Local invariant object localization based on a reduced color space

RÜDIGER HEINTZ, EDUARDO MONARI, GERHARD SCHÄFER
Fakultät für Elektro- und Informationstechnik
University of applied science
Moltkestr. 30, 76133 Karlsruhe
GERMANY
http://www.home.fh-karlsruhe.de/~heru0001

Abstract: - Invariant object localization is one of the challenging tasks in computer vision research. In this paper we present a robust rotation and scale invariant object localization method. A local Gabor filter space is treated as core of this method. Image rotation and scaling operations were transformed into shift operations along the Gabor filter space dimensions. This property enables efficient scale and rotation estimation without segmentation. The used Gabor Filter can deal with complex input data. This feature is used to analyze color images by reducing the color space. The method was tested with two standardized image databases and our own image database. Besides a good quality in localizing rotated and scaled objects, the method has a strong robustness against variations of lightning conditions and 3D viewpoint changes.

Key-Words: - Object recognition, Gabor filter, robust local features, rotation and scale invariant, object localization, feature space, color space reduction

1 Introduction

Object localization in a real world environment is one of the most challenging problems in computer vision. This paper describes a feature space that allows an efficient match of scaled and rotated objects.

This work based on the publication of Ville Kyrki and Joni Kristian Kamarainen[1], who introduced a Gabor feature space and used it for facial feature extraction. In this feature space scaling and rotation are transformed to shifts which can be simply handled. The proof that rotation and scaling can be transformed into shifts in this feature space was introduced in [1] and completed by us in [2].

In [3], we examined this feature space and determined the suitable parameters of the feature extraction. Based on the results in this paper a localization method for objects is described. The applied methods were tested under different environment conditions.

2 Gabor feature space

A number of authors are using a set of Gabor filters to extract local image features. Key applications are face recognition [4] and texture segmentation [5]. Basically an input image is convolved by a collection of 2-D Gabor functions and the convolution results for each pixel are stored as a feature vector.

2.1 Gabor filter

2D Gabor filter cores are characterized by equation

$$g_k(x,y) = \frac{1}{\sigma_x \sigma_y} \cdot e^{\left(-\frac{1}{2}\left(\frac{\breve{x}^2}{\sigma_x^2} + \frac{\breve{y}^2}{\sigma_y^2}\right)\right)} \cdot e^{\left(j\cdot\omega_k\cdot(\breve{x}+\breve{y})\right)} \quad (1)$$

$$\text{with } \begin{bmatrix} \breve{x} \\ \breve{y} \end{bmatrix} = \begin{bmatrix} \cos(\theta_n) & \sin(\theta_n) \\ -\sin(\theta_n) & \cos(\theta_n) \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}$$

The filter function can be interpreted as a damped complex wave. The parameters describe the expansion in x and y direction (σ_x, σ_y), the orientation of the complex wave θ_n and the frequency ω_k. The decomposition of rotation and scaling as independent values can be handled by a Gabor filter with restricted parameters [1], [2].

$$g_W(x,y,\omega_k,\theta_n) = \frac{\omega_k^2}{\pi^2} \cdot e^{\left(-\frac{(x^2+y^2)\cdot\omega_k^2}{2\cdot\pi^2}\right)} \cdot e^{\left(j\cdot\omega_k\cdot(\breve{x}+\breve{y})\right)} \quad (2)$$

$$\text{with } \sigma_x = \sigma_y = \frac{\pi}{\omega_k} \,, \, \begin{bmatrix} \breve{x} \\ \breve{y} \end{bmatrix} = \begin{bmatrix} \cos(\theta_n) & \sin(\theta_n) \\ -\sin(\theta_n) & \cos(\theta_n) \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}$$

