

A Featured Area-Based Image Registration

Youngsung Soh, Mudasar Qadir, Aamer Mehmood, Yongsuk Hae, Hadi Ashraf, and Intaek Kim

Abstract—Image registration is necessary when images from multiple viewpoints need be brought into common coordinate system. Image registration techniques can be classified as area-based methods and feature-based methods. In area-based methods, no features are selected and regularly tessellated areas are usually used for matching. In feature-based methods, features such as regions, lines, and prominent points are detected and used for matching. When image contains rich features, feature-based methods are preferred and when it does not, area-based methods are usually adopted. There are occasions where richness of features varies locally in the image. In this case, either area-based methods or feature-based methods alone may not generate successful results. In this paper, we propose a mixture of two methods termed as featured area-based method. In the proposed, we first tessellate the image into equal-sized areas, estimate richness of features of each area utilizing the edge direction histogram, choose only those areas with a certain level of richness, and use them for matching. We compared the proposed with well-known conventional methods such as Kanade-Lucas-Tomasi (KLT) method, speeded up robust features (SURF), and scale-invariant feature transform (SIFT), and showed that the proposed performs better than others.

Index Terms—Image registration, KLT, SIFT, SURF.

I. INTRODUCTION

Image registration is the process where multiple images containing (possibly distorted) various extent of 3D scene are brought into common coordinate system. Multiple images can be generated from multiple viewpoints, multiple (same or different) sensors, a single sensor with mirror(s), etc. There are several areas where image registration techniques can be used. They include image mosaicking and image rectification for stereo analysis. Zitova *et al.* [1] and Deshmukh *et al.* [2] gave a good review of image registration techniques. In [1] image registration techniques are classified as area-based methods and feature-based methods. In area-based methods, no features are selected and regularly tessellated areas are usually used for matching. Area-based matching methods can be categorized into three groups. They are cross correlation (CC)-like methods, Fourier transform-based methods, and mutual information methods.

CC-like methods include normalized cross correlation (NCC), sum of squared difference (SSD), sum of absolute difference (SAD), and structural similarity (SSIM). Fourier

transform-based methods include phase correlation and extended phase correlation [3]. Mutual information methods make use of joint and conditional entropy concepts. In feature-based methods, various features were used. They are regions [4], [5], lines [6]-[10], and points [11]-[13]. For feature-based matching, methods using spatial relations or invariant descriptors, relaxation method, and wavelets were proposed.

When image contains rich features, feature-based methods are preferred and when it does not, area-based methods are usually adopted. There are occasions where richness of features varies locally in the image. In this case, either area-based methods or feature-based methods alone may not generate successful results. In this paper, we propose a mixture of two methods termed as featured area-based method. In the proposed, we first tessellate the image into equal-sized areas, estimate richness of features of each area utilizing the edge direction histogram, choose only those areas with a certain level of richness, and use them for matching.

The paper is organized as follows. In Section II, related works for feature selection and matching are discussed. Section III presents the proposed method. Results are described in Section IV and Section V concludes the paper.

II. RELATED WORKS

In this section, we describe several well-known conventional techniques for feature selection and matching. They are KLT, SURF with fast library for approximate nearest neighbors (FLANN), SURF with brute force matching (BF), and SIFT with random sample consensus (RANSAC).

A. KLT

Lucas *et al.* [14] proposed KLT method for both feature selection and matching. For feature selection, gradient matrix G of each pixel is constructed, non-singularity of G is tested, and its eigenvalues are computed. If minimum eigenvalue of G exceeds predefined local and global thresholds, then that pixel is selected as a feature.

For feature matching, residual error function that measures the squared difference of selected window in the reference image and the corresponding candidate window in the target image is formulated. To minimize the error function, it is differentiated with respect to horizontal and vertical optical flow components and they are set to zero. First order Taylor expansion is used to solve the equation. Due to the use of first order Taylor expansion, standard KLT was valid only for small pixel displacement. To remedy iterative version of KLT was introduced.

Manuscript received December 5, 2013; revised February 27, 2014.

The authors are with the Myongji University, Yongin, 449-728, Korea (e-mail: soh@mju.ac.kr, mudasar.kalwar@gmail.com, aamergcu@hotmail.com; wise_sunys@nate.com, nothan111@gmail.com, kit@mju.ac.kr).

