

# Feature Matching under Region-based Constraints for Robust Epipolar Geometry Estimation

Wei Xu and Jane Mulligan

Department of Computer Science, University of Colorado at Boulder, Boulder, Colorado  
80309-0430 USA  
{Wei.Xu, Jane.Mulligan}@Colorado.edu

**Abstract.** Outlier-free inter-frame feature matches are important to accurate epipolar geometry estimation for many vision and robotics applications. We discover a set of high-level geometric and appearance constraints on low-level feature matches by exploiting reliable region matching results. A new outlier filtering scheme based on these constraints is proposed that can be combined with traditional robust statistical methods to identify outlier feature matches more reliably and efficiently. The proposed filtering scheme is tested in a real application of outdoor mobile robot navigation based on far-field scenes and of scenes that contain repeated structures.

## 1 Introduction

Epipolar geometry estimation (or relative pose estimation) between different views [1] is a hot topic in vision and robotics research and has many applications such as short-baseline/wide-baseline stereo, 3-D reconstruction, object tracking, mobile robot navigation, and visual odometry. Since the estimate is computed from image feature correspondences, researchers have developed and are still endeavoring to develop more and more prominent and stable image features. The focus has transited from the traditional Harris corner features [2]) that are computed from a simple local signature (i.e., local neighborhood in the image), to the more recently developed view independent features such as SIFT [3], GLOH [4] and MSER [5] that exploit high-level information from a much larger and more sophisticated signature aiming for global invariance. However, the preliminary feature matching schemes of the feature detectors are usually not outlier-free. To address this problem, robust statistical methods, such as RANSAC [6], MAP-SAC [7] and LMeds [8], are usually combined with the epipolar geometry estimator to refine the preliminary inter-frame feature correspondences and estimate the geometry at the same time. These statistical methods identify whether a correspondence is an inlier or an outlier based on whether or not it is consistent with the major structure of the whole data (i.e. all feature correspondences).

A characteristic of many view independent image features (e.g., SIFT) is that although the features are computed from a local signature the matching of them is global. This design is necessary for applications where the relative pose between a pair of views is large (e.g., wide baseline stereo). In this case, the matching scheme has to search over the whole peer image to find the optimal correspondences of the features in the base image. However, global matching may also arouse more outliers, especially in an outdoor

environment where similar texture patterns (e.g., trees and grasses), repeated structures (e.g., buildings) and lighting changes may all cause false matching. In this paper, we propose a filtering scheme to overcome or alleviate this shortcoming of global matching by applying region-based locality and appearance constraints to refine global feature matching results.

## 2 Related Work

As introduced earlier, robust statistics methods are usually used to refine global feature matching results. Among these methods, RANSAC [6] and its variants (e.g., MAP-SAC [7]) are most popular ones. They use random sampling and hypothesis-and-verification test repeatedly to locate the major structure of the data. However, the success of RANSAC and its variants is limited by the following practical issues: 1) They require the user to specify the fraction of outliers in the data ( $\epsilon$ ) and the residual threshold between inliers and outliers ( $t$ ), but in practice these parameters are usually unknown and have to be guessed. 2) They guarantee to locate the major structure of the data at a designated confidence level ( $\delta$ ) in  $N$  trials —  $N = \lceil \ln(1 - \delta) / \ln(1 - (1 - \epsilon)^m) \rceil$  where  $m$  is the minimal number of data points to generate a hypothesis model. If this level  $\delta$  is set up too high or fraction of outliers ( $\epsilon$ ) is high, hundreds of thousands trials may be needed to locate the major structure of the data. This is impractical for many vision and robotics applications. In practice, many applications (e.g., [9]) would rather trade performance for time and limit the number of trials ( $N$ ) to be several hundreds or a few thousands.

