Registration of Camera Captured Documents Under Non-rigid Deformation

Venkata Gopal Edupuganti
Dept. of Computer Science
New Jersey Inst. of Technology
Newark, New Jersey, USA
vge2@njit.edu

Vinayak A Agarwal
Stanford University
Stanford, California, USA
vinayak1@stanford.edu

Suryaprakash Kompalli
Hewlett-Packard Labs
Bangalore, India
kompalli@hp.com

Abstract

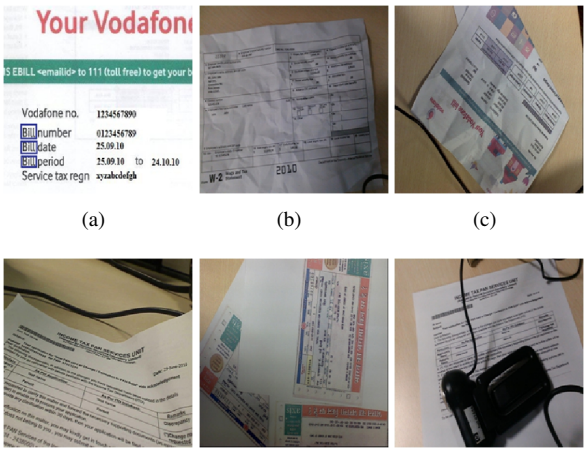
Document registration is a problem where the image of a template document whose layout is known is registered with a test document image. Given the registration parameters, layout of the template image is superimposed on the test document. Registration algorithms have been popular in applications such as forms processing where the superimposed layout is used to extract relevant fields. Prior art has been designed to work with scanned documents under affine transformation. We find that the proliferation of camera captured images makes it necessary to address camera noise such as non-uniform lighting, clutter, and highly variable scale/resolution. The absence of a scan bed also leads to challenging non-rigid deformations being seen in paper images. Prior approaches in point pattern based registration like RANdom SAmple Consensus (RANSAC) [4], and Thin Plate Spline-Robust Point Matching (TPS-RPM) [5, 6] form the basis of our work. We propose enhancements to these methods to enable registration of cell phone and camera captured documents under non-rigid transformations. We embed three novel aspects into the framework: (i) histogram based uniformly transformed correspondence estimation, (ii) clustering of points located near the regions of interest (ROI) to select only close by regions for matching, (iii) validation of the registration in RANSAC and TPS-RPM algorithms for non-rigid registration. We consider Scale Invariant Feature Transform (SIFT) [8] and Speeded-Up Robust Features (SURF) [1] as our features. Results are reported comparing prior art with our method on a dataset that will be made publicly available.

1. Introduction

Image registration [5, 6, 10, 13, 15] by establishing correspondence across interest points in image pairs has been well studied in computer vision. Registration becomes more challenging when outliers exist in the correspondence set. These outliers could arise from noise in image acquisition, feature extraction and/or matching. Several local invariant (e.g. scale, affine, and intensity) detectors and descriptors [1, 3, 8, 9] have been proposed to overcome the natural variations in image acquisition. Feature dissimilarity measures such as L2 norm, cosine distance etc together with outlier elimination techniques such as RANdom Sample Consensus (RANSAC) [4], Hough Transform [8], TPS-RPM [5, 6] have been applied to establish true correspondences. The goal for most techniques has been to estimate underlying transformation function across natural images for purposes such as image stitching, image augmentation, or camera geometry estimation.

Camera captured documents differ from natural images in a few critical areas: (i) Non-linear deformations such as folds are common in documents, (ii) Our problem targets forms and in most forms, the filled values are large in number and the amount of similar content between the test and template is a percentage of the document content, (iii) Content such as logos, and even text is repeated at multiple locations within the document. Fig. 1 shows a few forms that are quite sparse in content, where the same logo/text occurs at multiple locations in a document. This results in a new class of outliers that are similar in the domain of local features but correspond to a different region that is not aligned with the global image layout. The existence of correspondences from one region to multiple regions increases the number of outliers, and has an adverse effect on traditional iterative methods such as RANSAC. These challenges are in addition to known problems in camera capture such as lighting variations, camera equipment differences, and scale.

