

SEPARATING BACKGROUND AND FOREGROUND OPTICAL FLOW FIELDS BY LOW-RANK AND SPARSE REGULARIZATION

Tomoya Sakai* Hiroki Kuhara*

* Graduate School of Engineering, Nagasaki University

ABSTRACT

We present a method for separating background and foreground optical flow fields induced by observer's egomotion and motion of objects, respectively. Optical flow is a vector field of instantaneous apparent motion computed from successive images. An optical flow field can be assumed as a linear combination with a few basis fields caused by translational and rotational egomotion and a spatially sparse optical flow field by the moving objects. We represent two-dimensional optical flow vectors as complex numbers and stack the fields as columns of a complex matrix. The low-rank component naturally corresponds to the egomotional background optical flow fields and the sparse component captures the moving foreground objects. We show that these components are successfully extracted from optical flow sequences by the robust PCA applied to the complex matrix.

Index Terms— Complex PCA, vector field decomposition, ADMM, visual navigation

1. INTRODUCTION

The aim of this paper is to separately obtain optical flow fields induced by different causes of motion in an image sequence. Optical flow [1, 2, 3, 4, 5] refers to apparent motion of objects in a scene caused by the relative motion between an observer and the scene. Estimation and analysis of optical flow from a sequence of observed images are beneficial for understanding the motion of the observer and objects that drive the optical flow, especially in the development of mobile robot and autonomous vehicle navigation as well as in the field of visual biology.

Egomotion can be estimated from the optical flow of rigid background objects in a scene [6, 7, 8, 9, 5]. In the estimation of egomotion, the optical flow caused by moving foreground objects is considered as outliers to remove from the egomotional optical flow. Conversely, one needs to identify the foreground optical flow for motion-based object detection and tracking [10, 11, 12]. Since an optical flow field is a superposition of the foreground and background fields, decomposition of the fields would be a comprehensive approach to the motion analysis.

We present a method for separating background and foreground optical flow fields. Our method exploits low-rank and sparse properties of a sequence of two-dimensional optical flow fields. That is, the background fields are composed of a few basis fields caused by translational and rotational egomotion, while the moving foreground objects induce spatially sparse optical flow fields. We treat a two-dimensional optical flow field as a complex vector in order to introduce the robust PCA technique [13, 14, 15, 16, 17, 18, 19, 20] for the separation of the low-rank and sparse fields. We experimentally show that the low-rank and sparse components correspond to the optical flow fields respectively induced by egomotion and motion of objects in a scene.

2. TWO-DIMENSIONAL OPTICAL FLOW

Optical flow is a vector field describing displacement of image intensity in the manner of continuum mechanics. In this paper, we focus on the optical flow of a time-varying image given as a sequence of two-dimensional images. we denote the optical flow as

$$\mathbf{d}(x, y, t) = \begin{bmatrix} d_x(x, y, t) \\ d_y(x, y, t) \end{bmatrix} = \begin{bmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{bmatrix}, \quad (1)$$

where $[x, y]^T$ denotes the position in a two-dimensional image domain, and t refers to time. Given a spatio-temporal image $I(x, y, t)$, we have an equation of continuity where the image intensity I is assumed as conserved quantity.

$$\frac{\partial I}{\partial t} + \nabla^T(I\mathbf{d}) = 0 \quad (2)$$

In computer vision, the optical flow is often assumed or approximated as incompressible. That is, the image intensity is considered to remain constant through time and displacement. Introducing this incompressible property to Eq. (2) leads to the so-called optical flow constraint (OFC).

$$\frac{\partial I}{\partial t} + \mathbf{d}^T \nabla I = 0 \quad (3)$$

The equation for optical flow, Eq. (3), is not enough to uniquely determine the vector \mathbf{d} . To resolve this underdetermined issue, computation of optical flow needs to impose

This work was supported by MEXT KAKEN 25330200.

some known structure or characteristics on the flow, such as smoothness over the image domain [1], piecewise constant [2], minimal total variation [21, 22] and so on.