In addition to robust statistical methods, methods making use of high-level geometric objects (e.g., line segments [10], texture-invariant regions [11]) or constraints (e.g., the sidedness constraint [12]) have also been proposed to guide low-level feature matching. However, the efficacy of these methods is limited for outdoor vision and robotics applications: the texture-invariant regions only exist in texture abundant areas and thus usually have a sparse and uneven distribution in outdoor scenes; the matching scheme based on line segments only works for indoor environments; and the sidedness constraint needs several different views to apply. Our filtering scheme is motivated by the previous work of Tao et al. [13] that proposed a global matching framework of stereopsis which makes use of low-level segmentation in depth space to restrict the search space of dense stereo correspondences. Following its general idea, we propose to exploit high-level geometric and appearance constraints based on reliable region matches to refine preliminary correspondences of view independent features. However, unlike Tao’s work, we do not intend to integrate the region-based constraints into any particular feature detectors, but are more interested in developing a general scheme for filtering the preliminary correspondences generated by any view independent feature detectors or any feature detectors that employ global matching.

Another piece of previous work closely related to ours is the enhanced feature matching scheme using saliency region correspondences [14]. It integrated image segmentation and region matching in the joint image space and solved them together using the normalized cuts technique. The approach is more general than ours because it aims at wide-baseline applications. However, according to the examples provided in the paper, it may still get false region matches due to the dramatic appearance changes

between wide-baseline views. Comparing to [14], the approach proposed in this paper aims at more restricted short-baseline applications but tries to provide a more effective solution to it, since the region matches can be more reliability obtained given short-baseline. It also decouples image segmentation and region matching and solves them separately in a much simpler manner than normalized cuts. The target applications of the proposed scheme includes those in which the relative pose between a pair of views is small or moderate, such as short-baseline stereo and object tracking and mobile robot navigation at a high image sampling rate.

### 3 The Approach

To make the proposed filtering scheme work, we need to solve the following problems: 1) how to obtain good representations of regions and match them correctly, and 2) what kind of constraints from region matches can be used to filter low-level feature correspondences, and how? Details of our solutions are given as follows.

#### 3.1 Inter-frame region matching

We developed an inter-frame region matching scheme based on the similarities of a sequence of geometric and appearance statistics of image segmentation results. Considering our target task is epipolar geometry estimation for outdoor mobile robot navigation, we adopted the Color Structure Code (CSC) algorithm [15] for segmenting the sampled image frames. It is reported that CSC is good at segmenting natural color scenes in a test with over 5,000 outdoor natural images [15]. Our own experiments confirmed this report and showed that in most cases the segments generated by CSC are conceptually coherent with real-world regions when the scene is relatively distant or flat. Besides, CSC is very fast and was successfully used for real-time vehicle tracking [16].

We assume the image segments generated by CSC are reasonable representations of real-world regions in the scene, so we can match the regions by comparing the similarities between their corresponding segments. The pair-wise region similarities are computed from a sequence of statistics measuring regional geometric and appearance properties including color ( $S_{color}$ ), centroid ( $S_{centroid}$ ), area ( $S_{area}$ ), texture ( $S_{texture}$ ) and shape (including measurements on the proportion and orientation of the equivalent ellipse ( $S_{ee\_prob}$  and  $S_{ee\_ori}$ ) and on the bounding rectangle ( $S_{br}$ )).

Given a pair of regions A and B,  $S_{color}(A, B)$  and  $S_{centroid}(A, B)$  are calculated as the normalized Euclidean distances between the mean color vectors (in the CIE-Lab color space) and the mean position vectors of the pixels in A and B. The area similarity  $S_{area}$  is defined as:

$$S_{area}(A, B) = \frac{|s_A - s_B|}{s_A + s_B}$$

where  $s_A$  and  $s_B$  are the sizes of regions A and B respectively. The shape similarity is measured by the similarities of the equivalent ellipse ( $S_{ee\_prob}$  and  $S_{ee\_ori}$ ) and of the bounding rectangle ( $S_{br}$ ) which are defined as [17]:

$$S_{ee-prob}(A, B) = \frac{1}{2} \cdot \left( \frac{|m_A - m_B|}{m_A + m_B} + \frac{|n_A - n_B|}{n_A + n_B} \right)$$

where  $m_A$  and  $n_A$  ( $m_B$  and  $n_B$ ) are the lengths of the major and minor axes of the equivalent ellipse of region A(B).

