Journal of Scientific & Industrial Research
Vol. 70, April 2011, pp. 251-255

Video denoising without motion estimation using K-means clustering

C Ramya[1]* and S Subha Rani[2]
PSG College of Technology, Coimbatore 641 004, India

Received 16 September 2010; revised 28 February 2011; accepted 03 March 2011

This study presents a novel dictionary pruning algorithm that found dictionaries of optimized size for a given dataset, without compromising its approximation accuracy and performance. It is achieved by applying KSVD (K-means singular value decomposition) algorithm to patches of dictionary. This optimized dictionary selection will provide an increased convergence speed and performance of decomposition algorithm by ensuring minimum error as well as sparsity of representation. Proposed method optimized dictionary selection, and with KSVD yielded better video denoising than KSVD with fixed dictionary.

Keywords: Denoising, KSVD, OMP, Video

Introduction

Image enhancement operations by using denoising improve image's contrast and brightness characteristics, reduce noise content, and sharpen edge details etc. For video denoising (VD), motion estimation between consecutive frames has been the basis for many denoising algorithms. However, several studies[1,2] suggest that motion detection is not necessary if state-of-art in denoising is achieved. KSVD (K-means singular value decomposition) algorithm[3-5] has been proposed to train a sparsifying dictionary for corrupted gray scale images. Then, this work was extended to color image sequence denoising[6-10]. Further, probabilistic approach[11-13] was used to mitigate sparse representation (SR) problem and dictionary learning was suggested. Another solution was reported[14] as convex optimization approach that minimizes sum of squared residues and a sparsity constraint on dictionary coefficients. An image denoising method[15] has been developed based on the principle of sparse and redundant representation.

This study presents VD with a novel dictionary pruning algorithm (DPA), which gives better convergence speed and denoising performance of decomposition algorithms. Hence KSVD discovers an optimum number of dictionary elements by reducing redundancies in learned dictionary and leads to a substantial benefit in denoising performance of KSVD than KSVD with fixed dictionary.

*Author for correspondence
E-mail: ramya_meena@rediffmail.com

Experimental Section

Orthogonal Matching Pursuit (OMP)

SR finds a set of prototype signals, referred as atoms, which form a dictionary Φ. Sparse linear combination of atoms in Φ represent a particular set of given signal. Let Y denotes prototype signal, to find Φ such that $y_i \sim \Phi x_i$, where x_i is sparse vector. Denoising of prototype signal is given as

$$\min_{\Phi, x_i} \|x_i\|_0 \quad s.t. \quad \|y_i - \Phi x_i\|_2 < e \quad \ldots(1)$$

where $\|.\|_0$ denotes l_0 norm, which counts number of nonzero element of vector and e is tolerable limit of error in reconstruction.

Matching pursuits[15-19] (MP) for SR is a greedy algorithm that finds linear approximations of signals by iteratively projecting them over a redundant, possibly non-orthogonal set of signals called dictionary. Also, it may give a suboptimal approximation. Sub optimality of MP for finite atoms takes more iteration to reduce residue below a given threshold. It is removed by OMP, which gives optimal approximation with respect to selected subset of dictionary elements by making residue orthogonal to all chosen dictionary elements. OMP is

Step 1: select dictionary matrix in n clusters { $\mu_1, \mu_2, \mu_3, \ldots, \mu_n$ }

Step 2: Calculate cluster centroids $\overline{m}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} m_{ij}, \quad i = 1, \ldots, N$

Step 3: In each iteration, assign each $x^{(i)}$ to closest cluster centroid

μj $c^{(i)} = \arg\min_j \| x^{(i)} - m_j \|^2$

Moving each cluster centroid μj to mean of the points assigned to it.

$$m_j = \frac{\sum_{i=1}^{m} 1\{c^{(i)} = j\} x^{(i)}}{\sum_{i=1}^{m} 1\{c^{(i)} = j\}}$$

Repeat until convergence

Step 4: If cluster membership is stabilize then stop else go to step 3.

Fig.1—Proposed dictionary pruning algorithm

easy to implement and provides a satisfactory stable result.

