A METHOD OF LOCAL FEATURES BASED IMAGE MATCHING

Naga Sudha

Research Scholar ,Dept of ECE JNTUCEH

Abstract: The challenges in local-feature-based image matching are variations of view and illumination. Many methods have been recently proposed to address these problems by using invariant feature detectors and distinctive descriptors. However, the matching performance is still unstable and inaccurate, particularly when large variation in view or illumination occurs. In this paper, a view and illumination invariant image-matching method is proposed. Estimation the relationship of the relative view and illumination of the images, transform the view of one image to the other, and normalize their illumination for accurate matching is done iteratively. This method does not aim to increase the invariance of the detector but to improve the accuracy, stability, and reliability of the matching results. The performance of matching is significantly improved and is not affected by the changes of view and illumination in a valid range. The proposed method would fail when the initial view and illumination method fails, which gives a new sight to evaluate the traditional detectors. Two novel indicators for detector evaluation, namely, valid angle and valid illumination, which reflect the maximum allowable change in view and illumination, respectively are proposed.

Introduction: Image matching is a fundamental issue in computer vision. It has been widely used in tracking , image stitching , 3-D reconstruction , simultaneous localization and mapping (SLAM) systems , camera calibration , object classification, recognition, and so on. Image matching aim to find the correspondence between two images of the same scene or objects in different pose, illumination, and environment. Focus on local feature-based image matching. The challenges of this work reside in stable and invariant feature extraction from varying situations and robust matching. In image matching, key region or point of interest is often used as the local feature due to its stable performance in detection and description. A region feature is usually derived from a circle or ellipse with certain location and radius and is effective and efficient, compared with other types of features such as edges and contours. Therefore, region features are extensively used in real applications.

Interesting points are extracted from images,and the region of interest is the associated circuilar or elliptical region aroung the interesting point.GenerallyHarris,SUSAN,CSSetc or centre of silent region SIFT,SURF,DoH as the interesting point since they are stable and easy to locate and describe. The radius of the region is determined by a priori setting Harris corner or the region scale invariant features.The total number of features detected is the minimum number of the features extracted from the matched image.

Color,structure and texture are widely used to describe images.Descriptors with edge orientation information SIFT and HOG are used since they are more robust to scale,blur and rotation.Matching features:Local features from two images are first matched when they are the nearest pair. A handful of distances can be used in practice such as L1 distance,L2 distance,Histogrma intersection distance and earth mover’s distance.If the nearest distance is higher than k times of the second nearest distance, the nearest matching pair will be removed.These are the very initial matching results.Then the priori hypothesis of the object transform filters the un-uniform transformed matches. imply use planar objects to show the effectiveness of the proposed method.

For the multi-transform problem, the proposed method could be also integrated. Random sample consensus (RANSAC),is used to select the uniform or multiple transformations set from all the matches.

The three parts of the detect–describe–match (DDM) framework determine the performance of image matching.

The first step is the basis of this framework. Unstable and variant features increase the difficulties of the next two steps. Researchers mostly focus on the first step for invariant feature extraction and have proposed many excellent detectors .However, an important experience of a pervious work is that all the aforementioned feature detectors are not strictly invariant to the changes of view and illumination. The same interestingregions extracted from the matching images tend to be fewerand fewer when increasing the variation of view or illumination. For larger changes, there would be few invariant features that can be extracted from both images to be matched.

This Describingregions. Color, structure, and texture are widely used to describe images in the recent literature. Descriptors with edge orientation information (SIFT and HOG

Matching features: Local features from two images are first matched when they are the nearest pair. A handfulof distances can be used in practice, such as L_1 distance, L_2 distance, histogram intersection distance, and earth mover's distance . If the nearest distance is higher than k times of the second nearest distance, the nearest matching pair will be removed. These are the very initial matching results. Then the priori hypothesis of the object transform filters the un-uniform transformed matches.

