Novel Approach In Wide Baseline Matching: Detecting Unique Features In a Scene

Xing Yifan A0105591J National University of Singapore
Thesis Supervisor: Associate Professor Ng Teck Khim

Abstract. This paper presents a novel approach in feature matching and feature descriptor constructing under the context of wide baseline matching. In the current pool of literature, several robust methodology in detecting interesting points between pairs of images have been proposed. These include SIFT [1] and SURF [2] features. However, one of the drawbacks of these methodologies includes the ignorance of color information in the images, which contains a large amount of information in terms of distinctiveness of the images. Several approaches have been proposed to combine color information with the robust intensity feature detectors like SIFT and SURF [14],[15],[16],[19],[25],[27],[31]. However, during the feature detection phase of these proposed approaches, they still mainly depend on the concept of high local intensity contrast from the SIFT approach. In this paper, a novel approach of looking for unique features out of the whole image is proposed. The methodology is further compensated with a specially designed feature descriptor based on color and intensity profile of the detected unique feature image patches. Experimentations are carried out to examine the precision of this proposed approach of image correspondence matching.

Keywords: image correspondence matching, feature descriptors, feature detection, color histogram, histogram of gradients, KL divergence, Earth-Mover distance.

1. Introduction

Image correspondence matching has become ubiquitous in the areas of computer vision and image retrieval. Further, the image correspondence matching problem under a wide baseline context has become a popular area of study. Under a context of wide baseline movement of the cameras, the scene type usually displays the following properties Tell, D.[13]. For any point in the scene, its corresponding point may be anywhere in the other image. The neighborhoods around corresponding points may exhibit similarity, but due to the wide baseline movement of the cameras, it is not possible to directly compare the neighborhoods of the corresponding points. Further, in some cases, objects that are visible in one image may be occluded in the other. To tackle the problem of wide baseline matching, the majority of the methods work on local feature detection and matching between images. Such methodology of local image feature detection and matching is increasingly being applied in applications of real-time object recognition [42], 3D reconstruction [43], panorama stitching [44], robotic mapping [45], visual servoing [13] and video tracking [46].

A necessary property of the local feature detection approaches is that the features detected must be repeatable identified. Further, it is desired that as many such features to be extracted from images with different viewpoints. On the other hand, the detected features must be robust to photometric and geometric changes with the assistance of the descriptors of these features. Existing methodologies cover a wide range of properties of the image pixel and its neighborhood to build up the feature detector and descriptor. SIFT [1] has been proven to be a robust feature descriptor in terms of geometrical changes, especially scale change. It makes use of an approximation to Laplacian space, i.e., Difference of Gaussian at different scale space to find feature points with high local intensity contrast and achieved invariance to scale changes. SURF [2] speeds up the detection phase of SIFT by making use of an integral image and it simulates Difference of Gaussian /Laplacian space with a Haar wavelet response by using box filters. There are several extensions of SIFT out there in the literature pool. One of them makes use of a connected graph structure of the SIFT features detected and forms a more robust feature by clustering the features to form connected component [4]. Lindeberg, T [3] makes use of a newly defined Hessian feature strength measure for feature detection and then he links image features to form feature trajectory for better matching performance. Further, there are methodologies in constructing more feature matches based on an initial set of feature matches from standard feature matching algorithms. Ramalingam, S., Antunes, M., Snow, D., Lee, G., & Pillai, S.[11] proposed a method to generate and match new points using virtual lines constructed using pairs of key points, which are obtained using standard feature point detectors. Pritchett, P., & Zisserman [8] makes use of a set of initially found good matches and fine tune the homograhy matrix then expand out to obtain more matches in an iterative manner. Other intensity based feature detection and description methodology include HOG and MSER. In [40], a histogram of gradients over the cells divided from the whole image is computed based on the horizontal and vertical pixel gradient. A good performance precision in human detection is achieved by this HOG methodology. In [38], a feature descriptor named MSER is obtained via a watershed approach. Connected regions with certain intensity threshold range will be selected as a good feature if they remain connected/stable over a set of threshold values applied to the intensity of the image pixels. While this approach has a good performance in runtime, MSER is less stable under varying lighting conditions and contrast, and it is sensitive to parameter settings.

However, many of the above mentioned approaches are purely intensity based and ignored the important role that color plays in feature detection and matching. It is believed that color exhibits significant information in terms of the distinctiveness of the images. There are many methodologies out there in the literature pool that rely on color descriptors for feature matching. For instance, color histograms [22] constitute a broad category of color descriptors and many variations are built based on it, such as Opponent histogram [22], HSV histogram [29], color co-occurrence histogram [27], color local contrastive descriptor [21] and color coherent vector [48], etc. Another broad category of color descriptor is based on color moments [26] information where the first moment usually refers to the average of pixel values over a color channel, second moment refers to the standard deviation and third moment refers to the skewness. It is noted that color moments with a high order tends to be more sensitive to photometric or geometric changes. Some of these color descriptors do not include spatial information of the pixels, such as color histogram while some do, such as color moments with order greater than 1 and color coherent vector etc. While these color descriptors perform well as computed from different color channels and color spaces, in terms of robustness or invariance to photometric and geometric changes, intensity based approaches seem to be more sophisticated [7].

As a result, there are several approaches combining the intensity contrast information and color information to build more robust descriptors. In [23], a concept of color ratio over R/G/B color channel is combined with Canny edge detector to form a descriptor including both edge and color information. In [10], a purely intensity based approach (with local maxim and maximum value suppression) is adopted for feature detection and then a color moment descriptor is built upon the detected features. Further, by adding color information to the previously mentioned robust local intensity contrast descriptor such as SIFT and SURF improves the matching performance significantly. CSIFT [14] makes use of a color invariant space and computes the normal SIFT descriptor based on that color invariant space developed by Geusebroek et.al [49]. PCSIFT [19] transforms RGB color space to a perception color space and runs SIFT on the color channes of the perception color space. With a similar concept, HSV-SIFT [24] works on the HSV color space and Hue-SIFT [29] runs SIFT specifically on Hue channel of the transformed HSV color space. In [25], color-moments are combined with SIFT to form a more robust descriptor. CIC-SIFT [15] makes use of color independent component analysis to transform RGB color space to the independent components and runs SIFT on these independent components. CCH-SIFT [27] appends the color co-occurrence histogram to the normal 128 dimensional SIFT feature descriptor to form a combined feature descriptor for better matching accuracy. CH-SIFT [28] makes use of a color local kernel histogram and combines this histogram with SIFT feature descriptor by a weighted factor. In addition, combing SURF feature detector and color information has also exhibited a better matching performance. CW-SURF [31] makes use of a color invariant that is a derivative of the one used in CSIFT and is experimented to be of the most distinctiveness from Geusebroek et.al [49], and then runs normal SURF detector on that computed color invariant value. CSURF [32] appends a local kernel color histogram (similar to the one used in CH-SIFT) over the chromatic color space to the normal SURF descriptor with a weighting factor to achieve a more robust feature descriptor.

(Detailed discussion of the descriptors of the above approaches can be found in section 3.)

However, the detection phase of these approaches that combine SIFT/SURF with color information still largely depend on the intensity contrast approach exhibited in SIFT. In this paper, a novel method of detecting unique features from the scene is proposed. Sometimes, a good feature to detect and match between two images of the same scene might not exhibit a good local intensity contrast as required by most of the feature detection methodologies including SIFT and its derivatives. On the other hand, such a feature point might just be very unique as compared to other parts of the scene. The uniqueness can be captured by color profile or intensity profile. For example, as shown in fig1, the marked green patch does not exhibit a good local intensity contrast and may not be picked up as a good feature by algorithms like SIFT/SURF, however, as viewed from the entire image’s perspective, this image patch displays a uniqueness in terms of color profile – it is the only image patch that purely consists of green color.

fig1: demonstration of a unique feature patch that does not necessarily exhibit a strong local intensity contrast.

2. Objectives

The objective of this paper is to examine the performance of the proposed approach of looking for unique features for image correspondence matching. The robustness of the unique features and their corresponding descriptors will be tested under changing photometric environments, i.e., changing illuminance and changing lighting source etc as well as changing geometric environments, i.e., rotation, scale and viewpoint change etc. Further, different feature criteria for determining the uniqueness of the feature image patches will be compared and examined. Future expansion on selecting feature criteria for determination of unique image patches will be discussed.
The main flow and contribution of this paper can be summarized as below:

1) A detailed literature survey of existing methodology of feature detection and descriptors will be discussed in terms of their pros and cons. This will include the examination of feature descriptors on gray scale/pure intensity profile, feature descriptors based on purely color information and feature descriptors that cover both intensity information and color information.
2) The algorithm design of our proposed approach of detecting unique feature patches out of the whole image will be presented. A detailed description of the feature descriptor proposed based on color histogram and intensity to cope with the unique feature patches extracted during the detection stage will be presented.
3) A feature matching procedure is proposed with the incorporation of image pyramid. Two histogram dissimilarity measures, Jensen Shannon Divergence and Earth Mover Distance will be discussed for matching the feature profile histogram from the descriptors established in step (2).
4) Experimentation result on test images under illuminance change, rotational change and viewpoint change on a wide baseline basis will be presented to examine the performance of the proposed feature descriptor.

3. Literature review

Different approaches, which describe nearby regions of the feature points, have been introduced into the literature. Most of the feature descriptors work with the aim to increase the distinctiveness of the feature region described for the later matching phase. Several widely used feature descriptors are introduced in a detailed manner in this section. They include descriptors focusing on intensity contrast, descriptors working on color information and ones combine both intensity information and color profile.

3.1 Intensity based descriptors:

a. HOG

First introduced in [40], histogram of image pixel gradient with respect to horizontal and vertical direction captures local object appearance and shape information within an image patch since such information can be described by the distribution of intensity gradients. Typically, the image will be divided into cells which can either be rectangular or circular cells. Within each cell, histogram of gradients are binned into several directions and weighted by the magnitude of the gradient. Final HOG descriptor will be the concatenation of histograms of the cells. HOG descriptor is typically invariant to geometric and photometric transformations but not for object orientation. It is particularly suited for human detection in images.

b. SIFT

SIFT [1] feature is widely accepted as one of the most robust feature detectors and descriptors since it is scale invariant and invariant to rotational, illuminance change to a certain degree. For feature detection phase, a DoG is applied as approximation to LoG at different scale space over the different octaves of a Gaussian image pyramid. A local feature is compared not only to its 8 neighbors on the same scale level but to the 9 pixel neighbors on adjacent scale level as well. This helps detection of feature point in a scale invariant manner.

