Report Number: WUCSE-2004-30

2004-06-03

Road extraction from motion cues in aerial video

Robert Pless and David Jurgens

Aerial video provides strong cues for automatic road extraction that are not available in static aerial images. Using stabilized (or geo-referenced) video data, capturing the distribution of spatio-temporal image derivatives gives a powerful, local representation of scene variation and motion typical at each pixel. This allows a functional attribution of the scene; a “road” is defined as paths of consistent motion --- a definition which is valid in a large and diverse set of environments. Using a classical relationship between image motion and spatio-temporal image derivatives, road features can be extracted as image regions that have significant image variation... **Read complete abstract on page 2.**

Road extraction from motion cues in aerial video

Robert Pless and David Jurgens

Complete Abstract:

Aerial video provides strong cues for automatic road extraction that are not available in static aerial images. Using stabilized (or geo-referenced) video data, capturing the distribution of spatio-temporal image derivatives gives a powerful, local representation of scene variation and motion typical at each pixel. This allows a functional attribution of the scene; a “road” is defined as paths of consistent motion --- a definition which is valid in a large and diverse set of environments. Using a classical relationship between image motion and spatio-temporal image derivatives, road features can be extracted as image regions that have significant image variation and a motion consistent with its neighbors. The video pre-processing to generate image derivative distributions over arbitrarily long sequences is implemented in real time on standard laptops, and the flow field computation and interpretation involves a small number of 3 by 3 matrix operations at each pixel location. Example results are shown for an urban scene with both well-traveled and infrequently traveled roads, indicating that both can be discovered simultaneously. This method works robustly in scene with significant traffic motion and is thus ideal for urban traffic scenes, which often are difficult to analyze using static imagery.

Department of Computer Science & Engineering

Washington University in St. Louis
School of Engineering & Applied Science

2004-30

Road extraction from motion cues in aerial video

Authors: Pless, Robert; Jurgens, David

June 3, 2004

Abstract: Aerial video provides strong cues for automatic road extraction that are not available in static aerial images. Using stabilized (or geo-referenced) video data, capturing the distribution of spatio-temporal image derivatives gives a powerful, local representation of scene variation and motion typical at each pixel. This allows a functional attribution of the scene; a firoad?is defined as paths of consistent motion --- a definition which is valid in a large and diverse set of environments. Using a classical relationship between image motion and spatio-temporal image derivatives, road features can be extracted as image regions that have significant image variation and a motion consistent with its neighbors. The video pre-processing to generate image derivative distributions over arbitrarily long sequences is implemented in real time on standard laptops, and the flow field computation and interpretation involves a small number of 3 by 3 matrix operations at each pixel location. Example results are shown for an urban scene with both well-traveled and infrequently traveled roads, indicating that both can be discovered simultaneously. This method works robustly in scene with significant traffic motion and is thus ideal for urban traffic scenes, which often are difficult to analyze using static imagery.

WUCSE-2004-30: Road extraction from motion cues in aerial video

Robert Pless and David Jurgens

Washington University, Department of Computer Science and Engineering

Box 1045, One Brookings Ave., St. Louis, MO, 63130

{pless, daj2}@cse.wustl.edu

Abstract:

Aerial video provides strong cues for automatic road extraction that are not available in static aerial images. Using stabilized (or geo-referenced) video data, capturing the distribution of spatio-temporal image derivatives gives a powerful, local representation of scene variation and motion typical at each pixel. This allows a functional attribution of the scene; a "road" is defined as paths of consistent motion --- a definition which is valid in a large and diverse set of environments. Using a classical relationship between image motion and spatio-temporal image derivatives, road features can be extracted as image regions that have significant image variation and a motion consistent with its neighbors. The video pre-processing to generate image derivative distributions over arbitrarily long sequences is implemented in real time on standard laptops, and the flow field computation and interpretation involves a small number of 3 by 3 matrix operations at each pixel location. Example results are shown for an urban scene with both well-traveled and infrequently traveled roads, indicating that both can be discovered simultaneously. This method works robustly in scene with significant traffic motion and is thus ideal for urban traffic scenes, which often are difficult to analyze using static imagery.

1. Introduction

Automatically populating databases with current information about road networks is important in the automatic acquisition and update of geographic information systems (GIS). Both civic planning and tactical response to emergency situations require the data to reflect current conditions. The extraction of roads from image data has led to significant scientific inquiry within the Computer Vision community developing tools for scale-invariant detection of features amidst significant and highly varied background clutter. The complexity of this problem requires image analysis systems to include significant semantic modeling, allowing context based reasoning to be used in image areas with ambiguous image data. Hinz states:

> *"It is clear that these techniques must be integral parts of an extraction system to attain reasonably good results over a variety of scenes" [Hinz 2003]*

While this assertion may be valid for extracting roads from single images, we argue here that the ambiguities are largely eliminated in the analysis of video data from a scene. Video data is particularly beneficial for urban scenes, where roads tend to be more difficult to identify from a single image but there is a high traffic volume and therefore consistent motion cues.

