ILLUMINATION COMPENSATION VIA LOW RANK MATRIX COMPLETION FOR MULTIVIEW VIDEO CODING

Xiaopeng Zhang, Hongkai Xiong

Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

ABSTRACT

A multi-view video system would capture the same scene from different viewpoints, and suffer from significant illumination variations due to inaccurate camera calibration and varying light conditions. It deteriorates the inter-view correlations and the quality of synthesized views at decoder side. By converting the problem of illumination compensation to noise removing, this paper proposes a low-rank matrix completion algorithm to reduce the effect of illumination variation. Diverged from the existing work which chooses a central view as reference and keeps views consistent, it is dedicated to compensating all the views to match the low-rank structure of views. The discrepancies among views are regarded as mixed noise, and constructed as an incomplete matrix with low-rank. It is solved by a stable matrix completion and obtains a mapping function for color correction. It is robust to outliers since only plausible corresponding points are involved with completion. Experimental results show that the proposed algorithm can increase coding efficiency by up to 0.7dB for luminance component and up to 2.1dB for chrominance components.

Index Terms— Illumination compensation, multi-view video coding, matrix completion, low-rank structure

1. INTRODUCTION

With the breakthrough of 3-D vision field, multi-view video coding has been drawn more attention to decorrelate spatio-temporal samples within views [1]. However, significant luminance and chrominance discrepancies might occur due to inaccurate camera calibration and varying light condition among views. It would deteriorate the correlations among views and impose negative impacts on the synthesis and rendering of virtual views at decoder [2]. In this sense, to compensate the discrepancies becomes vital importance for desirable coding performance.

In 2008, Fecker *et al.* [3] developed a well known illumination compensation approach by histogram matching (H-M) which aims to match the histogram of the target view to the reference view. Because all the other views should be corrected in alignment with the center view as reference, the phase information of pixels is ignored. An alternative way is to correct multi-views to the average color of all views, and each view was corrected by polynomial functions [4]. In an analytic form, Chen [5] constructed a global linear model to compensate for the discrepancy based on modulation and translation parameters. However, it fails to take the local discrepancy into consideration.

Impressively, this paper is to convert the problem of illumination compensation to noise removing via low-rank matrix completion. An incomplete matrix with each column representing one view at one time and each row representing the same point, is constructed where the discrepancies among views can be regarded as mixed noise. The matrix is supposed to be low-rank if the missing entries are filled in correctly, and it could be recovered from noise by a stable matrix completion algorithm. In turn, a mapping function would be attained for color correction. Compared to the existing approach making views consistent, it is robust to outliers with lower modification.

The rest of this paper is organized as follows: Sec. II reviews the preliminary knowledge of matrix completion. The proposed correction approach is described in Sec. III with regard to the construction of the incomplete matrix, the matrix completion and the mapping process. The experimental validation is shown in Sec. IV.

2. BACKGROUND OF MATRIX COMPLETION

Matrix completion is a well studied but still challenging problem, which solves the problem of completing a low-rank matrix from only a small portion of observations [6]. Let $X = U\Sigma V^T$ be the singular value decomposition of matrix $X \in R^{m*n}$ with rank r, the singular values are arranged in a non-increasing order, i.e., $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$, and $\sum = diag(\{\sigma_i\}_{1\leq i\leq r})$. The nuclear norm of X is defined as $\|X\|_* = \sum\limits_i \sigma_i$. For each $\tau > 0$, the soft threshold operator D_τ is defined as: $D_\tau(X) = US_\tau(\Sigma)V^T$, where $S_\tau(\Sigma) = diag(max(\sigma_i - \tau, 0))$. The observed partial matrix D can be completed by solving the problem:

$$\begin{aligned} &\min \|A\|_* \\ &\textit{s.t.} \quad A_{ij} = D_{ij}, \quad (i,j) \in \Omega \end{aligned} \tag{1}$$

The work has been partially supported by the NSFC grants No. U1201255, No. 61271218, No. 61228101.

