THINK INDIA JOURNAL

ISSN:0971-1260

Vol-22-Issue-22-December-2019

SPARSE CORRESPONDENCE METHODS

Patil Nitin Santosh[1] and Dr. V. Balakrishna Reddy[2]

Research Scholar[1] and Professor[2], Swami Vivekanand University, Sagar, Madhya Pradesh

ABSTRACT

The majority of computational stereo methods require that the input images are rectified to make the challenges of stereo-vision more tractable. The rectification process, introduced to brings the rows of each image into alignment by warping the image captured by each viewpoint. Although rectification is often accomplished through manual calibration, automatic alignment using sparse feature-based methods is commonly performed if the stereo-head is not reachable or if frequent re-alignment is required.

Keywords: Convential Feature, Phase Congruency, Directional Filters, Thermal Images, Spectral, Aligenment

INTRODUCTION

This paper presents an investigation into the feasibility of conventional feature-based methods for automatic alignment of distinct multi-spectral viewpoints. It provides an overview of the terminology and role of conventional feature-based methods in viewpoint alignment, which is concluded with a brief discussion of the core assumptions of these methods. Conventional feature methods are reported to perform poorly with multi-spectral images; a literature survey is identify where and how these methods fail. The survey serves as a problem statement for the development of an adapted measure which incorporates the invariant elements of phase congruency describes the approach to evaluating conventional and the adapted method. The results of this investigation are provided.

Feature-based alignment

An image feature is a broad term referring to a point in an image that is deemed significant by some measure. For example, an edge is a structural feature usually indicated by a high gradient magnitude. This section introduces the traditional approaches to feature-based alignment through the detection and matching of corresponding observations from different viewpoints.

Conventional approaches

Detection and matching of sparse image features is vital in many applications and is the conventional approach to automatic alignment of images taken from different viewpoints. Numerous methods of feature detection and description have been presented in literature; this section presents a representative sample of common approaches to this task.Many of the earliest detection methods identified corners as stable interest points. The robust Harris corner detector was first presented in 1988 as a refinement of existing approaches to corner detection it identified corners as points in an image where an autocorrelation function was maximised. The Shi-Tomasi corner detector (or the Good Features (GF) corner detector from the title of the paper \Good Features to Track") provided further refinement to this method by reducing the frequency of edge responses, and is an important part of the investigation later in this chapter. Also included is the modern Features from Accelerated Segment Test (FAST) corner detector which is commonly used with the Oriented Fast and Rotated BRIEF (ORB) and Binary Features from Robust Orientation Segment Tests (BFROST) binary descriptors.

Binary descriptors have recently received a large amount of interest due to the computational and memory limitations of mobile devices Binary descriptors are compact binary strings formed by pair-wise intensity comparisons of pixels radially distributed about an interest point; each bit in the descriptor represents the logical outcome of an inequality between the intensity values of each pair. The Binary Robust Invariant Scalable Keypoint (BRISK) descriptor operates on corners detected with the Adaptive and Generic Accelerated Segment Test (AGAST) detector to provide rotation and scale invariance. Binary strings are compared with the Hamming distance: each element is compared with a logical XOR operation, and the sum of the resulting bits represents the binary difference

ISSN:0971-1260

Vol-22-Issue-22-December-2019

Possibly the most common approach found in literature and in practice is Lowe's Scale invariant feature transform (SIFT) method of detection, description and matching SIFT is prominent in literature as it has become a robust benchmark against which the performance of novel methods can be measured The detection stage of SIFT localises blob-like interest points as maxima in scale-space in a Difference of Gaussian (DoG) image pyramid. The descriptor itself is a 16 16 pixel window in which each pixel contributes a vote, weighted by gradient magnitude, to a histogram of discrete orientations used to describe the region. The notable Speeded Up Robust Features (SURF) method is based on the SIFT descriptor and provides an efficient alternative at the cost of matching performance. Both SURF and SIFT feature vectors are compared with the L2-Norm measure, a floating-point arithmetic operation.