Document image processing has earlier used registration techniques for forms processing. The motivation has frequently been that information from a small part of the document is critical for most user applications. For example, the amount and date on a receipt is all that is needed as an input to a tax software. Limiting downstream processing to relevant regions is known to be useful both from the view of accuracy and speed. To extract only relevant regions, a test image is registered with a template image that has



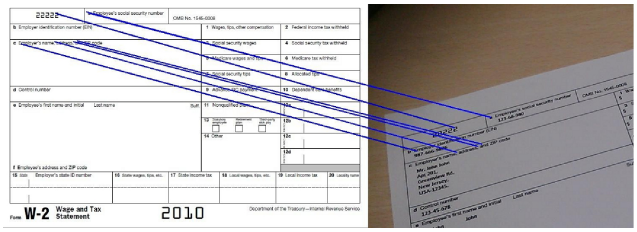
(a) (b) (c)

(d) (e) (f)

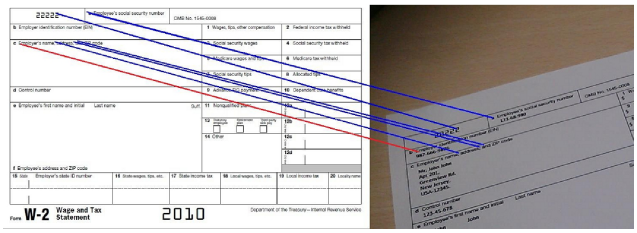
Figure 1. (a) Regions (in blue) are similar(i.e. "Bill"), (b)-(f) captured document images with non-planar deformations.

known layout. Here, selected regions of text are extracted from a filled in form (test image) using information about the form layout (template image). Registration parameters are used to overlay the layout of the template image onto the test image. The layout specifies geometric positions of the relevant fields, which are then extracted from the test image. While several prior techniques [12, 14] fall in this area, these methods address only affine transformations and assume a high-quality image. We will show that camera images have several other variations that prove challenging to these methods.

Approaches such as RANSAC [4] and Hough clustering [8] estimate true correspondence by fitting a transformation function to existing correspondences. They are designed to eliminate correspondences between points that have high feature similarity but do not agree with the global image geometry. By design, outliers would also influence the transformation function, and would be considered as inliers if they conform to the underlying transformation. For example, Fig. 2 shows an outlier that conforms to global geometry has been considered as an inlier. This, and similar outliers deviate the region of interest from the desired location. Applying these methods for non-rigid registration is not acceptable as the underlying transformation function varies at different parts of the image. Several non-rigid registration frameworks [5, 6, 10, 13, 15] have been developed for the non-rigid registration of medical images. One of them is the Robust Point Matching (RPM) [6, 15] algorithm, which formulates the registration problem as a maximum likelihood estimation problem using mixture models. Chui and Rangarajan [6] embedded the (Expectation Maximization) EM-like frame work in a deterministic annealing scheme by considering the soft-assignment of point sets to allow partial matches. They applied Thin-Plate Splines (TPS) [2] for the estimation of underlying transformation function, as TPS can be decomposed into affine and non-affine sub spaces. The Robust Hybrid Deformable Matching (RHDM) framework [15] incorporates feature dissimilarity measure into the TPS-RPM framework. Sofka et al. [13] pointed out that the TPS-RPM algorithm would fail on extraneous structures (such as H-shape point sets), as it tries to align the center of mass of the point sets in the early iterations of the algorithm leading to a bias in the estimate which it can not overcome in the later stages of the algorithm. Recently, Myronenko et. al. [10] incorporated motion coherence theory into the framework in the place of TPS. All these methods fail to consider a few factors: (i) Initial correspondences are not taken into account, (ii) New correspondences which are not in the initial correspondence set are created during matching, (iii) It is assumed that template points (points in the source image) are sparsely distributed, and (iv) The same search range parameter is considered for all template point clusters. We show that accounting for these factors in image registration is central to non-affine document registration.



(a)



(b)

Figure 2. (a) Correspondences before RANSAC, (b) Correspondences after RANSAC. Wrong correspondences are shown in red color. This outlier deviates the region of interest from the desired location.

2. Document Image Registration Methodology

Fig. 3 presents an overview of our method that is designed to address some of the drawbacks described earlier. The rectangular boxes (blue color) in the template image of Fig. 3 indicate the regions of interest to be extracted from a test image. We form clusters of template points using k-means [7] in the template image. Clusters that satisfy a