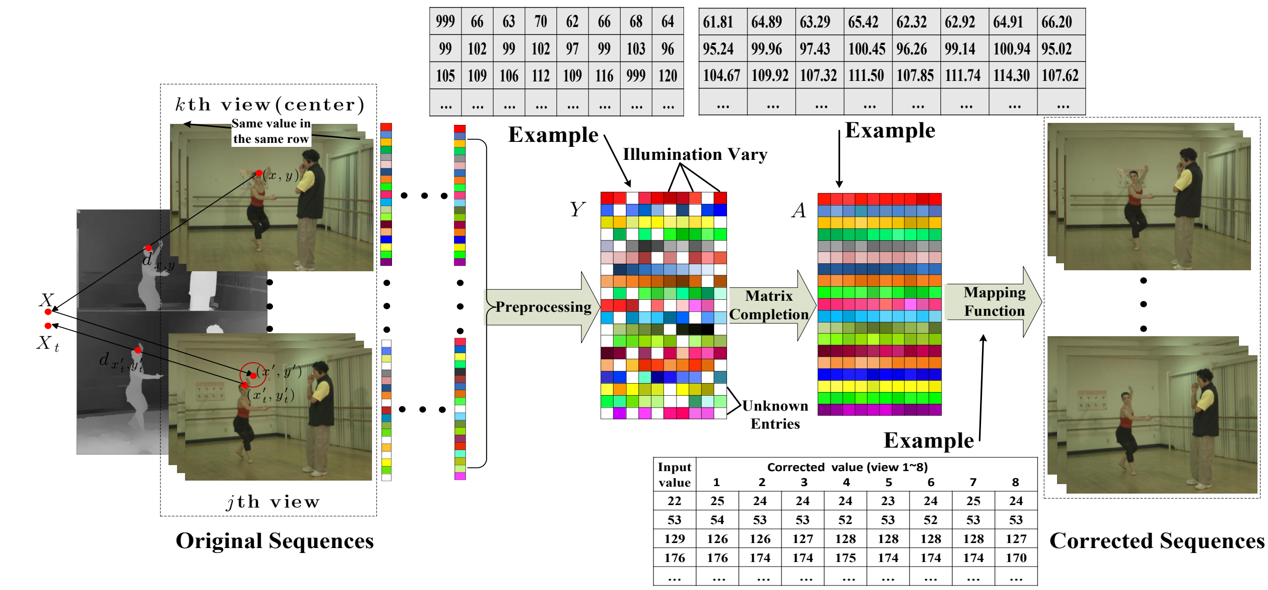

Fig. 1: The diagram of the proposed algorithm

where Ω is the index of observed entries.

In real applications, the observed data is more or less affected by noise. Hence, a stable matrix completion method is developed which guarantees reasonable accurate results from noisy sampled entries [7]. Considering the following minimization problem:

$$\begin{aligned} &\min \|A\|_* \\ &s.t. \quad \|P_\Omega(A-Y)\|_F^2 \leq \#(\Omega)\sigma^2 \end{aligned} \tag{2}$$

where Y is the noisy sampled matrix, P_Ω is a linear operator that keeps the entries in Ω unchanged and sets those outside Ω zeros, $\#(\Omega)$ is the number of samples and σ is a given parameter which is related to the standard deviation of noise.

