MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Depth-weighted group-wise principal component analysis for Video foreground/background separation

Tian, D.; Mansour, H.; Vetro, A.

TR2015-111 September 2015

Abstract

We propose a depth-weighted group-wise PCA (DG-PCA) approach to separate moving foreground pixels from the background of a video acquired by a moving camera. Our approach utilizes a corresponding depth signal in addition to the video signal. The problem is formulated as a weighted l2,1- norm PCA problem with depth-based group sparsity being introduced. In particularly, dynamic groups are first generated solely based on depth, and then an iterative solution using depth to define the weights in l2,1-norm is developed. In addition, we propose a depth-enhanced homography model for global motion compensation before the DG-PCA method is executed. We demonstrate through experiments on an RGBD dataset the superiority of the proposed DG-PCA approach over conventional robust PCA methods.

2015 IEEE International Conference on Image Processing (ICIP)

DEPTH-WEIGHTED GROUP-WISE PRINCIPAL COMPONENT ANALYSIS FOR VIDEO FOREGROUND/BACKGROUND SEPARATION

Dong Tian, Hassan Mansour, Anthony Vetro

Mitsubishi Electric Research Labs (MERL)
Cambridge, Massachusetts, USA
{tian, mansour, avetro}@merl.com

ABSTRACT

We propose a depth-weighted group-wise PCA (DG-PCA) approach to separate moving foreground pixels from the background of a video acquired by a moving camera. Our approach utilizes a corresponding depth signal in addition to the video signal. The problem is formulated as a weighted $l_{2,1}$-norm PCA problem with depth-based group sparsity being introduced. In particularly, dynamic groups are first generated solely based on depth, and then an iterative solution using depth to define the weights in $l_{2,1}$-norm is developed. In addition, we propose a depth-enhanced homography model for global motion compensation before the DG-PCA method is executed. We demonstrate through experiments on an RGB-D dataset the superiority of the proposed DG-PCA approach over conventional robust PCA methods.

***Index Terms*—** Foreground /Background separation, principal component analysis, depth-based group sparsity, global motion compensation

1. INTRODUCTION

Video foreground / background (FG/BG) separation provides advanced functionality in applications such as video surveillance, human-computer interaction, and panoramic photography, where moving foreground objects are separated from the background of the video signal. For example, it can help improve object detection / classification, trajectory analysis, and unusual motion detection leading to high level understanding of events in an image sequence.

Among other statistical representation based approaches, robust Principal Component Analysis (RPCA) [1] stands out and has attracted a lot of interest from researchers in recent years. The RPCA problem assumes that an observed video signal $\mathbf{B} \in \mathbb{R}^{m\times n}$ can be decomposed into a low rank component $\mathbf{X} \in \mathbb{R}^{m\times n}$ and a complementary sparse component $\mathbf{S} \in \mathbb{R}^{m\times n}$, and thus the FG/BG separation is formulated as an optimization problem for $\mathbf{X}$ and $\mathbf{S}$, e.g. in [1],

$$(\mathbf{X}, \mathbf{S}) = \arg\min_{\mathbf{X},\mathbf{S}} \|\mathbf{X}\|_* + \lambda\|\mathbf{S}\|_1, \text{ s.t. } \mathbf{B} = \mathbf{X} + \mathbf{S}, \quad (1)$$

where $\|.\|_*$ is the nuclear norm of a matrix and $\|.\|_1$ is l_1-norm of a vectorization of the matrix. The solution to the RPCA problem involves computing a full or partial singular value decomposition (SVD) at every iteration. To avoid the resulting complexity, several techniques, such as, Low-Rank Matrix Fitting (LMaFit) [2, 3] have proposed using low rank factors and optimize over the factors in order to limit the computational complexity. Due to the significant reduction in computing complexity, in this paper, we adopt the idea of factorization on the low-rank component by representing $\mathbf{X} = \mathbf{L}\mathbf{R}^T$, where $\mathbf{L} \in \mathbb{R}^{m\times r}$, $\mathbf{R} \in \mathbb{R}^{n\times r}$, and $r \geq \text{rank}(\mathbf{X})$.

