Nonparametric Information Fusion for Motion Estimation

Dorin Comaniciu
Real-Time Vision and Modeling Department, Siemens Corporate Research
755 College Road East, Princeton, NJ 08540, USA
comanici@scr.siemens.com

Abstract

The problem of information fusion appears in many forms in vision. Tasks such as motion estimation, multimodal registration, tracking, and robot localization, often require the synergy of estimates coming from multiple sources. Most of the fusion algorithms, however, assume a single source model and are not robust to outliers. If the data to be fused follow different underlying models, the traditional algorithms would produce poor estimates. We present in this paper a nonparametric approach to information fusion called Variable-Bandwidth Density-based Fusion *(VBDF). The fusion estimator is computed as the location of the most significant mode of a density function which takes into account the uncertainty of the estimates to be fused. A novel mode detection scheme is presented, which relies on variable-bandwidth mean shift computed at multiple scales. We show that the proposed estimator is consistent and conservative, while handling naturally outliers in the data and multiple source models. The new theory is tested for the task of multiple motion estimation. Numerous experiments validate the theory and provide very competitive results.*

1 Introduction

Proper information fusion [16] is a critical step for many vision tasks [19, 20, 21]. Fusion is also an important topic across modalities, for applications such as collision warning and avoidance or speaker localization . Most often a classical estimation framework such as the (extended) Kalman filter [5] is employed to derive an estimate from multiple sensor data.

We assume that each sensor measurement is characterized by its mean vector and a covariance matrix defining the uncertainty of the mean. When the processing of all measurements takes place at a single location, the fusion is called *centralized*. In centralized fusion the sensor measurement errors are usually considered independent across sensors and time. A construction with improved reliability and flexibility is provided by *distributed* fusion [24], represented by a collection of processing nodes that communicate with each other. Such architecture handles the information in two steps: the sensor measurements are evaluated first, then, the state information from a local neighborhood is fused. A major topic in distributed fusion is the handling of cross-correlation, which is difficult to evaluate. The *Covariance Intersection* algorithm [18] provides a consistent and conservative solution to this problem.

The distributed fusion architecture is suitable for the task of motion estimation from image sequences. Indeed, for motion estimation we also need two processing steps. The assumption that some image property (such as the brightness) is conserved locally in time only constraints the component of the motion field in the direction of the spatial image gradient. A second step is necessary to fuse the initial motion estimates from a given neighborhood, to exploit spatial coherence [7].

The connection between motion flow computation and information fusion has been first recognized by Singh and Allen [34]. By assuming a single source model and statistical independence between the estimates to be fused, they employed the best linear unbiased estimator (BLUE) to combine local information. Later on, Simoncelli, Adelson and Heeger [33, 32] developed a Bayesian framework, which was also restricted to single motion estimation.

This paper addresses a fundamental issue in distributed fusion: how to deal with *multiple source models*, while maintaining a *consistent and conservative approach in handling cross-correlation*. Our solution is robust, nonparametric in nature, employing an adaptive kernel density function that exploits the uncertainty of the initial estimates. The new technique is called *Variable-Bandwidth Density-based Fusion* (VBDF). It defines the fusion estimator as the location of the most significant mode of the density function. The mode is computed using a novel multiscale optimization framework based on variable-bandwidth mean shift. Interestingly enough, the VBDF expression is close to that of Covariance Intersection, although our formulation relies on kernel density estimation theory.

We show that the application of the VBDF framework to multiple motion computation yields very competitive results. Many other vision tasks that require the fusion of some initial estimates computed in a given neighborhood (window) can benefit from our framework.

Section 2 formally introduces the fusion problem and presents its most common solutions. Variable-bandwidth kernel density estimation is discussed in section Section 3. The VBDF estimator and its properties are presented

in Section 4. In Section 5 we discuss the application of the new estimator to the optical flow problem, while experiments and comparisons are presented in Section 6.

2 Previous Work in Information Fusion

In this section we discuss the information fusion problem and show that its solution depends on how much we know about about cross-correlation. Let $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$ be two estimates that are to be fused together to yield an "optimal" estimate $\hat{\mathbf{x}}$. The error covariances are defined by

$$\mathbf{P}_{ij} = \mathrm{E}\left[(\mathbf{x} - \hat{\mathbf{x}}_i)(\mathbf{x} - \hat{\mathbf{x}}_j)^\top\right] \tag{1}$$

for $i = 1, 2$ and $j = 1, 2$. To simplify notation denote $\mathbf{P}_{11} \equiv \mathbf{P}_1$ and $\mathbf{P}_{22} \equiv \mathbf{P}_2$.

