More effective image matching with Scale Invariant Feature Transform

Cosmin Ancuti[*] , Philippe Bekaert[*]
Hasselt University
Expertise Centre for Digital Media
Transnationale Universiteit Limburg- School of Information Technology
Wetenschapspark 2, 3590 Diepenbeek, Belgium

Abstract

Feature matching is based on finding reliable corresponding points in the images. This requires to solve a twofold problem: detecting repeatable feature points and describing them as distinctive as possible. SIFT (Scale Invariant Feature Transform) has been proven to be the most reliable solution to this problem. It combines a scale invariant detector and a very robust descriptor based on gray image gradients. Even if in general the detected SIFT feature points have a repeatability score greater than 40 %, an important proportion of them are not identified as good corresponding points by the SIFT matching procedure. In this paper we introduce a new and effective method that increases the number of valid corresponding points. To improve the distinctness of the original SIFT descriptor the color information is considered. In our method we compute the cross correlation and the histogram intersection between neighbour keypoints regions that has been adapted to scale and rotation. Finally, the experimental results prove that our method outperforms the original matching method.

Keywords: SIFT, descriptor, detector, feature point, scale space, matching, cross correlation, color histogram.

1 Introduction

The matching problem of two images that represents the projection of the same 3D scene that is viewed from different position can be defined by two main statements. The first one reveals the need of detecting the same points independently in both images and is accomplished by an invariant and repeatable detector. The second one emphasizes the necessity of a robust and distinctive descriptor that creates the possibility of correct matching of the detected feature points.

Detecting invariant local feature points manifests increasingly attention in the last two decades due of their potential to solve a wide variety of problems, like texture and object classifying, image retrieval, 3D reconstruction, camera calibration, robot localization etc. In general the feature points (keypoints, interest points) are extracted from regions where the image has significant variation in at least two directions.

Invariance of the keypoints to different transformations (scale, rotation, translation, view angle, illumination variations) represents a crucial problem in the process of matching.

After detection phase, the keypoints need to be described as distinctive as possible. A good descriptor should be robust and distinctive, enabling to transfer compactly the information contained by the neighbor region around the feature points. These properties of descriptors are caused by the information content and by the invariance of the local region.

In general as much as the detected regions are more invariant, the less information is transported by the descriptor. The relation between detector and descriptor is very important and not all the combination between detectors and different descriptors give similar results.

Scale Invariant Feature Transform (SIFT) [Lowe 2004] has been proven to be the most robust combination of detector and descriptor [Mikolajczyk et al. 2004]. SIFT keypoints are identified in the scale space using the Difference of Gaussian (DoG) a close approximation of the LoG (Laplacian of Gaussian).

The SIFT keypoints are described by an orientation histogram built on the gray image gradient magnitudes and orientations sampled around the feature point location. Due to the fact that the SIFT is designed mainly for gray image, it may be vulnerable in case of the color images.

Color provides powerful information in the classification process that in general simplifies the procedure of object identification in a considered scene. Human visual system can discern thousands of color values, compared to only about tens of gray intensities. Considering images from the Figure 1, one may observe that the extracted keypoints represent closely the same gray geometric information. Adding the color information reveals that the centered regions of the keypoints describe completely different structures.

To overcome this drawback of SIFT, in our method we use the color histogram [Swain et al. 1991] as an additional information, in order to increase the efficiency of matching process. Color histograms, extensively used in object recognition problems, are explored in this paper to increase the distinctness of the regions around the extracted keypoints.

The signature of SIFT is a vector (mainly of 128 length) that comprises the information of computed gradient orientation histogram of the keypoints neighbor patches. For matching, a simple Euclidian distance is performed between keypoints descriptors. The good correspondences are identified only in the case when the first best matched distance is less than 0.8 times than the distance to the second best matched.

Even if the detected keypoints in DoG space have in general a repeatability score grater than 40 % (depending on the transformation complexity of images) [Mikolajczyk et al. 2004] , the matching approach used in SIFT (ambiguity rejection method based on the Euclidean distance) is not able to match an important amount of repeatable extracted keypoints.

In this paper we introduce an effective method to increase the number of correct matched keypoints extracted with the SIFT operator .The approach computes the normalized cross correlation (NCC) between the regions centered on the SIFT keypoint.

NCC is a classical measure but has been used efficiently only in the small baseline case (when transformations between images is minimal). In addition, because of the high computation time, this measure is not so attractive in the process of matching. For our problem (where rotation, scale and small affine transformation affects images) to find correct correspondences the NCC is adapted to scale and rotation based on the computed characteristic scales and prominent orientations of the feature points.

----

[*]e-mail: {Cosmin.Ancuti, Philippe.Bekaert}@uhasselt.be

Moreover, to make it computational efficient, only a very reduced part of the feature points (5% of the extracted feature points) are compared with the adapted NCC measure.

Our method is relative straightforward to implement and the experimental results show that the new procedure outperforms the original one.

Overview: The paper is organized as follows: in the following section the related work is presented. The next section contains a briefly review of the SIFT algorithm. In section 4 we present how the new matching approach is implemented. Finally, in the last two sections the experimental results and the conclusions are summarized.

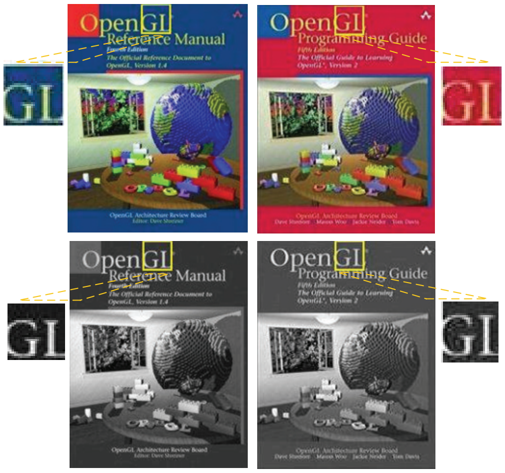

Figure 1. The selected gray regions look very similar when the color information is discarded.

