Image Matching Techniques for Vision-based Indoor Navigation Systems: Performance Analysis for 3D Map Based Approach

Xun Li, Jinling Wang
School of Surveying and Geospatial Engineering
University of New South Wales
Sydney, Australia
xun.li@student.unsw.edu.au

Abstract— **In this paper, we give an overview of image matching techniques for various vision-based navigation systems: stereo vision, structure from motion and map-based approach. Focused on map-based approach, which generally uses feature-based matching for localization, and based on our early developed system, a performance analysis has been carried out and three major problems have been identified: being vulnerable to illumination changes, drastic viewpoint changes and good percentage of mismatches. By introducing ASIFT into the system, the major improvement takes place on the epoch with large viewpoint changes. In order to deal with mismatches that are unable to be removed by RANSAC, we propose to use cross-correlation information to evaluate the quality of homography model and help select the proper one. The conducted experiments have proved that such an approach can reduce the chances of mismatches being included by RANSAC and final positioning accuracy can be improved.**

Keywords-component; image matching; vision; navigation; cross-correlation

I. INTRODUCTION

Image matching techniques have been used in a variety of applications, such as 3D modelling, image stitching, motion tracking, object recognition and vision based localization. Over the past few years, many different methods have been developed, which can be generally classified into two groups: area-based matching (intensity based, like cross-correlation and least-squares matching [1]) and feature-based matching (e.g. SIFT [2]). Area-based methods [3], may be comparatively more accurate, because they take into account a whole neighbourhood around the imagine points being analysed to establish correspondences. Feature based methods on the other hand uses symbolic descriptions of the images that contain certain local image information to establish correspondence. No single algorithm, however, has been universally regarded as optimal for all applications since they all have their pros and cons. In this paper, we explore the performance of various image matching techniques according to the specific needs of vision-based positioning systems for indoor navigation applications.

The main purpose of vision-based navigation is to determine the position (and orientation) of the vision sensor carried by the moving platform, then mobile vehicle‘s motion/trajectory can be recovered. However, it is not without its limitation. Vision sensor can measure relative position with derivative order of 0 but senses only a 2D projection of the 3D world – direct depth information is lost. Several approaches have been made to tackle such a problem. One way is to use stereo cameras by which the distance to a landmark can be directly measured. Another way is to use monocular vision with the integration of data from multiple viewpoints (e.g. structure from motion), or rely on the prior knowledge of the navigation environment such as maps, or models. Despite the variety form of vision-based navigation systems, they all need the same basic but essential function to support their self-localisation mechanism: image matching. The way they use such function is however different, which in terms affect their choice of image matching methods.

For stereo vision based approaches, stereo matching is employed to create a depth map (i.e. disparity map) for navigation. Area based algorithms solve the stereo correspondence problem for every single pixel in the image. Therefore, these algorithms result in dense depth maps as the depth is known for each pixel [4]. Typical methods include Census [5], SAD (Sum of Absolute Differences), and SSD (Sum of Squared Differences). The common drawback is that they are computational demanding. To deal with the problem, some efforts have been made. In [19] the authors proposed a quad-camera based system which used a custom tailored correspondence algorithm to keep the computation load within reasonable limits. Meanwhile, feature based methods are less error sensitive and require less work load. But the resulting maps will be less detailed as the depth is not calculated for every pixel [4]. Therefore, how to achieve a disparity map which is both dense, accurate while the system maintains reasonable refresh rate is the cornerstone of its success, and still remains to be an open question.

For monocular vision sensor, the two approaches also differ from each other. For structure from motion (SFM), consecutive frames present a very small parallax and small camera displacement. Given the location of a feature in one frame, a common strategy for SFM is to use feature tracker to find its correspondence in the consecutive image frame. Kanade-Lucas-Tomasi (KLT) tracker [7] is widely used for small

978-1-4673-1954-6/12/$31.00 ©2012 IEEE

baseline matching. The methodology for a feature tracker to track interest points through image sequence is usually based on a combined use of feature-based and area-based image matching methods. First interest points are extracted by operators from the first image, such as [2, 8, 9]. Then due to the very short baseline, positions of corresponding interest points in the second image are predicted and matched with cross-correlation, which can be further refined using least squares matching. Some approaches perform outlier rejection based either on epipolar geometry [10] or RANSAC [7] for the last step.

