Robust Matching of Building Facades under Large Viewpoint Changes

Jimmy Addison Lee, Kin-Choong Yow, and Alex Yong-Sang Chia
School of Computer Engineering, Nanyang Technological University
50 Nanyang Avenue, Singapore 639798
{jimm0002, kcyow, chia0106}@ntu.edu.sg

Abstract

This paper presents a novel approach to finding point correspondences between images of building facades with wide viewpoint variations, and at the same time returning a large list of true matches between the images. Such images comprise repetitive and symmetric patterns, which render popular algorithms e.g.*, SIFT to be ineffective. Feature descriptors such as SIFT that are based on region patches are also unstable under large viewing angle variations. In this paper, we integrate both the appearance and geometric properties of an image to find unique matches. First we extract hypotheses of building facades based on a robust line fitting algorithm. Each hypothesis is defined by a planar convex quadrilateral in the image, which we call a "q-region", and the four corners of each q-region provide the inputs from which a projective transformation model is derived. Next, a set of interest points are extracted from the images and are used to evaluate the correctness of the transformation model. The transformation model with the largest set of matched interest points is selected as the correct model, and this model also returns the best pair of corresponding q-regions and the most number of point correspondences in the two images. Extensive experimental results demonstrate the robustness of our approach in which we achieve a tenfold increase in true matches when compared to state of the art techniques such as SIFT and MSER.*

1. Introduction

One of the most challenging tasks in computer vision is the locating of true matches between images of building facades taken from arbitrary viewpoints. The presence of repetitive and symmetric patterns in these man-made objects introduces a significant number of false matches; while a substantial dissimilarity in viewpoints hampers the number of matches to a modest number. In the past, several appearance-based object recognition techniques have been proposed, where most of them include a feature detector searching for salient features (*e.g.*, points or regions) characterized by *e.g.*, corners, edges, or entropy; followed by a feature descriptor to describe the features. To try obtaining invariance under arbitrary viewing conditions, they either extract invariant features or they compute invariant descriptors based on non-invariant features. Nevertheless none of the existing approaches (*e.g.*, by Baumberg [1]; Kadir *et al*. [5]; Lee *et al*. [7]; Lowe [10]; Matas *et al*. [11]; Obdrzálek and Matas [18]; Tell and Carlson [22]; Xiao and Shah [24]; *etc*.) can handle large viewpoint changes very well, *e.g*. for view angles that separate images higher than 40° [15]. The features corresponding over two wide baseline frames cannot be effectively determined via traditional matching techniques due to the large geometric transformation changes [24]. The problem of obtaining a large list of true correspondences between images of building facades with large viewpoint difference is yet to be solved.

We present a novel approach to obtain a substantial number of true matches between images of building facades even under large viewing variations. The fundamental idea behind our approach is that if two regions are true corresponding regions in the two images, they will fit a projective transformation model. As a result, all points within the pair of regions will also fit this transformation model. Therefore, our approach is to first find possible pairs of regions in two images that are corresponded closely to some planar convex quadrilaterals which are often used in man-made objects, *e.g*. construction of buildings. For each pair of regions obtained, we compute a projective transformation model that describes the transformation. Subsequently, interest points within the regions are extracted to evaluate the correctness of each transformation model. The transformation model with the largest set of correctly corresponded interest points thereby has the best pair of corresponding regions in the two images. In contrast with appearance-based matching approaches, our approach works well even under large geometric transformation changes.

The remainder of the paper is organized as follows. Section 2 discusses related work. In section 3, our methodology is presented. Experiments with comparisons follow in section 4 and section 5 concludes the paper.

2009 IEEE 12th International Conference on Computer Vision (ICCV)
978-1-4244-4419-9/09/$25.00 ©2009 IEEE

2. Related work

Detecting and matching features across two images typically involves a feature detector and a feature descriptor. A wide variety of them have already been proposed in the literature. The most widely used feature detector probably is the Harris corner detector [4]. It is not selecting just corners, but rather any image location that has large gradients in all directions at a predetermined scale. However, this detector is not invariant to scale and affine transformations. The Harris-Laplace and Hessian-Laplace detectors [12] are then proposed to deal with affine transformations. They use a (scale-adapted) Harris measure or the determinant of the Hessian matrix to select the location, and the Laplacian to select the scale. Focusing on speed, Lowe [9] approximates the Laplacian of Gaussian (LoG) by a Difference of Gaussians (DoG) filter, but the drawback of LoG and DoG detectors is that local maxima can also be detected in the neighborhood of contours or straight edges, where the signal change is only in one direction. These maxima are less stable because their localization is more sensitive to noise or small changes in neighboring texture. Schaffalitzky and Zisserman [20] extend the Harris-Laplace detector [12] by affine normalization proposed by Baumberg [1]. However, the location and scale of features are provided by the scale invariant Harris-Laplace detector [12] which is not invariant to large affine transformations.