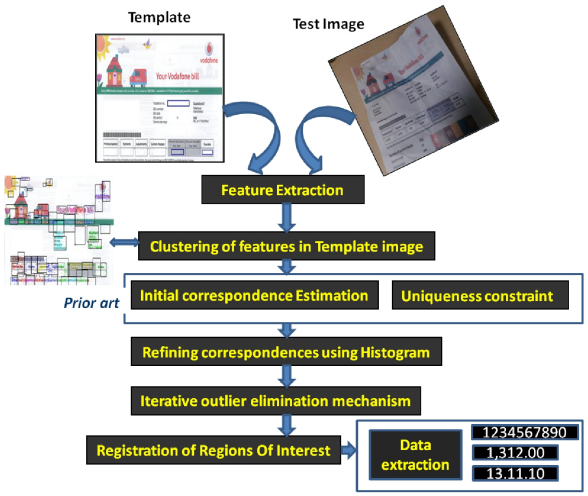


Figure 3. Overview of Document Image Registration. The template image can be a scanned image or electronically generated where the regions of interest (ROIs) are known. Expected output is ROIs in the test image.

proximity criterion with respect to the Regions Of Interest (ROI) are selected for registration. We incorporate a histogram based uniformly transformed correspondence estimation into the framework to speed up iterative correspondence estimation. We further show how prior knowledge of correspondences can be integrated into TPS-RPM framework, and enhance RANSAC and TPS-RPM by minimizing the registration error computed using local gradient information. We demonstrate that this framework enhances the performance of these algorithms for non-rigid deformations.

The rest of this section describes our methodology in detail. Section 2.3 presents the iterative approaches for outlier elimination and registration; here we discuss RANSAC, and enhanced RANSAC. We present the iterative approaches for non-rigid registration using TPS-RPM and enhanced TPS-RPM framework in sections 2.5 and 2.6 respectively. Section 3 presents the experimental results on various document images. Finally, Section 4 concludes the paper.

2.1. Template point selection and Initial correspondence

We extract invariant points from the template and test images using methods such as SIFT or SURF. Subsequently, we will refer to points in template and test image by X and Y respectively. Feature points in the template image are clustered by using k-means algorithm [7]. For each ROI r in the template image, points belonging to m clusters that are closest to the ROI are selected as the template point set for the ROI (X_r). The idea behind the selection of points only in the m closest clusters is that these points move closely with the ROI, and further it reduces the non rigidity among the points.

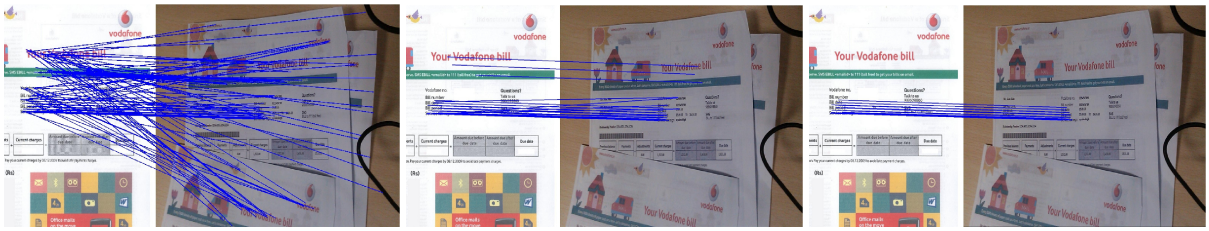
Lowe's [8] method of initial correspondence generation is used to map points in X_r onto feature points in Y [8]. For each $x_i \in X_r$, two closest points in Y are found by using Euclidean distance of the feature space. If the ratio of these distances is less than t, then the template point with lesser distance is added to the correspondence set $C = \{(x_i, y_j) | x_i \in X_r \text{ and } y_j \in Y\}$. The correspondences now have a many-to-one mapping from X to Y. For each test point $y_j \in C$, a new correspondence set C' is obtained by performing a reverse mapping. Each point in $y_j \in C$ is now mapped onto the points $x_i \in C$. Correspondences are retained only if the obtained mapping is already present in C. This ensures that for each $y_j \in Y$ there exists only one $x_i \in X_r$. The new correspondences are now $C' = \{(x_i, y_j) | x_i \in X_r, y_j \in Y, \text{ and } (x_i, y_j) \in C\}$.

2.2. Refine Correspondence Set using Histogram

We eliminate correspondences among outliers by using a histogram of Euclidean distances on the Cartesian coordinate space. The Euclidean distance between Cartesian coordinates of x_i and y_j for all $(x_i, y_j) \in C'$ is obtained and placed into histogram bins. Bin size is given by $(max_{dist} - min_{dist})/(\textit{number of bins})$, where max_{dist} and min_{dist} are the maximum and minimum Euclidean distances of the corresponding points $(x_i, y_j) \in C'$. Correspondences whose euclidean distances fall in the peak bin and the bins that are within the threshold t_e of the height of the peak bin are selected in a new set C''. This step operates under the assumption that while local distortions in document images can be non-planar, these distortions will not grossly alter the relative distribution of corresponding points. The results section will discuss how this step eliminates gross outliers, improving the convergence rate of iterative mechanisms. Fig. 4(a) shows the one-to-one correspondences (C') obtained for a sample test/template image, Fig. 4(b) shows refined correspondences (C''), and Fig. 4(c) shows the correspondences after RANSAC under affine transformation.

