A Multi-scale Phase Method for Content Based Image Retrieval

Xingxing Chen, Rong Zhang, Zhengkai Liu, Lei Song
Department of Electronic Engineering and Information Science,
University of Science and Technology of China, Hefei 230027 P.R.China
Email: zrong@ustc.edu.cn Tel: +86-551-3603262 Fax: +86-551-3601342

Abstract

In this paper, we present a method using phase features for content based image retrieval (CBIR). Two related key issues of CBIR are feature extraction and similarity measure. However, most traditional methods treat them respectively and prevent further performance improvement. The method proposed here is based on the multi-scale local phase feature (MLPF) and local weighted phase correlation which combines the above two issues together by phase. And phase data is often locally stable with respect to noise, scale change and common illumination change. Moreover, we implement steerable filters to obtain rotation invariant. Finally, experiments have been conducted on image retrieval to show the effectiveness of the proposed method.

Keywords: *CBIR, multi-scale local phase feature, local weighted phase correlation*

1. Introduction

Over the last decade, content-based image retrieval has received a lot of attention and the field has grown tremendously especially after the year 2000 [1]. It's a new area of multimedia informatics covering techniques for retrieval of images from image databases based on their visual contents. The essential content of retrieval is usually represented by simple image features such as color, shape and texture. A few CBIR systems are already known in practice as QBIC [2], NETRA [3], Photobook [4] and a more general survey on image retrieval systems could be found in [5].

CBIR Systems such as QBIC, NETRA and Photobook employ different low-level image features for retrieval. In particular, the QBIC system adopts the color histogram along with a texture descriptor implemented using the Tamura's feature set. In NETRA project, Gabor filters have been used to extract texture features for the purpose of retrieval. Similarly, color, shape and texture information is implemented in Photobook. However, most of the proposed methods assume that the images have the same orientation or scale [6]. This assumption is not realistic for most practical application and the performance of these methods becomes worse when this underlying assumption is no longer true.

There is no doubt that the performance of a CBIR system mostly depends on the way in which the image visual content is represented. An efficient visual content descriptor is achieved not only from the appropriate features extracted, but also through a proper organization of them. Generally, these organization schemes aim at providing a more physical interpretation of the visual content in an image, which is much closer to the human perception. Other than image visual content representation, the similarity measure is another key related aspect of CBIR. Similarity is an interpretation of the image based on the difference with another image. For each feature types, the best performance can only be achieved by an according similarity measure.

In this paper, we propose a multi-scale phase method which takes into consideration the above problems. Feature extraction process firstly involves a region of interest (ROI) detection based on scale-invariant keypoints [11]. Then a powerful orientation analyzing tool, namely, complex steerable filter (CSF) is applied to the detected regions so as to extract the multi-scale phase feature which is scale invariant and 2D rotation invariant. Moreover, a special organization of the feature is performed. In this way, the problem caused by the wrong assumption of image orientation and scale is solved. What's more, we implement a local weighted phase correlation (LWPC) as the similarity measure for our phase feature. The LWPC combines the robustness of phase difference and the voting strategy of phase correlation and appears consistent with physiological data [19].

The remainder of this paper is organized as follows. Sec. 2 provides a detail description of the proposed multi-scale local phase features, including our ROI detection and specially designed CSF, while Sec. 3 describes the similarity measure we put forward. In Sec. 4 we conduct an experiment to demonstrate the results of our algorithm for retrieval tasks. Finally, Sec. 5 concludes the paper and provides an overview on our future work.

2. Multi-scale local phase feature

0-7695-2929-1/07 $25.00 © 2007 IEEE
DOI 10.1109/ICIG.2007.4

Feature extraction is a preprocessing step for most CBIR systems. Considering the goal of CBIR systems, to provide the user with "similar" images in some user-define sense, it may be unnecessary that content-based retrieval relies on describing the content of an image in its entirety. A scale- invariant keypoints based ROI descriptor would be sufficient to represent the image content. Another reason for using ROI is for reducing computational cost. As mentioned in the abstract, phase data provide a lot of advantages over other features. So, we implement our specially designed CSF to the ROI detected in order to extract proper phase features. In this section, we proposed a method to extract multi-scale local phase feature for a given query image and images in the database. MLPF of an image can be obtained by applying ROI detection, followed by a complex steerable filtering.