3. MATRIX COMPLETION-BASED ILLUMINATION COMPENSATION SCHEME

It is difficult to exactly model and analyze all error sources and their effects on the pixel values under illumination variation. For a given point in frame k, its value $f_k(x,y)$ consists of two components: the actual value $g_k(x,y)$ and the deviation $n_k(x,y)$, which can be regarded as addictive noise:

$$f_k(x,y) = g_k(x,y) + n_k(x,y) \tag{3}$$

where (x,y) denotes the location of the point. For every frame k, we stack it column-wisely as a vector F_k, and construct a matrix with each column representing one view at one time [8]. We permutate the points in each column to put the corresponding points into the same row. When there are n frames used for correspondence, we obtain the matrix:

$$Y = [F_1, F_2, ..., F_n] = A + Z \tag{4}$$

where A denotes the matrix of actual values and Z denotes the noise matrix. Due to occlusion, the matrix Y is incomplete. We construct the matrix with the visible points regarded as known entries, and the invisible points as unknown entries(large number, say 999), let Ω be the index set of the visible points, the incomplete matrix can be represented as:

$$Y_{ij} = \begin{cases} A_{ij} + Z_{ij}, & \text{for } (i,j) \in \Omega \\ 999, & \text{otherwise} \end{cases} \tag{5}$$

As shown in Fig. 1, the goal is to recover the low-rank matrix A from the observed incomplete noisy matrix Y. A matrix completion solver could be regarded as a mathematical filter, then a mapping function is generated for illumination compensation.

3.1. Incomplete matrix construction among views

By a projection process from the corresponding depth maps and camera parameters, a 3-D warping method is used to obtain the corresponding points [9]. The matching point is detected by projecting the pixel of one view to its 3-D coordinate and reprojecting it onto the target view. Let I_k represent the kth view at one time and d_k its corresponding depth map; A_k, R_k and t_k denote, respectively, the intrinsic matrix, the rotation matrix and the translation vector of camera I_k. Hence, the 3-D coordinate X of point $I_k(x,y)$ is obtained by:

$$X = R_k A_k^{-1} [I_k(x,y)\ 1]^T d_k(x,y) + t_k \tag{6}$$

The point X is reprojected into the target view (refer to Fig. 1) to get the matching point by:

$$I_j(x', y') = A_j R_j^{-1}(X - t_j) \tag{7}$$

where (x', y') is the location of matching point in jth view. Due to the inaccuracy of the depth maps and camera parameters, the point $I_j(x', y')$ may not be the accurate correspondence. Thus, we choose the nearest integer pixel to $I_j(x', y')$ as center, and its 8-neighborhood pixels as candidate matching points (marked as red circle in Fig. 1). All the 9 points are back-projected to the 3-D space to get the point set X_{set}, and the nearest 3-D point X_t is chosen by the following principal:

$$X_t = \underset{X' \in X_{set}}{\arg\min} \parallel X' - X \parallel_2 \tag{8}$$

Once obtaining its projected points $I_k(x_t, y_t)$ in the kth view, in view of occlusion, if the Euclidean distance between $I_k(x, y)$ and $I_k(x_t, y_t)$ is less than an empirical threshold σ (e.g. 2), the two points would be regarded as corresponding points, and otherwise an unknown entry.

In the 3-D warping, the center view is chosen as reference and all the points in the center view are warped column-wisely to all the other views to find corresponding points. It can be repeated for a couple of successive time instances to get more corresponding points. For the views from different time instead of viewpoints, the points with the same value would be put into the same row in accordance with the center view as reference.

3.2. Low rank extraction via matrix completion

To recover the low-rank matrix from the incomplete noisy matrix, we adopt a preprocessing step to choose only reliable elements for completion where the elements are chosen based on their deviation to the mean of all known entries in the same row. When the deviation of a pixel is larger than a tentative threshold, it is set to be unknown. Referring to the matrix Y for completion in Fig. 1, each column represents one view at one time. The color indicates the similarity of the corresponding points, and the white entry represents the unknown.