The use of gradient magnitude as a significance measure is seen in the majority of detector and descriptor algorithms. While its use in SIFT to weight votes is explicit, the pair-wise intensity comparisons of the binary descriptors are essentially encoding the sign (i.e. increasing or decreasing intensity) of the gradient. The assumption that gradient direction remains constant is fundamental to describing elements in visible spectrum images; however, the non-linear relationship between visible and infrared modalities breaks this assumption. This chapter presents an investigation into the performance of conventional feature-based methods in a multi-spectral context and proposes an alternative means of describing image elements that is invariant to the variations between the spectra.

Multi-spectral image features

Traditional approaches to sparse feature correspondence operate under assumptions that hold within the visible spectrum. This section presents a brief background, the purpose of which is to determine where traditional methods fail, identify the practical limits of their application and establish a context to the investigation carried out in this chapter. A survey of the performance of conventional feature-based methods reported in literature is presented .Observations from this survey are used to formulate the objectives of the investigation, and in the development of a multi-spectral descriptor.

Problem statement

There is a clear consenus that traditional feature correspondence methods perform poorly or fail entirely to detect and match thermal and visible features , although there has been limited investigation focused on analysing this behaviour beyond the visible spectrum.

In order to identify where traditional methods fail, start with the first assumption that these methods are only suitable for use with visible spectrum images. To dispute this, Ricaurte provides an extensive investigation of the intra-band behaviour of SIFT, SURF, BRISK, Binary Robust Independent Elementary Feature (BRIEF) , ORB and Fast Retina Keypoint (FREAK) algorithms on thermal image pairs compared to their visible spectrum performance of the same scene. While noticeably weaker than operation in the visible spectrum, the traditional methods (SIFT, in particular) performed well enough in the thermal spectrum to warrant practical use. This intra-spectrum performance is used in applications presented in the literature .

The second logical assumption is that traditional methods will only perform well within the same spectral range. This is partially true; minor adaptations on the traditional SIFT descriptor have been required to enable VS-NIR feature matching . However, as the difference of wavelength between the spectral modalities increases, the images become increasingly dissimilar and traditional feature-based correspondence methods fail entirely . This failure is attributed to the low spatial resolution, noise and the lack of common descriptive elements (i.e. colour or texture) in thermal images, as well as the increasingly disparate information captured by the distinct sensors.

The result of this dissimilarity is that feature detectors do not locate the same observations and feature descriptors that rely on gradient direction fail due to the non-linear relationship between the intensity values of each spectral modality . Contrast reversals are common in the distinct observations captured by these sensors; for example, the edges of a dark, warm object on a light, cool background would appear highly dissimilar to SIFT due to the gradient directions being 180 degrees out of phase (i.e. contrast is reversed). Numerous adaptations of the SIFT algorithm have been made to compensate for and correct contrast reversal (e.g. Gradient Direction Invariant SIFT (GDI-SIFT) , Oriented SIFT (OR-SIFT) Uniform Robust SIFT (UR-

ISSN:0971-1260

Vol-22-Issue-22-December-2019

SIFT) however, these methods are designed for matching visible with near infrared images and not suited to matching across the large spectral gap between visible and thermal information.

Hypothesis and objectives

The goal of the remaining sections of this chapter is to present an approach to multi-spectral feature detection and description that is better suited than traditional methods. This goal is based on two hypotheses: first, that traditional approaches are fundamentally unsuited to matching multi-spectral features, and second, that the detection and description of structural features (such as edges and corners) provide a more repeatable and matchable similarity measure than gradient-based methods.

The specific objectives are therefore to develop and present a set of methods that detect and describe structural features such that they can be matched and to demonstrate the effectiveness of the approach by comparing matching performance to traditional methods.

Development of theory

Object and material boundaries manifest as edges and ridges and are often described in terms of their orientation in the image. These structural features are commonly detected at pixels with a high gradient magnitude and described in terms of the direction of the gradient, perpendicular to the edge. The method proposed in this section has close parallels to the work of Park et al. and was closely influenced by the work of Aguilera and, in particular, Mouats and Aouf .

Conventional methods rely on gradient magnitude to indicate significance in detection and gradient orientation as a means of description; however, the disparate information captured by thermal and visible sensors means that these assumptions, based on intensity information, do not hold in most cases. Methods that utilize frequency-domain analysis for detection and description are proposed in this work as a means for repeatable detection and inter-spectrum feature description. Phase congruency, introduced in this paper is a recurring focus of this work. It is used here as robust corner and edge detector, invariant to image contrast, to replace the gradient magnitude approach of traditional methods.