In recent years, the standard sparsity concept in Compresive Sensing was extended into the development of RPCA methods to incorporate structured sparsity. This was mainly motivated by the observation that the sparse data are often not randomly located but tend to cluster together. For example, Huang et al. [4] proposed a learning formulation called dynamic group sparsity (DGS) that uses a pruning step in selecting the sparse components that favor local clustering. Another approach proposed in [5] and [6] enforces group sparsity by replacing the l_1-norm in (1) with a mixed $l_{2,1}$-norm defined as,

$$\|\mathbf{S}\|_{2,1} = \sum_{g=1}^{s} w_g \|\mathbf{S}_g\|_2, \quad (2)$$

where $\mathbf{S}_g$ is the component corresponding to group g, $g = 1, ..., s$, and w_g's are weights associated to each group. The resulting problem formulation is given by

$$(\mathbf{X}, \mathbf{S}) = \arg\min_{\mathbf{X},\mathbf{S}} \|\mathbf{X}\|_* + \lambda\|\mathbf{S}\|_{2,1}, \text{ s.t. } \mathbf{B} = \mathbf{X} + \mathbf{S}. \quad (3)$$

Though the most recent FG/BG separation approaches in the PCA-family have been shown to be quite effective for sequences with static background, their separation performance degrades for image sequences with a moving camera, even with limited jitter motion. A global motion compensation (MC) was proposed in [7] to align the images before applying a RPCA-based FG/BG separation method.

On the other hand, video with corresponding depth maps have become ubiquitous, especially with the rapid growth of depth sensors like Microsoft Kinect and the advancement of depth estimation algorithms from stereo images. In fact, depth has been utilized in the FG/BG separation tasks for over a decade. Since 1999, [8] and [9] have reported that jointly using depth and color data produces superior separation results. More recent work e.g. [10] demonstrated that a depth-enhanced method DECB can better deal with illumination changes, shadows, reflections and camouflage, than their conventional counterpart [11]. Camplani, et al. [12] proposed to jointly consider color data and its corresponding dense depth data in a classifier approach. However, to the best of our knowledge, these depth assisted methods have not been studied for FG/BG separation in a moving camera sequence.

The remainder of the paper is organized as follows. We propose a depth-weighted group-wise PCA method (DG-PCA) to implement a novel depth group sparsity approach in Section 2. In order to address the challenges for a camera-moving sequence, we propose a novel global motion model refined by depth to align the images in Section 3. In Section 4, we conduct experiments to compare the proposed DG-PCA method with a conventional method w/o depth as input. Finally, we conclude the work in Section 5.

2. DEPTH-WEIGHTED GROUP-WISE PCA

In the RPCA problem formulation, the video background is assumed to have small variations that can be modeled using the low rank component $\mathbf{X}$. Foreground objects, represented by $\mathbf{S}$, are assumed to be sparse and have a different type of motion than the background. Existing FG/BG separation algorithms, e.g. [2, 3], do not incorporate the foreground object structure in the separation. Here, we propose a structured group-sparsity based PCA method that can overcome some larger variations in the background, e.g. from misalignment in global motion compensation on a camera moving sequence.

2.1. A Factorized RPCA method

First, we discuss a benchmark unstructured sparsity algorithm that characterizes the performance of RPCA techniques. This benchmark method, which we will refer to as factorized RPCA, uses an augmented Lagrangian alternating direction method (ADM) similar to LMaFit in [2, 3] to solve the following problem:

$$\begin{aligned}(\mathbf{L},\mathbf{R},\mathbf{S},\mathbf{Y}) = &\arg\min_{\mathbf{L},\mathbf{R},\mathbf{S},\mathbf{Y}}(\frac{1}{2}\|\mathbf{L}\|_F^2 + \frac{1}{2}\|\mathbf{R}\|_F^2 + \\ &\lambda\|\mathbf{S}\|_1 + <\mathbf{Y},\mathbf{E}> + \frac{\mu}{2}\|\mathbf{E}\|_F^2),\end{aligned} \tag{4}$$

where $\mathbf{Y} \in \mathbb{R}^{m\times n}$ is the Lagrangian multiplier, λ and μ are weighting factors, and $\mathbf{E} = \mathbf{B} - \mathbf{L}\mathbf{R}^T - \mathbf{S}$. Note that the nuclear norm $\|\mathbf{X}\|_*$ in (1) is replaced by $\frac{1}{2}\|\mathbf{L}\|_F^2 + \frac{1}{2}\|\mathbf{R}\|_F^2$ in (4), where $\mathbf{X} = \mathbf{L}\mathbf{R}^T$, based on the observation in [13] that

$$\|\mathbf{X}\|_* = \inf_{\mathbf{L},\mathbf{R}} \frac{1}{2}\|\mathbf{L}\|_F^2 + \frac{1}{2}\|\mathbf{R}\|_F^2, \text{ s.t. } \mathbf{X} = \mathbf{L}\mathbf{R}^T. \tag{5}$$