The method by Tuytelaars and Van-Gool [23] is purely intensity-based and starts with extraction of local intensity extrema. Next, the algorithm investigates the intensity profiles along rays going out of the local extremum. An ellipse is fitted to the region determined by significant changes in the intensity profiles. A similar approach based on local intensity extrema was introduced by Matas *et al*. [11], named Maximally stable extremal regions (MSER) detector. They use the water-shed algorithm to find intensity regions and fit an ellipse to the estimated boundaries. The MSER detector, in particular, has been evaluated in [15] to have often better performance than other affine invariant detectors, followed by Harris-Affine and Hessian-Affine detectors [13].

For feature descriptors, Mikolajczyk and Schmid [14] have evaluated a variety of them and identified the Scale-invariant feature transform (SIFT) descriptor [9] as the most resistant to common image deformations. In Lowe's original scheme, he uses the DoG detector [9] to detect features which are invariant to scale changes. In a recent evaluation [16], the SIFT descriptor is found to be the best descriptor for the MSER detector. In our experimental section in section 4, we will compare our matching results with the following approaches: Harris-Affine detector with SIFT descriptor; MSER detector with SIFT descriptor; and the DoG detector with SIFT descriptor.

Other than using feature detectors and descriptors, vanishing points have been used to determine camera orientation in urban environments. However, the assumption that the majority of detected line segments comes from principal vanishing directions associated with the world coordinate frame has to be valid. Furthermore, the detecting of the vanishing points is rather tedious. In Robertson and Cipolla [19]'s approach, the images are transformed into a canonical frame after the dominant vanishing points are recovered. Image features are then detected, followed by pose recovery between views. However, they are unable to distinguish between some viewpoints without additional information, *e.g*. extra query views, due to the similarity of buildings or parts of buildings. Košecká and Zhang [6] proposed a similar approach in which each rectangular structure is defined by four line segments coming from two different orthogonal line's groups. A three stage matching strategy is used to match these rectangular structures, *e.g*. using the ratios of rectangular sizes and normalized cross correlation of canonical views. Given that these ratios and appearances change significantly under large viewpoint changes, their matching approach may hence be unreliable. Additionally, no quantitative data are available to qualify their results.

3. Our methodology

We find quadrilaterals all around us, in everything from fabrics to buildings. We assume that the man-made objects investigated (*e.g*., buildings, windows, doors, etc.) comprise some planar quadrilaterals. Each quadrilateral does not have to have four right angles nor does it have to have four equal sides, as the shape of, *e.g*. a square, will look different under different viewing angles. In addition, for each quadrilateral there must be an edge between two adjacent corners, and each corner is formed by the intersection between two edges. In images, the edges of man-made objects (*e.g*. building facade) can be effectively characterized by line segments [17].

Since the line segments of a quadrilateral will form four sides and corners, our assumption is that at least one of these corners can be detected in order for our algorithm to be able to locate this quadrilateral. We call this corner a "geometric corner". Fig. 1 illustrates the finding of geometric corners in images of building facades under different cases (*e.g*. viewpoint changes). The reason why we choose geometric corners over any appearance-based corners (*e.g*. those detected by Harris corner detector [4]) is that geometric corners are always real corners in man-made objects where it comes from an intersection of two line segments (edges).

Four line segments can be connected to form a planar quadrilateral which we called a "q-region". This q-region is a region within the image and formed from line segments instead of corners. It corresponds closely to a planar structure in the real world so that we can derive a projective transformation (homography) model from each respective pair in two images. We generate all possible q-regions from the de-