2.3. Iterative Approaches for Outlier Elimination: RANSAC

RANSAC is an iterative optimization algorithm that repeats two phases: (i) generation of hypothesis by randomly sampling the data and (ii) hypothesis verification on data. Termination is done after a fixed number of iterations or when a termination condition is met, with several variants of RANSAC depending on the speed, accuracy, and robustness [4]. In our method, each RANSAC iteration selects three random non-collinear points from $x_i \in X$ such that $(x_i, y_j) \in C''$. Using the correspondence between x_i and y_j, we compute an affine transformation matrix. The trans-



(a) (b) (c)

Figure 4. Correspondences at different stages of the framework (a) after Lowe's [8] method and one-one mapping, (b) after Euclidean distance based histogram, and (c) after RANSAC.

formation matrix M is applied on $\forall x_i | x_i, y_j \in C''$, to obtain $\bar{x_i}$. If $\bar{x_i} \equiv y_j$, x_i is marked as inlier else x_i is marked as outlier. If the number of inliers in a particular iteration are greater than inliers in a previous iteration, the current set of inliers are accepted. The algorithm is terminated after a fixed number of iterations.

2.4. Enhanced RANSAC for Robust Registration

RANSAC is able to eliminate correspondences that do not conform to global geometry, and obtain a gross match between the template and test images. However, we find that additional verification is needed to eliminate outliers arising from locally non-affine distortions (Fig. 2). Since specific regions of the image (ROIs) are of interest to us, we limit processing to these ROIs. We additionally assume that there will be image regions near the ROI that are similar across the test and template image. In each iteration of RANSAC, when the transformation matrix M is obtained, we use M to warp the test image onto the template. Histogram of gradients (HOG) [8] is computed from image regions surrounding the ROI in the template and the test images. A modified RANSAC is performed using Chi-square [11] similarity of the HOG as the matching criterion. The method is described in algorithm 1.

Algorithm 1 Outlier elimination using Enhanced RANSAC

Input: Set of input correspondences C'', Test image B; m_r the number of fixed regions for the registration of ROI.
$HOG_i : i = 1, 2, ..., m_r$; HOG of fixed nearby regions.
HOG_{dist}: maximum positive integer.
Output: Refined correspondence set C''' with inliers, Transformation matrix M.
Initialization: $iterations$=0; $inliers$=0; $outliers$=0; MAX_{iter}=maximum number of iterations.
while $iterations < MAX_{iter}$ **do**
  Hypothesis generation: Randomly select three correspondences among non-collinear points from C''. Compute the transformation matrix $Current_M$ from the three correspondences.
  Hypothesis evaluation: Warp the test image B with $Current_M$ to align with the template image. Compute HOG of the fixed regions in the warped image $HOG_j : j = 1, 2, ..., m_r$
  Compute the chi-square distance between HOG_i and $HOG_j : i, j = 1, 2, ..., m_r$, average it with m_r, and denote it as $Curr_{dist}$.
  if $Curr_{dist} < HOG_{dist}$ **then**
    Update:
    $HOG_{dist} \leftarrow Curr_{dist}$
    $M \leftarrow Current_M$
  end if
end while
Update Correspondence set C''' with the correspondences that agree with M.

2.5. Thin Plate Spline-Robust Point Matching

While we find that enhanced RANSAC is capable of addressing some of the deformations, methods like TPS-RPM have been specifically designed to derive non-rigid transformation functions [6, 15]. This section describes the TPS-RPM algorithm, and a few drawbacks of the method when applied to our case. We provide enhancements to TPS-RPM in the subsequent section. Let $X = x_i : i = 1, 2, ..., N$ be a sparsely distributed template point set and $Y = y_j : j = 1, 2, ..., M$ be a relatively dense test point set. Both point sets are projected on a normalized Cartesian coordinate plane. TPS-RPM [6, 15] uses Gaussian mixture density to model the distribution of test points, while Gaussian cluster centers are determined by the template points. In order to robustly align the two sets, the algorithm performs deterministic annealing, where the temperature T of the annealing process acts as a search range parameter. At high temperatures the algorithm aligns the two point sets by preserving global structure of the template points, when T goes down the search becomes local where it accounts local deformations. It starts the annealing process with a larger T such that all the test points will be in the vicinity of template point clusters. At each T, it alternately estimates the

