Joint Metric Learning on Riemannian Manifold of Global Gaussian Distributions

Qinqin Nie , Pengfei Zhu , Qinghua Hu , Hao Cheng

College of Intelligence and Computing Tianjin University

Introduction

In many computer vision tasks, images or image sets can be modeled as a Gaussian distribution to capture the underlying data distribution. The challenge of using Gaussians to model the vision data is that the space of Gaussians is not a linear space. From the perspective of information geometry, the Gaussians lie on a specific Riemannian Manifold. In this paper, we present a joint metric learning (JML) model on Riemannian Manifold of Gaussian distributions. The distance between two Gaussians is defined as the sum of the Mahalanob is distance between the mean vectors and the log-Euclidean distance (LED) between the covariance matrices. We formulate the multi-metric learning model by jointly learning the Mahalanob is distance and the log-Euclidean distance with pairwise constraints. Sample pair weights are embedded to select the most informative pairs to learn the discriminative distance metric. Experiments on video based face recognition, object recognition and material classification show that JML is superior to the state-of-the-art metric learning algorithms for Gaussians. promising tracking performance on several publicly available datasets.

Global Gaussian Distributions Modeling

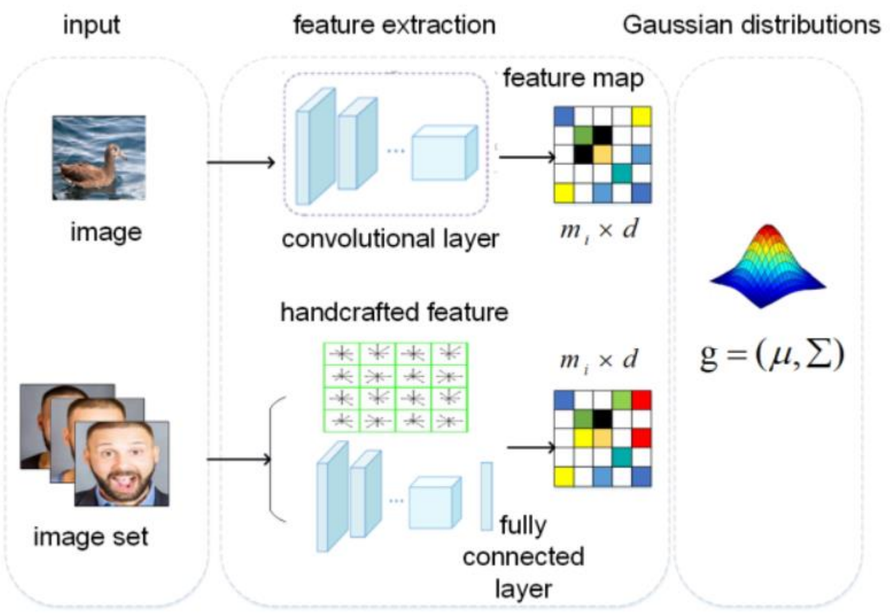

Fig. 1: Global Gaussian distribution modelling of images and image sets. For an image set, handcrafted features and deep features (fully connected layer) can be extracted for each image. For an image, we use convolutional neural network to extract the output of the final convolution layer as deep features. Thus, a feature matrix $X \in R^{m_i \times d}$ can be extracted for both an image or image set.

To build a Gaussian distribution , the first order statistics (mean) μ and the second order statistics (covariance matrix) Σ can be computed as follows:

$$\mu = \frac{1}{m}\sum_{i=1}^{m} x_i,$$

$$\Sigma = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu)^T(x_i - \mu),$$

Joint Metric Learning

Given an image or image set, a feature matrix $X_i^{m_i \times d}$ can be extracted by using hand-crafted features or deep features. By computing the mean and covariance matrix of X_i, an image or image set can be modelled by a Gaussian distribution $N(\mu, \Sigma)$. The Mahalanob is distance between the means can be defined as:

$$d_u = (\mu_i - \mu_j)^T A(\mu_i - \mu_j) = tr(AT_{ij}),$$

geodesic distance between Σ_i , Σ_j on original SPD manifold can be represented as:

$$d_{\mathbf{\Sigma}}(\mathbf{\Sigma}_i, \mathbf{\Sigma}_j) = \left\| \mathbf{M}^T \log(\mathbf{\Sigma}_i)\mathbf{M} - \mathbf{M}^T \log(\mathbf{\Sigma}_j)\mathbf{M} \right\|_F^2,$$

Considering the difference between the first-order and second-order statistics, the weighted joint distance between two Gaussians is defined as follows:

$$d(g_i, g_j) = r d_\mu(g_i, g_j) + (1-r) d_{\mathbf{\Sigma}}(g_i, g_j),$$

For metric learning, a large number of sample pairs are generated firstly. However, not all sample pairs are necessary for metric learning and the importance of samples also varies greatly during the metric learning process. To this end, we embed the weight of sample pairs into the metric learning mode as follows:

$$\min \left\{ \begin{array}{l} \sum_{ij\in P}(rtr(\alpha_{ij}^v \mathbf{A}\mathbf{T}_{ij}) + (1-r)tr(\alpha_{ij}^v \mathbf{B}\mathbf{H}_{ij}))+ \\ \sum_{ij\in N}(rtr(\beta_{ij}^v \mathbf{A}^{-1}\mathbf{T}_{ij}) + (1-r)tr(\beta_{ij}^v \mathbf{B}^{-1}\mathbf{H}_{ij})) \end{array} \right\}$$

$$s.t. \sum_{ij\in P}\alpha_{ij} = 1, \sum_{ij\in N}\beta_{ij} = 1, \forall ij, \alpha_{ij} \geq 0, \beta_{ij} \geq 0.$$

where α_{ij} and β_{ij} are the weights for similar pairs and dissimilar pairs, respectively.