2 Related work

In the last decades a lot of techniques have been applied in order to obtain more distinctive and more robust descriptors. A first class is the distribution-based descriptors. One representation is the spin image introduced in [Johnson et al. 1997] in the context of object recognition. The spin descriptor has been adapted to images in [Lazebnik et al. 2003]. Another important class is based on the frequency content of the image. The Fourier transform cannot represent distinctive local regions because the spatial relation between points is not enough explicit. The Gabor transform [Gabor 1946] is maybe the most representative but this technique requires a large number of Gabor filter to capture the signal disparity.

Steerable filters introduced by Freeman [Freeman et al. 1991] guides the derivatives in particular direction and are obtained by convolution with Gaussian derivatives. Another technique is based on the generalized moment invariants [Van Gool et al. 1996]. The moments can be easy to compute but they are sensitive to geometric and photometric distortions.

Scale invariant feature transform SIFT has been introduced by [Lowe 1999] and has been proved to be the most robust and distinctive among the other local invariant descriptors taking into consideration different geometric changes [Mikolajczyk et al. 2004]. Despite of its impressive capacity to distinct similar local features, it does not take into consideration the color information, being developed mainly for gray images.

However, recently a color version of SIFT has been introduced in vision literature. CSIFT [Hackim et al. 2006] combines SIFT with color invariance introduced by [Geusebroek et al. 2001]. Even if they use a relative complex mathematical approach the presented results have not improved considerable the original version. Moreover CSIFT has been tested only on the images affected by the variation of the illumination (more complex transformation like scale and rotation are ignored).

Our method also used color but in a different way and relative more easy to implement. After adapting the neighbor patches to rotation and scale invariance the color histogram of the regions combined with NCC computed for a part of the feature points is used as additional information to increase the distinctness of the SIFT descriptor.

Due to the high popularity it is not surprising that a number of SIFT variants have been developed meanwhile. One of them is PCA-SIFT [Ke et al. 2004] that apply the Principal Component Analysis to the normalized gradient patch in order to reduce the size of the descriptor vector from 128 to 36.

Gradient location and orientation histogram GLOH [Mikolajczyk et al. 2004] is an extension of the SIFT. GLOH computes SIFT descriptor for a log-polar location grid with three bins in radial direction and then descriptor size is reduced with the PCA.

Speed Up Robust Features (SURF) [Bay et al. 2006] uses Hessian matrix and relies on the integral images to increase the computation efficiency for extracting feature points. Feature points are described by the responses of the Haar-wavelets responses within the interest point neighborhood. Even if it is faster then SIFT, In general the number of good corresponding points matched by SURF is lower than SIFT.

For detecting feature point SIFT is based on the DoG, a close approximation of the LoG. The LoG has been used by Lindeberg [Lindeberg 1998] and [Lindeberg 1999] to find blob like feature points. More recently [Mikolajczyk et al. 2003] the Harris Laplacian detector has been introduced. As its name disclosed, this operator is based on Harris detector [Harris et al. 1988] that is used for 2D localization of features and than using the multi scale representation the scale invariant features are extracted.

All this detectors are scale invariant and are based on the automatic scale detection principle which represents the base for the majority scale invariant detectors. The Harris Laplacian seems to have the highest repeatability score among the other invariant detectors.

Color is a very important property of images that make them more distinctness. Several color system have been proposed during the time[Wyszecki et al. 1982]. Various search system schemes based on the color has been proposed. The more used methods include color histograms [Swain et al. 1991], invariant moments [Geusebroek et al. 2001], color co-occurrence matrix [Seong-O Shim et al. 2003].

A common method to describe the texture features used mainly for image retrieval, object recognition and classification is the color histograms [Swain et al. 1991].

Color histograms are efficient and relative easy to compute being robust to translation and rotation transformations and quasi invariant to scale and view angle transformations. Even if is widely used in image retrieval and indexing problems, color histograms have a major drawback: do not take the spatial information of pixels into consideration, thus very different images can have similar color distributions.

To overcome this problem in our approach is used the invariant to scale and to rotation normalized cross correlation. To make the method computational efficient only a very reduced part of the extracted keypoints are compared with NCC measure.

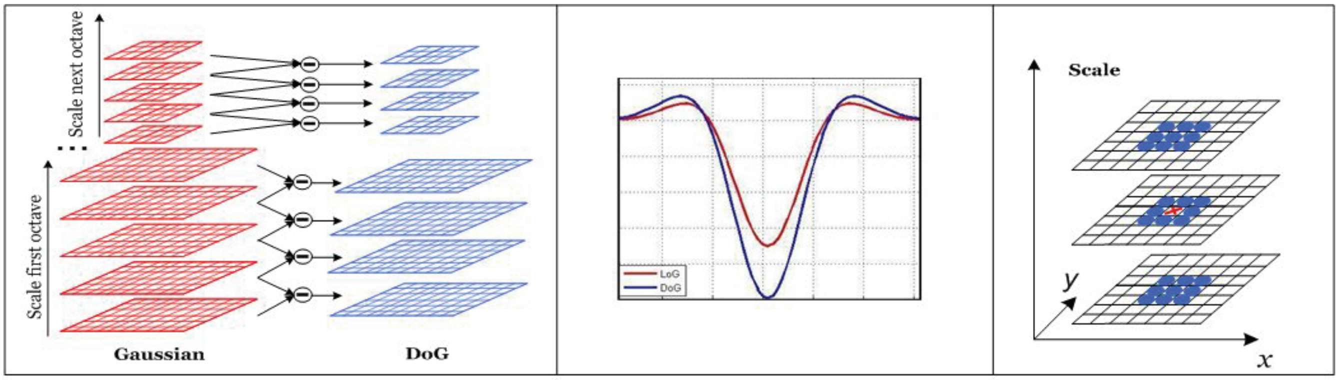

Figure2. a)Construction of the Difference of Gaussian(DoG) scale space; b) similarity between DoG and LoG (Laplacian of Gaussian); c)search for extrema in DoG scale space images