correspondence and computes the underlying transformation function. It computes the probabilities of all test points being assigned to the template point clusters and computes the probable location of the matching point for each template point. With the template points and the corresponding probable matching points it estimates the transformation function f using TPS [2] to ensure smoothness in the transformation function. It repeats the annealing process with the template point clusters centered at $f(x_i)$ until T reaches final temperature T_{final} (i.e. average of the squared distance between the nearest neighbors of the test points). To handle outliers in both point sets it maintains two additional clusters centered at the center of mass of the both point sets with large temperature T_0.

Drawbacks:

- The assumption of template point set as a sparsely distributed one is not true in the case of document images with multi-scale local features, as operators like SIFT, SURF generates dense points in a given region.
- TPS-RPM aligns the template point set to the test point set by considering only the geometry of the template point set. Apart from geometry there is an initial correspondence set which gives additional information to prevent template points being assigned to irrelevant test points.
- Each iteration of TPS-RPM generates new correspondences which are not in the initial correspondence set. The new irrelevant correspondences penalize the estimated transformation function.

2.6. Enhanced TPS-RPM

We design the enhanced TPS-RPM algorithm to overcome the drawbacks of TPS-RPM. Apart from the template point set X_r and test point set Y, the algorithm takes into account the correspondence set C''. To prevent each template point being moved towards the irrelevant test point we assign different temperature T_i to each Gaussian cluster center x_i. Finally, the algorithm refines the new correspondences with nearby identical correspondences in C''. Remaining parts of this section present the problem formulation, enhanced TPS-RPM algorithm, and the refinement of new correspondences.

Let $C'' = (x_i, y_j)|x_i \in X_r, y_j \in Y$ be the set of input correspondences computed using the methodology in Section 2.2, where $X_r = x_i : i = 1, 2, ..., N$ and $Y = y_j : j = 1, 2, ..., M$ are the template and test point sets respectively. As we enforce one-one mapping in the correspondence set, N is equal to M. Let f be the underlying Thin-Plate Spline [2] based non-rigid transformation function, and the transformed template point set is $X'_r = x'_i = f(x_i) : i = 1, 2, ..., N$. Construct a correspondence matrix P to store the probabilities of each test point being assigned to each template point with dimension $(N+1) \times (M+1)$.

$$P = \left(\begin{array}{ccc|c} p_{11} & \cdots & p_{1M} & p_{1,M+1} \\ . & \cdots & . & . \\ . & \cdots & . & . \\ . & \cdots & . & . \\ p_{N1} & \cdots & p_{1M} & p_{1,M+1} \\ \hline p_{N+1,1} & \cdots & p_{N+1,M} & 0 \end{array}\right) \quad (1)$$

The inner $N \times M$ sub-matrix defines the probabilities of each x_i being assigned to y_j. The presence of an extra row and column in the matrix handles outliers in both point sets. Each p_{ij} is computed as

$$p_{ij} = \frac{1}{T_i} e^{-\frac{(y_j - f(x_i))^T (y_j - f(x_i))}{2T_i}} \quad (2)$$

Where $T_i : i = 1, 2, ..., N$ is the temperature of each template point cluster. For outlier clusters, the temperature T is kept at maximum throughout the annealing process. As discussed in Section 2.5, when T_i reaches T_{final} the correspondence is almost binary. If x_i is mapped to y_j then $p_{ij} \approx 1$. Similarly, if x_i is an outlier then $p_{i,M+1} \approx 1$, and if y_j is an outlier then $p_{N+1,j} \approx 1$. The matrix P satisfies the following row and column normalization conditions.

$$\begin{array}{l} \sum_{i=1}^{N+1} p_{ij} = 1, for\ j = 1, 2, ..., M,\ and \\ \sum_{j=1}^{M+1} p_{ij} = 1, for\ i = 1, 2, ..., N \end{array} \quad (3)$$

The goal of the framework is to find an optimal transformation matrix P' and the optimal transformation function f' that minimizes the energy function $E(P, f)$ as defined below.

