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

Moving object segmentation using depth and optical flow in car driving sequences

Kao, J.-Y.; Tian, D.; Mansour, H.; Vetro, A.; Ortega, A.

TR2016-126 August 2016

Abstract

Segmentation of moving objects in a scene is difficult for non-stationary cameras, and especially challenging in the presence of fast and unstable egomotion, e.g., as encountered with car-mounted cameras or wearable devices. Based on an analysis of motion vanishing points of the scene and estimated depth, a geometric model that relates extracted 2D motion to a 3D motion field relative to the camera is derived. Observing that the 3D motion field is piecewise smooth, a constrained optimization problem that considers group sparsity is formulated to recover the 3D motion field from the 2D motion. The recovered 3D motion field is then clustered to provide the segmentation of moving objects. Experiments are performed using the KITTI Vision Benchmark Suite and demonstrate that the proposed framework provides a dense segmentation of moving objects that is robust to the challenging conditions inherent with car driving sequences.

IEEE International Conference on Image Processing (ICIP)

MOVING OBJECT SEGMENTATION USING DEPTH AND OPTICAL FLOW IN CAR DRIVING SEQUENCES

Jiun-Yu Kao[2*] *Dong Tian*[1] *Hassan Mansour*[1] *Anthony Vetro*[1] *Antonio Ortega*[2]

[1] Mitsubishi Electric Research Labs (MERL),
201 Broadway, Cambridge, MA 02139, USA
[2] Department of Electrical Engineering, University of Southern California,
3740 McClintock Ave., Los Angeles, CA 90089, USA

ABSTRACT

Segmentation of moving objects in a scene is difficult for non-stationary cameras, and especially challenging in the presence of fast and unstable egomotion, e.g., as encountered with car-mounted cameras or wearable devices. Based on an analysis of motion vanishing points of the scene and estimated depth, a geometric model that relates extracted 2D motion to a 3D motion field relative to the camera is derived. Observing that the 3D motion field is piece-wise smooth, a constrained optimization problem that considers group sparsity is formulated to recover the 3D motion field from the 2D motion. The recovered 3D motion field is then clustered to provide the segmentation of moving objects. Experiments are performed using the KITTI Vision Benchmark Suite and demonstrate that the proposed framework provides a dense segmentation of moving objects that is robust to the challenging conditions inherent with car driving sequences.

Index Terms— Motion segmentation, foreground / background separation, freely moving cameras, pixel-wise foreground labeling, 3D motion reconstruction

1. INTRODUCTION

Detecting or segmenting moving objects from videos is an essential step for video analysis and understanding and thus there is abundant literature dealing with this problem. Due to the large variety in the video types and contents, an effective algorithm is not always guaranteed for all types of videos. When the videos are captured via static cameras, traditional background subtraction techniques usually provide quite effective solutions to the detection and segmentation of targets. However, more and more of today's videos are captured from moving platforms, e.g., cameras mounted on vehicles or drones, where the traditional background subtraction algorithms are not applicable.

When the camera motion is known or the scene geometry is restricted, such as pan-tilt-zoom cameras, then some extensions of the traditional background subtraction methods via motion compensation can be effective. For example, if a planar scene is monitored, a geometric transformation between images of a plane can be characterized as a homography, which includes 8 parameters. Applying RANSAC [1] or its variants can perform robust estimation for the homography with matches of points and can thereby establish an extended background map. Then adopting any traditional methods such as Gaussian models or mixture of Gaussians can statistically model the pixel process and furthermore detecting the objects.