3 Scale Invariant Feature Transform (SIFT)

Finding extrema points: In the first stage the image pyramid is built convolving a given image *I(x,y)* with a Gaussian kernel *G(x,y,σ)*. The pyramid can be seen as a 3-dimensional space where the image represents the first two dimensions and the standard deviation *σ* of the Gaussian kernel is the 3rd dimension.
Scale space can be represented by :

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{1}$$

where the Gaussian kernel is :

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

The standard deviation of the Gaussian kernel controls the level of the image blurring.
Even if in the previous work [Lindeberg 1998][Mikolajczyk et al.2004] scale normalized Laplacian of Gaussian (LoG) is used for true scale invariance the SIFT keypoints are extracted searching for the extrema in Difference of Gaussian (DoG) space. DoG represents a close approximation of LoG and mathematical relation between them can be found solving the heat diffusion equation. DoG is computed subtracting neighbor scales. The construction of the DoG space is represented in Figure 2**.** Image *I(x,y)* is incrementally convolved with Gaussian kernel creating scale levels separated by a constant ***k***. Then the image is down sampled with a factor of 2, hence each octave has half of the previous octave size. Octaves of the scale space are divided in ***s*** intervals where the relation between ***k*** and ***s*** is *k*=$2^{1/s}$:
Every keypoint points is identified comparing every sample point with its 8 neighbor pixels from the same scale image level and with the 9 pixels from both adjacent scale levels. The sample point is selected only in case when its value is greater or lesser than all the 26 neighbor points (see Error! Reference source not found.).

Feature localization: In the next stage all selected candidate points are verified for the stability. In the initial implementation [Lowe 1999] the candidate keypoints are identified at the location and scale of the central sample point. Following the results of [Brown et al. 2003] an important improvement for stability has been achieved fitting a 3D quadratic function to the local sample points in order to determine the interpolated location of the maximum. As a function is used the quadratic Taylor expansion of the DoG function *D(x,y,σ)* centered on the sample point *x*=$(x,y,\sigma)^T$:

$$D(x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 D}{\partial x^2} x \tag{3}$$

The following equation is solved in order to set the offset:

$$\hat{x} = -\frac{\delta^2 D^{-1}}{\delta x^2} \frac{\delta D}{\delta x} \tag{4}$$

where $\hat{x}$ is the offset vector ,and *D* derivatives are computed using local pixel differences around the point location.
Lowe find very useful to filter out the unstable points with low contrast and also the points with strong response along edges. Therefore an approach inspired from Harris feature points detection [Harris et al. 1988] is used. Keypoints that have a small value of the principal curvatures will be rejected. The principal curvature is estimated computing the trace and determinant of the Hessian matrix H of the *D* function.

Features descriptor: The SIFT signature is based on the image gradient magnitudes and orientations sampled around the feature point location after the characteristic scale of the feature point is used to select the level of the image in the scale space. The original implementation computes a 4x4 grid of orientation histogram built on the 4x4 sub-region of the feature point from a 16x16 region centered on the feature point location. Each histogram has 8 bins corresponding for every 45º(see Figure 3).
To avoid boundary influence, trilinear interpolation is used to distribute the value of each gradient sample into adjacent histogram bins. In order to have invariance to rotation a dominant orientation need to be assigned to every feature point. First, the image gradient orientation and magnitude of every sample point is computed.

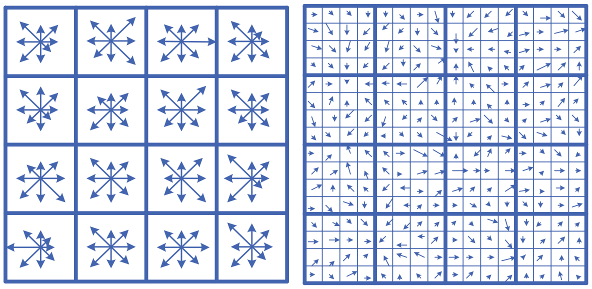

Figure 3. a)16x16 patch gradients ;b) 4x4 descriptor(128 vector values)

Then, a 36 bin (one bin for every 10º) orientation histogram is constructed using the gradient orientation in the neighbor region around the sample point. Every points from that region contributes to the histogram bins by its gradient orientation weighted by its gradient magnitude and by a Gaussian-weighted circular window.

4 More effective matching images with SIFT

Even if in general the detected SIFT feature points have a repeatability score greater than 40 % important proportion of them are not identified as good corresponding points in the SIFT matching procedure (depends on the complexity of the applied image transformation).

Matching SIFT feature points is performed computing the Euclidean distance between all the 128 length descriptor vectors of the keypoints. Unfortunately only the minimum distance criterion is not enough to identify good corresponding points in the images.

Lowe observed that a better matching technique is to compare the distance of the first best matched to that of the second-best matched feature point.

In this paper we focus on the problem to identify the mismatched SIFT points that are still repeatable detected. To make SIFT descriptor more distinctive we use the color information (that is completely ignored in the original version) by comparing the color histogram of the region centered on the extracted feature points. Additionally, the NCC is computed between a limited feature points color neighbor regions. To avoid the problem of memory and time consuming our method computes the NCC only a for a very reduced parts of the keypoints (only about 5% of the total feature points).

4.1 Color as an additional discriminative measure

Color received a significant attention in computer vision literature due of its powerful discrimination properties.

Color space is in general specified in three dimensions which makes it higher distinctive comparing with the gray images (one dimension representation of the color).