2.2 Gabor filter bank

Using Gabor fiters for a special application normally requires a variation of ω_k and θ_n to extract significant features. The convolution results of a set of Gabor filters (Gabor filter bank) from a given image can be represented as a matrix where the rows are formed by variations of θ_n and columns by variations of ω_k.
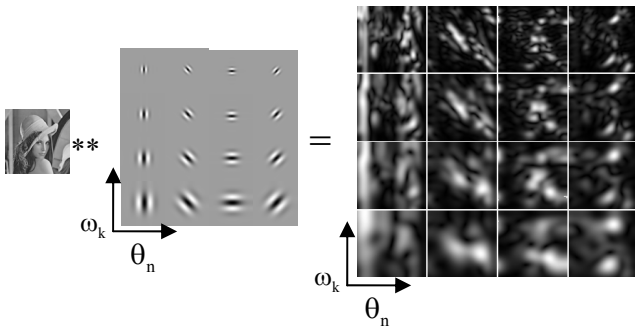
Fig.1 Result of a Gabor filters convolution

Fig.1 shows the real part of the filter functions and the magnitude of the convolutions. Like formula (3) the convolution results for each pixel can be combined to a matrix (Fig.2). This matrix is described as Jet matrix.

$$J(x,y)=$$
$$\begin{bmatrix} g_W(x,y,\omega_1,\theta_1) & g_W(x,y,\omega_1,\theta_2) & \cdots & g_W(x,y,\omega_1,\theta_N) \\ g_W(x,y,\omega_2,\theta_1) & g_W(x,y,\omega_2,\theta_2) & \cdots & g_W(x,y,\omega_2,\theta_N) \\ \vdots & \vdots & \ddots & \vdots \\ g_W(x,y,\omega_M,\theta_1) & g_W(x,y,\omega_M,\theta_2) & \cdots & g_W(x,y,\omega_M,\theta_N) \end{bmatrix} \quad (3)$$
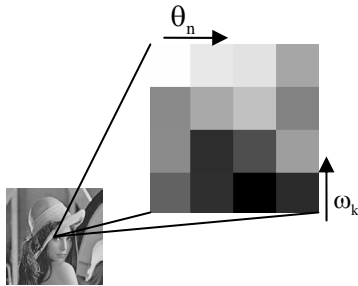

Fig.2 Jet matrix of an image pixel

The Jet matrix can be interpreted as a discretization of a feature space spanned by ω_k and θ_n. This feature space is defined as Gabor feature space.

The parameters of the sample points are calculated by formulas (4) and (5). The formulas are based on theoretical [1], [2] and experimental results [3].

$$\omega_k = 2^{-\frac{m}{2}} \quad \text{with } m = \{1,...,M\} \; M>1 \quad (4)$$

$$\theta_n = n \cdot \frac{\pi}{N-1} \quad \text{with } n = \{1...N\} \; N>1 \quad (5)$$

Formula (5) restricts θ_n to the range 0 to π. Due to the symmetry properties (6) of the Gabor filters with restricted parameters (2) this range is sufficient.

$$g_W(x,y,\omega_k,\theta_n) = g_W^*(x,y,\omega_k,\theta_n+\pi) \quad (6)$$

As described in [1] and [2] scaling and rotation of an environment changes the associate Jet matrix by shift operations. By negletion of signs from the imaginary values in the Jet matrix, rotation of an environment is transformed into circular shifting in the associate Jet matrix.

2.2 Fast computation

The calculation of each Value in the Jet matrix requires a complex convolution operation. A realization of this complex convolution with 2D filter structure is less efficient. The operation can be accelerated by making use of the separation property of the Gabor filters. So the 2D filter structure can be transformed into two 1D filter structures. This results in a better performance. A further improvement can be found by realizing the convolution via FFT. At least the fastest algorithm used here was the recursive Gabor filter [6]. This algorithm works with linear complexity. On our test system (P4, 2.66 GHz, 512MB, WinXP, VS.net 7.1) the recursive Gabor filter was approximate two times faster than an optimized FFT based [7] convolution.

3 Color image analysis

The original localization system was operating with grayscale images. Values in grayscale images vary only along one dimension.