For feature descriptors, a 4x4 intensity gradient histogram is built on 4x4 sub regions of the region computed from the feature scale. Each histogram has 8 bins of orientations. This constitutes the 3D histogram with dimensions of orientation bin and (x,y) coordinate of the pixels. Final descriptor is the concatenation of the 4x4 sub regions histogram which is of dimensionality 4*4*8 = 128. It is noted that though this descriptor is similar to HOG, a significant difference is that SIFT feature descriptor are rotated to align orientation before computing the sub region histograms.

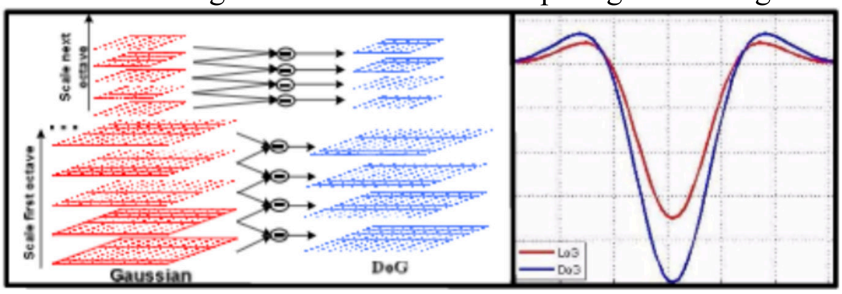


fig2: DoG scale space; DoG function as approximation to LoG function

c. GLOH

GLOH [7] descriptor is a one-step-further extension of SIFT feature. Instead of the 4x4 rectangular sub regions, GLOH picks a log-polar location grid with three bins in radial direction (the radius set to 6, 11, and 15) and 8 in angular direction, which results in 17 location bins (the central bin is not divided in angular directions, shown in fig3). The gradient orientations are then quantized in 16 bins. This gives a 272 bin histogram. Further, the size of this descriptor is reduced with PCA. The covariance matrix for PCA is estimated on 47,000 image patches collected from various images. Finally, GLOH descriptor is made up of the 128 largest eigenvectors from PCA. GLOH descriptor is experimented to out perform the original SIFT descriptor based on a measurement of recall, precision and number of correspondences found [7].

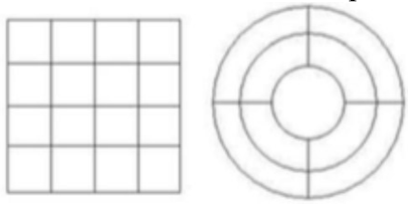


fig3: sub regions used by SIFT(rectangular) and GLOH(circular)

d. SURF

SURF [] works with the aim of speeding up the feature detection and matching of SIFT approach. Instead of DoG, determinant of Hessian is used for detection of blobs of interest. In addition, determinant of Hessian is further approximated by box filtering which can be computed very fast using the concept of integral image.

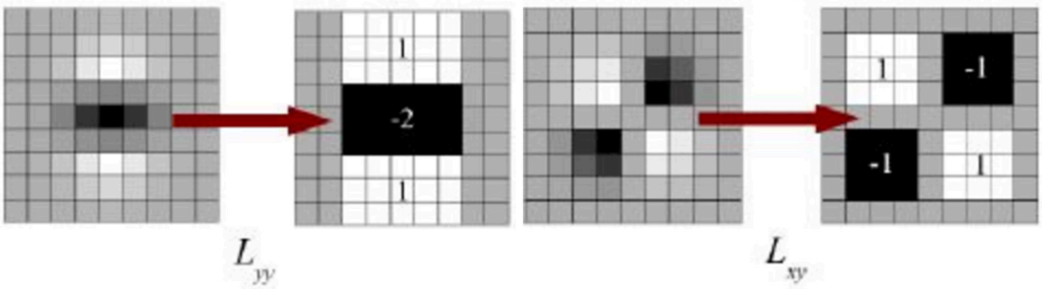


fig4: box filtering approximation to determinant of Hessian; L(x,y,t) = Gaussian(x,y,t) * I(x,y), * is for convolution, t is scale level.

In SURF, the scale space is implemented by applying box filtering at different sizes. For feature descriptor, SURF uses Haar Wavelet responses in horizontal and vertical direction and the process can be greatly speeded up using integral image. Then, a pixel neighborhood of size 20scale x 20scale is taken. The neighborhood is further divided into 4x4 regions similar to SIFT and for each sub region, horizontal (dx) and vertical (dy) wavelet responses are measured. A feature vector over each sub region is formed as following. Thus, final SURF descriptor will be of dimensionality 4*4*4 = 64.

$$v = \left(\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|\right)$$

It is noted that SURF works well for images with blurring and rotational changes while it might not perform well at view-point changes and illumination changes.

e. MSER

MSER [38] defines a feature descriptor of maximally stable extremal regions that are defined solely by an extremal property of the intensity function in the region and on its outer boundary. With a predefined set of intensity thresholds, the image will be filtered based on these thresholds. A pixel will appear as ‘white’ if its intensity is above the threshold. It is expected that by decreasing the intensity thresholds, more ‘white’ regions will appear and merge. The set of all connected components throughout the set of thresholds is the set of all maximal regions (minimal regions is defined similarly). Such sets of connected components constitute the MSER descriptor. It is experimented that MSER is robust to affine transformations and viewpoint changes.

f. ORB

In [50], a feature detector and descriptor named ORB is built based on modifications to the FAST [51] feature detector and BRIEF [52] feature descriptor.
FAST feature detector works by considering a segment test on a circle of 16 pixels (a Bresenham circle of radius 3) to classify if a candidate point p is a corner point. High-speed test with pixels with label 1, 9, 5 and 13 and incorporation with machine learning algorithms with training image patches obtained from the target application makes it possible to generate a very fast feature detector. ORB further extends FAST detector by adding rotational invariance via assigning an orientation to each corner and aligning them. ORB also adds scale invariance by detecting FAST feature at different octaves of a Gaussian image pyramid.
For feature descriptor, ORB builds its descriptor based on BRIEF [52], where a vector of binary tests of pixel intensity of pairs sampled from an image patch is captured. ORB makes modification of the BRIEF descriptor by 'steering' it, i.e., by incorporating an orientation alignment (obtained from modified FAST detector) for the testing points included in the original BRIEF descriptor. Finally ORB runs a greedy search among all possible binary tests for BRIEF to find the ones having highest variance, which is a desired property for a discriminative feature descriptor. This final descriptor is named rBRIEF.
The performance result of ORB is computationally faster than those of SIFT and SURF while maintaining a relatively strong robustness to photometric and geometric changes. Still, ORB works on purely intensity information of the image patches.

3.2 Color Descriptors:

a. Color Histogram

Color histograms have been widely used as feature point's neighborhood descriptor due to its computational efficiency and insensitivity to small changes of camera viewpoint. However, color histogram in general does not contain spatial information of the pixels. Thus, there are certain derivatives [21] [27] [48] of color histogram which incorporate spatial information to achieve a better performance

RGB histogram The most general color histogram will be from R,G,B channel, which is a combination of the 1D histograms from each of the channel of R,G,B space. Unfortunately, this histogram performs poorly in terms of invariance property [22].
Opponent histogram Opponent histogram is based on the following opponent color space. O1 and O2 channel are invariant to light intensity change while O3 contains intensity information.

$$\begin{pmatrix} O_1 \\ O_2 \\ O_3 \end{pmatrix} = \begin{pmatrix} \frac{R-G}{\sqrt{2}} \\ \frac{R+G-2B}{\sqrt{6}} \\ \frac{R+G+B}{\sqrt{3}} \end{pmatrix}. \tag{49}$$

HSV histogram HSV histogram is based on the transformed Hue, Saturation and Value channel, where Value channel corresponds to the intensity information, Saturation channel corresponds to colorfulness information and Hue channel represents the different colors. It is noted that Hue channel is invariant both to shading and to specular effects as it is perpendicular both to the shadow–shading direction and to the specular direction. However, it becomes unstable when it is close to the end of saturation channel (large or small value of saturation). To address this issue, Weijer, J [29] proposes an approach of weighting Hue with its corresponding Saturation value to obtain a more robust color model.

LCCD Local color contrast descriptor In [21], a color descriptor that adds spatial information to the color histograms by considering the local contrast of color channels measured by F Divergence is proposed. Two local contrastive feature descriptors are appended to the normal SIFT feature descriptor to increase robustness. The first one is an F Divergence difference between Opponent histograms of the central patch and its surround 8 neighborhood patches. The second feature descriptor of the feature region is the F Divergence between different channels of the R,G,B histogram (R and G, R and B, B and G). It is verified by experimentation that a compensation of the newly added two color descriptor to the SIFT descriptor improves the performance of individual SIFT.

Color Co-occurrence histogram In [27], a concept of color co-occurrence histogram (CCH) is introduced, where CCH captures the counts of number of pixels that are a certain distance away and fall within particular color bins. CCH is actually equivalent to the statistical second order of the color probability distribution. CCH not only captures the pixel color distribution information but also the spatial correlation of the neighborhood pixels.

b. Color Moments

Another broad category of color descriptor work on color moments by treating the color channel values (for example, R,G,B values) as data points from a certain distribution. Then, moments can be defined on this distribution. In [26], the following general color moments are defined.

$$M_{pq}^{abc} = \int \int x^p y^q [I_R(x,y)]^a [I_G(x,y)]^b [I_B(x,y)]^c dx dy.$$

p+q defines the moment order and a + b + c defines the degree. Usually, color moments up to the first order and second degree are used. Further, color moment invariants can be constructed from the above generalized color moments. It is mentioned in [26] that color moment invariants with invariant property to light intensity/light color change or shift can be obtained.

c. Color Coherence Matrix/Vector

In Pass, G., Zabih, R., & Miller, J [48], a concept of Color Coherence Vectors (CCV) is developed to make up the drawback of Color Histogram's lack of spatial information by measuring the spatial coherence of the pixels with a given color. The paper classifies each pixel in a quantized color histogram bin as either coherent or incoherent, based on whether or not it is part of a large similarly colored region. A color coherence vector (CCV) stores the number of coherent versus incoherent pixels with each color. However, the time complexity of this approach rises if we want to apply concept of CCV over each of the dense small patches rather than on the whole image.

d. Color Autocorrelogram

In [47], a color correlogram is defined to express how the spatial correlation of pairs of colors changes with distance. The correlogram is defined based on the counts of pairs of pixels that are a certain distance away with the pairs of pixels binned into two different color quantization bins. This color autocorrelogram idea is similar to that of the color co-occurrence histogram where spatial information of the color pixels are incorporated by considering the Euclidean distance between them.