If the cross-correlation can be ignored, i.e., $\mathbf{P}_{12} = \mathbf{P}_{21}^\top = \mathbf{0}$, the best linear unbiased estimator (BLUE) is also called Simple Convex Combination [10] and is expressed by

$$\hat{\mathbf{x}}_{CC} = \mathbf{P}_{CC}\left(\mathbf{P}_1^{-1}\hat{\mathbf{x}}_1 + \mathbf{P}_2^{-1}\hat{\mathbf{x}}_2\right) \tag{2}$$

$$\mathbf{P}_{CC} = \left(\mathbf{P}_1^{-1} + \mathbf{P}_2^{-1}\right)^{-1} \tag{3}$$

When the initial estimates are correlated ($\mathbf{P}_{12} = \mathbf{P}_{21}^\top \neq \mathbf{0}$) and the noise correlation can be measured, the BLUE estimator $(\hat{\mathbf{x}}_{BC}, \mathbf{P}_{BC})$ is derived according to Bar-Shalom and Campo [4] using Kalman formulation. The most general case of BLUE estimation also assumes prior knowledge of the covariance of $\mathbf{x}$ [23].

A conservative approach to information fusion has been proposed by Julier and Uhlman in the form of Covariance Intersection algorithm [18]. Their objective was to obtain a *consistent* estimator of the covariance matrix when two random variables are linearly combined and their cross-correlation is unknown. Consistency means that the estimated covariance is always an upper-bound, in the positive definite sense, of the true covariance, no matter what the cross-correlation level is. The intersection is characterized by the convex combination of the covariances

$$\hat{\mathbf{x}}_{CI} = \mathbf{P}_{CI}\left(\omega\mathbf{P}_1^{-1}\hat{\mathbf{x}}_1 + (1-\omega)\mathbf{P}_2^{-1}\hat{\mathbf{x}}_2\right) \tag{4}$$

$$\mathbf{P}_{CI} = \left(\omega\mathbf{P}_1^{-1} + (1-\omega)\mathbf{P}_2^{-1}\right)^{-1} \tag{5}$$

where $\omega \in [0, 1]$. The parameter ω is chosen to optimize the trace or determinant of $\mathbf{P}_{CI}$.

Covariance Intersection has a very suggestive geometrical interpretation: if one plots the covariance ellipses $\mathbf{P}_1$, $\mathbf{P}_2$ and $\mathbf{P}_{BC}$ (as given by the Bar-Shalom/Campo formulation) for all choices of $\mathbf{P}_{12}$, then $\mathbf{P}_{BC}$ always lies within the intersection of $\mathbf{P}_1$ and $\mathbf{P}_2$. It results that a strategy that computes a $\mathbf{P}_{CI}$ that encloses the intersection region is consistent even for unknown $\mathbf{P}_{12}$. It has been shown in [18] that the difference between $\mathbf{P}_{CI}$ and the true covariance of $\mathbf{x}$ is a semipositive matrix. More recently, Chong and Mori [10] examined the performance of Covariance Intersection, while Chen, Arambel and Mehra [9] analyze the optimality of the algorithm.

Observe that the Covariance Intersection can be generalized to the fusion of n estimates as

$$\hat{\mathbf{x}}_{CI} = \mathbf{P}_{CI}\sum_{i=1}^{n}\omega_i\mathbf{P}_i^{-1}\hat{\mathbf{x}}_i \tag{6}$$

$$\mathbf{P}_{CI} = \left(\sum_{i=1}^{n}\omega_i\mathbf{P}_i^{-1}\right)^{-1} \tag{7}$$

with $\sum_{i=1}^{n}\omega_i = 1$. In equations (6) and (7) the weights ω_i are also chosen to minimize the trace or determinant of $\mathbf{P}_{CI}$.

Although very important from theoretical viewpoint, Covariance Intersection has two major weaknesses: it assumes a single source model and is not robust to outliers. In the next sections we show that these problems can be overcome by using adaptive kernel density estimation. The new VBDF estimator is defined as the *sample mode* of a density function constructed using kernels with variable-bandwidth.

3 Adaptive Density Estimation

Adaptive density estimation with variable kernel bandwidth [8] has been only recently applied in computer vision [12, 14]. The motivation for variable-bandwidth is to improve the performance of kernel estimators by adapting the kernel scaling and orientation to the local data statistics .