$$\begin{array}{rcl} [P', f'] & = & \underset{P,f}{\text{argmin}}\ E(P, f), \\ E(P, f) & = & E_g(P, f) + \lambda E_s(f) + E_a(P), \text{ Where} \\ E_g(P, f) & = & \sum_{i=1}^{N} \sum_{j=1}^{M} p_{ij} ||y_j - f(x_i)||^2 \\ E_s(P, f) & = & \int\int \left[\left(\frac{\partial^2 f}{\partial u^2}\right)^2 + 2\left(\frac{\partial^2 f}{\partial u \partial v}\right)^2 + \left(\frac{\partial^2 f}{\partial v^2}\right)^2 \right] \\ E_a(P, f) & = & T \sum_{i=1}^{N} \sum_{j=1}^{M} p_{ij} \log p_{ij} - \zeta \sum_{i=1}^{N} \sum_{j=1}^{M} p_{ij} \end{array} \quad (4)$$

In the energy function E (Eqn. 4), $E_g(P, f)$ is the geometric feature-based energy term defined by Euclidean distance. $E_s(P, f)$ is the smoothness energy term with λ being the regularization parameter that controls smoothness of the transformation function. To favor rigid transformations at higher temperatures and local non-rigid transformation at lower temperatures, the framework reduces λ using an annealing schedule (i.e. $\lambda_i = \lambda_{init} T_i$ where λ_{init} is a constant, $i = 1, 2, ..., N$). $E_a(P, f)$ is a combination of two terms; the first term controls fuzziness of P and the last term prevents too many points being rejected as outliers.

The transformation function f uses TPS [2], which can be decomposed into affine and non-affine subspaces, thereby accommodating both rigid and non-rigid transformations.

$$f(x_i, d, w) = x_i.d + \phi(x_i).w \quad (5)$$

Where x_i is the homogeneous point representation of the 2D point x_i, d is a $(D+1) \times (D+1)$ affine transformation matrix of the D-dimensional image (For 2D images D=2), and w is a $N \times (D+1)$ warping coefficient matrix representing non-affine deformation. $\phi(x_i)$ is the TPS kernel of size $1 \times (N+1)$, where each entry $\phi_k(x_i) = ||x_k - x_i||^2 \log||x_k - x_i||$.

Algorithm 2 Enhanced TPS-RPM Pseudo code

Input: Template point set X_r, Test point set Y, and the correspondence set C''.
Output: Correspondence matrix P and transformation $f = d, w$.
Initialize: Temperature $T_i : i = 1, 2, ..., N$ of each template point cluster with the Euclidean distance between the template point and the corresponding test point y_j specified in C'', T_{final} as average of the squared distance between the nearest neighbors of the test points.
Initialize: smoothness parameter $\lambda_i \leftarrow \lambda_0 T_i : i = 1, 2, ..., N$
Initialize: d with identity matrix, P using eq. 2, and w with a zero matrix.
while $max(T_i) > T_{final}$ **do**
  repeat
    Update Correspondence: Compute P using eq. 2
    Normalize P using eq. 3 iteratively.
    Update transformation Update w and d using QR decomposition([5, 6])
  until P, d and w converged
  Update $T_i \leftarrow T_i\gamma$, update $\lambda_i \leftarrow \lambda_0 T_i; i = 1, 2, ..., N$; ($\gamma$ is the annealing rate)
end while

2.7. Refining New Correspondences

Even though we take the correspondence set into account, the set of correspondences after the Algorithm 2 contains new correspondences which are not in C'', as the set C'' contains correspondences of the dense points. The new correspondences introduced by TPS-RPM lead to inaccurate transformation of the ROI (blue box) (Example in Fig. 5(b)). To overcome this we refine the new correspondences with the correspondences of C'' that fall in the $h \times h$ window (yellow boxes) of the new correspondence (Fig.5(a)). We refine the registration parameters obtained in section 2.6 by minimizing the histogram of gradients error. The methodology is similar to enhanced RANSAC described in section 2.4. Fig.5(c) and Fig.5(d) show the correspondences after enhanced TPS-RPM and the projected ROI respectively.

3. Results and Discussion

We experiment with twelve template images falling in two different categories; one set is made of colored documents that have rich graphics and the other contains black and white documents that have minimal or no graphics (Fig. 6). Our test set consists of 480 images collected using two capturing devices (iPhone3GS and Logitech webcam Pro 9000)[1]. For each template image, we collected 20 test images with each of the two capturing devices. We consider SIFT and SURF features to evaluate different registration methodologies. The experiments use an AMD Athlon Dual core 2.69GHz machine with 1.75GB memory, taking on average 2 seconds to register each image.

