Feature Detection with an Improved Anisotropic Filter

Mohamed Gobara and David Suter

Department of Electrical and Computer Systems Engineering,
Monash University, Clayton, 3800 Victoria, Australia
{osman.gobara, d.suter}@eng.monash.edu.au

Abstract. The problem of detecting local image features that are invariant to scale, orientation, illumination and viewpoint changes is a critical issue in many computer vision applications. The challenges involve localizing the image features accurately in the spatial and frequency domains and describing them with a stable analytical representation. In this paper we address these two issues by proposing a new non-linear scale-space implementation that improves the localization accuracy of the SIFT [3] local features. Furthermore we propose a simple adjustment to the standard SIFT descriptor and show that the modified version is more robust to affine changes.

1 Introduction

Interest point detection is a key issue in many computer vision applications including motion tracking, object recognition and 3D reconstruction. An interest point is any point in the image that is characterized by distinctive neighboring features. This includes L-corners, T-junctions, Y-junctions and highly textured areas. The detection of interest points is a dual stage process, (a) localization and (b) representation. In the localization phase we detect the position and the scale of each interest point and in the representation phase we use an analytical model to describe the local shape or pattern at each interest point. The goodness of a model (i.e. also known as a local descriptor) is measured in terms of its degree of invariance over transformations caused by viewpoint and illumination changes. A good model (i.e. highly invariant descriptor) would identify a local pattern, before and after being transformed, with the same numeric measure.

Schmid and Mohr [1] examined a wide variety of interest point detectors and categorized them, based on their localization criteria, into three main groups: Contour-based, Intensity-based and Parametric-model based methods. The Contour-based methods define interest points either at the intersections of grouped line segments or at the maximum curvature of approximated contours. Intensity-based methods define interest points through the illumination distribution of the neighborhood. In most cases these algorithms are based on the second moment matrix, which is a mathematical measure for the distribution of the local image gradients. Parametric-based methods on the other hand define interest points at regions that fit a predefined analytical intensity model. This paper focuses on a group of Intensity-based detectors [3, 4], which define the interest points as the local peaks of grayvalue derivatives in scale-space. In most cases these detectors are capable of identifying local patterns independent from any scale changes. In this paper we propose a new non-linear

P.J. Narayanan et al. (Eds.): ACCV 2006, LNCS 3852, pp. 643 – 652, 2006.

scale-space representation, which improves the localization accuracy of the aforementioned detectors [3, 4].

In all our experiments we used the SIFT descriptor [3] to define the local patterns at each interest point. Mikolajczyk and Schmid [7] proved that the SIFT descriptor is more robust to affine changes than many other descriptors including steerable filters [8], differential invariants [2, 9], complex filters [11] and moment invariants [10]. We did also use a modified version of the SIFT descriptor which is more distinctive and in many cases leads to a much better matching results.

Overview. Section 2 presents different implementations for the scale-space including a new proposal, which in general uses the non-linear spatial filter of Köthe [6]. Section 3 reviews the main features of the detectors and descriptors used in our tests. Section 4 introduces the evaluation criteria. Section 5 and 6 present the experimental results and the conclusion.

2 Scale-Space Representations

A linear scale-space is defined by the solution of the following diffusion equation;

$$\frac{\partial L(z,s)}{\partial s} = \frac{1}{2}\nabla^2 L(z,s) = \frac{\partial_{xx} L(z,s) + \partial_{yy} L(z,s)}{2} \tag{1}$$

with the initial condition that $L(z,0)$ (i.e. initial scale s=0) is equal to the original image $I(z)$, ∇^2 is the Laplacian kernel and z is the spatial coordinates of the interest point. Equivalently a linear scale-space can be defined by convolving $I(z)$ with the Gaussian kernel G(z,s).

$$G(z,s) = \frac{1}{2\pi\sqrt{s}} e^{-z^2/2s^2} \tag{2}$$

To reduce the amount of smoothing around edges Perona and Malik [5] proposed the use of anisotropic diffusion as a generalization of the linear scale-space representation.

$$\frac{\partial L(z,s)}{\partial s} = \frac{1}{2}\nabla^2 \left(h(z,s)\ \nabla L(z,s) \right) \tag{3}$$

where $h(z, s)$ is defined to be dependant on the image gradient. A possible solution for $h(z, s)$ is presented by eq.4 where k defines the range of gradients in an image and thus controls the amount of smoothing at point z.