B. SURF with FLANN

Bay *et al.* [15] proposed SURF method for feature selection. It is based on three steps. They are interest point detection, descriptor generation, and descriptor matching. In interest point detection step, integral images are formed for fast computation with box type filters. Hessian matrix is built by adopting approximated second order derivatives with box filters. Scale space is formed and non-maximum suppression and subpixel accuracy interpolation are used to detect blob-like features. To generate descriptors, orientation assignment is performed, dominant orientation is selected, and Haar wavelet is used to generate 64 dimensional normalized feature descriptors. Finally matching is done by comparing the sign of Laplacian and finding nearest neighbor using Euclidean distance.

Muja *et al.* [16] proposed FLANN method. They introduced fast nearest neighbor search technique based on priority search in multiple hierarchical clustering trees. It is used with SURF for matching the descriptor vector from reference image with the one from target image.

C. SURF with BF

SURF method described in section II.B is used for feature selection. Brute-force matcher is used for matching the descriptor vector provided by SURF between the reference and the target images using K nearest neighbor technique effectively. This tries all the possibilities to find the best match.

D. SIFT with RANSAC

Lowe *et al.* [17] proposed SIFT method. It is based on four steps. They are key point detection, key point localization, orientation assignment, and key point descriptor generation. Key point detection step detects scale-space extrema using DoG instead of Gaussian to speed up the process. In key point localization step, Hessian matrix is adopted to compute the principal curvatures and low contrast points are eliminated by comparing computed curvatures. Next, gradient orientation histogram is formed and used to make features rotation-invariant. In the last step, 4x4 array of histograms with 8 bins is used to produce 128 dimensional feature descriptor vectors.

Fischler *et al.* [18] proposed RANSAC technique. First hypothetical inliers are randomly selected from data set, a model is fitted to them, and model parameters are estimated. All other data are checked for possible inclusion in the model. If the estimated model contains sufficient number of data, then it is considered good. Otherwise random hypothetical inlier selection is performed again and the above procedure is repeated. The process stops when the model contains sufficient number of data and total fitting error is below some threshold.

III. THE PROPOSED METHOD

The proposed method assumes that there are two images, the reference and the distorted, and tries to bring the distorted into the coordinate of the reference. The proposed method consists of three steps. They are image tessellation, selection of featured area by measuring feature richness, and matching.

Detailed description of each step follows.

A. Image Tessellation

In this step, the reference image is tessellated into equal-sized areas as in Fig. 1. Those areas located in the periphery of the image (shown in solid white box in Fig. 1) are not used since they may not be visible in the distorted due to various kinds of motion and/or geometric distortion.

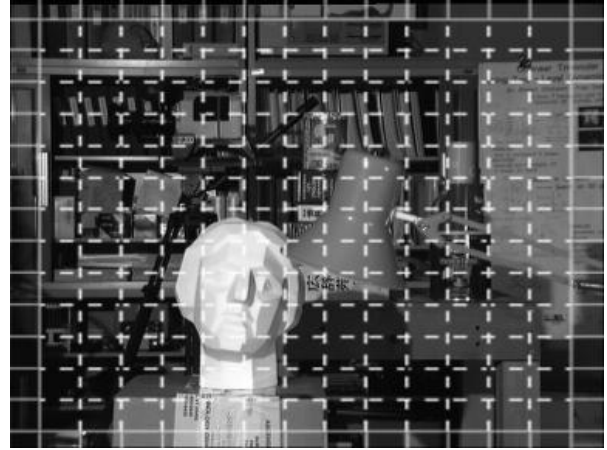


Fig. 1. Image tessellation.

B. Selection of Featured Area by Measuring Feature Richness

In this step, we try to select the area which has enough detail in it. By enough detail we mean that the brightness pattern of the area is not uniform and the structure in the area is not uniform either. For example, we reject the areas containing simple single directional structures like table legs since there may be many matches in the distorted. First Sobel operator is applied to each pixel in the reference image in order to get the horizontal and vertical edge strength and gradient directions. If the absolute values of horizontal and vertical edge strength are greater than some threshold, the pixel is considered as a strong pixel. For each area, if the area contains enough number of strong pixels, it becomes a candidate featured area. Next the gradient direction histogram is built for strong pixels in the candidate area. We use 19 bins for directions ranging from 0° to 180°. Thus each bin is of width 10°. Absolute value is taken for negative directions. If the cumulative sum of any three consecutive bins does not exceed a predefined threshold, then that area is finally selected as a featured area. This is for rejecting areas with simple single directional structures. For single directional structures, we allow up to 30° difference in directions to account for possible distortions such as rotation and aliasing in digitization. The number three for consecutive bins is empirically chosen.