Given a sequence of discrete images $I(x_m, y_m, t_n)$ ($m \in \{1, \dots, M\}$ and $n \in \{1, \dots, N + 1\}$), one can compute a pixel-wise dense optical flow field at each time t_n from the subsequent images at t_n and t_{n+1} . We do not go into the detail of the optical flow computation, because our interest is in what consist of optical flow fields, and in how to separate the fields into the components. See [3, 4, 5] for the review of optical flow computation methods.

3. EXTRACTING LOW-RANK AND SPARSE FIELDS FROM TWO-DIMENSIONAL OPTICAL FLOW SEQUENCE

3.1. Optical flow field as superposition

In many situations, an observed optical flow field can be represented as a linear combination of a small number of optical flow fields caused by essentially distinct grounds. Consider, for instance, an image sequence of a visual scene comprising of foreground objects moving in different directions. Optical flow of the static background is induced by camera egomotion and the motion of objects.

$$\mathbf{d} = \mathbf{d}^{\text{ego}} + \mathbf{d}^{\text{obj}} \quad (4)$$

Here, \mathbf{d}^{ego} and \mathbf{d}^{obj} are the optical flows by the egomotion and the object motion, respectively.

While the object motion can be arbitrary, the instantaneous egomotion is described by the translational velocity $\boldsymbol{\tau} \in \mathbb{R}^3$ [m/s] and rotational velocity $\boldsymbol{\omega} \in \mathbb{R}^3$ [rad/s] in the camera-centered coordinates. The egomotional optical flow \mathbf{d}^{ego} can be further modeled as a superposition of optical flows by the translation and rotation [6, 5]. For a pinhole camera, we have

$$\mathbf{d}^{\text{ego}} = \frac{1}{Z} \begin{bmatrix} -f & 0 & x \\ 0 & -f & y \end{bmatrix} \boldsymbol{\tau} + \frac{1}{f} \begin{bmatrix} xy & -(f^2 + x^2) & fy \\ f^2 + y^2 & -xy & -fx \end{bmatrix} \boldsymbol{\omega}. \quad (5)$$

Here, f is the focal length, and a point with the camera-centered coordinates $[X, Y, Z]^T$ is projected to $[x, y]^T = f/Z \cdot [X, Y]^T$ on the two-dimensional image plane.

An optical flow field in the entire image domain, due to camera egomotion in particular, is often so well-structured as to be linearly combined with a few basis fields such as \mathbf{d}^{ego} in Eq. (5). In contrast, optical flow by the object motion, \mathbf{d}^{obj} , can be assumed to be spatially sparse unless the objects occupy the image domain.

3.2. Complex-number representation

In order to find the linear combination of an optical flow field, it is useful to represent the field as an element of a

linear space. We propose to denote a two-dimensional optical flow vector $\mathbf{d} = [d_x, d_y]^T \in \mathbb{R}^2$, as a complex number $d = d_x + id_y \in \mathbb{C}$ where $i = \sqrt{-1}$ is the imaginary unit. Since a dense optical flow field at each time t_n computed from a sequence of discrete images is a collection of optical flow vectors at M individual pixels, it can be treated as a point in an M -dimensional complex space, i.e., $\mathbf{c}^{(n)} = [d(x_1, y_1, t_n), \dots, d(x_M, y_M, t_n)]^T \in \mathbb{C}^M$.

The complex-number representation of optical flow fields brings about some benefits when analyzing field components.

- The geometry of the two-dimensional vector space is reflected in the algebraic structure of the complex values. Operations on the optical flow vectors, namely, addition, scaling and rotation, correspond to addition and multiplication of complex values, respectively. The representation of scaling and rotation by complex values is more compact and faster to compute than the vector and matrix representations.
- Vector/matrix decomposition techniques are applicable to the complex-valued representation of optical flow fields. A sequence of optical flow fields is treated as a trajectory in the complex vector space \mathbb{C}^M . Egomotional optical flow fields depict a low-dimensional subspace of \mathbb{C}^M in which the complex vectors of basis optical flow fields reside. Sparse outlying components are related to moving objects.