Historically, several problems have limited the use of video data in photogrammetry --- the relatively low resolution of video and the massive and highly redundant form of the data set. These problems have been ameliorated with the wider availability of mega-pixel video cameras and algorithmic advances that allow real time stabilization (registering a video to an internally consistent coordinate system or geo-registration), and anomaly detection as tools for extracting efficient representations of the data. For the remainder of this paper we will assume that the video has been stabilized, so that motion within the video is caused by (1) objects moving in the scene, (2) the background motion of fixed objects in the scene (trees, water motion), or (3) residual (unstabilized) motions of objects that are significantly above the ground plane, shadows of these objects and so on,

This work is inspired by recent work in video surveillance – anomaly detection algorithms that are effective at modeling consistent background motions (eg. trees waving in the wind, water waves, or consistent traffic patterns) in order to trigger an alarm or to save data when an unusual event occurs or an object moves through the scene in an unusual manner. The construction of these models, which are intended to capture the typical behavior of a scene, turns out to be an ideal pre-processing and video-data summarization step in terms of identifying roads in a scene. Several addition points make this approach particularly compelling:

- The anomaly detection method is based on capturing the joint distribution of spatio-temporal image derivatives, not tracking of objects over long time periods. Therefore the data can be generated by many short time sequences (at least pairs of images) and does require continuous imaging of the same area. This gives flexibility in the data capturing process and allows road extraction data analysis to be piggy-backed on data captured for other purposes (such as aerial surveillance).
- The processing of the video data gives, for each pixel or small image region, the best fitting motion direction, and a measure of how consistent the image derivatives are with a single motion direction. This serves a new pre-processing step and can be integrating with current techniques including contextual cues as well as tools such as snakes, condensation, or particle filtering.
- This method is effective at seeding static image analysis methods. Detecting roads based on consistent motion patterns is highly effective, but only for roads with visible traffic. The

parameters of the roads detected in this manner (the image size, color, typical curvature, etc) can be used to seed image based methods with parameters *specific* to the given data set, in order to detect the remaining roads in the scene.

- Finally, additional information may be available depending on the length of time a scene is observed. The volume or frequency of travel along the road and the distribution of vehicle speeds may be captured from video data and are important in some applications.

The following section attempts to place this work in the context of recent approaches to road extraction. Section 3 introduces the real-time approach to spatio-temporal image processing and techniques to maintain a representation of the motion distributions at each pixel. Section 4 discusses the analysis of these models in the context of road extraction, and Section 5 gives several experiments demonstrating the technique.

2. Background

Many of the systems for road extraction can be categorized in terms of their (1) front end sensors, (2) initial data filtering and analysis at the pixel or local level, and (3) methods to define extended paths on the basis of initial image data. Here we present a sparse survey of recent literature on road extraction methods as a means of putting our proposed approach into context. Although our approach is defined explicitly in subsequent sections, for comparison purposes, it would be categorized in this framework as using (1) aerial *video* and (2) extended spatio-temporal filtering. We are explicitly agnostic about the third component, and emphasize that the front end processing we propose is relevant for multi-resolution, active testing, or snake based models of extended roads.

Geman and Jedanyk discuss road extraction from satellite imagery [Geman96]. They argue that immediately classifying pixels as “road” or “background” is infeasible because the local region surrounding a road pixel and a background pixel may appear identical (even in multi-spectral LandSat imagery). Thus they propose a particular, brightness invariant local operator and use an active testing approach that follows the road appearance and path by minimizing an entropy measure. Additional methods improve road detection by integrating larger local windows of image appearance. Examplars for this approaach advocate for multi-scale analysis for the extraction of road network from multi-spectral imagery [Wiedemann98], and using snakes as a method of finding long regions that are straight or curve slowly [Laptev00].

Focusing more on the data analysis at the pixel level, Porkili proposes a set of line-filters that measures both how likely a particular pixel is a road, as well as the direction of the possible road at that pixel. As

the algorithm progresses, the ends of currently detected roads can be extended to areas that have a very low likelihood of being a road, as long as they have the correct orientation [Porkili03]. Related methods define the road as a probabilistic contour, and use color and gradient information to extend contours across occlusions or shadows [Bicego03].

Synthetic aperture radar (SAR) has been considered as a front end sensor to simplify the process of road extraction. Tupin considers the problem of road extraction in urban areas, and proposes a 2-step algorithm that extracts line features from the speckle radar image and subsequently uses a Markov random field to impose contextual knowledge to cluster the detected segments into roads [Tupin98]. Wessel argues that road extraction from SAR imagery is effective for highways where there are no scattering objects (signs, or bridges) that interfere with the road, but are ineffective in industrial areas (which tend to always have scattering objects) or for secondary roads (which tend to have insufficient signal return).