In this we simply use planar objectsto show the effectiveness of the proposed method. Forthemultitransform problem, the proposed method could be also integrated. Random sample consensus (RANSAC),is used to select the uniform or multiple transformations set from all the matches.Te three parts of essential difference of images with different view and illumination. Normally, a question need to be answered: whether an object in two images with different views and illumination looks like the same one, supposing there are two images with a large view change, The two top images are the same object in different views. They are so different in appearance that they can be considered as two different objects. Do not attempt to find invariant local feature detectors as in a previous work but focus on a better framework for image matching

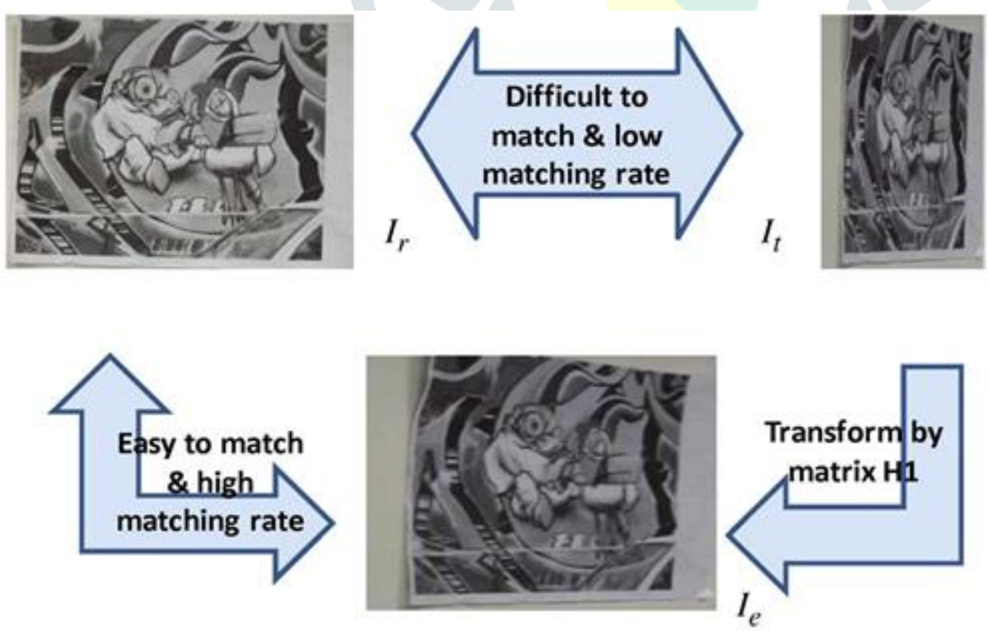


Fig:1Illustration of the proposed matching algorithm. I_r and I_t are the images to be matched. I_e is simulated from I_t by transformation L . I_r is difficult to I_t match with for the difference of view point and illumination, whereas I_e is easier to match with I_t since they are closer in the parameter space.

An iterative image-matching framework that iterates the estimation of pose and illumination to improve the matching performance. First, transform the view and illumination of the image by estimating the pose and illumination correspondence between the matching pair by an initial detector, e.g., Harris, SIFT, SURF , and HLSIFD.

Then, extract local features from the simulated image and match them with the features in another image. With this framework, the repeatability score (RS) and the number of correct matches (NCMs) could be stabilized under heavy variations in a valid range. Out of the valid range (larger view or illumination change), our method will fail to obtain correct matching result. Find that every feature detector under framework has a considerable tolerance to the changes of view and illumination.

SIFT:SIFT is an algorithm in computer vision to detect and describe local features in images. Applications include object recognition, robotic mapping and navigation, image stitching, 3D modeling, gesture recognition, video tracking, individual identification of wildlife and match moving.It is used for detection and extracting local features of an image.the steps are as follows:

1. Generate a Difference of Gaussian(DoG) or a laplacian pyramid
2. Extrema detection from the DoG pyramid which is the local maxima and minima, the point found is an extrema
3. Eliminate low contrast or poorly localized points, what remains are the keypoints
4. Assign an orientation to the points based on the image properties
5. Compute and generate keypoint descriptors

The main SIFT implementation consists of 4 major stages described by Lowe . The first step is finding a scale space extrema. This is done via image pyramids. The process involves repeatedly smoothing an image by convolving it with a gaussian operator and subsequently sub-sampling in order to achieve higher levels of the pyramid.