A variation of the Edge Orientation Histogram (EOH) descriptor is developed in this dissertation. The EOH descriptor, as presented in predominantly used in the MPEG-7 standard is noted for its efficiency and efficacy in image content search and retrieval, and it is often applied as a global or semi-global descriptor in order to retrieve similar images from a database .. This class of edge orientation methods influenced the Histogram of Oriented Gradients (HOG) descriptor which has been particularly effective in hand gesture recognition, a task requiring invariance to photometric qualities such as illumination, skin colour and background colour when comparing an observed gesture to a database.

The EOH feature vector is constructed with an edge orientation voting process in which the orientation of each edge pixel is used to construct a histogram of orientation votes that describe the local region. The square N N pixel descriptor footprint is divided into 16 sub-regions; each sub-region holds a voting process to construct a histogram of the four directional 0, 45, 90, 135 degree bins and the no orientation bin. The histograms from each sub-region are separately normalised and concatenated to form the 80 element EOH feature vector (i.e. 16 sub-regions each contributing a 5 bin orientation histogram). An illustration of this process is shown in Fig.

ISSN:0971-1260

Vol-22-Issue-22-December-2019

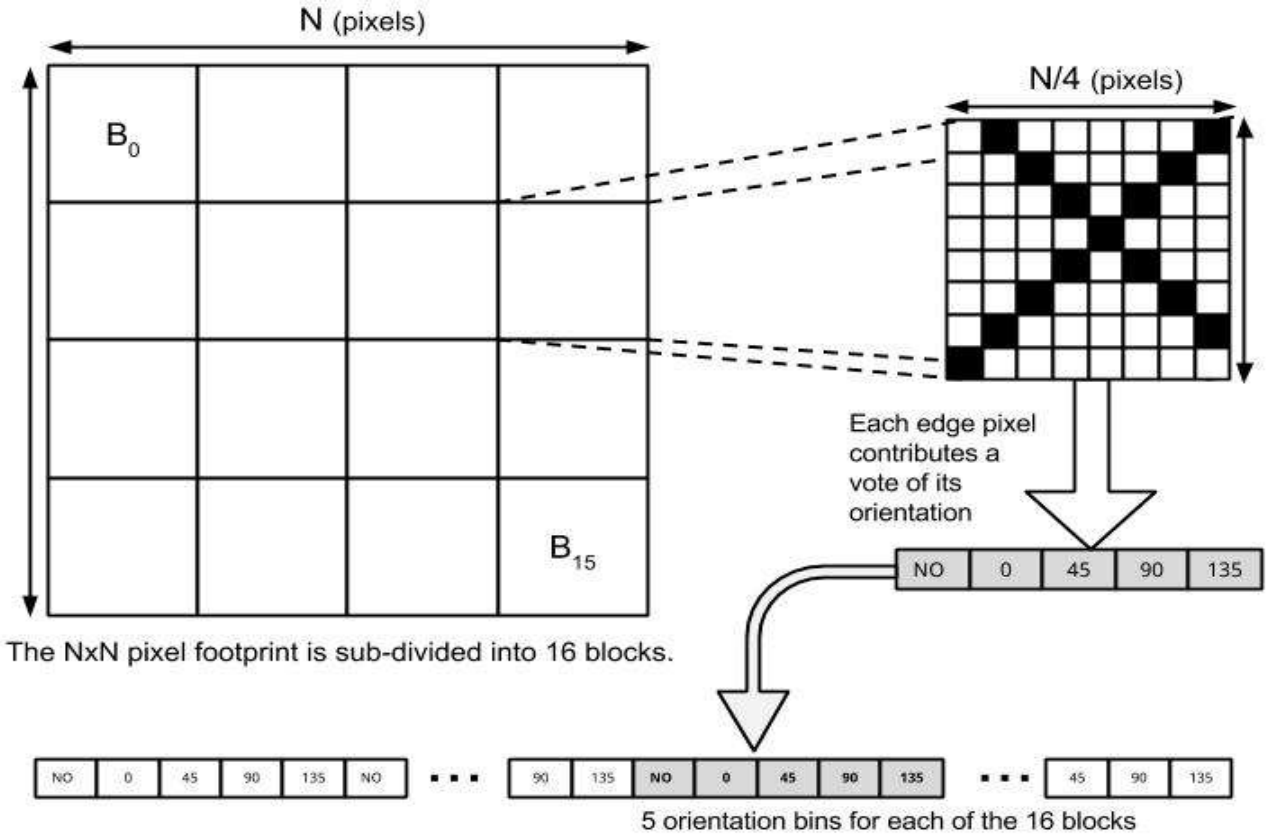


Figure: The edge orientation histogram (EOH) is an N N pixel square patch divided into 4 4 blocks, each characterised by five bins containing the votes for the gradient direction of its pixels .

Edge orientation voting is a task common to many descriptor algorithms; the SIFT descriptor, for example, uses 36 orientation bins. The vote of each edge pixel of the EOH descriptor is based on which of the five 3 spatial filters is maximised. Figure illustrates two commonly used spatial filter configurations. The numerical values in each cell represent coefficients assigned to pixel values in the convolution operation with the underlying edge map.