However, when the cameras are freely moving, motion compensation based methods cannot robustly estimate the image mosaic and thereby are not applicable. Instead, recently proposed methods mostly use motion segmentation approaches as the building blocks. Motion segmentation methods can be roughly categorized into four categories: statistical, factorization-based, algebraic decomposition and spectral clustering techniques. Statistical approaches alternate between assigning trajectories to subspaces and refitting subspaces to their assigned points; this may be achieved using the EM algorithm in [2]. Factorization-based approaches such as [3][4] directly factorize the matrix of trajectories. Algebraic decomposition approaches such as GPCA [5] formulate motion segmentation as a problem of subspace separation. Another alternative is robust PCA where the background scene is modeled as a low-dimensional subspace. The foreground objects can be segmented from the background in [6][7] when the camera is stationary. Finally, spectral clustering based approaches first utilize local information to compute pairwise similarity between keypoint trajectories, from which an affinity matrix is generated and followed by clustering the trajectories into separate subspaces as described in [8]. One such example is sparse subspace clustering (SSC) [9] where the affinity matrix is constructed by the coefficients when attempting to represent each trajectory as a sparse linear combination of others. Some of the above mentioned motion segmentation approaches, such as SSC, can handle the situation when small camera movement exists, e.g., with a hand-held camera.

Unfortunately, all the above motion segmentation approaches are not directly applicable to the challenging situation when the cameras are freely and fast moving, such as in the car driving sequences. However, these approaches have been used as fundamental building blocks in recent works which attempt to tackle the motion segmentation problem for this more difficult case. One example is [10], which first utilizes RANSAC to robustly estimate a compact trajectory basis to model the background, so that the background is subtracted by removing those trajectories lying in the subspace spanned with the estimated basis. An optimization problem to maximize a posterior function is then solved for a pixel-wise foreground/background labeling. Their method works with freely moving camera but fails in car driving sequences since it highly deviates from the orthographic projection assumption. A more recent work by Elqursh and Elgammal [11] also maintains representations for both the background and the foreground by formulating motion segmentation as a manifold separation problem. Long term trajectories and a Bayesian filtering framework are utilized to include both the motion dependencies and pixel-level appearance. Although moving objects can be successfully segmented with their method, long term trajectories are re-

*This work was done when Jiun-Yu Kao worked at MERL.

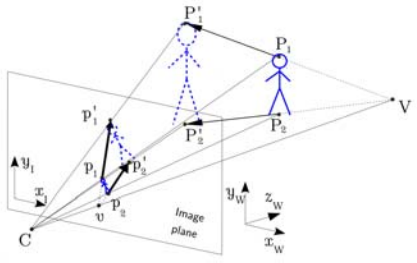

Fig. 1. Illustration of *motion vanishing point* in 3D world coordinate system and image coordinate system. [12]

quired in their algorithm, which may lead to issues in high memory consumption and large latency.

Our proposed approach is also based on motion segmentation method. In contrast to [11], our method only requires pixel-wise motion vectors rather than long-term trajectories. Besides, we assume the pixel-wise disparity can also be provided or estimated. We first derive a model to relate the given 2D motion field to the 3D motion relative to the camera associated with each pixel, which is what we would like to solve for. As the model only provides an under-determined set of equations, we further include two desired properties to restrict the solution set to what is desired. After solving for the pixel-wise 3D motions, standard spectral clustering [8] is utilized to provide the final labeling.

The rest of the paper is organized as follows. Section 2 illustrates the insights and derivation of the geometric model between two motion fields. Section 3 explains the desired properties in the solved 3D motions and thereby the formulation of the optimization problem, along with the algorithm to solve it. Section 4 provides the experimental results on KITTI benchmark along with briefly explaining the segmentation process. Section 5 concludes this paper.

2. 3D MOTION ANALYSIS

2.1. Motion Vanishing Point

Strong perspective effects appear all the time for outdoor sequences, especially for car driving sequences, captured by a camera mounted on a moving car. According to [12], the concept of *motion vanishing point* (MVP) originates from the observation on how 3D motions in the scene project to the image plane as described in Figure 1. Due to the perspective effect, a pair of parallel 3D motion vectors, $P_1 \rightarrow P_1'$ and $P_2 \rightarrow P_2'$, will intersect at point V at infinity in the 3D coordinate system. When instead looking at their projections to the image plane of a camera centered at C, the corresponding pair of 2D motion vectors, $p_1 \rightarrow p_1'$ and $p_2 \rightarrow p_2'$, will intersect at a motion vanishing point v. In fact, the 2D motion vectors of all points on the same object will share the same motion vanishing point v in the image plane once if they exhibit the same motion in the 3D world. From the above observations, we notice that there should be a relationship between 2D and 3D motion vectors. Having a clearer idea about this relationship can provide a way to recover/estimate the 3D motions out of the 2D motion vectors. In the following section, this relationship is mathematically derived and stated. We also demonstrates how a motion vanishing point is represented, which was not provided in [12].

