Learning General Optical Flow Subspaces for Egomotion Estimation and Detection of Motion Anomalies

Richard Roberts and Christian Potthast and Frank Dellaert
School of Interactive Computing, Georgia Institute of Technology
Atlanta, GA 30332
richard.roberts@gatech.edu, potthast@gatech.edu, dellaert@cc.gatech.edu

Abstract

This paper deals with estimation of dense optical flow and ego-motion in a generalized imaging system by exploiting probabilistic linear subspace constraints on the flow. We deal with the extended motion of the imaging system through an environment that we assume to have some degree of statistical regularity. For example, in autonomous ground vehicles the structure of the environment around the vehicle is far from arbitrary, and the depth at each pixel is often approximately constant. The subspace constraints hold not only for perspective cameras, but in fact for a very general class of imaging systems, including catadioptric and multiple-view systems. Using minimal assumptions about the imaging system, we learn a probabilistic subspace constraint that captures the statistical regularity of the scene geometry relative to an imaging system. We propose an extension to probabilistic PCA (Tipping and Bishop, 1999) as a way to robustly learn this subspace from recorded imagery, and demonstrate its use in conjunction with a sparse optical flow algorithm. To deal with the sparseness of the input flow, we use a generative model to estimate the subspace using only the observed flow measurements. Additionally, to identify and cope with image regions that violate subspace constraints, such as moving objects, objects that violate the depth regularity, or gross flow estimation errors, we employ a per-pixel Gaussian mixture outlier process. We demonstrate results of finding the optical flow subspaces and employing them to estimate dense flow and to recover camera motion for a variety of imaging systems in several different environments.

1. Introduction

This paper deals with estimation of ego-motion and dense optical flow from sparse flow measurements in a generalized imaging system, in the case where the imaging system moves through an environment with some degree of statistical regularity.

We are interested in approximate visual odometry as a replacement for more costly or limited methods of obtaining incremental platform motion. Laser scanners, or LIDAR, with which one can perform laser scan matching, tend to be heavy, expensive, and require much power. Wheel odometry, while very accurate indoors, is unreliable due to wheel slippage in outdoor environments. Inertial Measurement Units (IMU) accumulate velocity errors, and therefore must be fused with absolute position or velocity measurements.

Geometric stereo visual odometry, as described by Matthies [14], relies on finding the camera motion that best explains the motion of image features between cameras and frames in a stereo head. Most current systems employ fast hypothesis testing in a RANSAC harness, followed by pose optimization using inlying feature matches, for example in [17, 16, 1]. Nistér *et al.* report errors of about 12 m over a 360 m course [17]. Others have used a ground plane assumption to compute monocular visual odometry, for example [20, 4]. Geometric methods, while extremely accurate, often assume perspective cameras for which lens distortion can be modelled. Additionally, they are computationally demanding, making them challenging to implement on low-power systems.

Optical flow estimation in general suffers from the *aperture problem*. Because each neighborhood of pixels can have a different motion in the image, optical flow is typically computed independently on small windows throughout the image (e.g. [13]), but then cannot be evaluated in regions with ambiguous texture. Global smoothness constraints (e.g. [8]) propagate flow to these ambiguous regions, but make strong assumptions about the scene and the imaging system.

More recent work in computing optical flow has applied global constraints that arise from the optics of the imaging system and the structure of the scene. These usually assume a perspective camera, and model motion in the image as an affine or perspective transformation, for exam-

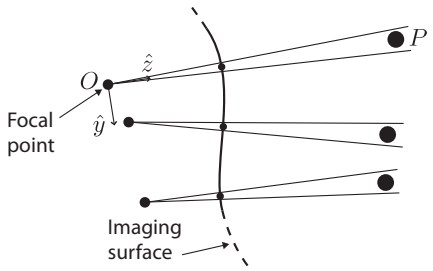


Figure 1: Schematic illustration of a generalized imaging system as a collection of local cameras. Each local camera images a narrow cone of rays on the imaging surface, contributing one pixel to the imaging system. Each local camera is defined by its focal length and its pose with respect to the imaging system.

ple [2, 10]. Some of these methods can estimate differing motion in separate regions of the image, for example [3, 11]. Irani exploits linear subspace constraints that hold over several frames [9]. Others use PCA to find the linear subspace automatically, e.g. [6], but still require optical flow to be computed over entire frames.

In contrast to the above work, we recover egomotion and dense optical flow directly from sparse optical flow using robust subspace constraints that hold over extended motion of a generalized imaging system, while simultaneously identifying outliers in the sparse flow. The subspace constraints hold when scene depth in the majority of the image is nearly constant over time, as is typically the case for planar mobile robots in outdoor and urban environments. In these environments, the ground plane, and even buildings and walls, are often at constant depth relative to the robot.

