Journal of Southeast University (English Edition) Vol. 23, No. 2, pp. 202 – 205 June 2007 ISSN 1003—7985

Estimation of illumination chromaticity via adaptive reduced relevance vector machine

Ding Errui[1] Zeng Ping[1] Yao Yong[2] Wang Yifeng[1]

([1] Research Institute of Peripherals, Xidian University, Xi' an 710071, China)

([2] Research Center of Computer Information Applications, Xidian University, Xi' an 710071, China)

Abstract: A new regression algorithm of an adaptive reduced relevance vector machine is proposed to estimate the illumination chromaticity of an image for the purpose of color constancy. Within the framework of sparse Bayesian learning, the algorithm extends the relevance vector machine by combining global and local kernels adaptively in the form of multiple kernels, and the improved locality preserving projection (LLP) is then applied to reduce the column dimension of the multiple kernel input matrix to achieve less training time. To estimate the illumination chromaticity, the algorithm is trained by fuzzy central values of chromaticity histograms of a set of images and the corresponding illuminants. Experiments with real images indicate that the proposed algorithm performs better than the support vector machine and the relevance vector machine while requiring less training time than the relevance vector machine.

Key words: color constancy; illumination estimation; chromaticity histogram; adaptive reduced relevance vector machine

In color object recognition and content-based image retrieval, a variation of illumination results in a color shift of images which causes color descriptors to be too unstable for analysis. Without color stability, most applications mentioned above will be adversely affected even by small changes in the illumination. Therefore, it is necessary to estimate the illumination of images by automatic means to retain color constancy[1-4].

The color constancy processing[1] can be defined as the transformation of a source image taken under an unknown illuminant for an identical target image obtained in the same scene under a standard illuminant. Thus there are two steps in the process of color constancy[2,4], as illustrated in Fig. 1. The first step estimates the illuminant chromaticity. The second step corrects the image pixel wise with the estimated illuminant chromaticity. Many algorithms such as the finite-dimensional linear model of surface reflectance have successfully solved the problem of the second step. Special attention is paid to the problem of first step here.

In this paper, we propose an illumination estimation scheme based on a chromaticity histogram and an adaptive reduced relevance vector machine (AR-RVM). First, a simple but efficient chromaticity histogram in the form of fuzzy central values is obtained acquired. Secondly, the adaptive reduced relevance vector machine is constructed to map the relationship between chromaticity histograms of a set of images and the corresponding illuminants. Experimental results of 321 real images[4] indicate that the proposed scheme of estimation is feasible and the performance of the adaptive reduced relevance vector is better than that of the support vector machine and the relevance vector machine[5].

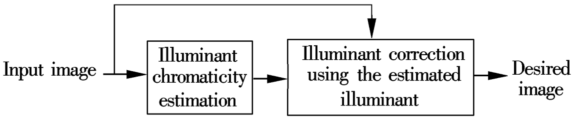

Fig. 1 Diagram of color constancy processing

1 Estimation of Illumination Chromaticity

As in Refs. [2 – 4], all the pixels in the images are projected into a chromaticity space. The color of images here is assumed to be an RGB signal. The rg chromaticity space is also used in the proposed algorithm.

$$r=\frac{R}{R+G+B},\qquad g=\frac{G}{R+G+B} \tag{1}$$

The space has the advantage that it requires no additional preprocessing since it is bounded between 0 and 1. If necessary, the implicit blue chromaticity component can easily be recovered: $b=1-r-g$.

Received 2006-11-28.

Foundation items: The National Natural Science Foundation of China (No. 60573139), the Innovation Foundation of Xidian University for Graduates (No. 05008).

Biographies: Ding Errui (1980—), male, graduate; Zeng Ping (corresponding author), male, professor, zp8637@126.com.