$$h(z,s) = e^{-\frac{|\nabla L(z,s)|}{K}}. \tag{4}$$

2.1 Hourglass Representation

Köthe [6] proposed an oriented non-linear spatial filter that looks like an hourglass. The new filter modulates the Gaussian so that it becomes zero at a perpendicular dis-

tance from the local edge direction ϕ_0. The output of the filter at point (x,y) is given by the following equation:

$$h_{\sigma,\rho}(z,\phi,\phi_0) = \frac{1}{N} e^{-\frac{z^2}{2\sigma^2}} e^{-\frac{\tan^2(\phi-\phi_0)}{2\rho^2}} \tag{5}$$

where z and ϕ are the polar coordinates of point (x, y), ρ defines the width of the Hourglass filter, the larger the value of ρ the more the filter tends to become uniform, and N is a normalization factor that sums the weights of the filter to 1. Köthe recommended that ρ should be set to a value between 0.3 and 0.7.

The dimension of the Hourglass scale-space is defined by an initial scale σ_0, final scale σ_f, and a factor k of scale change between successive levels. At each scale level σ a local direction ϕ_o is calculated for each sample point using a simple derivative function. Next the Hourglass kernel is rotated by ϕ_o degrees and applied to the sample point.

3 Experiment Setup

In the following we will review the implementation details of two interest point detectors and two descriptors used in our experimental tests. The detectors are invariant to scale and rotation changes. The descriptors on the other hand are distinctive and relatively robust to common image transformations.

3.1 Interest Point Detectors

The detection scheme in the following two algorithms starts with an appropriate implementation of the scale-space.

SIFT: first, local peaks are selected from a Difference of Gaussian pyramid. A 3D quadratic function is fitted at each local peak and an interest point location is calculated up to a sub-pixel /sub-scale accuracy at the extremum value of this quadratic function. Finally interest points with low contrast values and points located along edges are considered unstable and rejected.

Harris-Laplacian [4]: a scale-space is built for the Harris function using the second moment matrix C(z,s,s$^-$). At each scale-space level s the local peaks of the Harris function are selected as possible interest point candidates. Finally, candidates with the local scale-space maximum of the Laplacian function are identified as interest points.

$$\text{Harris function} = \det(C) - \alpha\,\text{trace}^2(C$$

$$\textit{Where} \quad C(z,s,s^-) = s^2 G(z,s^-) * \begin{bmatrix} L_x^2(z,s) & L_x L_y(z,s) \\ L_x L_y(z,s) & L_y^2(z,s) \end{bmatrix}, \tag{6}$$

L_z and L_y are the gradients along the x and y axis respectively.

3.2 Descriptors

The descriptors used in our tests are: (1) the standard SIFT and (2) a modified version of the SIFT. In the remaining part of this section we will review the design aspects of these two descriptors.

SIFT: A descriptor is calculated for each interest point with a spatial location z and scale s through to the following steps:

1. A dominant orientation angle θ is calculated from the local neighborhood of p, which is defined by a circular region of radius 1.5s. The method of detecting θ is explained in detail in [3].
2. A local window W of size 16x16 is fitted at location z and scale s.
3. A gradient orientation and magnitude are calculated for each sample point that lies within W.
4. To achieve rotation invariance, the coordinates and the gradient orientations of W are rotated by angle $-\theta$.
5. The gradient magnitudes of W are smoothed with a uniform Gaussian kernel of scale k=1.5 the width of W. This step is meant to reduce the effect of sample points that lie away from z as they are considered the most likely affected points with misregistered errors.
6. The local window W is divided into 16 different 4x4 sample regions.
7. The weighted gradient magnitudes of each sample region are summed in an orientation histogram with eight directions as shown in figure.1.