2. Generalized imaging system and optical flow subspace

In this section we describe a class of generalized imaging systems, and show that the optical flow therein lies near a linear subspace, constant for all incremental platform motion and time-invariant depth. Also, we show how to recover platform motion from a linear optical flow subspace.

2.1. Collection of local cameras

We are interested in generalized imaging systems with near-arbitrary optics, including multiple viewpoint systems, catadioptric systems, and projective cameras with distortion. Grossberg and Nayar [7] formalize generalized imaging systems in terms of "ray pixels", or *raxels*, and show that a caustic ray surface yields a smooth mapping from pixel locations (u, v) on a 2-dimensional imaging surface to imaged rays. Generalized imaging systems may also be modelled as a collection of local perspective cameras, each of which contributes one pixel to the system, as shown in Figure 1. Each local camera images a single ray, and is defined by its focal length and its pose with respect to the imaging system of the local camera.

Pless [18] and Neumann *et al.* [15] derive generalized optical flow and motion constraints for generalized imaging systems. A piece-wise smooth mapping from pixel locations to imaged rays allows optical flow to be well-defined, except at discontinuities in this mapping. Our method automatically ignores these discontinuities, such as occur at the boundaries of tiled images from multiple cameras.

2.2. Linearity of the subspace

We aim to recover a subspace for optical flow that is linear in the platform motion and can be assumed *constant over the entire camera trajectory.* Irani shows that for incremental motion over several frames, the optical flow in a projective camera lies in a linear subspace for arbitrary structure and camera motion [9]. This subspace approximation holds for a few frames and depends on the scene structure. In order to assume a constant subspace for extended motion, we exploit the regularity of scene depth with respect to the camera, which actually holds in many applications.

We now show that the optical flow at the pixel (u_j, v_j), corresponding to the j^{th} local camera, is linear in the incremental 3D platform motion $\delta \in \mathbb{R}^6$. In other words, there is an optical flow function $t_j(\delta, Z)$ such that

$$t_j(\delta, Z) \triangleq \begin{bmatrix} \dot{u}_j \\ \dot{v}_j \end{bmatrix} \approx \mathrm{T}_j \delta, \tag{1}$$

where Z is the scene depth and $\mathrm{T}_j \in \mathbb{R}^{2\times 6}$ is the matrix defining the linear subspace in which the flow t_j lives. To obtain T_j, consider the flow induced by the incremental translation $\tau_j^{\mathsf{T}} = \begin{bmatrix} \tau_{jx} & \tau_{jy} & \tau_{jz} \end{bmatrix}$ and rotation $\phi_j^{\mathsf{T}} = \begin{bmatrix} \phi_{jx} & \phi_{jy} & \phi_{jz} \end{bmatrix}$ of a *local* camera. Assuming incremental rotation [9, 12], we have

$$t_j(\delta, Z) \approx \frac{1}{Z_j} f_j \begin{bmatrix} \tau_{jx} \\ \tau_{jy} \end{bmatrix} + f_j \begin{bmatrix} \phi_{jy} \\ -\phi_{jx} \end{bmatrix}, \tag{2}$$

where f_j is the focal length of the local camera.

The key point is that, while Eq. 2 depends on scene depth, we can write it with a *typical flow matrix* associated with the *expected* inverse depth $\mu_j = \left\langle Z^{-1} \right\rangle$:

$$t_j(\delta, Z) \approx \begin{bmatrix} f\mu_j & 0 & 0 & 0 & f & 0 \\ 0 & f\mu_j & 0 & -f & 0 & 0 \end{bmatrix} \begin{bmatrix} \tau_j \\ \phi_j \end{bmatrix} \tag{3}$$

In turn, the incremental motion τ_j and ϕ_j of the local camera is a function of the incremental motion of the base frame, which can be approximated by a 6×6 linear mapping F_j:

$$\begin{bmatrix} \tau_j \\ \phi_j \end{bmatrix} = f_j(\delta) \approx F_j \delta \tag{4}$$

Hence, concatenating

$$t_j(\delta, Z) \triangleq \begin{bmatrix} \dot{u}_j \\ \dot{v}_j \end{bmatrix} \approx \begin{bmatrix} f\mu_j & 0 & 0 & 0 & f & 0 \\ 0 & f\mu_j & 0 & -f & 0 & 0 \end{bmatrix} F_j\delta = \mathrm{T}_j\delta \quad (5)$$

The optical flow in a generalized imaging system is thus in a linear subspace, with dimensionality equal to the degrees-of-freedom of the platform motion. This linear subspace is constant for constant depth and optics. Skew, scaling, focal point offset, and non-equal major and minor focal lengths in a local camera may all be approximated by a local affine transformation, and thus do not affect the linearity of the optical flow.