Let $\mathbf{x}_i$, $i = 1 \ldots n$, be n data points in the d-dimensional space R^d. By selecting a different bandwidth matrix $\mathbf{H}_i = \mathbf{H}(\mathbf{x}_i)$ (assumed full rank) for each $\mathbf{x}_i$ we define the *sample point* density estimator

$$\hat{f}_v(\mathbf{x}) = \frac{1}{n(2\pi)^{d/2}}\sum_{i=1}^{n}\frac{1}{\mid \mathbf{H}_i \mid^{1/2}}\exp\left(-\frac{1}{2}D^2\left(\mathbf{x}, \mathbf{x}_i, \mathbf{H}_i\right)\right) \tag{8}$$

where

$$D^2\left(\mathbf{x}, \mathbf{x}_i, \mathbf{H}_i\right) \equiv (\mathbf{x} - \mathbf{x}_i)^\top\mathbf{H}_i^{-1}(\mathbf{x} - \mathbf{x}_i) \tag{9}$$

is the Mahalanobis distance from $\mathbf{x}$ to $\mathbf{x}_i$. The variable-bandwidth mean shift vector at location $\mathbf{x}$ is given by [11]

$$\mathbf{m}_v(\mathbf{x}) \equiv \mathbf{H}_h(\mathbf{x})\sum_{i=1}^{n}\omega_i(\mathbf{x})\mathbf{H}_i^{-1}\mathbf{x}_i - \mathbf{x} \tag{10}$$

where $\mathbf{H}_h$ is the data-weighted harmonic mean of the bandwidth matrices computed at $\mathbf{x}$

$$\mathbf{H}_h(\mathbf{x}) = \left(\sum_{i=1}^{n}\omega_i(\mathbf{x})\mathbf{H}_i^{-1}\right)^{-1} \tag{11}$$

and

$$\omega_i(\mathbf{x}) = \frac{\frac{1}{|\mathbf{H}_i|^{1/2}} \exp\left(-\frac{1}{2} D^2\left(\mathbf{x}, \mathbf{x}_i, \mathbf{H}_i\right)\right)}{\sum_{i=1}^{n} \frac{1}{|\mathbf{H}_i|^{1/2}} \exp\left(-\frac{1}{2} D^2\left(\mathbf{x}, \mathbf{x}_i, \mathbf{H}_i\right)\right)} \tag{12}$$

are weights satisfying $\sum_{i=1}^{n} \omega_i(\mathbf{x}) = 1$. It can be shown that the iterative computation of the mean shift vector (10) always moves the point $\mathbf{x}$ to a location where the density (8) is higher or equal to the density at the previous location. As a result, an iterative hill-climbing procedure is defined, which converges to a stationary point (i.e., zero gradient) of the underlying density. We will give the convergence proof for the variable-bandwidth mean shift in the journal version of this paper.

4 VBDF Estimator

The VBDF estimator is defined as the *location of the most significant sample mode of the data*. In this section we present a multiscale framework for the detection of the most significant mode, discuss the properties of new estimator, and show performance comparisons of the VBDF against the BLUE and Covariance Intersection algorithms.

4.1 Computation Through Multiscale Optimization

Assume that the data points $\mathbf{x}_i$, $i = 1 \ldots n$ are each associated with a covariance matrix $\mathbf{C}_i$ that quantifies uncertainty . The location of the most significant mode is obtained in a multiscale fashion, by tracking the mode of the density function across scales. More specifically:

- We perform first mode detection using large bandwidth matrices of the form $\mathbf{H}_i = \mathbf{C}_i + \alpha^2 \mathbf{I}$, where the parameter α is large with respect to the spread of the points $\mathbf{x}_i$. The mode detection algorithm is based on mean shift and involves the iterative computation of expression (10) and translation of $\mathbf{x}$ by $\mathbf{m}_v(\mathbf{x})$ until convergence. At the largest scale, the mode location does not depend on the initialization (up to some numerical approximation error) since for large α the density surface is unimodal.

- In the next stages, the detected mode is tracked across scales by successively reducing the parameter α and performing mode detection again. At each scale the mode detection algorithm is initialized with the convergence location from the previous scale.

Note that for the last mode detection procedure, the bandwidth matrix associated with each data point is equal to the point covariance matrix, i.e., $\mathbf{H}_i = \mathbf{C}_i$, $i = 1 \ldots n$. Denote by $\hat{\mathbf{x}}_m$ the location of the most significant mode. Since the gradient at $\hat{\mathbf{x}}_m$ is zero we have $\mathbf{m}_v(\hat{\mathbf{x}}_m) = \mathbf{0}$ which means

$$\hat{\mathbf{x}}_m = \mathbf{H}_h(\hat{\mathbf{x}}_m) \sum_{i=1}^{n} \omega_i(\hat{\mathbf{x}}_m) \mathbf{H}_i^{-1} \mathbf{x}_i \tag{13}$$

$$\mathbf{H}_h(\hat{\mathbf{x}}_m) = \left(\sum_{i=1}^{n} \omega_i(\hat{\mathbf{x}}_m) \mathbf{H}_i^{-1} \right)^{-1} \tag{14}$$

4.2 Properties

Equations (13) and (14) define the VBDF estimator, which has the following properties:

- The covariance (14) of the fusion estimate is a convex combination of the covariances of initial estimates. Thus, the expression of the new estimator resembles that of Covariance Intersection, although its derivation has a different motivation, based on density estimation theory.