Color images are based on different color systems depending on application. Practical color spaces are dealing at least with 3 dimensions. So grayscale images can be interpreted as dimension reduced color images.

Due to the fact, that the used Gabor filter algorithm can handle complex input data with little additional effort not only grayscale images can be processed. The reduction of the color space from 3 dimensions to 2 dimensions is less radical then reducing the data to one dimension. So more information can be use for object localization with little additional effort.

The dimension reduction is carried out under the condition to lose minimal information about the reference object. Therefore the reference object is examined by Principal Component Analysis (PCA)[10]. PCA is a linear transformation that allows a dimension reduction under minimal lost of cumulative variance. Variance is a measure for the information content.

By inverse transform the resulting information content of the dimension reduction can be clarified.

Fig.3 Original and image after dimension reduction

In Fig.3 an example for a dimension reduction with PCA is shown. The right image is the result of the inverse transformation of the dimension reduced left image. The mean relative error is 2.46%.

For object recognition only the object information must remain. Therefore the reference object is used for calculating the transformation parameters.

The calculation effort for the search image is described by only 6 multiplication and 6 addition operations per pixel.

Fig.4 Reference object and search image.

Fig.5 Grayscale and reduced color space search image

Fig.5 shows a comparison between information content of grayscale image and reduced color space image. The sign in the reduced color space image has a strong contrast to the background.

4 Object localization

This object localization method is using only two images, a reference image with the reference object and a search object in a search image, which can be scaled and rotated. First of all, significant points of the reference object have to be marked. The positions of these significant points and their Jet matrices will be extracted. The complete object localization is based on this information.

The procedure itself is described within three subsections. The first subsection deals with the localization estimation of significant point environments in the search image. The second subsection shows that prefiltering results in a better performance for localization estimation of significant point environments. The last subsection describes how the estimated point localization is used for object localization.

4.1 Environment localization

The matching of the reference object in the search image is done by the comparison of Jet matrices. For that all shifts between the Jet matrices are checked to determine the shift with maximum correlation.

For robust localization under various scales it's important to compensate the shift in scale direction of the Jet matrices. This will be done by a reference Jet matrix which was expanded. In rotation direction this isn't needed because the signs of the imaginary parts are ignored and so the Jet matrix is circular shifted in rotation direction.

The scale range can be calculated by the amount of additional rows K

$$\text{Scale: } S = 2^{-\frac{K}{2}} \dots 1 \tag{7}$$

4.1.1 Standardized complex cross correlation

The standardized complex cross correlation of two Jet matrices JR and JS is calculated by formula (9). The size of the search matrix is MxN. The size of the reference matrix is (K+M)xN.

$$C_{m,n} = \frac{\sum_{i}^{M}\left(\sum_{j}^{N} JSA_{i,j} \bullet JRA^{*}_{i+m,(j+n)\bmod(N)+1}\right)}{\sqrt{\sum_{i}^{M}\sum_{j}^{N} JSA_{i,j} \bullet JSA^{*}_{i,j} \cdot \sum_{i}^{M}\sum_{j}^{N} JRA_{i+m,(j+n)\bmod(N)+1} \bullet JRA^{*}_{i+m,(j+n)\bmod(N)+1}}}$$

$$C_{m,n} = \frac{\sum_{i}^{M}\left(\sum_{j}^{N} JSA_{i,j} \bullet JRA^{*}_{i+m,(j+n)\bmod(N)+1}\right)}{\sqrt{\sum_{i}^{M}\sum_{j}^{N} |JSA_{i,j}|^{2} \cdot \sum_{i}^{M}\sum_{j}^{N} |JRA_{i+m,(j+n)\bmod(N)+1}|^{2}}} \tag{8}$$

$$\text{with } JxA_{i,j} = \text{Re}\left(Jx_{i,j}\right) + I \cdot \text{Abs}\left(\text{Im}\left(Jx_{i,j}\right)\right) \tag{9}$$

Formula (9) takes into account that the signs of the imaginary parts are ignored. The modulo function is needed for circular shifting. The size of the output matrix C is (K+1)xN.