$$S_{ee-ori}(A, B) = \left(1 - \frac{1}{e^{4(\Pi-1)}}\right) \cdot \frac{\lfloor \frac{\Delta\theta+90}{180} \rfloor \cdot 180 - \Delta\theta}{90}$$

where  $\theta_A$  ( $\theta_B$ ) denotes the angle of the major axis of the equivalent ellipse of region A(B),  $\Delta\theta = |\theta_A - \theta_B|$ ,  $\Pi = \min(\frac{m_A}{n_A}, \frac{m_B}{n_B})$ .

$$S_{br}(A, B) = \frac{1}{2} \cdot \left( \frac{|w_A - w_B|}{w_A + w_B} + \frac{|h_A - h_B|}{h_A + h_B} \right)$$

where  $w_A$  and  $h_A$  ( $w_B$  and  $h_B$ ) denote the width and height of the bounding rectangle of region A(B) respectively. The definition and computation of the equivalent ellipse and the bounding rectangle are described in [18].

The measurement of the texture similarity is in two steps. First, we compose a texture vector of six fields that measures the mean, standard deviation, smoothness, third moments, uniformity and entropy of the intensity distribution in a region [19]. Next, the texture similarity  $S_{texture}(A, B)$  is calculated as the normalized Euclidean distances between the texture vectors of A and B.

All the above similarity measurements are already normalized as defined. Finally, the overall similarity score is calculated as a weighted sum of the individual measures:

$$S(A, B) = \alpha_1 S_{color} + \alpha_2 S_{centroid} + \alpha_3 S_{area} + \alpha_4 S_{ee-ori} + \alpha_5 S_{ee-prob} + \alpha_6 S_{br} + \alpha_7 S_{texture}$$

The values of the weights  $\alpha_i$ 's in the above equation are determined by application and set up by experience. Specifically, we use a larger weight for color, area and centroid position and a smaller weight for the other features for our target applications that have a lot of outdoor far-field scenes.

Given the region set  $R^1 = \{r_p^1 \in I_1, p = 1 \dots m_1\}$  of a frame  $I_1$  and the region set  $R^2 = \{r_q^2 \in I_2, q = 1 \dots m_2\}$  of another frame  $I_2$ , the matching between  $R_1$  and  $R_2$  is the following combinatorics optimization problem:

$$O = \arg \min_{r_p^2, r_q^1} \left\{ \sum_{p=1}^{m_1} S(r_p^1, r_p^2) + \sum_{q=1}^{m_2} S(r_q^1, r_q^2) \right\}$$

subject to:  $r_p^2 \in R^2$  and  $\cup_{p=1}^{m_1} r_p^2 = R^2$ ,

$r_q^1 \in R^1$  and  $\cup_{q=1}^{m_2} r_q^1 = R^1$ ,

$r_{p_1}^2 \neq r_{p_2}^2$  if  $p_1 \neq p_2$ ,

$r_{q_1}^1 \neq r_{q_2}^1$  if  $q_1 \neq q_2$ .

In our practice, we revised the above problem a little considering practical factors. We defined two constraints on the color and area statistics of the regions: (a)  $S_{color} \leq \tau_1$ , (b)  $S_A > \tau_2$  and  $S_B > \tau_2$ , considering in an outdoor environment the measurement of regions of far-field scenes or under varying lighting conditions may not be accurate.  $\tau_1$  and  $\tau_2$  are thresholds determined by application. Constraint (a) requires the appearance change of matched regions is not too large; (b) excludes small image segments from matching because they are less reliable at representing a region. Due to constraint (b), the completeness constraints in the above problem (the first two constraints) have to be released since not all regions are now included. Also, considering in our target applications the relative pose is not too large we had another constraint on motion: (c)  $S_{centroid} \leq \tau_3$ , where the threshold  $\tau_3$  is co-determined by the speed of the vehicle and the image sampling rate. These practical constraints guarantee that the obtained region matches are reliable.