The gray value system (G) is computed from the *RGB* information that corresponds to red, green and blue images provided by a CCD color camera. The intensity (gray) is not perceived uniform and is highly vulnerable to the image conditions:

$$G = 0.299\ R + 0.587\ G + 0.144\ B \qquad (5)$$

Several color space representation have been proposed during the years [Wyszecki et al. 1982]. The most known is the *RGB* representation. *RGB* is a color system based on the primary colors red (*R*=700nm), green (*G*=546.1nm) and blue (*B*=435.8nm). *XYS* system inherits the *RGB* and is based on the three primaries colors *X,Y* and *Z* that cannot perceive or produced by human eyes. A different approach is *L*a*b** color space representation where *L** includes the luminance information, *a** the combined red-green components and *b** reflects the yellow-blue content. The NTSC (National Television System Committee) introduced the *YUV* color system that conveys separately the luminance of the color(*Y*), the hue(*I*) and the saturation(*Q*) of the color. The corresponding color system for Europe is *YUV* . A more closely related to human color perception than *RGB* color representation is *HIS* (Hue Saturation and Intensity) system which is perceptually more intuitive but not perceptually uniform.

In this paper we use the *rgb* color system. The *rgb* representation of the color is derived from the *RGB* system dividing the *R,G,B* values by their sum:

$$r = \frac{R}{(R+G+B)} \quad g = \frac{G}{(R+G+B)} \quad b = \frac{B}{(R+G+B)} \qquad (6)$$

Comparing with *RGB* the *rgb* presents several advantages being not sensitive to surface orientation, illumination direction and illumination intensity [Geusebroek et al 2001].This is because the normalized system *(rgb)* depends only to the ratio of the *RGB* coordinates.

A traditional method to describe the texture is color histogram [Swain et al. 1991]. Color histograms are efficient and relative easy to compute being robust to translation and rotation transformations and quasi invariant to scale and view angle transformations. Even if is widely used in image retrieval and indexing problems, color histograms have a major drawback: do not take the spatial information of pixels into consideration, thus very different images can have similar color distributions. A simple method to increase the distinctness of color histograms is to divide the image into sub-patches and calculate a histogram for each of them. Unfortunately, increasing the number of sub-patches, affects seriously the memory and computational time.

An alternative that eliminates this drawback is the color correlogram [Jing Huang et al. 1997]. Besides of the pixel color distribution information the color correlograms transport also the information of spatial correlation of the neighbor pixels. An equivalent measure that also keeps track of the number of colored pixel pairs that occurs in space images at certain distance is the color co-occurrence matrix [Peng Chang et al. 1999].

Comparing with the classical color histogram the color correlogram or color co-occurrence matrix provides more distinctness results but on the other hand the high computation time may diminish the attractiveness of these methods.

Our implementation is based on the color histograms [Swain et al. 1991] combined with the adapted cross correlation.

Gray level histograms contain the distribution of the image intensities in an image. A more powerful representation is the color histograms that capture the joint probabilities of the intensities of the color channels. The expression of color histogram is:

$$h_{r,g,b}(x,y,z) = N \cdot prob(x = r, y = g, z = b) \qquad (7)$$

where the color channels are *r,g,b* and *N* represents the number of pixels of the considered image.

The similarity of two color histogram can be compared by several distances [Puzicha et al. 1999]. In this paper the Minkowski-form distance L_2 is used. Considering two histograms $h_1$and h_2, the formal equation of the distance (known also as histogram intersection) is:

$$d(h_1, h_2) = \left(\sum_{r,g,b} | h_1 - h_2 |^2 \right)^{1/2} \qquad (8)$$

As discussed before, the color histograms suffer due of not considering the relation between adjacent pixels. To put out this disadvantage we compute the normalized cross correlation (NCC) between the regions centered on the feature points. Because NCC is computational expensive we avoid the computation of the whole sets of feature points based on a reliable and efficient method described in section 4.3.

4.2 Scale and rotation invariant Normalized Cross Correlation

Normalized cross correlation (NCC) is a classical measure to match images. It takes into consideration the simplest descriptor: the vector of patch pixels. Unfortunately it gives reliable results only for the small baseline case. When significant transformation (scale, rotation, translation, illumination, view angle) affects images this measure cannot find good correspondences.

Let I_1 and I_2 the intensity values of two regions (windows) W_1 and W_2 of the same size *(2r+1)*x*(2r+1)* centered on 2 keypoints p_1 and p_2 from a considered pair of images. The normalized cross correlation between the windows is computed with the following mathematical expression:

$$NCC(p_1,p_2)=\frac{\sum_{-r}^{r}\sum_{-r}^{r}\left[(I_1-\bar{I}_1)(I_2-\bar{I}_2)\right]}{\sqrt{\sum_{-r}^{r}\sum_{-r}^{r}(I_1-\bar{I}_1)^2}\sqrt{\sum_{-r}^{r}\sum_{-r}^{r}(I_2-\bar{I}_2)^2}} \qquad (9)$$

$\bar{I}_1, \bar{I}_2$ represent the means of intensity values of considered windows.

A similar measure is the sum of squared differences (SSD) that measures the squared Euclidean distances between I_1 and I_2. The relation between SSD and NCC is immediately and the properties are very similar. The values of NCC lies between 0 and 1, with 1 indicates a perfect matching. Even if SSD is computational more inexpensive the NCC is preferred in many situations because it is invariant to linear brightness and contrast variations.

Our approach is based on the NCC. The method uses the NCC to compare the neighbor patches as an additional constraint for the selected feature points. To reduce the computation time, only a very reduced part of the feature points (5% of the extracted feature points) is compared with NCC.

However, in our case where important transformation affect images the NCC should be adapted. Thus, the region vectors centered at every keypoint are adapted in order that corresponding patches to be invariant for rotation and scale transformations

Characteristic scale: One of the most challenging problem is the invariance to scale modifications. The scale space theory [Lindeberg 1998] addresses this problem and has been proved an reliable solution to estimate scale ratio between images.