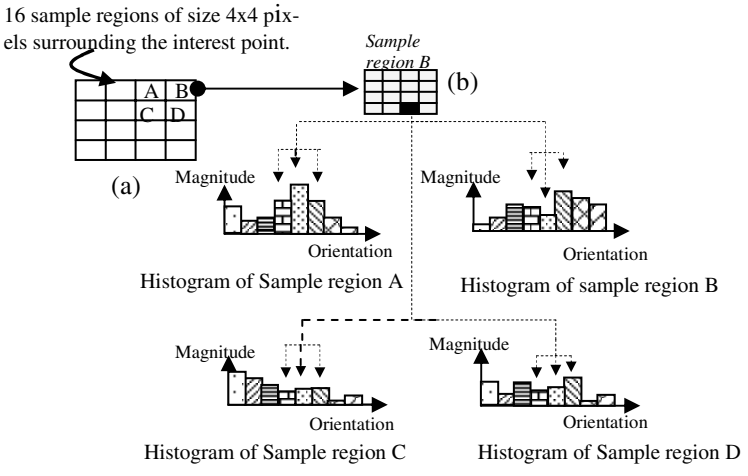

Fig. 1. (a) The neighborhood of the interest point is divided into 16 sample regions. (b) The gradients of each sample region (i.e. as in region B) are accumulated in an orientation histogram with 8 directions and distributed among the histogram bins of neighboring regions (i.e. regions A, C and D) through a tri-linear interpolation.

8. The descriptor is formed from a vector containing the values of all the 8x16=128 orientation histogram bins.
9. To reduce the effects of illumination change the vector elements are normalized to a unit length, then thresholded to values not greater than 0.2 and finally renormalized.

Modified-SIFT: Steps '1' and '5' in the above algorithm are modified and applied for each interest point z with scale s as follows:

- *Step 1*: In the SIFT algorithm the pixels at spatial distances less than $1.5s$ from z are defined as the local neighbors of z while in the modified-SIFT the pixels with both grayvalue and spatial distances less than $1.5s$ are defined as the local neighbors of z.
- *Step 2*: A Gaussian function with scale k is used to weight the gradients of the local neighbors of point z in the SIFT algorithm. The weight is set to decrease exponentially as the spatial distance between the local neighbor and point z increases. In the modified-SIFT a weight $w_i(c)$ is assigned for each local point i using the function of equation.7. The weight $w_i(c)$ is defined in terms of the gravalue distance c between i and point z.

$$w_i(c) = \frac{1}{2\pi\sqrt{k}} e^{-c^2/2k^2} \tag{7}$$

The reason behind the above modifications is that normally local regions are identified by their color distribution. The distribution is in most cases continuous and of size proportional to the scale of the local region.

4 Evaluation

We have conducted two matching tests to measure the performance of the interest point detectors of section 3.1 before and after applying the Hourglass scale-space representation and the SIFT descriptor before and after applying the modifications of section 3.2.

In the first test a number of synthetically transformed images were used for matching. These transformations included, scale changes, rotation, brightness changes and noise addition. In this test the Receiver Operating Characteristics (ROC) curves were used for evaluation as indicated by Carneiro and Jepson [12], where for each type of transformation and each feasible combination of the three different elements under test (i.e. scale-space representation, interest point detector and local descriptor) a detection rate versus a false positive rate is plotted.

Given a test image $\mathbf{I}$ and its transformed version $\mathbf{I}'$, where $\mathbf{I'= M\ I+b}$, a detection rate is defined as the ratio between the number of correct matches (correct-positives) and the total number of interest points of $\mathbf{I}$. A correct match is scored between two interest points $\mathbf{x}$ and $\mathbf{y}$, where $\mathbf{x} \in \mathbf{I}$ and $\mathbf{y} \in \mathbf{I}'$, if $\mathbf{y}$ is very close to the mapped point $\mathbf{x'=M\ x+b}$ (i.e. $\|\mathbf{y}-\mathbf{x'}\| < \varepsilon$) and has nearly the same local descriptor as $\mathbf{x}$ (i.e. $\|D(\mathbf{y})-D(\mathbf{x})\| < \tau$).

On the other hand given a database of images that doesn't include $\mathbf{I}$ nor $\mathbf{I'}$, a false positive rate is defined as the ratio between the number of false matches (false posi-

tives) and the total number of interest points of **I**. A false match is scored if there exists an interest point **z** in the database that is similar to **x** (i.e. $\|D(\mathbf{z})-D(\mathbf{x})\| < \tau$). In our tests ε was set to 3 pixels and τ was changed in regular steps of 0.03 to form the ROC curves.