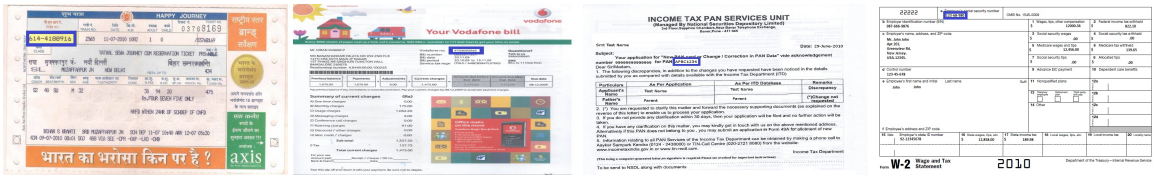
Five approaches have been compared in our experiments: RANSAC, RANSAC+Histogram, Enhanced RANSAC+Histogram, TPS-RPM, Enhanced TPS-RPM, and Enhanced TPS-RPM with refining new correspondences. In RANSAC we apply the RANSAC algorithm presented in Section 2.3 after obtaining initial correspondences(Section. 2.1). In RANSAC+Histogram, we apply RANSAC after refining the initial correspondences using histogram of Euclidean distances (Section 2.2). In Enhanced RANSAC+Histogram we apply the enhanced RANSAC algorithm presented in Section 2.4 after the Euclidean distance based histogram. For RANSAC, RANSAC+Histogram, and Enhanced RANSAC+Histogram we set threshold t of Lowe's approach to 0.9 and maximum RANSAC iterations to 100. In the case of the three methods based on TPS-RPM, we set Lowe's threshold to 0.6 and 0.8 for SIFT and SURF respectively to generate reasonably sparse points with enough correspondences. This difference in Lowe's threshold for SIFT and SURF comes from the fact that non-rigid registration depends on the selection of control points. Empirical evaluation shows that SIFT generates enough control points with small threshold value compared to SURF. Matching is restricted to points in the template image that fall in clusters close to an ROI. The set of template image points to be used for matching are selected in the following manner: (i) For each ROI, all clusters are marked as un-selected, (ii) As long as the number of match points for the ROI is less than 300, we mark the nearest un-selected cluster as selected and add points in this cluster to the set of matching points for the ROI.

Fig. 7 shows the performance of different registration methodologies with SIFT and SURF features on different template images. We measure the Registration accu-

[1]Data is available at URL: http://hidden for review

(a) (b) (c) (d)

Figure 5. Refining correspondences using enhanced TPS-RPM. (a) Correspondences after Enhanced TPS-RPM, (b)ROIs from correspondences of Fig.5(a), (c)Correspondences after refinement(Section 2.7), and (d) ROIs from correspondences of Fig. 5(c)



(a) (b) (c) (d)

Figure 6. Template images with blue rectangles as Regions of Interest (ROI). (a,b): Samples from colored documents with multiple graphics, and (c,d): Samples from black and white documents with no graphics

racies as *number of truly registered regions / total number of regions*. We find that Enhanced RANSAC+Histogram and Enhanced TPS-RPM with refinement of correspondences outperforms the other methods. Enhanced TPS-RPM with refinement performs slightly better than Enhanced RANSAC+Histogram as it has the advantage of non-rigid registration. TPS-RPM performs worst, which is likely due to its assumption of sparseness in the point sets. We find that in black and white images that are primarily white with sparse content (For example, W2 forms), SURF performs very poor on all the methods. Euclidean distance Histogram as a preprocessing step to RANSAC significantly improves the performance of RANSAC on all the test cases, and finally SIFT performs slightly better than SURF on all template types due to its superior repeatability. We find that repeatability becomes more critical in the case of non-rigid registration, where SIFT gives larger number of control points surrounding the ROI as compared to SURF.

4. CONCLUSIONS

This paper describes a framework for robust registration of camera captured document images. We embed four novel aspects into the framework; clustering of feature points using k-means, Histogram based outlier refinement to speed up iterative algorithms like RANSAC and TPS-RPM, enhanced RANSAC for robust registration of document images, and finally enhanced TPS-RPM with refined correspondences for registration of images under non-rigid deformation. Clustering of feature points enables selection of nearby regions for registration of ROI. Euclidean distance based Histogram not only eliminates outliers but also enhances the convergence rate of RANSAC. We find that enhanced RANSAC algorithm refines the global registration parameters to suit each ROI, accommodating non-affine deformations. Enhanced TPS-RPM incorporates prior knowledge of correspondences into TPS-RPM and leads to better registration of non-rigidly deformed images. One limitation of our method is that matching is applied to known regions of interest (ROI) in the template image. While this is a reasonable assumption for several document processing applications, it is not a valid assumption in general. We are exploring methods that derive a non-rigid transformation for the entire document image by considering each cluster of points in the template image as an ROI. Preliminary experiments show that the methodology is able to identify regions of interest in video frames. We are also working on parallel implementations of the algorithm to reduce execution time.