Returning to aerial imagery, the papers most closely related to this approach concentrate on road extraction in urban environments. Hinz points out that most work focuses on the easier problem of rural road extraction, and existing work on urban scenes make assumptions about the grid structure of many city streets [Faber00], or combines height models and high-resolution imagery to extract streets through residential areas [Price00]. To be effective in more general situations, a system is proposed that incorporates a great deal of detailed knowledge about roads and their context, uses explicit formulated scale-dependent models of road appearance, and continually performs hypothesis testing to ensure that the local context information is appropriate [Hinz03].

Finally, from the computer vision community, work on traffic monitoring using a “forest of sensors”, is effective at creating trajectories of objects tracked through an environment [Grimson98]. While this could form the basis for an approach similar to ours, Grimson does not consider the problem of road extraction, and their method requires continual long term surveillance to build trajectories, instead of capturing and integrating short term motion cues.

A technical issue directly related to our approach, and otherwise independent of road extraction, is that we require the input video sequence to be stabilized, so that collecting statistics of spatio-temporal filter responses over many frames at a single pixel gives motion information about the same scene point. Numerous algorithms for this process have been proposed using either the tracking of feature points [Wixson00], or based directly on spatio-temporal filter responses [Pless00, Dai01]. We adapt the method used in [Pless00], which involves, for each frame, computing spatio-temporal filters at a sparse set of image points, and solving for a general linear transformation (the image warping homography) that minimizes the change from the previous frame. The sequence of the warped images becomes the

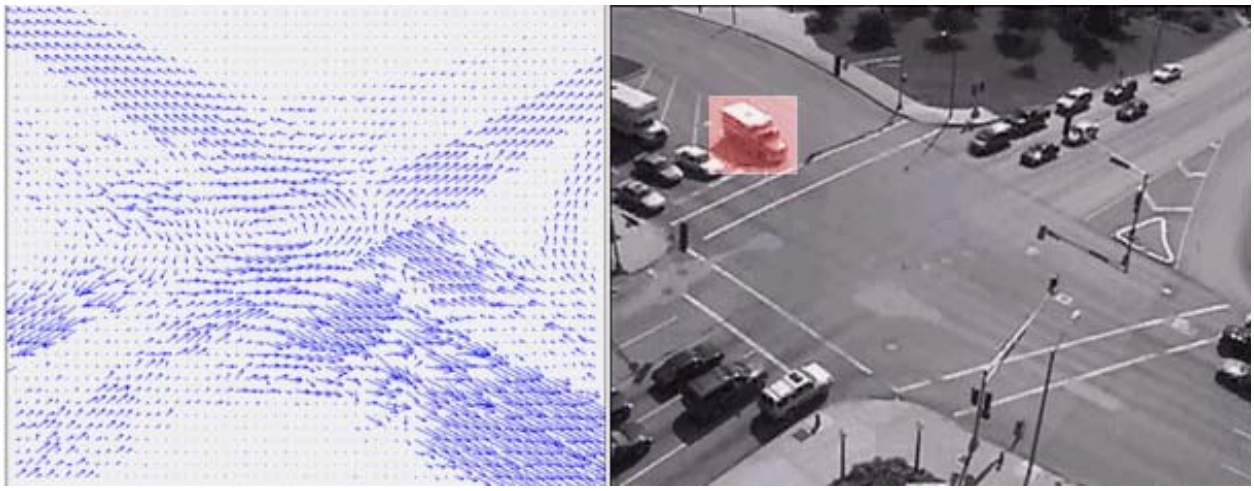

Figure 1. The optic flow field solved for the scene at the right, using 10 consecutive minutes of video data. Near the bottom right of the scene, specular reflections caused by the sunlight reflecting directly into the camera invalidate the assumptions of Equation 1, and cause inconsistent optic flow vectors. This model of typical motion on the scene allows identification of anomalous behavior, such as the ambulance highlighted because it has to go outside the normal traffic pattern to get through the scene. The inspiration for the present work is in the recognition that this background model has significant descriptive power about the scene that can be used for more than simply anomaly detection.

stabilized video used as input to the algorithm described below. Video data that is tagged with accurate knowledge of the 3D position and orientation of the camera in each frame permits warping and stabilization without additional image processing.

In summary, to our knowledge, no one has directly considered the question of using aerial video imagery in the detection of roads. Recent advances in computational power and algorithmic maturity make the use of video data feasible. Using video imagery to define roads based upon motion patterns is most effective in urban environments --- a domain that remains particularly challenging for both image and SAR based analysis.

3. Approach

The approach is based upon spatio-temporal image analysis. This approach explicitly avoids finding or tracking image features. Instead, the video is considered to be a 3D function I(x,y,t), defining the image intensity as it varies in space (across the image) and time. The fundamental atoms of the image processing are the value of this function and the response to spatio-temporal filters (such as derivative filters), measured at each pixel in each frame. Unlike interest points or features, these measurements are defined at every pixel in every frame of the video sequence. Appropriately designed filters may give robust measurements to form a basis for further processing. Optimality criteria and algorithms for creating derivative and blurring filters of a particular size have been developed by [Farid97], and lead to significantly better results than estimating derivatives by applying Sobel filters to raw images. For these

reasons, spatio-temporal image processing provides an ideal first step for streaming video processing applications.