2.1 ROI detection

As a first step to extract MLPF, scale-invariant keypoints are implemented to locate the potential region of interest. The scale-invariant keypoints were firstly introduced by David G. Lowe in 1999 [12] and further developed in 2004 [11]. And two major stages are needed to detect the scale-invariant keypoints: scale-space extrema detection and keypoints localization.

The first stage searches over all scales and image locations by using a famous function of scale known as scale space [13]. It is proposed by Koenderink [14] and Lindeberg [15] that under some reasonable assumption Gaussian function is the only possible scale space kernel. For this reason, scale space of an image is usually defined as follows:

$$L(x,y,\sigma) = G(x,y,\sigma) \otimes I(x,y) \quad (1)$$

In formula (1), L(x, y, σ) is scale space function, G(x, y, σ) is a variable scale Gaussian function and I(x, y) is an input image while $\otimes$ stands for 2D convolution in x and y.

In order to efficiently detect the stable keypoints that are invariant to scale, Lowe proposed a difference of Gaussian (DOG) function, D(x, y, σ), which can be computed from the difference of two nearby scales of scale-space separated by a constant factor k:

$$D(x,y,\sigma) = (G(x,y,k\sigma) - G(x,y,\sigma)) \otimes I(x,y)$$
$$= L(x,y,k\sigma) - L(x,y,\sigma) \quad (2)$$

It has been proved that D(x, y, σ) is a close approximation to the scale-normalized Laplacian of Gaussian which is scale invariant. And the extrema of scale-normalized Laplacian of Gaussian are the most stable image features according to detailed experimental comparisons [17]. So as to detect the local extrema of D(x, y, σ), each pixel is compared to its eight neighbors in the current scale and nine neighbors in the scale above and below.

After the candidate keypoints are found, a further step is taken to reject those that have low contrast or are poorly localized on an edge. This stage is known as accurate keypoints localization. The latest approach is developed by Brown [18]. He fits a 3D quadratic function to the pixels so as to determine the interpolated location of the maximum. This method uses the Taylor expansion of D(x, y, σ): D, which has been shifted so that the origin is at the current pixel being processed. The location of the extrema is calculated by taking the derivative of this function with respect to x, x=(x, y, σ)' is the offset from the current pixel, and set it to zero, thus we have,

$$\hat{x} = -\frac{\partial^2 D^{-1}}{\partial x^2}\frac{\partial D}{\partial x} \quad (3)$$

Finally, by substituting equation (3) into D, we have the function value at the extrema,

$$D(\hat{x}) = D + \frac{1}{2}\frac{\partial D^T}{\partial x}\hat{x} \quad (4)$$

All the extrema with a value of $|D(\hat{x})|$ less than 0.03 are rejected as unstable extrema with low contrast. For further stability, edge responses should be eliminated. In this case, we calculate the principal curvatures at the location and scale of the keypoint using a 2x2 Hessian matrix, H,

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{yx} & D_{yy} \end{bmatrix} \quad (5)$$

By using the approach by Harris and Stephens [16] we can easily obtain the ratio of the eigenvalues. Let α be the eigenvalue with the largest magnitude, β be the smaller one and r be the ratio between largest eigenvalue and the smaller one. So that r = α/β. we can compute the ratio by the trace and the determinant of H. Then we eliminate keypoints who have a ratio r above the threshold 10. This threshold is obtained from experiments.

Once the keypoints are detected, we simply apply a circular mask to the area where the keypoints are located, and take this mask covered regions as our ROIs.

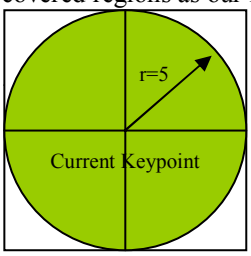

Figure 1. Region of interest.

As shown in Figure 1, our region of interest is centered at the current keypoint and has a radius of 5 pixels.