Algorithm 1 Factorized RPCA algorithm to solve problem (4) – Benchmark

```
Require: Input data $\mathbf{B}$, $\lambda$, $\mu$, error tolerance $\tau$, maximum iteration number $N$
 1: Init: $i \leftarrow 0$, $\mathbf{L}_i$ and $\mathbf{R}_i \leftarrow$ random matrix
 2: repeat
 3:     $\mathbf{L}_{i+1} = (\mu(\mathbf{B}-\mathbf{S}_i) + \mathbf{Y}_i)\mathbf{R}_i(\mathbf{I} + \mu\mathbf{R}_i^T\mathbf{R}_i)^{-1}$
 4:     $\mathbf{R}_{i+1} = (\mu(\mathbf{B}-\mathbf{S}_i) + \mathbf{Y}_i)^T\mathbf{L}_{i+1}(\mathbf{I} + \mu\mathbf{L}_{i+1}^T\mathbf{L}_{i+1})^{-1}$
 5:     $\mathbf{S}_{i+1} = \mathcal{S}_{\lambda/\mu}(\mathbf{B} - \mathbf{L}_{i+1}\mathbf{R}_{i+1}^T + \mu^{-1}\mathbf{Y}_i)$
 6:     $\mathbf{E} = \mathbf{B} - \mathbf{L}_{i+1}\mathbf{R}_{i+1}^T - \mathbf{S}_{i+1}$
 7:     $\mathbf{Y}_{i+1} = \mathbf{Y}_i + \mu\mathbf{E}$
 8:     $i \leftarrow i + 1$
 9: until $i \geq N$ or $\|\mathbf{E}\|_F \leq \tau$
10: return $\mathbf{L}$, $\mathbf{R}$, $\mathbf{S}$, $i$ and $\|\mathbf{E}\|_F$
```

Algorithm 1 shows the iterations used to solve (4). Note in step 5 the soft-thresholding operator is given,

$$\mathcal{S}_{\lambda/\mu}(\mathbf{r}) = \text{sign}(\mathbf{r})\max(|\mathbf{r}| - \lambda/\mu, 0), \tag{6}$$

with $\mathbf{r} = \mathbf{B} - \mathbf{L}\mathbf{R}^T + \frac{1}{\mu}\mathbf{Y}$, which does not impose structure on the sparse component.

2.2. Depth-weighted Group-wise PCA

In practical image sequences, the foreground objects (sparse components) tend to be clustered both spatially and temporally rather than evenly distributed. This observation led to the introduction of group sparsity into RPCA approaches by [4, 5, 6], pushing the sparse component into more structured groups. Our method utilizes the depth map of the video sequence to define the group structures in a depth-weighted group-wise PCA (DG-PCA) method.

In order to deal with structured sparsity, we replace the l_1-norm in the factorized RPCA problem with a mixed $l_{2,1}$-norm as defined in (2). The resulting problem is shown below:

$$\begin{aligned}(\mathbf{L},\mathbf{R},\mathbf{S},\mathbf{Y}) = &\arg\min_{\mathbf{L},\mathbf{R},\mathbf{S},\mathbf{Y}}(\frac{1}{2}\|\mathbf{L}\|_F^2 + \frac{1}{2}\|\mathbf{R}\|_F^2 + \\ &\lambda\|\mathbf{S}\|_{2,1} + <\mathbf{Y},\mathbf{E}> + \frac{\mu}{2}\|\mathbf{E}\|_F^2).\end{aligned} \tag{7}$$

Algorithm 2 describes the proposed DG-PCA framework. In order to define pixel groups $\mathbf{G}$ using the depth map $\mathbf{D}$, an operator $\mathcal{G}(\mathbf{D})$ segments the depth map into s groups using the following procedure. Suppose the depth level ranges from 0 to 255, a pixel with depth value d will be classified into group $g = \lfloor d/\frac{256}{s} \rfloor + 1$. Consequently, the input data $\mathbf{B}$ can be clustered into $\mathbf{B}_g$ with $g \in \{1, .., s\}$. Each $\mathbf{B}_g$ is composed of elements from $\mathbf{B}$ which is marked into segment g. In the same way, $\mathbf{L}_g$, $\mathbf{R}_g$, and Lagrangian multiplier $\mathbf{Y}_g$ are also grouped.