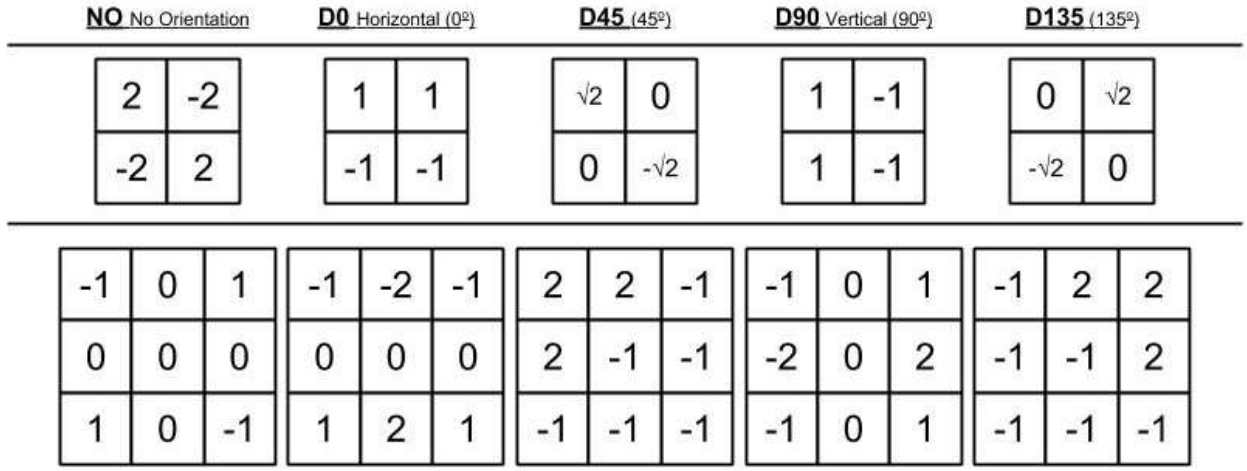


Figure: The directional filters shown above are applied to each pixel in the EOH. Edges are classified into four discretised directions, with an additional \no orientation" filter. The first row shows the 2 2 filters used by Park , the second row shows those used by Aguilera.

Approach to evaluation

ISSN:0971-1260

Vol-22-Issue-22-December-2019

This brief section presents the evaluation of conventional feature-based methods and the EOH descriptor method developed in the previous section. The approach to evaluating the different stages of feature-based methods is described and the tools (i.e. algorithms, software libraries and inputs) used are recorded .

Quantifying performance

In order to achieve the objectives set out for this investigation, feature correspondence methods are split into detection and matching stages and analysed separately where possible.Interest point detectors are analysed based on detection repeatability: the ability of a detector to identify and localise the same observations in the two viewpoints. Detecting the same points in both viewpoints is a crucial first step as it generates the pool of matchable points. Detector repeatability is evaluated to determine how many of the interest points detected in one modality are also found in the other; it is represented as a fraction of the simultaneously observed interest points over the total number of feature points detected. In order to avoid the use of a heuristic threshold on each detector's interest measure, a constant number of the strongest interest points across all detectors is taken. While this constant is also heuristically chosen, it standardises the way in which the detectors are evaluated.