- The matrix $\mathbf{H}_h(\hat{\mathbf{x}}_m)$ is a consistent and conservative estimate of the true covariance matrix of $\hat{\mathbf{x}}_m$, irrespective of the actual correlations between initial estimates. The proof is similar to the consistency proof of the Covariance Intersection [18].

- While the weights in the Covariance Intersection algorithm are chosen by minimizing the trace or determinant of the covariance, our criterion is based on the *most probable value* of the data, i.e., the most significant mode. This is a more appropriate criterion, especially when the data is multimodal, i.e., the initial estimates belong to different source models. Such property is common for motion estimation since the points in a local neighborhood may exhibit multiple motions. The most significant mode corresponds to the most relevant motion.

- The tracking of the density mode across scales insures the detection of the most significant mode. The use of the Gaussian kernel is essential for the continuity of the modes across scales .

- Finally, by selecting the most significant mode, the estimate is also robust to outliers.

4.3 Comparisons

In this subsection we compare experimentally the new VBDF estimator against the BLUE and Covariance Intersection.

A synthetic input data is shown in Figure 1a and consists of 8 initial bi-variate estimates expressed as location and covariance. Each covariance is displayed as an ellipse with 95% confidence. Observe that the input data has a clearly identifiable structure of 5 measurements, while the other 3 measurements can be considered outliers. In

addition, the uncertainty of the data is rather low and the mean vectors are rather far apart from each other. This creates a difficult mode estimation problem.

The same figure shows the VBDF estimate, having a mean equal to $(-0.3499, 0.1949)$ and its covariance, represented by an ellipse of thick line (the VBDF ellipse masks one of the input ellipses). We have also plotted the trajectory of the mode tracking across scales. Each small circle indicates the result of mode detection for one scale.

In Figure 1b we compare our result with that of the BLUE fusion ((2) and (3)) and Covariance Intersection ((6) and (7)). The kernel density estimate computed with $\mathbf{H}_i = \mathbf{C}_i + \alpha^2 \mathbf{I}$ is shown in Figure 1c for different values of α. A triangle marks the location of the most significant mode across scales. The lower right figure is obtained with $\mathbf{H}_i = \mathbf{C}_i$ and corresponds to the VBDF estimate.

The following conclusions can be drawn:

- The BLUE estimate produces the most confident result, however, the presence of outliers in the data has a strong, negative influence on this estimate. At the same time the BLUE estimate can be overly confident by neglecting the cross-correlation.
- The Covariance Intersection is also negatively influenced by outliers. We optimized the weights to minimize the trace of the covariance matrix. However, since the optimization regards only the covariance and not the location, the resulting estimate is rather poor.
- The best result is produced (with less confidence) by the VBDF algorithm. Note that by employing the variable-bandwidth mean shift and mode tracking across scales, we also rely on optimizing the weights. Observe that (as expected) the VBDF algorithm has not been influenced by outliers.
- A very important observation should be inferred from Figure 1c. The most significant mode across scales is not the highest mode computed with the bandwidths $\mathbf{H}_i = \mathbf{C}_i$! Note the highest location on the density landscape computed with $\mathbf{H}_i = \mathbf{C}_i$ is located at $(0.2380, -1.333)$, which is different from the VBDF estimate. This conclusion is in agreement to our own intuition, that the most significant mode should not be determined based solely on local information. The multiscale algorithm makes the right choice in selecting the right mode.