C. Matching

For matching, SAD in (1), is adopted. SAD is calculated for each featured area in the reference image I_{ref} and the corresponding area in the distorted image $I_{distorted}$. The area in the distorted with smallest SAD is chosen as a match.

$$\sum_{(i,j) \in W} |I_{ref}(i,j) - I_{distorted}(x+i, y+j)| \quad (1)$$

IV. EXPERIMENTAL RESULTS

In this section, we compare the proposed with other conventional methods as in Fig. 2. Given a reference image R, we geometrically transform it using OpenCV function with a known affine parameter set to get the distorted image D as in (2).

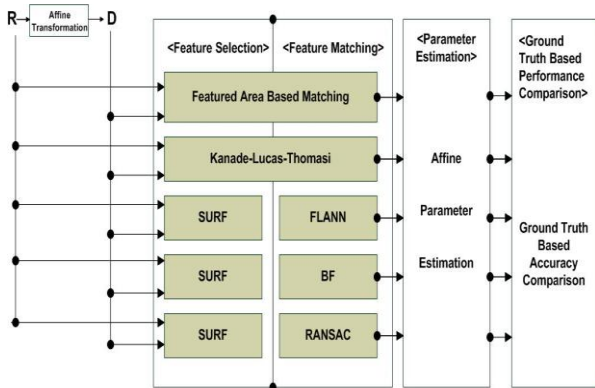


Fig. 2. Performance comparison procedure.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \end{bmatrix} \quad (2)$$

where $[x, y]$ and $[x', y']$ represent a reference point in R and a distorted point in D respectively. This parameter set will serve as a ground truth. Next we let each method get two input images R and D and do feature selection and matching to produce matching point pairs. Then these pairs are fed into affine parameter estimation program provided by OpenCV to get estimated parameter set for each method. This estimated parameter set will be compared with ground truth. The estimation accuracy of each parameter is computed using (3). The final accuracy is the sum of accuracies of all six parameters. Thus the accuracy of 600% is the best.

$$Accuracy = \frac{G-|G-E|}{E} \times 100 \% \quad (3)$$

where G and E represent ground truth value and estimated value respectively. Since the proposed is not scale and rotation invariant, we first test several affine parameter sets having only translation components. Later we release the parameter set to include rotation and scale components to verify the tolerance of the proposed. The first three affine parameter sets are as follows.

Test set 1

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 5 \\ 2 \end{bmatrix}$$

Test set 2

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} -5 \\ -2 \end{bmatrix}$$

Test set 3

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 5 \\ -2 \end{bmatrix}$$

Tables I, II, and III show the accuracy results of Test sets 1, 2, and 3 respectively. Each Table has seven columns for accuracy results. Each of the first six shows the accuracy of each affine parameter. The final column has overall accuracy. As can be seen in Tables I, II, and III, the proposed shows the “PERFECT” performance. Among conventional methods SIFT with RANSAC shows almost perfect result, whereas KLT performs a little worse than SIFT and SURF-based methods perform worst. As shown in the accuracy of each individual parameter, SURF-based methods have a difficulty in generating point pairs that give good estimation of translation parameters.

TABLE I: ACCURACY FOR TEST SET 1 (UNIT:%)

	A ₁₁	A ₁₂	A ₂₁	A ₂₂	T _x	T _y	Total
Proposed method	100	100	100	100	100	100	600
KLT	99.9	100	100	99.92	99.99	96.54	596.45
SURF with FIANN	99.94	100	100	99.93	97.47	91.42	588.77
SURF with BF	99.94	100	100	99.94	97.19	92.67	589.76
SIFT with RANSAC	99.97	100	100	99.98	99.66	99.38	599.01

TABLE II: ACCURACY FOR TEST SET 2 (UNIT:%)

	A ₁₁	A ₁₂	A ₂₁	A ₂₂	T _x	T _y	Total
Proposed method	100	100	100	100	100	100	600
KLT	99.95	100	100	99.99	96.74	97.38	594.08
SURF with FIANN	99.89	100	100	99.92	93.92	92.65	586.4
SURF with BF	99.88	100	100	99.93	93.79	91.64	585.23
SIFT with RANSAC	99.99	100	100	99.98	99.89	99.67	599.56

TABLE III: ACCURACY FOR TEST SET 3 (UNIT:%)

	A ₁₁	A ₁₂	A ₂₁	A ₂₂	T _x	T _y	Total
Proposed method	100	100	100	100	100	100	600
KLT	99.90	100	100	99.99	99.45	98.79	598.15
SURF with FIANN	99.9	100	100	99.94	96.23	92.25	588.34
SURF with BF	99.94	100	100	99.94	96.98	92.63	589.5
SIFT with RANSAC	99.99	100	100	99.99	99.84	99.93	599.77

Next we compare the performance when there are scale and rotation components in addition to translation components. The following two data sets are used.