The low-rank matrix A, would be recovered from Y by solving Problem (2) with its Lagrangian form:

$$\min \quad \frac{1}{2}\|P_\Omega(A - Y)\|_F^2 + \mu\|A\|_*. \tag{9}$$

It is a standard duality which is equivalent to (2) for μ. One can solve Eq. (9) by searching for an optimal value of μ which makes the solution A satisfy $\|P_\Omega(A-Y)\|_F^2 \approx \#(\Omega)\sigma^2$. and followed by a heuristic arguments in [7]. Here, μ is chosen as:

$$\mu = (\sqrt{n_1} + \sqrt{n_2})\sqrt{p}\sigma \tag{10}$$

where p is the ratio of known entries against the total number elements in Y, and n_1, n_2 are the dimensions of Y. The fixed point iterative algorithm [12] is adopted to solve Eq. (9) for its accuracy [10], and the details are summarized in Algorithm 1.

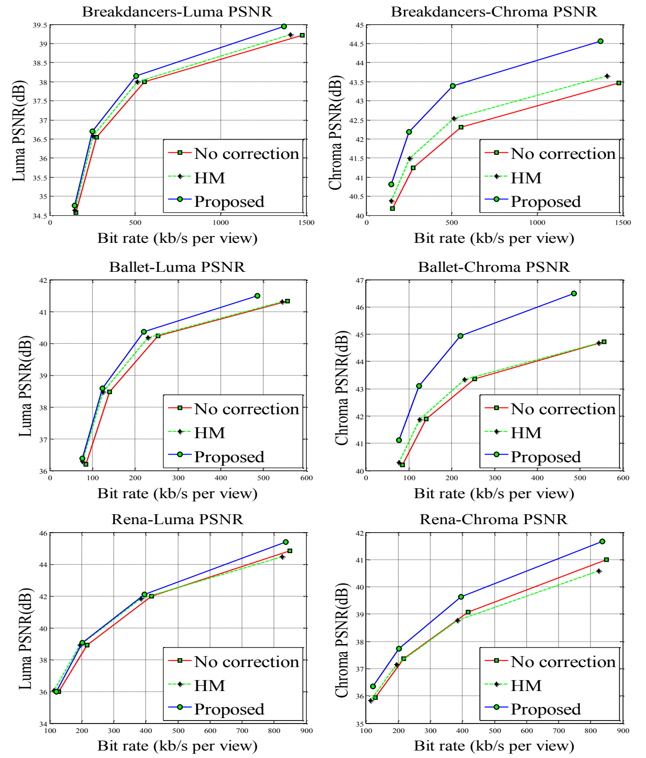

Fig. 2: The coding performance comparison of the original coding scheme without correction, the histogram-matching (HM) compensation scheme, and the proposed algorithm

3.3. Illumination compensation via mapping function

With the incomplete noisy matrix Y and low-rank matrix A, a mapping function would be generated for correction. Let Y_i and Y^j denote, respectively, the ith row and the jth column of Y. For a given view k, find its K corresponding columns $\{Y^{j_k}\}_{k=1}^K$ in Y, record the M rows $\{i_k\}_{k=1}^M$ which have the pixel value $n(0 \leq n \leq 255)$ in sub-matrix $\{Y^{j_k}\}_{k=1}^K$, and average row vectors $\{A_{i_k}\}_{k=1}^M$ in A to get corrected value $n'(0 \leq n' \leq 255)$. It can be represented as:

$$map(n, k) = round(\frac{\sum_{k=1}^M A_{i_k}}{Mn_2}) \tag{11}$$

The same mapping function would compensate all frames in a view, and see an example in Fig. 1.

Algorithm 1 Fix point iteration for solving Problem (9)

Initialization: $A^{(0)} = 0, k = 0$, given μ, σ, τ

while $\|A^k - A^{(k-1)}\|_F \geq \epsilon$, **do**

$$\begin{cases} R^{(k)} = A^{(k)} - \tau P_\Omega(A^{(k)} - Y) \\ A^{(k+1)} = D_{\tau\mu}(R^{(k)}) \end{cases}$$

end while

Output: $A = A^{(k)}$

Seq.	Original		HM			Proposed		
	Temporal	Spatial	Temporal	Spatial	Increased	Temporal	Spatial	Increased
Break	88.32%	11.68%	84.73%	15.27%	3.59%	85.95%	14.05%	2.37%
Rena	85.75%	14.25%	82.24%	17.76%	3.51%	81.78%	18.22%	3.97%
Ballet	96.35%	3.65%	94.69%	5.31%	1.66%	94.02%	5.98%	2.33%

