A Novel Graph-based Invariant Region Descriptor for Image Matching

Francis Bowen[a] *Student Member IEEE*, Eliza Y. Du[a], *Senior Member, IEEE*, and Jianghai Hu[b], *Member, IEEE*
[a]*Dept. of Electrical and Computer Engineering, Indiana University-Purdue University Indianapolis, 723 W. Michigan St. SL160, Indianapolis, IN 46202;*
[b]*Dept. of Electrical and Computer Engineering, Purdue University, 465 Northwestern Ave., West Lafayette, Indiana 47907*

***Abstract*—Identification of invariant image descriptors is an integral task for many computer vision applications such as image registration, object recognition, and object tracking. The detected features should be invariant to geometric transformations such as rotation and translation, as well photometric variations due to differing lighting conditions. In this work, we propose a unique and effective region descriptor that couples invariant features and texture information. The descriptor relies on spatial relationships of invariant SURF features to create a graph-based descriptor for image matching. Additionally, a novel method is proposed for matching region descriptors through the definition of an efficient similarity measure that couples invariant features and their spatial relationships. Several examples are presented to illustrate the effectiveness of the proposed region descriptor while the results of the proposed approach outperform SURF feature point matching.**

***Index Terms*—Feature extraction, image matching, region descriptors, Speeded-up Robust Features (SURF).**

I. Introduction

At the core of reliable image matching is the accurate identification of features within an image. The detected features should be invariant to geometric and photometric effects. Furthermore, the detected features should be unique enough so that ideally there will exist a strict one-to-one mapping across images. Extracting invariant features or regions can also aid in computer vision applications that require image registration, segmentation, or object recognition.

Typical approaches that rely on invariant features can be thought of in three phases. Initially, invariant features are detected, followed by invariant feature description, and then feature matching is performed between two images. Common keypoint detection algorithms include SIFT [1], which uses a scale-space from a pyramid of difference of Gaussian images, and the Harris Corner Detector [2] that attempts to identify image corners through the analysis of eigenvalues for the first order gradients of local image neighborhoods. The original Harris Corner Detector algorithm is not scale invariant and thus several variations have been proposed [3-4], however these proposed methods do not address the issue where one invariant feature of the reference image maps to several features in the query image.

Due to the scale, rotation, and translation invariance of the Scale Invariant Feature Transform (SIFT) descriptor, much work has been based on the SIFT detector and feature point descriptor, however the complexity of the algorithm limits its applications. In [5], Bay proposes an alternative to SIFT called Speeded-Up Robust Features (SURF) that is also scale, rotation and translation invariant. SURF is advantageous over SIFT in several areas. The algorithm utilizes integral images and discrete box filters for fast, integer-based computation of the scale-space. Moreover, the SURF descriptor is based on an integer approximation of the Haar wavelet responses within a local neighborhood of each feature point. Due to these enhancements and promising results, SURF has quickly become the basis for many computer vision applications [6-8]. Although SURF greatly reduces the complexity of SIFT, the discriminative power of the SURF descriptor does not always ensure a one-to-one mapping between invariant feature points from two images. One such scenario where this issue may arise is that of buildings whose structure includes many repetitive characteristics, such as windows. It is the aim of this work to use SURF detected feature points to create a more discriminative descriptor that will have a one-to-one mapping across images. In doing so, the overall solution is computationally efficient and may be a suitable candidate for real-time applications or easily mapped to hardware using reconfigurable logic.

In this work, a novel region descriptor is proposed that is invariant to both photometric and geometric variations. The computational efficiency of the SURF algorithm is leveraged to create a region descriptor that couples spatial, textural, and color information. In doing so, it is shown that the discriminative power of the proposed descriptor is superior to the SURF feature points alone. The overall structure of the region descriptor forms a connected graph which provides an efficient structure for rapid matching. Additionally, a unique similarity measure is proposed for comparing two region descriptors.

The rest of the paper is organized as follows. Section II reviews previous work pertinent to the proposed method, while Section III outlines the region descriptor algorithm and matching process. Section IV provides results and discussion

978-1-4673-0818-2/12/$31.00 ©2012 IEEE

of the proposed descriptor and Section V gives the final conclusions.