Although it may seem at first to be quite limiting, approximating inverse depth as a constant works well when there is considerable regularity of depth relative to the robot in the environment. First, the relative depth to the ground plane is always a constant for planar motion. Hallways and urban canyons also often meet this requirement. Because our method detects and ignores outliers, points that violate the regularity do not significantly affect results.

2.3. Recovering egomotion

Because the mapping from incremental motion to optical flow is linear, we can learn this mapping to recover egomotion from the optical flow. Suppose we have already found some q-dimensional linear subspace (q is near 4 in our experiments) in the optical flow $t_i \in \mathbb{R}^d$ (d typically on the order of 3000) for every i^{th} frame, then there is some $x_i \in \mathbb{R}^q$ such that,

$$t_i = \mathrm{W}x_i + \mu + \epsilon, \quad (6)$$

where $\mathrm{W} \in \mathbb{R}^{d\times q}$ and $\mu \in \mathbb{R}^d$ define the linear mapping from this subspace to optical flow, and ϵ is Gaussian noise.

There is also a linear mapping from platform motion to optical flow, shown in Eq. 5, so it follows that there is as well a linear mapping between the subspace coordinates x and the incremental platform motion $\delta \in \mathbb{R}^\eta$ (in the previous section we had $\eta = 6$, but this is not a requirement). To learn this mapping, we find the matrix $\mathrm{M} \in \mathbb{R}^{\eta\times q}$ and the vector $m \in \mathbb{R}^\eta$ that satisfy

$$\mathrm{M}x_i + m = \delta_i, \quad (7)$$

for every frame i. To find M and m, we rearrange the previous equation into η linear systems, each of the form $\mathrm{A}y = b$:

$$\begin{bmatrix} x_1^{\mathrm{T}} & 1 \\ \vdots & \\ x_m^{\mathrm{T}} & 1 \end{bmatrix} \begin{bmatrix} \mathrm{M}_k^{\mathrm{T}} \\ m_k \end{bmatrix} = \begin{bmatrix} \delta_{1k} \\ \vdots \\ \delta_{mk} \end{bmatrix}, \quad k \in \{1..\eta\}, \quad (8)$$

where M_k is the k^{th} row of M, m_k is the k^{th} element of m, and δ_{ik} is the k^{th} element of the known platform motion for frame i. Each of these linear systems deals with one of the dimensions k of δ. We solved these systems using iteratively-reweighted least squares, as there were frames in our data sequences with completely incorrect optical flow and ground truth.

After learning M and m from training data, we can estimate the egomotion for new frames by evaluating Eq. 7 with the subspace coordinates x_i for that frame. We obtain these subspace coordinates by iterating only the E-step, in Eqs. 15 and 16, of the EM algorithm we present next.

3. Finding the flow subspace

We present a robust extension of probabilistic PCA [19] to find the principal subspace and thus exploit global correlations in optical flow. The robustness of our principal subspace method is not limited to the traditional case where entire samples (i.e. frames) are discarded as outliers. Instead, we can find dimensions (i.e. individual flow vectors) in each sample whose values are not consistent with the other dimensions. De la Torre and Black [5] also developed an "intra-sample" robust PCA method, in which they minimize an energy function that weights each pixel with a per-image-pixel analog outlier process. We instead present a generative model for intra-sample robust PCA, which models each pixel as a Gaussian mixture model of either the expected value on the subspace, or zero-mean noise. The idea of exploiting correlations in optical flow using PCA has arisen in other contexts. Fleet *et al.*, for example, used PCA to learn a model of optical flow on images containing deforming bodies. They used this model to predict optical flow and classify movements in the subspace [6].

3.1. Training data

Given a training video sequence, we compute a downsampled sparse optical flow field over each pair of frames. We divide each image into a grid of 20×20 pixel cells, then track the strongest Harris corner in each cell using the Lucas-Kanade algorithm [13]. A threshold on corner response that prevents tracking textureless regions makes the flow fields sparse.

Each pair of frames yields an observation vector t_i, filled with the concatenated horizontal and vertical optical flow components from each local camera. For flow fields of size $w \times h$, for example, the length of each observation vector is $d = 2wh$. For each frame there is also a set S_i of the "seen" indices of t_i, where optical flow is available. Indices not in S_i are missing, and we save computation by ignoring them in calculations, as described below.

3.2. Generative model

In our generative model for optical flow, each vector is either an inlier, correlated to the other optical flow vectors in the frame, or an outlier of zero-mean Gaussian noise:

$$t_{ij} = \begin{cases} \mathrm{W}_j x_i + \mu_j + \epsilon_v & , z_{ij}^v \\ \epsilon_f & , z_{ij}^f \end{cases}, \quad (9)$$
$$\text{where } \epsilon_v \sim \mathcal{N}(0,\, \sigma_v) \text{ and } \epsilon_f \sim \mathcal{N}(0,\, \sigma_f)$$

where t_{ij} is the j^{th} optic flow component from the i^{th} frame. z_{ij}^v and z_{ij}^f are mutually exclusive binary variables indicating whether the optic flow component is an inlier or outlier (v for valid and f for false). ϵ_v and ϵ_f are zero-mean Gaussian random variables modelling the expected noise for inliers, which should be relatively small, and the expected noise for outliers, which should be relatively large.