Interest point matching demonstrates a descriptor's effectiveness in encoding the information at interest points such that it can be correctly matched. Evaluation of descriptors is a little more complicated in that a number of steps must be taken to standardise the matching process. Once two sets of feature vectors have been extracted from the images, a simple brute force matching algorithm is used: each feature vector of one set is compared to every feature vector of the other set and associated (matched) with the vector that optimises the similarity metric. In this work, the declared matches of the descriptors are limited to the matches that have a distance less than twice the best (smallest distance) match. Descriptor performance is evaluated based on the number of true positives (correctly identified as a match) and false positives (incorrectly identified as a match) within the declared set of matches for each descriptor.

Tools and parameters

In order to achieve the objectives set out for this investigation, a representative sample of detectors and descriptors was chosen. The stand-alone detectors evaluated are the Shi-Tomasi and phase congruency corner detectors, as well as the detectors of the SIFT, SURF and FAST methods. The conventional SIFT, SURF and FAST algorithms were then analysed (with their respective detection methods) alongside the EOH descriptor (with the Shi-Tomasi corner detector) using a brute force matching algorithm.

The algorithms (with the exception of the EOH descriptor) were implemented using the OpenCV library and the Python programming language. The parameters suggested by the respective authors of each method were used. The EOH descriptor uses the phase congruency tools, described and the numpy library for implementing efficient linear filtering and edge orientation voting. Automated analysis was done using two image pairs from the OSU (OCTBVS) Color-Thermal Dataset with a 2 pixel ground truth. Although the corresponding points were removed from the two sets once matched, clustering of interest points resulted in the detection results not always being symmetric across the spectral modalities. A maximum of 200 interest points was taken from each image in the analysis of detector repeatability, although this was increased to 300 when descriptor performance was analysed in an e ort to increase the number of potential matchable points. This maximum is heuristically chosen, but it helps to standardise comparison of the detection and matching algorithms.

The proposed EOH descriptor was then investigated under a maximum feature displacement constraint of 40 pixels coupled with outlier rejection with Random Sample And Consensus (RANSAC) [80] (OpenCV implementation). These constraints are commonly used in practice [43, 100] and are used here to better characterise the practical performance of the proposed method. In order to demonstrate its performance on a different dataset, a sample from the CVC Multimodal Stereo Dataset (2) was manually evaluated and included.

