Optik 124 (2013) 3806–3811

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Extended Shape of Gaussian: Feature descriptor based on element set of matrix Lie group

Feng Cheng*, Zuxi Wang, Dehua Li

Institute for Pattern Recognition & Artificial Intelligence, Huazhong University of Science and Technology, Wuhan 430074, Hubei, PR China

ARTICLE INFO

Article history:
Received 22 June 2012
Accepted 15 November 2012

Keywords:
SOG
Region covariance
Feature descriptor
Lie group
Visual tracking

ABSTRACT

In this paper, we extend the feature descriptor known as Shape of Gaussian (SOG) and we call the new descriptor Extended Shape of Gaussian (ESOG). SOG has a matrix Lie group structure, it use the geodesic distance to measure the difference between two features. First, we decompose geodesic distance on the Lie algebra into two orthogonal components. By adjusting the weights of components, we get a distance sequence. Then we identify that every element in the sequence corresponds to an element of the original Lie group, a matrix. All these matrices form ESOG. Thus the new descriptor utilizes a matrix set rather than one matrix to describe feature. In this view, SOG and region covariance are both special element of ESOG. So we can choose different element from it for different application. Noting that different elements in the ESOG describe a signal in a different view, we propose an adaptive method to select appropriate ESOG element for visual tracking. The element selected by this method is called Adaptive SOG (ASOG). ASOG keeps the advantages of both SOG and region covariance and has better accuracy and robustness under different conditions. Experiments show the tracking results compared with SOG.

1. Introduction

Feature descriptor is always important for machine learning and computer vision. An appropriate descriptor brings the advantages such as better discrimination, robustness and the decrease of calculating cost.

In order to describe the characteristics of signal, some statistical methods have been proposed, such as histogram [1–3] and a variety of other methods based on it, such as histogram of oriented gradients [4] and LBP histogram [5] and so on. However, when we try to use more feature, the computational cost will grow exponentially and become unbearable.

In 2006, Tuzel et al. [6] proposed a descriptor called region covariance. This method use covariance matrix to describe signal. When using n features to represent a signal, the covariance matrix only need a $(n+1)n/2$ dimensional vector, however, the histogram method needs n^b dimension, where b is the number of bins of histogram. Tuzel also proposed a method to calculate the covariance matrix using integral image, which reduces calculating cost greatly. It has been used widely in tracking, detection and classification [7–9]. In 2009, Gong et al. [10] proposed a descriptor called SOG. This method is a natural extension of region covariance. It combines mean vector and covariance matrix together while preserves the Riemannian structure. Because of the introduction of mean vector, SOG has an improved performance, especially when region covariance is not reliable.

However, the mean vector and covariance describe signal from different perspectives, the reliability of them changes in different circumstances. Thus the analysis of SOG and an appropriate adjustment strategy is necessary. So we analyze the properties of SOG, factorize it in two orthogonal directions. Based on that, we proposed extend SOG descriptor (ESOG) and an adaptive method to select the appropriate descriptors from ESOG. The selected descriptors are called as Adaptive SOG (ASOG).

The rest of the paper is organized as follows: In Section 2 we present a short review of the SOG feature descriptor. In Section 3 we analyze SOG and give the definition of ESOG. In Section 4 we propose the method to choose ASOG. Experiments and conclusion are given in Sections 5 and 6 respectively.

2. Shape of Gaussian (SOG)

SOG descriptor is the product of Shape of Signal Probability Density Function (SOSPDF). SOSPDF treats a *pdf* (probability density function) as a geometry object (i.e. a curve of a surface) and then characterize the *pdf* with the shape of the object. Histogram and region covariance are both some kind of SOSPDF.

In a SOSPDF view, we need to specify two component to describe a signal. The first one is the signal channels. A n dimensional channel signal is modeled as a n dimensional feature vector $X \in \Re^n$. The

* Corresponding author. Tel.: +86 27 87544769.
E-mail address: chfwh@qq.com (F. Cheng).

0030-4026/$ – see front matter
http://dx.doi.org/10.1016/j.ijleo.2012.11.071

element of the vector may be the raw original data or values of other features. We choose the vector elements according to our purpose. For example, in order to describe an image, the typical X may have a form as

$$X = [x, y, \mathrm{R}, \mathrm{G}, \mathrm{B}, |I_x|, |I_y|, \sqrt{I_x^2 + I_y^2}] \tag{1}$$

where x and y are horizontal and vertical coordinate respectively. R, G and B correspond to values of the three color channel. I_x and I_y are the gradients in horizontal and vertical direction. The second component of SOSPDF is the *pdf* estimation. Different choice of estimation method will produce different SOSPDF features. SOG estimate the shape of *pdf* using full parameterized multivariate Gaussian:

$$f(X) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(X-\mu)^T \Sigma^{-1}(X-\mu)} \tag{2}$$