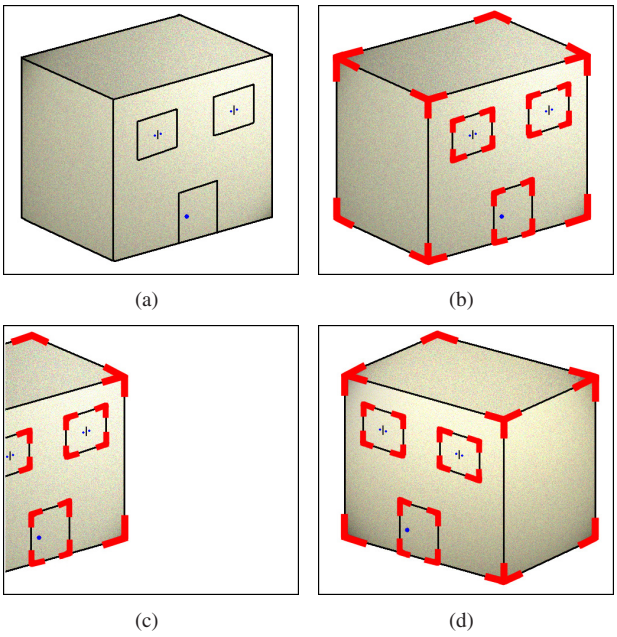


Figure 1. Geometric corners found in images of building facades under different cases. **(a)** A block represents a building on a 2D image plane. **(b)** Geometric corners (represented in red) found in block. **(c)** Block is partially shifted out of the image but geometric corners can still be found. **(d)** Block changed in viewpoint but geometric corners can still be found.

tected line segments to provide robustness against partially occluded or shifted out of images building facades. In order to reduce the chance of obtaining false matches in scenes comprising repetitive and symmetric patterns (*e.g.* building facades), we try to eradicate unnecessary scene in one image which does not exist in the other image by constricting the pair of images to the q-regions enclosing the same scene. This differs from many existing approaches which blindly match the entire images and inevitably obtain a significant number of false matches.

To this end, we extract Harris points [4] within the q-regions. The projective transformation (homography) model with the largest set of corresponded Harris points within two q-regions thereby has the best pair of corresponding q-regions and the most number of point correspondences in the two images.

There are three key stages in our approach. First, line segments are extracted from an edge image. Second, we hypothesize possible q-regions in the image. And lastly, we locate the best projective transformation model derived from a pair of q-regions in two images that returns the largest set of corresponded interest points.

3.1. Line fitting algorithm

We extract line segments from an edge image by a strip fitting algorithm [8]. Let $L = \{\ell_1, \cdots, \ell_n\}$ denote a set of line segments found from [8]. Due to the well-known fragmentation of edge maps of real images, missing edge pixels will result in the breaking of a line segment. To circumvent this problem, a post-processing step is applied to identify and join broken line segments. Two line segments ℓ_i and ℓ_j are joined if the following condition is satisfied,

$$Cond_{join}(\ell_i, \ell_j): \\ \theta > \tau_\theta \; and \; \min\left(\|P_{k,i}, P_{k,j}\|_2\right) < \tau_d, \forall k \in \{1,2\}, \quad (1)$$

where θ is the smaller internal angle between line segments ℓ_i and ℓ_j, and τ_θ denote the allowed internal angle threshold between the line segments. $P_{k,i}$ denote the k^{th} end point of line segment ℓ_i, and τ_d denote the allowed distance threshold between the shortest end-to-end points of the line segments. We denote $\ell_{i,j}$ as the line segment formed by joining ℓ_i with ℓ_j, and $\hat{L}$ to be the set of new line segments that are created. The complete set of line segments that will be used in the next stage is $\left\{L \cup \hat{L}\right\}$, where all line segments before and after the post-processing step are retained. Thus there is no risk of joining two line segments that are not supposed to be joined (not broken) and end up losing them permanently.

3.2. Q-region hypothesis extraction

The outputs from the line fitting approach in the previous stage are used to locate possible q-regions in the images. The approach is presented below.

3.2.1 Initial geometric corner detection

As discussed, our assumption is that at least one geometric corner, described as an intersection point between two line segments, can be found within each image of building facades. Given two line segments ℓ_i and ℓ_j (represented in red in Fig. 2(a)) as first and second line segments with their end-to-end points as P_1P_2 and P_3P_4 respectively, $P_1 = (x_1, y_1)$, $P_2 = (x_2, y_2)$, $P_3 = (x_3, y_3)$, and $P_4 = (x_4, y_4)$, the intersection point $(x_{i,j},y_{i,j})$ between ℓ_i and ℓ_j in parametric form is defined as

$$x_{i,j} = x_1 + t_i(x_2 - x_1), \\ y_{i,j} = y_1 + t_i(y_2 - y_1), \quad (2)$$

where $(x_{i,j},y_{i,j})$ is within the image. Solving for t_i yields

$$t_i = \frac{(x_4 - x_3)(y_1 - y_3) - (y_4 - y_3)(x_1 - x_3)}{(y_4 - y_3)(x_2 - x_1) - (x_4 - x_3)(y_2 - y_1)}. \quad (3)$$

If the denominator is not 0, and if $0 \leq t_i \leq 1$, then the line segments intersect. An initial region $R_{i,j}$ (shaded green in Fig. 2(b)) detected at this stage is formed by the intersection point $(x_{i,j},y_{i,j})$ together with the the first and second line segments ℓ_i and ℓ_j extended to the end of the image as shown in Fig. 2(b).