According to Eq. (1), the chromaticity of each pixel in an input image can be calculated and then a chromaticity histogram can be built. However, a common image usually contains tens of thousands of pixels, which will cause a massive chromaticity histogram that is, therefore, difficult to deal with. Refs. [2 - 3] proposed a sampling method to reduce the size of the chromaticity histogram, in which a 0 or a 1 of the final chromaticity of histogram indicates whether an RGB of rg chromaticity is present in the sampling bin. The sampling method can reduce the histogram to a large degree in the form of a sparse matrix. Nevertheless, the sampling step for different images is hard to be determined and even the sparse matrix usually contains hundreds of non-zero elements. Ref. [1] proposed a novel way of representing the original histogram using the central values. The algorithm needs to project the histogram into the x-axis and the y-axis and detect the central value using an ideal low pass filter, which is cumbersome in calculation. In this paper, a new method based on the latter method and fuzzy logic is presented, that is, a membership is assigned to the chromaticity of each pixel and a fuzzy central value is calculated. These fuzzy central values are believed to fully represent the color distribution and can easily be detected using fuzzy clustering algorithms such as FCM. In addition, the chromaticity of the illuminant is assumed to be the same as the chromaticity of a reference white patch under the same illuminant[2-3].

After the preparation of the training data is made, all that remains is to find a suitable regression tool to map the relationship between these two sets of data, namely, the chromaticity histogram and the chromaticity of the illuminant. A neural network was first proposed in Refs. [1 - 2] and it has proven to be superior to the previous traditional color constancy algorithms such as gray-world, white-patch, max-RGB and gamut mapping. Based on the principle of structural risk minimization (SRM), Xiong et al. [3] proposed an SVM-based algorithm, the results of which indicate its advantages over those of a neural network. On the basis of sparse Bayesian learning[5] (SBL), the relevance vector machine (RVM) proposed by Tipping has been proved to be obviously superior to the SVM in its prediction accuracy of regression with such advantages as the needlessness of cross validation despite its longer training time. Within the framework of SBL, a new regression algorithm is proposed here to improve the prediction accuracy and lessen the training time for illumination chromaticity estimation.

2 Relevance Vector Machine for Regression

Given a data set of input-target pairs $\{\boldsymbol{x}_j, t_j\}_{j=1}^N$, it is assumed that t is independent and data noise is subject to a Gaussian distribution with the variance σ^2 and the mean being zero. The prediction form of the relevance vector machine for regression[5] is given by

$$\hat{f}(\boldsymbol{x}) = \beta_0 + \sum_{j=1}^{N} k(\boldsymbol{x}, \boldsymbol{x}_j)\beta_j \tag{2}$$

where $k(\boldsymbol{x}, \boldsymbol{x}_j)$ are a kernel function and β_j are the regression coefficients. Based on Eq. (2), the likelihood of the complete data set can be written as

$$p(\boldsymbol{t} \mid \boldsymbol{\beta}, \sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \|\boldsymbol{t} - \boldsymbol{\Phi\beta}\|^2 \right\} \tag{3}$$

where $\boldsymbol{\beta}$ is the regression coefficient in a vector form, and $\boldsymbol{\Phi}$ is the $N(N+1)$ design matrix with $\boldsymbol{\Phi}_{nm} = k(\boldsymbol{x}_n, \boldsymbol{x}_m)$ and $\boldsymbol{\Phi}_{n1} = 1$.

From the prior Gaussian and Bayes' rule, the posterior distribution over the weights $\boldsymbol{\beta}$ is given by

$$p(\boldsymbol{\beta} \mid \boldsymbol{t}, \boldsymbol{\alpha}, \sigma^2) = (2\pi)^{-\frac{N+1}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \cdot \exp\left\{ -\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}) \right\} \tag{4}$$

where $\boldsymbol{\Sigma} = (\boldsymbol{\Phi}^{\mathrm{T}} \boldsymbol{B\Phi} + \boldsymbol{A})^{-1}$, $\boldsymbol{\mu} = \boldsymbol{\Sigma\Phi}^{\mathrm{T}} \boldsymbol{Bt}$ are the posterior covariance and mean, respectively; $\boldsymbol{A} = \mathrm{diag}(\alpha_0, \alpha_1, \ldots, \alpha_N)$ and $B = \sigma^{-2} I_N$; $\boldsymbol{\alpha} = \{\alpha_0, \alpha_1, \ldots, \alpha_N\}$ is a vector of hyperparameters.