To give an intuitive view of the ROI detection, we carry out an experiment here using two images: "qb1.bmp" is the original image, and "qb1_r.bmp" is the same image which has been rotated 90 degrees. And the green circles in the result images are our detected ROI. We convert our color images into gray and don't show all the ROI in the below figure for some display reason.

qb1.bmp qb1_r.bmp

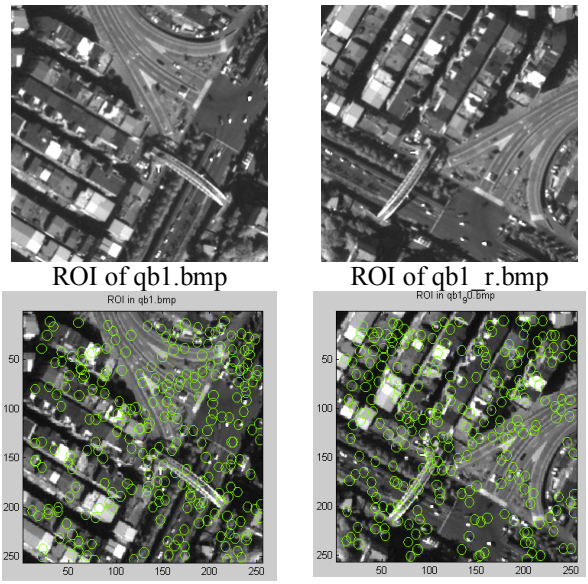

Figure 2. Some ROI detected in two "similar" images.

2.2 Complex steerable filtering

The concept of steerable filter is first proposed by William T. Freeman and Edward H. Adelson [9]. It describes a class of filters in which a filter of arbitrary orientation can be synthesized as a linear combination of a set of "basis filters". Moreover, it may be designed in quadrature pairs to allow adaptive control over phase as well as orientation.

To extract the phase of ROI detected in section 2.1, the easiest way is to use complex-valued quadrature-pair steerable filters. A pair of filters is said to be quadrature if they are Hilbert transforms of each other, i.e. they have identical amplitude spectra but differ in phase by π/2 radians. The reason for choosing quadrature filter pairs is that they allow for synthesizing filters of a given frequency response with arbitrary phase. The second order steerable quadrature filter pair can be written as follows:

$$G_2 = 0.9213(2x^2-1)e^{-(x^2+y^2)}; H_2 = (-2.205x + 0.9780x^3)e^{-(x^2+y^2)} \quad (6)$$

By using the CSF designed in formula 6, we can easily calculate the phase.

In order to prevent high computational costs, steerable quadrature filter pair is further designed to be x-y separable. For any function f which can be written as polynomial in x and y, it may be decomposed in x-y separable basis. Applying the rotation formula to each x and y term of the polynomial will produce a sum of powers of x and y with coefficients which are function of the rotated angle:

$$f^{\theta}(x,y) = \sum_i \sum_j k_{ij}(\theta) x^i y^j \quad (7)$$

Each x and y product in the rotated polynomial can be thought of as an x-y separable basis, with its coefficients $k_{ij}(\theta)$ the interpolation function. The detailed realization of the second order basis filters could be found in [9].The x-y separable quadrature pair basis filters are used to synthesize the filter steer to an arbitrary orientation θ:

$$G_2^{\theta} = k_{ga}(\theta)G_{2a} + k_{gb}(\theta)G_{2b} + k_{gc}(\theta)G_{2c};$$
$$H_2^{\theta} = k_{ha}(\theta)H_{2a} + k_{hb}(\theta)H_{2b} + k_{hc}(\theta)H_{2c} + k_{hd}(\theta)H_{2d} \quad (8)$$

Thus we have our complex steerable filter (CSF) turned to a specific orientation, θ:

$$CSF(x,y,\theta) = G_2^{\theta} + iH_2^{\theta} \quad (9)$$

And we define the orientation energy of our detected ROI along a particular direction, θ, as E(θ):

$$E_2(\theta) = [G_2^{\theta}]^2 + [H_2^{\theta}]^2 \quad (10)$$

It can be decomposed to a Fourier series in angle, where only even frequencies are present:

$$E_2(\theta) = C_1 + C_2\cos(2\theta) + C_3\sin(2\theta) + \ldots \quad (11)$$

Then we use the lowest frequency term to approximate the main direction of ROI, θ_M, the orientation which maximize the orientation energy,

$$\theta_M = \frac{1}{2}\arctan(\frac{C_3}{C_2}) \quad (12)$$

Finally, we apply our CSF to all the ROI at according scale along main direction to extract the multi-scale local phase feature. The complex steerable filtering is performed by 2D convolution:

$$CSF(x,y,\theta_M) \otimes R(x,y,\sigma) = G_2^{\theta_M} \otimes R(x,y,\sigma) + iH_2^{\theta_M} \otimes R(x,y,\sigma) \quad (13)$$

where CSF is the filter, R(x, y, σ) is the ROI at scale σ.

