

# FUNCTIONAL VANISHING POINT ESTIMATION VIA A FILTERED-RADON OPERATOR

William Mantzel, Justin Romberg

Georgia Institute of Technology  
 {willem, jrom}@ece.gatech.edu

## ABSTRACT

When available, vanishing points in a scene are a key factor in effectively recovering absolute camera orientation, thus simplifying the structure-from-motion problem. We present a novel method for estimating vanishing points without explicitly detecting line features. This approach first maps images into line-space with a filtered-Radon operator, allowing subtle line textures to contribute, and improving the angular resolution of broken or occluded segments of the same line. Then, we use a robust coarse-to-fine method to jointly estimate the three vanishing points. We evaluate our method on video sequences, demonstrating robustness to clutter lines as well as the ability to effectively utilize subtle edge-texture information.

**Index Terms**— vanishing points, Radon transform, absolute orientation

## 1. INTRODUCTION

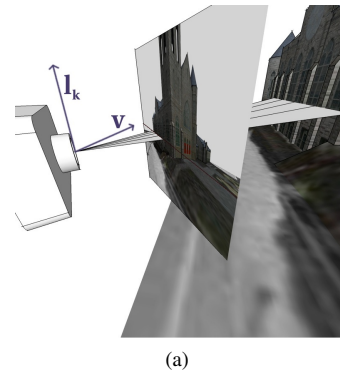
Though there have been significant advances in the area of structure-from-motion, especially with the advent of recent feature-descriptors for reliable matching, the estimation of vanishing points (as depicted in Figure 1) remains a key step in recovering absolute orientation for a camera path. Absolute orientation estimates may help to merge independent data sets, or may help correct the accumulation of small errors in the relative orientation.

Most vanishing point estimation techniques fall into one (or in some cases both) of two camps. The first camp uses some form of Hough transform, first partitioning the set of all directions in  $\mathbb{R}^3$  into bins (e.g. by sampling the faces of a cube or the  $z = 1$  plane), and then incrementing each bin count that contains a point on a line, for all candidate lines [1, 2, 3]. Then the set of bins with largest count values are good candidates for vanishing points. As is pointed out by [3], the computational complexity is linear in the number of lines instead of the quadratic complexity characteristic of earlier approaches that considered all intersections of lines. However, when taking into account that the number of bins generally increases as the number of lines increases, it is apparent that the computational complexity is likely greater than claimed.

The second camp of techniques uses a set of lines  $l_k \in L$  that are already known to be closely related to a candidate vanishing point (generated from the first camp for example). These lines may be used to estimate the vanishing point by maximizing a function of the form:

$$\operatorname{argmax}_{\|v\|=1} \sum_{l_k \in L} f(l_k)g(v^T l_k), \quad (1)$$

where  $f$  reflects the strength of the line, possibly after passing through a thresholding function of some sort, resulting in a binary decision. The function  $g$  has maximum at 0 so that the optimization favors a vanishing point that is orthogonal to the normal vectors  $l_k$  of the strongest lines (i.e. a vanishing point that lies on the strongest



**Fig. 1.** A line with unknown depth is known to lie on a plane (with normal  $l_k$ ) that contains the vanishing point  $v = [-1 \ 0 \ 1]^T$  in homogeneous coordinates. A collection of such  $l_k$  then lie in  $v^\perp$  allowing  $v$  to be estimated with two or more corresponding lines. After at least two vanishing points are estimated, the orientation  $R$  can be recovered.

lines). A common choice  $g(c) = -c^2$  reflects a form of Gaussian maximum likelihood estimation and is easily solved using a least-squares eigenvalue technique [4]. This set  $L$  is generally determined iteratively in an Expectation-Maximization framework that alternates between computing the vanishing point  $v$  and classifying relevant lines by thresholding their inner product distance to this vanishing point. This approach generally works well except that  $\ell_2$  minimization approaches are notoriously sensitive to outliers, which in this case are the lines most likely to enter and leave the set of relevant lines  $L$ , resulting in a potentially unstable vanishing point recovery.