The regression coefficient $\boldsymbol{\beta}$ of RVM can be estimated by the mean of the posterior using the maximum a posterior (MAP) which depends on the marginal likelihood for hyperparameters and noise variance. The marginal likelihood is given by

$$p(\boldsymbol{t} \mid \boldsymbol{\alpha}, \sigma^2) = (2\pi)^{-\frac{N}{2}} |\boldsymbol{B}^{-1} + \boldsymbol{\Phi A}^{-1} \boldsymbol{\Phi}^{\mathrm{T}}|^{-\frac{1}{2}} \cdot \exp\left\{ -\frac{1}{2} \boldsymbol{t}^{\mathrm{T}} (\boldsymbol{B}^{-1} + \boldsymbol{\Phi A}^{-1} \boldsymbol{\Phi}^{\mathrm{T}})^{-1} \boldsymbol{t} \right\} \tag{5}$$

3 Adaptive Reduced Relevance Vector Machine Approach

3.1 Adaptive combination of global and local kernels

In the relevance vector machine, only one type of kernel is used, the RBF kernel for instance. Research on kernels indicates that every kernel has its own advantages and disadvantages. Kernels are classified into two main types, namely local and global kernels[6]. Local kernels (for example, the RBF kernel) are only influenced by the data points near or close to each other while global kernels (for example, a polynomial kernel) are also influenced by data points far away from each other. A combination of these two types of kernels can retain the individual merits and overcome the demerits. A mixture of kernels is proposed to improve the performance of the SVM[6].

$$k_{\text{mixture}} = \rho k_{\text{global}} + (1-\rho) k_{\text{local}} \tag{6}$$

where k_{mixture} is the mixture of kernels, k_{global} is the global kernel, k_{local} is the local kernel and ρ is the mixing coefficient to be determined ($0 \leqslant \rho \leqslant 1$).

Usually, ρ is predetermined by experience and fixed in the implementation. However, from the theory of basis, either the global kernel or a local kernel is a candidate basis of solution space for any algorithm of regression, and a specific set of suitable bases are selected for signal construction in terms of an individual signal. Consequently, a fixed and bounded ρ is not reasonable. An adaptive mixture of kernels is then proposed, where a ratio variable s is used to control the relationship between global kernel and a local kernel.

$$k_{\text{mixture}} = k_{\text{global}} + s k_{\text{local}} \tag{7}$$

where s is the ratio variable with $s \geqslant 0$. Substituting Eq. (7) into Eq. (2), the new prediction of RVM becomes

$$\begin{aligned} \hat{f}(\boldsymbol{x}) &= \beta_0 + \sum_{j=1}^{N} k_{\text{mixture}}(\boldsymbol{x}, \boldsymbol{x}_j)\beta_j = \\ & \beta_0 + \sum_{j=1}^{N} [k_{\text{global}}(\boldsymbol{x}, \boldsymbol{x}_j) + s_j k_{\text{local}}(\boldsymbol{x}, \boldsymbol{x}_j)]\beta_j = \\ & \beta_0 + \sum_{j=1}^{N} \sum_{i=1}^{2} k_i(\boldsymbol{x}, \boldsymbol{x}_j)\beta_{ji} \end{aligned} \tag{8}$$

where

$$k_i(\boldsymbol{x}, \boldsymbol{x}_j) = \begin{cases} k_{\text{global}}(\boldsymbol{x}, \boldsymbol{x}_j) & i = 1 \\ k_{\text{local}}(\boldsymbol{x}, \boldsymbol{x}_j) & i = 2 \end{cases}$$

$$\beta_{ji} = \begin{cases} \beta_j & i = 1 \\ s_j \beta_j & i = 2 \end{cases}$$

The matrix form of Eq. (8) is

$$\hat{F}(X) = \boldsymbol{\Phi\beta} \tag{9}$$

It can be found that a regression algorithm based on an adaptive mixture of kernels can be reformulated into a regression algorithm based on multiple kernels (two kernels here). The matrix $\boldsymbol{\Phi}$ in Eq. (9) is different from that in Eq. (3). The size of $\boldsymbol{\Phi}$ here is $N(2N+1)$, whose number of columns doubles. Despite the improvement of prediction accuracy of the adaptive mixture of kernels, a larger $\boldsymbol{\Phi}$ tends to lengthen the training time.