We model inliers as arising from linear combinations of the basis flows, or principal axes, in the columns of W, with the latent variables, x. We also center the data by finding the robust mean, μ. As with PPCA, we estimate the variance σ_v^2 of the inlier noise, which measures how well the model explains the training data.

The key observations we make about this model are:

1. The model is similar to PPCA: if all flow vectors were measured and were inliers, the basis flows would span the principal subspace, as shown in [19].

2. When conditioned on the parameters and the latent variables, each observation vector is drawn from a mixture model of two *spherical* normal distributions.

3. Although the flow vectors of each frame are correlated through the basis flows and latent variables, all flow components are mutually *independent* when conditioned on the basis flows and latent variables.

3.3. Maximum likelihood formulation

As we observed above, when conditioned on W, μ, and x, t_{ij} for every (i, j), i.e. all optical flow vectors, are independent. This permits solving for the basis flows and latent variables while simply ignoring missing data, as we will show; estimating the missing data is not necessary. The conditional distribution over a single component of a single flow vector is the mixture model

$$t_{ij}|x_i z_{ij} \theta \sim \mathcal{N}\left(t_{ij};\, \bar{t}_{ij},\, \sigma_v^2\right)^{z_{ij}^v} \mathcal{N}\left(t_{ij};\, 0,\, \sigma_f^2\right)^{z_{ij}^f}, \quad (10)$$

where $\bar{t}_{ij} = \mathrm{W}_j x_i + \mu_j$, $\theta = \left\{\mathrm{W}, \mu, \sigma_v^2, \sigma_f^2, \pi_v, \pi_f\right\}$, and π_v and $\pi_f = 1 - \pi_v$ are the mixing coefficients of the mixture models. This is a mixture of two Gaussians: the first is the *inlier* distribution, whose mean $\bar{t}_{ij}$ is the expected optical flow from Eq. 9, and whose variance σ_v^2 converges to the expected squared-error of the PPCA model. The second is the *outlier* distribution, whose zero-mean and large variance model erroneous optical flow vectors. There is in fact a different inlier distribution for every flow vector, because each comes from a different row of the matrix W, but they are related through the latent variables x. As with a traditional Gaussian mixture model, $p\left(z_{ij}^v\right) = \pi_v$ and $p\left(z_{ij}^f\right) = \pi_f$.

As with PPCA, there is a Gaussian prior on x:

$$x \sim \mathcal{N}(\mathbf{0},\, \mathbf{I}_{q\times q}). \quad (11)$$

We are now ready to formulate the maximum likelihood problem of inferring the parameters and hidden variables when conditioned on the observations by maximizing

$$p(\theta x z|t) \propto p(\theta) \prod_i p(x_i) \prod_{j\in S_i} L\left(t_{ij}|x_i z_{ij} \theta\right) p\left(z_{ij}\right). \quad (12)$$

Note the product over $j \in S_i$, by which we only consider vectors in the training data that were "seen", and ignore those that were missing. Because of the conditional independence between flow vectors, we need not explicitly marginalize out these missing values, nor estimate their values, and this saves significant computation.

3.4. EM algorithm

The complete log-likelihood of (12), using the distributions above is

$$\begin{aligned}\mathcal{L}(\theta x z|t) = &\sum_i \log \mathcal{N}(x_i;\, 0,\, 1) + \sum_{j\in S_i} \\ &z_{ij}^v \left(\log \pi_v + \log \mathcal{N}\left(t_{ij};\, \bar{t}_{ij},\, \sigma_v^2\right)\right) \\ &+ z_{ij}^f \left(\log \pi_f + \log \mathcal{N}\left(t_{ij};\, 0,\, \sigma_f^2\right)\right). \end{aligned} \quad (13)$$