Next, the operator $\mathcal{S}_{\lambda/\mu,g}$ in Algorithm 2 is a group-wise soft-thresholding, as shown below,

$$\mathcal{S}_{\lambda/\mu,g}(\mathbf{r}_g) = \max(\|\mathbf{r}_g\|_2 - w_g\lambda/\mu, 0)\frac{\mathbf{r}_g}{\|\mathbf{r}_g\|_2 + \epsilon}, \tag{8}$$

Algorithm 2 Depth-weighted group-wise PCA algorithm to solve problem (7) – Proposed in this paper

```
Require: Input data $\mathbf{B}$, $\lambda$, $\mu$, error tolerance $\tau$, maximum
    iteration number $N$, and depth map $\mathbf{D}$
 1: Init: $i \leftarrow 0$, $\mathbf{L}_i$ and $\mathbf{R}_i \leftarrow$ random matrix, $\mathbf{G} \leftarrow \mathcal{G}(\mathbf{D})$
 2: repeat
 3:     $\mathbf{L}_{i+1} = (\mu(\mathbf{B} - \mathbf{S}_i) + \mathbf{Y}_i)\mathbf{R}_i(\mathbf{I} + \mu\mathbf{R}_i^T\mathbf{R}_i)^{-1}$
 4:     $\mathbf{R}_{i+1} = (\mu(\mathbf{B} - \mathbf{S}_i) + \mathbf{Y}_i)^T\mathbf{L}_{i+1}(\mathbf{I} + \mu\mathbf{L}_{i+1}^T\mathbf{L}_{i+1})^{-1}$
 5:     $\mathbf{S}_{i+1,g} = \mathcal{S}_{\lambda/\mu,g}(\mathbf{B}_g - \mathbf{L}_{i+1,g}\mathbf{R}_{i+1,g}^T + \mu^{-1}\mathbf{Y}_{i,g})$
 6:     $\mathbf{E} = \mathbf{B} - \mathbf{L}_{i+1}\mathbf{R}_{i+1}^T - \mathbf{S}_{i+1}$
 7:     $\mathbf{Y}_{i+1} = \mathbf{Y}_i + \mu\mathbf{E}$
 8:     $i \leftarrow i + 1$
 9: until $i \geq N$ or $\|\mathbf{E}\|_F \leq \tau$
10: return $\mathbf{L}$, $\mathbf{R}$, $\mathbf{S}$, $i$ and $\|\mathbf{E}\|_F$
```

where $\mathbf{r}_g = \mathbf{B}_g - \mathbf{L}_g\mathbf{R}_g^T + \frac{1}{\mu}\mathbf{Y}_g$, and ϵ is a small constant to avoid division by 0, and w_g defines group weights in (2). Since a foreground object has higher chances to be closer to the camera, i.e., to have a higher depth value than a background object, we propose the following equation to set group weights,

$$w_g = c^{1-\frac{d_g}{255}}, \tag{9}$$

where c is some constant, and d_g is the mean depth value of pixels in group g. w_g is equal to 1 for objects nearest to the camera, $d = 255$, and it is equal to c for objects farthest to the camera, $d = 0$. The choice of c controls the value of the threshold that permits foreground pixels to be selected based on their location in the depth field. Finally, after $\mathbf{S}_g$ is calculated for each group g, the sparse component $\mathbf{S}$ is obtained by summing up all $\mathbf{S}_g$ together.

Note that the above setup favors group structures where the foreground objects are closer to the camera. It is also possible within our framework to define the groups as the sets of pixels that are spatially connected and have a constant depth, or connected pixels where the spatial gradient of the depth is constant.