3.3 Combined Color and Intensity descriptors:

a. Color SIFT

There are couples of approaches in the literature pool that combines SIFT with different color descriptors for a more robust feature descriptor. The following section includes a detailed discussion of them.

CSIFT In [14], a CSIFT descriptor is proposed based on a color invariance model developed by Geusebroek et.al [49]. The color invariant used is calculated from RGB space with a linear transformation to the spectral differential quotients and further, to spatial differential quotients by a convolution with Gaussian filters. Further, normal SIFT feature descriptor is built based on the color invariant derived from the spatial differential quotients. In [53], a C-invariant that can be viewed as the normalized component of the Opponent color space (O1/O3, O2/O3) is used. Similarly, normal SIFT feature descriptor is then build on the C-invariant value derived. Though the CSIFT descriptor works well under photometric variation due to the color invariance used, it is not efficient in practical use where there are many geometric (rotation, scale) changes.

PC-SIFT In [19], a perception-based color space [20] which is invariant to illumination changes is used. It is mentioned that the distance of colors in perception space matches well with the perceptual metric used by human viewers. SIFT features are detected on the three channels of the perception color space respectively. For feature descriptor, instead of using a gray scale intensity gradient in the normal SIFT descriptor. PC-SIFT makes used of 3D color vector's gradient and magnitude for building the descriptor where the angle between the two 3D gradient vectors (**gx**, **gy**) is the orientation and norm of the difference of two 3D gradient vectors is the magnitude (weighted with Gaussian window).

HSV-SIFT In [24], Anna Bosch builds a HSV SIFT descriptor based on the HSV channels. Further, a different sub region binning methodology is applied in the descriptor. Instead of a 4x4 rectangular sub region, HSV-SIFT is built based on circular sub patches with circular area support of 4 different radiuses. A good scene classification result is obtained from this descriptor.

Hue-SIFT In [29], SIFT features are extended by adding a color descriptor K to the SFIT descriptor S with a weighting factor λ (shown as B in the following).

$B = \left(\widehat{K}, \lambda\widehat{S}\right)$

For the color descriptor, a Hue value histogram is used. However, the certainty of the hue is inversely proportional to the saturation. Hue will become unstable to ends of saturation channel (small or large values of saturation). Thus, Hue histogram is further made stable by weighting each sample against its saturation values.

SIFT-Color Moments In [25], a color moment descriptor is appended to normal SIFT descriptor at each of the RGB channel. The 3 color moments used are mean of the pixel value, standard deviation of the pixel values and skewness of the pixel values over the 3 channels of the RGB space. The concatenated moment descriptor (of dimensionality 9) is appended to SIFT with a weighting factor λ.

CIC-SIFT In [15], an independent component analysis is deployed to learn the transformation matrix **W** from RGB space to the independent component space. The transformation is learnt such that values in vector **y** = (y1, y2, y3) = **W** · (R, G, B) will be as independent as possible. Lastly, normal SIFT descriptors are build on the 3 independent components and CIC-SIFT descriptor is the concatenation of them.

CCH-SIFT In [27], a color co-occurrence histogram is appended to the SIFT descriptor built based on gray scale intensity values of the image patch with a weighting factor α (shown as following).

$$V_{SIFT-CCH} = [\alpha V_{SIFT}, (1-\alpha)V_{CCH}]$$

Color co-occurrence histogram counts the number of pixel pairs that are d = $(\Delta x, \Delta y)$ distance away and falls into the two different color bins. (as discussed in section 3.2). The following graph illustrates the spatial dependency information captured by CCH.

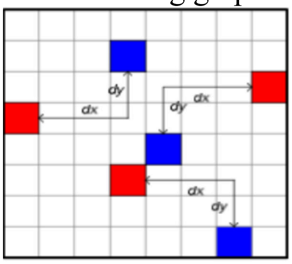


fig5: spatial dependency of CCH on color bins of red and blue.

CH-SIFT In [28], a local kernel color histogram is appended to SIFT normal descriptor to capture the color information of the feature patch. It is mentioned [28] that instead of binning each pixel to a single color bin, each pixel's effect is taken over a neighborhood of the target color bin. A color Gaussian kernel and chrominance color histogram (YUV color space) is used such that the contribution *W(Uk,Vk)* of a pixel *(xk, yk, Yk, Uk, Vk)* to the color bin of *(Uj, Vj)* is shown as following

$$d_U = U_k - U_j \;\;,\;\; d_V = V_k - V_j$$

$$W_j(U_K, V_K) = \frac{K_j}{2\pi\sigma} exp\left(-\frac{d_U^2 + d_V^2}{2\sigma^2}\right)$$

Finally, this 2D histogram over U, V channel of YUV color space is appended to the intensity based SIFT descriptor with a weighting factor to form CH-SIFT.

b. Color SURF

Similar to Color SIFT, there are approaches combine color information and intensity contrast approach by combing color descriptors and SURF descriptors.

CSURF In [32], CSURF makes use of a similar mechanism used in CH-SIFT to derive a local kernel color histogram. The mere difference is that instead of appending it to the SIFT feature descriptor; CSURF is obtained by appending the local kernel color histogram to SURF descriptor.

CW-SURF In [31], CW-SURF descriptor is built based on a similar approach to CSIFT. It also makes use of the color invariance model developed by Geusebroek [49]. The difference is that CW-SURF further extends the C-invariant defined by Geusebroek by taking the Euclidean norm of the different C-invariants defined over the spatial differential quotients [49] to obtain Cw-invariant. The comparison of the different invariants and the derivation of these color invariants can be found in [49] [53]. Finally, a standard SURF detector and descriptor is applied on this derived Cw-invariant value over the pixels in the feature patch. However, similar to CSIFT, one draw back of this CW-SURF descriptor is that it does not handle large geometric changes well since it mainly focuses on photometric invariance derived from the color invariance model.

c. Combined Color Information and Intensity Information
Besides Color SIFT and Color SURF descriptors, there are other methodologies in the literature pool that combines color information and intensity information in a different manner.

In [10], a purely intensity based approach is used for feature region detection where local intensity maxima is detected with non-maximum suppression algorithm. Further, an ellipse constitutes of maximum intensity points around the region is fitted as feature patches. Then, for feature description, 18 color moment invariants are computed up to second order powers of (R,G,B) channel values and up to first order power of (x,y) coordinates. For feature matching phase, Mahalanobis-distance is used as dissimilarity measure between the two feature descriptors. Lastly, rejection of false matches using a geometry constraint and a photometric constraint is performed rather than using RANSAC with known epipolar geometry.

In Diplaros, A., Gevers, T., & Patras, I [23], RGB color ratio (ln(R/G), ln(R/B) and ln(G/B) images) around a neighborhood of a pixel is used as color descriptor; Then Canny's edge detector is applied to get the shape information including magnitude and orientation over the different ratio channels (ln(R/G), ln(R/B) and ln(G/B))to form histograms of color ratios.

4. Algorithm design
In this section, firstly, distance measures between the color profiles and intensity profiles that are used in the unique feature detection phase feature and matching phase are discussed. Secondly, the feature detection phase of unique patches are presented in a detailed manner. Thirdly, two feature descriptors that are used to compensate for the unique feature patches are introduced. Lastly, the feature matching process and an overview of the algorithm flow is presented.

4.1 Distance measure between feature profiles (feature histograms)
Jensen-Shannon Divergence The first distance measure introduced is derived from the F-divergence family [21] from the information theory, which is Jensen-Shannon

Divergence (defined as follows, where D is the KL divergence between the two distributions).

$$\mathrm{JSD}(P \parallel Q) = \frac{1}{2}D(P \parallel M) + \frac{1}{2}D(Q \parallel M) \qquad M = \frac{1}{2}(P+Q)$$

$$D_{\mathrm{KL}}(P\|Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}.$$

It is noted in [21] that the above f-divergence between two distributions is invariant under an invertible transformation g on the space R. Thus, it is believed to be robust against some local image distortions.

Earth Mover Distance Earth mover distance is defined as minimum flow needed to shift from one histogram to the other to make them equal. Definition is shown below (WORK(*P,Q*) with its constraints).

$$\mathrm{WORK}(P, Q, \mathbf{F}) = \sum_{i=1}^{m}\sum_{j=1}^{n} d_{ij} f_{ij}$$

$$\begin{aligned} f_{ij} &\geq 0 && 1 \leq i \leq m,\ 1 \leq j \leq n \\ \sum_{j=1}^{n} f_{ij} &\leq w_{p_i} && 1 \leq i \leq m \\ \sum_{i=1}^{m} f_{ij} &\leq w_{q_j} && 1 \leq j \leq n \\ \sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} &= \min\left(\sum_{i=1}^{m} w_{p_i}, \sum_{j=1}^{n} w_{q_j}\right), \end{aligned}$$

It is required that a distance weighting matrix to be predefined (dij's) and under the context of image patch profile histogram comparison. The distance matrix will be defined based on the profile information used, i.e., if it is HSV histogram or Harris corner histogram or HOG. For instance, if EMD is used against Hue histogram, then the distance-weighting matrix should not be simply the Euclidean distances between the Hue values bins since Hue is 360 degree circular (the 1st bin and last bin are actually very close and should have a small distance weight indeed). One disadvantage of this metric is a relatively high computational time and complexity.

4.2 Unique Feature Detection

As discussed in section 1 and 3, there are many feature detectors in the literature pool such as Harris Corner Detector, SIFT detector, SURF detector and ORB detector. The majority of the methodologies work on gray scale intensity based image and ignored color information mainly due to the large amount of variations in the real world scenes which significantly increases the difficulty in a precise color measurement [7]. While there are attempts to expand these descriptors to color space by running these intensity based detectors on a transformed color channel, such as CSIFT, CWSURF, PC-SIFT, HueSIFT, etc., as mentioned in section 3. Still, these methodologies rely on local intensity contrast based detectors as they are built based upon the gray scale detectors. In this section, a detector that captures the uniqueness of potential good feature patches is introduced. As discussed in section 1, a good feature might not exhibit a good local intensity contrast. On the other hand, it might be just displaying uniqueness in terms of certain feature profiles, such as color profile, shape profile or intensity profile. Thus, the current proposed feature detector works by comparing the patch-by-patch color profile and intensity profile and then selecting the image patches with the highest distinguishability score from a weighted score of color profile and intensity profile. The difference measure between the color profile and intensity profile are captured by Jensen Shannon Divergence/ Earth mover distance between the color distribution and intensity distribution of the two image patches. The following graph presents the detecting procedure.