Our approach essentially belongs to this latter camp but is different to these prior approaches in at least 3 key aspects. First, rather than explicitly detecting a set of lines we instead utilize a *filtered-Radon operator* that maps line gradients in the image domain to small singularities in an otherwise sparse Radon domain. This method is remarkably better at handling subtle edge information in edge-dense environments, and is roughly the same in environments where there are a few dominant lines. Secondly, while other approaches iteratively classify or assign the lines to the relevant vanishing points before recomputing these vanishing points, our approach effectively defers this assignment decision, creating a robustness to outliers. Lastly, because of the difficulty combining the orthogonality constraint with a joint least-squares computation over all three vanishing points, these other approaches first compute these vanishing points independently of each other, and then enforce orthogonality *a-posteriori*. Our approach in contrast searches directly over the 3-D set of rotations.

## 2. FILTERED-RADON OPERATOR

Rather than explicitly detect line features, we defer detection and classification as much as possible by utilizing a functional description of line space.

Assuming that our camera has been calibrated (i.e. the intrinsic parameters estimated), let  $I$  be a rectified image so that  $I(x, y)$  is the intensity observed in direction  $[x \ y \ 1]^T$ . (For standard field-of-view cameras, the domain of this image is some subset of the region  $\{[-1 \ 1] \times [-1 \ 1]\}$ .) The filtered-Radon operator is given as:

$$Q(\theta, m) = \iint I(x, y)h(m - m') dx dy : \quad (2)$$

$$m' = x \cos(\theta) - y \sin(\theta). \quad (3)$$

If  $h(m) = \delta(m)$ , this is the standard Radon transform. By using other choices for  $h$ , this operator effectively “filters” in the direction orthogonal to  $\theta$  before taking the usual Radon integral in the  $\theta$  direction. By choosing  $h$  as a derivative operator, this operation is similar to taking the Hough transform of the gradient of the image, except that this approach has the advantage of preserving the direction and sign of the gradient (instead of only considering the magnitude).

To illustrate this concept for a derivative operator  $h$ , consider an image with a vertical edge through  $x = 0$  with small width  $2\epsilon$ . Now we have  $|Q(\theta, m)| \simeq \min(|\cot(\theta)|, \frac{1}{\epsilon})$  for  $|m| < |\sin(\theta)|$  and zero otherwise. Although  $Q$  decays away from  $\theta = 0$  like  $\theta^{-1}$ , we would ideally like to choose  $h$  so that  $Q$  vanishes everywhere but the edge location (i.e.  $\theta$  and  $m$  close to zero). This is achieved by choosing  $h$  as a second derivative operator. In the above example now,  $Q$  is much sparser and vanishes except when  $|\theta| < \epsilon$  and  $|m| < \epsilon$ . Because of this localized response in the filtered-Radon domain and the linearity of the construction, line features interfere very little with each other and can be easily distinguished.

$Q(\theta, m)$  then describes the intensity of the line satisfying  $x \cos(\theta) - y \sin(\theta) = m$  on the  $z = 1$  image plane. Because each of these points is a perspective projection of its 3D point, the corresponding 3D line lies somewhere in the plane spanned by these points, as shown in Figure 1. Here, it is easily seen that  $l(\theta, m) \triangleq [-\cos(\theta) \ \sin(\theta) \ m]^T$  is orthogonal to these points (in practice we normalize  $\|l\| = 1$ ). Because there is a one-to-one correspondence between these lines and their planes, it is equivalent to describe each line by its plane’s normal. We then describe each line sample in the filtered-Radon domain  $(\theta, m)$  by its associated normal vector  $l$ , so that finding vanishing points lying in the intersection of these line-induced planes is equivalent to finding a vanishing point (vector) orthogonal to a set of these normal vectors.