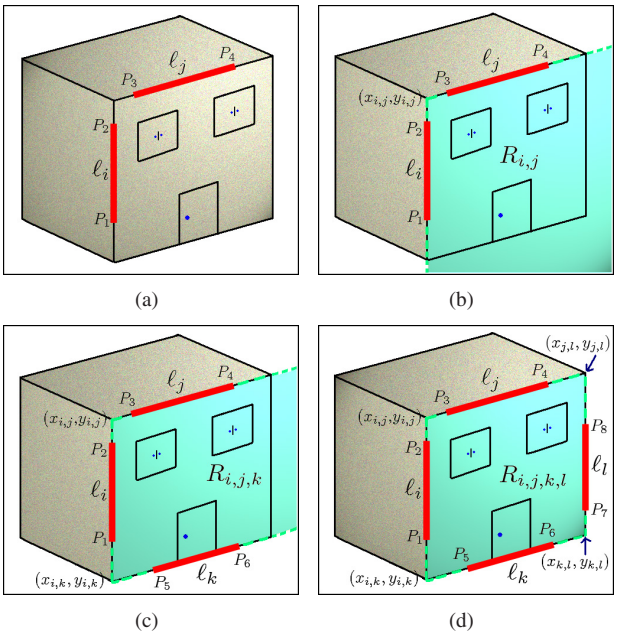


Figure 2. Illustration of q-region formation. **(a)** Two line segments ℓ_i and ℓ_j (represented in red) detected. **(b)** Initial region $R_{i,j}$ (shaded green) formed by an intersection point $(x_{i,j},y_{i,j})$, and two line segments ℓ_i and ℓ_j. **(c)** Refined region $R_{i,j,k}$ formed by two intersection points $(x_{i,j},y_{i,j})$ and $(x_{i,k},y_{i,k})$, and three line segments ℓ_i, ℓ_j, and ℓ_k. **(d)** Q-region $R_{i,j,k,l}$ (shaded green) formed by four intersection points $(x_{i,j},y_{i,j})$, $(x_{i,k},y_{i,k})$, $(x_{j,l},y_{j,l})$, and $(x_{k,l},y_{k,l})$, and four line segments ℓ_i, ℓ_j, ℓ_k, and ℓ_l.

3.2.2 Q-regions formation

We have to find from one geometric corner all the other supporting line segments to form q-regions. Let ℓ_k with end-to-end points P_5P_6 be the line we are going to check if it can be a possible third line segment to form a q-region. We hypothesize a possible third line segment for a q-region if at least one point P_5 or P_6 is within $R_{i,j}$, and $y_1 \leq y_{i,k} \leq y_2$ where $(x_{i,k}, y_{i,k})$ is the point where ℓ_i and ℓ_k intersect. Fig. 2(c) shows a third line segment is found and we have our refined region as $R_{i,j,k}$ (shaded green). The intersection point can be achieved by applying the same equations as Eq. (2-3). The only difference is that the intersection point $(x_{i,k}, y_{i,k})$ here does not have to be within the image. This is for dealing with building facades that are partially shifted out of the image.

Finally, we are to find the fourth line segment in a similar fashion as the third line segment ℓ_k found previously. Let ℓ_l with end-to-end points P_7P_8 be the line we are going to check if it can be a possible fourth line segment to form a q-region. However instead of finding only one point where ℓ_i and ℓ_k intersect, now we find two intersection points, one between ℓ_k and ℓ_l and the other between ℓ_j and ℓ_l. We hypothesize a possible fourth line segment for a q-region if at least one point P_7 or P_8 is within $R_{i,j,k}$. In addition, $x_3 < x_{j,l} \leq x_4$ where $(x_{j,l}, y_{j,l})$ is the point where ℓ_j and ℓ_l intersect, and $x_5 < x_{k,l} \leq x_6$ where $(x_{k,l}, y_{k,l})$ is the point where ℓ_k and ℓ_l intersect. Fig. 2(d) shows the formation of a q-region $R_{i,j,k,l}$ (shaded green) which the entire face of the block is found. This process is repeated to find all possible third and fourth line segments for generating all possible q-regions using this geometric corner.