For the map based approach, first a map in the form of image collection (or model) is built by a learning step. Then self-localisation is realized by matching the query image with the corresponding scene in the database/map, whenever a match is found, the position information of this reference image is transferred directly to the query image and used as user position. A major difference between map-based approach and two previous methods lies in that the real time query image might be taken at substantially different viewpoint, distance or different illumination conditions from the map images in the database. Besides, the two images might be taken using different optical devices. In other words, the two matching images may have a very large baseline, large scale difference and big perspective effects, which lead to a wide range of image transformation while transformation parameters are unknown. Due to such significant changes, most image corresponding algorithms working well for short baseline (e.g. stereo, or video sequence) images will fail in this case. For area-based approach, cross correlation method can't get a good performance when rotation is greater than 20° or scale difference is greater than 30% [11]; an iterative search for LSM will require a good initial guess of the two corresponding locations, which is not applicable in situations where image transformation parameters are unknown. Moreover, early matching methods based on corner detectors [8] would fail because of the big perspective effects [10]. Therefore, more distinctive and invariant features are needed. The first work in the area was by Schmid and Mohr [12] who used a jet of Gaussian derivatives to form a rotationally invariant descriptor around a Harris corner. Lowe extended this approach to incorporate scale invariance [2]. Then, these newly developed invariant features were applied to image matching systems [13,14,15] for accurate location estimation.

Map based visual system using invariant feature matching for localisation is a common approach in today's research domain. The most popular algorithm is Lowe's SIFT. However, certain limitations still exist. In this paper, we mainly discuss the performance of various image matching methods used by such systems and their influence on final positioning. The aim is to find the bottleneck and possible improvement. Moreover, we propose to integrate intensity-based method into the feature matching process to strengthen the robustness of the matching algorithm against mismatches and noise.

The paper is constructed as follows: the first section discusses image matching methods in the context of vision-based navigation systems and the selection of different algorithms to cope with different needs of various systems; in the second section, limitations of general image matching method used in map-based visual systems have been revealed through evaluation based on our own vision-based positioning system, as being vulnerable to illumination changes, drastic viewpoint changes and mismatches (after using RANSAC [16]); in the third section ASIFT [6] is introduced into the system to address viewpoint changes and the next section we propose SIFT based method with cross correlation information to reduce the chance mismatches to be included for positioning; we present our conclusion and final thoughts in the last section.

II. MAP-BASED VISUAL POSITIONING WITH THE USE OF GEO-REFERENCED SIFT FEATURES

A. System Methodology

The development of map-based positioning and navigation system mainly consists of two steps: mapping, positioning and navigation, and both contains image matching procedure. The SIFT features are invariant to image translation, scaling, rotation, and partially invariant to illumination changes and affine or 3D projection, therefore we believe SIFT is a suitable choice for image matching at positioning stage. Due to the nature of our system's methodology: the very same geo-referenced features need to be matched for positioning; the matching for tie point extraction at the mapping stage has to be consistent with the later one.

More detailed explanation of the system is as follows. First, mapping is carried out. Images of the navigational environment are collected and SIFT matching between images with overlapped areas is performed. The aim is to produce geo-referenced images of the navigation environment. More specifically, SIFT feature points on map images will be geo-referenced through photogrammetric bundle adjustment (indirect geo-referencing). Two major inputs of bundle adjustment are ground control points and tie points. The first dataset come from ground control survey and image measurement of these control points, while the second are common SIFT feature points produced by the previous matching process. The quality of the map depends on the accuracy of geo-referencing. At the real time positioning stage, when real time images are taken by the vision sensor mounted on the (moving) vehicle, another image matching based on SIFT is carried out between the real time image and the map images. When any of the SIFT feature points from the map image(s) finds its correspondence on the real time image, the geo-information it carried can be transferred to its counterpart. Therefore, matched SIFT features on the real time image obtain both image coordinates from matching process and 3D coordinates from map images, which can later serve as pseudo ground control points (PGCPs) for space resection based positioning at the final stage. More specifically, modified space resection is utilized to calculate vision sensor’s external orientation in 6DOF. Outliers including mismatches are detected and removed by the system outlier detection mechanism based on RANSAC.

B. *Evaluating the Performance of SIFT Matching for the Image-based Positioning System*

A controlled experiment is designed to evaluate the performance of SIFT matching for the image-based positioning system. Major factors that influence image matching and their impact on final positioning have been investigated: illumination and viewpoint changes. On top of this, mismatches, which has long been a bottleneck for visual systems have been studied as well. First, mapping is performed in the target environment. All geo-referenced map images were taken with adequate lighting and viewing direction perpendicular to the wall (mapping area with visual features). Then a calibrated CCD camera (Canon EOS4500) with a fixed focal length at 24.18 mm was used as vision sensor of the positioning system. On positioning stage, three stable camera sites were deployed facing different mapping areas with X, Y, Z coordinates of the three camera stations surveyed by a total station, and angular changes at each camera site roughly measured. A total 8 pairs of images (16 in total) were taken at the 3 sites, with each pair consists of an image with adequate lighting and the other which covers the same scene but receives limited lighting. The performance of image matching is compared, both before and after RANSAC processing. Furthermore, the number of PGCP generated by each matching process is also studied along with the geometry of these points, which will directly affect the precision of positioning. Finally a manual check for the PGCP locations from the two matched images is performed to verify the correctness of image matching after RANSAC.