OMP attempts to find atoms iteratively such that in each iteration, error in representation is reduced. This is achieved by selection of that atom from dictionary, which has largest absolute projection on error vector. This implies that atom adds maximum information and hence maximally reduces error in reconstruction. Given a signal vector y and a dictionary Φ, algorithm attempts to find code vector x in following steps: i) Select atom that has maximal projection on residual; ii) Update $x^n = \arg\min_{x^n} \| y - \Phi x^n \|_2$; and iii) Update residual $r^n = y - y^n$.

But, in each iteration, OMP is still choosing dictionary element by MP criteria. Since there is no constraint on orthogonality of dictionary; the dictionary elements chosen by greedy approach of OMP can be highly correlated. This will adversely affect convergence speed of OMP. So dictionary size is an important factor for storage considerations and computational speed. To mitigate this problem, a novel dictionary pruning algorithm is adapted to reduce total number of OMP iterations and size of dictionary to smaller, more manageable size. Thus convergence speed is increased.

Proposed Method [Dictionary Pruning Algorithm (DPA)]

For all kinds of signals, fixed dictionary is not suitable. So the best option is a data dependent dictionary. An optimal size dictionary will increase speed and performance of decomposition algorithms used. Optimal size of dictionary is obtained by KSVD. Optimized dictionary selection is essential for an efficient dictionary learning algorithm. Scalability in terms of time and space is a desirable property of clustering algorithm applied to dictionary so that every patch can be classified into different classes. Thus size of dictionary gets optimized. Consider dictionary of { $x^{(1)}, \ldots, x^{(m)}$ } patches to group this data into few clusters. Here $x^{(i)} \in R^n$ as usual. Optimizing algorithm (Fig. 1) is as follows:

Initialize cluster centroids $\mu_1, \mu_2, \mu_3, \ldots, \mu_n \in R^n$ randomly.
Repeat until convergence:

For every i, set, $c^{(i)} = \arg\min_j \| x^{(i)} - m_j \|^2$

for every j set, $$m_j = \frac{\sum_{i=1}^{m} 1\{c^{(i)} = j\} x^{(i)}}{\sum_{i=1}^{m} 1\{c^{(i)} = j\}} \qquad \ldots(2)$$

where n is number of clusters and cluster centroids μj represent positions of centers of clusters. Inner-loop of algorithm repeatedly carries out two steps. Assign each $x^{(i)}$ to closest cluster centroid μj and then moving each cluster centroid μj to mean of the points assigned to it.

In this study, VD is done with KSVD for DPA. KSVD uses a two phase approach to update values of dictionary coefficients and dictionary: i) dictionary coefficients are updated using OMP; ii) dictionary coefficients are assumed to be fixed and only one column of dictionary is updated at a time.

Denoising Using KSVD

For a given image[15], which can now be thought of as a set of signals Y, denoising problem can be stated as ill-posed problem of finding a set of patches P as

$$Y = P + h \qquad \ldots(3)$$

where η is assumed to be zero mean Gaussian noise that corrupts patches. In order to denoise image patches P, a patch based approach has to be adopted.

$$\hat{X}, \hat{P} = \arg\min_{X,P} \| P - Y \|_2^2 + l \| \Phi X - P \|_F^2 + \sum_i m_i \| x_i \|_0 \qquad \ldots(4)$$

where $\| . \|_F^2$ denotes Frobenius norm square, which is defined as square of every element in matrix. This can be viewed as solving a set of smaller optimization problems as

$$\left\{ \hat{x}_i, \hat{P} \right\} = \arg\min_{x_i, P} \| P - Y \|_2^2 + l \| \Phi x_i - R_i P \|_F^2 + \sum_i m_i \| x_i \|_0 \qquad \ldots(5)$$

where R_i is a matrix that selects ith patch from P i.e. $p_i = R_i P$. This cost function allows to minimize error between restored image and input noisy one, under assumption that each patch in input image can be represented as a sparse linear combination of patches in Φ. Ideally, for denoising first term should be rewritten as $\| P - Y \|_2^2 < Cs^{2}$, where C is a constant and σ^2 is variance of noise. However, this term is implicitly incorporated into cost function in selection of λ, which will depend on noise variance. Closed form solution to this cost function is given as