II. Previous Work

A. Speeded-Up Robust Features

In [5], Bay proposed a novel alternative to SIFT called Speedup Robust Features (SURF) which aimed to compute multi-scale feature points that are invariant to scale, rotation and translational deformations. The algorithm relies on integral images for fast computation of discrete image integrals, which is defined in (1).

$$I_{\Sigma}(x,y) = \sum_{i=0}^{x} \sum_{j=0}^{y} I(x,y) \qquad (1)$$

By utilizing the integral image, the area within a bounded region (A,B,C,D) of the original image can be computed using four memory accesses and three additions.

$$\Sigma = I_{\Sigma}(A) - I_{\Sigma}(B) - I_{\Sigma}(C) + I_{\Sigma}(D) \qquad (2)$$

Interest points are determined by calculating the determinant of the Hessian matrix for a particular pixel location (x,y).

$$\det\left(\mathrm{H} = \begin{bmatrix} L_{xx}(x,y,\sigma) & L_{xy}(x,y,\sigma) \\ L_{xy}(x,y,\sigma) & L_{yy}(x,y,\sigma) \end{bmatrix}\right) = L_{xx}(x,y,\sigma)L_{yy}(x,y,\sigma) - 0.81\left(L_{xy}(x,y,\sigma)\right)^2 \qquad (3)$$

Where $L_{xx}(x,y,\sigma) = \frac{\partial^2}{\partial x^2} g(\sigma) * I(x,y)$ for a given scale, σ. The Gaussians are estimated using discrete box filters. The following images illustrate the discrete box filters for L_{yy} and L_{xy}.

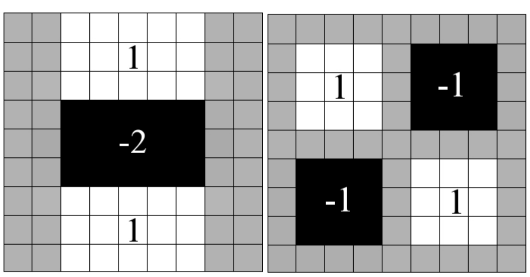


Figure 1 Box filters for L_{yy} and L_{xy}

To create the scale-space representation, the discrete box filters are increased in size while the image maintains a constant size, which is in contrast to the typical Gaussian pyramid scheme where the Gaussian kernel is kept constant and the image size is scaled. With SURF, the computational time for any filter size remains constant, thus increasing the efficiency over SIFT. From the scale-space, a search is performed in a 3x3x3 neighborhood while non-maximal suppression identifies candidate feature points.

The SURF feature descriptor is created in a similar vein as the scale-space. Haar wavelet responses are calculated using the integral image and approximated integer filters. For each detected keypoint, a local neighborhood centered at the point, of size 20σ, where σ is the detected scale, is used to estimate the Haar wavelet responses in the x and y directions. The responses are computed in equally sized 4x4 subregions of the local neighborhood. Each subregion produces a vector $h_i \in \mathbb{R}^{4x1}$ that contains components Σd_x, Σd_y, $\Sigma|d_x|$, and $\Sigma|d_y|$. The descriptors for all subregions are concatenated to form the 64-dimensionsed descriptor. Σd_xrepresents the Haar response in the x-direction and $\Sigma d_y$is the response in the y-direction.

$$d = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_{16} \end{bmatrix}, \qquad d \in \Re^{64} \qquad (4)$$

Several methods have been proposed for matching SURF descriptors. For object recognition and matching, Bay [6] describes a Bayes classifier coupled with a bag-of-words approach for comparing descriptors, however such an approach requires training. A simpler approach simply uses the Euclidian distance as a similarity metric for two descriptors, d_k and d_j. Matching is accomplished through an exhaustive comparison between every feature point and taking the best matched pair with minimized distance. Figure 2 illustrates the ineffectiveness of feature point matching by solely relying on the Euclidean distance.

$$S_E = \sqrt{\sum_{i=1}^{64} \left(d_k[i] - d_j[i]\right)^2} \qquad (5)$$

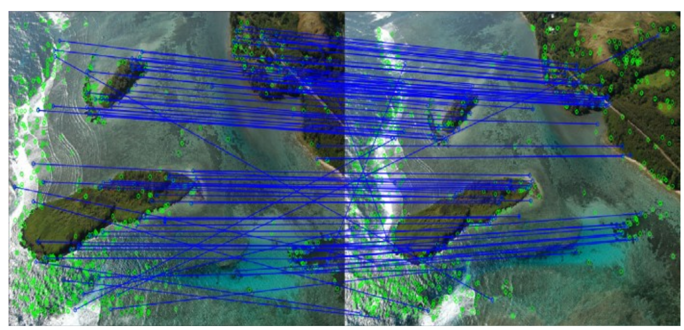


Figure 2. Matching SURF feature points using the Euclidian distance. Using point to point matching alone may create several mismatched pairs.