The second test involved matching real images taken from different viewpoints. In this test the evaluation of the matching results of each image pair (**I**, **I**') was based on the following criteria: for each interest point **x** that belongs to **I** the two points ($\mathbf{x_1}$ and $\mathbf{x_2}$) with the most similar descriptors to **x** are identified in **I**', where $\|D(\mathbf{x_1})-D(\mathbf{x})\| < \|D(\mathbf{x_2})-D(\mathbf{x})\|$. Next $\mathbf{x_1}$ is considered a valid match to **x** if $\|D(\mathbf{x_1})-D(\mathbf{x})\|$ is less than 90% of $\|D(\mathbf{x_2})-D(\mathbf{x})\|$. For further validation the matching results of this test were visually inspected and reported in table.3.

5 Results

The 8 test images of figure.4.a and a database of 60 different images representing a collection of natural scenes were used to create the ROC curves of figure 2, 3 and 5. These curves were designed to evaluate the performance of the five different techniques of table.1. In this test a total of 1.04 million interest points were detected according to the distributions of table.2.

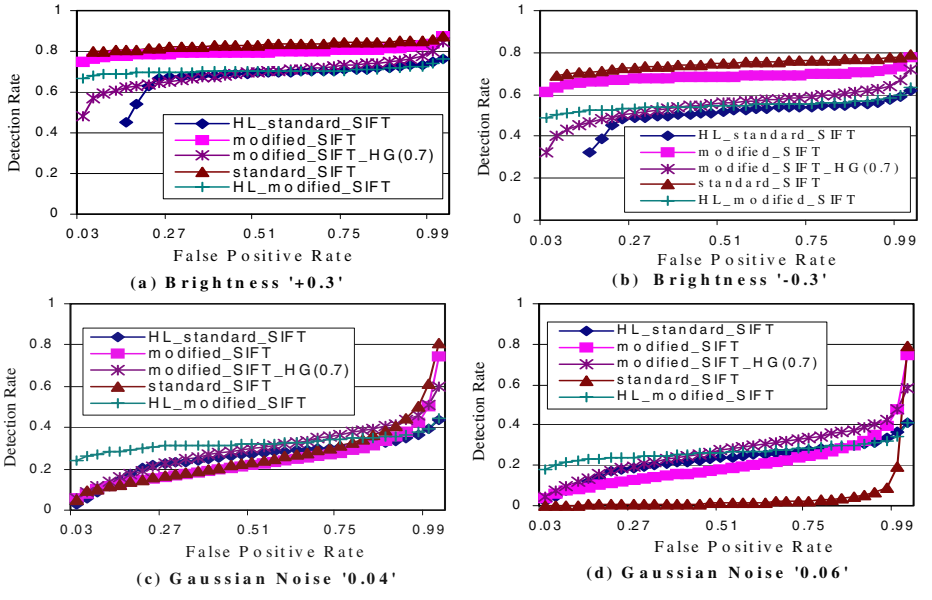

Fig. 2. ROC curves for simple image transformations that include (a) an increase in the illumination by a factor of 0.3 and (b) a decrease in the illumination by a factor of 0.3, and an addition of Gaussian noise with variances of (c) 0.04 and (d) 0.06. The curves were plotted for interest points detected by the SIFT and the Harris_Laplacian(HL) detectors and matched through the SIFT and *modified*_SIFT descriptors.

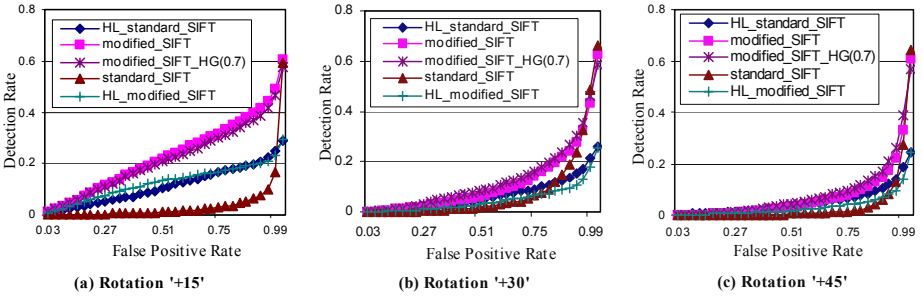

Fig. 3. ROC curves for image rotations of 15, 30 and 45 degrees

Fig. 4. Test images including the (a) original series and (b) an affine-transformed version

