coarse search is performed utilizing an intensity-based approach for determining similar image patches. From the similar image patches, invariant features are extracted, matched and designated as registration control points.

Using the original reference and query images, the query image is segmented into $n \times n$ segments. Each segment is compared to the entire reference image through a sliding window. For each comparison, the normalized cross correlation metric is used to associate a similarity score with the two image patches. The normalized cross-correlation equation is given as,

$$NCC = \frac{1}{m \times n} \sum_{x,y} \frac{[I(x,y) - \overline{I}][t(x,y) - \overline{t}]}{\sigma_I \sigma_t}$$ \hspace{1cm} (13)

where $I(x,y)$ is the grayscale reference image, $t(x,y) \in \mathbb{R}^{m \times n}$ is the image patch from the query image, whereas $\overline{I}$ and $\overline{t}$ are the means of $I(x,y)$ and $t(x,y)$, respectively. $\sigma_I$ and $\sigma_t$ denote the standard deviation of pixel intensities for $I(x,y)$ and $t(x,y)$. The NCC similarity metric is the most common correlation function for intensity-based registration and can be computed quickly, however the function is susceptible to variations in pixel intensity, such as lighting condition differences, as well noise [38].

In this work, the query image is segmented such that each segment is 50 pixels wide and 50 pixels tall. Of all the matched pairs, the set of images with the 8 lowest scores are chosen for feature point matching, unless the total number of patches is less than 8 in which case all matched pairs are utilized.

Invariant feature points are then extracted and matched between the identified similar patches, independent of any other patches. The proposed method exploits the symmetric matching scheme to ensure a one-to-one mapping of feature points. Symmetric matching can be accomplished by determining the best match pair from the reference to query image and from the query to reference image. If the best matched pairs are identical in both scenarios, the pair is understood to be symmetrically matched. The set of matched feature points is then used as the set of control points for transformation parameter estimation.

E. Unrestricted Search

If the comprehensive registration approach produces an inaccurate result, the final attempt to perform registration uses the proposed graph-based region descriptor without a limitation of the search space. A brute force matching scheme is executed on the original reference and query images, comparing the detected feature points from the limited window approach. Since this is the last stage of the registration hierarchy, the resulting parameter estimation is regarded as the final output.

V. RESULTS AND DISCUSSION

A. Region Descriptor Results

In order to evaluate the effectiveness of the proposed region descriptor, a ground truth is established for each image in the test set. The manually selected points are used to compute a transformation which is used to evaluate the expected location of each feature point. Given two images, $I_1$ and $I_2$, that vary under a projective transform, there exists a transformation matrix, $H$, such that $I_2 = HI_1$. If the transformation matrix is known, the location of a feature point in $I_1$ can be identified in $I_2$ according to,

$$\begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = H \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix},$$

where $(x_1, y_1, 1)$ and $(x_2, y_2, 1)$ are the spatial coordinates of feature points in $I_1$ and $I_2$, respectively. The following process was used to identify the number of correctly matched feature points for a group of 41 matched images.
- For each image pair, manually select 10 matching control points across both images.
- Estimate the transformation matrix using the Direct Linear Transform.
- For each feature point in the reference image, calculate the projected point in the query image using the estimated transformation matrix.
- Calculate the spatial distance between the projected point and the matched feature point of the query image. If the two points are within 10 pixels, count the pair as a match.

The following graph in Figure 7 compares the matching accuracy distribution of the proposed region descriptor to SURF, SIFT and BRISK distributions. As shown, BRISK demonstrated the poorest performance, as depicted by the majority of test cases exhibiting less than 20% matching accuracy. For the aerial and street view images, SIFT outperformed SURF, while the proposed region descriptor is observed to have a higher matching accuracy where most test cases demonstrated a matching accuracy of at least 60%. The distribution parameters are summarized in Table 1, where it is evident that the proposed region descriptor provides a higher average matching rate than the traditional feature descriptors. As shown in Table 1, the graph-based region descriptor produced an average matching accuracy of 62.19%, where traditional approaches using SURF, SIFT and BRISK had matching accuracies of 24.65%, 35.18%, and 14.2%, respectively, which indicates the effectiveness of the graph-based region descriptor in situations involving disaster scenes.

The overall distributions depicted in Figure 7 indicate the effectiveness of the proposed region descriptor by illustrating the descriptor’s superior distribution for higher matching rates. Similarly, for lower matching rates, the proposed descriptor is observed to have the lowest distribution, further indicating the proposed method’s robustness in situations that involve large variations which adversely affect the traditional approaches.