3.2 Dimension reduction using improved LPP

As mentioned above, the column number of $\boldsymbol{\Phi}$ increases after the adaptive mixture of kernels is extended, so the straightforward way of lessening training time pressure is dimension reduction. Locality preserving projections[7] (LPP) is a new dimension reduction algorithm which has been proposed recently. Its aim is to find the inherent manifold embedded in a Euclidean way. Compared with traditional reduction algorithms such as PCA, LPP preserves the local neighborhood information and performs better than PCA.

Given a local similarity matrix $\boldsymbol{S}$ and an input matrix $\boldsymbol{X}$ (here $\boldsymbol{X} = \boldsymbol{\Phi}^{\mathrm{T}}$), LPP tries to find the optimal projections by solving the following minimization problem.

$$\begin{aligned} \boldsymbol{w}_{\text{opt}} &= \arg\min_{\boldsymbol{w}} \sum (\boldsymbol{w}^{\mathrm{T}}\boldsymbol{X}_i - \boldsymbol{w}^{\mathrm{T}}\boldsymbol{X}_j)^2 S_{ij} = \\ & \arg\min_{\boldsymbol{w}} \boldsymbol{w}^{\mathrm{T}}\boldsymbol{XLX}^{\mathrm{T}}\boldsymbol{w} \\ \text{s. t.} \quad & \boldsymbol{w}^{\mathrm{T}}\boldsymbol{XDX}^{\mathrm{T}}\boldsymbol{w} = 1 \end{aligned} \tag{10}$$

where $\boldsymbol{X}_i$ is the i-th column of $\boldsymbol{X}$, $\boldsymbol{D}$ is a diagonal matrix with $D_{ii} = \sum_j S_{ij}$, and $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{S}$ is the Laplacian matrix. As usual, the local similarity matrix $\boldsymbol{S}$ is created by an adjacency graph using either ε-neighborhoods or k nearest neighbors. Each method, however, has its own pros and cons when the data distribute non-uniformly. When the distribution is sparse, the method of ε-neighborhoods covers fewer data points and thus less local information while the method of k nearest neighbors does not. By the same token, when the distribution is dense, the method of ε-neighborhoods appears to be better than the method of k nearest neighbors. With a view to this, an improved method using both k nearest neighbors and ε-neighborhoods, meanwhile, is applied here to avoid the problem mentioned above.

Simple transformations can turn Eq. (10) into

$$\boldsymbol{XLX}^{\mathrm{T}}\boldsymbol{w} = \lambda \boldsymbol{XDX}^{\mathrm{T}}\boldsymbol{w} \tag{11}$$

where λ is the eigenvalue. Using Eq. (11), the reduced multiple kernel input matrix is

$$\boldsymbol{\Phi}_{\text{reduced}} = \boldsymbol{\Phi W}_{\text{opt}} \tag{12}$$

where $\boldsymbol{W}_{\text{opt}}$ is the matrix composed of the first Q eigenvectors of Eq. (11) and Q is the number of principal components. The size of $\boldsymbol{\Phi}_{\text{reduced}}$ is NQ ($Q < 2N+1$). Substituting $\boldsymbol{\Phi}_{\text{reduced}}$ for $\boldsymbol{\Phi}$ in Eq. (3), the integration of Eqs. (3) to (5) is the proposed algorithm here.

4 Experiments

In this section, the SVM and the RVM are trained together with the proposed algorithm of the ARRVM to assess the performance. The kernel for the SVM and the RVM is the RBF kernel with a width parameter equaling 0.20 while the kernels for the ARRVM are RBF kernels with the same width parameters and polynomial kernels with the degree being 2. The 321 SONY real images[4] taken under 11 different illuminants are used here and 3/4 (241) and 4/5 (257) of them are respectively selected as training samples while the remaining are test samples. Two basic error measures, the root mean square (RMS) errors for distance and angle [3-4], are used. The reduced column dimension of $\boldsymbol{\Phi}_{\text{reduced}}$ is 96 and 92, respectively. The estimation errors and training time comparisons are listed in Tab. 1 and Tab. 2.