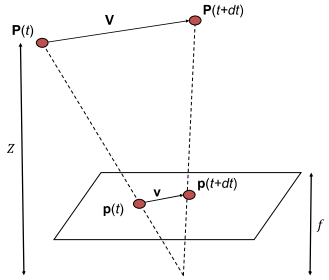

Fig. 2. Illustration of the relationship between 2D and 3D motion vectors.

2.2. Motion Field Model

In order to derive the model between 2D motions in the image plane and the 3D motions in the scene, we have $\mathbf{P}(t)$ as a function of time which denotes the coordinates of a moving 3D point. The velocity of this point will be $\mathbf{V} = \frac{\mathrm{d}\mathbf{P}(t)}{\mathrm{d}t} = (V_x, V_y, V_z)$. Further, $\mathbf{p}(t) = (x(t), y(t), f)$ denotes the projection of $\mathbf{P}(t)$ onto the image plane in a 3D coordinate system, where f is a constant for the focal length. Consequently, the velocity of $\mathbf{p}(t)$ in the image plane can be represented as $v_x = \frac{\mathrm{d}x(t)}{\mathrm{d}t}$, $v_y = \frac{\mathrm{d}y(t)}{\mathrm{d}t}$, and $\mathbf{v} = (v_x, v_y, 0)$ of each point is the 2D motion field of the image. The relationship between 2D and 3D motion field is illustrated in Figure 2.

When perspective projection is assumed, we have,

$$\mathbf{p}(t) = \frac{f}{Z} \cdot \mathbf{P}(t), \tag{1}$$

where Z is the depth of 3D point from the camera center. Then we can derive the motion vector $\mathbf{v}$ by differentiating $\mathbf{p}(t)$ with respect to t as follows,

$$\mathbf{v} = \frac{\mathrm{d}\mathbf{p}(t)}{\mathrm{d}t} = f \cdot \frac{\mathbf{V} \cdot Z - V_z \cdot \mathbf{P}(t)}{Z^2} = \frac{f}{Z}\mathbf{V} - \frac{V_z}{Z}\mathbf{p}(\mathbf{t}) \tag{2}$$

Writing the expressions for v_x and v_y respectively leads to the following,

$$v_x = \frac{f\,V_x - V_z\,x}{Z} \qquad v_y = \frac{f\,V_y - V_z\,y}{Z} \tag{3}$$

Furthermore, (2) can be reorganized as

$$\mathbf{v} = \frac{1}{Z}(\mathbf{v}_0 - V_z\,\mathbf{p}(t)) \tag{4}$$

where $\mathbf{v}_0 = (fV_x\,, fV_y)$.

From (4) we can see that, for a group of points exhibiting the same 3D motion, i.e., having constant $V = (V_x, V_y, V_z)$, will also have the same $\mathbf{v}_0$. When the 3D motion is parallel to the plane, i.e. $V_z = 0$, all the motion vectors of this group of points will be parallel to each other as $\mathbf{v} = \frac{1}{Z}\mathbf{v}_0$. When $V_z \neq 0$, all the motion vectors belonging to this group will point towards or away from a common *motion vanishing point*, where at this point the 2D motion field is zero. And the image plane coordinate of the MVP will be $\mathbf{p} = \frac{1}{V_z}\mathbf{v}_0 = (f\frac{V_x}{V_z}\,, f\frac{V_y}{V_z})$, as derived from (4).

We observe that the model given in (4) provides a rigorous explanation of the MVP concept and could be used to recover the 3D motion from the 2D motion vector field with knowledge of the scene depths Z. In next section, we will first show how to formulate the motion segmentation problem and then present a solution to the proposed formulation.