TABLE I. PERFORMANCE OF SIFT MATCHING IN THE SYSTEM

Site ID	ID	Angular change	All SIFT Matc-hes	Reliable Matches	Percent of Reliable Matches	Num of PGCPs	Num of False PGCPs	PDOP	ADOP
	1	0	578	202	34.95 %	51	1	2399	583
	2	0	515	141	27.38 %	28	0	2575	622
	3	-30	372	103	27.76 %	32	2	1149	223
	4	-30	357	119	33.45 %	34	1	1072	208
	5	50	267	59	22.19 %	2	0		
1	6	50	210	37	17.38 %	1	1		
	7	0	427	145	34.02 %	35	0	2106	530
	8	0	346	96	27.75 %	24	0	2795	701
	9	-20	401	193	48.07 %	59	1	969	235
2	10	-20	386	145	37.56 %	60	2	195	50
	11	0	417	112	26.74 %	11	0	5711	1177
	12	0	295	56	19.07 %	9	0	822	4034
	13	-30	457	146	32.00 %	22	0	372	2133
	14	-30	340	66	19.41 %	6	0	424	2471
	15	20	241	65	26.87 %	5	0	5813	28658
3	16	20	206	47	22.69 %	5	0	6624	32951

Firstly, it can be observed with ease, in each pair the image with good lighting condition (e.g. No.1) is able to find more common SIFT matched features when matched with reference map image than its counterpart with limit lighting (e.g. No.2). As a result, more PGCPs are generated and a better geometry (smaller DOP values) is provided. This test proves that lighting variation will influence the precision of final positioning by its impact on the geometric strength of our adjustment system. For image-based positioning& navigation systems alike, which depend on visual information and image matching techniques for localization, one limitation is that illumination changes, which is especially common for outdoor environment, may affect a navigation solution. The reason behind is that most existing local descriptors including the SIFT are based on luminance information rather than color information. Some color descriptors have been proposed recently to increase illumination invariance like C-SIFT [20], which can be studied further. It also noted that in an indoor vision-based positioning scenario, lighting condition can be easily controlled and kept in consistency.

The third column in Table 1 indicates the angular changes of viewpoint at κ (around Z-axis) for each camera site, other angles remain stable. Assuming the viewing direction perpendicular to the mapping area (wall with geo-referenced features) to be 0, a clockwise rotation to be positive changes. It's easy to note that the only epochs that fails to give a positioning result is the pair with the most drastic angular change, real time image No.5 (6). Not only does the total number of SIFT matches decreases, the percentage of correct matches filtered using RANSAC has also been reduced. Actually it is the pair with lowest correct rate. As a result, too few PGCPs are generated for positioning.

Figure 1. Real time image No.5

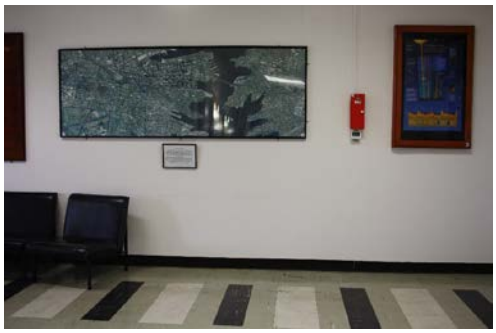

Figure 2. Real time image No.11

For better comparison, we choose two epochs with similar coverage of the scene: No.5 and No.11 shown in Fig.1&2, both of which were taken under amble light. It can be easily observed that Image No.5 include more features, but less SIFT matches were found. Moreover, the correct rate of No.5 is lower than that of No.11. It indicates that when the two matched images suffer from large viewpoint variation, less SIFT matches will be found, and there's a higher chance to generate false matches. But if we take a look at other images with smaller angular changes, such role does not apply. The reason is that the performance of SIFT based matching only drops under substantial viewpoint changes.

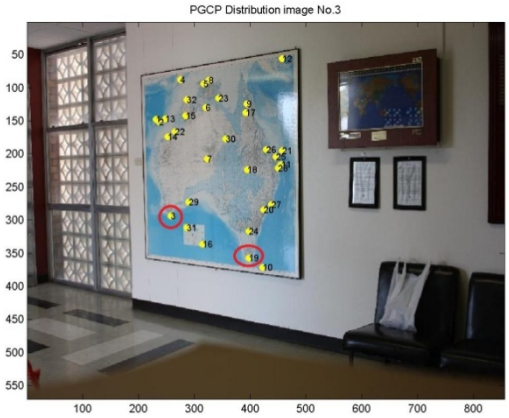

Figure 3. Real time query image No.3 with PGCPs,false PGCPs have been circled.