Tab. 1 Comparison of various algorithms (ratio = 3/4)

Algorithm	RMS dis	RMS angle	CPU time/s
SVM	0.089 9	2.465 4	4.000 0
RVM	0.087 1	2.444 3	27.656 3
ARRVM	0.081 5	2.276 6	6.296 9

Tab. 2 Comparison of various algorithms (ratio = 4/5)

Algorithm	RMS dis	RMS angle	CPU time/s
SVM	0.086 8	2.429 5	4.625 0
RVM	0.078 8	2.300 3	34.406 3
ARRVM	0.071 0	2.227 6	4.859 4

Obviously, the RVM performs better than the SVM in estimation accuracy as verified by Ref. [5], but it has a longer training time than the SVM. The proposed algorithm is superior to both the SVM and the RVM in estimation accuracy while having the less training time than the RVM.

5 Conclusions

The application of machine learning tools such as the SVM to the estimation of illumination chromaticity has been proved to be better than that of previous traditional methods. In this paper, within the framework of sparse Bayesian learning, a new algorithm based on the relevance vector machine for the illumination estimation is proposed. Compared with the SVM and the RVM, the proposed algorithm has the following advantages.

1) Adaptive combination of global kernel and local kernel improves the estimation accuracy while aggravating the training time.

2) The application of the improved LPP reduces the input matrix column dimension and thus shortens the training time.

References

[1] Lin Chin-Teng, Fan Kan-Wei, Cheng Wen-Chang. An illumination estimation scheme for color constancy based on chromaticity histogram and neural network[C]//*Proceedings of* 2005 *International Conference on Systems, Man and Cybernetics*. Hawaii, USA, 2005: 2488 - 2494.

[2] Cardei Vlad C, Funt Brian, Barnard Kobus. Estimating the scene illumination chromaticity by using a neural network [J]. *Journal of the Optical Society of American A: Optics and Image Science, and Vision*, 2002, **19**(12): 2374 - 2386.

[3] Xiong Weihua, Funt Brian. Estimating illumination chromaticity via support vector regression [J]. *Journal of Imaging Science and Technology*, 2006, **50**(4): 341 - 348.

[4] Barnard Kobus, Cardei Vlad, Funt Brian. A comparison of computational color constancy algorithms—part Ⅰ: methodology and experiments with synthesized data [J]. *IEEE Transactions on Imaging Processing*, 2002, **11**(9): 972 - 984.

[5] Tipping M E. Sparse Bayesian learning and the relevance vector machine [J]. *Journal of Machine Learning Research*, 2001, **1**(3): 211 - 244.

[6] Smits G F, Jordaan E M. Improved SVM regression using mixtures of kernels[C]//*Proceedings of the International Joint Conference on Neural Networks*. Honolulu, 2002: 2785 - 2790.

[7] He Xiaofei. Locality preserving projections [D]. Chicago: The University of Chicago, 2005.

基于自适应约简相关向量机的光照色度估计

丁二锐[1]　曾　平[1]　姚　勇[2]　王义峰[1]

([1]西安电子科技大学外部设备研究所,西安 710071)

([2]西安电子科技大学计算机信息应用研究中心,西安 710071)

摘要:提出了一种新的自适应约简相关向量机回归算法来估计图像的光照色度以达到色彩一致性目的. 在稀疏贝叶斯学习的框架下,该算法首先以多核形式自适应结合全局核函数和局部核函数扩展相关向量机,然后应用改进的保局投影来约简多核输入矩阵的列维数以减少训练时间. 为了估计光照色度,通过图像色度直方图的模糊中心值和其相应光源值训练算法. 基于真实图像的实验表明所提算法优于支持向量机和相关向量机且其训练时间小于相关向量机.

关键词:色彩一致性;光照估计;色度直方图;自适应约简相关向量机

中图分类号:TP391;TP181