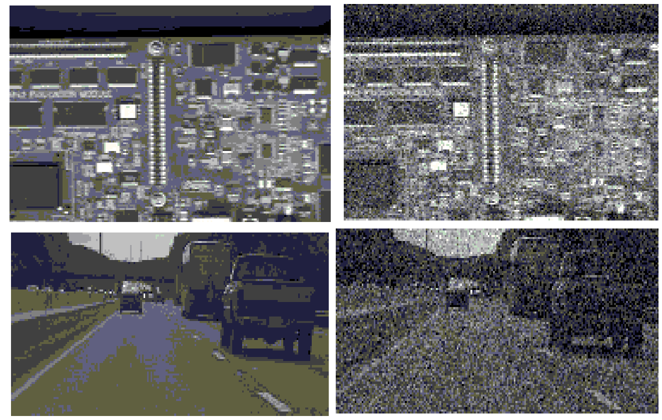
$$\hat{P} = \frac{(lY + \sum_i R_i^T \Phi x_i)}{(lI + \sum_i R_i^T R_i)} \qquad \ldots(6)$$

Solution to this problem thus involves averaging of overlapping patches after each patch has been sparse coded along with a weighted sum of original noisy patch. Each pixel in a patch is hence a weighted linear combination of different pixels, weights being derived from sparse coding. Since patches are overlapping, final value of each pixel is thus an average of all representations obtained from sparse coding stage. For image sequences[1], consider Y and P be noisy and clean videos respectively. Adding an index t in the range [1,T] to account for time dimension, one gets

$$f_{Video}^{All}(\{x_{ijt}\}_{ijt}, \hat{P}, \Phi) = l \| P - Y \|_2^2 + \sum_{ij}\sum_{t=1}^{T} m_{ijt} \| x_{ijt} \|_0 + \sum_{ij}\sum_{t=1}^{T} \| \Phi x_{ijt} - R_{ijt} P \|_2^2 \qquad \ldots(7)$$

Minimization of this function generates a single dictionary for entire sequences, and cleans all images at once. But training of a single dictionary for entire sequences is problematic. For dictionary to be able to suit all images, it is required to change along sequence. So dictionary must suit to the scene also, as it exploits benefits of temporal locality and redundancy. Thus it leads to modified version as

$$f_{Video}^{(t\pm\Delta t)}(\{x_{ijk}\}_{ijk}, P_t, \Phi_t) = l \| P_t - Y_t \|_2^2 + \sum_{ij}\sum_{k=t-\Delta t}^{t+\Delta t} m_{ijk} \| x_{ijk} \|_0 + \sum_{ij}\sum_{k=t-\Delta t}^{t+\Delta t} \| \Phi_t x_{ijk} - R_{ijk} P \|_2^2 \qquad \ldots(8)$$



Original frame Noisy frame

Fig. 2—Board and plane sequences (1 in 30 frames) with σ=25

Table1—Comparison of MSE values of noisy, KSVD and proposed method (σ=25)

Test	Noisy sequence	KSVD with fixed dictionary	KSVD with proposed dictionary
News	25.1484	9.9238	9.7286
Board	25.1899	8.6179	7.4142
Plane	25.0445	10.3706	9.8615

defined for t = 1,2,…T. $R_{ijt}P$ extract a patch of fixed size from volume X in time t and spatial location [i , j]. Thus KSVD algorithm make use in build noise averaging capability to denoise large set of noisy patches and creates a small, relatively clean representative set.

Results and Discussion

To test DPA, select a set of four different image sequences (plane, board, ish and News) with standard set of parameters. Each of the four test sets is superimposed with synthetic zero mean Gaussian noise, using noise levels σ =5, 10, 15, 20, 25, 30, 35, 40, and 50. Input was corrupted by zero mean Gaussian noise of σ =25 (Fig. 2). Here contaminated frames are denoised using KSVD. In the implementation, overlapping patches were selected with each patch (size, 8 x 8). Patches used for training dictionary were taken from noisy image itself. KSVD is iterated 15 times with standard dictionary. But by DPA, iterations are reduced to 10 times with better denoising performance (Fig. 3). Denoising obtainded for ish color image sequences are shown in Fig 4. MSE (mean square error) for noise level of σ=25 (Table 1) of DPA is considerably low compared to fixed dictionary.