Real world objects appear in different ways depending of the selected observation scale. Scale space representation is defined as a solution to the diffusion equation which is equivalent with the convolution of the signal with Gaussian kernel.

It has been proved [Koenderink 1984] [Lindeberg 1999] that under a variety of reasonable assumption Gaussian is the unique kernel for generating a scale-space. The neurophysiologic studies in [Young et al. 2001] shown that mammalian retina and visual cortex present sensitive fields that can be properly modeled by Gaussian derivatives up to order four.

The scale space is constructed by convolving the initial image with Gaussian kernels that has different values of the standard deviation.

To estimate the ratio scale, the automatic scale selection principle is used. The principle [Lindeberg 1998] is built on the existent relation between images at different resolution levels. Its applicability is extremely important due to the fact that in general images contains sharp and diffuse features and it is almost impossible to identify all kinds of features at the same scale level. Lindeberg [Lindeberg 1998] postulates that in absence of other evidence, the selected scale (characteristic scale) is the scale where the function of some combinations of normalized derivatives attain a local maximum. The idea behind characteristic scale is borrowed from physics and estimates the characteristic length of corresponding image structure. The characteristic scale is independent by the image scale and the ratio between selected scales of two extrema that represent the same image feature is the same with the ratio between image scales.

Dominant orientation: The rotation invariance problem is handle by computing a dominant orientation of the extracted feature points neighbor regions. The approach is inspired by [Thurnhofer et al. 1996] and is used also in SIFT descriptor [Lowe 2004]. To determine a prominent orientation for every keypoint the gradient magnitude and orientation is precomputed at every level of scale using pixel differences. Let be δ_x and δ_y the finite differences across *x* and *y* directions for a considered pixel.

The magnitude *m* and orientation φ can be calculated using the following expressions:

$$m(x,y)=\sqrt{\delta_x(x,y)^2+\delta_y(x,y)^2}$$
$$\varphi(x,y)=\arctan\left(\frac{\delta_y(x,y)}{\delta_x(x,y)}\right) \qquad (10)$$

The gradient orientation histogram is computed in the neighborhood of every keypoint. In our experiments the histogram is composed by 36 bins with every bin covering 10 degree range of orientations.

The number of bins can vary and represents a tradeoff between computation time and accuracy of the final results. The contribution of each neighbor pixel is weighted by the gradient magnitude and a Gaussian window with σ that is 1.5 times of the respective characteristic scale. Dominant orientation of keypoints is determined by the highest peak of the histogram.

Due to the fact that the characteristic scale and dominant orientation is an approximate approach, in our method the ratio scale and difference in orientation of considered images are estimated. Besides of our practical experiments this approach is motivated also by two existing problems. First the SIFT keypoints are extracted with DoG which is only an approximation of the LoG(that has been proved in [Mikolajczyk 2003] and [Lindeberg 1999] to be the most reliable derivative function for scale space). Second, in some cases, because more than 15% of the extracted SIFT keypoints have assigned multiple orientations may cause confusions.

Our approach is based on the computed characteristic scale and dominant orientation of the keypoints. The estimated ratio scale between images is computed using the values of the best matched 80% SIFT keypoints characteristics scales. A similar procedure is used to estimate the difference in orientation of considered images relying to the assigned dominant orientation of the 80 % best matched keypoints.

Finally, in order to compute the NCC, the patches around keypoints from the first image are rotated according to the estimated difference in orientation between images. The same principle is used to estimate the ratio scale.

According to the automatic scale selection principle [Lindeberg 1999] the sizes of the centered keypoints regions of two corresponding keypoints are proportional with the ratio scale of the images.

Therefore, the image with higher computed value of characteristic scale will have a larger region around keypoints. Considering two points in the images, to be able to compute the NCC of the neighbor regions the keypoint with higher characteristic scale is interpolated to the dimension of the smaller region.

4.3 Algorithm

Our approach compares the information of the neighbour regions centered on the extracted keypoints. First, the keypoints are tried to be matched using the procedure based on the ambiguity rejection [Lowe 2004].

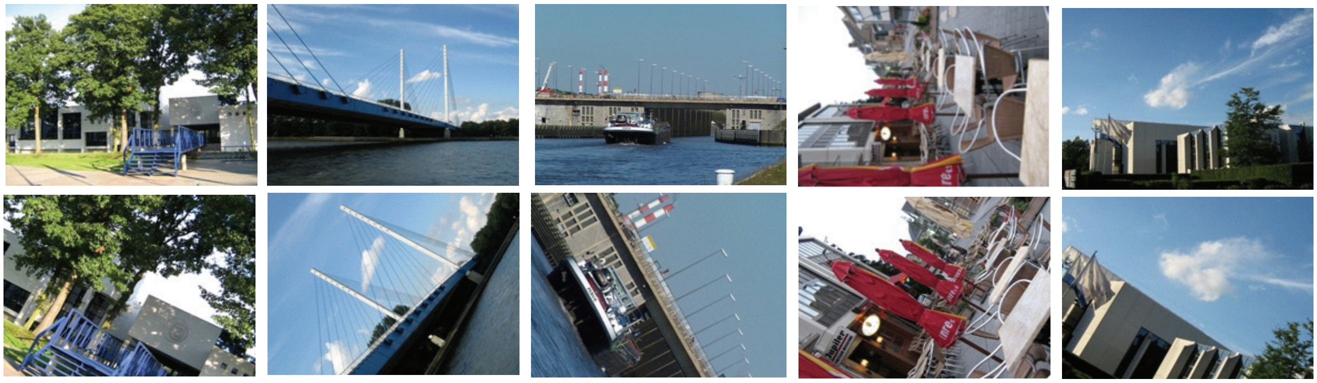

Figure 4.Pairs of testing images

This procedure selects good matches only if the ratio of the closest match to the second one is lesser than 0.8. For the rest of the unmatched keypoints, we increase the discriminative information of their neighbour regions by computing the color histograms and attempting to match them as was detailed in the previous subsections. The results of this paper have been generated using a size of 16x16 pixels of the neighbour patches centered on the SIFT keypoints. In our experiments, taken into consideration the size of the keypoints regions, we quantized the *rgb* color space using 5 bits (32 levels of colors). We observed that this level of quantization is a good compromise between computation and matching performance.