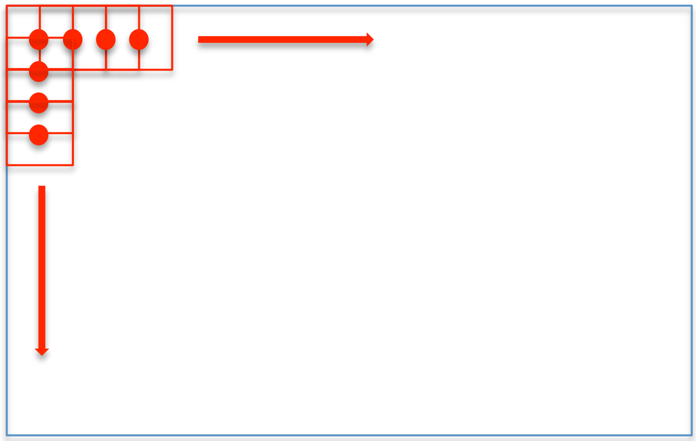


fig6: demonstration of patch extraction of feature detection phase

Patches of a fixed size is extracted over the original image with a shift of half of the patch size. Then, color profile and intensity profile are computed over the patches. Here, a Hue and Saturation color histogram (HS histogram) over the patch extracted is used to measure the color profile and a Harris corner response over the patch is used to measure an intensity profile. Then, Jensen Shannon divergence/ Earth mover distance is applied to compare the two patches' profiles. Final distinguishability score is obtained from a weighted average of the distinguishability score in HS histogram and Harris response histogram. Detected 'unique' feature patches are those with high distinguishability scores as compared to the rest of the patches extracted. Further, it is expected that the profile information used will be able to distinguish 'unique' patches from the non-'unique' patches. For experimentation demonstration, please refer to section 5.

4.3 Build Feature Descriptor

There are pros and cons of all kinds of feature descriptors discussed in section 3. In this paper, I would like to combine both the color information based feature descriptor and the intensity information based descriptor to compensate our uniqueness feature detector. Since Color profile, HOG profile and Harris corner response profile within an image patch has been considered for unique patch detection. It is natural to build descriptor based on these information respectively. In the next section, two feature descriptors are proposed, one based on color histogram and the other based on HOG. During matching phase, the two descriptors will contribute a weighted score to the feature correspondence.

a. Color histogram based

The first descriptor is based on color histogram. Here, Hue and Saturation Channel histogram are used as they exhibit invariance to certain shading and specular effects. Value channel is left out, as intensity is largely sensitive to illuminance changes. Both Hue and Saturation Channel are binned with 16 bins. Further, to tackle the issue that color histogram in general does not incorporate spatial information of the color pixels. I divide the feature patch to a 2x2 sub regions and compute the HS histogram on the sub regions as well to incorporate spatial information in addition to the full patch HS histogram. Finally, a Gaussian weighting function is applied to the full feature patch to obtain rotational invariance to a certain degree. The graph below elaborates the HS descriptor built. Final HS histogram descriptor will be (HueHist_fullpatch, HueHist_subpatches, SaturationHist_fullpatch, SaturationHist_subpatches) with each histogram of dimensionality 16 and thus final dimensionality of the proposed HS Histogram will be of dimensionality 5 (full_patch + 2x2 subpatches) * 2(Hue and Saturation) * 16 for each image patch.

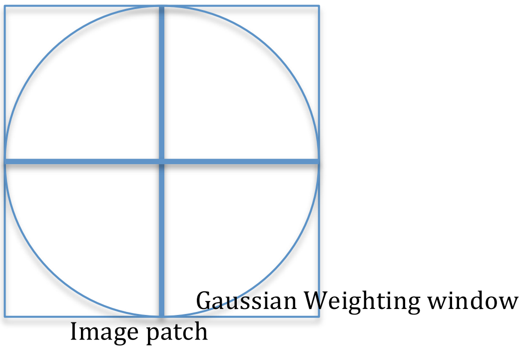


fig7: HS histogram descriptor

b. Histogram of Gradient based

As it is desired to combine intensity information with color information, the 2nd descriptor proposed here is based on HOG of the intensity information. Instead of computing HOG on the 2x2 sub patch, circular support regions are applied here where the HOGs are computed on the sub patches of different radius centered at the center of the patch. The radius is decreased by a step of 1.2 from the original window size. In particular, 4 sub region circular patches are considered as illustrated in the following diagram. For this Circular HOG, 16 bin orientation is applied and the magnitude of the gradient is weighted using a Gaussian window similar to the case in the HS histogram descriptor. Thus, final HOG descriptor will be of dimensionality 5 * 16.

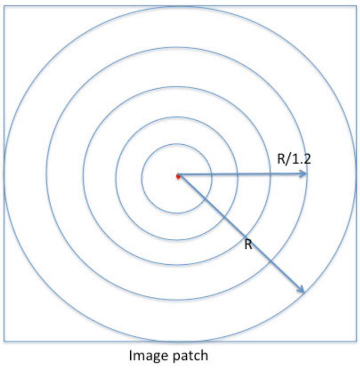


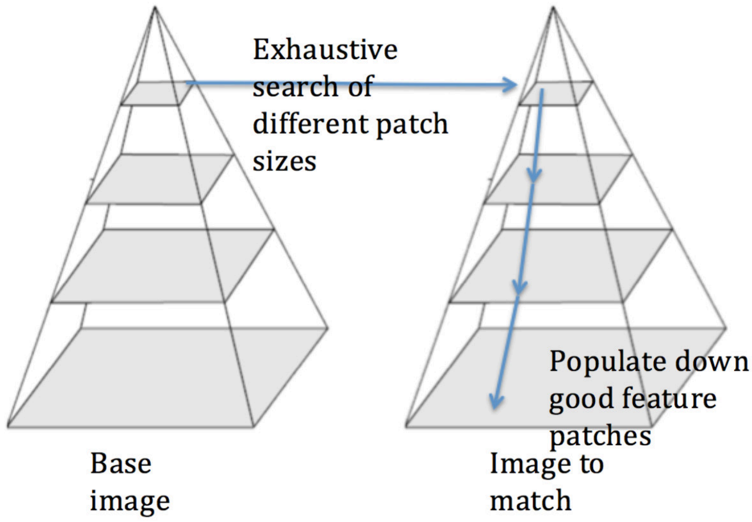
fig8: Circular HOG descriptor

4.4 Feature Matching using Descriptors based on Color and Intensity Profile

For feature descriptor matching, scale invariance is taken into consideration by 1) exhaustive search of patches of different size 2) by adopting a Gaussian Image Pyramid. Further, the distance between two patches are captured by the Jensen Shannon Divergence/ Earth Mover Distance between the corresponding HS histograms/ HOG.

- **Matching with Image Pyramid**
  Firstly, image pyramids are built based on the base image and the target image to match. From the unique feature patches detected from the feature detection phase,

feature patches are first populated up to the top of the image pyramid (smallest resolution). Then, exhaustive search is done with different patch sizes (here I use 5 different patch sizes with a scale change step of 1.2) at the top image pyramid level, where the potential good feature matches are found. Gradually, potential feature matches are populated down to lower levels of image pyramid of target matching image and patch comparisons are done on that lower level to obtain a more accurate set of potential matches. The process is repeated until the lowest level of target image's pyramid is achieved (of original image resolution).



- **Jensen Shannon Divergence (JSD)/Earth-Mover Distance(EMD) between Histograms**

From the two feature descriptor built, 1) HS histogram 2) HOG, dissimilarity between match patches are measure by the pairwise Jensen Shannon Divergence/Earth-Mover Distance of the full patch histogram plus the sub patch histograms. For instance, in the case of HS histogram descriptors of the two image patches, shown as following, call them ***P1*** and ***P2***

$\boldsymbol{P1} = \begin{pmatrix} P1_{HueHists} \\ P1_{SaturationHists} \end{pmatrix}$

$\boldsymbol{P2} = \begin{pmatrix} P2_{HueHists} \\ P2_{SaturationHists} \end{pmatrix}$

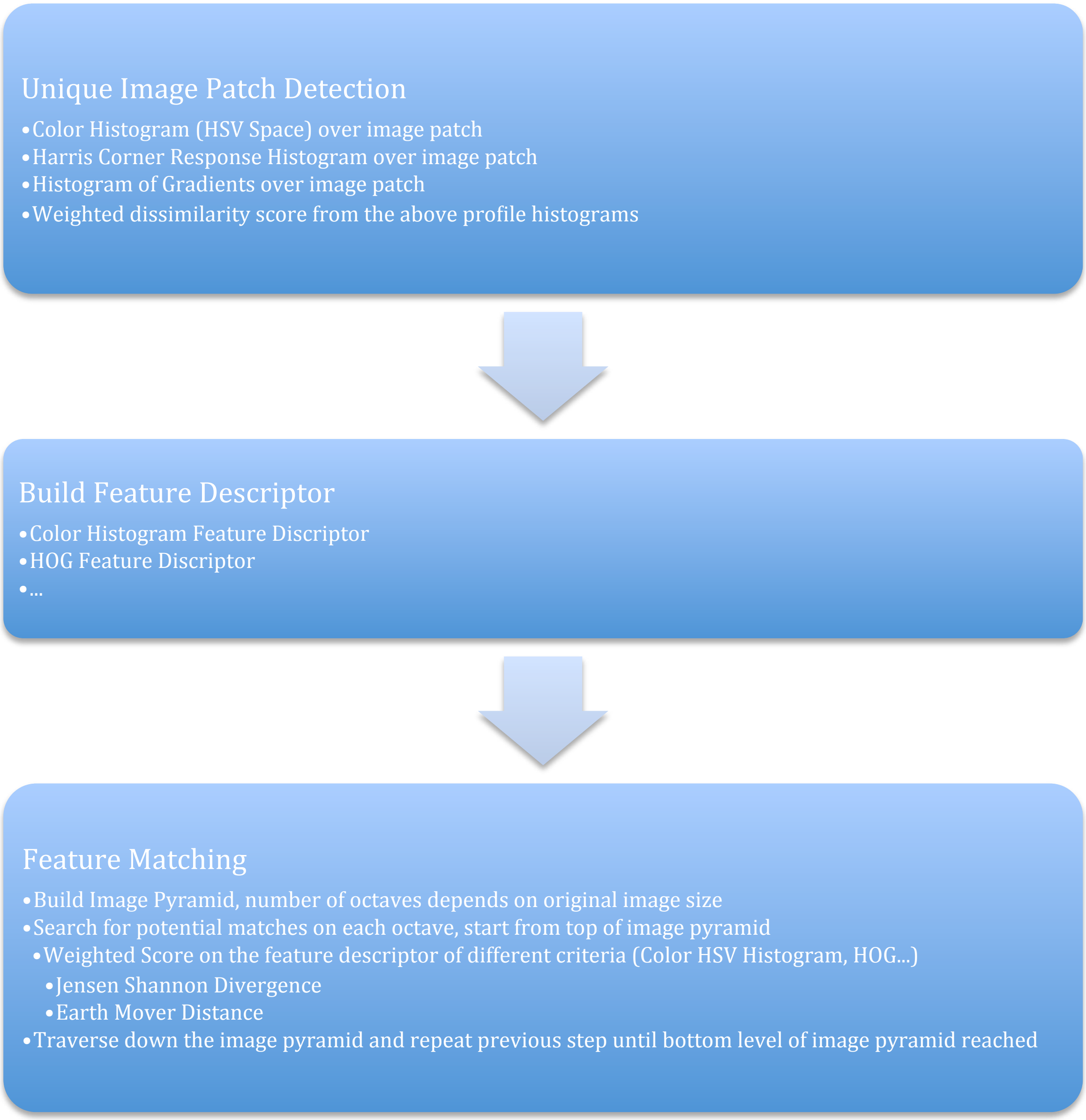
$$P1_{HueHists} = \begin{pmatrix} HueHist_{fullpatch} \\ HueHist_{subpatch1} \\ HueHist_{subpatch2} \\ HueHist_{subpatch3} \\ HueHist_{subpatch4} \end{pmatrix}, P1_{SaturationHists} = \begin{pmatrix} SaturationHist_{fullpatch} \\ SaturationHist_{subpatch1} \\ SaturationHist_{subpatch2} \\ SaturationHist_{subpatch3} \\ SaturationHist_{subpatch4} \end{pmatrix}$$