We adopted the “perfect-matching” scheme proposed by Rehrmann [17] to solve the revised optimization problem: For a region in one image frame, the optimal match of it is the region peer in the other frame with the highest similarity score to it. This examination is performed for every inter-frame pair of regions. Region pairs whose members are mutually optimal to each other with very high similarity scores are classified as reliable “perfect matches”. Previous work [16] and our own experience show the “perfect-matching” scheme works well for vehicle tracking and outdoor mobile robot navigation applications with a normal image sampling rate (15 fps of 320x240 frames). However, it may be less effective (i.e., only a few “perfect matches” are obtained) when there occurs a dramatic viewpoint change between a pair of frames. Even this situation occurs, it only reduces the number of region-based constraints that can be used to refine the low-level feature matches but would not affect the correctness of those constraints.

### 3.2 Region-based constraints

In general conditions the geometric transformation between a pair of views of a static scene is a similarity transformation under which the geometric properties of a region including position, area, orientation and shape are not consistent. However, if the relative pose between a pair of views is relatively small these properties may only change a little and can still be used as measures for region matching, as our region matching scheme does. Also, we consider that the appearance of a region may change with time in outdoor vision and robotics applications — uncontrolled lighting, change of viewpoints and non-diffuse reflections may cause the same scene point to have different color/intensity values in different images. Taking both of these geometric and appearance aspects into consideration, we bring forward the following region-based constraints on low-level feature correspondences:

**Region ownership constraint** — Feature matches where the two features reside in non-matched regions, called “cross-region” matches, are detected and identified as outliers. Considering region boundaries may not be accurately detected in practice, we do not count as outliers those “cross-region” matches in which any of the two features is close to the boundary of the matching region of its owner region.

```

Input: Image  $I_1$  and its segments  $R_p^1 \in I_1, p = 1 \dots m_1$ ; image  $I_2$  and its segments
 $R_q^2 \in I_2, q = 1 \dots m_2$ ; inter-frame region matches
 $RM = \{(R_i^1, R_i^2) \mid R_i^1 \in I_1, R_i^2 \in I_2 \text{ and } i = 1 \dots m\}$ ; inter-frame feature
matches  $FM_{all} = \{(X_j^1, X_j^2) \mid X_j^1 \in I_1, X_j^2 \in I_2 \text{ and } j = 1 \dots n\}$ ; adjacency
threshold  $\tau$ 
Output: Feature match inlier set  $FM_{inlier} \in FM_{all}$  and outlier set  $FM_{outlier} \in FM_{all}$ 
Update  $FM$  using  $RM$  using the appearance constraint;
 $FM_{inlier} = \emptyset, FM_{outlier} = \emptyset$ ;
for  $j=1$  to  $n$  do
  foreach  $(X_j^1, X_j^2) \in FM_{all}$  do
    find regions  $R_p^1 \in I_1$  and  $R_q^2 \in I_2$  such that  $X_j^1 \in R_p^1$  and  $X_j^2 \in R_q^2$ ;
  end
  if  $(R_p^1, R_q^2) \in RM$  then
     $FM_{inlier} = FM_{inlier} \cup (X_j^1, X_j^2)$ ;
  else
    find region  $R_p^2 \in I_2$  such that  $(R_p^1, R_p^2) \in RM$ ;
    find region  $R_q^1 \in I_1$  such that  $(R_q^1, R_q^2) \in RM$ ;
     $d_1 = \text{computeMinDist}(X_j^1, R_q^1)$ ;
     $d_2 = \text{computeMinDist}(X_j^2, R_p^2)$ ;
    if  $\min(d_1, d_2) < \tau$  then
       $FM_{inlier} = FM_{inlier} \cup (X_j^1, X_j^2)$ ;
    else
       $FM_{outlier} = FM_{outlier} \cup (X_j^1, X_j^2)$ ;
    end
  end
end

```

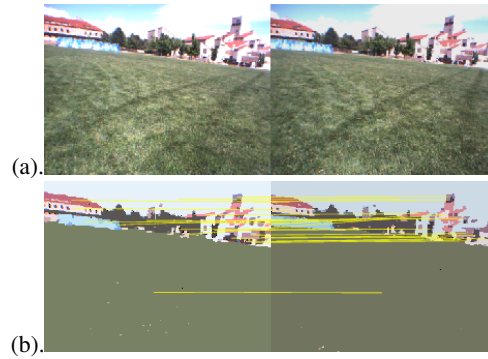
**Algorithm 1:** An outlier filtering scheme based on the proposed region-based constraints.