The algorithm can be resumed in the following pseudo code lines:

```
for i=1 to no_keypoints_img1
    for j=1 to no_keypoints_img2
        compute distance;
        if (dist_first_best_matched/dist_second_bestmatched>0.8)
                SIFT_match(Ambiguity Rejection);
        else
                find_first_best_matches;
                no_best_matches=5% of keypoints;
                for k=1 to no_best_matches
                        compute_dist_Color_Histograms(i,k);
                        compute_addapted_NCC(i,k);
                        if (max_NCC== NCC(i,k)) &&
                           (min(dist_Color_Hist)==dist_Color_Hist(i,k))
                                Select(i,k)= Good Match;
                        end
        end
```

The method proposed in this paper is focused only on the keypoints that are not matched by SIFT approach. Additional good matches are identified in the keypoints that have minimum distance between their vector descriptors. Only for the best 5% of the keypoints are computed the color histograms and adapted NCC. A good match is selected in case when for a keypoint both the minimum of the color histogram and maximum of NCC is attained.

5 Experimental results

To validate our method a number of real images have been tested. The color images have been taken with an ordinary digital camera(Kodak Z740) and transformations like rotation and scale have been synthetically simulated. A part of the tested images are presented in the Figure 4. The number of valid correspondence are identified based on the computed homography (rates a unique relation between points from two images), estimated using the algorithm from [Hartley and Zisermann 2003]. The analyze is limited only to the planar scenes. Supposing that x_1 and x_2 represent the projected points of a 3D space point the relation between them is given by the homography expression:

$$x_2 = H_{12}x_1 \tag{11}$$

Whether the homography matrix is known the repeatability criterion [Schmid et al. 2000] determines the number of good correspondences. The score of repeatability for a pair of images represents the ratio between the number of point-to-point correspondences and the minimum number of points detected in images. Note that only the points located in the same scene part visible in both of the images are considered. For evaluation, we compare the performance of our method with the original SIFT. The SIFT is computed for the gray version of the considered color images applying the equation 5 transformation.

To evaluate our approach the ratio between the number of good detected matched points and the number of maximum possible matches of two images is computed.

The image groups from the Figure 4 are evaluated and the results are presented in the following graphic.

As can be seen our approach performs better than original SIFT, finding in general with approximately 7% more valid matches. Additionally, in the figure 6 and figure 7 (next pages) more examples with detailed information are examined.

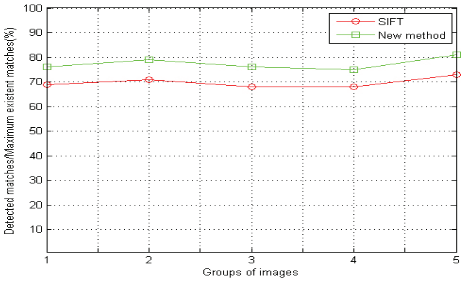

Figure 5.Comparative matching results

6 Conclusions and feature work

In this paper we presented a new method that increase the efficiency of matching images using SIFT operator. Because the original SIFT do not use the color information being mainly designed for gray images we introduce the color as additional discriminative information. Therefore the normalized cross correlation between neighbor patches centered on the feature points is computed. NCC is a very efficient measure only for the small baseline case. When more complex transformation affects images the NCC may be adapted the compared patches to scale and orientation after the ratio scale and difference of orientation is computed. Additionally, to increase the distinctness of the SIFT descriptor the color histogram of considered feature points regions are compared using the histogram intersection measure.

The experimental results show that our approach improves the number of good matched feature points using the SIFT detector/descriptor and also the stability of extracted keypoints.

In future work we would like to focus more on the color in order to explore more efficient the color space properties. Moreover we would like to extend our approach for more important affine transformations.

Acknowledgements

The authors acknowledge financial support on a structural basis from the ERDF (European Regional Development Fund) and the Flemish Government and the Flemish Interdisciplinary institute for BroadBand Technology(IBBT). The first author also acknowledges financial support by a research grant from the Hasselt university research fund.

References

Bay, H. and Tuytelaars, T. and Van Gool, L.J., 2006. *SURF: Speeded Up Robust Features*, ECCV06 pp. 404-417

Freeman, W. and Adelson, E. 1991. *The Design and Use of Steerable Filters*. IEEE Transactions on Pattern Analysis and Machine Intelligence,vol. 13, no. 9, pp. 891-906.

Gabor, D. 1946. *Theory of Communication*, J. Inst. Electr. Engineering London 93, vol. 3, no. 93, pp. 429-457.

Richard Hartley, Andrew Zisserman, *Multiple View Geometry in Computer Vision*, Publisher Cambridge University Press; 2Rev Ed edition (30 Nov 2003)

Geusebroek, J. M., Burghouts, G. J.and Smeulders, A. W. M. 2005. *The Amsterdam Library of Object Images*. Int. J. Comput. Vision, 61(1):103–112.

Geusebroek, J. M., van den Boomgaard, R. Smeulders, A. W. M. and Geerts, H. 2001. *Color Invariance*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(12):1338–1350.

Abdel-Hakim, A.E.; Farag, A.A., 2006. *CSIFT: A SIFT Descriptor with Color Invariant Characteristics* , In Proceedings of Computer Vision and Pattern Recognition Conference, pp. 1978- 1983.