The maximum of absolute values in C is a similarity measure with ranges between 0 and 1 where 1 means identical. The shift of the maximum corresponds to the shift of the Jet matrices and therefore to the rotation and scale parameters between the matched environments.

The calculation of the cross correlation is effortful. The sum of the reference Jet matrix in the denominator is equal for all Jet matrix comparisons and therefore it must be calculated only once. The

sum of the search Jet matrix in the denominator is independent of m and n, and therefore it needs to be charged again for every Jet matrix only once.

The large computation effort arises from the sum in the numerator. This must be calculated m x n times.

An acceleration of this method was found by correlating of the mean average values. The procedure can be used for shift estimation but the cross correlation result isn't selective enough as similarity measure.

$$C_{m,n} = \frac{\sum_{j}^{N}\left(\sum_{i}^{M} JSA_{i,j} \cdot \sum_{i}^{M} JRA_{i+m,(j+n)\bmod(N)+1}^{*}\right)}{\sqrt{\sum_{i}^{M}\sum_{j}^{N}\left|JSA_{i,j}\right|^{2} \cdot \sum_{i}^{M}\sum_{j}^{N}\left|JRA_{i+m,(j+n)\bmod(N)+1}\right|^{2}}} \qquad (10)$$

To solve this, the cross correlation for the estimated shift calculated by formula (8) is used as similarity measure. This method is approximately 3 times faster than the cross correlation without shift estimation and the results are almost equal (Table 1).

4.1.2 Test of the environment localization methods

The test dataset consists of images from the SIPI Image Database [8], the Surrey Object Image Library (SOIL) [9] and own images.

Reference images from each images library were selected. From this reference images 800 search images were generated by different rotation and scale values. Extra search images were generated by changing the 3D viewpoint position. In each test images 2-5 significant points were marked. Each of these significant points was searched in the associate search images. The test dataset contained 2000 positions which should be found.

The results are shown in table 1. The size of the reference Jet matrices was 9x6 and the size of the search Jet matrices 6x6. Best results are given by standardized complex cross correlation on color images.

Method	determination procedure	
	best point	best of five
complex cross correlation based on color images	1839	1933
complex cross correlation based on color images with estimated shifts	1817	1915
complex cross correlation based on grey images	1621	1687
complex cross correlation with estimated shifts	1602	1653

Table 1: Comparison of environment localization methods

Two different determination procedures were used. Best point function checks the correctness of the point position with the greatest similarity measure. The Best of five points function checks whether one of the estimated best five points meets the correct position.

4.2 Prefiltering

Before starting the search, areas which are unusable for object localization should be cut out. Due to the reduction of comparisons this results in better performance.

The object localization method is based on localizations of significant points. Each Jet matrix contains information about distribution of frequencies inside a local environment. This information can be used to classify unique environments.

Edges and uniform areas are nearly unrecognizable. This is because edges and uniform areas aren't unique. Along edges as well as inside uniform areas local environments are almost identical.

So the prefiltering is a rough classification between unique and non unique environments. This classification is possible by evaluating the Jet matrix distribution.

Fig.6 shows representative Jet matrix distributions for edges, unique environments and uniform areas.

Fig.6 Jet matrix of arbitrary positions in image

Fig.6 shows a significant feature of an edge. In a wide frequency range one dominant direction exists. The Jet matrices of unique environments have random distribution. A uniform area which is shown in Jet matrix 2 is characterized by small values. These criteria can be used for the classification of unique environments.

The classification was tested through different feature: contrast, variance, mean, row variance etc. A lot of features provided good results. Often two features were necessary to suppress edges and areas. The Minimum Column Variance was used at least.

This provides good results; the characterization value can simply be calculated and suffices as a feature for the classification of a Jet matrix.