$$G\ (x, y, \sigma) = \frac{1}{2\pi\sigma^2} - (x^2 + y^2)/2\sigma^2$$

$$L\ (x, y, \sigma) = G(x, y, \sigma) * I\ (x, y)$$

After obtaining a full pyramid, the difference of each octave results in an approximate solution to the Laplacian of Gaussian

$$m(x,y) = \sqrt{(L\ (x+1,y) - Lx-1,y))^2 + (L(x,y+1) - L\ (x,y-1))^2}$$

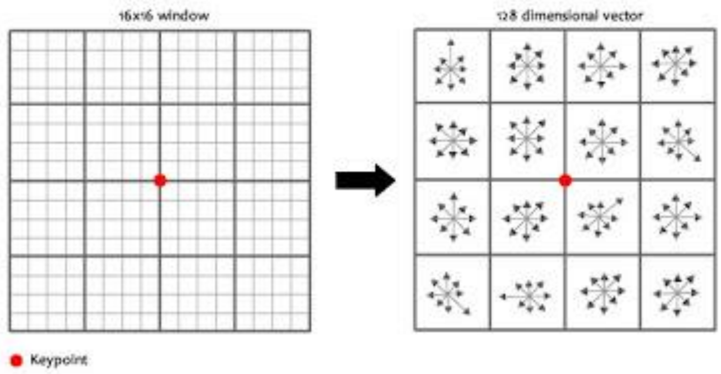
$$O(x,y) = tan^{-1}((L(x,y+1) - L(x,y-1))/(L(x+1,y) - L(x-1,y)))$$

The next step is to iterate through each pixel and compare it with its 8 surounding neighbors and 9 neighbors at one scale higher and lower. If the pixel value is higher or lower amongst its neighbors then it is considered as a keypoint.

After determining the keypoint, tylor expansion is used to generate subpixel values from the pixel data.

$$D\ (x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 D}{\partial x^2} x$$

Doing so, will generate a lot of keypoints. These keypoints are low contrast points or points along an edge. To eliminate these points, tylor expansion is used to get an intensity value for each subpixel. It is then compared against a threshold. If the value is higher or lower than the threshold, the point is accepted or rejected. Following this step a series of stable keypoints will be left that are scale invariant. To make the keypoints rotation invariant, a weighted histogram of local gradient direction and magnitudes around each keypoint is computed at a selected scale and the most prominent orientation in that region is selected and assigned to the keypoint.



David Lowes method:SIFT key points of objects are first extracted from a set of reference images and stored in a database. An object is recognized in a new image by individually comparing each feature from the new image to this database and finding candidate matching features based on Euclidean distance of their feature vectors. From the full set of matches, subsets of key points that agree on the object and its location, scale, and orientation in the new image are identified to filter out good matches.

SIFT:SIFT is an algorithm in computer vision to detect and describe local features in images. Applications include object recognition, robotic mapping and navigation, image stitching, 3D modeling, gesture recognition, video tracking, individual identification of wildlife and match moving.It is used for detection and extracting local features of an image.the steps are as follows:

1. Generate a Difference of Gaussian(DoG) or a laplacian pyramid
2. Extrema detection from the DoG pyramid which is the local maxima and minima, the point found is an extrema
3. Eliminate low contrast or poorly localized points, what remains are the keypoints
4. Assign an orientation to the points based on the image properties
5. Compute and generate keypoint descriptors

The main SIFT implementation consists of 4 major stages described by Lowe . The first step is finding a scale space extrema. This is done via image pyramids. The process involves repeatedly smoothing an image by convolving it with a gaussian operator and subsequently sub-sampling in order to achieve higher levels of the pyramid.

$$G\,(x, y, \sigma) = \frac{1}{2\pi\sigma^2} - (x^2 + y^2)/2\sigma^2$$

$$L\,(x, y, \sigma) = G(x, y, \sigma) * I\,(x, y)$$

After obtaining a full pyramid, the difference of each octave results in an approximate solution to the Laplacian of Gaussian

$$m(x,y) = \sqrt{(L\,(x+1,y) - Lx-1,y))^2 + (L(x,y+1) - L\,(x,y-1))^2}$$

$$O(x,y) = tan^{-1}((L(x,y+1) - L(x,y-1))/(L(x+1,y) - L(x-1,y)))$$

The next step is to iterate through each pixel and compare it with its 8 surounding neighbors and 9 neighbors at one scale higher and lower. If the pixel value is higher or lower amongst its neighbors then it is considered as a keypoint.

After determining the keypoint, tylor expansion is used to generate subpixel values from the pixel data.