**Appearance constraint** — For all pairs of feature matches residing in the same pair of matched regions, re-compute the matches of these features taking into account the color difference between these two regions. This may result in some features being assigned different conjugates to the original matches.

In our implementation, the appearance constraint was enforced by first compensating the average color difference of a pair of matched regions and then updating the feature matches within this pair. It is not able to identify outlier feature matches but may change the original assignment. The region ownership constraint directly identifies outlier feature matches that break it. An outlier filtering scheme based on these constraints are summarized in Algorithm 1.

## 4 Experimental Results

We have tested the proposed constraint filtering scheme with respect to epipolar geometry estimation for different target applications. View independent SIFT features, eight-point algorithm and MAPSAC [7] (which is a maximum a posteriori (MAP) variant of

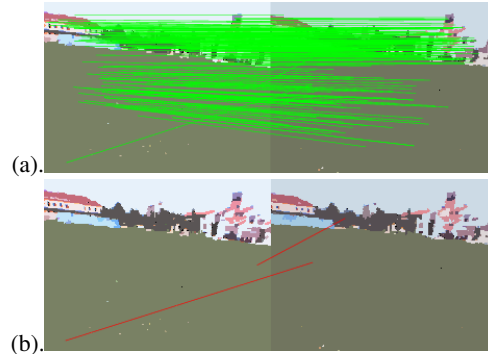


**Fig. 1.** (a) Original image pair from a motion sequence. (b) Color segments with yellow lines connecting matching segment centroids.

RANSAC) are adopted to compute the epipolar geometry between a pair of frames. The evaluation criterion of the estimated epipolar geometry is the sum of the back-projection errors of the scene points onto the image planes. Our evaluation strategy is to compare the accuracy of the estimated epipolar geometry with and without applying the filtering scheme on the preliminary feature matches before running MAPSAC. The comparison results will show whether or not there is a performance gained by incorporating the region-based constraints.

Fig. 1a shows a pair of images of different views of an outdoor far-field scene. The image pair are collected in an application of mobile robot navigation based on far-field scenes. The images are first segmented using the CSC algorithm, and then the inter-frame region match set  $RM$  are computed using the proposed region matching scheme. In far-field scenes, textured regions are usually smoothed out by distance and agglomerated into large color coherent regions, so both the color-based CSC segmentation algorithm and the proposed region matching scheme work very well, as Fig. 1b shows. SIFT features are detected on each image and their preliminary matches are examined by MAPSAC (Fig.2a). Fig.2b shows two obvious outlier matches in Fig.2a that have passed the examination of MAPSAC. As introduced earlier, this problem can be overcome if the user can provide the true values of the outlier fraction  $\epsilon$  and the inlier identification threshold  $t$ . However, in practice the true values of these parameters are usually unknown and vary for different scenes, and guessed values have to be used instead. So, this "missing outlier" problem of robust statistics methods can often happens if the guess is wrong, and other efforts in addition to robust statistical methods such as the proposed region-based constraints are needed to refine the matches.