We get the closed-form solution to α_{ij} and β_{ij} as follows:

$$\alpha_{ij} = \frac{(rtr(\mathbf{A}\mathbf{T}_{ij}) + (1-r)tr(\mathbf{B}\mathbf{H}_{ij}))^{\frac{-1}{v-1}}}{\sum_{ij\in P}(rtr(\mathbf{A}\mathbf{T}_{ij}) + (1-r)tr(\mathbf{B}\mathbf{H}_{ij}))^{\frac{-1}{v-1}}}.$$

$$\beta_{ij} = \frac{(rtr(\mathbf{A}^{-1}\mathbf{T}_{ij}) + (1-r)tr(\mathbf{B}^{-1}\mathbf{H}_{ij}))^{\frac{-1}{v-1}}}{\sum_{ij\in N}(rtr(\mathbf{A}^{-1}\mathbf{T}_{ij}) + (1-r)tr(\mathbf{B}^{-1}\mathbf{H}_{ij}))^{\frac{-1}{v-1}}}.$$

Last , We get the final solution to *A* and *B* as follows:

$$\mathbf{A}_{final} = \left(\mathbf{P1} + \lambda \mathbf{A}_0^{-1}\right)^{-1} \#_t \left(\mathbf{N1} + \lambda \mathbf{A}_0\right),$$

$$\mathbf{B}_{final} = (\mathbf{P2} + \lambda \mathbf{A}_0^{-1})^{-1} \#_t (\mathbf{N2} + \lambda \mathbf{A}_0).$$

The algorithm of joint metric learning on Riemannian manifold of Gaussian distributions is summarized in Algorithm 1.

Algorithm 1 JML

Input: A set of Gaussians, g_i, $i = 1, 2, ..., n$, $\varepsilon = 0.001$, $T = 5$
Output: Distance Metric matrix $\mathbf{A}$, $\mathbf{B}$;
Step1: Initialized $\mathbf{A}$, $\mathbf{B}$ using an identify matrix;
Step2: For t =1,2,…,T, repeat
2.1 Computer α_{ij} and β_{ij} using (14), (15) ;
2.2 Computer $\mathbf{A}$ and $\mathbf{B}$ using (26), (27) ;
2.3 If $t \geq 2$ and $\left|f(\mathbf{A}, \mathbf{B})^t - f(\mathbf{A}, \mathbf{B})^{t-1}\right| < \varepsilon$ go to **Step3**
Step3: Output the matrix $\mathbf{A}$, $\mathbf{B}$.

Results

To ensure a broad assessment of the different approaches, four common public datasets are used to conduct comparative experiments. ETH-80 contains 3,280 high-resolution color images. UIUC contains216 images and 18 categories, each category includes 12 images. YouTube Celebrities (YTC) dataset is a collection of celebrities from YouTube. Flickr material dataset (FMD) contains 1000 images and 10 categories.

Accuracies of different methods on four datasets

Method	ETH-80	FMD	UIUC	YTC
MMD	85.75	60.60	62.78	69.60
MDA	87.75	63.50	67.31	64.72
AHISD(linear)	72.50	46.72	55.37	64.65
AHISD(non-linear)	72.00	46.72	55.37	64.65
CHISD(linear)	79.75	47.52	65.09	67.24
CHISD(non-linear)	72.50	63.90	65.65	68.09
SPDML-AIRM	90.75	63.42	62.00	67.50
SPDML-Stein	90.75	66.80	61.12	68.10
LEML	93.50	66.60	62.96	69.85
JML(A)	77.50	68.40	75.56	70.96
JML(B)	90.00	64.88	66.05	61.99
JML(A+B)	**100.0**	**70.13**	**78.47**	**73.76**

Table 1: Accuracies of different methods on four datasets

Here JML(A) means that we only learn the distance metric using the fist-order statistics while JML(B) means that the distance metrics are learned for covariance matrices. JML(A+B) means that the distance metrics for both the mean and covariance matrices are learned jointly. LEML and SPDML learn distance metric based on SPD manifold. MMD and MDA are based on nonlinear manifold assumption while AHISD and CHISD are linear subspace based methods. Compare with the state-of-the-art metric learning algorithms, our proposed JML achieve superior performance.

Conclusions

In this paper, we proposed a joint metric learning (JML) model for global Gaussian distributions. The distance between Gaussians are defined as the sum of the sum of the Mahalanob is distance of the first-order statistics and the log-Euclidean distance(LED) of the second order statistics. JML effectively combines the information of the means and covariance matrices by joint metric learning and embeds the weights of Gaussian pairs into the learning model.