The example discussed assumes a geometric corner is initially found on the top-left corner of a building facade. However, the initial geometric corner can also be found on the bottom-left, top-right, or bottom-right corner of a building facade if the building facade is not partially occluded or partially shifted out of the image. The steps will remain very similar and the resulting q-region will still be the same. For example, if the initial geometric corner is found on the bottom-left corner of a building facade, the initial region $R_{i,k}$ (shaded green) and the final resulting q-region $R_{i,j,k,l}$ formed will be as shown in Fig. 3(a) and 3(b) respectively. Duplicated q-regions will be removed even if they were generated from different initial geometric corner.

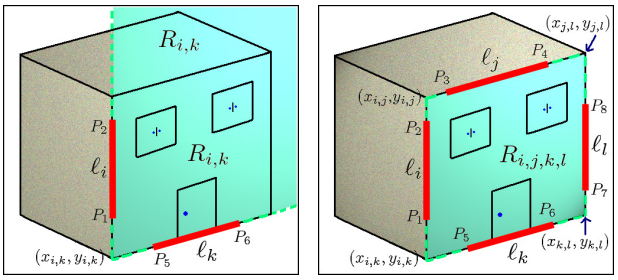


Figure 3. Finding q-region with a different initial geometric corner (bottom-left corner). **(a)** Initial region $R_{i,k}$ (shaded green) formed by an intersection point $(x_{i,k},y_{i,k})$, and two line segments ℓ_i and ℓ_k (represented in red). **(b)** Resulting q-region $R_{i,j,k,l}$.

3.2.3 Additional q-regions generation

In order to better deal with the issue that part of the building facade may be occluded or shifted out of the image, we generate all possible combinations of q-regions using all the vertical and horizontal line segments detected, not necessarily to use the original four line segments. Fig. 4 illustrates an example where the building facade is partially shifted out of the image and there are four vertical line segments including the first line segment ℓ_i detected, a total of ${}^4C_2 = 6$ q-regions is generated.

3.3. Best model selection

Since each q-region will have four corner points $(x_{i,j},y_{i,j})$, $(x_{i,k},y_{i,k})$, $(x_{j,l},y_{j,l})$, $(x_{k,l},y_{k,l})$, we take the respective four corner points to instantiate a projective trans-