We then applied the "Constraint filtering + MAPSAC" scheme to identify outlier feature matches and estimate the epipolar geometry for this far-field scene example. Fig. 3(a) shows the mean estimation errors of 100 runs of both the "MAPSAC only" and the "Constraint filtering + MAPSAC" scheme. It is shown that the "Constraint filtering + MAPSAC" scheme can achieve lower estimation errors than applying MAPSAC alone, and the improvement are more significant for noisy data which correspond



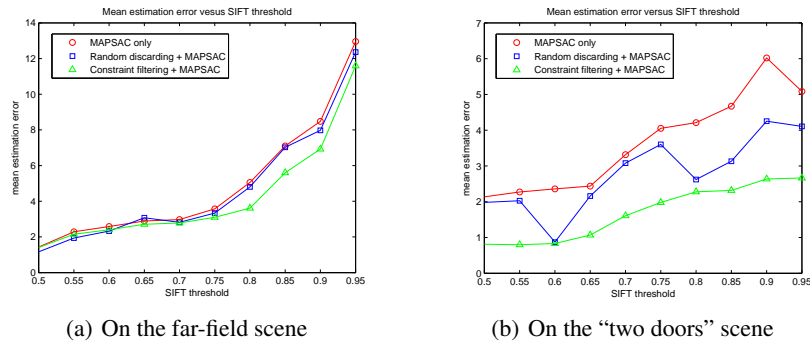
**Fig. 2.** (a) Inlier matches of SIFT features (connected by green lines) identified by a run of MAPSAC with  $N = 1,000$  trials. (b) The two outliers (connected by red lines) that have passed the examination of MAPSAC.

to higher SIFT thresholds. To verify that the performance improvement is not purely due to the drop of matches, we also tested another “Random discarding + MAPSAC” scheme that randomly discards the same amount of preliminary matches as that of the outliers that the proposed constraint filtering scheme identifies before going to MAPSAC. The performance of this “Random discarding + MAPSAC” scheme is also shown in Fig. 3(a).

It should be noted that the power of the proposed constraint filtering scheme is not fully exerted in this example of far-field scenes because there only exist a few outliers in total in the preliminary SIFT feature matches, which leaves little space for the proposed filtering scheme to improve upon. Fig. 4 shows another example of a close-range scene containing repeated structures where the proposed constraint filtering scheme can better exert its power. The two doors in the scene have the same structure and similar local texture patterns. The global matching scheme of the SIFT feature detector is easily confused with this kind of scenes of repeated structures. Fig. 3(b) shows the the proposed constraint filtering scheme can greatly improve the accuracy of the estimated epipolar geometry for this kind of scenes — it reduced the mean estimation errors (of 100 runs) by around 50% for different SIFT threshold levels.

## 5 Discussion and Conclusion

In this paper we propose a set of high-level region-based constraints for refining low-level image feature matches. We use color image segmentation techniques to extract coherent regions over a pair of images of the same scene. High-level geometric and appearance information, in the format of region-based constraints, are extracted from reliable region matching results. These constraints are used to identify and remove outliers of preliminary feature matches before a robust statistical method, MAPSAC, is applied to estimate the epipolar geometry. Experiments with different applications show that combining a pre-filtering scheme based on the proposed *ownership* (spatial) and



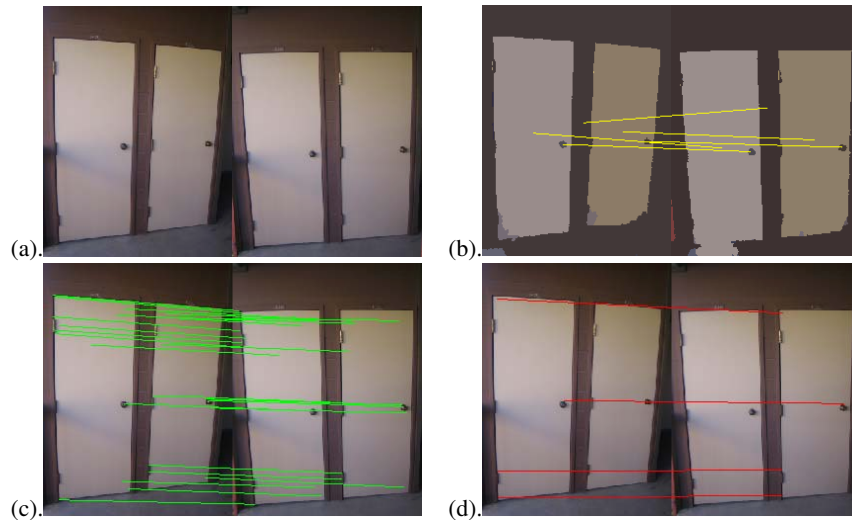
**Fig. 3.** Comparison with MAPSAC. (a) Epipolar geometry estimation errors under different SIFT threshold levels for the far-field scene. (b) Experiment results with SIFT features for the “two doors” scene — Incorporating the proposed constraint filtering scheme helps reduce the epipolar geometry estimation errors by around 50% than applying MAPSAC alone for this scene.