Column mean:
$$m_n = \frac{1}{M}\sum_{i=1}^{M} JSA_{i,n} \qquad (13)$$

Column variance:
$$v_n = \frac{1}{M}\sum_{i=1}^{M}\left(JSA_{i,n} - m_n\right)^2 \qquad (14)$$

Column variance vector:
$$v = \begin{bmatrix} v_1 & v_2 & \cdots & v_K \end{bmatrix} \qquad (15)$$

Minimum Column variance:
$$mcv = \min(v) \qquad (16)$$

4.3 Object localization

In chapter 3.1 the environment localization of significant points was introduced. The object localization is based on the selection of more significant environments.

The localization uses a hierarchal structure for fast object localization. The level of the hierarchy denotes the number of significant points commonly assigned between environments in reference and search image. Each assignment reduces the search area in the next level.

If the search fails in the current level, the search procedure returns to the previous level and selects a new assignment. A search fails if the value of the cumulative correlation of all assigned environments in reference and search image will be below a certain level.

Localization of partially hidden Objects can simply be handled by the limitation of searched assignments. The hierarchal approach accelerates the comparison of the Jet matrices, too. The estimated scaling and rotation value of the previous assigned points reduces the potential shift values which must be considered in the Jet matrices.

5 Examples for object localization

This chapter shows the object localization of different objects. For the better comparison all images are scaled to 40% of original size.

The object localization is nearly rotation and scale invariant, can localize overlapped objects and is robust against 3D viewpoint variation and different lighting.

All object localizations were made with the same parameter values. The size of the reference Jet matrices was 9x6 and the size of the search Jet matrices 6x6.

Referenz Object with marked points	Search Results

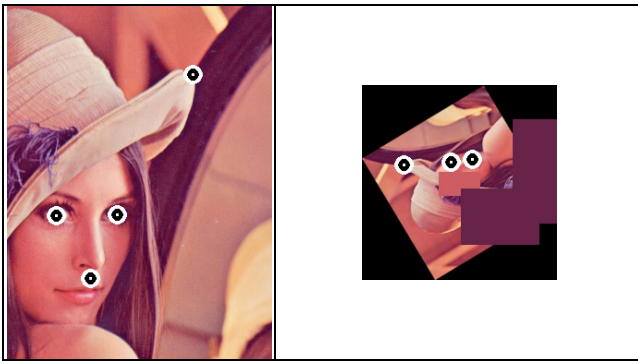
Fig.7 Test images from the SOIL-47 Database

