Acta Automatica Sinica
Vol. 40, No. 6 June, 2014

Available online at www.sciencedirect.com
ScienceDirect

A Global Image Feature Construction Method Based on Local Jet Structure

XIE Jin[1, 2] CAI Zi-Xing[1]

Abstract This article presents a novel and robust feature descriptor called the multi-scale autoconvolution on local jet structure (MSALJS), which is quasi-invariant to affine transformation. The MSALJS, a global image feature descriptor, is based on the derivatives that describe the image local structure to compute the multi-scale autoconvolution moment. Experimental data demonstrate that the MSALJS can be used in practical applications in which the object is deformed in various ways, such as particular occlusion, view angle change, and so on.

Key words Multi-scale autoconvolution (MSA), affine transformation, local jet structure, invariant

Citation Xie Jin, Cai Zi-Xing. A global image feature construction method based on local jet structure. *Acta Automatica Sinica*, 2014, **40**(6): 1148−1155

The recognition of objects and patterns that are deformed in various ways has been a goal of many recent studies. The typical object deformation includes geometric transformation and illumination change. There are many approaches to recognize these deformed objects. The approach using invariant features seems to be the most promising. To classify features into local features and global features, there are many local feature descriptors and global feature descriptors that are invariant to deformation and applied successfully to two-dimensional or three-dimensional object recognition and matching.

A more general situation in the practice is the projective transformation, which means photographing a planar scene using a pinhole camera whose optical axis is not perpendicular to the scene. If an object is small or far from the camera, a local perspective transformation of a smooth surface can be approximated by affine transformation. Moment invariants are special functions of image moments. Since the study of Hu[1] in 1962, many novel approaches to global affine invariance moments have been put forward[2−6]. Under the scale space theory, local affine method and derived differential invariants have been addressed[7−13]. Specifically, Koenderink et al. thoroughly studied the properties of local derivatives and derived differential invariants[14−17].

On the basis of the probabilistic interpretation of the image function, Rahtu et al.[6] presented an affine invariant image transform called multi-scale autoconvolution (MSA). Because it is calculated from the whole image, moments are very sensitive to some typical deformations. At the same time, these invariant features are absolute to affine transformation and cannot be used well in practical applications. To overcome these faults, our idea is to get a novel feature quasi-affine invariant, which is not sensitive to particular occlusion, illumination change, view angle change, and so on. In this article, we proposed a novel feature descriptor, the multi-scale autoconvolution on local jet structure (MSALJS) by analyzing the Hessian matrix of each point and by computing MSA moment, without direct intensity computing like Rahtu et al.[6].

The organization of the paper is as follows: Section 1 introduces the MSA moment, Section 2 analyzes the local jet structure and illustrates how to build a global feature descriptor, Section 3 provides some experiments, and Section 4 draws the conclusion.

Manuscript received September 2, 2011; accepted August 23, 2013
Supported by National Natural Science Foundation of China (90820302) and Scientific Research Fund of Hunan Provincial Education Department (12C0202)
Recommended by Associate Editor DAI Qiong-Hai
1. School of Information Science and Engineering, Central South University, Changsha 410083, China 2. College of Polytechnic, Hunan Normal University, Changsha 410081, China

1 MSA moment

In this section, we define some interrelated concepts and introduce the basic idea of MSA moment construction[6].

An affine transformation ψ is defined as $\psi(\boldsymbol{x}) = T\boldsymbol{x} + \boldsymbol{t}$, where $\boldsymbol{x}$, $\boldsymbol{t} \in \mathbf{R}^2$ and T is a 2×2 nonsingular matrix. The inverse transformation of ψ is presented as $\psi^{-1}(\boldsymbol{x}) = T^{-1}\boldsymbol{x} - T^{-1}\boldsymbol{t}$. Suppose $f(\boldsymbol{x})$ is a gray image intensity function. When applying an affine transformation ψ to the image f, we get a new gray image function f', where $f'(\boldsymbol{x}) = f \circ \psi^{-1}(\boldsymbol{x}) = f(T^{-1}\boldsymbol{x} - T^{-1}\boldsymbol{t})$.

From a mathematical point of view, the affine invariant feature I is a function defined on f, which produces the same value for f and f' (the affine transformation ψ version of f). This characteristic can be also expressed as $I(f) = I(\psi(f))$ for any affine transformation ψ.

Consider three points ($\boldsymbol{x}_0$, $\boldsymbol{x}_1$, and $\boldsymbol{x}_2$) of the image to define a new point $\boldsymbol{u}$ (the corresponding variable also noted as $\boldsymbol{u}$):

$$\boldsymbol{u} = \alpha(\boldsymbol{x}_1 - \boldsymbol{x}_0) + \beta(\boldsymbol{x}_2 - \boldsymbol{x}_0) + \boldsymbol{x}_0 \tag{1}$$

where (α, β) are the coordinates for point $\boldsymbol{u}$ in the space spanned by vectors $\boldsymbol{x}_1 - \boldsymbol{x}_0$ and $\boldsymbol{x}_2 - \boldsymbol{x}_0$ and with the origin at $\boldsymbol{x}_0$.