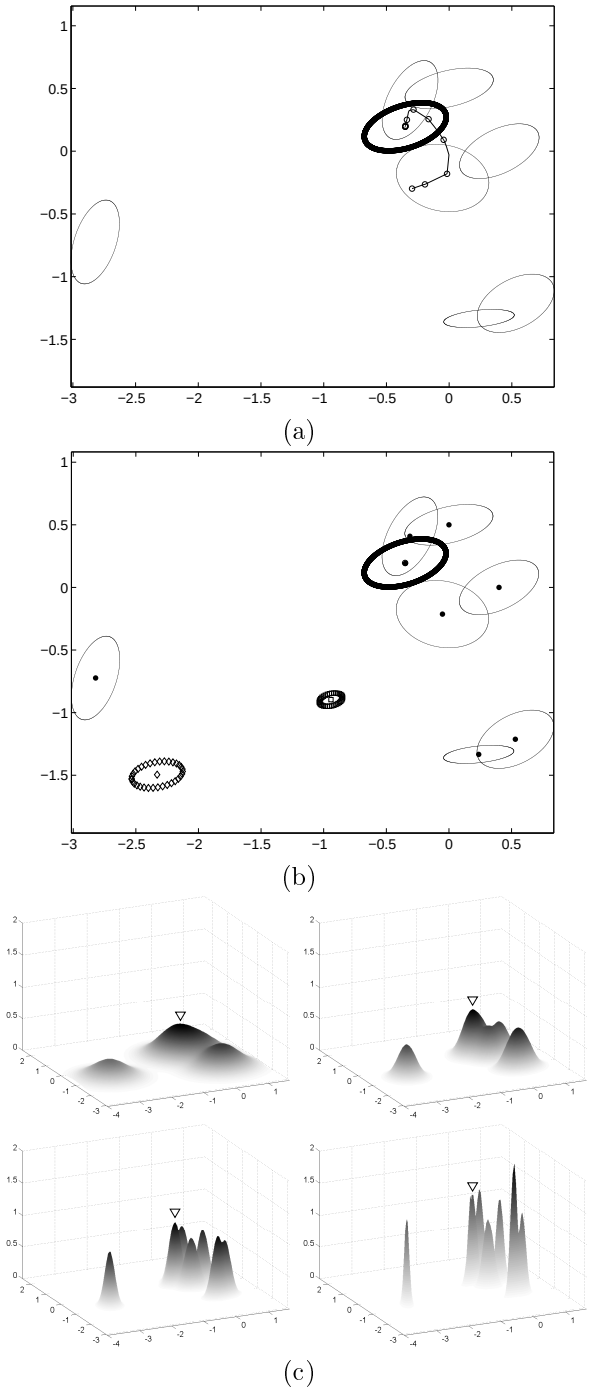


Figure 1: *VBDF Estimation.* (a) Input data represented as ellipses with 95% confidence. Trajectory of mode tracking across scales is shown. Ellipse corresponding to the VBDF estimate is drawn with thick line. (b) Fusion results overlayed on input data. Ellipses are represented with squares for BLUE estimate (smallest ellipse in the figure), diamonds for Covariance Intersection, and thick line for VBDF estimate. (c) Density surfaces corresponding to formula (8) with different values for α (see text). A triangle marks the mode that is tracked across scales. Lower right figure is the final result.

5 Estimation of Multiple Motion

This section presents the application of the VBDF estimator to the computation of multiple motion. We start with a short discussion on motion estimation, then explain how to compute the initial motion estimates and how to fuse them.

5.1 Previous Work in Motion Estimation

Detailed reviews on motion estimation are given by Aggarwal and Nandhakumar [1], Mitiche and Bouthemy [30], and Nagel [31]. Three main approaches to motion estimation can be identified, based on spatial gradient, image correlation, and regularization of spatio-temporal energy. The motion is commonly assumed to be locally constant, affine, or quadratic.

Most of the techniques based on spatial gradient embrace a two step approach for the computation of motion flow. An initial estimate of the flow is determined for each image location using the brightness constancy. The initial estimates are then fused locally in the hope for a better fusion estimate. The presence of multiple motions, however, makes the second task difficult since the initial estimates are generated by multiple and unknown source models. Multiple motions can be generated by objects moving with different velocities, but can also be the result of transparency, highlights or shadows.

One of the most popular and efficient optical flow techniques has been developed by Lucas and Kanade [26] in the context of stereo vision. They neglected the uncertainty of initial estimates and use (weighted) least squares in a neighborhood to fuse them. Later on, Weber and Malik [36] employed the total least squares for the same task. Simoncelli, Adelson and Heeger [33, 32] improved the method by computing and using the uncertainty of initial estimates. Nevertheless, they assume that the initial estimates are independent and do not model multiple motions. Black and Anandan [7] approached the motion estimation problem in a robust framework, being able to deal with multiple motions.

The first benchmarking effort on the evaluation of motion estimation algorithms has been conducted by Barron, Fleet and Beuchemin [6]. Since then, most of the newly proposed algorithms are compared using their methodology. We will do the same in this work.

5.2 Initial Estimates

For a given image location we extract an initial motion estimate from a very small NxN neighborhood using Biased Least Squares (BLS) [17, 27]

$$\hat{\mathbf{x}} = \left(\mathbf{A}^\top \mathbf{A} + \beta \mathbf{I}\right)^{-1} \mathbf{A}^\top \mathbf{b} \tag{15}$$

where $\mathbf{A}$ is the $N^2 \times 2$ matrix of spatial image gradients, and $\mathbf{b}$ is the N^2-dimensional vector of temporal images, as in [35, p.196].

The BLS solution has a covariance matrix $\mathbf{C}$ that is proportional to the variance σ^2 of the image noise. The advantage of BLS is that it avoids instability problems in the regular Least Squares solution by allowing a small amount of bias. The technique is also called *ridge regression* or *Tikhonov regularization* and various solutions have been proposed to compute the regularization parameter β from the data [15].