Finding the expectation of this log-likelihood with respect to x and z simultaneously would be intractable, but we can instead first find the expectation with respect to x, and then with respect to z, comprising a generalized EM algorithm. We obtain update equations by taking the derivatives of (13) w.r.t. the parameters, and then solving for each parameter:

$$\begin{aligned}\mu_j &= \left(\sum_{i\in S_j} \left\langle z_{ij}^v \right\rangle\right)^{-1} \sum_{i\in S_j} \left\langle z_{ij}^v \right\rangle \left(t_{ij} - \mathrm{W}_j x_i\right) \\ \sigma_v^2 &= \left(\sum_{i,\, j\in S_i} \left\langle z_{ij}^v \right\rangle\right)^{-1} \\ &\quad \sum_{i,\, j\in S_i} \left\langle z_{ij}^v \right\rangle \left(y_{ij}^2 - 2 y_{ij} \mathrm{W}_j \left\langle x_i \right\rangle + \mathrm{tr}\left(\left\langle x_i x_i^{\mathsf{T}} \right\rangle \mathrm{W}_j^{\mathsf{T}} \mathrm{W}_j\right)\right) \\ \mathrm{W}_j &= \sum_{i\in S_j} \left\langle z_{ij}^v \right\rangle \sigma_v^{-2} \left\langle x_i \right\rangle^{\mathsf{T}} y_{ij} \left(\sum_{i\in S_j} \left\langle z_{ij}^v \right\rangle \sigma_v^{-2} \left\langle x_i x_i^{\mathsf{T}} \right\rangle\right)^{-1}, \end{aligned} \quad (14)$$

where $y_{ij} = t_{ij} - \mu_j$. S_i is as above, while S_j is the set of *frame* indices whose j^{th} flow component was not missing. In our experiments, we compute σ_v^2 but constrain $\sigma_f^2 = 10$, though this outlier variance could easily be estimated online as well using a similar update equation.

We find the expectation $\langle x_i \rangle$ with respect to its posterior $p(x|tz\theta) \propto p(t|xz\theta)\, p(x)$ using Bayes' law and the distributions (10) and (11). Combining the conditional distributions over each t_{ij} into a multivariate distribution over t_i with diagonal covariance Λ_i^{-1} simplifies finding the distribution over x_i. Following, $p(x|tz\theta)$ has mean and covariance

$$\langle x_i \rangle = \left(\mathrm{W}_{Si}^{\mathsf{T}} \Lambda_i \mathrm{W}_{Si} + \mathbf{I}_{q\times q}\right)^{-1} \mathrm{W}_{Si}^{\mathsf{T}} \Lambda_i y_i, \qquad (15)$$

$$\left\langle x_i x_i^{\mathsf{T}} \right\rangle = \left(\mathrm{W}_{S_i}^{\mathsf{T}} \Lambda_i \mathrm{W}_{S_i} + \mathbf{I}_{q\times q}\right)^{-1} + \langle x_i \rangle \langle x_i \rangle^{\mathsf{T}},$$

$$\text{where } \Lambda_i = \sigma_v^{-2} \mathbf{I}_{q\times q} z_i.$$

The expectation $\left\langle z_{ij}^v \right\rangle$ is easier, as it is the simply the proportion of probability of the flow component belonging to the inlier distribution:

$$\left\langle z_{ij}^v \right\rangle = \frac{\mathcal{N}\left(t_{ij};\, \hat{t}_{ij},\, \sigma_v^2\right)}{\mathcal{N}\left(t_{ij};\, \hat{t}_{ij},\, \sigma_v^2\right) + \mathcal{N}\left(t_{ij};\, 0,\, \sigma_f^2\right)}, \qquad (16)$$

where $\hat{t}_{ij} = \mathrm{W}\left(\mathrm{W}^{\mathsf{T}}\mathrm{W}\right)^{-1}\left(\mathrm{W}^{\mathsf{T}}\mathrm{W} + \sigma_v^2 I_q\right)\langle x_i \rangle$.

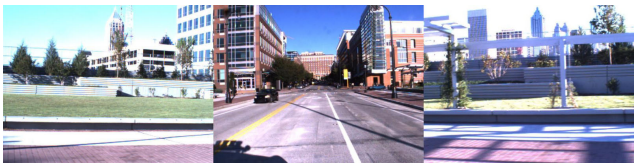
4. Experiments

In this section, we present experiments that illustrate the linear flow subspace for several imaging systems, the robustness of our method to outliers, the estimation of dense flow, and the recovery of platform motion.

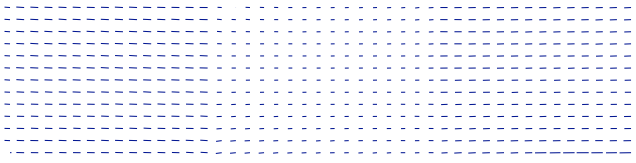
4.1. The flow subspace

Figure 2 shows the basis flows spanning the 2-dimensional linear optical flow subspace for the 3-camera outdoor driving sequence. The vehicle had 2 degrees-of-freedom, arising from its steering rate and speed. The first basis flow corresponds approximately to yawing, while the second corresponds approximately to driving forwards. In the second basis flow, the flow vectors at the bottom of the image are larger than those at the top because they correspond to points on the ground plane closer to the camera, thus moving more in the image when the vehicle translates. On the other hand, optical flow magnitude is invariant to depth in rotations, yielding the approximately equal-length vectors in the first basis flow.