Let

$$G(x,y,\theta_M,\sigma) = G_2^{\theta_M} \otimes R(x,y,\sigma);$$
$$H(x,y,\theta_M,\sigma) = H_2^{\theta_M} \otimes R(x,y,\sigma) \quad (14)$$

A complex polar representation (i.e. MLPF) is derived as:

$$\rho(x,y,\theta_M,\sigma)e^{j\phi(x,y,\theta_M,\sigma)} = G(x,y,\theta_M,\sigma) + iH(x,y,\theta_M,\sigma) \quad (15)$$

where $\phi(x,y,\theta_M,\sigma)$ is the multi-scale local phase. And then a registration step is taken, which rotates the 2D complex feature clockwise by θ_M degree.

So as to demonstrate the effectiveness of steerable filters' ability to determine the main direction, we conduct an experiment as follows: "qb1.bmp" is the original image, and "qb1_r.bmp" is the same image which has been rotated 90 degrees. The image is randomly selected. Then we calculate the main directions of these two images respectively. From the figure below, we find out that the two main directions detected from qb1 and qb1_r differ in phase by exactly 90 degrees.

qb1.bmp qb1_r.bmp

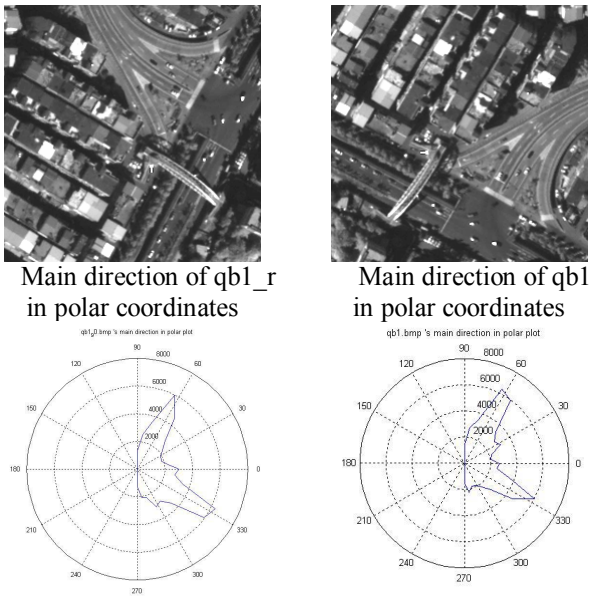

Figure 3. Main direction extracted from two "similar" images.

After all these steps, we come to a multi-scale local phase feature of the image. The most exciting part of this MLPF is that it is scale invariant and 2D rotation invariant due to its multi-scale analyzing ability and self registration for rotation. What's more, the computational cost of our complex steerable filtering won't be too high due to its separable filter design.

3. Similarity measure

Once our MLPF is extracted, the similarity between two images is computed through the features. Borrowing from the method proposed by David J. Fleet [19], we use local weighted phase correlation instead of Euclidean distance. And we have modified the original formula in [19] so as to calculating the similarity measure coefficient SR.

Let the complex-valued steerable filtering results of ROI on query image Q(x, y), and database image D(x, y) be,

$$C_q(x,y,\theta_M,\sigma) = CSF(x,y,\theta_M) \otimes R_q(x,y,\sigma);$$
$$C_d(x,y,\theta_M,\sigma) = CSF(x,y,\theta_M) \otimes R_d(x,y,\sigma) \quad (16)$$

Thus the registered MLPF can be written as:

$$C_q(x,y,\sigma) = \rho_q(x,y,\sigma)e^{i\phi_q(x,y,\sigma)}; C_d(x,y,\sigma) = \rho_d(x,y,\sigma)e^{i\phi_d(x,y,\sigma)} \quad (17)$$

The similarity of two ROIs from Q and D is calculated by:

$$SR = \frac{\sum_x \sum_y W(x,y) \otimes [C_q(x,y,\sigma) C_d^*(x,y,\sigma)]}{1 + \sqrt{\sum_x \sum_y W(x,y) \otimes |C_q(x,y,\sigma)|^2} \sqrt{\sum_x \sum_y W(x,y) \otimes |C_d(x,y,\sigma)|^2}} \quad (18)$$

where W(x, y) is a small localized window, and its size changes with the scales. Here we noticed that the above formula for calculating SR is a little bit different from Equation (13) in [19].