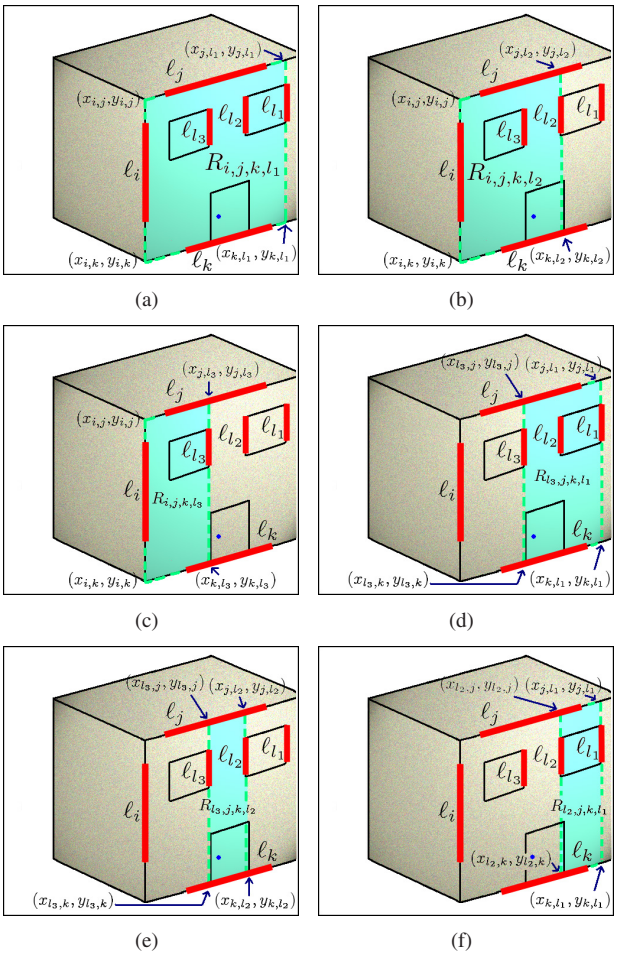


Figure 4. All possible combinations of q-regions (shaded green) formed from 4 vertical line segments ℓ_{l_1}, ℓ_{l_2}, ℓ_{l_3}, and ℓ_i, giving ${}^4C_2 = 6$ q-regions. **(a)** Q-region R_{i,j,k,l_1} generated from line segments ℓ_i, ℓ_j, ℓ_k and ℓ_{l_1}. **(b)** Q-region R_{i,j,k,l_2} generated from line segments ℓ_i, ℓ_j, ℓ_k and ℓ_{l_2}. **(c)** Q-region R_{i,j,k,l_3} generated from line segments ℓ_i, ℓ_j, ℓ_k and ℓ_{l_3}. **(d)** Q-region R_{l_3,j,k,l_1} generated from line segments ℓ_{l_3}, ℓ_j, ℓ_k and ℓ_{l_1}. **(e)** Q-region R_{l_3,j,k,l_2} generated from line segments ℓ_{l_3}, ℓ_j, ℓ_k and ℓ_{l_2}. **(f)** Q-region R_{l_2,j,k,l_1} generated from line segments ℓ_{l_2}, ℓ_j, ℓ_k and ℓ_{l_1}.

formation (homography) model between them. If the two q-regions are true corresponding regions, they should return many corresponded locations of detected Harris points [4]. The homography equation is given by

$$\begin{bmatrix} x_i^{'} \\ y_i^{'} \\ 1 \end{bmatrix} \cong \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}. \quad (4)$$

Let P_k and $P_k^{'}$ denote a Harris corner point in the first and second image respectively. We represent the homography model as H_m and denote the projection of point P_k by H_m to be $H_m(P_k)$. The correctness of the homography model and hence the confidence between a pair of point correspondences of the two q-regions is then computed as

$$Conf(H_m) = \sum_k \Big[\|H_m(P_k), P_*^{'}\|_2 \leq dist_{offset} \Big], \quad (5)$$

where

$$P_*^{'} = \underset{i}{\operatorname{argmin}} \Big(\|H_m(P_k), P_i^{'}\|_2 \Big), \quad (6)$$

$dist_{offset}$ is the distance offset, and [.] is a binary indicator function. The Harris detector [4] detects a very high number of interest points and the repeatability is high. This is appropriate for our approach for locating corresponded points in two images using projective transformation (homography).

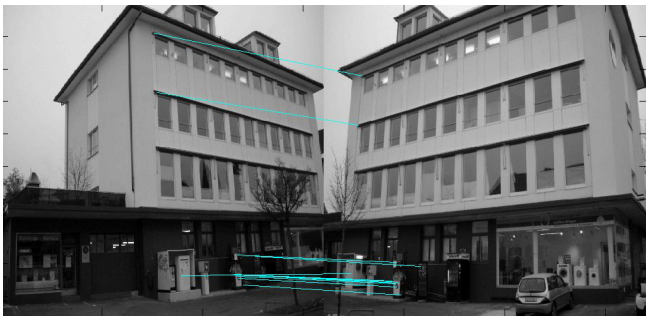
4. Experiments

Although our approach is inherently invariant to image scale, translation, and rotation, experiments on these are not reported due to limitation of space in this paper. Our main purpose in this paper is to illustrate the ability to deal with large viewpoint changes in images of building facades. In the following, we validate our approach by matching images of building facades under different viewpoint change conditions. We select a total of 125 pairs of images from the ZuBud database [21] which comprises of views of building scenes. The 125 pairs of images are deliberately selected in which they can be equally divided into 5 sets of 25 pairs each. Each set is categorized by difference in viewpoint angle between 1° to 20°, 21° to 30°, 31° to 40°, 41° to 50°, and above 50°. Furthermore, we compare our matching results with the matching results obtained from three of the currently most popular approaches, which are the (1) Harris-Affine detector with SIFT descriptor, (2) MSER detector with SIFT descriptor, and (3) DoG detector with SIFT descriptor. We visually inspect the correctness of each match as homographies cannot be used to inspect the correctness of the matches in our experiments because our approach is based on homography. We have performed separate experiments with and without RANSAC homography [3] and find that RANSAC homography can improve on the above three approaches by rejecting false matches. It has been proven to work very well in many applications (*e.g.* in [2, 10, 23]) and will only fail when the number of false matches is too severe, *e.g.* above 80%. The results discussed here will include only the ones with RANSAC homography for the three approaches since the results are significantly improved.