5.3 Fusion

We combine the motion flow information in a local image neighborhood of dimension $n = M \times M$ using the VBDF estimator (13) and (14). Denoting by $(\hat{\mathbf{x}}_i, \mathbf{C}_i)$, $i = 1 \ldots n$ the initial flow estimates produced through BLS, their fusion results in

$$\hat{\mathbf{x}}_m = \mathbf{C}(\hat{\mathbf{x}}_m) \sum_{i=1}^{n} \omega_i(\hat{\mathbf{x}}_m) \mathbf{C}_i^{-1} \hat{\mathbf{x}}_i \tag{16}$$

$$\mathbf{C}(\hat{\mathbf{x}}_m) = \left(\sum_{i=1}^{n} \omega_i(\hat{\mathbf{x}}_m) \mathbf{C}_i^{-1} \right)^{-1} \tag{17}$$

where

$$\omega_i(\hat{\mathbf{x}}_m) = \frac{\frac{1}{|\mathbf{C}_i|^{1/2}} \exp\left(-\frac{1}{2} D^2\left(\hat{\mathbf{x}}_m, \hat{\mathbf{x}}_i, \mathbf{C}_i\right)\right)}{\sum_{i=1}^{n} \frac{1}{|\mathbf{C}_i|^{1/2}} \exp\left(-\frac{1}{2} D^2\left(\hat{\mathbf{x}}_m, \hat{\mathbf{x}}_i, \mathbf{C}_i\right)\right)} \tag{18}$$

and $\hat{\mathbf{x}}_m$ is determined through mode tracking across scales, as discussed in Section 4.1.

6 Experiments

A standard procedure was employed to construct a three level image pyramid using a five-tap filter [0.0625 0.25 0.375 0.25 0.0625]. For the derivative filters in both spatial and temporal domain we used the simple difference. As a result, the optical flow was computed from *only three frames*, from coarse to fine. Initial flow estimates were obtained in a neighborhood of 3×3 (i.e., N=3) and the regularization parameter $\beta = 1$. Estimation errors were evaluated using the software [6] that computes the average angular error μ_e and its standard deviation σ_e. We only discuss flow estimated with a density of 100.

Our first test involved the sequence *New-Sinusoid1* introduced by Bab-Hadiashar and Suter [3]. This sequence (see Figure 2a) has spatial frequencies similar to *Sinusoid1* from [6] but has a central stationary square of 50 pixels, thus containing motion discontinuities. The correct flow for *New-Sinusoid1* is shown in Figure 2b. The robust motion estimation method described in [3] has errors in the range ($\mu_e = 1.51 - 2.82, \sigma_e = 5.86 - 8.82$).

Using VBDF estimation we obtained a remarkable decrease in errors to ($\mu_e = 0.57, \sigma_e = 5.2$) and the estimated motion has sharp boundaries (Figure 2c). In Figure 2d we show the angular error multiplied by 100 (white corresponds to large errors). These results were obtained with a 7×7 analysis window (i.e., M=7) and variable bandwidth mean shift applied across 5 scales. The noise variance used in BLS has been assumed to be $\sigma^2 = 0.08$, equal to that of the quantization noise. An average number of 3 mean shift iterations per scale per window were executed.

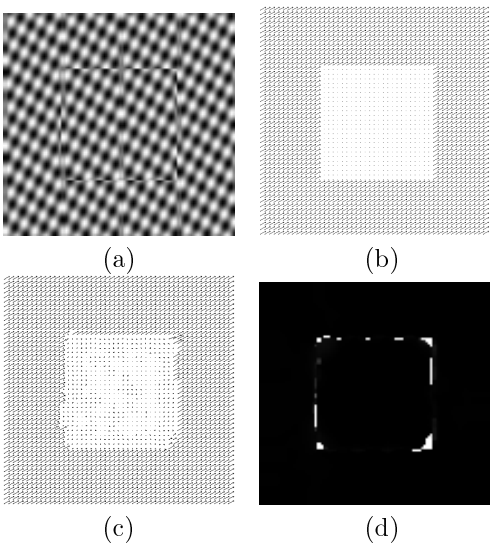


Figure 2: *New-Sinusoid1* sequence. (a) Frame 9. (b) Correct flow. (c) VBDF flow ($\mu_e = 0.57, \sigma_e = 5.2$). (d) Error corresponding to the new algorithm.