Applying an affine transformation ψ to points $\boldsymbol{x}_0$, $\boldsymbol{x}_1$, and $\boldsymbol{x}_2$, the affine transformation versions of these sample points are noted as points $\boldsymbol{x}'_0$, $\boldsymbol{x}'_1$, and $\boldsymbol{x}'_2$, where $\boldsymbol{x}'_0 = T\boldsymbol{x}_0 + t$, $\boldsymbol{x}'_1 = T\boldsymbol{x}_1 + t$, and $\boldsymbol{x}'_2 = T\boldsymbol{x}_2 + t$. When carrying out affine transformation ψ for point $\boldsymbol{u}$, one can get a corresponding point $\boldsymbol{u}'$:

$$\boldsymbol{u}' = \alpha(\boldsymbol{x}'_1 - \boldsymbol{x}'_0) + \beta(\boldsymbol{x}'_2 - \boldsymbol{x}'_0) + \boldsymbol{x}'_0 = T\boldsymbol{u} + \boldsymbol{t} \tag{2}$$

When applying an affine transformation ψ to image $f(\boldsymbol{u})$, we can get

$$f'(\boldsymbol{u}') = f(T^{-1}\boldsymbol{u}' - T^{-1}\boldsymbol{t}) = f(\boldsymbol{u}) \tag{3}$$

The result $f'(\boldsymbol{u}') = f(\boldsymbol{u})$ in (3) shows that any order moment of $f(\boldsymbol{u})$ is invariable in an affine transformation of f. The general kth-order moment of $f(\boldsymbol{u})$, denoted by M_f^k, is defined as

$$M_f^k = \int_{\mathbf{R}^2} f(\boldsymbol{u})^k \cdot p(\boldsymbol{u}) \mathrm{d}\boldsymbol{u} = \mathrm{E}[f(\boldsymbol{u})^k] \tag{4}$$

where $f(\boldsymbol{u})$ is the probability density function. Therefore, the MSA moment $M(\alpha,\beta)$, defined as the first moment of $f(\boldsymbol{u})$ as shown in (5), is an affine invariant moment,

$$M(\alpha,\beta) = \mathrm{E}[f(\boldsymbol{u})] = \int_{\mathbf{R}^2} f(\boldsymbol{u})p(\boldsymbol{u})\mathrm{d}\boldsymbol{u} \tag{5}$$

To compute $M(\alpha,\beta)$ in (5), one needs to know the probability density function $p(\boldsymbol{u})$ of variable $\boldsymbol{u}$. Assuming that $\gamma = 1-\alpha-\beta$, the random variable $\boldsymbol{u}$ in (1) is computed as follows:

$$\boldsymbol{u} = \alpha\boldsymbol{x}_1 + \beta\boldsymbol{x}_2 + \gamma\boldsymbol{x}_0 \tag{6}$$

Let $p(\boldsymbol{x}) = \frac{1}{||f||}f(\boldsymbol{x})$ be the probability density function of f; then the probability density function $f(\boldsymbol{u})$ of variable $\boldsymbol{u}$ in (6) can be calculated using

$$p(\boldsymbol{u}) = \int_{\mathbf{R}^2}\int_{\mathbf{R}^2}\left(\frac{1}{\alpha^2}\frac{1}{||f||}f\left(\frac{\boldsymbol{x}}{\alpha}\right)\right)\left(\frac{1}{\beta^2}\frac{1}{||f||}f\left(\frac{\boldsymbol{y}}{\beta}\right)\right)\times \left(\frac{1}{\gamma^2}\frac{1}{||f||}f\left(\frac{\boldsymbol{u}-\boldsymbol{x}-\boldsymbol{y}}{\gamma}\right)\right)\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y} \tag{7}$$

By substituting (7) into (5), the MSA moment $M(\alpha,\beta)$ can be written as

$$M(\alpha,\beta) = \frac{1}{\alpha^2\beta^2\gamma^2}\frac{1}{||f||^3}\int_{\mathbf{R}^2}\int_{\mathbf{R}^2}\int_{\mathbf{R}^2} f(\boldsymbol{u})f\left(\frac{\boldsymbol{x}}{\alpha}\right)\times f\left(\frac{\boldsymbol{y}}{\beta}\right)f\left(\frac{\boldsymbol{u}-\boldsymbol{x}-\boldsymbol{y}}{\gamma}\right)\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}\mathrm{d}\boldsymbol{u} \tag{8}$$

The Fourier transform of $f(\boldsymbol{x})$ is defined as

$$F(\boldsymbol{\omega}) = \int_{\mathbf{R}^2} f(\boldsymbol{x})\mathrm{e}^{-\mathrm{j}2\pi\boldsymbol{x}\omega}\mathrm{d}\boldsymbol{x} \tag{9}$$

On the basis of the Fourier transform and its similarity property, the MSA moment $M(\alpha,\beta)$ can be rapidly calculated using a single integral as follows:

$$M(\alpha,\beta) = \frac{1}{F(0)^3}\int_{\mathbf{R}^2} F(-\boldsymbol{\omega})F(\alpha\boldsymbol{\omega})F(\beta\boldsymbol{\omega})F(\gamma\boldsymbol{\omega})\mathrm{d}\boldsymbol{\omega} \tag{10}$$

According to (10), each pair (α,β) can produce one invariant for f, which does not change under affine transformation.