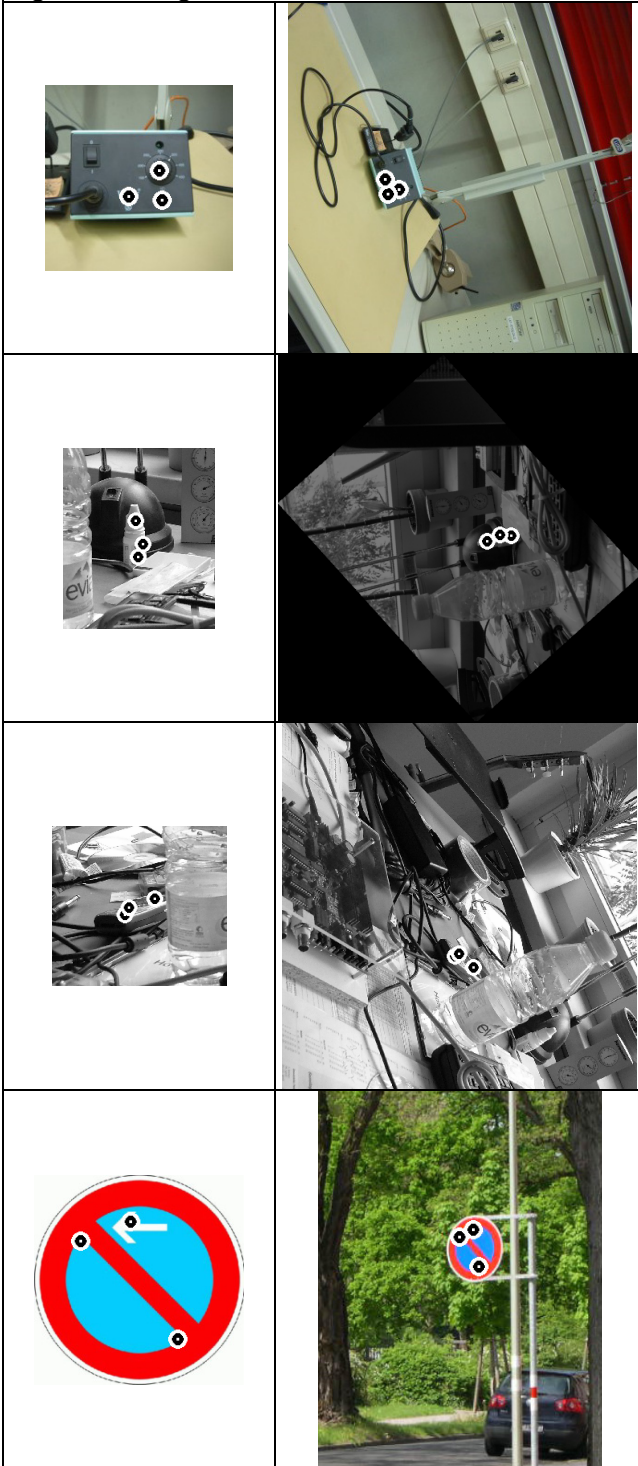Fig.8 Test images from the SIPI Database

Fig.9 Test images from our database

6 Conclusion

As shown in chapter 4, the Jet matrix based object localization method provides excellent results. The object localization is robust and flexible. Different object localization problems can solved without changing the parameters. By use of color information the object localization can be improved with little additional effort.

On our test system (P4, 2.66 GHz, 512MB, WinXP, VS.net 7.1) object localization in an Image with 320x240 Pixel needs approximately 0.5 sec. The calculation of the Jet matrixes takes approximately 70% of the complete calculation time. The Gabor filter convolution algorithm in use can be accelerated by parallel computation. Current research activities are dealing with the transformation of the Gabor filter algorithm into Field Programmable Gate Array.

References:

[1] Kyrki, V.; Kamarainen, J-K.; Kälviäinen, H., Simple Gabor feature space for invariant object recognition, Pattern Recognition Letters, Vol.25, No.3, 2004, pp.311-318

[2] Heintz, R; Invariant Properties of Gaborfilter; Internal Report; 2005

[3] Heintz, R; Schäfer, Gerhard; Lokale invariante Objektlokalisierung mittels Gaborfiltern; GMA-Kongress; Vol.X, 2005, pp. XX-XX

[4] Messer, K.; Kittler, J.;Sadeghi, M. and others, Face Authentication Test on the BANCA Database, 17th International Conference on Pattern Recognition, Vol.4, 2004, pp.523-532

[5] Grigorescu, S.E.; Petkov, N.; Kruizinga, P., Comparison of texture features based on Gabor filters, IEEE Transactions on Image Processing, Vol.11, No.10, 2002, pp. 142-147

[6] Young, Ian T., vanVliet, Lucas J., van Ginkel, Michael, Recursive Gabor Filtering, IEEE Trans. Sig.Proc. ,Vol.50, No.11, 2002, pp.2799–2805

[7] Frigo, Matteo, Johnson, Steven-G., The Design and Implementation of FFTW3, IEEE Proc. Sig.Proc., Vol.93, No.2, 2005, pp.216–231

[8] Weber, G. Allan, The USC-SIPI Image Database: Version 5, USC-SIPI Image Database, USC-SIPI Report, Vol. 315,1997

[9] Burianek, Jan; Ahmadyfard, Alireza, Kittler, Josef, SOIL-47: Surrey Object Image Library, http://www.ee.surrey.ac.uk/Research/VSSP/demos/colour/soil47/, 2005

[10] Jolliffe, Ian T., Principal Component Analysis, Springer-Verlag, 1986