Harris, C.G. and Stephens, M. 1988. *A Combined Corner and Edge Detector*, In Proceedings of Fourth Alvey Vision Conference, vol. 18, pp.147-151.

Jing Huang Kumar, Mitra, S.R., Wei-Jing Zhu, Zabih, R. 1997. *Image Indexing Using Color Correlograms*. In Proceedings of Conference IEEE Computer Vision and Pattern Recognition, pp. 768-762.

Johnson, A. and Hebert, M. 1997. *Object Recognition by Matching Oriented Points.* In Proceedings of Computer Vision and Pattern Recognition Conference, pp. 684-689.

Ke, Y. and Sukthankar, R. 2004. *PCA-SIFT: A More Distinctive Representation for Local Image Descriptors*. In Proceedings of Conference IEEE Computer Vision and Pattern Recognition, vol. 2, pp. 506-513.

Koenderink, J.J. 1984. *The structure of images*. Journal Biological Cybernetics, vol. 50, pp. 363-370.

Young, R. A. and Lesperance, R. M. 2001. *The Gaussian Derivative Model For Spatial-Temporal Vision: II. Cortical Data.* Spatial Vision, 14(3-4): pp. 321-389.

Lazebnik, S., Schmid, C. and Ponce, J. 2003. *Sparse Texture Representation Using Affine-Invariant Neighborhoods.* In Proceedings of Computer Vision and Pattern Recognition, pp. 319-324.

Lindeberg, T. 1998. *Feature Detection with Automatic Scale Selection,* International Journal of Computer Vision, vol. 30, no. 2, pp. 77-116.

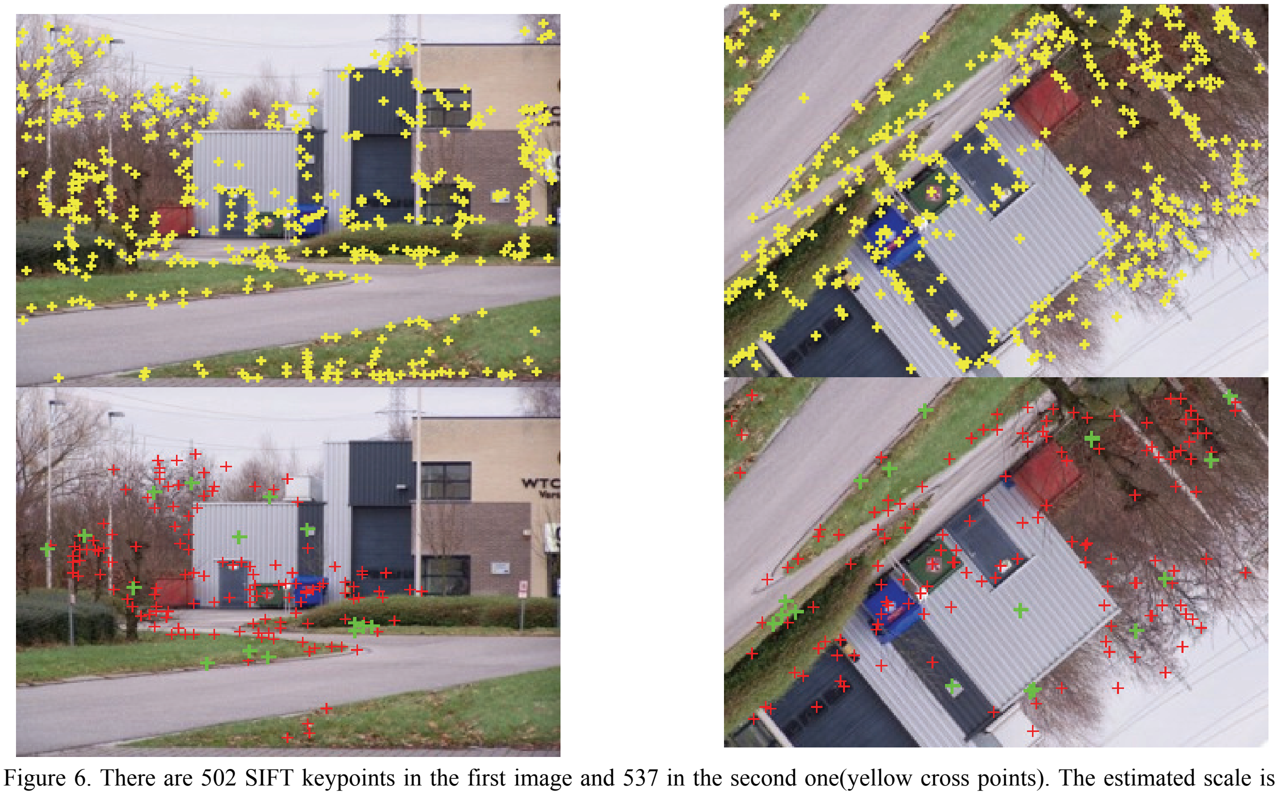

Figure 6. There are 502 SIFT keypoints in the first image and 537 in the second one(yellow cross points). The estimated scale is 1.48 and estimated difference in orientation is 220º. The SIFT algorithm identifies 166 matches (red cross points)and with our approach additional 15 good matched points are identified(green cross points).

Lindeberg, T. 1999. *Methods for Automatic Selection*. Handbook on Computer Vision and Applications, volume 2, Academic Press, Boston, pp. 239-274.

Lowe, D. 2004. *Distinctive Image Features from Scale-Invariant Keypoints*, International Journal of Computer Vision, vol. 2, no. 60, pp. 91-110.

Lowe, D. 1999. *Object Recognition from Local Scale-Invariant Features*, In Proceedings of Seventh International Conference on Computer Vision, pp. 1150-1157.