3. MOTION SEGMENTATION BASED ON 3D MOTION CLASSIFICATION

3.1. 3D Motion Structures

Once the 3D motions are estimated, it would build a foundation to segment the scene into regions with different motions. However, this model could only provide an under-determined set of equations, and hence the solutions, i.e., the recovered 3D motions, are not uniquely determined. Therefore, additional properties of the motion structure need to be utilized in order for the solution to be consistent with constraints impose by the physical world.

According to the observations on the real-world car driving sequences, there are two properties in the structure of 3D motions relative to the camera. Firstly, the reconstructed 3D motions $\mathbf{V}$ are desired to have piece-wise smoothness property because the motions should be smooth and near-constant within each object. Secondly, the 3D motions $\mathbf{V}$ in a scene are typically distributed sparsely in terms of l_1-norm assuming that there are only a few dominant motions in the scene at each time. Furthermore, group sparsity in terms of $l_{2,1}$-norm may be utilized considering that there are only a quite limited number of moving objects in the scene, which was utilized in 2D motion in [13]. It is proposed to utilize the two types of properties by integrating two new constraints when solving the 3D motions out of a given 2D motion field, which is to be described in Section 3.2.

3.2. 3D Motion Segmentation Problem Formulation

In order to overcome the under-determined problem in the geometric model in (3) or equivalently in the motion vanishing analysis in (4), we propose to formulate the problem as an energy minimization task after considering the two additional properties: piece-wise smoothness and group sparsity of $\mathbf{V}$. The first energy functional $\mathbf{E}_g$ tries to satisfy the geometric model in (3). For the purpose of piece-wise smoothness, a total variation regularizer $\mathbf{E}_p$ is included in the objective function. Furthermore, an $l_{2,1}$-norm regularizer $\mathbf{E}_s$ is added in order to achieve group sparsity. Therefore, the optimization problem is formulated as follows.

$$\underset{\substack{V_{x_i},V_{y_i},V_{z_i}\\ i=1,\cdots,N}}{\text{minimize}} \quad \mathbf{E}_g + \lambda\mathbf{E}_p + \mu\mathbf{E}_s \tag{5}$$

where the energy functionals $\mathbf{E}_g, \mathbf{E}_p$, and $\mathbf{E}_s$ are given by

$$\begin{aligned} \mathbf{E}_g &= \|\mathbf{AV}-\mathbf{b}\|_2^2 \\ \mathbf{E}_p &= TV_2(\mathbf{V}) \\ \mathbf{E}_s &= \|\mathcal{G}(\mathbf{V})\|_{2,1} \end{aligned} \tag{6}$$

Here, N is the number of pixels in the frame and
$\mathbf{V} = \left(fV_{x_1}\cdots fV_{x_N} fV_{y_1}\cdots fV_{y_N} V_{z_1}\cdots V_{z_N}\right)^T$,
$\mathbf{b} = \left(v_{x_1}\cdots v_{x_N} v_{y_1}\cdots v_{y_N}\right)^T$ as the given 2D motion field, and $\mathbf{A} = \begin{pmatrix} \mathbf{A}_Z & \mathbf{0}_{N\times N} & -\mathbf{A}_x \\ \mathbf{0}_{N\times N} & \mathbf{A}_Z & -\mathbf{A}_y \end{pmatrix}$ where $\mathbf{A}_Z = diag(\frac{1}{Z_1},\cdots,\frac{1}{Z_N})$, $\mathbf{A}_x = diag(\frac{x_1}{Z_1},\cdots,\frac{x_N}{Z_N})$, and $\mathbf{A}_y = diag(\frac{y_1}{Z_1},\cdots,\frac{y_N}{Z_N})$.

$TV_2(\mathbf{x})$ represents the sum of 2-dimensional total variation of V_x,V_y and V_z separately. And the group $l_{2,1}$-norm is defined as $\|\mathcal{G}(\mathbf{x})\|_{2,1} = \sum_{g=1}^{s}\|\mathbf{x}_g\|_2$ where $\mathbf{x}_g$ is the component belonging to group g, $g = 1,\cdots,s$. A pixel with disparity value d will be classified into group $g = \lfloor d/\frac{256}{s}\rfloor + 1$, assuming that the disparity ranges from 0 to 255.