We performed the second test on *Yosemite* sequence (Figure 3a). This synthetic sequence contains many challenges, including multiple motions and aliasing. Numerous results have been reported on *Yosemite* involving either the complete sequence, or the partial sequence, with the sky and clouds discarded. For the complete sequence, our algorithm resulted in ($\mu_e = 4.25, \sigma_e = 7.82$) for the middle frame. The estimated flow is shown in Figure 4a and Figure 4b presents the angular error. The best results reported yet on complete *Yosemite* at full density were obtained by Memin and Perez [28] ($\mu_e = 5.38, \sigma_e = 7.73$), Alvarez, Weickert and Sanchez [2] ($\mu_e = 5.53, \sigma_e = 7.40$) and Liu *et al.* [25] ($\mu_e = 7.52, \sigma_e = 13.72$).

In Figure 5 we show a fusion example for Yosemite corresponding to the location (49,13) at the top of the image pyramid. This location is situated at the border between the sky and mountain. The initial location of the mode detection algorithm is marked by a large dot. The VBDF ellipse is drawn with a thick line.

For the skyless *Yosemite* we obtained ($\mu_e = 1.55, \sigma_e = 1.65$). According to our knowledge, the best results reported for the skyless sequence were obtained by Memin and Perez [29] ($\mu_e = 1.58, \sigma_e = 1.21$), Bab-Hadiashar and Suter [3] ($\mu_e = 1.97, \sigma_e = 1.96$) , and Lai and Vemuri [22] ($\mu_e = 1.99, \sigma_e = 1.41$). Recently, Farneback [13] reported errors of ($\mu_e = 1.14, \sigma_e = 2.14$), by processing the entire Yosemite sequence at once. This, however, involves large computational effort, memory space, and delay in response.

In comparison to the techniques from above, our method is simpler, easy to implement, and efficient, being based on the detection of the most significant mode of the density of some initial estimates. For *Yosemite* we used a 15×15 analysis window and $\sigma^2 = 0.08$. In addition, the distances between the initial flow vectors were weighted according to the intensity difference between the corresponding image pixels by a Gaussian kernel of standard deviation equal to 12. This assured that we grouped together flow vectors similar in direction and magnitude and coming from locations with similar intensity.

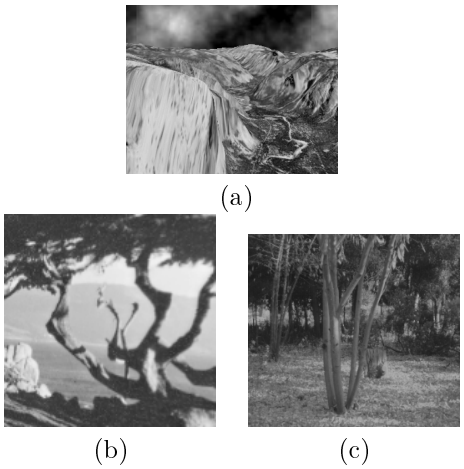


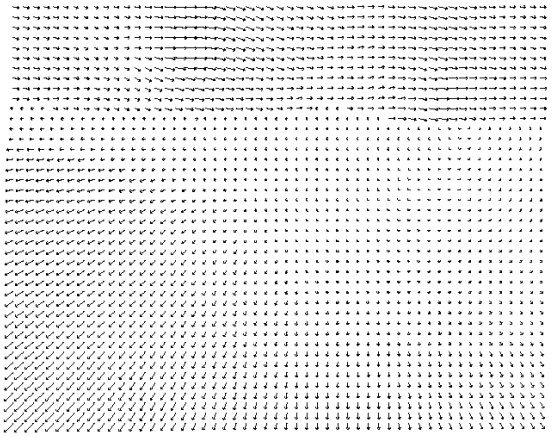
Figure 3: Test sequences (a) *Yosemite*, 9th frame. (b) *Tree* Translating/Diverging, 20th frame. (c) SRI Tree, 10th frame.

For the *Translating Tree* we obtained ($\mu_e = 0.19, \sigma_e = 0.17$) (see Figure 6a). The closest result is that of Lai and Vemuri [22] ($\mu_e = 0.40, \sigma_e = 0.28$). Finally, for the *Diverging Tree* (Figure 6b) our approach resulted in ($\mu_e = 1.10, \sigma_e = 0.73$), a result which compares favorably to the best available ($\mu_e = 1.34, \sigma_e = 1.05$), again from Lai and Vemuri [22]. Resulting flow for the SRI sequence is presented in Figure 7. Observe the sharp flow boundaries. The same parameters as in Yosemite were used for these sequences, but without intensity weighting.

7 Discussions

This paper introduced the VBDF estimator as a powerful tool for information fusion based on adaptive density estimation. We showed the ability of the new estimator to deal with multiple source models and to handle cross-correlation in a consistent way. We compared the VBDF framework with the BLUE fusion and Covariance Intersection and showed that the new estimator can be used to construct a very effective motion computation algorithm.