Table 1: The optimal prediction ratio of three schemes ($QP = 22$)

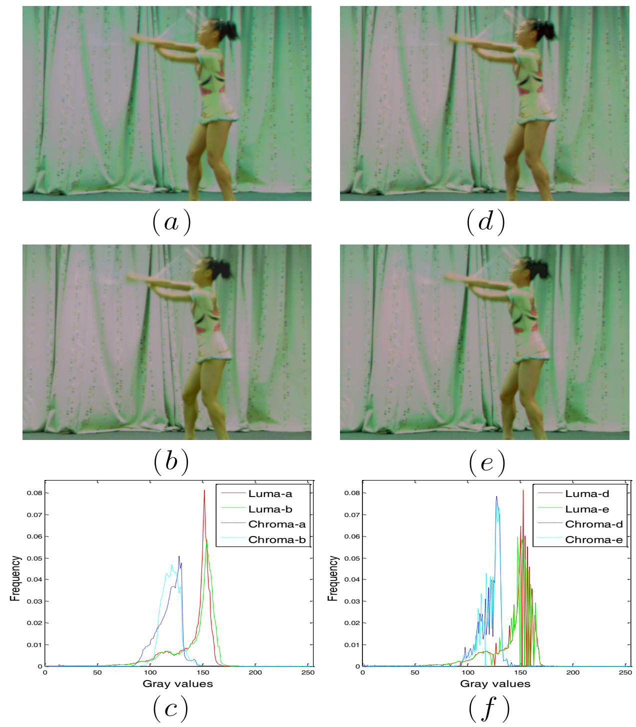

Fig. 3: The subjective performance of Rena:(a), (b) original views (d), (e) compensated views (c), (f) corresponding histograms

4. EXPERIMENTAL RESULTS

In experiments, we implement the proposed algorithm with the JMVC(version 8.3.1) reference encoder. Three multi-view sequences, Ballet (8 views), Breakdancers (8 views) [14] and Rena (8 views), are extensively tested where the GOP size is 12 and the quantization parameter (QP) is evaluated as the typical conditions: 22, 27, 32, and 37 [13]. Also, every 10th frame per view are able to construct the incomplete matrix, and the threshold selecting reliable pixels is set to make the ratio p around $50\% \sim 60\%$. In Algorithm 1, the parameter τ is 1.95 and the stopping criterion ϵ is 10^{-4}.

The proposed algorithm is compared with 1) the original coding scheme without correction and 2) the histogram-matching (HM) compensation scheme [3]. When different compensation schemes are compared, we use their corresponding inputs as references for measuring the PSNR since there are no ground truth video that has the perfect color. Fig. 2 shows their coding performance comparison of luminance and chroma components, where the chroma PSNR is the average PSNR of U and V components. It can be seen that the proposed scheme gets a PSNR gain of luminance component ranging from 0.4 to 0.7 dB over the original coding without correction and 0.3 to 0.8 dB over the HM compensation scheme. For chroma components, it would range from 1.1 to 2.1 dB and 0.9 to 1.9 dB, respectively. Fig. 3 illustrates the subjective view effects of the proposed scheme, where the histogram of the compensated views are closer than the original views. Obviously, the colors of the views are different before compensation but look consistent after compensation. For clarity, we evaluate the correlations (percentage in Table 1) among views by counting the number of temporal and spatial reference blocks from block matching [16]. Furthermore, the similarity between the compensated sequences and the original ones are evaluated in terms of the SSIM criterion [15]. From the average SSIM over all frames in Table 2, the proposed algorithm is witnessed to get lower modification than the HM scheme.