Table 1. The five techniques under test

Method Title	Detector	Descriptor	Scale-Space
HL_*standard*_SIFT	Harris Laplacian	SIFT	Linear
*modified*_SIFT	SIFT	*modified*_SIFT	Linear
*modified*_SIFT_HG(0.7)	SIFT	modified_SIFT	Hourglass ρ=*0.7*
*standard*_SIFT	SIFT	SIFT	Linear
HL_*modified*_SIFT	Harris Laplacian	*modified*_SIFT	Linear

In case of the Hourglass scale-space, experimental results showed that the number of detected interest points is directly proportional to the size of the smoothing kernel and inversely proportional to the value of the ρ-parameter (see equation.5), where in general an increase of 0.2 in the value of ρ results in the reduction of the number of points by a factor of 0.81. Making use of this fact and in order to speed up the process of building the Hourglass scale-space the SIFT algorithm was slightly modified, where instead of expanding the input image by a factor of 2 the first level of the Gaussian pyramid was sampled at the same rate of the input image and the smoothing

kernel was increased from size 7 to 13. This automatically implies that in case of the Hourglass scale-space no interest points can be detected with a scale less than 0.5.

The ROC curves of figures 2a and 2b show that under illumination changes the highest two detection rates were scored for the *standard*_SIFT and the *modified*_SIFT consequently. The HL_*modified*_SIFT was ranked third up to a false positive rate of 0.27. At false positive rates greater than 0.27 the *modified*_SIFT_HG was ranked third and both the HL_*modified*_SIFT and the HL_*standard*_SIFT were ranked fourth.

The curves of figures 2c and 2d show that the HL_*modified*_SIFT is the most resistant to noise at lower false positive rates while the *modified*_SIFT_HG performs much better at higher false positive rates.

To evaluate the performance for orientation changes the test images were rotated at 15, 30 and 45 degrees and the ROC curves were plotted for each angle change. The results of figure 3 show that the *modified*_SIFT and the *modified*_SIFT_HG worked much better than the other three techniques for all the three angle changes with an exceptional performance at angle 15.

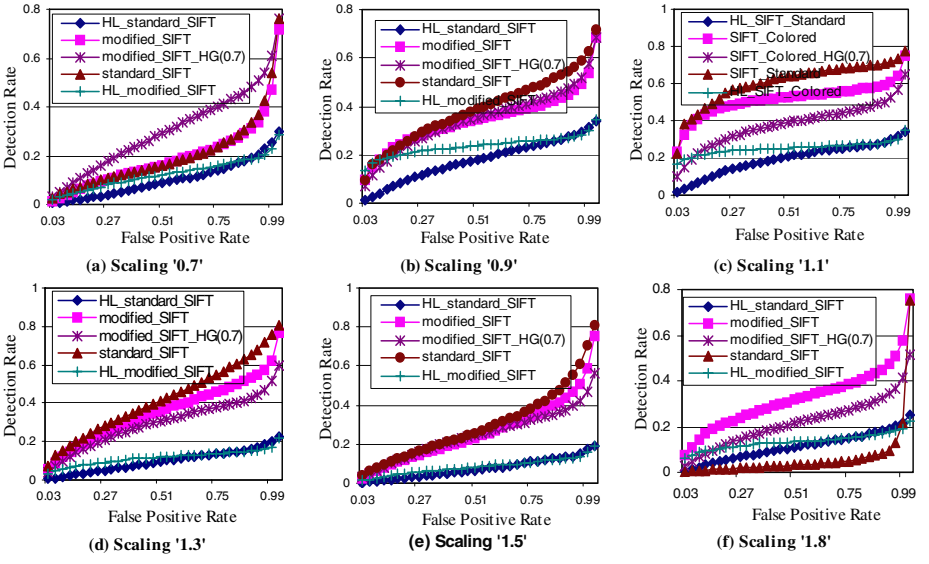

Fig. 5. ROC evaluation curves for scale changes between 0.7 and 1.8

The matching results of figure 5 involve a wide range of scale changes starting from a factor *f* of 0.7 and increasing in steps of 0.2 up to a factor of 1.8. The ROC curves show that the *modified*_SIFT_HG performed outstandingly well at *f*=0.7, the *standard*_SIFT dominated the range between 0.9 and 1.5 and the *modified*_SIFT had the highest detection rates at *f*=1.8. Moreover in the range between 0.9 and 1.1 the HL_*modified*_SIFT worked much better than the HL_*standard*_SIFT.