$$P2_{HueHists} = \begin{pmatrix} HueHist_{fullpatch} \\ HueHist_{subpatch1} \\ HueHist_{subpatch2} \\ HueHist_{subpatch3} \\ HueHist_{subpatch4} \end{pmatrix}, P2_{SaturationHists} = \begin{pmatrix} SaturationHist_{fullpatch} \\ SaturationHist_{subpatch1} \\ SaturationHist_{subpatch2} \\ SaturationHist_{subpatch3} \\ SaturationHist_{subpatch4} \end{pmatrix}$$

Then, the HS histogram *Dissimilarity(**P1**,**P2**)* will be the Euclidean l2 norm of the pairwise distances of the corresponding Hue and Saturation histograms. The distance of individual Hue/Saturation histogram can be measured either by JSD or EMD. Dissimilarity of two image patches in terms of HOG is similarly defined.

- **Weighted score from designed HS Histogram feature descriptor and designed HOG descriptor.**
  Finally, the distance/dissimilarity between two image patches will be a weighted average of the dissimilarity score from HS histogram and score from HOG descriptor.

4.5 Summarized Flow



5. Experiment results

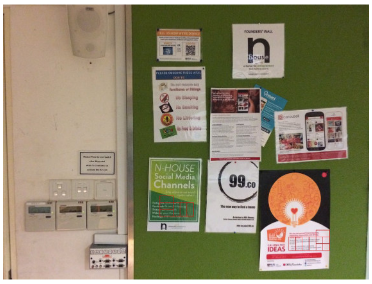
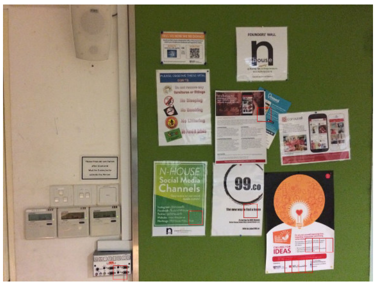
In the following section, experimentation results on different sets of images are present for an initial analysis of the color detector and color descriptor built so far. Photometric changes (illumination changes), geometric changes (view point changes, scale changes and rotational changes) are all present in the below image test sets.

5.1 'Unique' Feature Detection (no matching involved)


HELLO
PANDA

HELLO
PANDA





N-HOUSE
Social Media
Channels
99.co
IDEAS

N-HOUSE
Social Media
Channels
99.co
IDEAS



99.co
IDEAS

99.co
IDEAS

fig9: Left column: unique patches detection purely based on HS histogram profile; Right column: unique patches detection with a weighted score of 0.7 from HS profile and 0.3 from Harris response profile

In feature detection phase, the use of Jensen Shannon Divergence or Earth-Mover distance does make a big impact in the final detection results. The above results are generated using Jensen Shannon Divergence between histograms as dissimilarity measure. In general, the usage of HS profile and Harris response profile are able to sort out patches that appear 'unique' from the entire image's perspective. Further, it can be seen that when we adjust Harris response score to be higher, patches containing corner points will be more likely to be detected. However, there are certain noises in 'unique' feature detection and further investigation and improvements are needed. Please refer to section 7 for further extensions.

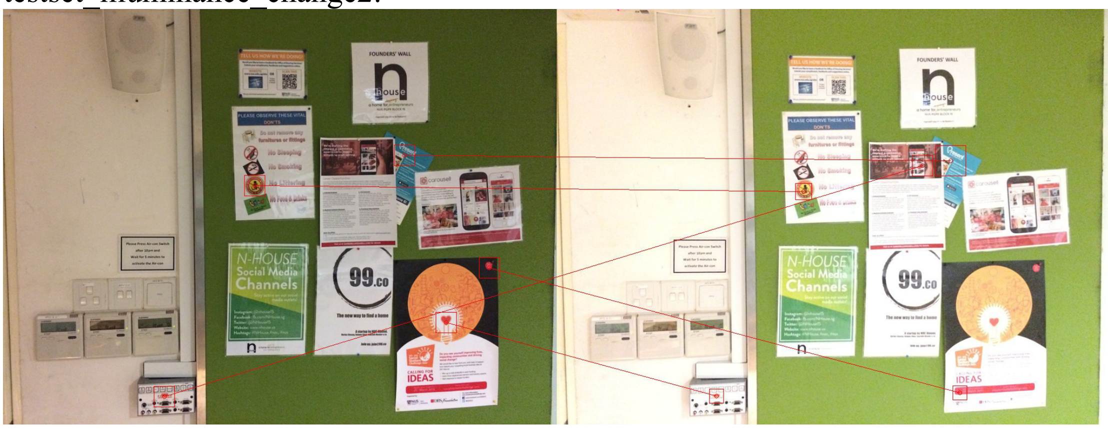
5.2 Feature Description and Matching

Note that the feature patches displayed in this section are picked from a manual process instead of from the feature patches derived from Feature Detection phase, as the purpose is to test the robustness of the feature descriptors alone.

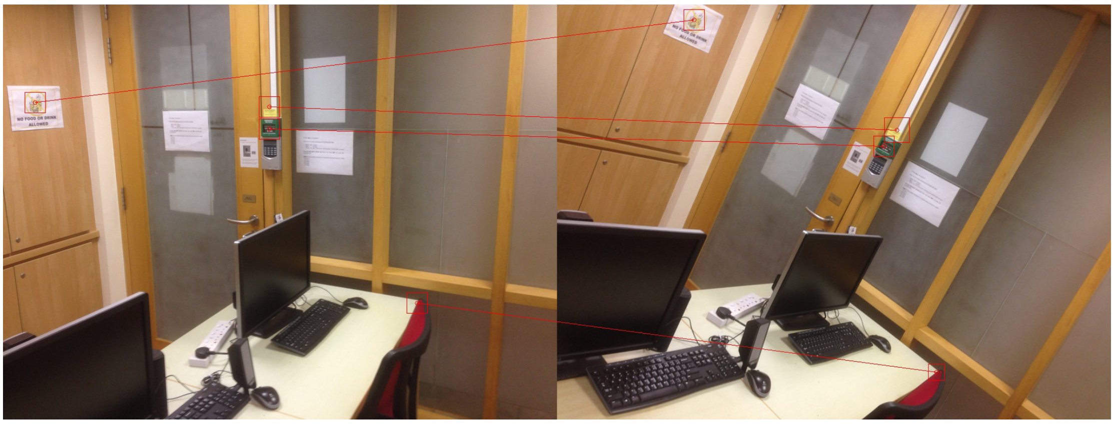
a. HS histogram descriptor alone

testset_illuminance_change1:

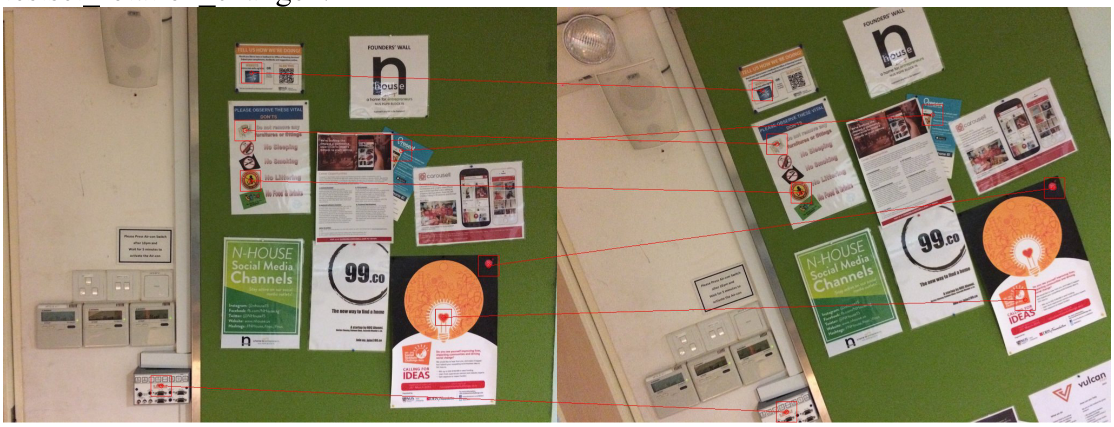
testset_illuminance_change2:



