Original papers

Towards feature points based image matching between satellite imagery and aerial photographs of agriculture land

Sajid Saleem [a,*], Abdul Bais [b], Robert Sablatnig [c]

[a] Department of Electrical Engineering, University of Wah, Quaid Avenue, Wah Cantt, Pakistan
[b] Faculty of Engineering and Applied Science, University of Regina, Regina, Saskatchewan, Canada
[c] Computer Vision Lab, Institute of Computer Aided Automation, Vienna University of Technology, 1040 Vienna, Austria

ARTICLE INFO

ABSTRACT

This paper focuses on image matching between satellite imagery and aerial photographs of agriculture land. Feature points are used for image matching. The satellite imagery and aerial photographs were acquired at different times, viewpoints, sensors, and altitudes. Therefore, they possess very high temporal, photometric, and projective differences. When feature points, such as Scale Invariant Feature Transform (SIFT) and Speeded Up Robust Features (SURF) are applied on such images, they demonstrate poor performance. This paper aims at evaluating the performance of SIFT, SURF, and other state of the art feature points in order to determine well deserving feature points for the images of agriculture land. We also propose a new feature point descriptor, i.e. Modified Normalized Gradient SIFT, which obtains on average 1.73–2.37% better performance than other state of the art descriptors.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Image matching is a popular technique in Computer vision (Mikolajczyk and Schmid, 2005). It computes visual similarity between the images of the same scene taken at different times, viewpoints, scale, rotation, illumination, or sensors, etc. It is used in wide range of applications, such as image retrieval, image segmentation, objection recognition, scene classification, and camera localization. It is based on three fundamental steps: (i) feature point (corners, blob, or intersections of lines) detection, (ii) feature point description, i.e. assigning a feature vector to the neighborhood of each feature point, and (iii) feature vector matching.

In literature, various algorithms for the detection and description of feature points have been proposed. Each algorithm computes feature points in such a manner, so that scale, rotation, illumination, projective, and photometric variations between the images can be overcomed and feature point correspondences can be established reliably and accurately (Lowe, 2004).

The feature point detector algorithm can be traced back to Moravec (1980). He uses a local image window to determine changes in the average pixel intensities by shifting the window by a small amount in various directions for corner detection. Harris and Stephens (1988) revisit this corner detection by using a smoothed circular window to make the corner detection stable under noise and other image variations. Since then the corner detection is known as Harris corner and has been widely used in different fields of the computer vision. Various modifications to corner detection have also been proposed (Mikolajczyk and Schmid, 2001; Mikolajczyk et al., 2005). In this regard the scale space theory of Lindeberg (1993) has been widely used.

Mikolajczyk and Schmid (2001) use the scale space theory for the detection of scale invariant Harris corners, which are called Harris Laplace (HarLap) feature points. Similarly, Lowe (2004) uses the scale space theory for the detection of Scale Invariant Feature Transform (SIFT) keypoints. But he applies the theory by making use of Difference of Gaussian (DoG) filters with Hessian function.

Mikolajczyk et al. (2005) modify the detection of SIFT keypoints. They use Laplacian filters with Hessian function to detect Hessian Laplace (HesLap) feature points. HesLap points are similar to SIFT keypoints and represent blob like image structures, but HesLap demonstrates better scale space accuracy than SIFT. Mikolajczyk et al. (2005) also propose affine invariant features points, which are called Harris Affine (HarAff) and Hessian Affine (HesAff) (Mikolajczyk et al., 2005).

Bay et al. (2006) modify the SIFT keypoint detection. They use integral images (Simard et al., 1999) with Hessian function to detect Speeded Up Robust Features (SURF). They show that SURF demonstrates better performance than SIFT under scale, rotation, affine, and illumination variations and can be detected at near the frame rate.

* Corresponding author.
   E-mail address: sajidslm@yahoo.com (S. Saleem).

Matas et al. (2004) propose Maximally Stable Extremal Regions (MSER) algorithm. This detector algorithm resists continuous transformation of image coordinates, image intensities, and detect regions at near the frame rate.

Rosten and Drummond (2006) also focus on the speed issue. They propose Features from Accelerated Segment Test (FAST) by using a machine learning technique. Rublee et al. (2011) revisit the FAST detector and propose an Oriented FAST and Rotated Binary Robust Independent Elementary Features (ORB) detector, which is a speeded up and scale invariant version of the FAST detector and provides accurate estimation of the keypoint orientations. Similarly, Leutenegger et al. (2011) propose Binary Robust Invariant Scalable Keypoints (BRISK) by making use FAST detector, image pyramid and a saliency criterion.

In the literature, there are numerous feature point detectors. Each detector contributes a large number of feature points for image matching. Such feature points carry no distinct information, which can be used to establish feature point correspondences. To achieve this, various descriptor algorithms have been proposed, which assign feature vectors to the neighborhood of feature points. Such feature vectors are also known as descriptors.

The descriptor algorithm can be traced back to Zhang et al. (1995). They use correlation windows centered at Harris corners. This method, because of its simplicity results in a large number of false correspondences, which are then removed using fundamental matrix (homography). However, the ground-breaking work in the area of descriptor algorithms is by Schmid and Mohr (1997). They compute rotationally invariant descriptors for Harris corners to obtain high precision scores in image matching and image to database matching tasks.