Spatio-temporal image derivative filters are particularly meaningful in the context of analyzing motion on the image. Considering a specific pixel and time (x,y,t), we can define $I_x(x,y,t)$ to be the derivative of the image intensity as you move in the x-direction of the image. $I_y(x,y,t)$, and $I_t(x,y,t)$ are defined similarly. Dropping the (x,y,t) component, and the optic flow constraint equation gives a relationship between I_x, I_y, and I_t, and the optic flow, (the 2d motion at that part of the image) [Horn81]:

$$I_x\, u + I_y\, v + I_t = 0.$$

This classical equation in computer vision holds true for smoothly varying images, when the motion (the magnitude of the <u,v> vector) is relatively small, and the only reason that the intensity at a pixel changes is because of local motion in the image. Since this gives just one equation, which has two unknowns (u,v), it is not possible to directly solve for the optic flow. Many optic flow algorithms therefore assume that the optic flow is constant over a small region of the image, and use the (I_x, I_y, I_t) values from neighboring pixels to provide additional constraints. This assumption does not hold true at the boundaries of objects, leading to consistent errors in the optic flow solution.

The gain from stabilized video. The advantage of stabilized video is that instead of combining data from a spatially extended region of the image, we can instead combine equations through time. This allows one to compute the optic flow at a single pixel location without any spatial smoothing. Figure 1 shows one frame of a video sequence of a traffic intersection, and the flow field that best fits the data for each pixel over time. The key to this method is that the distribution of intensity derivatives, (I_x, I_y, I_t) --- simply the distribution, and not, for instance the time sequence --- encodes several important parameters of the underlying variation at each pixel. Fortunately, simple parametric representations of this distribution have the dual benefits of (1) the parameters are efficient to update and maintain, allowing real-time systems, and (2) the set of parameters for the entire image efficiently summarize and encode features of interest to GIS applications.

3.1 Representations.

Although we make no claim that the (I_x, I_y, I_t) measurements are drawn from a Gaussian distribution, we choose to model the background distribution by assuming a mean of 0 and modeling the complete covariance of the best fitting Gaussian model. Concretely, we maintain (independently for each pixel), the free parameters of the (symmetric) co-variance matrix Σ. For our measurement vector which has

length 3, this is a total of 6 parameters for each pixel. The co-variance can be exactly updated online, so the total storage required is 6 × (# of pixels in each image), regardless of the overall length of the video.

Properties of the covariance matrix of intensity derivatives.

The covariance matrix of the intensity derivatives has a number of interesting properties that are exposed through computation of its eigenvalues and eigenvectors. In particular, suppose that a matrix Σ has eigenvectors (v_1, v_2, v_3) corresponding to eigenvalues (e_1, e_2, e_3), and suppose that the eigenvalues are sorted by magnitude with v_1 as the largest magnitude. The following properties hold:

- The vector v_3 is a homogenous representation of the total least squares solution [VanHuffel91] for the optic flow, that is the total least squares solution for optic flow is $v_3(1)/v_3(3)$, $v_3(2)/v_3(3)$. We call this 2d vector (f_x, f_y), for flow.
- If, for all the data at that pixel, the set of image intensity derivatives exactly fits some particular optic flow, then e_3 is zero.
- If, for all the data at that pixel, the image gradient is in exactly the same direction, then e_2 is zero. (this is the manifestation of the aperture problem)
- The value $(1 - e_3/e_2)$ varies from 0 to 1, and is an indicator of how consistent the image gradients are with the best fitting optic flow, with 1 indicating perfect fit, and 0 indicating that many measurements do not fit this optic flow. We call this measure **c**, for consistency.
- The ratio e_2/e_1, varies from 0 to 1, and is an indicator of how well specified the optic flow vector is, when this number is close to 0, the image derivative data could fit a family of optic flow vectors with relatively low error, when this ratio is closer to 1, then the best fitting optic flow is better localized. We call this measure **s**, for specificity.

These properties are not entirely new, but in the typical context of computer vision are less important, because the covariance matrix is made from measurements in a region of the image that is assumed to have constant flow. Since this assumption breaks down as the patch size increases, there is strong pressure to use patches as small as possible, instead of including enough data to validate the statistical analysis of the covariance matrix. However, in the persistent vision paradigm, we can collect sufficient data at a pixel by aggregating measurements through time, and this analysis becomes more relevant.

In summary, the interpretation of the covariance matrix results in the following scalar information at each pixel (x,y):

- $\Sigma(x,y)$ – the tensor field consisting of the covariance matrix at each pixel,

- $\langle f_x(x,y), f_y(x,y) \rangle$ --- the 2d vector field of the optic flow field (total least squares solution) at each pixel,
- **s**(x,y) --- the specificity of the optic flow solution at that pixel,
- **c**(x,y) --- the consistency of that optic flow solution at that pixel.