As depicted in Figure 2, two issues exist with the SURF point matching. First, several mismatches are present due to the similar texture of the surrounding water. Second, there exist numerous SURF feature points which are matched to the same point in the query image. Due to the inaccuracies of the feature point matching, the proposed descriptor will couple textural and spatial information to provide a more discriminative characteristic to the image regions.

Several improved similarity metrics have been proposed in recent literature. Fan *et al.* [9] outlines a method in which the discriminative power of the SURF descriptor is improved with the inclusion of color information. The authors employed the

Bhattacharyya distance for comparing the descriptor's color histograms and created a weighted descriptor from the original SURF descriptor and color information. In [10], the Normalized Cross Correlation similarity measure is presented, however results are not favorable in situations of high variations in illumination.

B. Graph-based Descriptors

Graph-based analysis has recently become a popular method for determining global features from invariant local characteristics. The overall discriminative power of invariant feature points is potentially increased if the graph is also invariant to geometric or photometric variations. Simacek and Unsalan proposed a connected graph from SIFT points in [11] to detect buildings in an urban scene, while Kisku *et al* used a similar approach for a facial recognition system [12]. In both instances the authors show an increased matching accuracy when comparing subgraphs from two images than using the feature points alone. Neither approach utilizes color information while both methods rely on the complex SIFT descriptor which may limit the scope in which the algorithms may be used.

III. Proposed Method

Image matching requires the identification of common features or regions between images. If enough similarity exists, the scene or objects are matched. The detected features of each image should be distinct and the method should be invariant to photometric and geometric inconsistencies. The proposed algorithm creates a descriptor from the spatial and color relationships of invariant descriptors. Because of object deformations and occlusions, the matching process must calculate similarities between regions to determine all commonalities between images.

To mitigate these issues, it is the aim of this paper to propose a region descriptor that is formed from a connected graph that contains SURF feature points as its vertices. Figure 3 illustrates an overview of the proposed method.

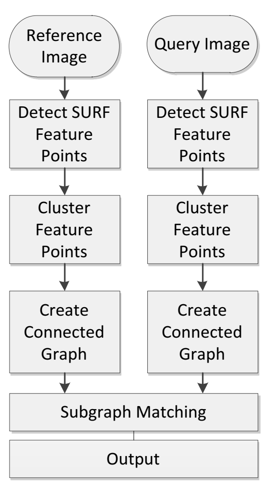


Figure 3. Region descriptor overview.

For a given reference and query image, the SURF feature points are computed then a clustering process is used to identify similar feature points. The clustering algorithm ensures all feature points within a group share common characteristics and they are invariant to rotation, translation and affine transformations. Each cluster will also share common color properties as determined by the area's color distribution.

A connected graph is then constructed from each feature point within a cluster. To accomplish this, a proposed similarity measure is used to exploit the spatial and textural characteristics of each feature point while Dijkstra's algorithm is employed to create a connected graph of feature points.

Matching two images involves matching region descriptors across images. Since each descriptor is a connected graph, subgraph matching schemes are explored. Often exact matches are not possible due to deformations, occlusions or variations in color; therefore the problem of region matching must be accomplished using a proposed similarity measure. This measure compares two graphs on the basis of the individual nodes and their spatial relationships in feature-space.

A. Feature Point Clustering

The initial feature point detection is accomplished using the SURF detector. Two descriptors are compared by calculating their Euclidean distance in feature space according to (5). Let S_c be the similarity of two feature points in RGB color space defined as,

$$S_c = \sqrt{(I_R^1 - I_R^2)^2 + (I_G^1 - I_G^2)^2 + (I_B^1 - I_B^2)^2} \tag{6}$$

where I_c^x is the color channel for feature point x, with $c \in \{R, G, B\}$. Coupling color and descriptor similarity, a potential cluster of similar descriptors is expressed as,

$$S(i,j) = \{S(i,j) \mid S_E(i,j) \le T_E, S_c(i,j) \le T_C\} \tag{7}$$

Let Ψ^{nxn} be a search window centered at a feature point. The set of descriptors within the search window is the set C, defined as,

$$C = \{d_i \mid d_i \in \Psi^{nxn}\} \tag{8}$$

The descriptors within a search window are clustered if 75% of the descriptors are similar to all other descriptors within the window. We define the set of clustered descriptors as