*appearance* constraints with MAPSAC can achieve better performance than applying MAPSAC alone.

The effectiveness of the proposed constraints relies on the quality of the image segmentation and region matching results. Since our target application is short-baseline applications, reliable region matching results can be obtained in most cases. How to extend the proposed scheme to wide-baseline applications is a possible direction for future work.

## References

1. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. second edn. Cambridge University Press (2004)
2. Harris, C., Stephens, M.: A combined corner and edge detector. In: Proc. 4th Alvey Vision Conference. (1988) 147–151
3. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60** (2004) 91–110
4. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence **10** (2005) 1615–1630
5. J. Matas, e.a.: Robust wide baseline stereo from maximally stable extremal regions. In: Proc. 9th European Conference on Computer Vision (ECCV’02). Volume 1. (2002) 384–393
6. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Comm. of the ACM **24** (1981) 381–395
7. Torr, P., Murray, D.: The development and comparison of robust methods for estimating the fundamental matrix. International Journal of Computer Vision **24** (1997) 271–300
8. Zhang, Z.: Estimating motion and structure from correspondences of line segments between two perspective images. IEEE Transactions on Pattern Analysis and Machine Intelligence **17** (1995) 1129–1139
9. Zhang, W., Kosecka, J.: A new inlier identification procedure for robust estimation problems. In: Proc. Robotics: Science and Systems Conference 2006 (RSS’06). (2006)



**Fig. 4.** (a) Original image pair of the “two doors” scene. (b) Color segments with yellow lines connecting matching segment centroids. (c) Inlier matches of SIFT features identified by using the “Constraint filtering + MAPSAC” scheme. SIFT feature match threshold is 0.50. (d) Outliers identified by the proposed constraint filtering scheme alone. Note that these outliers are not always correctly identified by using MAPSAC alone (see also Fig. 3(b)).

10. Bay, H., Ferrari, V., Gool, L.V.: Wide-baseline stereo matching with line segments. In: Proc. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR’05). Volume 1. (2005) 329–336
11. Schaffalitzky, F., Zisserman, A.: Viewpoint invariant texture matching and wide baseline stereo. In: Proc. 18th IEEE International Conference on Computer Vision (ICCV’01). (2001) 636–643
12. Ferrari, V.: Wide-baseline multiple-view correspondence. In: Proc. 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR’03). Volume 1. (2003) 718–725
13. Tao, H., Sawhney, H., Kumar, R.: A global matching framework for stereo computation. In: Proc. 18th IEEE International Conference on Computer Vision (ICCV’01). (2001) 532–539
14. Toshev, A., Shi, J., Daniilidis, K.: Image matching via saliency region correspondences. In: Proc. 2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR’07). (2007) 1–8
15. Rehrmann, V., Priese, L.: Fast and robust segmentation of natural color scenes. In: Proc. 3rd Asian Conference on Computer Vision. (1998) 598–606
16. Ross, M.: Segment clustering tracking. In: Proc. 2nd Europ. Conf. on Colour in Graphics, Imaging, and Visualization. (2004) 598–606
17. Rehrmann, V.: Object oriented motion estimation in color image sequences. In: Proc. 5th European Conference on Computer Vision (ECCV’98). Volume 1. (1998) 704–719
18. Hornberg, A.: Handbook of Machine Vision. WILEY (2006)
19. R.C.Gonzalez, Woods, R., S.L.Eddins: Digital Image Processing Using MATLAB. Prentice Hall (2003)