The work of Lowe (2004) is also considered to be ground breaking. He uses the biological vision model of Edelman et al. (1997) to construct SIFT descriptors. He computes directional gradients on image patches (regions) centered at SIFT keypoints and then spatially divides the gradients into 4×4 location bins. For each location bin, he computes a histogram of orientated gradients. Finally, he concatenates histograms over all the location bins to obtain 128 dimensional SIFT descriptor. Since then, the SIFT descriptors have been widely used and also various modifications to SIFT descriptor have been proposed.

Mikolajczyk and Schmid (2005) evaluate different feature point descriptors. They observe that SIFT descriptor obtains the best performance. They also propose a modified version of the SIFT descriptor, which is called Gradient Location and Orientation Histogram (GLOH). The GLOH descriptor is based on log polar location binning scheme but the descriptor size is 272, which is reduced to 128 dimensions by using Principle Component Analysis (PCA).

Ke and Sukthankar (2004) also use PCA for the construction of PCA-SIFT descriptors. They resize image patches centered at SIFT keypoints to 41×41 pixels size. Then concatenate the directional gradients of the resized patches to obtain a PCA-SIFT descriptor. The descriptor dimension is then reduced with PCA. The PCA-SIFT with dimensions less than 128 can be obtained and still it outperforms SIFT in image matching tasks (Ke and Sukthankar, 2004). This also reduces the time complexity and the computational cost associated with the descriptor matching step of the image matching task.

Bay et al. (2006) propose a speeded up version of the SIFT descriptor which is called SURF. This descriptor is based on Haar wavelength responses and the 4×4 location binning scheme of the SIFT. It is a 64 dimensional descriptor, which reduces the computational cost of descriptor matching compared to SIFT.

Heikkilä et al. (2009) focus on monotonic and illumination variations. They use Local Binary Patterns (LBP) scheme (Ojala et al., 2002) to compute Center Symmetric Local Binary Patterns (CS-LBP), which are similar to image gradients but their computation is fast and offer more robustness towards intensity and illumination changes than image gradients. The CS-LBP features result in 256 dimensional descriptors and make the descriptor matching step computationally intensive.

Calonder et al. (2010) propose Binary Robust Independent Elementary Features (BRIEF) descriptor by using a sampling pattern consisting of 128, 256, or 512 intensity comparisons with the sample points selected randomly around the keypoint locations from an isotropic Gaussian distribution. Rublee et al. (2011) revisit the BRIEF descriptor to obtain ORB descriptors. The ORB descriptor is in fact a rotation invariant BRIEF descriptor, which is obtained by analyzing the variance and the correlation of the BRIEF and applying a learning method for de-correlating BRIEF to improve the performance against rotation variations. Leutenegger et al. (2011) propose BRISK descriptor by retrieving gray values around the feature points with a sampling pattern. This pattern also generates the orientation information to make the BRISK descriptor rotation invariant.

Yi et al. (2008) focus on non linear intensity changes, which occur between multisensor images. They show that SIFT descriptor is not invariant to such intensity variations. To cope with this problem they propose Gradient Orientation Modification (GOM), which restricts gradient orientations between 0 and π radians and then uses the restricted orientations in the SIFT algorithm to compute GOM-SIFT descriptors. GOM-SIFT demonstrates 7.04% better performance than SIFT under non linear intensity changes (Yi et al., 2008).

The GOM method on one hand improves the performance of SIFT, but on the other hand effects the rotation invariance of the SIFT descriptor. To cope with this problem, Vural et al. (2009) propose Orientation Restricted (OR) method, which computes SIFT descriptors and then combines the elements of the SIFT descriptors in the opposite orientation directions to obtain the OR-SIFT descriptors.

Saleem and Sablatnig (2013a) modify both gradient magnitudes and orientations to cope with non linear intensity changes. They use CS-LBP scheme for the modification of gradients. This results in Local Binary Patterns of Gradient (LBPG) features. These features are used in the SIFT algorithm to compute LBPG descriptors. The LBPG method is computationally intensive and results in 256 dimensional descriptors. In order to reduce the computational cost and the size of LBPG descriptors, Local Contrast (LC)-SIFT and Differential Excitation (DE)-SIFT have been proposed (Saleem and Sablatnig, 2013b). These methods do not modify the gradient orientations, but only replace the gradient magnitudes with LC (Su et al., 2010) and DE (Chen et al., 2010) magnitudes, respectively, in the SIFT algorithm. This results in 128 dimensional LC-SIFT and DE-SIFT descriptors.

The LC-SIFT and DE-SIFT method is further modified in Saleem and Sablatnig (2014). It is shown that gradient magnitudes can be normalized to obtain binary Normalized Gradient (NG) features, which carry either 0 or 1 magnitude. These NG features are highly distinct and result in NG-SIFT descriptors.

This paper focuses on image matching between Satellite imagery and aerial photographs of the agriculture land. These images were acquired at different times, viewpoints, altitudes, and sensors, therefore, they possess textural, projective, photometric, and non linear intensity variations. Such types of variations make image matching based on feature points difficult to accomplish. This paper evaluates the performance of state of the art feature points in order to determine well deserving feature points for the images of agriculture land. To the best of our knowledge, such an evaluation has not been reported in the literature.

The rest of this paper is organized as follows: Section 2 presents materials and methods. It describes the evaluation criteria, the test images, and our proposed Modified Normalized (MN) gradient SIFT