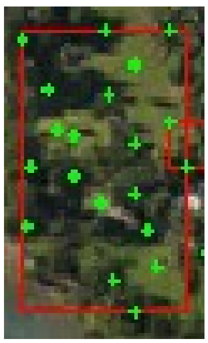
$$B_i = \left\{ d \in C \mid \frac{|S(i,j)|}{|C|} \ge 0.75 \right\} \tag{9}$$

If we define the set of descriptors in cluster n as $d^n = \{d \mid d \in B_n\}$ then two clusters are merged if the number of common feature points between clusters is at least 2. The resulting set is defined in (10).

$$B_i^* = \left\{ B_i \cup B_j \mid \left(\left| d^i \in B_j \right| \ge 2 \right) \right\} \tag{10}$$

Initial clustering results are provided in Figure 4 where SURF points are marked in green and the clusters are outlined in red.

a.



b.

Figure 4. Detected feature point clusters (a), and a close view (b) of the single cluster indicated by the blue outline in (a).

B. Connected Graph Region Descriptor

If we assume a cluster of feature points represent the vertices of a graph, we can formulate an algorithm for the creation of a region descriptor from a connected graph. Dijkstra's algorithm is used to determine the shortest path from one point to any other point in a connected graph given a list of weighted edges [13]. Let Q_G be defined as the queue representing our final connected graph. The first node in the queue will be regarded as the origin of the graph. Initially the queue is empty; $Q_G = \{\emptyset\}$.

- We define C_F as the set containing the cluster of feature points.
- Select the feature point closest to image's origin, FP_0 and push into the queue, $Q_G = \{FP_0\}$ and remove FP_0 from the set of feature points; $C_F = C_F \setminus FP_0$
- Let FP_i represent a neighbor of FP_0. Assign a distance $\|FP_0 - FP_i\|$ to FP_i. Repeat for all neighbors (ie every remaining feature point in the cluster).
- Select FP_j such that $\min_{FP_j}\|FP_0 - FP_j\|$.
- Push FP_j onto the queue; $Q_G = Q_G \cup FP_j$.
- Remove FP_j from the feature point set; $C_F = C_F \setminus FP_j$.
- For the node at the tail of the queue, repeat from step 4 until $C_F = \{\emptyset\}$.

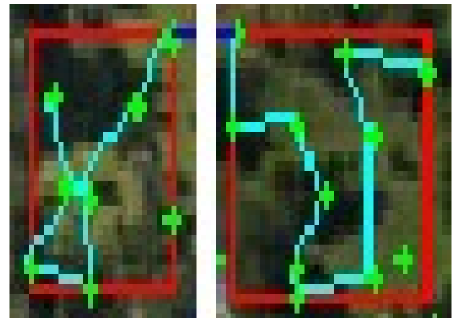


Figure 5. Two example connected graphs from invariant feature points. The green markers indicate the feature points while the blue lines are the connected arcs of the graph.

C. Descriptor Matching

Region matching can be reduced to the problem of determining whether or not G_1 is a sub-graph of G_2, with $|V_2| > |V_1|, G_1 = (V_1, E_1), G_2 = (V_2, E_2)$. For simplicity, we are not searching for exact matches but rather similar sub-graphs where the similarity is a calculation based on the feature point's descriptor and spatial information of the graph nodes. The graph from B_i^* is compared to the graph which spans B_j^*, and the lowest similarity score from the comparison is stored in a scoring matrix, $\mathcal{M}^{|B_i^*|\times|B_j^*|}$, where B_i^* represents a cluster of feature points.