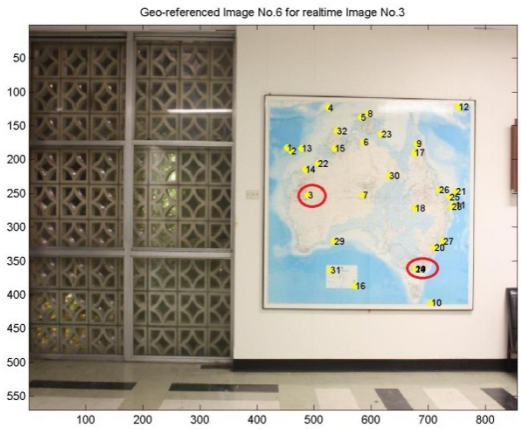

Figure 4. Map image No.6 with correspondences of PGCPs on query image No.3, false correspondences have been circled

Thirdly, it is noticed that after using RANSAC to reject mismatches, there is still a small chance that mismatches been left untreated, which might later generate false PGCP to jeopardise the final positioning process. As shown in Fig. 3 &4, when the real time query image No.3 is matched with map image No.6, 32 PGCPs are generated from the reliable matches provided by RANSAC. However, 2 false matches were still been spotted during manual check.

In summary three major weaknesses for image matching in the system have been found: invariant feature matching could not deal with drastic illumination changes and large viewpoint shift, furthermore, the current popular outlier detection mechanism RANSAC cannot guarantee the correctness of every pair. These three problems affect the final positioning by deteriorating the precision from inadequate number of matches or bringing in false matches.

III. USING ASIFT FOR VIEWPOINT CHANGES

In order to tackle the problem for unsatisfactory performance of SIFT subject to dramatic viewpoint distinction, some approaches have been recently proposed by some researchers to extend scale and rotation invariance to affine invariance, such as MSER [17] and Harris / Hessian Affine [18]. Although these methods have been proved to enable matching with a stronger viewpoint change, all of them are prone to fail at a certain point [19]. A better idea is to simulate viewpoint changes in order to reach affine invariance, the most successful algorithm using such method is named ASIFT (affine-SIFT). It is introduced by Morel and Yu in 2009 [6] to explicitly deal with extreme angle changes (up to 36 and higher). SIFT is only partially invariant to viewpoint changes because it is invariant to four out of the six parameters of an affine transform. Affine-SIFT (ASIFT), on the other hand, simulates all image views obtainable by varying the two camera axis orientation parameters, namely, the latitude and the longitude angles, left over by the SIFT method. Then it covers the other four parameters by using the SIFT method itself [6].

In this paper, we introduce ASIFT into our vision based navigation system to replace SIFT in order to achieve a more robust positioning result against viewpoint variation. At both mapping and positioning stage, ASIFT based image matching is used in the same way SIFT is utilized. In order to evaluate its performance and compare it with that of SIFT, datasets from the same controlled experiment is used. Therefore, not only matching reliability in terms of matched number and correct rate can be compared, more importantly, their influence on final positioning accuracy is evaluated against each other. It is noted that the distance ratio threshold used to select the best correspondence is set to 0.6 in ASIFT, in SIFT we keep the parameter in consistent.

In Fig. 5 and 6, we showed the matching between real time query image No.5 with map image No.10 using SIFT and ASIFT respectively. Under dramatic view changes, ASIFT produce more reliable matches while although SIFT get as many tentative matches, most of which are mismatches and filtered out by RANSAC. When dealing with images without much angular difference, however, ASIFT can hardly outperform its counterpart. In some cases it even gets a much worse result as shown in Fig. 7 and 8.

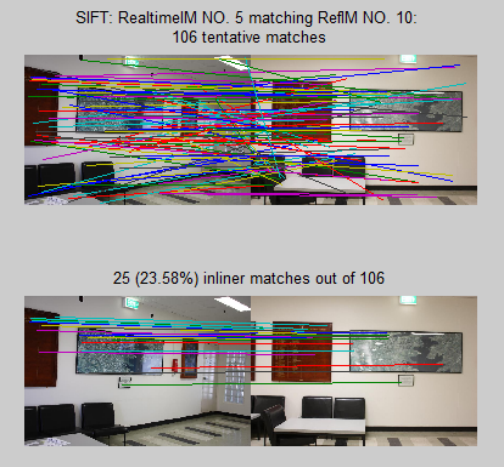

Figure 5. Matching between real time query image No.5 with map image No.10 using SIFT+RANSAC, dramative angular change, 25 reliable matches (inlier)