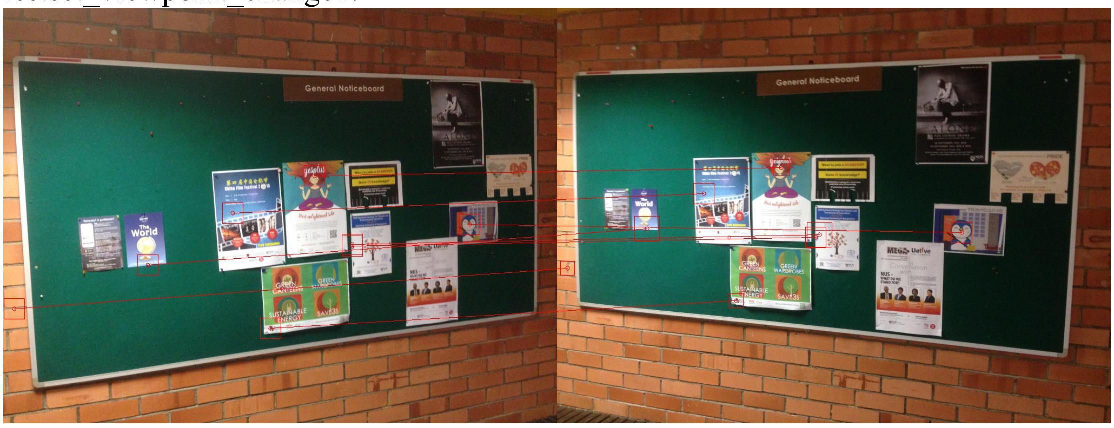
testset_rotation_change1:



testset_rotation_change2:



testset_viewpoint_change1:



testset_viewpoint_change2:

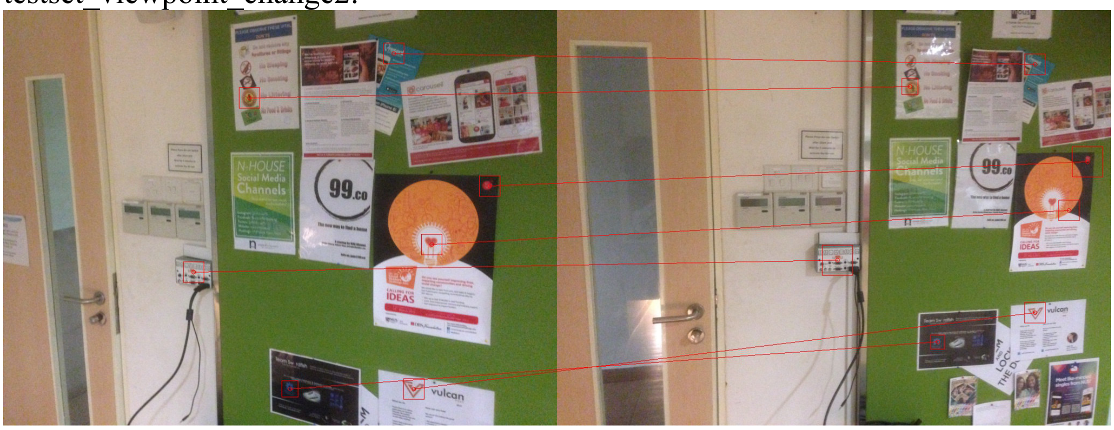
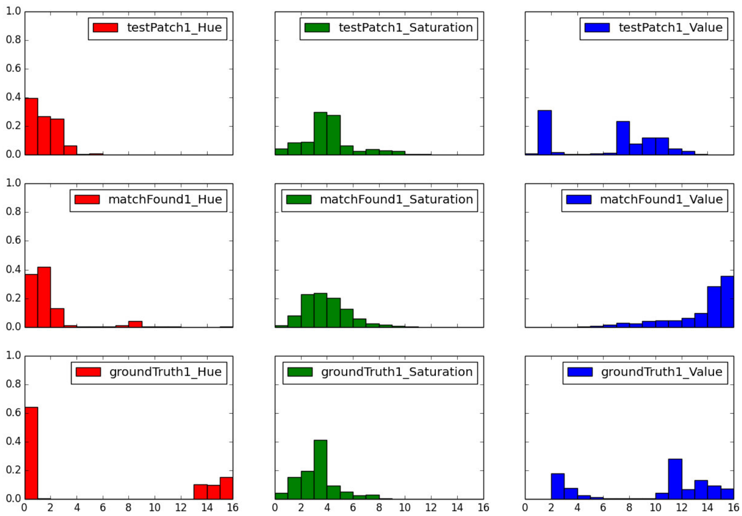


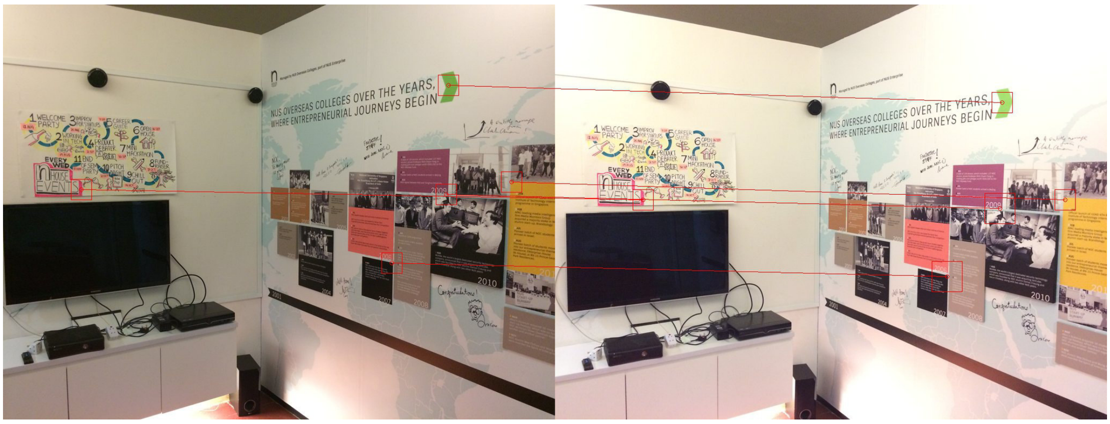
fig10: matching performance of HS histogram descriptor

It is experimented that with separate Hue and Saturation histogram as feature descriptor, a Jensen Shannon Divergence between the 16 bin histogram out performances Earth Mover Distance. This is largely due to the difficulty in getting an accurate distance weight matrix for EMD in the case of Hue histogram.
Further, HS histogram works pretty well with rotational and view point changes. However, when it comes to photometric change such as illuminance changes. It does not perform very well. For example, for the mismatch around the pink patch (shown below) in testset_illuminace_change1, the following statistics shows the actual Hue, Saturation and Value histogram of the mismatched patch (matchFound1) and the ground truth (groundTruth1) as compared to the feature patch (testPatch1) in the base image.

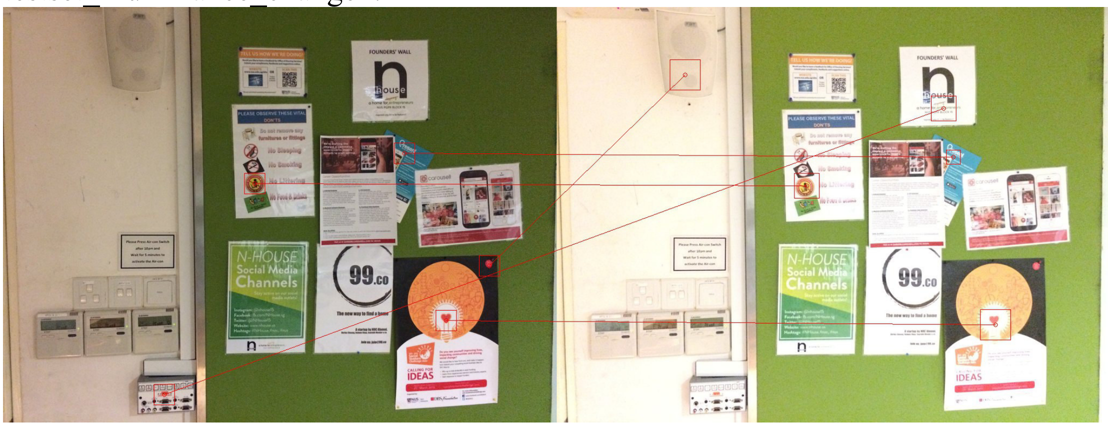


It is noted that the Hue channel histogram looks very dissimilar even between the ground truth and feature patch. This is a possible indication of the instability of Hue value [29] close to the ends of Saturation channel. With further investigation, it might be possible to make improvements by weighting the Hue histogram by its corresponding Saturation channel value or by compensating Hue histogram with other features.

b. Circular HOG alone
testset_illuminance_change1:



testset_illuminance_change2:



testset_rotation_change1:

testset_rotation_change2:

testset_viewpoint_change1:

testset_viewpoint_change2:

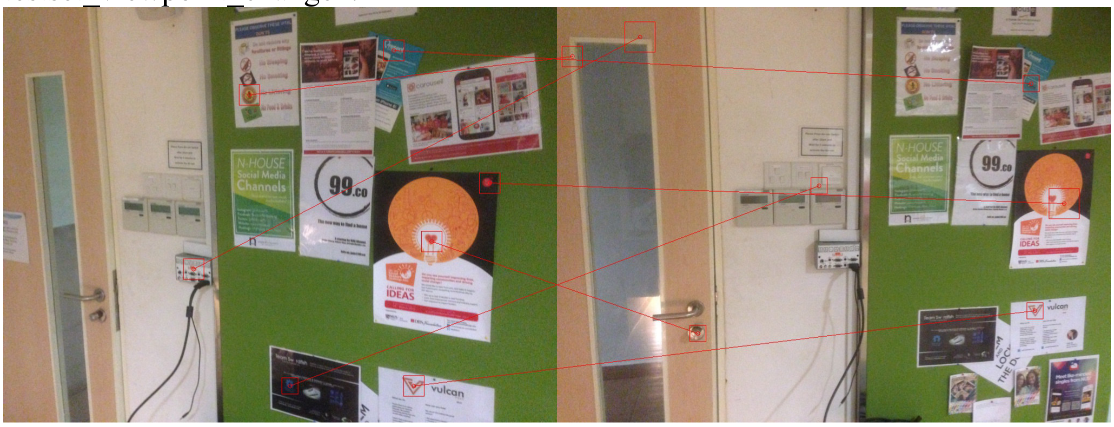


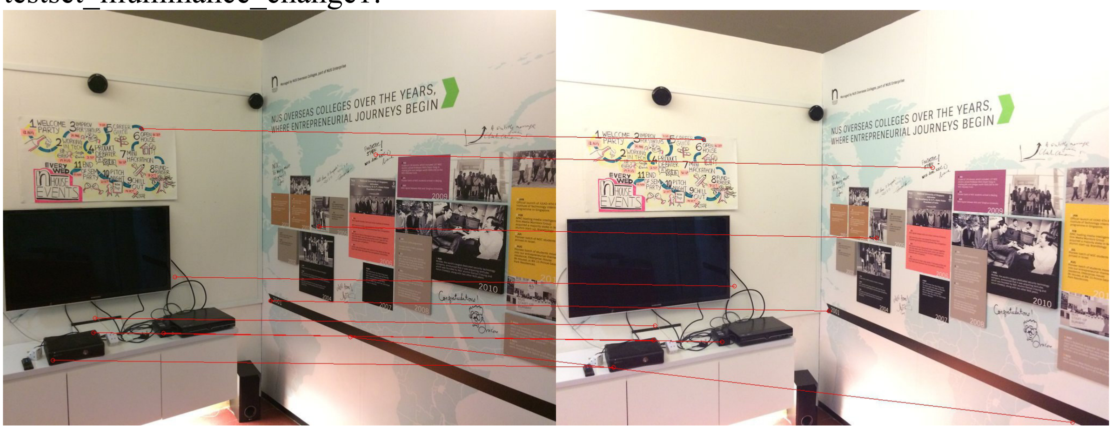
fig11: matching performance of Circular HOG descriptor

It is noted that for Circular HOG, though matching performance is better than HS histogram with illumination changes, matching performance is poor in terms of rotational and viewpoint changes. Even though we have the circular HOG of different scales within the image patch, the invariance to rotational changes is still poorly achieved. Further improvement to this can be achieved by adding an orientation assignment procedure in building the Circular HOG so that invariance to rotational changes could be captured.