This filtered-Radon operation can be performed for each  $\theta$  simply by filtering in the direction  $\theta + \pi/2$  and then integrating along direction  $\theta$ . Because this operation is essentially a shift-invariant filter, it could in principle be performed in the Fourier domain by drawing weighted samples along oriented lines, utilizing the Fast Fourier Transform to speed up computation. However, for sampled discrete images it is not necessarily obvious how to draw these samples. One such method for computing the discrete Radon transform called the Fast Slant Stack (FSS) manages to achieve (among other properties) algebraic exactness with its continuous counterpart [5]. That is, the result of this operation is equivalent to constructing a continuous image via sinc interpolation and subsequently integrating along specified lines. Or equivalently, this operation shears the image using Fourier resampling and sums the columns of the resulting sheared image (hence slant stack). Moreover, the method utilizes the Fractional Fourier Transform to achieve a fast transform, computing

$2N \times 2N$  output samples from  $N \times N$  samples in  $O(N^2 \log(N))$  time [6]. This is within a constant factor of the FFT (roughly 10 times as many flops as the  $2N \times 2N$  FFT). (For reference, most other discrete Radon transforms require  $O(N^3)$  computations for comparable output resolution). As a small price to pay, the directions are quantized uniformly in slope instead of angle, similarly to the Cascaded Hough Transform [7].

This Fast Slant Stack partitions directions into mostly vertical and mostly horizontal lines (the sets  $[-\pi/4 \ \pi/4)$  and  $[\pi/4 \ 3\pi/4)$ ). Line integrals of approximately vertical lines are computed via an efficient sinc interpolation for each column in the 2-D frequency domain, in order to utilize a discrete version of the projection-slice theorem. The corresponding filtering operation we employ amounts to multiplication in the frequency domain with a function that only varies horizontally and is constant in the vertical direction (and hence commutes with the interpolation). Therefore, the filtering and Radon operations commute with each other. Consequently, because the FSS is invertible, this filtered-Radon operator is invertible if and only if the filter  $h$  is invertible. In our proposed example of a second order difference operator, this operator is invertible when also given some additional information, for the same reason that a function is recoverable from its second derivative when also given the information of the function’s value and its derivative’s value at any point.

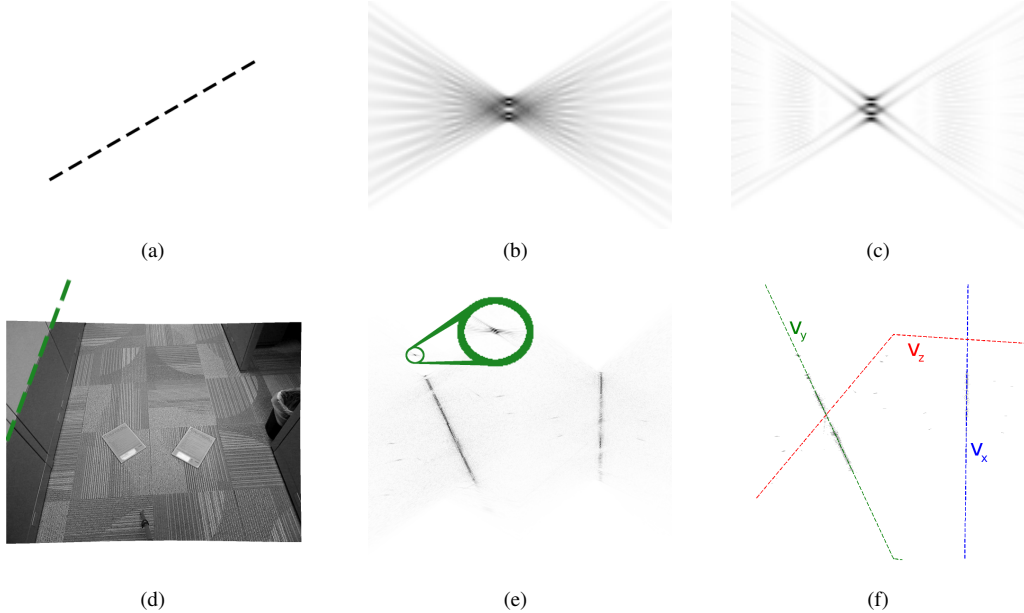
Although this operator is “endpoint-agnostic”, a corresponding advantage is the ability to “link” line segments together to achieve better angular resolution. Each segment of a dashed line (for example) may be too short to precisely determine the angle, but the characteristics of the whole dashed line may be determined to very high angular resolution. In practice, such “dashed lines” may arise from lines that are partially occluded in several places, aligned windows, or similar coincident architectural constructs.