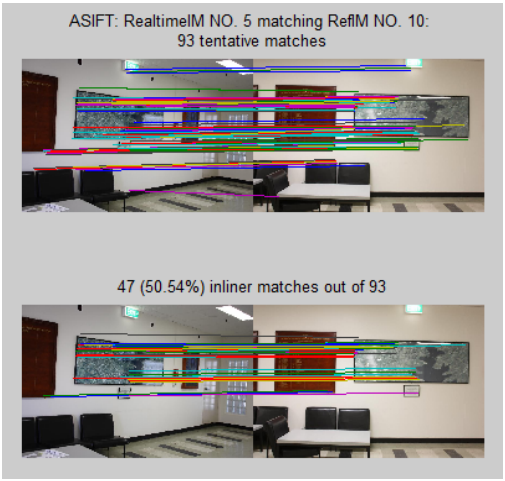

Figure 6. Matching between real time query image No.5 with map image No.10 using ASIFT+RANSAC dramative angular change, 47 reliable matches (inlier)

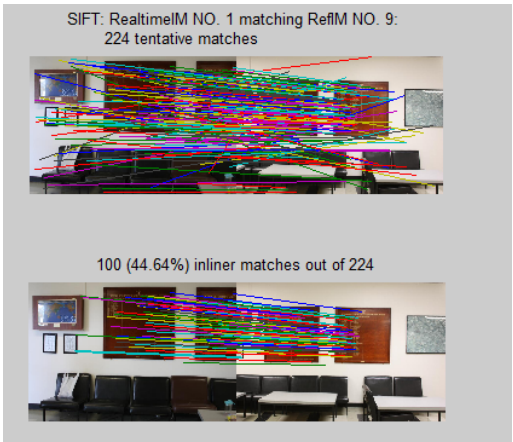

Figure 7. Matching between real time query image No.1 with map image No. 9 using SIFT+RANSAC, small angular change, 100 reliable matches (inlier)

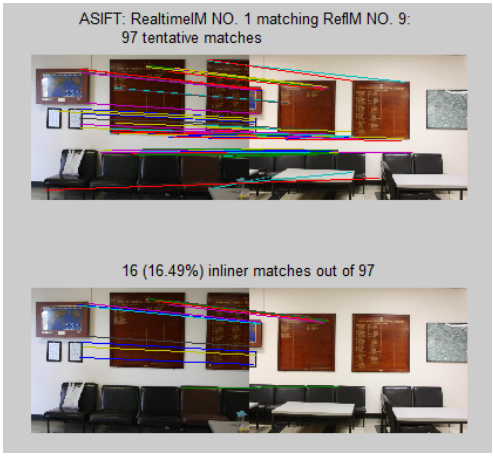

Figure 8. Matching between real time query image No.1 with map image No. 9 using SIFT+RANSAC, small angular change, 16 reliable matches (inlier)

TABLE II. PERFORMANCE OF ASIFT MATCHING IN THE SYSTEM

ID	Real time ImagID	Angular change	All ASIFT Matches	Reliable Matches	Percent of Reliable Matches	PGCP Number	Num of False PGCP	PDOP	ADOP
	1	0	297	72	24.14 %	6	0	5099	121-2
	2	0	190	37	19.19 %	4	0		
	3	-30	219	98	44.69 %	15	0	2319	459
	4	-30	150	69	46.15 %	8	0	2638	525
	5	50	385	193	50.10 %	13	0	2147	458
1	6	50	116	65	56.03 %	0	0		
	7	0	263	66	25.19 %	11	0	1232	305
	8	0	132	45	34.03 %	5	1	1337-6	351-1
	9	-20	187	119	63.42 %	12	2	1894	444
2	10	-20	185	115	61.94 %	10	0	2279	566
	11	0	702	520	74.13 %	12	0	12846	2610
	12	0	325	243	74.79 %	7	0	1773-5	3651
	13	-30	480	285	59.34 %	8	0	1275-2	2391
	14	-30	171	88	51.53 %	1	0		
	15	20	416	254	60.90 %	18	0	9313	1890
3	16	20	248	148	59.72 %	12	0	2165-8	435-8

1

We further compare the overall performance of the two algorithms. In order to mitigate the effect of occasionally extreme results, for every matching pair the algorithm runs 4 times and an average is used. Both matching go through