RESULTS

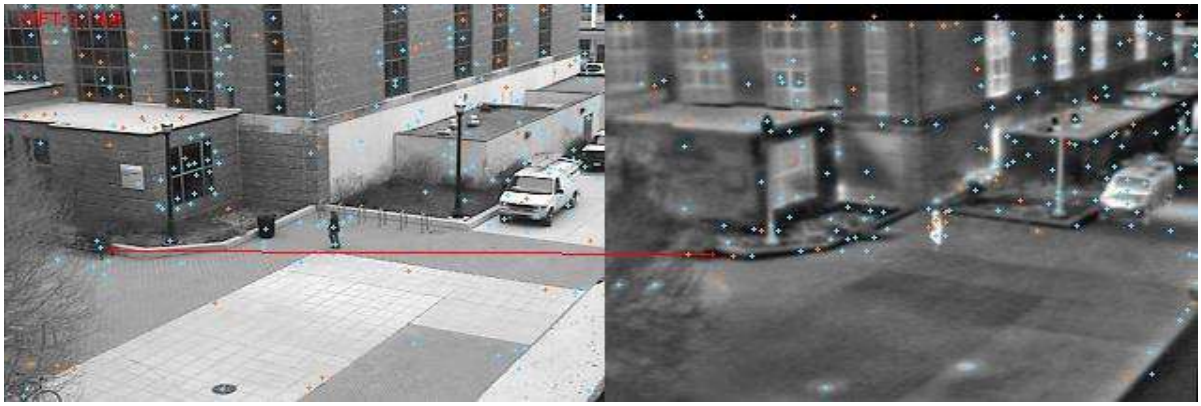
ISSN:0971-1260

Vol-22-Issue-22-December-2019

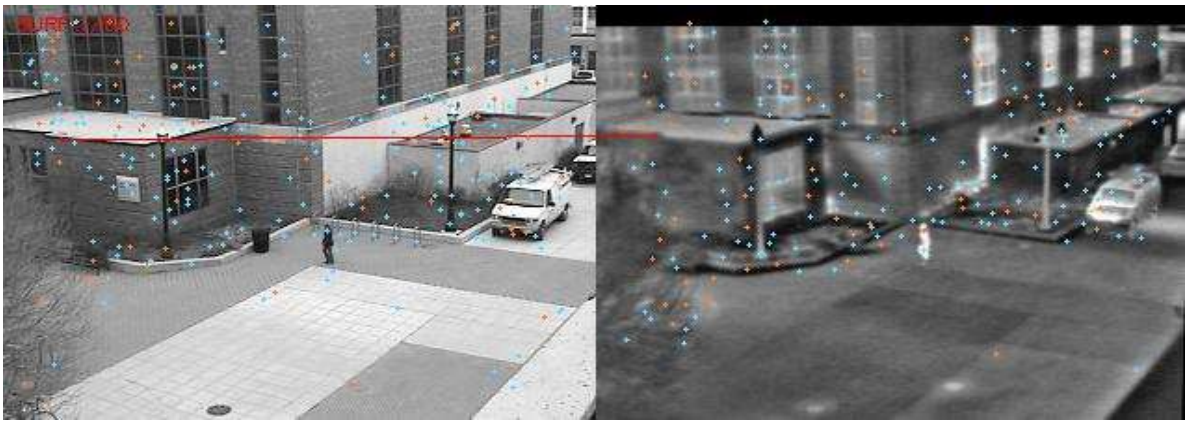
The Shi-Tomasi GF corner detector performed the best on both sample pairs, providing 35% repeatability from visible to infrared. Despite showing similar performance in one case, the phase congruency detector suffered due to a large percentage of detected interest points clustering along different edges in the two spectral modalities. Due to the 3 pixel non-maximal suppression of these methods, the results were symmetric. The FAST detector demonstrated repeatability of thermal interest points in the visible image (i.e. a high percentage of interest points detected in the thermal image were found in the visible image), but showed very poor repeatability of visual interest points in the thermal image. The SIFT and SURF interest point detection methods performed very poorly; less than 10% of interest points were common to both viewpoints.

The conventional SIFT, SURF and BRISK detection and description algorithms were matched using the methods described in the previous section. SIFT matched 4.3% (4 correct, 94 declared) as did SURF (3 correct, 69 declared). BRISK performed the wofirst at around 1.2% (3 correct, 249 declared). Under the same conditions, the EOH-GF algorithm matched 18.1% (39 correct, 215 declared). Figure illustrates one such case to emphasise the performance of the EOH-GF algorithm.

The EOH-GF method was analysed with a maximum displacement constraint (40 pixels) and RANSAC outlier rejection. The combined results showed that 25% of the interest points (115 of 460 detected) were common to both modalities. Brute force matching resulted in 34% (84 of 247 declared) of features being correctly matched, which increased to 58% (51 of 88 declared) after RANSAC was used to reject outliers. Figure 4 demonstrates the performance of the EOH-GF algorithm (under the constraints discussed) on a higher resolution thermal/visible image pair.The image pair is rectified although no ground truth exists; therefore, correspondences were manually analysed and circled in green (correct) or red (incorrect).



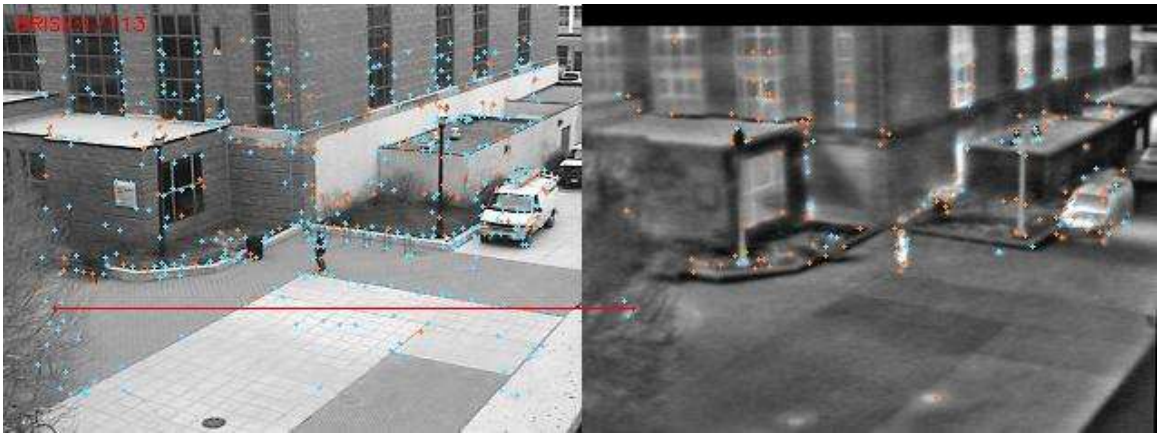
(a) SIFT (1 correct of 46 declared).