The claim is that these variables are en effective summary of information contained in a video sequence, and that the analysis of these scalar, vector, and tensor fields is an effective method for extracting road features from stabilized video. The following section explores some initial work in this direction.

4. Demonstration

The formal evaluation is limited by the lack of standardized test data nd the absence of other algorithms which compute GIS features from stabilized aerial video. To address this problem in the future, the code and data sets presented in the following section are publicly available at (http://www.cse.wustl.edu/~pless/videoGIS.html). Here we completely define the approach taken for an example data set, indicate the results, point out limitations and indicate areas of productive future research.

4.1 Methods

We have found that the method is quite robust to the specifics of the implementation details, but for concreteness we describe here the exact choices used in the results. Consecutive pairs of images from the video sequence are decompressed to create 2D arrays of intensity values. The image is convolved with a discrete 11 by 11 filter approximating a Gaussian with standard deviation of three pixels to create a blurred image. The I_x and I_y values are computed with appropriately oriented Sobel filters convolved with the blurred image. The I_t value is estimated as the difference between pixel values in consecutive frames. This (I_x, I_y, I_t) measurement is maintained for every pixel in the image, but is ignored at pixels whose distance from the boundary is less than 6 pixels, as the results of the convolution filters at these points depends upon assumptions about pixel values outside the image. At each pixel, a 3 x 3 covariance matrix is maintained by storing 7 parameters, (ΣI_x^2, ΣI_y^2, ΣI_t^2, $\Sigma I_x I_y$, $\Sigma I_x I_t$, $\Sigma I_y I_t$, n). To isolate the effects of image motion from intensity gradients that exist in the static image, these sums, and the value of n, is only updated when $|I_t| > 1$. The number n records the number of measurements that have been used in each of the sums. We emphasize that the above sums are taken through time, and these parameters are recorded

separately for each pixel, so some pixels may have more measurements that define the covariance matrix than others.

4.2 Results

The original image and several of the results of the pre-processing methods described above are shown for a 451 frame stabilized aerial video. This video is taken at approximately 3 frames per second. The video was shot from an aerial platform and geo-registered. Two frames of the geo-registered video are shown at the top of Figure 2, the black areas at the bottom corners arise because these images are warped to the coordinate system of the reference frame and these areas were outside the image. For each pixel, a score is calculated to measure how likely that pixel is to come from a road. This score function is:

$$\mathbf{s}\,\mathbf{c}\,\Sigma_{\cdot} I_t^2,$$

which is the intensity variance at that pixel, modulated by the previously defined scores that measure how well the optic flow solution fits the observed data (**c)** and how unique that solution is (**s**). This score is thresholded (threshold value set by hand), and overlayed on top of the original image in the bottom left of Figure 2.

Where cars passed during the duration of the video clip, roads were detected in regions where image based detection systems would fail, especially in the upper of the two curved roads in the middle of the image. This justifies the earlier assertion that this system is ideal in urban areas, where the image based cues are less clear but where the frequency of traffic is sufficient that motion cues are consistently available.

However, the motion cues provide more information than simply a measure of whether the pixel lies on a road. The best fitting solution for the optic flow also gives the direction of motion at each pixel. The components of the motion vectors are shown as the top row of Figure 3. There is significant noise in this motion field because of substantial image noise and the fact that for some roads the data included few moving vehicles. A longer image sequence would provide more data and make flow fields that are well constrained and largely consistent. The method would continue to fail in regions that contain multiple different motion directions or where the optic flow constraint equations fail (see the flow field and discussion in the caption of Figure 1). To make this analysis feasible with short stabilized video segments, it is necessary to combine information between nearby pixels.