RANSAC to remove mismatches. Comparing Table 1 and 2, an obvious improvement happens on epoch No.5, one with big angular change. Using ASIFT it is able to produce a positioning result with reasonable number of PGCP (13). But for epoch No.6, image with the same view as No.5 but limited lighting, is still unable to get a result. Compare every pair of adjacent images, it is easy to deduce that ASIFT shares the same shortcoming with SIFT: being sensitive to illumination changes. On top of this, we compare the total number of matches and reliable matches remained after the RANSAC process. Generally, ASIFT has a higher correct rate but less tentative matches as well as reliable matches compared with SIFT. Without shown in the table, at mapping stage less geo-referenced feature points are generated by ASIFT. As a result, less PGCPs are produced at final positioning, which means a worse geometry and a lower position precision. One reason the author believe is that ASIFT already include an outlier detection mechanism (ORSA) in its algorithm. Therefore less tentative matches but a higher correct rate can be produced. However, the total number of reliable matches cannot outperform that of SIFT. In cases where the number and distribution of reliable matches directly affect positioning accuracy, as ours, we will favor the previous approach, but ASIFT can still be used as a backup plan when SIFT fails to get a result because of large viewpoint changes.

IV. FEATURE BASED MATCHING WITH INTEGRATION OF CROSS- CORRELATION INFORMATION

It has been noticed that both SIFT and ASIFT produce a substantial number of mismatches. The reason behind is that such feature-based methods depend on the choice of correspondence on local information and fail to consider global context. When an image has repeated patterns, ambiguities will occur when the local information for the similar parts is identical.

A general solution is to use robust estimation method like RANSAC to remove outliers. For a number of iterations, a random sample of 4 correspondences is selected and a homography H is computed from those 4 correspondences. Every other correspondence is then classified as an inlier or outlier depending on its concurrence with H. After all of the iterations have finished, the iteration that contained the largest number of inliers is selected. H can then be recomputed from all of the correspondences that were considered as inliers in that iteration. While this method can remove most of the mismatches, experiments have proved that it cannot guarantee the correctness of every pair of ' inliers'. The reason for it is that the iteration can only run limited times (1000 at our system), and there is a certain chance that the iteration that have largest number of inliers still produces an erroneous or very inaccurate H since most of the inliers it includes are mistaken. As a result, mismatches are selected as reliable matches, which will further affect the positioning accuracy.

In this study, we propose to integrate intensity-based method into the feature matching process to strengthen the robustness of the matching algorithm against mismatches and noise. More specifically, cross-correlation information is used as an analysis and selection criterion for the matching. Instead of identifying mismatch(es) after it has been generated, it determines how good the homography model (H) is for the two matching images and discard bad H to reduce chances that mismatches are included. In RANSAC, we use 2-D projective transformation H (planar homography) to approximate the geometric transformation between two images (e.g. I and I'). Any two corresponding SIFT features in image I and I' pass RANSAC will comply with the model, which ideally can be expressed as:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} a1 & a2 & a3 \\ b1 & b2 & b3 \\ c1 & c2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (1)$$

In (1) $p = [x, y, 1]^T$ and $p' = [x', y', 1]^T$ denotes the two points expressed using homogeneous coordinates, and $\begin{bmatrix} a1 & a2 & a3 \\ b1 & b2 & b3 \\ c1 & c2 & 1 \end{bmatrix}$ represents homography model H. $\begin{bmatrix} a1 & a2 \\ b1 & b2 \end{bmatrix}$ parameterized affine changes, $[a3 \quad b3]^T$ shift parameters and $[c1 \quad c2]$ projective deformation. After homography model has been generated by RANSAC processing, a local square patch in I with a size of $(2w+1) * (2w+1)$ centered on p is generated, denoted as $N(p)$. By using the estimated H, $N(p)$ is transformed into $N(p')$ and resampled on image I'. Then cross-correlation between the two window patches is calculated using (2), where G_{uv} and G'_{uv} represent the intensity values of the two correlation windows, respectively, whereas $\mu(G)$ and $\mu(G')$ denote their average intensity.

$$\rho(G, G') = \frac{\sum_{u=-w}^{w}\sum_{v=-w}^{w}(G_{uv}-\mu(G))(G'_{uv}-\mu(G'))}{\sqrt{\sum_{u=-w}^{w}\sum_{v=-w}^{w}(G_{uv}-\mu(G))^2 \cdot \sum_{u=-w}^{w}\sum_{v=-w}^{w}(G'_{uv}-\mu(G'))^2}} \quad (2)$$

In (2) $\rho(G, G')$ varies from -1 to 1, the closer to 1 the higher correlation, and here indicates the bigger similarity between two patches and greater possibility to be correct corresponding points. Therefore we calculate the cross-correlation for each pair of reliable matches (inliers) for every single matching process. An average correlation $\overline{\rho}$ is calculated for all the matched (reliable) points produced by one matching (one H is generated). If it is close to 1, it means the estimated homography model H is very accurate; vice verse. An example is shown in Fig.9 &10 (both from our experiment), which illustrates a bad $\overline{\rho}$ could include mismatch as inlier and on the contrary, a $\overline{\rho}$ closer to 1 has less chance to include mismatch as inlier .