Consequently, an optical flow field represented by a complex vector $\mathbf{c}^{(n)} \in \mathbb{C}^M$ can be linearly combined with a very small number of complex vectors of basis fields and a sparse field. A basis field is unique up to scaling and rotation, which prompts the subspace of basis fields to be low-dimensional.

3.3. Stable extraction of low-rank and sparse field components by robust PCA

Consider a complex matrix of optical flow sequence $\mathbf{C} = [\mathbf{c}^{(1)}, \dots, \mathbf{c}^{(N)}] \in \mathbb{C}^{M \times N}$. Its n -th column $\mathbf{c}^{(n)} \in \mathbb{C}^M$ is a complex vector whose entries are the optical flow vectors at time t_n in the complex-number representation. The matrix \mathbf{C} is assumed to be represented as

$$\mathbf{C} = \mathbf{L} + \mathbf{S} + \mathbf{E}. \quad (6)$$

Here, \mathbf{L} , \mathbf{S} , and \mathbf{E} are the complex matrices of optical flow sequences of background, foreground, and error, respectively. The matrix \mathbf{L} is supposed to be low-rank so that its principal components span the low-dimensional subspace of egomotional optical flow fields. The nonzero entries of \mathbf{S} indicate outlying optical flows possibly induced by moving foreground objects. The term \mathbf{E} carries errors caused by image noise and numerical computation of the optical flow fields.

In the error-free case, the decomposition of the complex matrix \mathbf{C} can be formulated as a convex problem

$$\min_{\mathbf{L}, \mathbf{S}} \|\mathbf{L}\|_* + \lambda_1 \|\mathbf{S}\|_1 \quad \text{subject to} \quad \mathbf{C} = \mathbf{L} + \mathbf{S}. \quad (7)$$

Here, $\|\cdot\|_*$ and $\|\cdot\|_1$ are the nuclear norm and the ℓ_1 norm, which promote the low-rankness and sparseness, respectively. $\lambda_1 > 0$ is an arbitrary balancing parameter. This is the same formulation as the so-called robust PCA [13, 14, 18]. One can easily confirm that efficient RPCA algorithms [15, 16, 17] designed for real-valued matrices can be applied as is to a complex-valued matrices. Note that the soft-thresholding (shrinkage) operation, used in the algorithms for solving an ℓ_1 -minimization subproblem, is defined for a complex value $z \in \mathbb{C}$ as

$$\text{soft}(z, \theta) = \exp(i \arg z) \max(|z| - \theta, 0). \quad (8)$$

Even in a non-error-free case, the low-rank and sparse components, \mathbf{L} and \mathbf{S} , can be extracted from the noisy matrix \mathbf{C} [23]. A convex problem for the decomposition can be written as

$$\begin{aligned} \min_{\mathbf{L}, \mathbf{S}, \mathbf{E}} \quad & \|\mathbf{L}\|_* + \lambda_1 \|\mathbf{S}\|_1 + \lambda_2 \|\mathbf{E}\|_F^2 \\ \text{subject to} \quad & \mathbf{C} = \mathbf{L} + \mathbf{S} + \mathbf{E} \end{aligned} \quad (9)$$

where λ_1 and λ_2 are balancing parameters. Again, current algorithms for this three-term decomposition, such as in [19, 20], can be adopted for complex matrices. Although the convergence of the alternating directions method of multipliers (ADMM) for three or more components is still open problem, we employ an ADMM-type algorithm [19, 20] that repeats

$$\{\mathbf{U}, \mathbf{K}, \mathbf{V}\} = \text{svd}(\mathbf{C} - \mathbf{S}_k - \mathbf{E}_k + \frac{1}{\beta_k} \mathbf{Y}_k), \quad (10)$$

$$\mathbf{L}_{k+1} = \mathbf{U} \text{soft}(\mathbf{K}, \frac{1}{\beta_k}) \mathbf{V}^H, \quad (11)$$

$$\mathbf{S}_{k+1} = \text{soft}(\mathbf{C} - \mathbf{L}_{k+1} - \mathbf{E}_k + \frac{1}{\beta_k} \mathbf{Y}_k, \frac{\lambda_1}{\beta_k}), \quad (12)$$