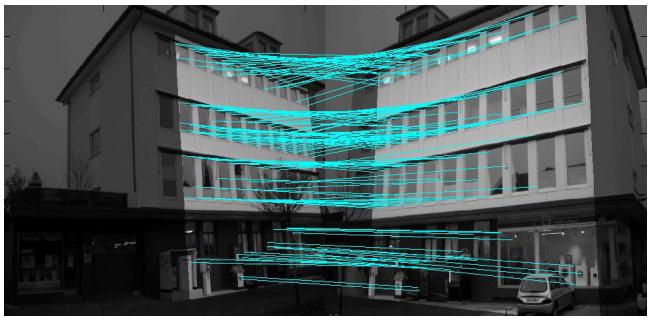
Figure 5 shows results of the four approaches for matching two images of building facades over 50° view angle difference. Matching images like these is extremely challenging due to the large viewpoint difference and the repetitive patterns in man-made objects. The three approaches (other than our approach) have as little as only 1 true match

(a) Harris-Affine + SIFT + RANSAC homography

(b) MSER + SIFT + RANSAC homography



(c) DoG + SIFT + RANSAC homography



(d) Our approach

Figure 5. Matching two images of building facades over 50° view angle difference. Note that the top-right corner of building facade in the right image is shifted out of the image. **(a)** Harris-Affine detector + SIFT descriptor + RANSAC homography: 9 matches, 3 pairs are true matches. **(b)** MSER detector + SIFT descriptor + RANSAC homography: 10 matches, 1 pair is true match. **(c)** DoG detector + SIFT descriptor + RANSAC homography: 11 matches, 10 pairs are true matches. **(d)** Our approach: 134 matches, 134 pairs are true matches.

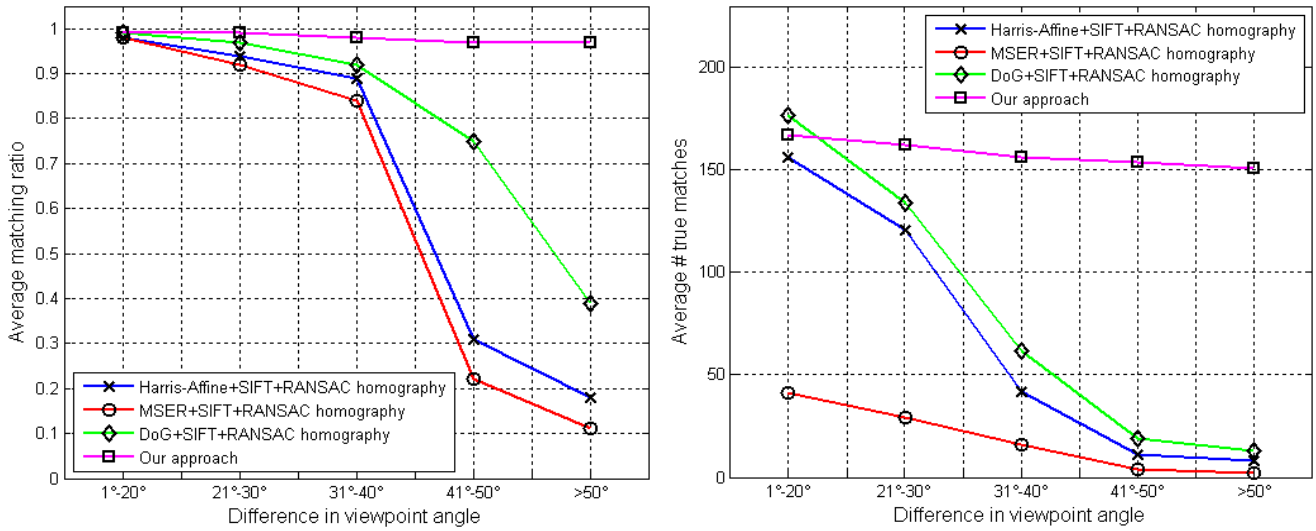
to a maximum of 10 true matches by the DoG + SIFT + RANSAC homography approach (see Fig. 5(a)-5(c)). Our approach obtains 134 matches and all are true matches as shown in Fig. 5(d). It has no problem finding the pair of corresponding q-regions in the two images even though the top-right corner of the building facade in the right image is shifted out of the image. It achieves above 10 times more true matches than the three approaches.

We have done a quantitative comparison on the entire sets of images, and the average results are presented in Fig. 6. In Fig. 6(a), we can see that the average matching ratio for all of the approaches are above 0.8 when the difference in viewpoint angle is $\leq 40^\circ$. Verification with RANSAC homography on the three approaches (not on our approach) plays an important part here by rejecting many false matches in order for them to obtain such good results. However, the average matching ratio for all except our approach start to decrease tremendously when the difference in viewpoint angle is $> 40^\circ$. Our approach remains stable even after reaching the $> 50^\circ$ limit.