The nodes have an associated descriptor, $d \in \mathbb{R}^{64}$ and we define the similarity of two nodes, V_i and V_j, to be $S_{FP} = \|d_i - d_j\|$. In a similar fashion, we can define the spatial similarity between two edges, E_i and E_j as $S_D = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$, where $V_i = (x_i, y_i)$ and $V_j = (x_j, y_j)$. The similarity between two graphs can then be computed as,

$$S = \frac{W_1 \sum_{|G_i|} S_{FP} + W_2 \sum_{|G_i|} S_D}{W_1 + W_2} \tag{11}$$

where G_i is the smaller of the two graphs being compared, $W_1$is the weight of the feature point descriptors while W_2 is the weight associated with the spatial distance between nodes. For this version, W_1 is set to 0.8 and W_2 is 0.4.

Given two graphs, G_i and G_j such that $|V_i| = n$, $|V_j| = m$, and $n > m$, we first compute a similarity score between $G_j$and a sub-graph of G_i, denoted as G_i' where G_i' contains the first m vertices of G_i. In successive iterations, we extract sub-graphs of size m, starting from the next node in the queue of G_i until we have $(n - m + 1)$ similarity scores. The smallest of these scores are chosen as the similarity score for $\mathcal{M}_{i,j}$. Once the exhaustive search is complete, $\mathcal{M}$ is traversed to find matches which are indicated by minimal scores.

If we are comparing two graphs, G_1 and G_2, where $n = |V_1|$, $m = |V_2|$, and $n \geq m$, we must make $n - m + 1$ comparisons to test G_2 against all sub-graphs of G_1. In the following figure, the vertices of G_1 are given as the set $\{V_1, .., V_5\}$ and the vertices of G_2 as $\{V_6, .., V_8\}$.

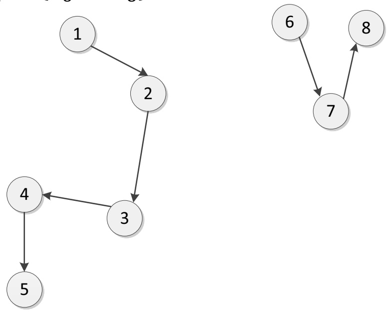


Figure 6. Partial graph matching

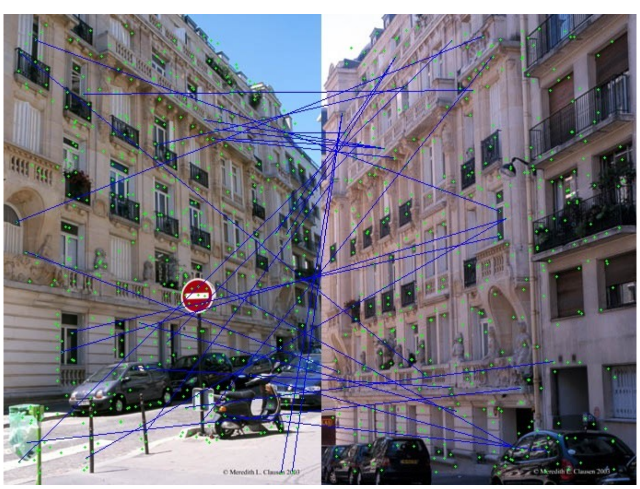
For the illustrated example, a matching score is generated for the comparison of sets $\{V_1, V_2, V_3\}$ and $\{V_6, V_7, V_8\}$, $\{V_2, V_3, V_4\}$ and $\{V_6, V_7, V_8\}$, and finally between $\{V_3, V_4, V_5\}$ and $\{V_6, V_7, V_8\}$. The minimal score represents the similarity between G_1 and G_2.

The complexity of matching G_1 and G_2 is determined as follows. Let $m = |G_1|$ and $n = |G_2|$, then the total number of sub-graph comparisons is given as $m - n + 1$, where $m \geq n$. The complexity of comparing region descriptors is then derived as follows.

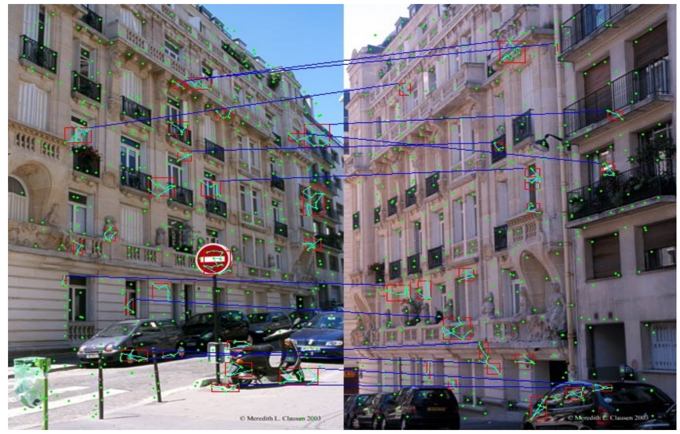
$$\begin{aligned} &O\big(n * (m - n + 1)\big) \\ &\leq O(nm) - O(n^2) + O(1) \\ &\leq O(nm) - O(n^2) \\ &\leq O(nm) \end{aligned}$$