Mikolajczyk, K. and Schmid, C. 2003. *A Performance Evaluation of Local Descriptors*. International Conference on Computer Vision and Pattern Recognition, pp. 257-263.

Mikolajczyk, K. and Schmid, C. 2004. *Scale and Affine Invariant Interest Point Detectors*. International Journal of Computer Vision, vol. 1, no. 60, pp. 63-86.

Puzicha, J., Rubner, Y., Tomasi, C. and Buhmann, J. 1999. *Empirical evaluation of dissimilarity measures for color and texture.* In Proceedings of the Seventh IEEE International Conference on Computer Vision , vol 2, pp. 1165-1172.

Peng Chang, Krumm, J. 1999.*Object recognition with color cooccurrence histograms* .In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition vol. 2, pp. 504.

Schmid, C., Mohr, R., and Bauckhage, C. 2000. *Evaluation of Interest Point Detectors*. International Journal of Computer Vision, 37(2): pp. 151–172.

Seong-O Shim, Tae-Sun Choi, 2003. *Image Indexing by Modified Color Co-occurrence Matrix*. IEEE International Conference on Image Processing, vol. 3, pp. 493-496.

Swain, M. and Ballard, D. 1991. *Color indexing*: International Journal of Computer Vision, vol. 7(1), pp.11-32.

Thurnhofer, S., Mitra, S. 1996. *Edge-enhanced image zooming.* Optical Engineering, 35(07): pp.1862–1870. Brian J. Thompson; Ed.

Van Gool, L., Moons,T. and Ungureanu, D. 1996. *Affine/Photometric Invariants for Planar Intensity Patterns*. In Proceedings of the Fourth European Conference on Computer Vision, vol. 1, pp. 642-651.

Wyszecki, G. and Stiles, W. S. 1982. *Color Science: Concepts and Methods, Quantitative Data and Formulae*. John Wiley & sons, second edition.

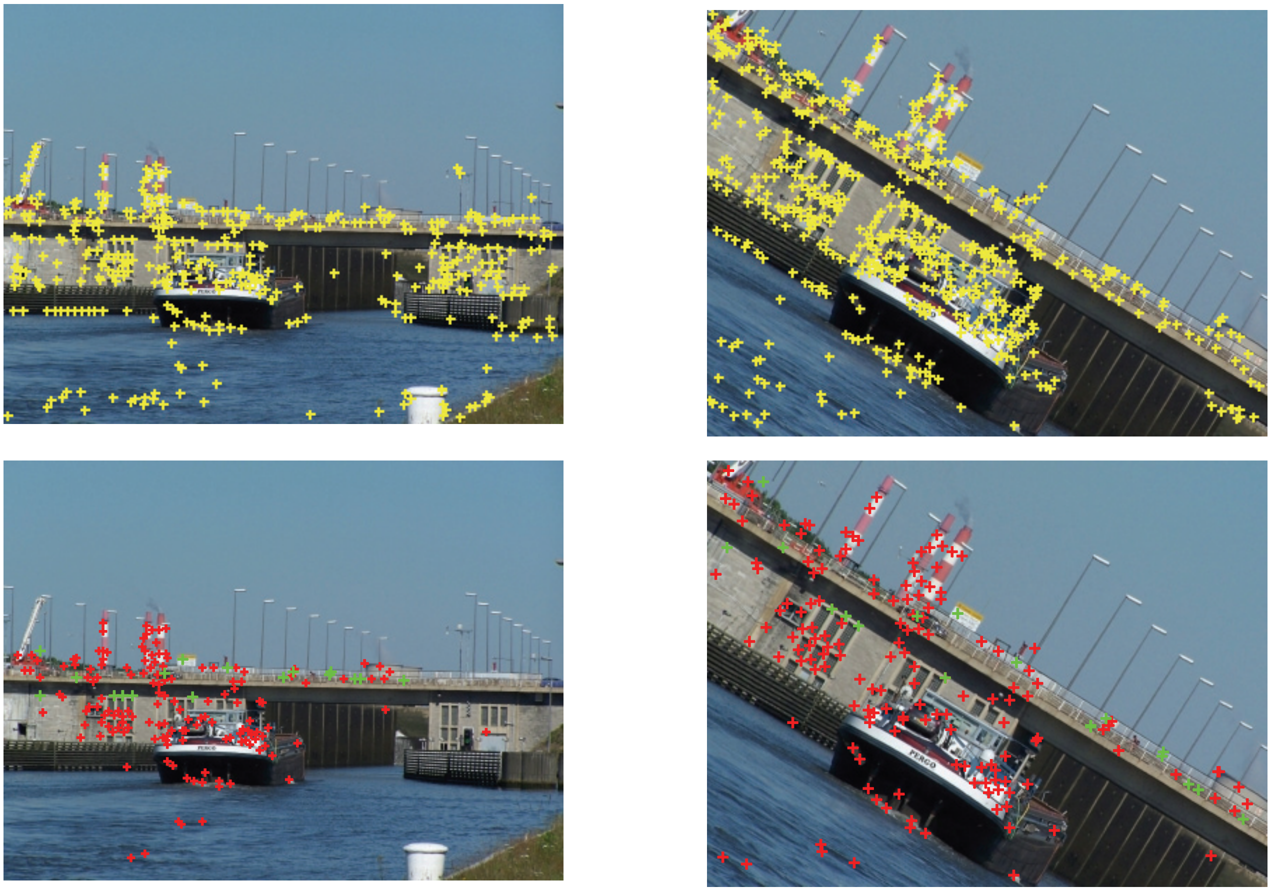

Figure 7. In the first image are identified 698 SIFT keypoints and 483 in the second one (yellow cross points). The estimated scale is 1.63 and estimated difference in orientation is 30º. The SIFT algorithm identifies 190 matches (red cross points) and with our approach additional 17 good matched points are identified (green cross points).