In Fig. 6(b), we can see that the number of true matches get lesser when the difference in viewpoint angle increases. There are barely any true matches from the three approaches when the difference in viewpoint angle goes $> 50^\circ$. However for our approach, the number of true matches still remain above the 100 true matches line even when the difference in viewpoint angle reached $> 50^\circ$. We achieved a tenfold or more increase in true matches when compared to the rest of the approaches. This shows the robustness of our approach in viewpoint changes for building facades.

5. Conclusion

The main contribution of this paper is a novel concept for finding a substantial number of true correspondences between images of building facades even under large viewing variations. The presented approach integrates both the appearance and geometric properties of an image to find true matches by first extracting line segments using a robust line fitting algorithm. Next, we find quadrilaterals in the images using these line segments since man-made objects (*e.g.* building facades) often comprise some planar convex quadrilaterals. We named the detected quadrilaterals as "q-regions". The fundamental concept behind our approach is that if two q-regions are true corresponding q-regions in two images, they will fit a projective transformation (homography) model. As a result, all points within the pair of regions will also fit this transformation model. We extract Harris points from the images to evaluate the correctness of each transformation model. The model which returns the largest set of corresponded Harris points is selected as the



(a) (b)

Figure 6. Results of matching vs. difference in viewpoint angle for images of building facades. **(a)** Average matching ratio (number of true matches/total number of matches). **(b)** Average number of true matches.

best transformation model.

We compared our performance with other popular state of the art techniques. The comparison has shown that the performance of others declines as the difference in viewpoint angle increases, and only ours has remained stable throughout with a tenfold increase in true matches compared to others.

References

[1] A. Baumberg. Reliable feature matching across widely separated views. In *CVPR*, pages 774–781, 2000.

[2] M. Brown and D. G. Lowe. Recognising panoramas. In *ICCV*, pages 1218–1225, 2003.

[3] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. In *ECCV*, pages 683–695, 1996.

[4] C. Harris and M. Stephens. A combined corner and edge detector. In *AVC*, pages 147–151, 1988.

[5] T. Kadir, A. Zisserman, and M. Brady. An affine invariant salient region detector. In *ECCV*, pages 228–241, 2004.

[6] J. Košecká and W. Zhang. Extraction, matching and pose recovery based on dominant rectangular structures. *CVIU*, 100(3):274–293, 2005.

[7] J. A. Lee, K.-C. Yow, and A. Sluzek. Image-based information guide on mobile devices. In *ISVC*, pages 346–355, 2008.

[8] M. K. Leung and Y.-H. Yang. Dynamic two-strip algorithm in curve fitting. *Pattern Recognition*, 23(1-2):69–79, 1990.

[9] D. G. Lowe. Object recognition from local scale-invariant features. In *ICCV*, pages 1150–1157, 1999.

[10] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60:91–110, 2004.

[11] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. In *BMVC*, pages 384–393, 2002.

[12] K. Mikolajczyk and C. Schmid. Indexing based on scale invariant interest points. In *ICCV*, pages 525–531, 2001.

[13] K. Mikolajczyk and C. Schmid. An affine invariant interest point detector. In *ECCV*, pages 128–142, 2002.

[14] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *PAMI*, 27(10):1615–1630, 2005.

[15] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van-Gool. A comparison of affine region detectors. *IJCV*, 65(1-2):43–72, 2005.

[16] P. Moreels and P. Perona. Evaluation of features detectors and descriptors based on 3d objects. *IJCV*, 73(3):263–284, 2007.

[17] V. S. Nalwa. Line-drawing interpretation: Straight lines and conic sections. *PAMI*, 10(4):514–529, 1988.

[18] S. Obdrzálek and J. Matas. Object recognition using local affine frames on distinguished regions. In *BMVC*, pages 113–122, 2002.

[19] D. Robertson and R. Cipolla. An image-based system for urban navigation. In *BMVC*, pages 819–828, 2004.

[20] F. Schaffalitzky and A. Zisserman. Multi-view matching for unordered image sets. In *ECCV*, pages 414–431, 2002.

[21] H. Shao, T. Svoboda, and L. Van-Gool. Zubud-zurich buildings database for image based recognition. Technical Report 260, 2003.

[22] D. Tell and S. Carlsson. Combining appearance and topology for wide baseline matching. In *ECCV*, pages 68–81, 2002.

[23] T. Tuytelaars and L. Van-Gool. Wide baseline stereo based on local, affinely invariant regions. In *BMVC*, pages 412–422, 2000.

[24] J. Xiao and M. Shah. Two-frame wide baseline matching. In *ICCV*, pages 603–609, 2003.