IV. Results and Discussion

The proposed descriptor was tested on a dataset consisting of scenes that are composed of buildings and other man-made structures, UAV-acquired images as well as general objects from the Amsterdam Library of Object Images [14]. Figure 7 provides an example of the detected region descriptors for the same dataset shown in Figure 2. In this scenario, it is observed that the proposed region descriptor provides a more accurate one-to-one mapping of descriptors. This is evidenced by the vectors connecting the matched descriptors across images. Most matched vectors have the same orientation which implies a more accurate matching.



a. SURF feature points matched using Euclidean distance.



b. Proposed region descriptor matching using the partial graph similarity measure.

Figure 7. SURF descriptor matching versus the proposed region descriptor.

Figure 8. Descriptor matching before and after a disaster.

As a special case, several images from the dataset were chosen to represent disaster scene analysis. In this scenario large photometric variation prevented the SURF detector from accurately identifying matched points. Figure 8 provides evidence that the region descriptor is more accurate. Several outlier matches are shown, but an algorithm such as RANSAC may be employed to identify a suitable matched set. As illustrated, the region descriptor was effective in identifying the same features under large changes in texture and structural properties.

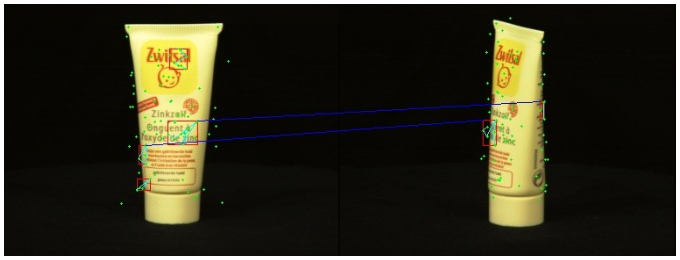


Figure 9. Descriptor matching on general objects.

The case involving general objects under varying rotations also demonstrates the effectiveness of the region descriptor; however as a result of the small number of detected SURF feature points, the overall number of region descriptors is small. This may increase the probability of incorrectly identified matches. An example of such a situation is shown in Figure 11. A region descriptor of the shoe is mismatched due to the incorrect assessment of two graph similarities.

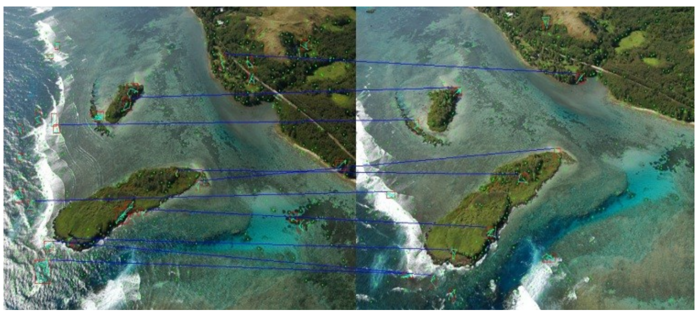


Figure 10. Descriptor matching in natural scene images.

Lastly, the region descriptor was tested on image sets obtained from an unmanned air vehicle (UAV) [15]. The UAV image sequence is taken from a high altitude and may be important for image stitching and registration applications. The descriptor further demonstrated its discriminative ability by correctly matching 83.4% of the detected regions. The proposed descriptor is shown to offer accurate matching in scenes that contain homogenous features, such as the water shown in Figure 10.

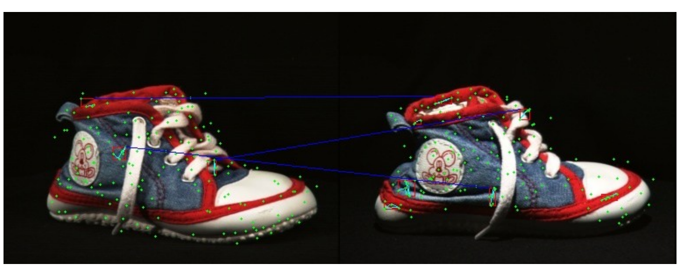


Figure 11. Mismatched local descriptor.