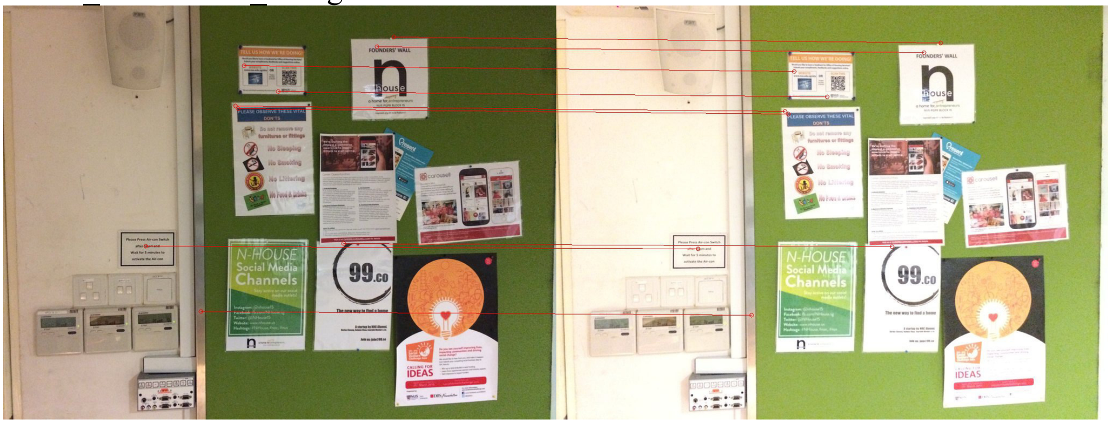
In addition, the weighting coefficients of the combined Descriptor of HS histogram and Circular HOG are yet to be investigated.

c. (*Comparison with SIFT for benchmark, top 10 best matches are plotted)

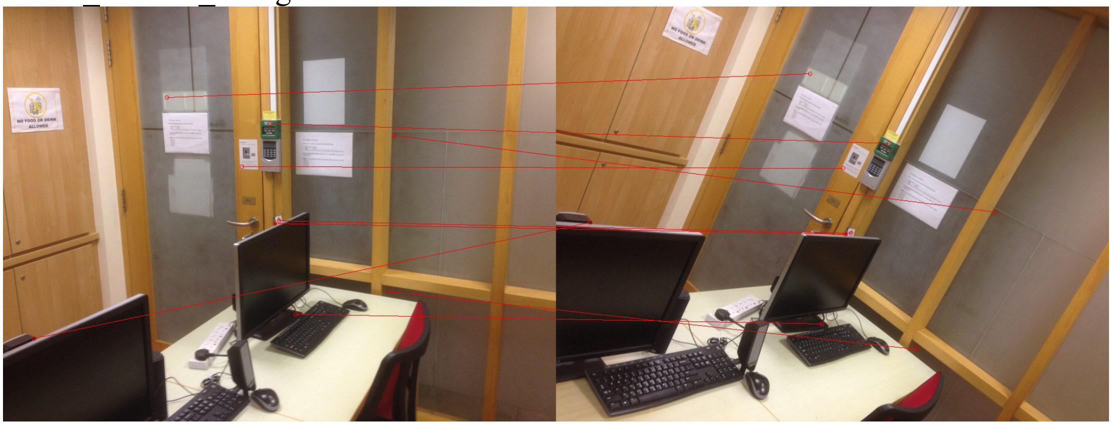
testset_illuminance_change1:



testset_illuminance_change2:



testset_rotation_change1:



testset_rotation_change2:

testset_viewpoint_change1:

testset_viewpoint_change2:

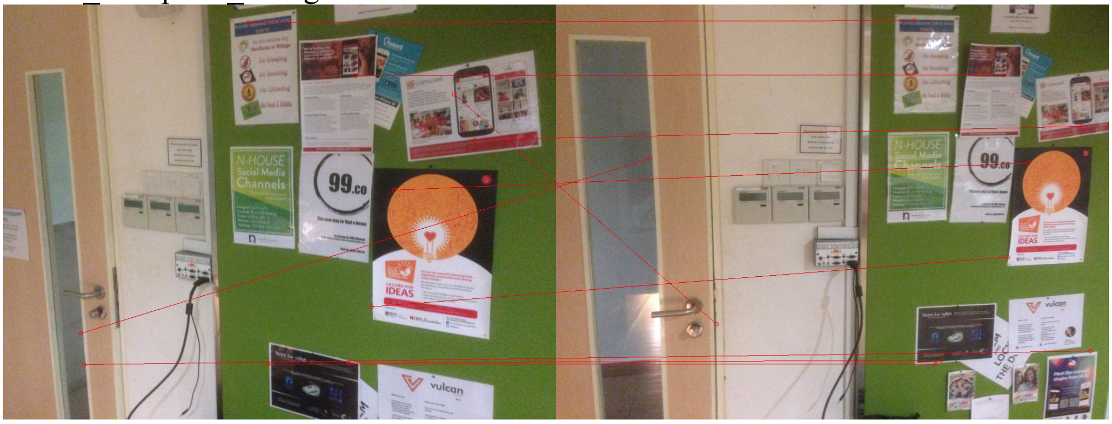


fig12: SIFT feature descriptor performance for benchmarking

It can be seen that for rotational and illuminance changes, our descriptors developed so far are not as robust as SIFT descriptors (Circular HOG performs relatively well for illuminance changes). But for the two viewpoint-change tests displayed above, HS histogram descriptor exhibits a relatively more robust matching accuracy as compared to SIFT.

6. Conclusion

In a nutshell, this paper presents a novel approach in wide baseline matching using the concept of 'unique' feature patches and builds corresponding feature descriptor based on that. Instead of looking at local intensity contrast, the dissimilarity of the patch's color profile and intensity profile are taken into consideration for selection of a good feature patch. In addition, two feature descriptors are proposed to compensate for the 'unique' feature patches detected for matching purpose. One of them is based on HS histogram with sub patches incorporating spatial information and the other one is based on a Circular HOG. Several experiments under photometric and geometric changes are presented to verify the robustness of these two descriptors. The final weighting of the contribution of the two descriptors is yet to be investigated. Further, more feature profiles could be investigated for the detection phase and it is worth looking into other kinds of descriptors for a more robust matching result.

7. Further Extensions and Future Work

There are couples of extensions that can be made to the current mechanism of feature detection and feature description methodology as shown in section 5.

(1) Feature Detection

a. Build image pyramid to incorporate feature detection at different scale

Similar to the feature-matching phase, an image pyramid can be used to help detect 'unique' feature patches of different scale during the feature detection phase.

b. Add more feature profile, such as shape, texture, etc

More feature profile can be considered other than just color and Harris response profile with the image patch. For example, shape, texture information could be captured as well.

c. Use integral image to boost up

Since we don't necessarily need Gaussian window in feature detection phase, a similar approach to the integral image used in SURF [2] could be adopted to boost up the calculation of the profile histograms. For example, we can build an integral image where Integral(x,y) captures the accumulation of color histograms of all pixels at position (<=x, <=y) (in SURF the sum of all pixel's intensity at position (<=x, <=y) is calculated). Then, it will be a cost of constant computation time to get any color histogram (without a Gaussian weighting window) of any rectangular regions in the image.

d. Adjust the weighting of the feature profiles used.

(2) Feature Descriptor

a. Add orientation assignment to the designed HOG descriptor to incorporate rotational invariance.

As mentioned in section 5, to achieve a more robust rotational invariance, one more step of orientation assignment can be added to building up the Circular HOG descriptor with a similar approach from SIFT / ORB descriptor where the orientation bin with the largest magnitude or the orientation towards the intensity centroid [50] can be captured as the feature patch's orientation. Then, during the descriptor-building phase, the feature patch will be aligned according to its assigned orientation to tackle the problem raised from rotational changes.

b. If rotation invariance is not required, we can use integral image to boost up the descriptor building procedure.

If we make the constraint that application of our wide baseline matching approach needs to have orientation stability, rotational invariance will not be required and integral image can be used to boost up the HS histogram descriptor building procedure by getting away from the Gaussian weighting window. The procedure will be similar to that of **(1) c** in this section since without the Gaussian weighting window, each pixel's contribution to the corresponding color bin will be constant (will not vary from patch to patch).

c. Adjust Hue Channel histogram by weighting against its Saturation value.

According to [29], it is noted that Hue channel becomes unstable when it is close to the end of saturation channel (large or small value of saturation). Weijer, J [29] proposes an approach of weighting Hue with its corresponding Saturation value to address this issue. Similar methodologies can be adopted here to investigate possible improvements to the current HS histogram in terms of illuminance changes.

d. Explore new feature descriptors for more robust feature matching

(3) Feature Matching

a. Run detector on target image as well to filter out potential good match places.

Instead of doing an exhaustive search at the top level of the Gaussian pyramids of the target match image, we can also run 'unique' feature detectors on the target image to filter out some potential good match places before we compare the patches using the feature descriptors. The runtime of the matching phase will be expected to further decrease with this implementation.