In order to solve (5), we develop an alternating direction method of multipliers (ADMM) algorithm [14]. The energy minimization is first recast as the following constrained optimization problem

$$\begin{aligned} \min_{\mathbf{V},\mathbf{S},\mathbf{Z}} \quad & \tfrac{1}{2}\|\mathbf{AV}-\mathbf{b}\|_2^2 + \lambda\|\mathbf{S}\|_{2,1} + \mu\|\mathcal{G}(\mathbf{V})\|_{2,1} \\ \text{subject to} \quad & \mathbf{S} = \mathbf{DZ}, \\ & \mathbf{V} = \mathbf{Z} \end{aligned} \tag{7}$$

where $\mathbf{D} : \mathbb{R}^{n_1 n_2} \rightarrow \mathbb{R}^{2n_1 n_2}$ is the two dimensional finite difference operator that acts on V_x, V_y, and V_z separately, and $n_1 \times n_2$ is the video frame size.

We then reformulate (7) into an unconstrained problem by forming the following augmented Lagrangian

$$\begin{aligned} \max_{\mathbf{y}_1,\mathbf{y}_2}\min_{\mathbf{V},\mathbf{S},\mathbf{Z}} \quad & \tfrac{1}{2}\|\mathbf{AV}-\mathbf{b}\|_2^2 + \lambda\|\mathbf{S}\|_{2,1} + \mu\|\mathbf{V}\|_{2,1} \\ & + \tfrac{\rho_1}{2}\|\mathbf{S}-\mathbf{DZ}+\mathbf{y}_1/\rho_1\|_2^2 \\ & + \tfrac{\rho_2}{2}\|\mathbf{V}-\mathbf{Z}+\mathbf{y}_2/\rho_2\|_2^2 \end{aligned} \tag{8}$$

where we introduced the Lagrange dual variables $\mathbf{y}_1$ and $\mathbf{y}_2$. The algorithm proceeds by alternating between minimizing the objective in (8) with respect to each of the variables $\mathbf{V}, \mathbf{S}, \mathbf{Z}$ separately.

4. EXPERIMENTS AND DISCUSSIONS

4.1. Experimental setup

We apply the proposed framework to the dataset provided in the KITTI Vision Benchmark Suite [15]. The dataset was recorded from a moving vehicle while driving in and around a city, which includes color and gray scale images from left and right cameras, Velodyne laser scans and GPS measurements. To generate the input 2D motion vectors, the software package provided in [16] is applied to every two consecutive frames recorded by a color camera on the left, which leads to the estimated optical flows indicating the 2D motion at each pixel. Under the proposed framework, the estimated depth at each pixel is also required. We therefore apply the approach proposed in [17] to recover the depth map from a single camera image and the Velodyne data, both provided in the dataset.

With the 2D motion vectors and depth map prepared, we solve the optimization problem as described in Section 3 for the 3D motions corresponding to each pixel. It's worth to note that a series of parameters must be defined to solve the optimization problem: s(number of groups for the $l_{2,1}$-norm), λ(weight of total variation regularizer), μ_x, μ_y, μ_z(weights of sparsity regularizer on V_x,V_y,V_z respectively), $clusterMethod$ and $numCluster$. Three types of $clusterMethod$ are applied to the solved 3D motions: simple k-means clustering, spectral clustering with 4-connected graph and spectral clustering with fully connected graph.

4.2. Experimental results and discussions

Figure 3 and Figure 4 shows the experimental results on two sequences recorded under city environment, *drive_0059* and *drive_0018* with parameters as $s = 10, \lambda = 0.01, \mu_x = \mu_y = 5, \mu_z = 0.01, numCluster = 5$, and $clusterMethod$ as simple k-means clustering. The intermediate results of solved V_z show that by solving the optimization problem stated in Section 3, the scene can be clearly separated into regions with different moving directions and speeds, which is why clustering on the estimated V_x, V_y, V_z can provide promising results. Comparison between the segmentation results using our proposed method and using the method utilizing robust PCA and label propagation [18] shows that our algorithm provides more robust detection for the foreground moving objects.