In a few cases the proposed region descriptor may be mismatched such as the shoe shown in Figure 11. Future work will aim to mitigate this issue by comparing the spatial relationships of the region descriptors themselves in a global manner. Additionally, more discriminative region descriptors can be formed by increasing the scale-space representation of the invariant feature points. As a result, we are creating larger connected graphs and thus increasing their uniqueness.

V. Conclusions

In this paper we have proposed a novel region descriptor that couples invariant local features with color information and spatial relationships. In doing so, the descriptor has shown to have a higher discriminative power than the SURF descriptor. The robust matching algorithm is computationally simple while maintaining high accuracy. Experimental results show the effectiveness of the descriptor in scenes with varying geometric and photometric properties. Such a descriptor can be helpful in content-based retrieval systems for general objects as well as man-made structures and natural scenes.

References

[1] D. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”, Int’l J. Computer Vision, vol. 2, no.60, 2004, p91-110.

[2] Ryu, J.-B.; Park, H.-H.; Park, J.; , "Corner classification using Harris algorithm," Electronics Letters , vol.47, no.9, pp.536-538, April 28 2011.

[3] Lee, J.; Roh, K.; Wagner, D.; Ko, H.; , "Robust local feature extraction algorithm with visual cortex for object recognition," Electronics Letters , vol.47, no.19, pp.1075-1076, September 15 2011.

[4] Fan Tiesheng; Niu Bing; Wang Qingsong; Wang Tao; Qu Dapeng; , "Novel Stereo Matching Method on Multi-scale Harris Corner Points," Computer Science and Computational Technology, 2008. ISCSCT '08. International Symposium on , vol.1, no., pp.167-170, 20-22 Dec. 2008.

[5] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool, “Speeded-Up Robust Features (SURF),” in International Journal of Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346-359, 2008.

[6] Zhi Li Song; Junping Zhang; , "Remote Sensing Image Registration Based on Retrofitted SURF Algorithm and Trajectories Generated From Lissajous Figures," Geoscience and Remote Sensing Letters, IEEE , vol.7, no.3, pp.491-495, July 2010.

[7] Flint, A.; Dick, A.; van den Hengel, A.; , "Local 3D structure recognition in range images," Computer Vision, IET , vol.2, no.4, pp.208-217, December 2008.

[8] Pinto, B.; Anurenjan, P.R.; , "Video stabilization using Speeded Up Robust Features," Communications and Signal Processing (ICCSP), 2011 International Conference on , vol., no., pp.527-531, 10-12 Feb. 2011.

[9] Peng Fan; Aidong Men; Mengyang Chen; Bo Yang; , "Color-SURF: A surf descriptor with local kernel color histograms," Network Infrastructure and Digital Content, 2009. IC-NIDC 2009. IEEE International Conference on , vol., no., pp.726-730, 6-8 Nov. 2009.

[10] Jian Wu; Heng-jun Yue; Yan-yan Cao; Zhi-ming Cui; , "Video Object Tracking Method Based on Normalized Cross-correlation Matching," Distributed Computing and Applications to Business Engineering and Science (DCABES), 2010 Ninth International Symposium on , vol., no., pp.523-527, 10-12 Aug. 2010.

[11] Sirmacek, B.; Unsalan, C.; , "Urban-Area and Building Detection Using SIFT Keypoints and Graph Theory," Geoscience and Remote Sensing, IEEE Transactions on , vol.47, no.4, pp.1156-1167, April 2009.

[12] Kisku, D.R.; Rattani, A.; Grosso, E.; Tistarelli, M.; , "Face Identification by SIFT-based Complete Graph Topology," Automatic Identification Advanced Technologies, 2007 IEEE Workshop on , vol., no., pp.63-68, 7-8 June 2007.

[13] E.W. Dijkstra, “A Note on Two Problems in Connexion with Graphs,” Numerische Mathematik, vol. 1, pp. 269-271, 1959.

[14] Amsterdam Library of Object Images, http://staff.science.uva.nl/~aloi/.

[15] Water and Environmental Research Institute of the Western Pacific (WERI) at the University of Guam, http:// http://www.weriguam.org/.