Coupled dictionary learning method for image decomposition

LI YaFeng\textsuperscript{1,2} & FENG XiangChu\textsuperscript{1}∗

\textsuperscript{1}Department of Applied Mathematics, Xidian University, Xi’an 710071, China;
\textsuperscript{2}Department of Computer Science, Baoji University of Arts and Science, Baoji 721007, China

Received November 9, 2010; accepted May 27, 2011; published online November 1, 2011

Abstract A novel variational model for image decomposition is proposed. Meanwhile a new cartoon-texture dictionary learning algorithm, which is guided by diffusion flow, is presented. Numerical experiments show that the proposed method has better performance than the existing algorithms in image decomposition and denoising.

Keywords dictionary learning, cartoon-texture decomposition, image denoising, total variation

1 Introduction

Extracting different features of image is one of important problems in the fields of image analysis and computer vision. A given image $f$ can be represented as $f = u + v$, where $u$ is the geometric (cartoon) component and $v$ is the oscillatory component representing texture or noise. Meyer [1] constructed the theory of cartoon-texture decomposition problem. He proposed that cartoon component should be modeled by using total variation norm \cite{2}, and oscillating texture part or noise should be modeled by using the G-norm, F-norm or E-norm in the special spaces of harmonic analysis. Based on this theory, Meyer proposed the following variational model for image decomposition:

$$\inf_{u,v} \{ |\nabla u|_1 + \lambda \|v\|_\ast, u + v = f \},$$

where $f$ is an observed image, $u$ is cartoon part, $v$ is texture part, the tuning parameter $\lambda$ specifies the trade-off between the two competing terms, and $\|\cdot\|_\ast$ denotes G-norm, F-norm or E-norm. But Meyer’s convex variational model cannot be solved directly due to the existence of the weaker norm. Thus, a lot of authors begin to study the practical methods of Meyer’s model. Vese and Osher [3] approximated the G-norm by the $\text{div}(L^p)$-norm (VO model). Osher et al. [4] proposed to use the $H^{-1}$-norm instead of $\|\cdot\|_\ast$, and discussed the following minimization problem (OSV model):

$$\inf_{u \in BV} \{ |\nabla u|_1 + \lambda \|f - u\|^2_{H^{-1}} \},$$

where $f$ is an observed image, $u$ is cartoon part, $v$ is texture part, the tuning parameter $\lambda$ specifies the trade-off between the two competing terms, and $\|\cdot\|_{H^{-1}}$ denotes $H^{-1}$-norm instead of $\|\cdot\|_\ast$. But Meyer’s convex variational model cannot be solved directly due to the existence of the weaker norm.
where \( \|f - u\|_{H^{-1}}^2 = \int_{\Omega} |\nabla^{-1}(f - u)|^2 \). The model (2) has the explicit Euler-Lagrange equation. Yin et al. proposed TV-L1 model for decomposing an image into features of different scales\[5,6]\, and compared several models (TV-L1, VO, Meyer) for image texture extraction in [7]. They formulated discrete versions of these models as second-order cone programs (SOCPs) which can be solved efficiently by interior-point methods. From a geometric point of view, Duval et al. discussed TV-L1 model in [8]. Recently, Buades et al. [9] proposed a pair of nonlinear filter, which retains the essential features of Meyer’s models. In [9], comparative experiments show that their method has a better and faster separation of cartoon from texture.

On the other hand, sparse representation method based on dictionary has been widely applied in different fields. Starck et al. [10] presented a sparse representation-based image decomposition method called morphological component analysis (MCA). The key of the method is to select two special analytic dictionaries to present cartoon and texture respectively. The success of this method is owing to this selection. However, it is difficult to choose proper dictionaries to describe the two different components. Image contents vary significantly across different images or different areas in an image. Hence, the selected analytic dictionaries lack the adaptivity. Although many authors discussed the sparse domain approximating, and improved algorithm and parameters choosing problems [11–14], how to choose the proper dictionaries [10] is still an open question.

Different from analytic dictionaries, dictionary learning (DL) method aims at optimizing some objective function in the training set and learning an adaptive dictionary. DL has been used to image denoising, image inpainting and face recognition. Many DL methods were proposed, such as MOD [15], KSVD [16], multiscale DL [17], DL for color image [18], etc. This paper proposes a novel DL based cartoon-texture decomposition model. A new algorithm to solve the model is presented and discussed in detail. Comparative experiments show that our method can separate cartoon and texture components and remove the noise better than the existing methods.

2 The proposed model

2.1 The model

A given image \( f \) is expired to be separated into \( f = u + v + w \), where \( w \) represents the Gaussian white noise, \( u \) and \( v \) are cartoon and texture parts, respectively. To make use of the advantages of variational methods and DL, two components of image are constrained in the proper function spaces, and meanwhile their sparse representation by learned dictionaries are required. So, our image decomposition model is defined as

\[
\arg \min_{\{u,v,D_1,D_2,\alpha,\beta\}} \left\{ \lambda_1 \|f - (u + v)\|^2 + \lambda_2 \|\nabla u\|_1 + \lambda_3 \|v\|^2_{H^{-1}} + \sum_i \|R_i u - D_1 \alpha_i\|^2 + \sum_i \mu_i \|\alpha_i\|_0 + \sum_i \|P_i v - D_2 \beta_i\|^2 + \sum_i \eta_i \|\beta_i\|_0 \right\},
\]

where \( \lambda_1, \lambda_2, \lambda_3 \) are trade-off parameters. The operators \( R_i \) and \( P_i \) denote extracting patch (with overlaps, the size of patches does not have to be the same) from location \( i \) in the cartoon part \( u \) and texture part \( v \), respectively. The two components are constrained in two function spaces, i.e. bounded total variation space \( BV \) for cartoon part, and negative Hilbert-Sobolev space \( H^{-1} \) for texture part. Meanwhile, the dictionaries \( D_1 \) and \( D_2 \) are used to represent two different components under the sparse constraints \( l_0 \) quasi-norm. The dictionaries \( D_1 = [d_1,d_2,\ldots,d_{k1}] \), \( D_2 = [s_1,s_2,\ldots,s_{k2}] \) satisfy the unit constraint conditions. The coefficients \( \mu_i, \eta_i \) depend on the noise level.

2.2 The model analysis

Model (3) is a joint optimization problem. As in many multi-variables optimization problems, we solve it by optimizing these variables alternatively. Model (3) is divided into the following four subproblems.