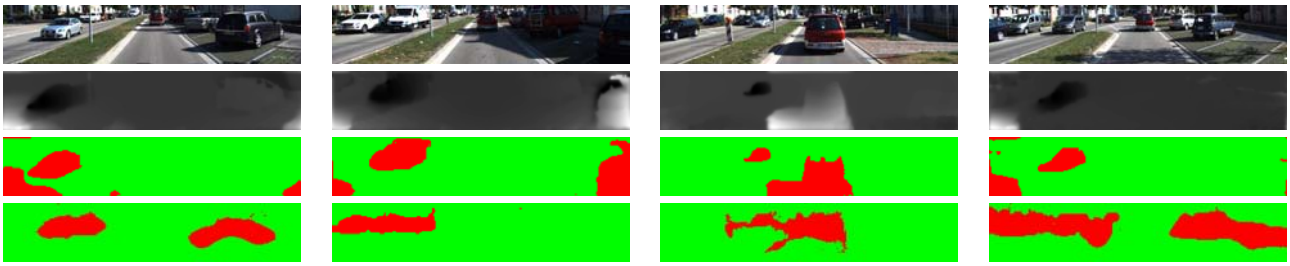

Fig. 3. Result on *drive_0059*. **From top to bottom:** Original images from left color camera, intermediate results of solved V_z, our segmentation results, and the benchmark provided with RPCA. **From left to right:** Frame #152, #174, #254 and #300.

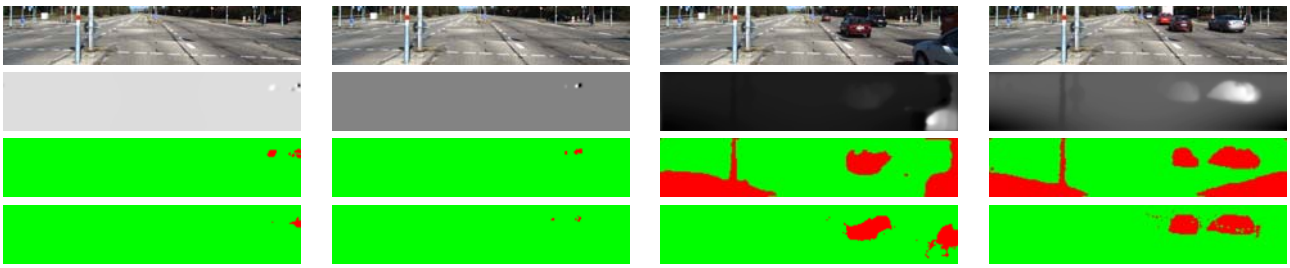

Fig. 4. Result on *drive_0018*. **From top to bottom:** Original images from left color camera, intermediate results of solved V_z, our segmentation results, and the benchmark provided with RPCA. **From left to right:** Frame #13, #20, #73 and #114.

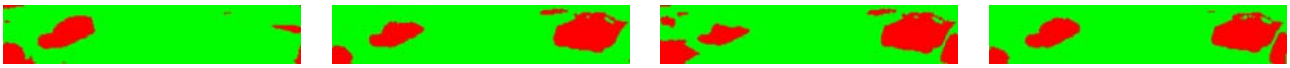

Fig. 5. Result on frame #300 of *drive_0059*. **From left to right:** Parameters set as in Fig. 3, only λ changed to 0.1, only μ_z changed to 5, and only s changed to 50.

Near to the picture boundaries (Figure 3) or for objects close to the camera (Figure 4), sometimes we could observe large motion away from the camera, which are likely to be segmented as a moving foreground. The reason behind is that the motion vectors toward the camera provided by the optical flow in those area are not large enough to reflect their actual motion; and hence those objects appear a motion away from the camera in the V_z domain. Providing an estimated optical flow with better quality will be able to resolve this type of mislabeling. Alternatively, the problem could be alleviated if the optical flow is allowed to be refined together with the segmentation procedure, which is subject to future improvements.