Each of the image sets can be categorized according to the difficulty associated with feature matching. The sources of such difficulties may be a combination of geometric variations from perspective differences, and image content variations from the effects of a disaster, object occlusion, or lighting conditions. 31.7% of the test cases exhibit examples of each difference that adversely affect the matching process. In these instances, the matching rate is low; however the proposed region descriptor still outperforms the traditional methods. Similarly, approximately 24.6% of the test cases provide examples where the variations are not as extreme and therefore the matching results are more favorable. Lastly, approximately 43.7% of the test data is comprised of images where distinct features are easily matched across images. In all three scenarios, the proposed region descriptor is shown to be more robust than the traditional methods.

<table>
<thead>
<tr>
<th>Matching Accuracy (%)</th>
<th>Region Desc (Proposed Method)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SURF</td>
<td>SIFT</td>
</tr>
<tr>
<td>24.65</td>
<td>35.18</td>
</tr>
</tbody>
</table>

Table 1. Matching accuracy distribution parameters.

For the proposed method's robustness in situations that involve large variations in image content between the original reference and query images, the effectiveness of the transformation parameter estimation, the root mean square error (RMSE) is calculated for each image set according to,

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=0}^{n-1} (x^i_e - x^i_m)^2},
\]

where \(n\) is the number of iterations (180), \(x^i_e\) is the \(i\)-th expected value, and \(x^i_m\) is the \(i\)-th measured value. When computing the RMSE of two images, \(x^i_e\) represents pixel \(i\) of the reference image while \(x^i_m\) is pixel \(i\) of the query image, where \(i \in \mathbb{R}^2\).

In the proposed approach, the rotation and translation parameters are estimated using the Canny edge maps for a given image pair. The edge maps are shown to be more accurate when determining the rotation between two. For each angle in the range 1° to 180°, the RMSE was calculated for each of the 41 image pairs. The average RMSE for rotation recovery is approximately 6° lower when the Canny edge map is used instead of the original images. The average RMSE increases proportionally to the angle variations; however the use of Canny edge maps reduces the rate in which the RMSE rises. Similarly the RMSE was calculated for query image offsets from 1 to 50 pixels in the x-direction and 1 to 50 pixels in the y-direction. The average RMSE is shown to be lower when Canny images are used for translation estimation than the original reference and query images.

An example coarse registration result is provided in Figure 8, which also includes an example result from registering the query image using SURF feature points and RANSAC for parameter estimation. For instances with poor coarse registration, large variations in image content between the
reference and query image adversely affect the parameter estimation. In an ideal situation, the coarse registration is most advantageous if only geometric variations are present, whereas photometric variations such as obstructions or large amounts of noise produce poor results. Figure 11 provides images that vary in lighting conditions; however, the edge maps are similar enough that the translation and rotation parameters are easily recovered.

For the given test database, 65.8% of the image sets rely on the limited window region descriptor registration technique. One example of a successful registration utilizing the initial coarse registration and limited window region descriptor approach is given in Figure 10. The example is an aerial view of a residential area that exhibits lighting condition differences and image content variations due to the house affected by a devastating fire. Images in Figures 10c and 10d provide reference registrations using common approaches utilizing RANSAC and manually selected control points.

C. Limited Window Region Descriptor Registration Results

The limited window graph-based region descriptor registration approach exploits the results of the log-polar phase correlation registration. If the coarse registration step is validated using the color histograms, invariant feature point clusters are used as a basis for forming k nearest neighbor (k-NN) shortest distance graphs, where \( k = 1 \). The registration technique’s success is directly related to the matching of the graph-based region descriptors. Since the coarse registration is verified to be accurate, the search space for matching graphs is greatly reduced. Figure 9 illustrates an example of matched graph-based region descriptors that are used to identify registration control points. As depicted, the green lines represent connected region descriptors where intersection of the green lines is minimized, which indicate strong evidence of accurate matching. Moreover, visual inspection confirms each match is one-to-one.

D. Comprehensive Intensity and Feature-based Registration Results

If the initial coarse registration or limited window registration approach produces invalid results, as determined by the histogram comparisons, a coarse template search is conducted using the normalized cross correlation metric to determine similar image regions. Each pair of matched regions is then matched locally using invariant feature points. For the proposed method, SURF feature points provide the basis for control point identification where a symmetric matching scheme is exploited. The following figure illustrates an example image set where the SURF feature matching and coarse initial template search provides an accurate registration.

Of the entire test database, 12.2% of the image sets were registered using the SURF feature points directly after identifying similar image patches through the use of the normalized cross-correlation metric. Figure 11 demonstrates
the effectiveness of the proposed method with a street view perspective of a building that has sustained significant damage. Although the textural properties are similar, large geometric differences are present due to the devastating effects of an earthquake. In the provided example, the SURF matches are shown, along with the registered query image. The image depicting the feature point matches contains the query image on the left and reference image on the right where the matched points are connected through colored lines.