The reason behind the results of figure 5.a is that in the linearly smoothed version of a downscaled image the nearby edges merge causing small structures to disappear and consequently affects the localization accuracy of the interest points. On the

contrary the *modified*_SIFT_HG preserves these structures through non-linear smoothing, which in turn lead to a more accurate localization and much better matching results. Moreover the inadequate performance of the *modified*_SIFT_HG at f >1 (i.e. see figures 5c - 5.f) was due to the fact that the *modified*_SIFT_HG usually ignores the local structures of very high spatial frequencies (i.e. scales less than 0.5) and in turn reduces the number of valid matches between the input image and its scaled version.

The results of figure 5.f show that the *modified*-SIFT descriptor is more robust to large scale changes than the *standard*-SIFT because it gives more emphasis to local neighbors with similar gray values to the interest point and consequently is affected by less misregistration errors. The matching results of table.3 further prove that the modified_SIFT_HG algorithm is more resistant to affine changes than the standard_SIFT algorithm.

Table 2. Distribution of the detected interest points

Image Group	%	Method	%
Image Database	41	HL_*standard*_SIFT	13
Test Images	4	*modified*_SIFT	25
Transformed Test Images	55	*modified*_SIFT_HG(0.7)	26
		*standard*_SIFT	20
		HL_*modified*_SIFT	16

Table 3. Visually inspected matching results for the test images of figures 4.a and 4.b

Image Title	Percentage of valid matches		
	standard_SIFT	modified_SIFT	modified_SIFT_HG (0.5)
Bottle	2.72	7.09	19.9
Child	5.36	13	38.1
Croc	5.88	16.8	18
Desk	8.14	17	36.6
Lamp	0.623	2.2	12.3
Pei	2.71	7.25	13.6
Toy	9.7	13.7	24.6
Car	12.8	24.9	44.8

6 Conclusion

In this paper we have presented an experimental evaluation for a new non-linear scale-space representation and a modified version of the SIFT descriptor. The evaluation was based on matching images with both synthetic and real geometric transformations. Two different techniques were used for evaluation including the Receiver Operating Characteristic (ROC) curves and an ordinary visual inspection method. The standard SIFT descriptor proved to have better matching results under illumination changes. The results of the proposed non-linear scale-space and the *modified*_SIFT descriptor were superior under orientation and large-scale changes.

The assumption of eliminating the local structures of very high spatial frequencies from the proposed non-linear scale-space proved to be a time saving step. On the other hand it underestimated the matching results of the *modified*_SIFT descriptor.

References

1. Schmid, C., Mohr, R., Bauckhage, C.: Evaluation of Interest Point Detectors. International Journal of Computer Vision, Vol. 37, Issue 2. (2000) 151–172.
2. Schmid, C., Mohr, R.: Local gray value invariants for image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, Issue 5. (1997) 530–535.
3. Lowe, D.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision. (2004).
4. Mikolajczyk, K., Shmid, C.: Indexing based on scale invariant interest points. Proceedings of the 8th International Conference on Computer Vision. Vancouver, Can., (2001) 525-531.
5. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12, Issue 7. (1990) 629-639.
6. Kothe, U.: Edge and Junction Detection with an Improved Structure Tensor. The 25thDAGM Symposium Mustererkennung. LNCS, Vol. 278. Springer (2003) 25-32.
7. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. Proceedings of Computer Vision and Pattern Recognition. (2003).
8. Freeman, W.T., Adelson, E.H.: The design and use of steerable filters. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 13, Issue 9. (1991) 891–906.
9. Koenderink, J., van Doorn, A.: Representation of local geometry in the visual system. Biological Cybernetics, Vol. 55. (1987) 367–375.
10. van Gool, L., Moons, T., Ungureanu, D.: Affine/photometric invariants for planar intensity patterns. Proceedings of European Conference on Computer Vision. (1996).
11. Schaffalitzky, F., Zisserman, A.: Multi-view matching for unordered image sets. Proceedings of European Conference on Computer Vision, Vol. 1. (2002) 414–431.
12. Carneiro, G., Jepson, A.D.: Phase-based Local Features. Proceedings of European Conference on Computer Vision, Vol. 1. (2002) 282–296.