$$C_j(x,\tau) = \frac{W(x) \otimes [O_l(x) O_r^*(x+\tau)]}{\sqrt{W(x) \otimes |O_l(x)|^2} \sqrt{W(x) \otimes |O_r(x)|^2}} \quad (19)$$

It is reasonable that we have made the modification. As we know, Equation (13) in [19] is for extracting disparity from two images. In that case, τ acts as a preshift of the right filter output and it stands for the candidate "disparity". But for our similarity measure, we focus on the final correlation result SR rather than the candidate disparity τ. Moreover, we add 1 in the denominator in case it might be 0.

The similarity SR is between 0 and 1, where 1 stands for 100% similar and 0 stands for not similar at all. Two ROIs are defined to be "similar" only when their SR >= τ1. We assume the number of similar ROI from two images to be N (Q, D), and the number of ROI detected in Q to be N (Q). Then the retrieved image D(x, y) is accepted as "similar" to Q(x, y) only when N (Q, D)/N (Q) >= τ2. Notice that τ1, τ2 are two user defined thresholds. This will give the user kind of ability to adjust the output retrieval results according to his demand.

4. Experiment results

To demonstrate the effectiveness of our proposed method, we carry out a CBIR experiment on the 234 categories from CD 8 of Corel photo gallery. Before we conduct the image retrieval on these 23,399 images, similar images are sorted into the same category manually. The 234 categories have specific names which have semantic meanings such as "agricltr", "architi", "castlei", "castleii" and etc.

Here are some examples of the images in the database. As there are too many image classes, we only choose a few categories in the following figure:

Figure 4. Some examples in Corel photo gallery.

Once the images are ready, we start our retrieval process. Firstly, we randomly choose a few images from the database as query images and retrieve the left image database. This is the typical QBE (Query by Example) mode. Some of the retrieved results are shown as follows:

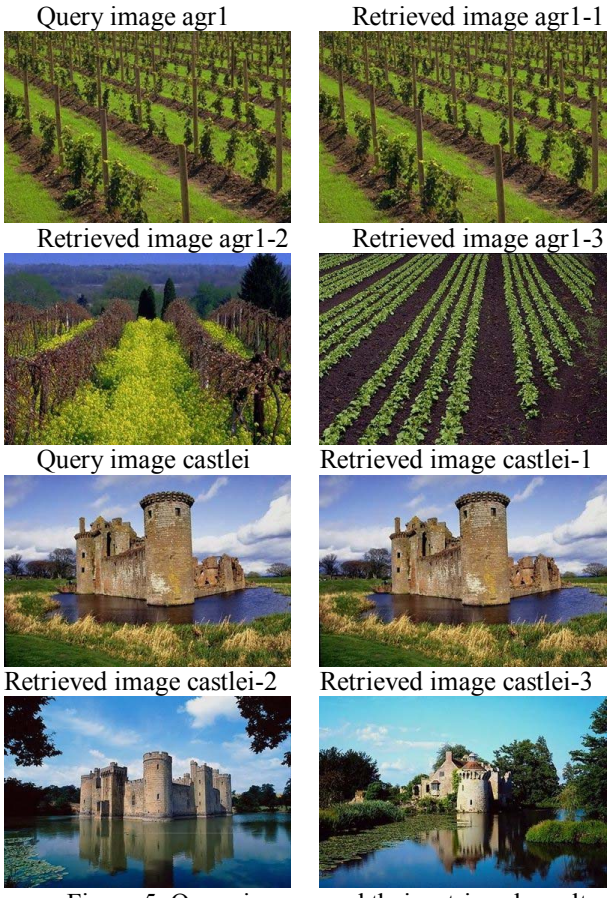

Figure 5. Query images and their retrieved result images.

In the above figure, we have given a retrieval example by using ar1 and castlei as query images. The "retrieved images" columns contain a part of the results from the database; what's more, the results are sorted in descending order according to their SR coefficient, i.e. the first retrieved image is the most "similar" result image.