$$\mathbf{E}_{k+1} = \left(1 + \frac{2\lambda_2}{\beta_k}\right)^{-1} \cdot (\mathbf{C} - \mathbf{L}_{k+1} - \mathbf{S}_{k+1} + \frac{1}{\beta_k} \mathbf{Y}_k), \quad (13)$$

$$\mathbf{Y}_{k+1} = \mathbf{Y}_k + \beta_k (\mathbf{C} - \mathbf{L}_{k+1} - \mathbf{S}_{k+1} - \mathbf{E}_{k+1}) \quad (14)$$

where soft works element-wise on matrices, svd is the singular value decomposition, and $\{\beta_k\}$ is a monotonically increasing positive sequence. Note that \mathbf{V}^H indicates the conjugate transpose of \mathbf{V} .

4. EXPERIMENTS

4.1. Simulated data

We test the extraction of low-rank and sparse components from a sequence of simulated vector fields. We generate two basis fields as shown in Fig. 1(a), and denote them as $\mathbf{u}^{(T)}$ and $\mathbf{u}^{(R)} \in \mathbb{C}^{10^2}$. The sequence is synthesized as $\mathbf{c}^{(n)} = \mathbf{l}^{(n)} + \mathbf{s}^{(n)} + \mathbf{e}^{(n)}$ for $n = 1, \dots, 300$. Here, we define the

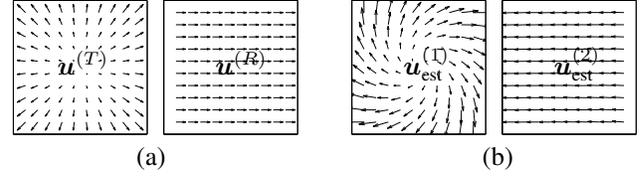


Fig. 1. Simulated basis optical flow fields $\mathbf{u}^{(T)}$ and $\mathbf{u}^{(R)}$, and the estimates $\mathbf{u}_{\text{est}}^{(T)}$ and $\mathbf{u}_{\text{est}}^{(R)}$.

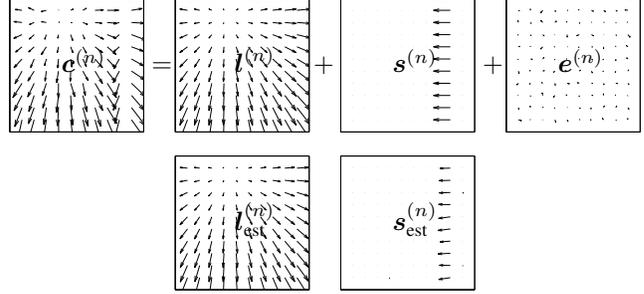


Fig. 2. A snapshot of simulated optical flow field $\mathbf{c}^{(n)}$, and the estimated low-rank and sparse components $\mathbf{l}_{\text{est}}^{(n)}$ and $\mathbf{s}_{\text{est}}^{(n)}$ ($n = 92$).

low-rank component as $\mathbf{l}^{(n)} = v_n^{(T)} \mathbf{u}^{(T)} + v_n^{(R)} \mathbf{u}^{(R)}$ with $v_n^{(T)} = 0.5(1 - 2n/300)$ and $v_n^{(R)} = 0.8 \exp(2\pi i n/10)$. The sparse component, $\mathbf{s}^{(n)}$, is a optical flow field of a set of vertically aligned pixels which horizontally takes a random walk. $\mathbf{e}^{(n)}$ is a Gaussian random noise with deviation 0.1. Figure 2 illustrates the simulated field components and extracted low-rank and sparse components $\mathbf{l}_{\text{est}}^{(n)}$ and $\mathbf{s}_{\text{est}}^{(n)}$ with $\lambda_1 = 3 \times 10^{-2}$ and $\lambda_2 = 5 \times 10^{-2}$. The angular error of $\mathbf{l}^{(n)}$ is 0.43° on average. The errors in flow endpoint are 8% and 3% for $\mathbf{l}^{(n)}$ and $\mathbf{s}^{(n)}$, respectively. The three-term decomposition successfully identifies the rank-two matrix \mathbf{L} , of which principal components, $\mathbf{u}_{\text{est}}^{(1)}$ and $\mathbf{u}_{\text{est}}^{(2)}$ as shown in Fig. 1(b), estimate $\mathbf{u}^{(T)}$ and $\mathbf{u}^{(R)}$ up to scaling and rotation, respectively.