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{3}$$

$$\Sigma = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)(x_i - \mu)^T \tag{4}$$

where μ is mean value of X, and Σ is the covariance matrix. So any n dimensional feature vector can be written as:

$$X = PX_0 + \mu \tag{5}$$

where X_0 is a n dimensional standard multivariate Gaussian distributed feature vector with zero mean and standard variance 1. Matrix P satisfies the equation $\Sigma = PP^T$. P is unique if it is defined as the solution of Cholesky factorization of Σ. Then the SOG of vector X is defined as a positive definite lower triangular affine transform (PDLTAT) matrix M.

$$M = \begin{bmatrix} P & \mu \\ 0 & 1 \end{bmatrix} \tag{6}$$

$$\begin{bmatrix} X \\ 1 \end{bmatrix} = \begin{bmatrix} P & \mu \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_0 \\ 1 \end{bmatrix} = M \begin{bmatrix} X_0 \\ 1 \end{bmatrix} \tag{7}$$

Thus, given feature vector X, a unique PDLTAT matrix $M(X)$ describes the relationship between X and X_0.

PDLTAT matrices form a matrix Lie group [10]. A Lie group is a group which is also a differentiable manifold such that the group operations, multiplication and inverse, are differentiable maps. The distance between two elements of a Lie group is measured by the minimum length curve between them. That curve is called geodesic. The tangent space of the identity element I forms Lie algebra. The exponential map relates elements of Lie group with points in Lie algebra. And the geodesic length d between two group elements M_1 and M_2 is given by

$$m = \log(M) \tag{8}$$

$$M = \exp(m) \tag{9}$$

$$d(M_1, M_2) = ||\log(M_1^{-1}M_2)|| = ||m_2 - m_1|| \tag{10}$$

$|| \; ||$ is L_2 norm.

3. Extended Shape of Gaussian (ESOG)

3.1. Orthogonal component of geodesic length $d(M_1, M_2)$

According to Lie group theory, Lie algebra of n dimensional PDLTAT is the set of matrices

$$m = \begin{bmatrix} U & v \\ 0 & 0 \end{bmatrix} \tag{11}$$

where U is a $n \times n$ lower triangular matrix, and v is a $n \times 1$ vector.

According to the definition of matrix exp operation and the exponential map between Lie algebra and Lie group, we know

$$M = \exp(m) = \sum_{k=0}^{\infty}\frac{m^k}{k!} = \begin{bmatrix} e^U & rv \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} P & \mu \\ 0 & 1 \end{bmatrix} \tag{12}$$

So we get

$$P = e^U \tag{13}$$

$$\mu = rv \tag{14}$$

where r is a function of matrix P. Here we select the inner product of two matrix $A^{m\times n}$ and $B^{m\times n}$ as

$$\langle A, B\rangle = \sum_{i=1}^{m\times n}\sum_{j=1}^{m\times n} a_{i,j}b_{i,j} \tag{15}$$

The norm of a matrix can be defined as Frobenius norm

$$||A||_F = \sqrt{\sum_{i=1}^{m\times n}\sum_{j=1}^{m\times n} a_{i,j}a_{i,j}} \tag{16}$$

Because the matrix norm is equivalent to each other, so we use Frobenius norm instead of L_2 norm, then the distance between matrix M_1 and M_2 can be written as the sum of two orthogonal vector as

$$d(M_1, M_2) = ||m_2 - m_1|| = \sqrt{||m_2 - m_3||^2 + ||m_1 - m_3||^2} \tag{17}$$

$$m_1 = \begin{bmatrix} U_1 & v_1 \\ 0 & 0 \end{bmatrix}, \quad m_2 = \begin{bmatrix} U_2 & v_2 \\ 0 & 0 \end{bmatrix}, \quad m_3 = \begin{bmatrix} U_1 & v_2 \\ 0 & 0 \end{bmatrix} \tag{18}$$

According to (15), it is easy to identify that $\langle m_2 - m_3, m_1 - m_3\rangle = 0$, which means they are orthogonal. Given X_1 and X_2, according to (12)–(14), we know that $||m_2 - m_3||$ only measures the difference between the covariance matrix Σ_1 and Σ_2. And $||m_1 - m_3||$ mainly measures the difference of μ_1 and μ_2, as compared to μ, the matrix r usually has a much smaller influence on vector v. Then the geodesic length d can be written as

$$d(M_1, M_2) = ||\vec{d}(m_1, m_3) + \vec{d}(m_3, m_2)|| = ||\vec{d}_\mu + \vec{d}_\Sigma||, \vec{d}_\mu \perp \vec{d}_\Sigma \tag{19}$$

3.2. Definition of ESOG

The most outstanding advantage of SOG, compared to region covariance, is that SOG take the mean vector into account while maintaining the Riemannian structure. When region covariance fails to distinguish two signals, information from mean vector will help, thus SOG is more effective than region covariance.

On the basis of Section 3.1, we can factorize the distance between SOGs in two orthogonal directions. So it is possible for us to adjust the weight of mean vector and covariance matrix according to their ability to discriminate signals. The distance between