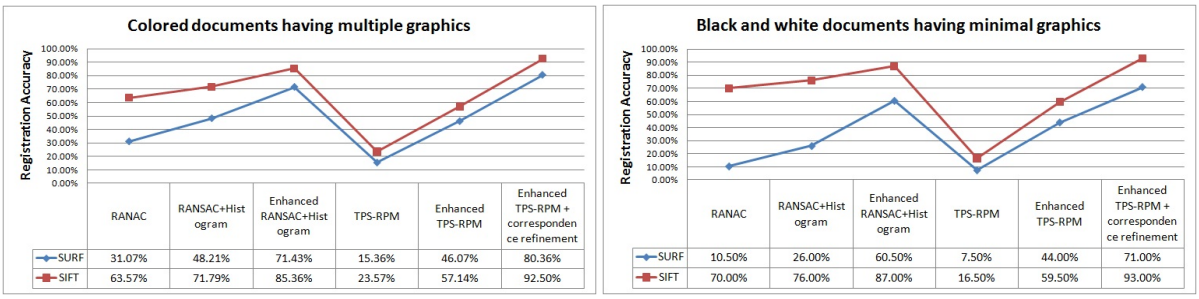
References

[1] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool. Speeded-up robust features (surf). *Computer Vision and Image Understanding*, 110(3):346–359, June 2008.

[2] F. L. Bookstein. Principle warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(6):567–585, June 1989.

[3] M. Calonder, V. Lepetit, and P. Fua. Keypoint signatures for fast learning and recognition. In *Proceedings of tenth European Conference on Computer Vision*, pages 58–71, Marseille, France, Oct. 2008.

[4] S. Choi, T. Kim, and W. Yu. Performance evaluation of ransac family. In *Proc. British Machine Vision Conference*, pages 1–12, London, Sept. 2009.



(a) (b)

Figure 7. Comparison of registration methodologies using SIFT and SURF point features on different image types.

[5] H. Chui and A. Rangarajan. A feature registration framework using mixture models. In *Proc. IEEE Workshop on Mathematical methods in Biomedical Image Analysis*, pages 190–197, Hilton Head Island, South Carolina, June 2000.

[6] H. Chui and A. Rangarajan. A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding*, 89(2-3):114–141, Feb. 2003.

[7] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. Piatko, R. Silverman, and A. Y. Wu. The analysis of a simple k-means clustering algorithm. In *Proc. the Sixteenth Annual Symposium on Computational Geometry*, pages 100–109, Hong Kong University of Science and Technology, June 2000.

[8] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Intl. Journal of Computer Vision*, 60(2):91–110, Nov. 2004.

[9] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. kadir, and L. V. Gool. A comparison of affine region detectors. *Intl. Journal of Computer Vision*, 65(1-2):43–72, Nov. 2005.

[10] A. Myronenko and X. Song. Point set registration: Coherent point drift. *IEEE Trans. On Pattern Analysis and Machine Intelligence*, 32(12):2262–2275, Dec. 2010.

[11] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance as a metric for image retrieval. *Intl. Journal of Computer Vision*, 40(2):99–121, Nov. 2000.

[12] R. Safari, N. Narasimhamurthi, M. Shridhar, and M. Ahmadi. Form registration: A computer vision approach. In *Proc. fourth Intl. Conference on Document Analysis and Recognition*, pages 758–761, Ulm, Germany, Aug. 1997.

[13] M. Sofka, Y. Gehua, and C. V. Stewart. Simultaneous covariance driven correspondence (cdc) and transformation estimation in the expectation maximization framework. In *Proc. IEEE Intl. Conference on Computer Vision and Pattern Recognition*, pages 1–8, Minneapolis, MN, June 2007.

[14] S. L. Taylor and R. Fritzson. Registration and region extraction of data from forms. In *Proc. eleventh Intl. Conference on Pattern Recognition*, pages 173–176, The Hague, Netherlands, Aug. 1992.

[15] J. Yang, R. S. Blum, J. P. Williams, Y. Sun, and C. Xu. Non-rigid image registration using geometric features and local salient region features. In *Proc. IEEE Intl. Conference on Computer Vision and Pattern Recognition*, pages 825–832, New York, NY, June 2006.