Seq.	HM			Proposed		
	Y	U	V	Y	U	V
Break	0.872	0.874	0.874	0.996	0.995	0.995
Rena	0.871	0.869	0.860	0.995	0.994	0.985
Ballet	0.869	0.873	0.875	0.996	0.994	0.996

Table 2: The SSIM evaluation between the original sequences and the compensated sequences

5. CONCLUSIONS

In this paper, we propose a matrix completion-based illumination compensation method as a pre-filtering step to compensate the illumination variations of multi-view video sequences. The incomplete matrix is constructed by finding corresponding points among views, followed by a matrix completion process aiming at recovering the low-rank matrix among views. According to the matrix before and after completion, the mapping function is obtained for illumination compensation. It favors lower modification to achieve visual consistent among views and is robust to outliers. Experiments show that it outperforms the uncompensated coding and the HM compensation.

6. REFERENCES

[1] A. Smolic, K. Mueller, N. Stefanoski, "Coding algorithms for 3DTV-A survey", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, pp. 1606-1621, Nov. 2007.

[2] S. Chan, H.Y. Shum, K.T. Ng, "Image-based rendering and synthesis", *IEEE Signal Processing Magazine*, vol. 24, no. 6, pp. 22-33, June 2007.

[3] U. Fecker, M. Barkowsky, and A. Kaup, "Histogram-based prefiltering for luminance and chrominance compensation of multiview video", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 9, pp. 1258-1267, Sep. 2008.

[4] C. Doutre, P. Nasiopoulos, "Color correction preprocessing for multiview video coding", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 9, pp. 1400-1406, Sep. 2009.

[5] Y. Chen, C. Cai, J. Liu, "YUV Correction for multi-view video compression", *Proc. IEEE Int. Conf. Pattern Recognition*, Hongkong, pp. 734-737, Aug. 2006.

[6] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization", *Foundations of Computational mathematics*, vol. 9, no. 6, pp. 717-772, 2009.

[7] E. J. Candès and Y. Plan, "Matrix completion with noise", *Proceedings of the IEEE*, vol. 98, no. 6, pp. 925-936, 2010.

[8] Y. Deng, Y. Liu, Q. Dai, et al., "Noisy Depth Maps Fusion for Multiview Stereo Via Matrix Completion", *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 5, pp. 566-582, May 2012.

[9] H. Y. Shum and S. B. Kang, "A review of image-based rendering techniques", *Proc. Visual Communications and Image Processing*, pp. 2-13, June 2000.

[10] H. Ji, C. Liu, Z. Shen, and Y. Xu, "Robust video denoising using low rank matrix completion", *Proc. IEEE Computer Vision and Pattern Recognition*, San Francisco, USA, pp. 1791-1798, June 2010.

[11] Z. Liu and L. Vandenberghe, "Interior-point method for nuclear norm approximation with application to system identification", *SIAM Journal on Matrix Analysis and Applications*, vol. 31, no. 3, pp. 1235-1256, 2009.

[12] S. Ma, D. Goldfarb, L. Chen, "Fixed point and Bregman iterative methods for matrix rank minimization", *Mathematical Programming*, vol. 128, no. 1, pp. 321-353, 2011.

[13] Y. Su, A. Vetro, and A. Smolic, "Common test conditions for multiview video coding", *JVT-T207*, Klagenfurt, Austria, 2006.

[14] C.L. Zitnick, S.B. Kang, M. Uyttendaele, et al., "High-quality video view interpolation using a layered representation", *ACM Transactions on Graphics (TOG)*, vol. 23, no. 3, pp. 600-608, 2004.

[15] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, "Image quality assessment: From error visibility to structural similarity", *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, Apr. 2004.

[16] A. Kaup and U. Fecker, "Analysis of multi-reference block matching for multi-view video coding", *Proc. IEEE Workshop on Digital Broadcasting*, pp. 33-39, 2006.