Last, it is worthwhile to mention that the nuclear norm equivalent items $\frac{1}{2}\|\mathbf{L}\|_F^2 + \frac{1}{2}\|\mathbf{R}\|_F^2$ in problem (7) has contributions to make Algorithm 2 more numerical stable. Without the nuclear norm, $(\mathbf{I} + \mu\mathbf{R}_i^T\mathbf{R}_i)^{-1}$ in step 3 of Algorithm 2 will become $(\mu\mathbf{R}_i^T\mathbf{R}_i)^{-1}$, which is unstable when the matrix $\mathbf{R}_i^T\mathbf{R}_i$ is singular, for example, when the image is dark with $\mathbf{B}, \mathbf{L}, \mathbf{R} \approx \mathbf{0}$.

3. DEPTH-ENHANCED HOMOGRAPHY MODEL

With moving camera sequences, the motion in the background no longer satisfies the low-rank assumption. Hence, in order to apply RPCA, global motion compensation using a homography model was proposed in [7] as a pre-processing step on the image sequence prior to using RPCA.

One approach for performing global motion compensation is to compute a homography model for the image sequence, see e.g. [14] and [15]. In an 8-parameter homography model $\mathbf{h} = [h_1, h_2, ..., h_8]^T$, the corresponding pixel $\mathbf{x}_1 = (x_1, y_1)^T$ in the current image and $\mathbf{x}_2 = (x_2, y_2)^T$ in its reference image are related as below,

$$x_2 = \frac{h_1 + h_3x_1 + h_4y_1}{1 + h_7x_1 + h_8y_1} \text{ and } y_2 = \frac{h_2 + h_5x_1 + h_6y_1}{1 + h_7x_1 + h_8y_1} \tag{10}$$

Given local motion information associating a pixel location $\mathbf{x}_1$ in the current image to its corresponding location $\mathbf{x}_2$ in the reference image, the homography model $\mathbf{h}$ can be estimated by solving a typical least square (LS) problem: $\mathbf{b} = \mathbf{Ah}$, where $\mathbf{b}$ is a vector composed by stacking the vectors $\mathbf{x}_2$'s, and the rows of $\mathbf{A}$ corresponding to each $\mathbf{x}_2$ is specified as follows:

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & x_1 & y_1 & 0 & 0 & -x_1x_2 & -y_1x_2 \\ 0 & 1 & 0 & 0 & x_1 & y_1 & -x_1y_2 & -y_1y_2 \end{pmatrix}. \tag{11}$$

In practice, the local motion information associating pixel locations is often inaccurate. In this case, the full 8-parameter model is shown to be sensitive to errors in the motion information. Hence, a reduced number of parameters in homography model is often preferred, thus limiting the types of motion in the scene. For example, 2-, 4- and 6-parameter models correspond to translational only, geometric and affine models, respectively, by setting some coefficients in $\mathbf{h}$ to be zero. In this paper, we select the 4-parameter geometric model as our starting point, where we have $\mathbf{h} = [h_1, h_2, h_3, 0, 0, h_6, 0, 0]^T$. Note that the proposed extension below based on depth is not limited to the geometric model.

However, motion in a video sequence is generally not planar. Therefore, even after a careful selection of the conventional homography model, it is still very common to find large motion estimation errors, which would dramatically degrade the detection rate in a subsequent PCA-like algorithm. Therefore, we propose a depth-enhanced homography model. Specifically, 6 new parameters related to depth are added, and we have $\mathbf{h} = [h_1, ..., h_8, h_9, ..., h_{14}]^T$. Let z_1 and z_2 stand for the depth of the corresponding pixels, and the proposed depth-enhanced homography model is given as follows,

$$\begin{aligned} x_2 &= \frac{h_1 + h_3x_1 + h_4y_1 + h_9z_1}{1 + h_7x_1 + h_8y_1}, \\ y_2 &= \frac{h_2 + h_5x_1 + h_6y_1 + h_{10}z_1}{1 + h_7x_1 + h_8y_1}, \\ z_2 &= \frac{h_{11} + h_{12}x_1 + h_{13}y_1 + h_{14}z_1}{1 + h_7x_1 + h_8y_1}. \end{aligned} \tag{12}$$

Note in the above equation, depth value 0 means the object is at ∞ from the camera. A larger depth value means that the object is closer to the camera. Certain simplification is possible for simpler sequences. For example, if $z_2 = z_1$ is assumed, the motion will be limited within the same depth plane.