As for how to choose an appropriate set of parameters to use, an experiment to analyze the effect of tuning each parameter is performed and the result is shown in Figure 5. First can we observe that, when enlarging λ, which strengthen the regularization on total variation, the solved segmentation attempts to be more piece-wise smooth. On the other hand, when enlarging μ, the regularization on the group sparsity is strengthened and will lead to preference for smaller values of solved V_x, V_y, V_z. Thus, it is obvious that an adequate choice for μ_x, μ_y, μ_z should highly correspond to the understanding about the amplitude distribution between V_x, V_y, V_z.

5. CONCLUSION

In this paper, we first attempt solving for the 3D motions at each pixel in the scene from the given 2D motion field and estimated depth map. Specifically, we derive a motion field model to relate 3D motions to the 2D motion vectors taken depth into consideration. An optimization problem solving for the 3D motions is furthermore formulated, which aligns with the presented motion field model and the imposed piece-wise smoothness and group sparsity constraints. Once the 3D motions are solved, several clustering schemes are applied to them in order to provide the final segmentation results. Our experiments on car driving sequences in the well-known KITTI Dataset demonstrate the ability of this approach to correctly estimate the 3D motions and further achieve a more accurate motion segmentation result.

6. REFERENCES

[1] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, June 1981.

[2] A. Gruber and Y. Weiss, "Multibody factorization with uncertainty and missing data using the em algorithm," in *Computer Vision and Pattern Recognition, 2004*, June 2004, vol. 1, pp. 707–714.

[3] K. Kanatani, "Motion segmentation by subspace separation and model selection," in *Computer Vision, 2001. Proceedings. Eighth IEEE International Conference on*, 2001, vol. 2, pp. 586–591 vol.2.

[4] N. Ichimura, "Motion segmentation based on factorization method and discriminant criterion," in *Computer Vision, 1999.*

The Proceedings of the Seventh IEEE International Conference on, 1999, vol. 1, pp. 600–605 vol.1.

[5] R. Vidal and R. Hartley, "Motion segmentation with missing data using powerfactorization and gpca," in *Computer Vision and Pattern Recognition, 2004*, June 2004, vol. 2, pp. 310–316.

[6] J. Wright, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Advances in Neural Information Processing Systems 22*, 2009.

[7] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *J. ACM*, vol. 58, no. 3, pp. 11:1–11:37, June 2011.

[8] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS*, 2001, pp. 849–856.

[9] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2765–2781, 2013.

[10] Y. Sheikh, O. Javed, and T. Kanade, "Background subtraction for freely moving cameras," in *Computer Vision, 2009 IEEE 12th International Conference on*, Sept. 2009, pp. 1219–1225.

[11] A. Elqursh and A. M. Elgammal, "Online moving camera background subtraction.," in *ECCV (6)*, 2012, vol. 7577 of *Lecture Notes in Computer Science*, pp. 228–241.

[12] D. Tian, J.-Y. Kao, H. Mansour, and A. Vetro, "Graph spectral motion segmentation based on motion vanishing point analysis," in *Multimedia Signal Processing (MMSP), 2015 IEEE 17th International Workshop on*, Oct 2015, pp. 1–6.

[13] D. Tian, H. Mansour, and A. Vetro, "Depth-weighted group-wise principal component analysis for video foreground/background separation," in *Image Processing (ICIP), 2015 IEEE International Conference on*, Sept 2015, pp. 3230–3234.

[14] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[15] J. Fritsch, T. Kuehnl, and A. Geiger, "A new performance measure and evaluation benchmark for road detection algorithms," in *International Conference on Intelligent Transportation Systems (ITSC)*, 2013.

[16] C. Liu, *Beyond Pixels: Exploring New Representations and Applications for Motion Analysis*, Ph.D. thesis, Massachusetts Institute of Technology, May 2009.

[17] J. Castorena, U. S. Kamilov, and P. T. Boufounos, "Autocalibration of LIDAR and optical cameras via edge alignment," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2016, Shanghai, China, March 20-25, 2016*, 2016.

[18] J.-Y. Kao, D. Tian, H. Mansour, A. Vetro, and A. Ortega, "Geometric-guided label propagation for moving object detection," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016.