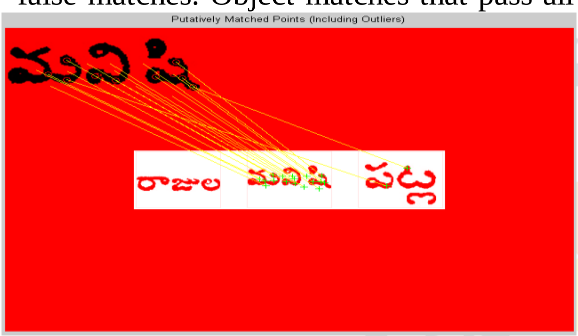
$$\boldsymbol{D\,(x) = D + \frac{\partial D^T}{\partial x}\,x + \frac{1}{2}x^T\frac{\partial^2 D}{\partial x^2}\,x}$$

Doing so, will generate a lot of keypoints. These keypoints are low contrast points or points along an edge. To eliminate these points, tylor expansion is used to get an intensity value for each subpixel. It is then compared against a threshold. If the value is higher or lower than the threshold, the point is accepted or rejected. Following this step a series of stable keypoints will be left that are scale invariant. To make the keypoints rotation invariant, a weighted histogram of local gradient direction and magnitudes around each keypoint is computed at a selected scale and the most prominent orientation in that region is selected and assigned to the keypoint.

The determination of consistent clusters is performed rapidly by using an efficient hash table implementation of the generalized Hough transform. Each cluster of 3 or more features that agree on an object and its pose is then subject to further detailed model verification and subsequently outliers are discarded.Finally the probability that a particular set of features indicates the presence of an object is computed, given the accuracy of fit and number of probable false matches. Object matches that pass all these tests can be identified as correct with high confidence.

Features: The detection and description of local image features can help in object recognition. The SIFT features are local and based on the appearance of the object at particular interest points, and are invariant to image scale and rotation. They are also robust to changes in illumination, noise, and minor changes in viewpoint. In addition to these properties, they are highly distinctive, relatively easy to extract and allow for correct object identification with low probability of mismatch.

They are relatively easy to match against a (large) database of local features but however the high dimensionality can be an issue, and generally probabilistic algorithms such as k-d trees with best bin first search are used .



RECALL	PRECISION
0.1429	1.0000
0.7143	0.8333
0.7143	0.4545
0.8571	0.3750
0.8571	0.2857
0.8571	0.1304
1.0000	0.0814
1.0000	0.0769
1.0000	0.0729
1.0000	0.0693
1.0000	0.0660

The query image is matched with more interest point in another image such that matching score is high in this case**.**

Conclusions : The SIFT features improve on previous approaches by being largely invariant to changes in scale, illumination, and local affine distortions. The large number of features in a typical image allow for robust recognition under partial occlusion in cluttered images. Scale-invariant edge groupings that make local figure-ground discriminations would be particularly useful at object boundaries where background clutter can interfere with other features. The indexing and verification framework allows for all types of scale and rotation invariant features to be incorporated into a single model representation. Maximum robustness would be achieved by detecting many different feature types and relying on the indexing and clustering to select those that are most useful in a particular image

References:

[1] J. Shi and C. Tomasi, "Good features to track," in Proc. Comput.Vision and Pattern Recognit., Jun. 1994, pp. 593–600.

[2] Y. Li, Y. Wang, W. Huang, and Z. Zhang, "Automatic image stitchingusing sift," in Proc. Audio, Lang. Image Process., Jul. 2008, pp.568–571.

[3] R. Szeliski, "Image alignment and stitching: a tutorial," Found. TrendsComput. Graph Vis. vol. 2, no. 1, pp. 1–104, 2006

[4] M. Brown and D. Lowe, "Unsupervised 3D object recognition and reconstruction in unordered datasets," in Proc. Int. Conf. *3-D Digit*. ImagModel., Jun. 2005, pp. 56–63.

[5] A. Davison, W. Mayol, and D. Murray, "Real-time localization andmapping with wearable active vision," in Proc.Int.Symp.Mixed AugmentedReality, 2003, pp. 18–27.

[6] B. Telle, M. J. Aldon, and N. Ramdani, "Camera calibration and 3dreconstruction using interval analysis," in Proc. Int. Conf. Image Anal.Process., 2003, pp. 374–379.