This concept is illustrated in Figure 2. Each line segment of the dashed line has limited resolution of about 200 milliradians as shown in (b), but the aggregate response of the whole line shown in (c) is much more compact. The filtered-Radon response of the the rectified image in (d) is shown in (e). In contrast, line detection approaches that take the Hough transform of the magnitude of the gradient (ignoring the direction) tend to give spurious lines in this textured region.

Note that although the uniformly spaced Radon transform maps a point to a sinusoid, the FSS maps each point to a broken line with a cusp at  $\theta = \pi/4$ . The 2 degrees of freedom of this line correspond to the degrees of freedom in each vanishing point. Over the 3 usual vanishing points, there are 6 parameters with 3 orthogonality constraints, yielding the 3 degrees of freedom of rotation.

Armed with a functional description of line feature intensity  $Q(\theta, m)$ , the goal is now to find such broken-line paths through the filtered-Radon domain such that line integrals along this path will account for maximum line-mass, while preserving the orthogonality constraint. This will enable us to recover the camera’s orientation  $\mathbf{R}$ . Throughout the rest of the paper, we will use the Rodrigues parameterization of the rotation matrix:

$$\mathbf{R}(\omega) = \exp \left( \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix} \right). \quad (4)$$



**Fig. 2.** The filtered-Radon operator of each segment of the dashed line shown in (a) is given as  $Q_k(\theta, m)$ . Each such segment has an angular resolution corresponding to its aspect ratio  $1/5 = 200$  milliradians. For this reason, the sum of individual responses  $\sum_k |Q_k(\theta, m)|$  shown in (b) gives a broader response in angle (horizontal axis) than the response of the whole dashed line  $|\sum_k Q_k(\theta, m)|$  shown in (c) that is closer to the resolution of the dashed line, 13 milliradians. The filtered-Radon response of the the rectified image in (d) is shown in (e). Although the dominant line marked by the dashed line is easily detected, there are also hundreds of other contributing lines from the carpet texture. Estimating the vanishing points shown in (f) involves finding the best path through this domain to account for this texture-line-mass. For example,  $v_y$  is the “forward” vanishing point.

### 3. ORIENTATION RECOVERY

Most commonly, the vanishing points of the most dominant directions correspond with the axes of the world coordinate system (East, North, Up). In this case, the rotation matrix  $\mathbf{R} = [v_x \ v_y \ v_z]$ , so estimating the orientation can be formulated as a joint search for all 3 vanishing points, using a variation of eq. (1):

$$\mathbf{R} = \operatorname{argmax}_{\mathbf{R} \in SO(3)} \sum_{\theta, m} f_\epsilon(Q(\theta, m)) \|g_\sigma(\mathbf{R}^T l(\theta, m))\|_\infty, \quad (5)$$

where  $g_\sigma$  is a Gaussian function with parameter  $\sigma$  that acts pointwise on the elements of the vector, and  $f_\epsilon(x) = \min(|x| - \epsilon, 0)$ , a soft thresholding function (e.g.  $\epsilon = .004$ ).

The  $\infty$ -norm (i.e. the maximum element by magnitude) implicitly assigns each line sample to its most likely vanishing point at that particular orientation  $\mathbf{R}$  (rather than using an *a-priori* assignment) enabling a joint search for all vanishing points simultaneously.