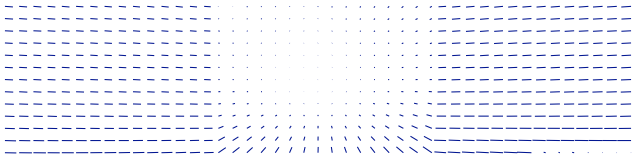
Figure 3 shows the basis flows for a catadioptric system in which part of the camera's view contains a curved mirror. The lower-left corner of the frame images the plate to which the camera and mirror are rigidly attached. The basis flows show the stark boundaries between these three image regions with differing optics.



(a) Typical frame (1920 × 480)



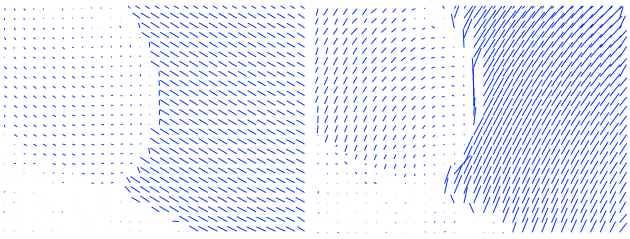
(b) Basis flow 1



(c) Basis flow 2

Figure 2: Basis flows spanning a two-dimensional linear flow subspace for the 3-camera outdoor driving sequence.

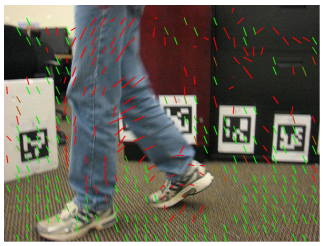
(a) Typical "ad-hoc catadioptric" frame



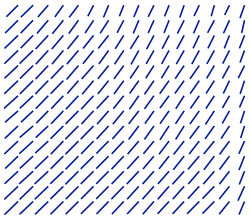
(b) Basis flow 1 (c) Basis flow 2

Figure 3: Typical frame (640 × 480) and basis flows for the ad-hoc catadioptric system. The basis flows show the irregular optics of the imaging system. Optical flow is coherent both in the mirror and in the view beyond the mirror, but is usually zero or erroneous in the lower-left part of the image where the camera sees only its mounting plate.

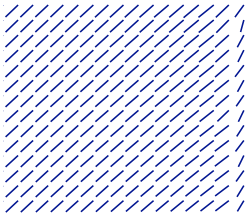
Figure 4 illustrates outliers in a sequence from a pan-tilt camera with a person walking in the view. 4a shows sparse flow measurements classified as inliers and outliers. 4b and



(a) Typical frame (640 × 480) showing outlying sparse flow vectors (in red) found by our method.



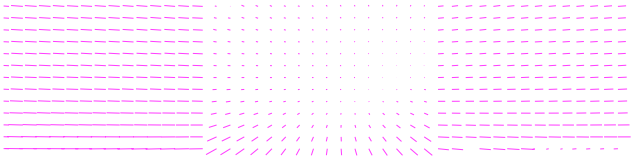
(b) One of the two basis flows found by standard PPCA.



(c) The same basis flow found by our method.

Figure 4: a) A pan-tilt camera, viewing a person walking while the camera moves. b) In a basis flow found by standard PPCA, the vectors are not parallel due to outliers (in red), as they should nearly be for this dataset. c) In the same basis flow found by our robust method, the vectors are properly much closer to parallel.

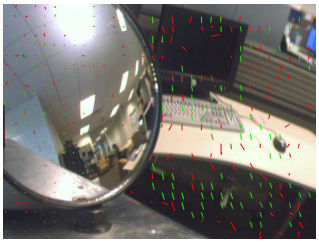
(a) Raw sparse optical flow.



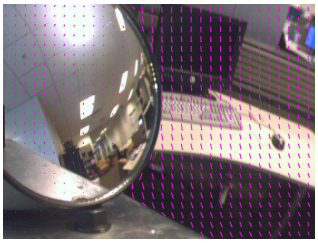
(b) Estimated dense optical flow.

Figure 5: Typical raw sparse optical flow field from the multi-camera outdoor driving dataset, and the corresponding dense flow estimated by our method.

4c show, respectively, the second bases found by non-robust PPCA, and by our method, which probabilistically ignores outliers when estimating bases. Because this is a narrow-FOV camera in pure rotation, the vectors in the basis flows should be nearly parallel.



(a) Raw sparse optical flow



(b) Estimated dense flow

Figure 6: Typical raw sparse optical flow field from the ad-hoc catadioptric system dataset, and the corresponding dense flow estimated by our method.