4.2. Synthetic roadscape

We demonstrate our method on a synthetic driving sequence in SET 2 of EISATS [24]. The ground truth of the optical flow sequence with resolution $M = 640 \times 480 = 307,200$ is available for the sequence 2 in SET 2. We made the sequence $\mathbf{c}^{(n)}$ ($n = 1, \dots, N = 395$) by adding Gaussian noise with deviation 1.0 [pixel/frame] to each component of the two-dimensional optical flow vector of the ground-truth fields.

The three-term decomposition takes about a minute on a modern computer with i7 CPU with 8GB memory. Figure 3 shows examples of separated fields with $\lambda_1 = 5 \times 10^{-3}$ and $\lambda_2 = 1 \times 10^{-3}$. One can clearly see that the low-rank fields express the egomotional optical flow by the forward translational movement of the camera, while the motion of the foreground vehicles is separated in the sparse fields.

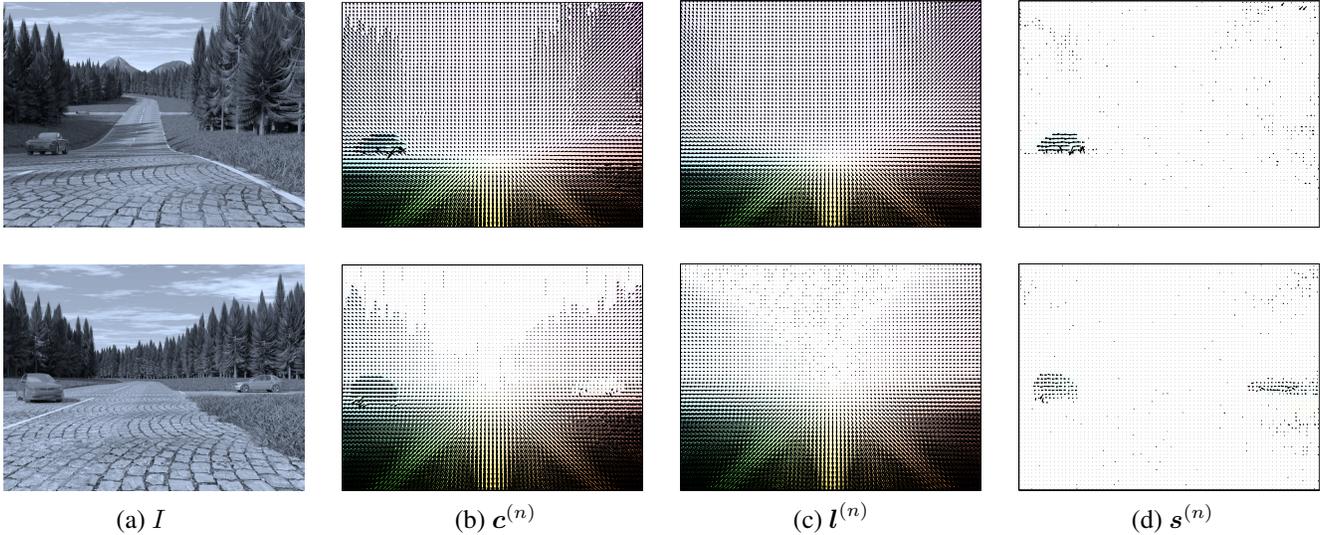


Fig. 3. Separation of optical flow fields of a driving sequence (first row: $n = 100$ and second row: $n = 217$). (a) Images from the sequence, (b) optical flow fields, (c) separated low-rank fields, and (d) sparse fields.