4. EXPERIMENTS

To evaluate the performance of the proposed DG-PCA approach, we selected four fr3/walking sequences under the "dynamic objects" category in the RGB-D benchmark provided by TUM [16]. The dataset contains dynamic objects with a low- to high-level global motion, so it serves the purpose of evaluating the FG/BG separation tasks, although the dataset was originally intended for other study purposes.

The accompanying depth in the dataset is captured by Microsoft Kinect sensor and denoted by z. In our work, the depth map d was computed from z as per (13) before being fed to our approach DG-PCA, where z_{near} and z_{far} denote the nearest and farthest depth extracted from the raw depth data z.

$$d = 255 \times \frac{\frac{1}{z} - \frac{1}{z_{far}}}{\frac{1}{z_{near}} - \frac{1}{z_{far}}}. \tag{13}$$

In order to perform FG/BG separation, a current image is processed together with one previous image in the following way. The two images are first aligned using global motion compensation with and without depth-enhanced homography model as described in Section 3. The two aligned frames are then processed using the factorized RPCA and the DG-PCA approaches to extract the background (low-rank component) $\mathbf{X} = \mathbf{L}\mathbf{R}^T$ and foreground (sparse component) $\mathbf{S}$. The rank of the $\mathbf{X}$ is set to 2. The pixels with value larger than 2 in $\mathbf{S}$ are marked as foreground and others as background.

We used sequence fr3/walking_static with minor camera motion to tune the algorithm parameters and then run tests on the other three sequences with higher motion. The parameter λ is chosen (7) at image level to be,

$$\lambda = 0.05(\|\mathbf{r}\|_2 / \sqrt{\text{size}(\mathbf{B})}) \times \mu, \tag{14}$$

where a constant 0.05 is selected empirically to limit the iteration step for a finer background subtraction. When updating group-wise sparse component in Algorithm 2, $\lambda_g = \lambda\sqrt{\text{size}(\mathbf{B}_i)}$ instead of the image level λ is used in (8). This choice of λ_g makes it dependent on the number of pixels in each group as evident in the multiplication $\sqrt{\text{size}(\mathbf{B}_i)}$ since the l_2 norm of the group is being thresholded instead of individual pixels. Moreover, we set $c = 10$ in (9).

The number of groups s from depth segmentation was found not very sensitive. We tested s in the range $[16, 32]$, and there is no significant differences in performance. We use $s = 32$ to report results herein. Moreover, we apply a 5×5 median filter to the depth-based grouping map $\mathbf{G}$ before proceeding to the group-wise soft-thresholding in order to limit the effect of noise in the depth map.

We studied the DG-PCA approach w/o and w/ depth-refined global motion compensation. Fig. 1 shows 5 snapshots across fr3/walking_rpy with 910 frames at VGA resolution which has the greatest global motion in the dataset. The figures show that the two DG-PCA methods (row 4 and 5) produce a much cleaner foreground segmentation compared to the factorized RPCA approach (row 3). For example, in the third snapshot, the person walking at a further distance behind the office partition can also be detected successfully by DG-PCA. Comparing DG-PCA without depth-enhanced global MC (row 4) and with depth-enhanced global MC (row 5), shows that the depth-enhanced homography model helps provide even better motion alignment compared to the conventional homography model which is evident in the improved foreground segmentation in snapshot 4 and the improved background suppression in snapshot 5.

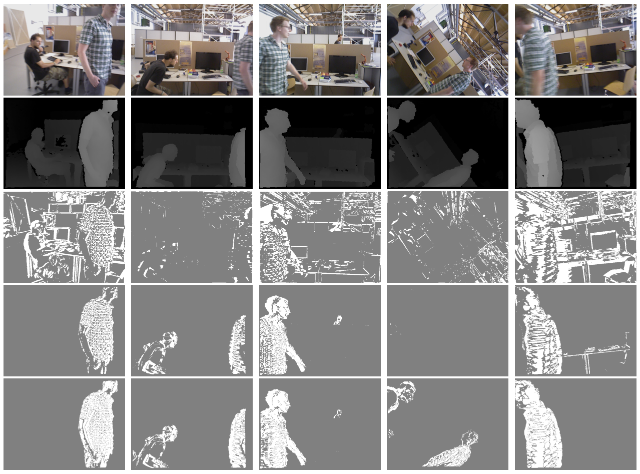

Fig. 1. Performance evaluation. Row 1: color images. Row 2: depth maps. Row 3: Factorized RPCA. Row 4: DG-PCA w/o depth-refined global MC. Row 5: DG-PCA with depth-refined global MC.