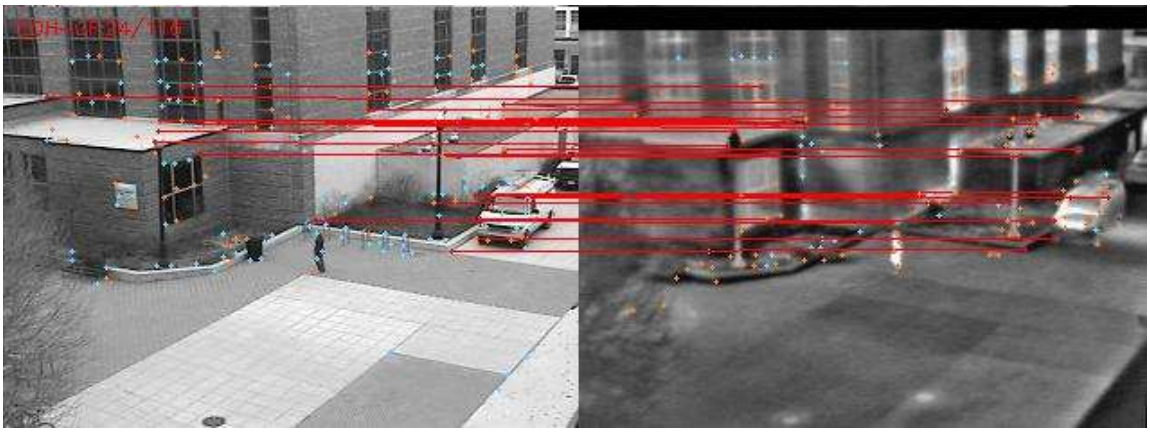
(b) SURF (1 correct of 60 declared).

ISSN:0971-1260

Vol-22-Issue-22-December-2019



c) BRISK (1 correct of 113 declared).



(d) EOH-GF (24 correct of 116 declared).

Figure: The inter-spectrum (visible to thermal) matching performance of the SIFT, SURF, BRISK and EOH-GF methods are illustrated in four image pairs. Correct matches are joined by red lines, while false positives and detected (but unmatched) interest points are shown as orange and blue crosses respectively. (Images from the OSU database .)