A certain threshold is set for $\overline{\rho}$ (0.75 in the experiment). After every matching process, a $\overline{\rho}$ is calculated. If it is smaller than the threshold, the matched points generated by the process will be discarded and the two images will be re-matched. The whole process is as shown in TABLE III.

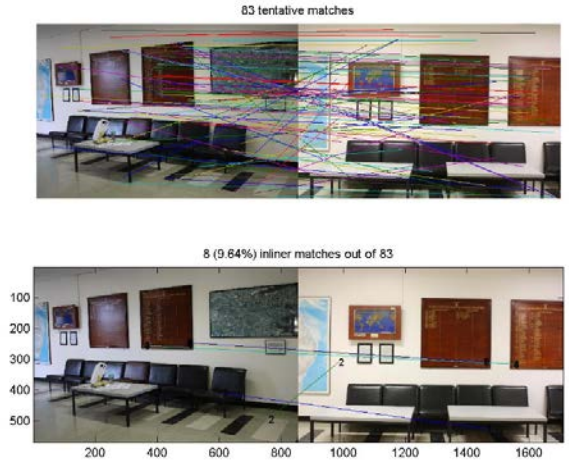

Figure 9. Matching between real time query image No.13 with map image No. 8 using SIFT+RANSAC+cross-correlation: $\overline{\rho} = 0.23$

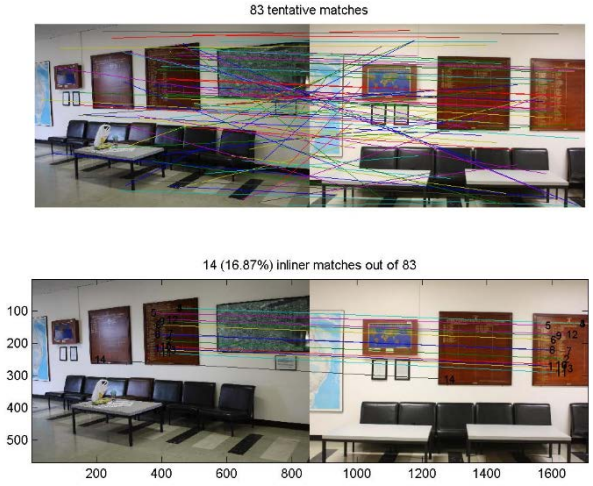

Figure 10. Matching between real time query image No.13 with map image No. 8 using SIFT+RANSAC+cross-correlation: $\overline{\rho} = 0.79$

TABLE III. FEATURE BASED MATCHING WITH INTEGRATION OF CROSS-CORRELATION INFORMATION

Step1: Extract feature points from the input images I and I′ using the SIFT method.

Step2: Perform an initial matching to get tentative matching points.

Step3: Use RANSAC to select reliable matches (p_i, p_i') and produce homography model H.

Step4: Using H, a square patch (N(p): G). around every reliable SIFT point p on image I is created and transformed and resampled on image I′, (N(p′): G′).

Step5: Calculate cross-correlation for every pair of square patch and get their average value $\overline{\rho}$.

Step6: If $\overline{\rho} < threshold$, go to step 2.

Step7: Find PGCP.

Step 8: Using PGCP to calculate 6DOF for the camera

By using such criteria for the system, the number of false PGCP has been reduced. In order to evaluate the impact on final positioning accuracy, root mean square error for every camera station is calculated and the results before and after using cross-correlation information are compared in TABLE IV, which reflects that positioning accuracy has been improved.

TABLE IV. COMPARISON OF THE RMSE OF THE POSITIONING RESULTS

RMSE		X(m)	Y(m)	Z(m)
Station1	SIFT	0.0477	0.0838	0.0598
	SIFT+ cross-correlation	0.0469	0.0824	0.0613
Station2	SIFT	0.0641	0.1192	0.2710
	SIFT+ cross-correlation	0.0644	0.0582	0.2481
Station3	SIFT	0.3397	0.7191	0.3278
	SIFT+ cross-correlation	0.3164	0.6439	0.2883

V. *CONCLUDING REMARKS*

In this paper, we give an overview of image matching techniques in the context of vision-based navigation systems. Based on a variety of systems, stereo vision, structure from motion and map-based approach, different methods are employed and discussed. Focused on map-based approach, which generally uses feature-based matching for localisation, and based on our early developed system, a performance analysis has been carried out and three major problems have been identified: being vulnerable to illumination changes, drastic viewpoint changes and good percentage of mismatches. The latter two problems have been addressed in this study.