4.2. Flow inference

The multi-camera outdoor driving data set consists of tiled images, and optical flow fields from it are sparse and contain numerous errors, as shown in Figure 5a. Our method estimates the latent variables for frames of observed sparse flow vectors, while identifying outlying flow vectors, and then reconstructs dense flow, shown in Figure 5b, using the estimated latent variables and the basis flows. Figure 6 shows similar results for the "ad-hoc catadioptric system".

In frames containing moving objects, gross sparse flow estimation errors, or structure that is unusually close to or far from the camera, robustness to outliers is very important. Figure 7 shows the inliers and outliers that are present in a typical frame. Green vectors are inliers, while red vectors are outliers. In the left-facing camera, the vectors on the moving pickup truck are labelled as outliers. In the forward-facing camera, the erroneous vectors on the textureless portion of the road are mostly ignored. In the right-facing camera, the structure of the wall that is very close and very far from the camera is ignored, while the vectors on the structure near the average distance are labeled as inliers. The colors, ranging from green to red, indicate the probability of a flow vector being an inlier.

4.3. Ego-motion estimation

We conducted an experiment using a differential-drive robot with two forward facing cameras, which instead of forming a stereo pair, face outwards at angles of approximately 20°. The robot was equipped with a well-tuned pose filter that fused wheel odometry with an inertial measurement unit (IMU). Figure 8 shows the trajectories estimated by this filters, as well as by our method, as described in Section 2.3. We trained our method using pose estimates from the pose filter on the short training sequence shown in the figure, and then switched to estimating pose with our method with no further training. The pose estimated from

Figure 7: Detection of outliers in the sparse optical flow in a frame from the 3-camera outdoor driving sequence. Colored lines are sparse optical flow, ranging from green when $p\,(\text{inlier}) = 1$, to red when $p\,(\text{inlier}) = 0$. Sparse flow vectors that are inconsistent with the linear flow subspace have low inlier probability. Vectors on the moving pickup truck, in the textureless regions of the road, and on the very close and very far structure of the wall are labelled as outliers.

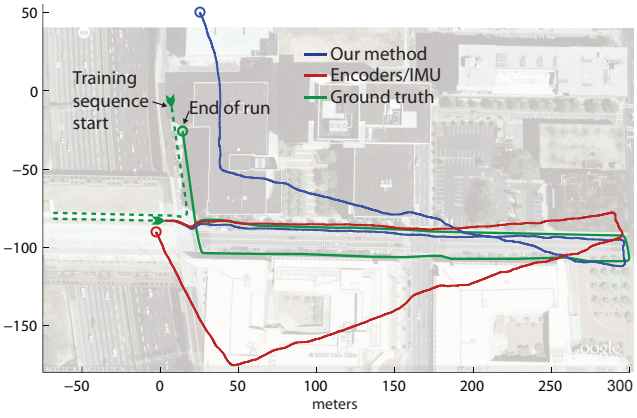


Figure 8: Platform trajectories estimated by our method (blue) and by integrating wheel odometry with an inertial measurement unit (red). We trained our method from ground truth over a 200 m training segment (dashed green), then switched to estimating pose for approximately 675 m with no further training. Each trajectory ends at the circle of the corresponding color.

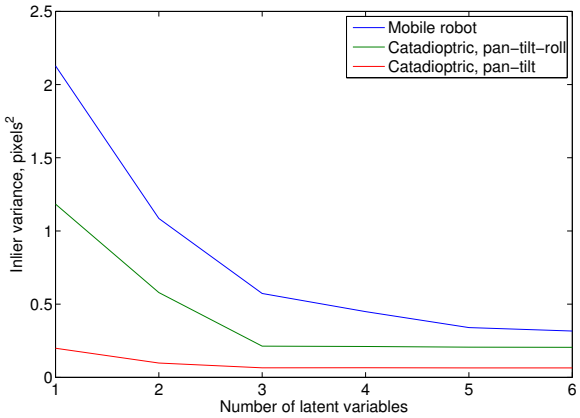


Figure 9: PPCA inlier distribution variance for models with various numbers of latent variables. The variance measures the prediction error of the robust PPCA model. The pan-tilt and pan-tilt-roll catadioptric systems are 2 and 3 DOF, respectively. The mobile robot has 2 controllable DOF's, but additionally pitches and rolls due to sloped ground.

optical flow is comparable in accuracy to the filtered pose from wheel odometry and IMU.

An informal timing evaluation revealed that our prototype code, with sparse flow extraction in C++ and dense flow and ego-motion estimation in MATLAB, runs faster than 30 Hz (or 33.3 ms per frame) after training. On 1920×480 frames, with 45×13 flow fields, computing sparse flow takes 19.1 ms, subspace coordinates 6.8 ms, and ego-motion 0.2 ms, for a total of 26.1 ms per frame.