![Matched feature points from the comprehensive registration stage.](image1)

b. Registered query image using SURF and RANSAC.

c. Registered query image using manually selected control points.

d. Registered query image using the proposed template matching technique.

e. Registered query image using the proposed unlimited search window method.

Figure 11. Street view of a registered building using NCC template matching and SURF.

E. Unrestricted Region Descriptor Registration Results

In the last mode of the proposed method, an unrestricted search is performed to match the proposed graph-based region descriptor. The resulting matching pairs are used as the basis for control point identification. This approach is only attempted if the previous two methods are determined to yield incorrect registrations. The remaining 22% of the test sets utilized the unrestricted technique.

For the example provided in Figure 12 the unlimited window search method is utilized to successfully register a street view image of a building before and during a fire. The fire and subsequent smoke provide significant obstruction in the query image. Moreover, the images were captured at different times and from different perspectives. This example exhibits the proposed approach’s robustness for street view scenes during a disaster and from different viewpoints.

![Registered query image using SURF and RANSAC.](image2)

b. Original query image.

c. Registered query image using SURF and RANSAC.

d. Registered query image using manually selected control points.

e. Registered query image using the proposed unlimited search window method.

Figure 12 Registration of a street view building during a disaster.

F. Overall Registration Results

As another baseline for comparison, the mutual information metric was used to determine the effectiveness of the proposed algorithm as demonstrated in [11]. The function is optimally 1.0 when two images are identical and therefore must be maximized for accurate registration. For comparison, the reference image and query image are converted to grayscale while the MI score is calculated from (15).

\[
MI(I_1, I_2) = H(I_1) + H(I_2) - H(I_1, I_2),
\]

where \(H(I_x)\) is the entropy measure for image \(x\) and \(H(I_x, I_y)\) is the joint entropy of images \(x\) and \(y\). \(H(I_x)\) and \(H(I_x, I_y)\) are defined as follows,

\[
H(I_x) = - \sum_{x \in I_x} p_{i_x}(x) \log(p_{i_x}(x)),
\]

\[
H(I_x, I_y) = - \sum_{x \in I_x} \sum_{y \in I_y} p_{i_x i_y}(x, y) \log(p_{i_x i_y}(x, y)).
\]

The probability density function, \(p_{i_x}(x)\), is estimated from the intensity histogram of image \(I_x\) while \(p_{i_x i_y}(x, y)\) is calculated from the joint histogram of images \(I_x\) and \(I_y\).
The test set was segmented into three classes where class 1 represents the set of images that are ideally registered using the limited window method, while class 2 and class 3 are image sets accurately registered using the comprehensive method and unlimited search window technique, respectively. The mutual information scores for each class using each approach is summarized in Table 3. As shown, the mutual information score is maximized for a particular class using the proposed method. Moreover, when a different approach is attempted, the mutual information is shown to be less than the score generated from the ideal approach. Table 3 indicates the proposed registration methodology is effective for a wider range of applications than if a single approach is used. Moreover, the overall mutual information score using the registered query image from any of the proposed methods is observed to be greater than the score without registration in 92.7% of the tested cases. It is for the class 1 and class 2 test sets, a higher mutual information score observed using the unlimited search approach, however this is coupled with significant latencies associated with an exhaustive search. A summary of the mutual information distribution parameters is provided in Table 2, which also presents strong evidence of the proposed region descriptor’s effectiveness for identifying suitable registration control points. The decrease in standard deviation indicates an improvement due to the proposed registration technique. Table 2 shows that the proposed method increased the mutual information to an average score of 0.46, which is an increase of 61.24% over the score using the original reference and query images.

This visual representation confirms that the registered images improved the score for scenarios including aerial views of commercial and residential buildings, as well as buildings from a street view. In each scenario the image content varies geometrically from perspective differences and photometric variations are present through differences in lighting conditions, obstructions and natural disasters. In the most challenging cases, successful registration is realized when image content varies greatly due to the devastating effects of a disaster, such as fire and flooding damage. From the evaluated image sets, it is shown that the proposed registration method is an effective approach for applications involving images before and after a disaster. Registration results could be further improved by incorporating a RANSAC procedure along with the direct linear transform for parameter estimation; however, this potential improvement would require additional latency.

Table 2. Mutual information distribution parameters for the MI score between the original reference and query image, \( I(1;2) \) and the MI score for the original reference and registered query images, \( I(1;\text{reg } 2) \)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average</strong></td>
<td>0.29</td>
<td><strong>0.512</strong></td>
<td><strong>0.40123</strong></td>
<td>0.5083</td>
</tr>
<tr>
<td><strong>Std Deviation</strong></td>
<td>0.17</td>
<td>0.16</td>
<td>0.0783</td>
<td><strong>0.4282</strong></td>
</tr>
</tbody>
</table>

*Registered image using proposed approach.