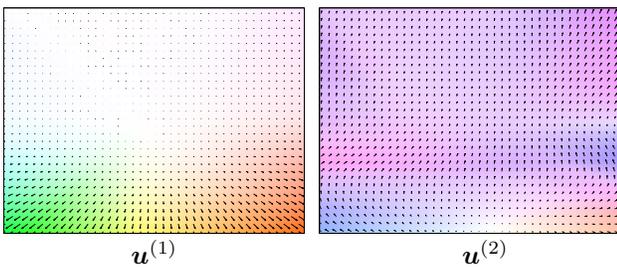


Fig. 4. Estimated 1st and 2nd principal components of the low-rank fields.

Since the sequence of the low-rank fields is expressed as $\mathbf{L} = \mathbf{U}\mathbf{K}\mathbf{V}^H$ by the singular value decomposition, the principal components stored in the columns of \mathbf{U} and \mathbf{V} describe the spatial and temporal variation of the low-rank fields. The first and second principal components, $\mathbf{u}^{(1)}$ and $\mathbf{u}^{(2)}$, well approximate the optical flow fields by the forward translation and rotation, respectively. The low-rank fields are synthesized mainly by these two principal components with varying amplitudes and phases indicated by the complex numbers in the first two columns $\mathbf{v}^{(1)}$ and $\mathbf{v}^{(2)}$ of \mathbf{V} as shown in Fig. 5. For example, $|v_n^{(2)}|$ and $\arg v_n^{(2)}$ indicate the scaling and rotation of the optical flow field $\mathbf{u}^{(2)}$ at time t_n . The second component $\mathbf{u}^{(2)}$ increases when the road slope is varying or the road is curving and the camera changes its direction. For example, the camera direction vertically changes because of the varying slope between $n \approx 40$ and 130 (decreasing slope) and between $n \approx 140$ and 200 (increasing slope). These changes respectively induce the upward and downward optical flow yielded by $\mathbf{u}^{(2)}$.

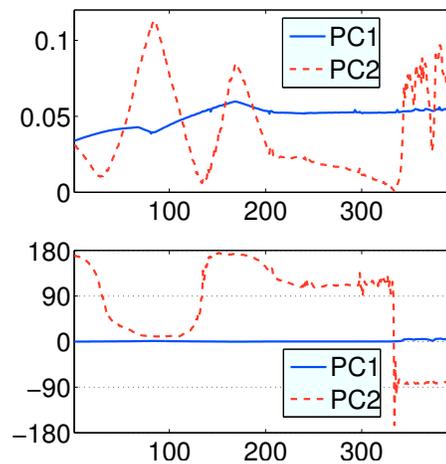


Fig. 5. Scaling (upper panel) and rotation (lower panel) of the 1st and 2nd principal components with respect to the time index n .

5. CONCLUDING REMARKS

We cast the separation of background and foreground optical flow fields induced by egomotion and object motion as a recovery of low-rank and sparse matrices from a noisy complex-valued matrix. Representing the two-dimensional optical flow vectors by complex numbers, existing RPCA algorithms are applicable to the optical flow fields.

Our method shows potential to extract the optical flow fields for egomotion estimation and detection/localization of moving objects in an unsupervised way without modeling the scenes and objects. Further research should include the performance evaluation of these tasks using the extracted fields.