In order to statistically evaluate our retrieval performance, and make a comparison with another algorithm, we also draw the precision-recall curve. Precision is the ratio of the number of relevant records retrieved to the total number of irrelevant and relevant records retrieved. Recall is the ratio of the number of relevant records retrieved to the total number of relevant records in the database. Precision and Recall are useful measures despite their limitations. Since LSP [10] is also a phase based method, we compare our work with it to show the performance improvement. The two precision-recall curves are given in figure below:

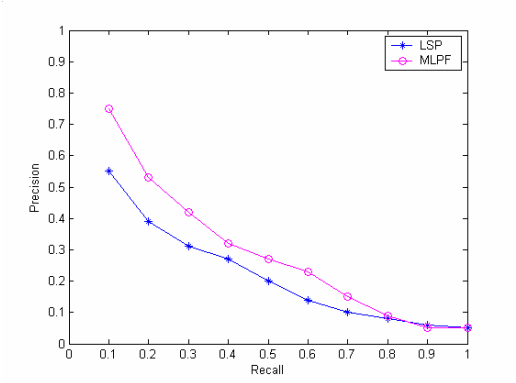

Fig. 6 Precision-Recall curve.

According to the curve in Fig. 6, we can clearly see the improvement of our proposed method, though the precision is a little bit lower than LSP when recall is near 90%. However, the performance under such high recall rate is meaningless. It is not hard to explain. As descried in Sec.3, if we set a loosen threshold $\tau 1$, most of the images relevant or irrelevant would likely be retrieved, which lead to a very high recall rate. But in the meantime the precision rate would drop sharply which results from a lot of irrelevant retrieved images.

5. Conclusions

Content-based image retrieval is a typical high-level image processing task and the two main issues are feature extraction and similarity measure. In this paper, we propose an efficient multi-scale phase method which combines the two issues into one integrated solution. The main contribution of this paper is that this viewpoint may provide a better retrieval performance. According to the experiment results, our method has outperformed LSP. However, the method present is totally based on local phase features. Further improvement could be achieved if we take into account some global features.

6. References

[1] Ritendra Datta, Jia Li, James Z. Wang, "Content-based Image Retrieval-Approaches and Trends of the New Age", *Proc. of ACM Multimedia Workshop on Multimedia Information Retrieval (MIR)*, pp. 253--262, Singapore, Nov. 2005.

[2] Flickner M., Sawhney H., Niblack W., et al., "Query by Image and Video Content: The QBIC System", *IEEE Computer*, 1995.

[3] Ma WY, Manjunath BS, "NeTra: A Toolbox for Navigating Large Image Databases", *Multimedia Systems*, Vol. 7, No. 3, 1999.

[4] Pentland A., Picard R. W. and Sclaroff S., "Photobook: Content-Based Manipulation of Image Databases", *International Journal of Computer Vision*, 18(3), 233-254, 1996.

[5] Remco C.Veltkamp, Mirela Tanase, "Content-Based Image Retrieval Systems: A Survey", *Technical Report UU-cs-2000-34*, October, 2000.

[6] C.M. Pun, "Invariant content-based image retrieval by wavelet energy signatures", *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing. (ICASSP)*, vol. 3, pp. 565–568, 2003.

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

[8] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-Based Image Retrieval at the End of the Early Years", *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2000.

[9] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891--906, 1991

[10] Xiaoxun Zhang, Yunde Jia, "Local Steerable Phase (LSP) Feature for Face Representation and Recognition", *CVPR*, 2006

[11] Lowe D G, "Distinctive Image Features from Scale-Invariant Keypoints", *International Journal of Computer Vision*, 2004

[12] Lowe D G, "Object Recognition from Local Scale-Invariant Features", *ICCV*, 1999

[13] Witkin A. P, "Scale space filtering", *Proc 8th International Joint Conference Artificial Intelligence*, 1983

[14] Koenderink, J.J, "The structure of images", *Biological Cybernetics*, 1984

[15] Lindeberg, T., "Scale-space theory: A basic tool for analyzing structures at different scales", *Journal of Applied Statistics*, 1994

[16] Harris, C. and Stephens, M., "A combined corner and edge detector", *Fourth Alvey Vision Conference*, 1988

[17] Mikolajczyk, K., and Schmid, C., "An affine invariant interest point detector", *European Conference on Computer Vision (ECCV),* 2002

[18] Brown, M. and Lowe, D.G., "Invariant features from interest point groups", *British Machine Vision Conference*, 2002.

[19] David J. Fleet, "Disparity from Local Weighted Phase-Correlation", *IEEE Systems, Man, and Cybernetics*, 1994