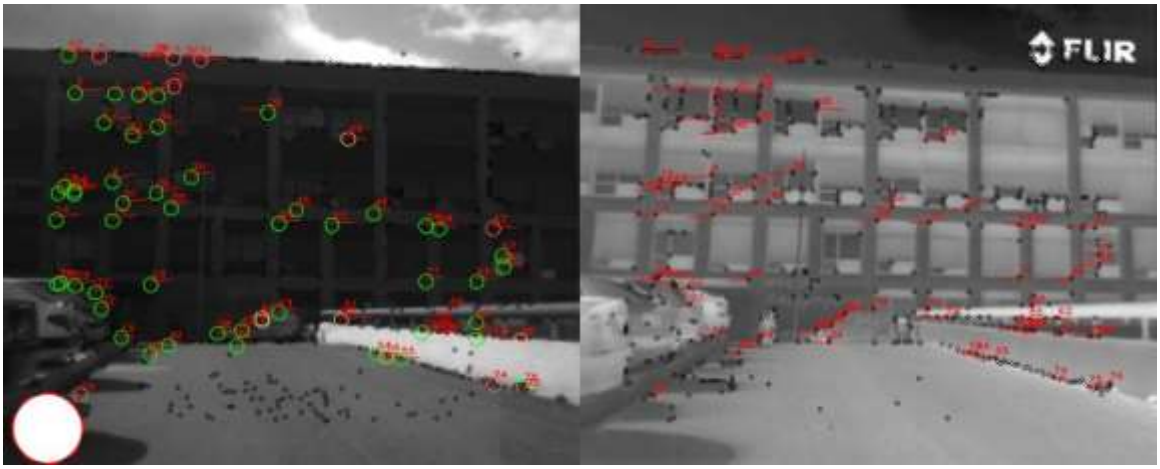


Figure : Of the RANSAC-declared EOH-GF 78 matches, 53 are correct (within 3 pixels) and 10 are incorrect. There are 4 matches of adjacent features and the remaining 7 are di cult to discern. The white circle on the bottom left hand side of the image indicates the extent of the disparity constraint for each match. (Image from the CVC dataset.

BIBLIOGRAPHY

ISSN:0971-1260

Vol-22-Issue-22-December-2019

[1] J. B. Campo, \Multimodal Stereo from Thermal Infrared and Visible Spectrum," PhD thesis, Universitat Autonoma de Barcelona, Spain, 2012

[2] M. Govender, K. Chetty, and H. Bulcock, \A review of hyperspectral remote sensing and its application in vegetation and water resource studies," Water SA, vol. 33, no. 2, 2009.

[3] H. Loh and M. Lu, \Printed circuit board inspection using image analysis,"

[4] IEEE Transactions on Industry Applications, vol. 35, no. 2, pp. 426{432, 1999.

[5] A. Wong and W. Bishop, \E cient least squares fusion of MRI and CT images using a phase congruency model," Pattern Recognition Letters, vol. 29, pp. 173{ 180, Feb. 2008.

[6] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, \Mul-timodality image registration by maximization of mutual information," IEEE transactions on medical imaging, vol. 16, pp. 187{98, Apr. 1997.

[7] J. P. W. Pluim, A. Maintz, and M. A. Viergever, \Mutual-information-based reg-istration of medical images: A survey," IEEE Transactions on Medical Imaging, vol. 22, no. 8, pp. 986{1004, 2003.

[8] Z. Liu, D. S. Forsyth, and R. Laganiere, \A feature-based metric for the quanti-tative evaluation of pixel-level image fusion," Computer Vision and Image Understanding, vol. 109, pp. 56{68, Jan. 2008.

[9] 8, L. G. Brown, A survey of image registration techniques, PhD thesis, Columbia University, 1991.

[10] 9,D. L. Hall and J. Llinas, \An introduction to multisensor data fusion," Proceed-ings of the IEEE, vol. 85, no. 1, pp. 6{23, 1997.

[11] R. Luo and M. Kay, \A tutorial on multisensor integration and fusion," in 16th Annual Conference of IEEE Industrial Electronics Society (IECON), 1990.

[12] A. Rogalski, \Infrared detectors: Status and trends," Progress in Quantum Elec-tronics, vol. 27, no. 2-3, 2003.

[13] A. Rogalski, \History of infrared detectors," Opto-Electronics Review, vol. 20, no. 3, pp. 279{308, 2012.

[14] M. Iqbal, An introduction to solar radiation. Elsevier, 1983.

[15] P. Ricaurte, C. Chilan, C. A. Aguilera-Carrasco, B. X. Vintimilla, and A. D. Sappa, \Feature point descriptors: Infrared and Visible Spectra," Sensors 2014 (Switzerland), vol. 14, pp. 3690{3701, 2014.

[16] L. Schaul, C. Fredembach, and S. S•usstrunk, \Color image dehazing using the near-infrared," in 2009 16th IEEE International Conference on Image Processing (ICIP), pp. 1629{1632, Citeseer, 2009.

[17] Y. M. Lu, C. Fredembach, M. Vetterli, and S. Susstrunk, \Designing color l-ter arrays for the joint capture of visible and near-infrared images," in 2009 16th IEEE International Conference on Image Processing (ICIP), pp. 3797{3800, 2009.

[18] R. Fattal, \Single image dehazing," ACM Transactions on Graphics, vol. 27, no. 3, p. 1, 2008.