It is also possible to extend this model to cases where there are multiple known (but not necessarily orthogonal) vanishing points in the world coordinate system  $\mathbf{V} = [v_1 \ v_2 \ \dots \ v_k]$  by simply making the substitution  $\mathbf{R} \leftarrow \mathbf{R} \mathbf{V}$  in (5).

The choice of  $g$  loosely reflects the type of least-squares maximum likelihood estimation from previous work after that work is modified to handle joint mixture models instead of presuming a-priori knowledge of line assignment. In particular, if we model  $v^T l$  as a Gaussian distribution given that  $l$  corresponds to  $v$ , we can easily construct a mixture model from this distribution and the uniform “noise” distribution. In this case, one can show that the log likelihood is  $g(v^T l) = \log(1 + A g_\sigma(v^T l)) + B \simeq A g_\sigma(v^T l) + B$  for

some constants A and B. Ignoring these trivial additive and multiplicative factors (that do not affect the maximization) gives our choice of  $g_\sigma$ . Although our choice uses a non-Gaussian observation model because of this loose approximation, this model will not perform significantly differently than a Gaussian model. In particular, note that this density that is the exponential of a Gaussian is shown (via a Taylor expansion) to be a Gaussian chosen from a mixture model with parameter probability  $P(\sigma^2 = \sigma_0^2/k) = (e^{-k})^{-1}$  where  $k = 0$  is the uniform model. We then argue that our induced distribution is no more arbitrary than assuming a Gaussian model with indeterminate  $\sigma$  parameter.

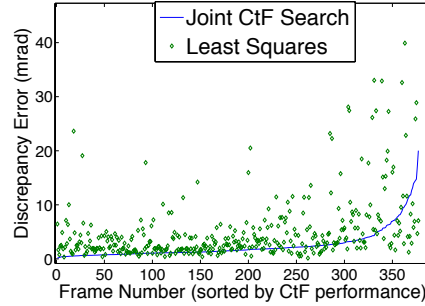
Although the nature of this optimization makes direct solutions intractable, the space of rotations only has 3 degrees of freedom in  $\omega$ , so it will be feasible to use a semi-exhaustive search. We use a coarse-to-fine approach, starting with an estimate  $\omega_0$  and sampling our optimization function with coarse  $g_\sigma$  (large  $\sigma_0$ ) in the neighborhood of  $\omega_0$  and choose the best candidate  $\omega_1$ , then using this candidate for the next finer scale, and so on while dividing  $\sigma$  by 2 at each iteration. As  $\sigma_k$  approaches zero, the optimization essentially tries to maximize a path integral through the filtered-Radon domain.

These samples for  $\omega_{k+1}$  are drawn randomly from  $\omega_k + v$ ,  $v \sim \mathcal{N}(0, \sigma_k)$ . In our work, we found 200 randomly generated samples per scale to work well. This is preferable to using a rectangular Cartesian grid because it is isotropic, and because there is a small but nonzero probability of making a large step and escaping from a local maximum from the previous coarser scale.

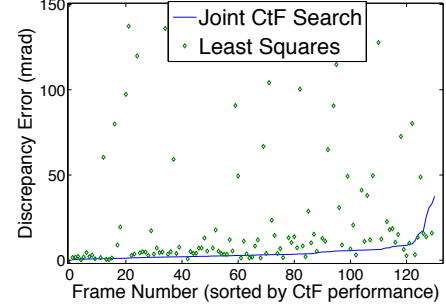
This initial estimate  $\omega_0$  may be obtained from an a-priori estimate (e.g. the previous frame). When no a-priori information is available, it is possible to do an exhaustive search by sampling the solid sphere  $\|\omega\| \leq \pi$ , requiring on the order of  $\sigma_0^{-3}$  samples.



(a)



(b)



(c)

**Fig. 3.** Here we show recovery performance for the (b) Biltmore and (c) office environment sequences with the frame indices reordered for clarity. On the Biltmore video sequence (e.g. (a)), our orientation recovery estimates the absolute orientation robustly against the distractor features from the cars and foliage and is able to “link” the windows together. For the office environment (e.g. Figure 2(d)), our approach not only incorporates strong lines, but also more subtle line textures.