Finally, we note that we observed some flickering effects with DG-PCA when playing the foreground mask over time. However, we believe that the problem can be alleviated through simple post-processing with temporal correlation.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed a novel PCA framework that utilizes depth-based group sparsity and depth-weighted $l_{2,1}$-norm for robust and efficient separation of foreground and background in video sequences. To improve robustness for video sequences with moving cameras, we also proposed to utilize the depth information in a depth-enhanced homography model for global motion compensation. Experimental results demonstrate that the proposed method, which combines motion and depth segmentation, significantly outperforms a conventional RPCA approach. One area of future work is to improve the temporal consistency of the results, which could be achieved through post-processing or accounting for the results of the prior frame in the optimization of the current frame.

6. REFERENCES

[1] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright, “Robust principal component analysis?,” *Journal of the ACM (JACM)*, vol. 58, no. 3, pp. 11, 2011.

[2] Zaiwen Wen, Wotao Yin, and Yin Zhang, “Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm,” *Mathematical Programming Computation*, vol. 4, no. 4, pp. 333–361, 2012.

[3] Yuan Shen, Zaiwen Wen, and Yin Zhang, “Augmented lagrangian alternating direction method for matrix separation based on low-rank factorization,” *Optimization Methods and Software*, vol. 29, no. 2, pp. 239–263, 2014.

[4] Junzhou Huang, Xiaolei Huang, and Dimitris Metaxas, “Learning with dynamic group sparsity,” in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 64–71.

[5] Wei Deng, Wotao Yin, and Yin Zhang, “Group sparse optimization by alternating direction method,” in *SPIE Optical Engineering+ Applications*. International Society for Optics and Photonics, 2013, pp. 88580R–88580R.

[6] Zhangjian Ji, Weiqiang Wang, and Ke Lv, “Foreground detection utilizing structured sparse model via l1, 2 mixed norms,” in *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on*. IEEE, 2013, pp. 2286–2291.

[7] Hassan Mansour and Anthony Vetro, “Video background subtraction using semi-supervised robust matrix completion,” in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 6528–6532.

[8] Gaile Gordon, Trevor Darrell, Michael Harville, and John Woodfill, “Background estimation and removal based on range and color,” in *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, 1999, vol. 2, pp. –464 Vol. 2.

[9] Michael Harville, Gaile Gordon, and John Woodfill, “Foreground segmentation using adaptive mixture models in color and depth,” in *Detection and Recognition of Events in Video, 2001. Proceedings. IEEE Workshop on*. IEEE, 2001, pp. 3–11.

[10] Enrique J Fernandez-Sanchez, Javier Diaz, and Eduardo Ros, “Background subtraction based on color and depth using active sensors,” *Sensors*, vol. 13, no. 7, pp. 8895–8915, 2013.

[11] Kyungnam Kim, Thanarat H Chalidabhongse, David Harwood, and Larry Davis, “Real-time foreground–background segmentation using codebook model,” *Real-time imaging*, vol. 11, no. 3, pp. 172–185, 2005.

[12] Massimo Camplani and Luis Salgado, “Background foreground segmentation with rgb-d kinect data: An efficient combination of classifiers,” *Journal of Visual Communication and Image Representation*, vol. 25, no. 1, pp. 122 – 136, 2014.

[13] Nathan Srebro, *Learning with matrix factorizations*, Ph.D. thesis, Citeseer, 2004.

[14] Yeping Su, Ming-Ting Sun, and Vincent Hsu, “Global motion estimation from coarsely sampled motion vector field and the applications,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 15, no. 2, pp. 232–242, 2005.

[15] Yue-Meng Chen and Ivan V Bajic, “A joint approach to global motion estimation and motion segmentation from a coarsely sampled motion vector field,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 21, no. 9, pp. 1316–1328, 2011.

[16] Jürgen Sturm, Nikolas Engelhard, Felix Endres, Wolfram Burgard, and Daniel Cremers, “A benchmark for the evaluation of rgb-d slam systems,” in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*. IEEE, 2012, pp. 573–580.