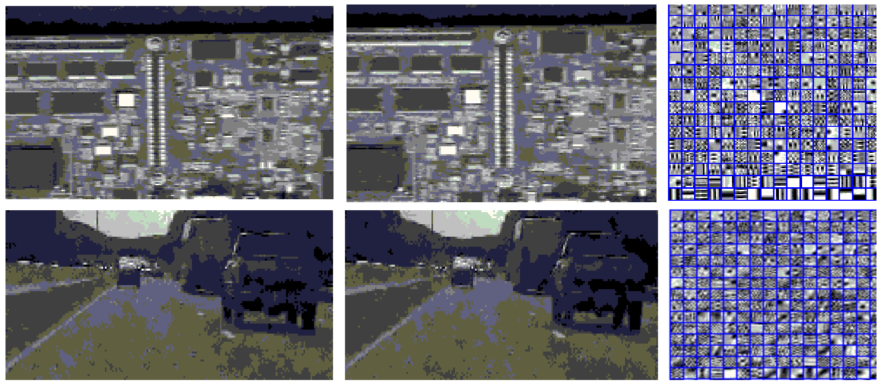


Fig. 3—Board and plane sequences (1 in 30 frames) with σ =25 (Left: cleaned by KSVD; Middle: cleaned by proposed method; Right: proposed method dictionary)

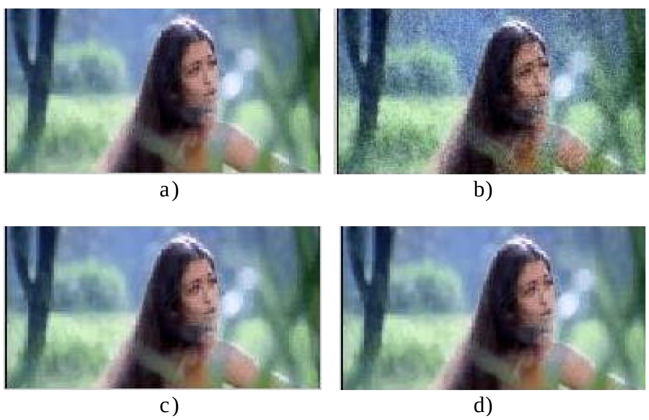


Fig. 4 — ish Colour image sequences (1 in 30 frames): a) Original frame; b) Noisy frame (σ=25); c) Cleaned by KSVD; and d) Cleaned by proposed method

Comparison of Peak Signal-to-Noise Ratio (PSNR) Values

PSNR values decreased when noise level increased for different image sequences. PSNR value of DPA sounds high compared to existing algorithm as observed in plane sequences of σ = 15; PSNR value of KSVD with fixed dictionary is 34.563 and with DPA is 36.523 (Table 2).

Structural Similarity (SSIM) Index

Structural similarity (SSIM) index is designed to improve on traditional methods (PSNR and MSE), which have proved to be inconsistent with human eye perception. SSIM index expresses quality by comparing local correlations in luminance, contrast, and structure between reference and distorted images. Mean structural similarity (MSSIM) index is calculated for DPA by considering σ =15 (Table 3). SSIM metric is calculated

Table 2—Comparison of PSNR values with different σ values

Test sequence	Input noise level (σ)	KSVD with fixed dictionary, dB	KSVD with proposed dictionary, dB
News	5	39.038	40.988
	10	34.358	36.240
	15	32.094	33.959
	20	30.426	32.283
	25	29.334	31.207
	30	28.037	29.880
	35	26.926	28.807
	40	25.919	27.768
	50	24.302	26.170
Board	5	38.339	40.128
	10	33.666	35.543
	15	30.534	33.6165
	20	28.865	31.894
	25	27.146	29.875
	30	26.066	29.78
	35	25.344	27.968
	40	24.192	26.305
	50	22.682	25.882
Plane	5	40.025	42.360
	10	36.732	37.048
	15	34.563	36.523
	20	32.608	33.882
	25	30.499	31.189
	30	31.337	33.374
	35	30.455	32.344
	40	30.10	31.982
	50	28.313	30.155

Table 3—MSSIM values of proposed method (σ = 15)

Test sequence	PSNR, dB	MSSIM
News	33.959	0.9379
Board	33.6165	0.9123
Plane	36.523	0.9524

on various windows of an image and measure between two windows X and Y of common size (N x N) is given as

$$SSIM(x,y) = \frac{(2m_x m_y + C_1)(2s_{xy} + C_2)}{(m_x^2 + m_y^2 + C_1)(s_x^2 + s_y^2 + C_2)}$$

$$MSSIM(X,Y) = \frac{1}{M}\sum_{j=1}^{M} SSIM(x_j, y_j) \quad \ldots(9)$$

where μ_x and μ_y are averages of x and y; s_x^2 and s_y^2 are variance of x and y; s_{xy} covariance of x and y; $C_1 = (k_1 L)^2$, $C_2 = (k_2 L)^2$ are constant; L dynamic range of pixel values.