Test set 4

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 0.98 & -0.015 \\ 0.015 & 0.98 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 5 \\ 2 \end{bmatrix}$$

Test set 5

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1.01 & 0.0075 \\ 0.0075 & 1.01 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 5 \\ 2 \end{bmatrix}$$

Tables IV and V depict the accuracy of Test sets 4 and 5 respectively. The proposed and SIFT with RANSAC show comparably good performance, whereas other methods perform worse. Among KLT and SURF-based methods, KLT performs a little bit better as in Tables I, II, and III. As was mentioned earlier, the proposed is not scale and rotation invariant. However test shows that the proposed can tolerate

small degree of rotation and scale due to excellent selection of areas for matching.

The image used for the experiment is from Tsukuba stereo image database [19]. We performed the same experiment for other images and got similar results.

TABLE IV: ACCURACY FOR TEST SET 4 (UNIT:%)

	A_{11}	A_{12}	A_{21}	A_{22}	T_x	T_y	Total
Proposed method	99.99	96.39	99.84	99.94	99.31	96.90	592.36
KLT	99.89	85.17	99.22	99.96	95.8	97.21	577.25
SURF with FIANN	99.85	95.85	97.56	99.89	90.81	90.76	574.71
SURF with BF	99.87	94.89	96.85	99.88	90.85	89.19	571.55
SIFT with RANSAC	99.91	97.86	98.7	99.89	97.44	96.05	589.85

TABLE V: ACCURACY FOR TEST SET 5 (UNIT:%)

	A_{11}	A_{12}	A_{21}	A_{22}	T_x	T_y	Total
Proposed method	99.97	98.96	93.53	99.99	99.02	98.60	590.09
KLT	99.94	99.12	93.24	99.93	97.86	99.12	589.21
SURF with FIANN	99.86	96.54	92.58	99.99	95.87	97.38	582.24
SURF with BF	99.87	95.1	92.3	99.97	95.52	95.94	578.7
SIFT with RANSAC	99.99	96.04	98.24	99.97	99.73	97.97	591.94

V. DISCUSSION

Image registration is necessary when images from multiple viewpoints need be brought into common coordinate system. In this paper we propose a mixture of area-based and feature-based methods, termed as featured area-based method. We first regularly tessellate the reference image into equal-sized areas, choose areas that satisfies feature richness criterion, and do matching in the distorted image using SAD. We obtained better results than conventional methods for various kinds of affine geometric distortions. The quality of image registration greatly depends on the selection of matching points pairs. The proposed shows the superior ability in selecting matching entities.

Though the proposed shows good tolerance for some affine transform settings containing rotation and scale components, it is inherently scale and rotation variant. We need to obtain the invariance and it is intended for future research.

ACKNOWLEDGEMENT

This work (Grants No. C0005448) was supported by Business for Cooperative R&D between Industry, Academy, and Research Institute funded by Korea Small and Medium Business Administration in 2012.

REFERENCES

[1] B. Zitova and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, pp. 977-1000, 2003.
 [2] M. P. Deshmukh and U. Bhosle, "A survey of image registration," *International Journal of Image Processing (IJIP)*, vol. 5, 2011.