6. REFERENCES

- [1] Berthold K.P. Horn and Brian G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185–203, 1981.
- [2] Bruce D. Lucas and Takeo Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings of Imaging Understanding Workshop*, 1981, pp. 121–130.
- [3] John L. Barron, David J. Fleet, and Steven Beauchemin, "Performance of optical flow techniques," *International Journal of Computer Vision*, vol. 12, no. 1, pp. 43–77, 1994.
- [4] Simon Baker, Daniel Scharstein, JP Lewis, Stefan Roth, Michael J Black, and Richard Szeliski, "A database and evaluation methodology for optical flow," *International Journal of Computer Vision*, vol. 92, no. 1, pp. 1–31, 2011.
- [5] F. Raudies, "Optic flow," *Scholarpedia*, vol. 8, no. 7, pp. 30724, 2013.
- [6] H Christopher Longuet-Higgins and Kvetoslav Prazdny, "The interpretation of a moving retinal image," *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 208, no. 1173, pp. 385–397, 1980.
- [7] David Nistér, "Preemptive RANSAC for live structure and motion estimation," *Machine Vision and Applications*, vol. 16, no. 5, pp. 321–329, 2005.
- [8] Richard JM Den Hollander and Alan Hanjalic, "A combined ransac-hough transform algorithm for fundamental matrix estimation.," in *British Machine Vision Conference*, 2007, pp. 1–10.
- [9] Oliver W Layton, Ennio Mingolla, and N Andrew Browning, "A motion pooling model of visually guided navigation explains human behavior in the presence of independently moving objects," *Journal of vision*, vol. 12, no. 1, pp. 20, 2012.
- [10] Wilfried Enkelmann, "Obstacle detection by evaluation of optical flow fields from image sequences," in *Computer Vision?ECCV 90*, pp. 134–138. Springer, 1990.
- [11] José Santos-Victor and Giulio Sandini, "Uncalibrated obstacle detection using normal flow," *Machine vision and applications*, vol. 9, no. 3, pp. 130–137, 1996.
- [12] N Andrew Browning, Stephen Grossberg, and Ennio Mingolla, "Cortical dynamics of navigation and steering in natural scenes: Motion-based object segmentation, heading, and obstacle avoidance," *Neural Networks*, vol. 22, no. 10, pp. 1383–1398, 2009.
- [13] Venkat Chandrasekaran, Sujay Sanghavi, and Pablo A. Parrilo, "Sparse and low-rank matrix decompositions," in *IFAC Symposium on System Identification*, 2009.
- [14] John Wright, Arvind Ganesh, Shankar Rao, and Yi Ma, "Exact recovery of corrupted low-rank matrices via convex optimization," in *Proceedings of Neural Information Processing System (NIPS)*, 2009.
- [15] Zhouchen Lin, Minming Chen, Leqin Wu, and Yi Ma, "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," in *UILU-ENG-09-2215, UIUC*, November 2009.
- [16] Xiaoming Yuan and Junfeng Yang, "Sparse and low-rank matrix decomposition via alternating direction methods," 2009.
- [17] Silvia Gandy and Isao Yamada, "Convex optimization techniques for the efficient recovery of a sparsely corrupted low-rank matrix," *Journal of Math-for-Industry*, vol. 2, no. 2010B-5, pp. 147–156, 2010.
- [18] Emmanuel J. Candès, Xiaodong Li, Yi Ma, and John Wright, "Robust principal component analysis?," *Journal of the ACM*, vol. 58, no. 3, pp. 11:1–11:37, 2011.
- [19] Min Tao and Xiaoming Yuan, "Recovering low-rank and sparse components of matrices from incomplete and noisy observations," *SIAM Journal on Optimization*, vol. 21, no. 1, pp. 57–81, 2011.
- [20] Omar Oreifej, Xin Li, and Mubarak Shah, "Simultaneous video stabilization and moving object detection in turbulence," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 2, pp. 450–462, 2013.
- [21] C. Zach, T. Pock, and H. Bischof, "A duality based approach for realtime TV-L1 optical flow," in *Proceedings of the 29th DAGM Conference on Pattern Recognition*, Berlin, Heidelberg, 2007, pp. 214–223, Springer-Verlag.
- [22] Andreas Wedel, Thomas Pock, Christopher Zach, Horst Bischof, and Daniel Cremers, "An improved algorithm for TV-L1 optical flow," in *Statistical and Geometrical Approaches to Visual Motion Analysis*, pp. 23–45. Springer-Verlag, Berlin, Heidelberg, 2009.
- [23] Z. Zhou, X. Li, J. Wright, E. Candès, and Y. Ma, "Stable principal component pursuit," in *IEEE ISIT Proceedings*, 2010, pp. 1518–1522.
- [24] Tobi Vaudrey, Clemens Rabe, Reinhard Klette, and James Milburn, "Differences between stereo and motion behavior on synthetic and real-world stereo sequences," in *23rd International Conference of Image and Vision Computing New Zealand (IVCNZ '08)*, 2008, pp. 1–6.