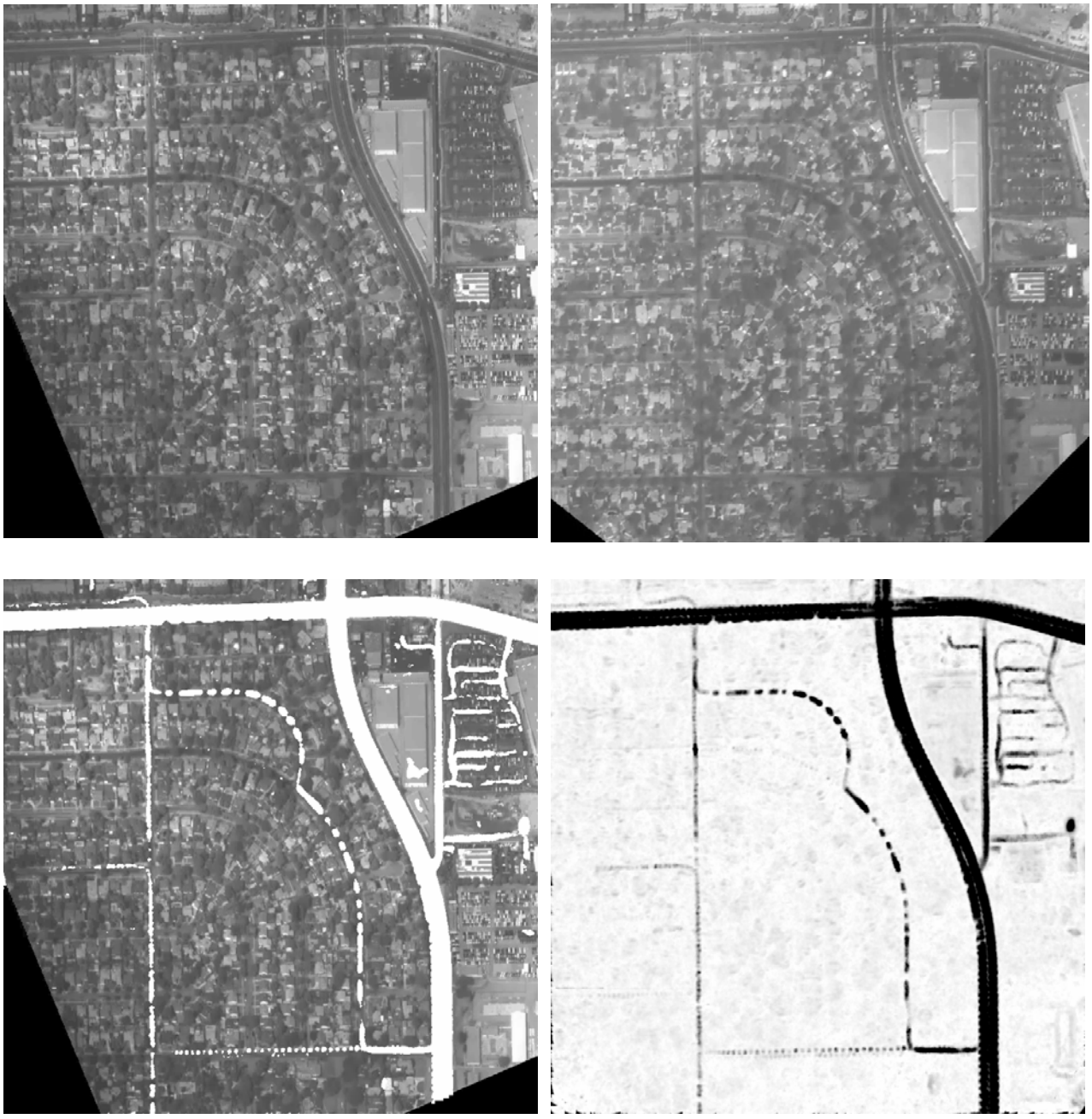

Figure 2. The top row shows frames 1 and 250 of a 451 frame stabilized aerial video (approximately 2:30 minutes long, 3 frames per second). The black in the corners are areas in this georegistered frame that are not captured in these images, these areas are in view for much of the sequence. The bottom right shows the amount of image variation modulated by the motion consistency --- a measure of how much of the image variation is caused by consistent motion as would be the case for a road (black is more likely to be a road).

Typically, combining information between pixels leads to blurring of the image and a loss of fidelity of the image features. However, the flow field that is extracted gives a best fitting direction of travel at each pixel. We use this as a direction in which we can combine data without blurring features – that is, we use the estimate of the motion to combine data along the roads, rather than across roads. This is a variant of motion oriented averaging [Nagel90].

Motion oriented covariance matrix smoothing:

Input: $\Sigma(x,y)$, $\langle f_x(x,y), f_y(x,y)\rangle$, defined at all x,y positions on the image.

Output: $\hat{\Sigma}(x,y)$, a covariance matrix that is the weighted sum of nearby covariance matrices, with more weight given to those that are in the direction of motion (either forward or backward).

After $\hat{\Sigma}(x,y)$ is computed, the flow field, and the flow field consistency and specificity measures can be again derived for the new covariance field as described in Section 3.1.

$$\langle d_x, d_y\rangle = \frac{\langle f_x(x,y), f_y(x,y)\rangle}{\sqrt{f_x^2(x,y) + f_y^2(x,y)}}$$

$$T = \begin{bmatrix} 30d_x & 30d_y \\ -3d_y & 3d_x \end{bmatrix}$$

$$M = T^\top T$$

$$w(a,b) = e^{-\langle a,b\rangle^\top M^{-1}\langle a,b\rangle}$$

$$\hat{\Sigma}(x,y) = \sum_{a=-5\ldots5} \sum_{b=-5\ldots5} w(a,b)\Sigma(x+a, b+y)$$

The bottom row of Figure 3 depicts the flow fields computed from the smoothed covariance matrix field. These fields are significantly cleaner than the flow field computed from the covariance matrices without smoothing and show clearly defined directions of travel for all roads with had motion. Furthermore, for large roads with significant motion cues, there is a clearly defined separation between different directions of travel.