[3] E. D. Castro and C. Morandi, "Registration of translated and rotated images using finite Fourier transform," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, pp. 700-703, Sept. 1987.
 [4] J. Flusser and T. Suk, "A moment-based approach to registration of images with affine geometric distortion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, pp. 382-387, Mar. 1994.
 [5] M. Roux, "Automatic registration of SPOT images and digitized maps", in *Proc. the IEEE International Conference on Image Processin*, Lausanne, Switzerland ICIP press, 1996, vol. 2, pp. 625-628.
 [6] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, pp. 679-698, Nov. 1986.
 [7] H. Li, B. S. Manjunath, and S. K. Mitra, "A contour-based approach to multisensor image registration," *IEEE Transactions on Image Processing*, vol. 4, pp. 320-334, Mar. 1995.
 [8] D. Marr and E. Hildreth, "Theory of edge detection", *Proceedings of the Royal Society of London*, vol. 207, pp. 187-217, Feb. 1980.
 [9] S. Moss and E. R. Hancock, "Multiple line-template matching with EM algorithm," *Pattern Recognition Letters*, vol. 18, pp. 1283-1292, 1997.
 [10] D. Shin, J. K. Pollard, and J. P. Muller, "Accurate geometric correction of ATSR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, pp. 997-1006, Jul. 1997.
 [11] S. Banerjee, D. P. Mukherjee, and D. D. Majumdar, "Point landmarks for registration of CT and NMR images," *Pattern Recognition Letters*, vol. 16, pp. 1033-1042, 1995.
 [12] D. Bhattacharya and S. Sinha, "Invariance of stereo images via theory of complex moments," *Pattern Recognition*, vol. 30, pp. 1373-1386, 1997.
 [13] Q. Zheng and R. Chellapa, "A computational vision approach to image registration," *IEEE Transactions on Image Processing*, vol. 2, pp. 311-325, Jul. 1993.
 [14] B. D. Lucas and T. Kanade, "An Iterative Image Registrstion Technique with an Application to Stereo Vision," in *Proc. International Joint Conference on Artificial Intelligence*, 1981, pp. 674-679.
 [15] H. Bay, A. E. T. Tuytelaars, and L. V. Gool, "SURF: Speeded Up Robust Features," *Computer Vision and Image Understanding (CVIU)*, vol. 110, pp. 346-359, 2008.
 [16] M. Muja and D. G. Lowe, "Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration," in *Proc. International Conference on Computer Vision Theory and Applications VISAPP*, Liasbon, Portugal, Feb. 2009, pp. 331-340.
 [17] D. G. Lowe, "Distinctive image features from scale invariant keypoints," *Int. J. Computer Vision*, vol. 60, issue 2, pp. 91-110. Nov. 2004.
 [18] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Comm. of the ACM*, pp. 381-395, 1981.
 [19] CVLAB. [Online]. Available: <http://www.cvlab.cs.tsukuba.ac.jp/dataset/tsukubastereo.php>



Youngsung Soh was born in Seoul, Korea on Mar. 4, 1956. He got BS in electrical engineering in 1978 from Seoul National University in Seoul, Korea. He obtained MS and PhD in computer science from the University of South Carolina in Columbia, South Carolina, USA in 1986 and 1989, respectively.

He served in the Korean army from June 1980 to Sept. 1982. He worked in Systems Engineering Research Institute in Korea as a senior researcher from Sept. 1989 to Feb. 1991. He joined Myongji University in Korea from Mar. 1991 and is currently a full professor in the Dept. of Information and Communication Engineering.

His current interest of research includes object tracking, stereo vision, and parallel algorithms for image processing.

Prof. Soh is a member of Korea Information Processing Society and Korea Signal Processing Systems Society.



Mudasar Qadir was born in Ghotki, Pakistan on Dec. 24, 1989. He got BS in computer science in 2011 from FAST-NU University in Peshawar, Pakistan. He entered a master course in information and communication engineering in Myongji University in 2013.

He worked in different software companies as a software engineer from Jul. 2011 to Feb. 2013.

His current interest of research includes object

tracking, stereo vision and parallel algorithms for image processing.



Aamer Mehmood was born in Kasur, Pakistan on May 1, 1988. He got BS in electrical engineering in 2011 from Govt. College University in Lahore, Pakistan. He entered a master course in information and communication engineering in Myongji University in 2012.

His current interest of research includes object extraction, stereo vision and parallel algorithms for image processing.



Yongsuk Hae was born in Mokpo, Korea on August 8, 1981. He got BS and MS in information and communication engineering from Myongji University in Yongin, Korea in 2009 and 2012, respectively. He entered a PhD course in information and communication engineering in Myongji University in Yongin, Korea in 2012.

He served in the Korean army from August 2001 to October 2003. Since December 2008, he has been working in Nain information company in Korea as a researcher.

His current interest of research includes object tracking, stereo vision, and parallel algorithms for image processing.



Hadi Ashraf was born in Lahore, Pakistan on Dec. 8, 1988. He got BS in electrical engineering in 2010 from Govt. University in Lahore, Pakistan. He entered a master course in information and communication engineering in Myongji University in 2012.

He joined a software company in Pakistan in Aug. 2010 and worked as a software engineer till April 2012.

His current interest of research includes object tracking, stereo vision and parallel algorithms for image processing.



Intaek Kim was born in Seoul, Korea in 1960. He received BS and MS in electronics engineering from Seoul National University in Seoul, Korea in 1980 and 1984 respectively. He obtained PhD in electrical engineering from Georgia Institute of Technology in Atlanta, Georgia, USA in 1992.

He worked for Goldstar central research lab from 1993 to 1995 as a senior engineer and joined Myongji University from 1995. He is now a professor in the Dept. of Information and Communication Engineering. His recent publications deal with the area of face recognition, hypersepctral image and MR imaging.

His research interest includes pattern recognition, image processing and smart grid area.

Prof. Kim is a member of Korean Institute of Electronics Engineer.