Conclusions

A DPA is presented to obtain a better convergence speed and performance of decomposition algorithms. It is based on sparse and redundant representation. By optimizing dictionary with K means clustering algorithm besides KSVD algorithm, a better denoising results were obtained than dictionary learning algorithms with fixed dictionary. DPA reduces total number of OMP iterations and size of dictionary to smaller, more manageable size. Also, dictionary learning technique like KSVD discovers an optimum number of dictionary elements by reducing redundancies in learned dictionary. All these extensions are thoroughly tested on an extensive set of image sequences and noise levels, and found that DPA dramatically improves speed and denoising performance.

References

1 Protter M & Elad M, Image sequence denoising via sparse and redundant representations, *IEEE Trans.Image Process*, **18** (2009) 27-35.

2 Mairal M J & Sapiro G, Learning multiscale sparse representations for image and video restoration, *SIMA, Multiscale Model Simul*, **7** (2008) 214-241.

3 Aharon M, Elad M & Bruckstein A M, The K-SVD: An algorithm for designing of over complete dictionaries for sparse representations, *IEEE Trans Signal Process*, **54** (2006) 4311-4322.

4 Elad M & Aharon M, Image denoising via sparse and redundant representations over learned dictionaries, *IEEE Trans Image Process*, **15** (2008) 3736-3745.

5 Elad M & Aharon M, Image denoising via learned dictionaries and sparse representation, *CVPR*, 1(2006) 895-900.

6 Mairal M J & Sapiro G, Sparse representation for color image restoration, *IEEE Trans Image Process*, **17** (2008) 53-69.

7 Li H & Liu F, Image denoising via sparse and redundant representations over learned dictionaries in wavelet domain, in ICIG '09 Proc *2009 Fifth Int Conf on Image and Graphics* (IEEE Computer Society Washington, DC, USA) 2009 754-758.

8 Wright J, Ma Y & Mairal J, Sparse representation for computer vision and pattern recognition, in *Proc IEEE Special Issue on: Applications of Sparse Representation and Compressive Sensing*, (2009) 1-10.

9 Boulanger J, Kervrann C & Bouthemy P, Space-time adaptation for patch based image sequence restoration, *IEEE Trans Pattern Anal Mach Intell*, **29** (2007) 1096-1102.

10 Dabov K, Foi A & Egiazarian K, Video denoising by sparse 3D transform-domain collaborative filtering, in *15th Eur Signal Process Conf, EUSIPCO 2007* (Poznan, Poland) 2007, 145-149.

11 Olshausen B A & Field D J, Sparse codingwith an over complete basis set: a stratergy employed by V1? *Vision Res*, **37** (1997) 3311-3325.

12 Olshausen B A & Millman K J, Learning sparse codes with a mixture –of –gaussian prior, *NIPS*,12 (1999) 841-847.

13 Kreutz-Delgado K, Murray J F, Rao B D, Engan K, Lee T W & Sejnowski T J, Dictionary learning algorithm for sparse representation, *Neural Comput*, **15** (2003) 349-396.

14 Lee H, Battle A, Rania R & Ng A Y, Efficient sparse coding algorithms, *Adv Neur Inform Process Syst*, **19** (2007) 801-808.

15 http://users.soe.ucsc.edu/~priyam/ksvd_report.pdf.

16 Mallat S, Davis G & Zhang Z, Adaptive time-frequency decompositions, *PIE J Optic Engg*, **33** (1994) 2183-2191.

17 Pati Y, Rezaiifar R & Krishnaprasad P, Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition, in *27th Conf Signals, Systems and Computers*, **vol 1** (Asilomar on Pacific Grove, CA, USA) 1993, 40-44.

18 Mallat S & Zhang Z, Matching pursuits with time-frequency dictionaries, *IEEE Trans Signal Process*, **41** (1993) 3397-3415.

19 Donoho D L, Elad M & Temlyakov V, Stable recovery of sparse overcomplete representations in the presence of noise, *IEEE Transaction on Information Theory*, **52** (2006) 6-18.