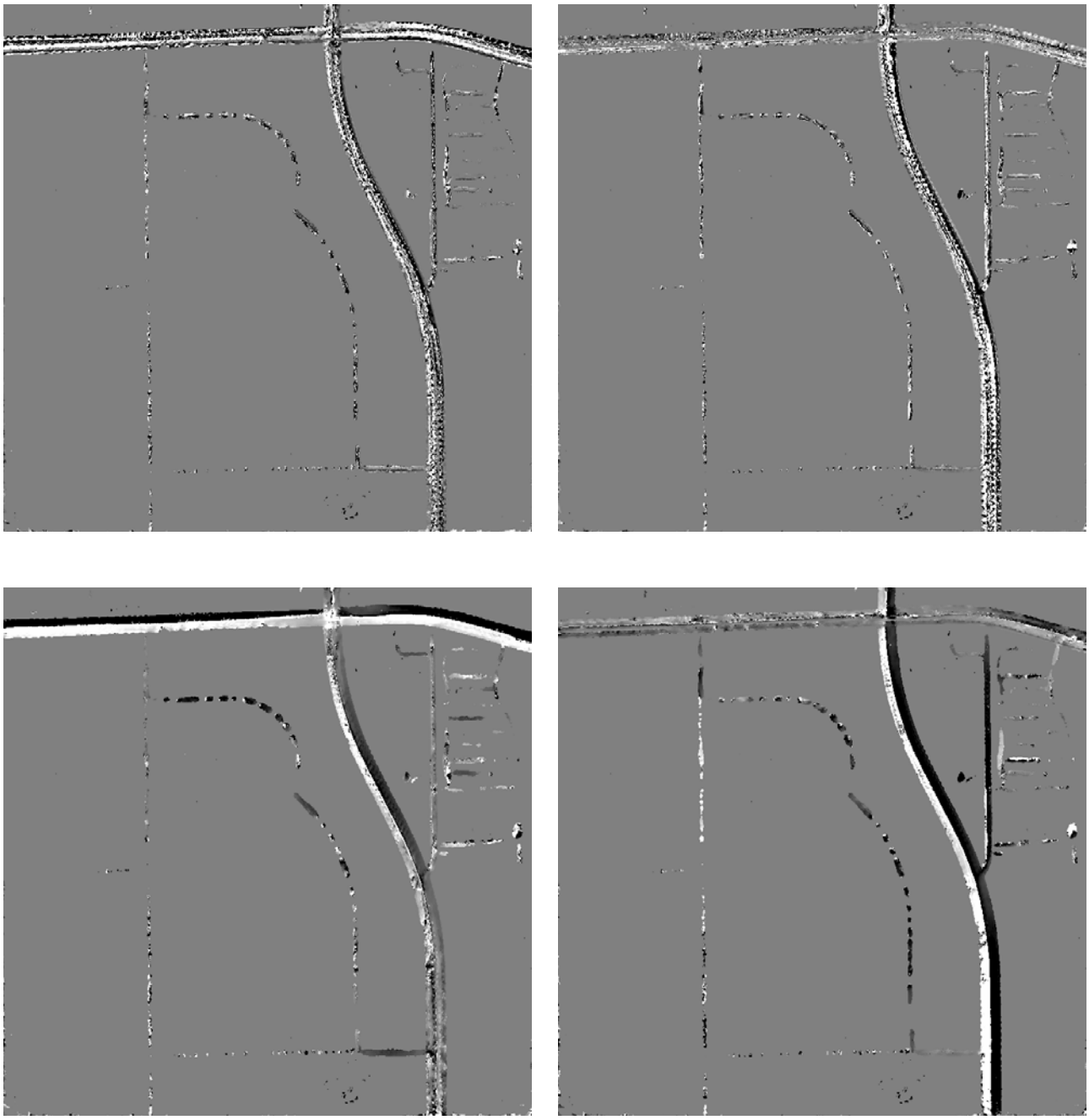

Figure 3. The top row show the x and y components of the best fitting optic flow vectors for the pixels designated as roads in figure 2. The flow fields are poorly defined, in part because of noisy data, and in part because there were few cars that move along some roads. These (poor) flow estimates were used to define the directional blurring filters that combine the image intensity measurements from nearby pixels (forward and backwards in the direction of motion). Using the covariance matrix data from other locations along the motion direction gives significantly better optic flow measurements (bottom row). In these images, black is negative and white is positive, relative to the origin in the top left corner of the image.

5. Discussion

This paper has presented straightforward algorithm for the analysis of stabilized video, and the marking of pixels whose intensity variation is consistent with motions along a road. Furthermore the direction of travel can be accurately identified. The motion cues are strongest for regions with significant traffic, a situation which is most relevant in urban settings. These settings are very challenging for current approaches based on the analysis of static images because of the typical complexity of the background.