The moment $M(\alpha,\beta)$ has the following symmetries:

$$M(\alpha,\beta) = M(\beta,\alpha) \tag{11}$$

$$M(\alpha,\beta) = \frac{1}{\alpha^2}M(\frac{1}{\alpha},-\frac{\beta}{\alpha}) \tag{12}$$

These symmetrical characteristics can be used to find small regions in the (α,β) plane, which determine all the MSA transform values. Rahtu et al.[6] selected 29 (α,β) pairs to obtain the feature vector in the classification experiments, in which the actual values of α and β are limited to a small $M(\alpha,\beta)$ plane region.

Without extracting boundaries or interest points, the transformed global values of an image can be directly used as a descriptor for affine invariant pattern classification. However, in the study of Rahtu et al.[6], $f(\boldsymbol{x})$ was an image intensity function, so this method cannot perform well under situations such as view change, illumination change, or partial occlusion.

2 Constructing a global feature descriptor based on local jet structure

2.1 Motivation

MSA has its discriminative capacity, which is applicable to affine invariant feature extraction. However, as an absolute affine invariant moment, MSA cannot be used well in practical applications, such as nonlinear illumination change and view change, because of the feature vector being directly calculated from the image intensity. One of the key ideas of our proposed feature descriptor is that the relative ordering of the pixel intensities in a local patch remains unchanged or stable under monotonically increasing brightness changes. The differential structure of the luminance at a point can describe the local geometry of the point. The Hessian matrix can capture important properties of a local image structure. Hence, we combined computing the Hessian matrix for every point with calculating the global MSA value to obtain a quasi-affine invariant descriptor.

2.2 Local jet structure analysis

The local geometry of an image point depends on the differential structure of the luminance at this given point. Let $L(\boldsymbol{w})$ be a two-dimensional smooth gray image, and let $L_{i_1 i_2\cdots i_k}$ be the tensor of rank k formed by the kth-order partial derivatives $\frac{\partial^k L}{\partial w_{i_1}\cdots\partial w_{i_k}}$, where i_j is labeled 1 or 2 for each $j = 1,\cdots,k$. According to the Taylor series expansion, the neighborhood luminance information of a given point with coordinate $\boldsymbol{w}$ can be described by a series of differential operator[15]

$$L(\boldsymbol{w}+\delta\boldsymbol{w}) = \sum_{k=0}^{\infty}\frac{1}{k!}L_{i_1\cdots i_k}(\boldsymbol{w})\delta w_{i_1}\cdots\delta w_{i_k} \tag{13}$$

The local jet of order r of the image at a given point with illumination $L(\boldsymbol{w})$ and notation $J^r[L](\boldsymbol{w})$ can be represented by a set of $L_{i_1 i_2\cdots i_k}$ as follows[16]:

$$J^r[L](\boldsymbol{w}) = \{L_{i_1\cdots i_k}(\boldsymbol{w}) | k = 0,\cdots,r\} \tag{14}$$

In the scale space theory, given any continuous signal f, the linear scale-space representation L of signal f can be gained by convolution with Gaussian kernels as $L(\cdot;\sigma) = g(\cdot;\sigma) * f(\cdot)$, where $L(\cdot;0) = f(\cdot)$, σ is a scale factor, and operator $*$ denotes convolution. Koenderink and van Doorn[15] demonstrated that the nth derivative of L, which is the blurred form of the signal f, equals the convolution of both f and the nth derivative of Gaussian kernel. Fig. 1 displays a two-dimensional Gaussian kernel (with $\sigma = 2$) and its second-order partial derivative forms. The partial derivative of $L(x,y)$ can be computed as

$$\frac{\partial^{n+m}}{\partial x^n\partial y^m}L(x,y;\sigma) = L(x,y;0) * \frac{\partial^{n+m}}{\partial x^n\partial y^m}g(x,y;\sigma) \tag{15}$$

The order of the jet can determine the amount of geometry represented. The zeroth-order jet $J^0[L]$ equals the local luminance. The first-order local jet, $J^1[L]$, consists of gradients L_x and L_y. A second-order jet allows to detect "line orientation", and a third-order jet would find "line curvature". The shape of a point neighborhood can be estimated by the Hessian matrix. On the basis of the second-order local jet, the Hessian matrix is defined by

$$H(x,y;\sigma_{\mathrm{D}}) = \begin{bmatrix} L_{xx}(x,y;\sigma_{\mathrm{D}}) & L_{xy}(x,y;\sigma_{\mathrm{D}}) \\ L_{xy}(x,y;\sigma_{\mathrm{D}}) & L_{yy}(x,y;\sigma_{\mathrm{D}}) \end{bmatrix} \tag{16}$$