By introducing ASIFT into the system, the major improvement takes place on the epoch with large viewpoint changes. More reliable matches have been produced and the system has been able to get a positioning result. However, our experiments have also revealed that ASIFT has a higher correct rate but less tentative matches as well as reliable matches compared with SIFT when dealing with images with small angular change. For the systems the number and distribution of reliable matches directly affect positioning accuracy, like ours, SIFT is still the more favourite option, but ASIFT can be used as a backup plan when SIFT fails to get a result because of large viewpoint changes.

Using RANSAC to remove mismatches has been a popular approach for feature-based matching in visual systems. But as has been proved by the performance analysis, some mismatches may still be included as inliers (reliable match) in final positioning process. The reason for this is that the iteration in RANSAC can only run limited times and there are chances that the one having the largest number of inliers still

produces an erroneous or very inaccurate Homography model and as a consequence, mismatches are selected as reliable matches. In order to deal with the problem, we have proposed to use cross-correlation information to evaluate the quality of homography model and the conducted experiments have proved that such an approach can reduce the chances of mismatches being included and final positioning accuracy can be improved.

Further research will be focused on improving the reliability and precision of image matching in an effort to improve the overall positioning accuracy. Meanwhile, the computation load of image matching will be considered for such real-time system operations.

REFERENCES

[1] Gruen, A.W., 1985. Adaptive Least Squares Correlation: A powerful Image Matching Technique. South Africa Journal of Photogrammetry Remote Sensing and Cartography, pp. 175-187.

[2] Lowe, D. 2004. Distinctive Image Features from Scale Invariant Key points. International Journal of Computer Vision pp. 91-110

[3] Trucco, E.; Verri, A., 1998, Introductory Techniques for 3-DComputer Vision, Prentice Hall.

[4] Kuhl, A. Comparison of Stereo Matching Algorithms for Mobile Robots. M.Sc. Thesis, Fakultät für Informatik und Automatisierung, Technische Universität Ilmenau, Ilmenau, Germany, 2004.

[5] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondence. Third European Conference on Computer Vision, Stockholm, Sweden, May 1994.

[6] ASIFT: J. M. Morel and G. Yu, “ASIFT: A new framework for fully affine invariant image comparison,” SIAM J. Imag. Sci., vol. 2, no. 2, pp. 438–469,Apr. 2009.

[7] G. Zhang, Z. Dong, J. Jia, T. T. Wong, and H. Bao, “Efficient non-consecutive feature tracking for structure-from-motion,” in Proceedings of the European Conference on Computer Vision (ECCV ’10), 2010.

[8] Harris, C.; Stephens, M., 1988. A combined corner and edge detector. In Fourth Alvey Vision Conference, Manchester, UK, pp. 147-151.

[9] Forstner, W., 1986. "A Feature Based Correspondence Algorithm for Image Matching," International Archives of Photogrammetry, Vol. 26-III, Rovaniemi, Finland, 1986.

[10] Remondino, F. & Ressl, C. 2006. Overview and experiencesin automated markerless image orientation. IAPRS, Vol. 36, Part 3, pp. 248-254.

[11] F. Lang and Forstner,W. Matching techniques. 1995. In Second Course in Digital Photogrammetry. Landesvermessungsamt NRW.

[12] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(5):530–535, May 1997.

[13] H. Bay, T. Tuytelaars, and L. V. Gool, “SURF: speed-up robust features”, 9th European Conference on Computer Vision,pp. 404-417 (2006)

[14] R. Boris, K. Effrosyni, and D. Marcin, “Mobile museum guide based on fast SIFT recognition”, 6th International Workshop on Adaptive Multimedia Retrieval, pp. 26-27 (2008)

[15] Se, S., Lowe, D., and Little, J. 2002. Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. International Journal of Robotic Research, 21(8):735–760.

[16] M.A. Fischler and R.C. Bolles. ”Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography.” Communications of the ACM, 24(6):381C395, 1981.

[17] Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide-baseline stereo from maximally stable extremal regions. Image and Vision Computing 22 (2004) 761-767

[18] Mikolajczyk, K., Schmid, C.: Scale & a_ne invariant interest point detectors. International Journal of Computer Vision 60 (2004) 63-86

[19] L. Nalpantidis, D. Chrysostomou and A. Gasteratos, "Obtaining Reliable Depth Maps for Robotic Applications with a Quad-camera System", ICRA09 Workshop on Safe navigation in open and dynamic environments Application to autonomous vehicles, 2009

[20] G. J. Burghouts and J.M. Geusebroek. Performance evaluation of local colour invariants. *CVIU*, 2009.