Many avenues are open for future research, including joining motion based algorithms with more standard appearance based algorithms, as well as fitting the motion cues into larger systems that output symbolic representations of roads rather then pixel scores. Concrete future directions include:

- Integrate appearance or motion models specific to intersections, parking lots, and other features of interest.
- Use the motion based analysis of motion to bootstrap appearance based road modeling. For example, in Figure 2, not all roads in the scene are discovered because some roads did not have cars pass along them during the input video. However, statistical appearance models of the roads identified through motion cues would give a *scene-specific* road appearance model to find those roads that were missed.
- Adapt snake based road following algorithms (such as [Laptev00]) to incorporate the flow field direction in the energy function minimized by the snake.
- Design algorithms that close the loop between region interpretation and anomaly detection and detect unusual events such as cars stopping in unusual places.

5. References

[Bicego03] M. Bicego, S. Dalfini, G. Vernazza and V. Murino: "Automatic road extraction from aerial images by probabilistic contour tracking", Proc. of IEEE Int. Conf. on Image Processing (ICIP03), Vol. III, pp. 585-588, 2003.

[Dai01] X.-T. Dai, L. Lu, and G. Hager, "Real-time Video Mosaicing with Adaptive Parameterized Warping", Demo Program, CVPR'2001: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001.

[Faber00] A. Faber and W. Forstner, “Detection of Dominant Orthogonal Structures in Small Scale Imagery”, International Archives of Photogrammetry and Remote Sensing, Vol. 33, Part B3/1, pp. 274–281, 2000.

[Farid97] H. Farid and E. P. Simoncelli. “Optimally rotation-equivariant directional derivative kernels”, Computer Analysis of Images and Patterns (CAIP), 1997.

[Geman96] D. Geman and B. Jedynak, "An Active Testing Model for Tracking Roads in Satellite Images”, IEEE Transactions on Pattern Analysis and Machine Intelligence, **18** (1), pp. 1-14, 1996.

[Grimson98] W. E. L. Grimson, C. Stauffer and R. Romano and L. Lee, “Using Adaptive Tracking to Classify and Monitor Activities in a Site”, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 22-27, 1998.

[Hinz03] S. Hinz and A. Baumgartner, “Automatic extraction of urban road networks from multi-view aerial imagery”, ISPRS Journal of Photogrammetry and Remote Sensing, 58, pp. 83-98, 2003.

[Horn81] B.K.P Horn and B.G. Schunck, “Determining optical flow”, Artificial Intelligence, **17**, pp.185—203, 1981.

[Laptev00] I. Laptev, H. Mayer , T. Lindeberg, W. Eckstein, C. Steger, and A. Baumgartner, “Automatic extraction of roads from aerial images based on scale space and snakes,” Machine Vision and Applications, **12** (1), pp. 23-31, 2000

[Nagel90] H. H. Nagel, "Extending the "Oriented Smoothness Constraint" into the Temporal Domain and the Estimation of Derivatives of Optical Flow", European Conference on Computer Vision, pp. 139-148, 1990.

[Pless00] R. Pless, T. Brodsky and Y. Aloimonos, ``Detecting Independent Motion: The Statistics of Temporal Continuity," IEEE Transactions on Pattern Analysis and Machine Intelligence, **22** (8), pp. 768-773, 2000.

[Porikli03] F.M. Porikli, "Road Extraction by Point-wise Gaussian Models", SPIE AeroSense Technologies and Systems for Defense and Security, Vol. 5093, pp. 758-764, 2003.

[Price00] K. Price, “Urban Street Grid Description and Verification”, IEEE Workshop on Applications of Computer Vision, pp. 148–154, 2000.

[Tupin98] F. Tupin, H. Maitre, J.-F. Mangin, J.-M. Nicholas, and E. Pechersky, Detection of Linear “Features in SAR Images: Application to Road Network Extraction”, IEEE Transactions on Geoscience and Remote Sensing, **36** (2), pp. 434--453, 1998.

[VanHuffel91] S. Van Huffel and J. Vandewalle. *”The Total Least Squares Problem: Computational Aspects and Analysis”*. Society for Industrial and Applied Mathematics, Philadelphia, 1991.

[Wessel03] B. Wessel and C. Wiedemann, “Analysis of Automatic Road Extraction Results from Airborne SAR Imagery”, Proceedings of the ISPRS Workshop 'PHOTOGRAMMETRIC IMAGE ANALYSIS’, 2003.

[Wiedemann98] C. Wiedemann, C. Heipke, H. Mayer and S. Hinz, "Automatic Extraction and Evaluation of Road Networks from MOMS-2P Imagery", International Archives of Photogrammetry and Remote Sensing, **32** (3), pp. 285-291, 1998

[Wildes01] R.P. Wildes, D.J. Hirvonen, S.C. Hsu, R. Kumar, W.B. Lehman, B. Matei, and W.Y. Zhao, "Video Georegistration: Algorithm and Quantitative Evaluation", Proceedings of the International Conference on Computer Vision, Vol II, pp. 343-350, 2001.

[Wixson00] L. Wixson, “Detecting Salient Motion by Accumulating Directionally-Consistent Flow”**,** IEEE Transactions on Pattern Analysis and Machine Intelligence, **22** (8), pp. 7774-780, 2000.