5. Discussion

The inlier distribution variance σ_v^2 is a measure of how well the robust PPCA model fits the training data. Specifically, it is the mean squared error between the predicted and observed optical flow in the training data, weighted by the inlier probability for each flow component. The number of degrees-of-freedom (DOF) of the motion of an imaging system is reflected in the relationship between the number of latent variables (i.e. principal components) in a robust PPCA model and the inlier variance. Figure 9 shows this relationship for three datasets with different numbers of motion DOF's, where the variance curve becomes almost flat as or shortly after the number of latent variables exceeds the DOF's. In the case of the mobile robot, only the velocity and steering angle were controllable, but ground slopes caused pitching and rolling. The additional decrease in variance in models with more than 4 latent variables in this case

suggests to us that there are additional unaccounted but predictable aspects of the optical flow or the depth of the environment that additional latent variables can explain.

The dense flow that we compute is not, in general, the true optical flow present everywhere in the image. Instead, it is an estimate of the flow that would occur in a static scene if the inverse depth at each pixel were equal to a robustly-computed *average* inverse depth for that pixel. Differences between the true optical flow and the predicted optical flow are mainly due to moving objects and objects that violate the constant depth assumption. Our method ignores most of these differences, but would fail when the majority of the scene coherently violates the depth assumption, for example if the robot became surrounded with nearby objects. In future work, instead of attempting to ignore depth variation and motion, a collection of learned subspaces could describe the depth in multiple typical environments, and yield an online estimate of the environment type from optical flow. Fitting different subspaces to different regions of images could lend itself towards understanding scenes with multiple coherent motions or depths.

References

[1] M. Agrawal and K. Konolige. Real-time localization in outdoor environments using stereo vision and inexpensive gps. In *IEEE International Conference on Pattern Recognition (ICPR'06)*, 2006.

[2] J. Bergen, P. Anandan, K. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *European Conference on Computer Vision*, volume 588, pages 237–252. Springer, 1992.

[3] M. J. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding*, 63(1):75–104, January 1996.

[4] J. Campbell, R. Sukthankar, I. Nourbakhsh, and A. Pahwa. A robust visual odometry and precipice detection system using consumer-grade single vision. In *IEEE International Conference on Robotics and Automation, Proceedings*, 2005.

[5] F. De la Torre and M. Black. Robust principal component analysis for computer vision. *International Conference on Computer Vision (ICCV'01)*, 1:362–369, 2001.

[6] D. Fleet, M. Black, Y. Yacoob, and A. Jepson. Design and Use of Linear Models for Image Motion Analysis. *International Journal of Computer Vision*, 36(3):171–193, 2000.

[7] M. D. Grossberg and S. K. Nayar. The raxel imaging model and ray-based calibration. *International Journal of Computer Vision*, 61(2):119–137, 2005.

[8] B. Horn and B. Schunck. Determining Optical Flow. *Artificial Intelligence*, 17(1-3):185–203, 1981.

[9] M. Irani. Multi-frame correspondence estimation using subspace constraints. *International Journal of Computer Vision*, 48(3):173–194, 2002.

[10] M. Irani and P. Anandan. Direct recovery of planar-parallax from multiple frames. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(11):1528–1534, November 2002.

[11] M. Irani, B. Rousso, and S. Peleg. Detecting and tracking multiple moving objects using temporal integration. In *European Conference on Computer Vision*, pages 282–287. Springer, 1992.

[12] H. Longuet-Higgins and K. Prazdny. The Interpretation of a Moving Retinal Image. *Proceedings of the Royal Society of London. Series B, Biological Sciences (1934-1990)*, 208(1173):385–397, 1980.

[13] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *International Joint Conference on Artificial Intelligence*, volume 3, 1981.

[14] L. Matthies. *Dynamic Stereo Vision*. PhD thesis, Carnegie Mellon University, 1989.

[15] J. Neumann, C. Fermuller, and Y. Aloimonos. Polydioptric camera design and 3d motion estimation. In *Computer Vision and Pattern Recognition, Proceedings of the IEEE Computer Society Conference on*, volume 2, 2003.

[16] K. Ni and F. Dellaert. Stereo tracking and three-point/one-point algorithms – a robust approach in visual odometry. In *International Conference on Image Processing*, 2006.

[17] D. Nister, O. Naroditsky, and J. Bergen. Visual odometry. In *IEEE Conference on Computer Vision and Patterns Recognition*, volume 1, pages 652–659, 2004.

[18] R. Pless. Discrete and differential two-view constraints for general imaging systems. In *Omnidirectional Vision, 2002. Proceedings. Third Workshop on*, pages 53–59, 2002.

[19] M. Tipping and C. Bishop. Probabilistic Principal Component Analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):611–622, 1999.

[20] H. Wang, K. Yuan, W. Zou, and Q. Zhou. Visual odometry based on locally planar ground assumption. In *Information Acquisition, 2005 IEEE International Conference on*, page 6, 2005.