#### 4. RESULTS

The filtered-Radon operator is flexible enough to effectively handle environments that would cause problems for the gradient-magnitude Radon (Hough) transform. However, the effectiveness of the orientation recovery remains to be seen.

Evaluating this approach presents a challenge, because (as is often the case in geometric computer vision) we are working beyond the precision of what most off-the-shelf instruments can reliably and practically measure. One method for evaluation would use the inner products of each pair of vanishing points as a type of consistency metric, since these values should be close to zero. However, this metric can be misleading when the estimation procedure is biased towards orthogonal vanishing points to begin with, and has no meaning when one strictly enforces orthogonality as we have done.

Instead, for our “ground truth” reference we rely on the still-more accurate estimates of relative orientation between adjacent frames of an image sequence, extracted from the essential matrix estimated from reliably matched points using the Scale Invariant Feature Transform (SIFT) [8]. We can then compute the rotational discrepancy between our estimated frame and the frame that was predicted from the previously estimated frame using the relative orientation between the two (from the essential matrix). This discrepancy gives a crude estimate for the orientation error.

The results are shown in Figure 3 for our approach and the least squares approach mentioned in the introduction.<sup>1</sup> There were many outliers when the least-squares procedure failed due to the loss of line-texture information to motion blurring. Notwithstanding these outliers, the least squares approach yields an average of 31.8 milliradians, while our joint coarse-to-fine search yields 4.7 milliradians. Because of the heavy-tail distribution of the least-squared approach, the median is a more reasonable 10.1 milliradians while ours has a median of 3.0 milliradians. The Biltmore sequence had a mean and median of 2.5 and 1.6 milliradians for our coarse to fine search and 6.6 and 2.8 milliradians for the least squares approach.

#### 5. CONCLUSION

In summary, we have introduced a functional approach to the characterization of line structure in images with the application of van-

ishing point estimation in mind. In addition to detecting lines, this operator also detects alignment. For example, text on the wall that is aligned in the horizontal direction is naturally detected by this operator and contributes to the horizontal vanishing point estimation. This operator may well prove to be useful in other applications as well. For example, collapsing  $Q$  to a function of  $\theta$  by summing over  $m$  produces a sort of histogram of gradient directions in the image, commonly used in pattern recognition. At present, this operator takes about 30 seconds (in Matlab) to compute for a 1 Megapixel image, with most of the bottleneck coming from the Fractional Fourier Transform.

#### 6. REFERENCES

- [1] C. Rother, “A new approach to vanishing point detection in architectural environments,” *Image and Vision Computing*, vol. 20, no. 9-10, pp. 647–655, 2002.
- [2] J.A. Shufelt, “Performance Evaluation and Analysis of Vanishing Point Detection Techniques,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 3, pp. 282–288, 1999.
- [3] M. Antone and S. Teller, “Scalable Extrinsic Calibration of Omni-Directional Image Networks,” *International Journal of Computer Vision*, vol. 49, no. 2, pp. 143–174, 2002.
- [4] J. Kosecka and W. Zhang, “Video compass,” *Proceedings of European Conference on Computer Vision*, pp. 657–673, 2002.
- [5] A. Averbuch, RR Coifman, DL Donoho, M. Israeli, and J. Walden, “Fast slant stack: a notion of Radon transform for data in a Cartesian grid which is rapidly computable, algebraically exact, geometrically faithful and invertible,” 2001.
- [6] D. Bailey and P. Swartztrauber, “The fractional Fourier transform and applications,” *SIAM Review*, vol. 33, no. 3, pp. 389–404, 1991.
- [7] T. Tuytelaars, L. Van Gool, M. Proesmans, and T. Moons, “The cascaded Hough transform as an aid in aerial image interpretation,” pp. 67–72, 1998.
- [8] D.G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

<sup>1</sup><http://www.youtube.com/ICIP2010willem>