8. Progress Timeline and Research Plan next semester

Research Progress Plan	Timeline
Feature Description: Fine Tuning on the current two feature descriptor	By November 2015
Feature Detection: Exploring of new feature profiles for 'unique' patch detection	By winter break 2015
Feature Detection: Weighting of different feature profiles' contribution to 'uniqueness' of feature patches detection.	By winter break 2015
Feature Description: Exploring more feature descriptors for more robust feature matching	By early next semester 2016
Feature Description: Weighting of different feature descriptor's dissimilarity score to form the combined final feature descriptor.	By early next semester 2016
Testing: Run a large number of tests on image databases, especially for wide baseline scenes.	In progress until completion of the feature detection and description methodology

9. References:

[1] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision,* 60, 2 (2004), pp. 91-110

[2] Bay, H., Ess, A., Tuytelaars, T., van Gool: Speeded up robust features (SURF). CVIU 110 (2008) 346–359

[3] Lindeberg, T. (2014). Image Matching Using Generalized Scale-Space Interest Points. Journal of Mathematical Imaging and Vision J Math Imaging Vis, 3-36

[4] Bowen, F., Du, E., & Hu, J. (n.d.). A novel graph-based invariant region descriptor for image matching. *2012 IEEE International Conference on Electro/Information Technology*.

[5] E. Tola, V. Lepetit, P. Fua, and S. Member, "Daisy: An efficient dense descriptor applied to wide baseline stereo," *IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 5, pp. 815–830, May 2010.*

[6] Wu, J., Cui, Z., Sheng, V., Zhao, P., Su, D., & Gong, S. (2013). A Comparative Study of SIFT and its Variants. *Measurement Science Review*, *13*(3). http://dx.doi.org/10.2478/msr-2013-0021

[7] K. Mikolajczyk, C. Schmid. A performance evaluation of local descriptors. PAMI 2005

[8] Pritchett, P., & Zisserman, A. (1998.). Wide baseline stereo matching. *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*

[9] Qin, L., & Gao, W. (2005). Image matching based on a local invariant descriptor. *IEEE International Conference on Image Processing 2005*

[10] Tuytelaars, T., & Gool, L. (2000). Wide Baseline Stereo Matching based on Local, Affinely Invariant Regions. *Procedings of the British Machine Vision Conference 2000.*

[11] Ramalingam, S., Antunes, M., Snow, D., Lee, G., & Pillai, S. (2015). Line-sweep: Cross-ratio for wide-baseline matching and 3D reconstruction.*2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[12] Dennis Tell and Stefan Carlsson. Combining topology and appearance for wide baseline matching. *European Conference on Computer Vision (ECCV)*, to appear, 2002.

[13] Tell, D. (2002). Wide baseline matching with applications to visual servoing

[14] Abdel-Hakim, A., & Farag, A. (n.d.). CSIFT: A SIFT Descriptor with Color Invariant Characteristics. *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (CVPR'06)*

[15] D. Ai; X. Han; X. Ruan and Y-W. Chen, "Adaptive Color Independent Components Based SIFT Descriptors for Image Classification," *Pattern Recognition (ICPR), 2010 20th International Conference on*, pp. 2436-2439, 23-26, Aug. 2010

[16] Zhang, Y., Zhaoxing, Z., & Han, X. (2009). Category specific SIFT descriptor and its combination with color information for content-based image retrieval. *Proceedings of the 2nd International Conference on Interaction Sciences Information Technology, Culture and Human - ICIS '09*

[17] Rassem, T., & Khoo, B. (n.d.). Object class recognition using combination of color SIFT descriptors. *2011 IEEE International Conference on Imaging Systems and Techniques*.

[18] Schügerl, P., Sorschag, R., Bailer, W., & Thallinger, G. (2007). Object Re-detection Using SIFT and MPEG-7 Color Descriptors. *Multimedia Content Analysis and Mining Lecture Notes in Computer Science,* 305-314.

[19] Cui, Y., Pagani, A., & Stricker, D. (2010). SIFT in perception-based color space. *2010 IEEE International Conference on Image Processing*.

[20] CHong, H., Gortler, S., & Zickler, T. (2008). A perception-based color space for illumination-invariant image processing. *ACM SIGGRAPH 2008 Papers on - SIGGRAPH '08.*

[21] Guo, S., Huang, W., Xu, C., & Qiao, Y. (2014). F-divergence based local contrastive descriptor for image classification. *2014 4th IEEE International Conference on Information Science and Technology*.

[22]Koen E. A. Van De Sande, Gevers, T., & Snoek, C. (2008). Evaluation of color descriptors for object and scene recognition. *2008 IEEE Conference on Computer Vision and Pattern Recognition.*

[23] Diplaros, A., Gevers, T., & Patras, I. (2006). Combining color and shape information for illumination-viewpoint invariant object recognition. *IEEE Transactions on Image Processing IEEE Trans. on Image Process.,* 1-11.

[24] Bosch, A., Zisserman, A., & Muoz, X. (2008). Scene Classification Using a Hybrid Generative/Discriminative Approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence IEEE Trans. Pattern Anal.* Mach. Intell., 712-727.

[25] Bo, L., & Whangbo, T. (2014). A SIFT-Color Moments Descriptor for Object Recognition. *2014 International Conference on IT Convergence and Security (ICITCS)*.

[26] Mindru, F., Tuytelaars, T., Gool, L., & Moons, T. (2003). Moment invariants for recognition under changing viewpoint and illumination. *Computer Vision and Image Understanding,* 3-27

[27] Ancuti, C., & Bekaert, P. (2007). SIFT-CCH: Increasing the SIFT distinctness by Color Co-occurrence Histograms. *2007 5th International Symposium on Image and Signal Processing and Analysis*

[28] Alilvand Ali, Hamidreza Shayegh Boroujeni, Charkari Nasrollah Moghadam "CH-SIFT: A local kernel color histogram SIFT based descriptor", *International Conference on Multimedia Technology (ICMT), 2011*

[29] Weijer, J., & Schmid, C. (2006). Coloring Local Feature Extraction. *Computer Vision – ECCV 2006 Lecture Notes in Computer Science,* 334-348

[30] Stottinger, J., Hanbury, A., Sebe, N., & Gevers, T. (2012). Sparse Color Interest Points for Image Retrieval and Object Categorization. *IEEE Transactions on Image Processing IEEE Trans. on Image Process.,* 2681-2692.

[31] Jalilvand, A., Boroujeni, H., & Charkari, N. (2011). CWSURF: A novel coloured local invariant descriptor based on SURF. *2011 1st International EConference on Computer and Knowledge Engineering (ICCKE).*

[32] Fan, P., Men, A., Chen, M., & Yang, B. (2009). Color-SURF: A surf descriptor with local kernel color histograms. *2009 IEEE International Conference on Network Infrastructure and Digital Content*

[33] Wafy, M., & M., A. (2015). Increase Efficiency of SURF using RGB Color Space. *International Journal of Advanced Computer Science and Applications IJACSA*

[34] Sidibe, D., Montesinos, P., & Janaqi, S. (2007). Matching Local Invariant Features: How Can Contextual Information Help? *2007 14th International Workshop on Systems, Signals and Image Processing and 6th EURASIP Conference Focused on Speech and Image Processing, Multimedia Communications and Services*

[35] Manjunath, B.S., Ohm, J.-R., Vasudevan, V.V., Yamada, A. (2001) MPEG-7 color and texture descriptors. *IEEE Trans. Circuits and Systems for Video Technology 11, 703–715*

[36] Sebe, N., Gevers, T., Weijer, J., & Dijkstra, S. (n.d.). Corner Detectors for Affine Invariant Salient Regions: Is Color Important? *in Proc. CIVR, 2006*, vol. 4071, pp. 61–71

[37] Sebe, N., Gevers, T., Dijkstra, S., & Weije, J. (2006). Evaluation of Intensity and Color Corner Detectors for Affine Invariant Salient Regions. *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*

[38] Matas, J., Chum, O., Urban, M., & Pajdla, T. (2002). Robust Wide Baseline Stereo from Maximally Stable Extremal Regions. *Procedings of the British Machine Vision Conference 2002*

[39] *Forssen, P. (2007*). Maximally Stable Colour Regions for Recognition and Matching*. 2007 IEEE Conference on Computer Vision and Pattern Recognition*

[40] Dalal, N., & Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*

[41] Bileschi, S., & Wolf, L. (2007). Image representations beyond histograms of gradients: The role of Gestalt descriptors. *2007 IEEE Conference on Computer Vision and Pattern Recognition*

[42] G. Takacs, V. Chandrasekhar, N. Gelfand, Y. Xiong, W.-C. Chen, T. Bismpigiannis, R. Grzeszczuk, K. Pulli, and B. Girod (2008). Outdoors Augmented Reality on Mobile Phone using Loxel-Based Visual Feature Organization *ACM International Conference on Multimedia Information Retrieval (MIR)*

[43] N. Snavely, S. M. Seitz, and R. Szeliski (2006). Photo Tourism: Exploring Photo Collections in 3D, *SIGGRAPH Conference Proceedings. New York, NY, USA: ACM Press, 2006, pp. 835–846.*

[44] M. Brown and D. Lowe. (2007). Automatic Panoramic Image Stitching Using Invariant Features, *International Journal of Computer Vision, vol. 74, no. 1, 2007, pp. 59–77. 1*

[45] S. Se, D. Lowe, and J. Little (2007). Vision-Based Global Localization and Mapping for Mobile Robots. *IEEE Transactions on Robotics, vol. 21, no. 3, 2007, pp. 364–375*

[46] G. Takacs, V. Chandrasekhar, B. Girod, and R. Grzeszczuk. (2007) Feature Tracking for Mobile Augmented Reality Using Video Coder Motion Vectors *ISMAR '07: Proceedings of the Sixth IEEE and ACM International Symposium on Mixed and Augmented Reality, 2007*

[47] Huang, J., Kumar, S., Mitra, M., Zhu, W., & Zabih, R. (1997). Image indexing using color correlograms. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition.*

[48] Pass, G., Zabih, R., & Miller, J. (1996). Comparing images using color coherence vectors. Proceedings of the Fourth ACM International Conference on Multimedia - MULTIMEDIA '96.

[49] Geusebroek, J., Boomgaard, R., Smeulders, A., & Geerts, H. (2009). Color invariance. IEEE Transactions on Pattern Analysis and Machine Intelligence IEEE Trans. Pattern Anal. Machine Intell., 1338-1350

[50] Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011). ORB: An efficient alternative to SIFT or SURF. *2011 International Conference on Computer Vision.*

[51] E. Rosten and T. Drummond. (2006) Machine learning for highspeed corner detection. In *European Conference on Computer Vision*, volume 1

[52] M. Calonder, V. Lepetit, C. Strecha, and P. Fua. (2010) Brief: Binary robust independent elementary features. In *In European Conference on Computer Vision*

[53] G.J. Burghouts and J.M. Geusebroek. (2009).Performance Evaluation of Local Color Invariants *Computer Vision and Image Understanding, vol. 113, pp. 48-62, 2009.*