The limited window search stage is shown to be effective in scenarios where content differences are minimal, such as noise or object obstructions. This stage is robust against illumination, translation, scale and rotational variations, but is adversely affected by structural differences. Many of the aerial disaster images in the test set were successfully registered at this stage. The comprehensive stage is robust in situations that involve content differences, such as object obstruction or structural differences; however, the method is negatively affected by illumination differences and structural repetitions, such as the windows in an urban scene. Lastly, the unrestricted search approach can be utilized in situations involving larger translation or rotation variations and scenarios that exhibit photometric differences, such as lighting conditions. The overall result of coupling all three stages is an approach that can be used with disaster images that have perspective, structural or textural differences.

The results provided in Figure 8 illustrates an example coarse registration that is performed at the initial stage of the propose method. The limited window registration method is attempted after evaluating the effectiveness of the coarse registration, which is shown in Figure 10. If the coarse registration or the limited window technique is determined to be ineffective, the intensity-based approach using NCC and SURF feature points is attempted. An example of the intensity-based method is given in Figure 11. Lastly, if the NCC approach is not successful, the unlimited search window scheme is executed where all detected region descriptors from both images are exhaustively compared. Figure 12 highlights an example registration performing the proposed unlimited window search technique. For every registration example in Figures 8, 10, 11 and 12, the proposed method is compared to the common feature-based approach utilizing SURF and RANSAC, as well as registration results resulting from the homography estimation of manually selected control points. In each of the provided examples, the proposed method is shown to provide a more accurate registration than the two common approaches.

The average latency of the proposed registration method is determined experimentally to be 5.82s for the limited window approach, 5.7s for the SURF and normalized cross-correlation technique, and 10.62s for the unrestricted region descriptor method. It has been shown that the main tasks of the proposed method can greatly benefit from a GPU implementation. With
the aid of a CUDA-enabled GPU, it has been estimated that latencies are reduced to 172ms, 114ms, and 428ms, respectively.

VI. CONCLUSIONS

In this work a novel multi-stage registration process is proposed which utilizes an effective graph-based region descriptor. The proposed approach attempts several methods for registration while evaluating each registration result between stages. This approach is shown to be a viable solution for registering images of scenes before and after a disaster. In such a scenario, the images to be registered may exhibit great variation in photometric and geometric characteristics. The application of disaster scene analysis and registration was used to show the effectiveness of the proposed technique.

The proposed method attempts an initial coarse registration, by estimating the translation and rotation parameters that relate two images, exploiting the shift properties of the image’s Fourier transforms. It is shown that the root mean square error (RMSE), calculated between the recovered rotation angle and expected angle, is lower when using the Canny edge maps and morphological operators for angles between 1° and 180°. Furthermore, the RMSE is shown to be consistent when estimating the translation offsets, regardless of whether or not the edge maps are utilized. Although the utilization of edge maps improves the original phase correlation method, there still exists room to improve the registration. Experimental results show that an RMSE of 4° is realized for angle differences of 180°.

Clusters of invariant feature points are used as the basis for creating the graph-based region descriptor, where the keypoints represent the nodes of the graph. Experimental results validate the matching ability of the proposed region descriptor. It is shown that the graph-based descriptor provided a higher matching rate than SIFT, SURF and BRISK for most scenarios in the test set. The proposed region descriptor is shown to be effective in scenarios that exhibit large variation in pixel intensities and structural differences. Moreover, the proposed descriptor is shown to identify image features within urban scenes and general objects with applications ranging from general registration to registration of an aerial imagery of an urban setting for such uses as scene stitching and disaster assessment.

For the given test set, 12.2% of the image pairs were registered using SURF feature points with the proposed initial coarse search using the normalized cross-correlation. The experimental results indicate the SURF-based approach produces accurate registrations when photometric variation is minimized. An example application where this phase would be advantageous is urban building registration where lighting conditions are similar between the two images.

It is shown that 22% of the test images failed through the SURF matching and therefore required the unrestricted approach. Of the scenarios that employ the unrestricted technique, 66% were registered accurately. The proposed approach is demonstrated to be effective in natural scenes, aerial images and urban scenarios.

Overall, the proposed registration technique is shown to improve registration accuracy when compared to traditional techniques in scenarios where large variations exist in pixel intensities, such as lightning conditions or damage from a natural disaster, as well as geometric differences, such as perspective variations. The mutual information metric was used to show experimentally that 92.68% of the query images were successfully registered.

REFERENCES