Finally, we underline the importance of mode tracking accross scales, for the detection of the most sigificant mode of a density function. In the context of motion estimation, the most significant mode corresponds to the most relevant motion in the local neighborhood. The same concepts can be naturally extended to other vision domains such as stereo, tracking, or robot localization.



(a)

(b)

Figure 4: Results for *Yosemite*. (a) Flow obtained by our method, including the sky, ($\mu_e = 4.25, \sigma_e = 7.82$). (b) Error corresponding to the new algorithm

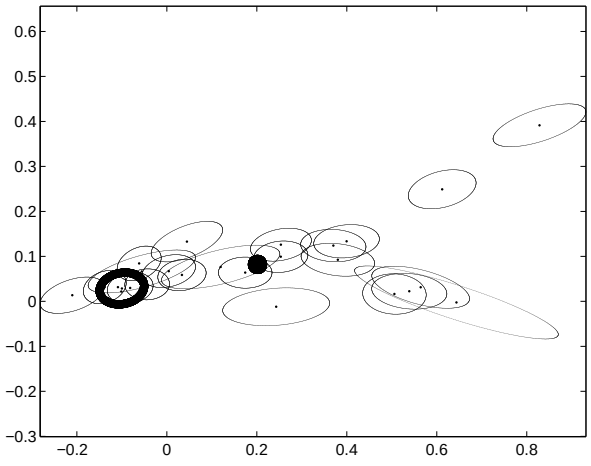
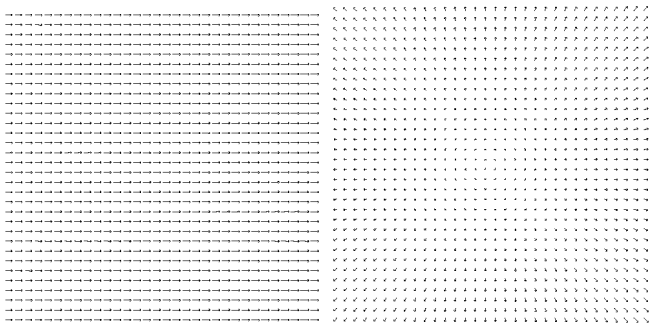


Figure 5: Ellipses with 95% confidence representing initial flow estimates for the location (49,13) of the top level of Yosemite pyramid. A window of $M = 5$ has been used to collect 25 initial estimates. The starting point of the algorithm is represented by a large dot in the center, while the VBDF estimate is drawn with a thick line.



(a) (b)

Figure 6: Results for *Tree*. (a) Flow for Translating sequence, ($\mu_e = 0.19, \sigma_e = 0.17$). (b) Flow for Diverging sequence, ($\mu_e = 1.10, \sigma_e = 0.73$).

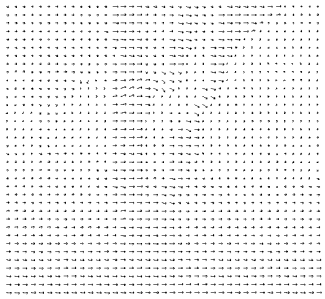


Figure 7: Results for *SRI*. Flow computed by our method.

Acknowledgments

I would like to thank Visvanathan Ramesh from Siemens Corporate Research and Peter Meer from Rutgers University for valuable discussions on this work. The comments of the anonymous reviewers were very helpful in refining the paper. Preliminary results were presented in the IEEE Workshop on Motion and Video Computing held in December 2002 in Orlando, Florida.

References

[1] J. Aggarwal and N. Nandhakumar, "On the computation of motion from sequences of images - A review," *Proceedings of IEEE*, vol. 76, pp. 917–935, 1988.

[2] L. Alvarez, J. Weickert, and X. Sanchez, "Reliable estimation of dense optical flow fields with large displacements," *Intl. J. of Computer Vision*, vol. 91, no. 1, pp. 41–56, 2000.

[3] A. Bab-Hadiashar and D. Suter, "Robust optical flow computation," *Intl. J. of Computer Vision*, vol. 29, no. 1, pp. 59–77, 1998.

[4] Y. Bar-Shalom and L. Campo, "The effect of the common process noise on the two-sensor fused track covariance,"

IEEE Trans. Aero. Elect. Syst., vol. AES-22, no. 22, pp. 803–805, 1986.

[5] Y. Bar-Shalom and T. Fortmann, *Tracking and Data Association*. Academic Press, 1988.

[6] J. Barron, D. Fleet, and S. Beuchemin, "Performance of optical flow techniques," *Intl. J. of Computer Vision*, vol. 12, no. 1, pp. 43–77, 1994.

[7] M. Black and P. Anandan, "The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields," *Computer Vision and Image Understanding*, vol. 63, no. 1, pp. 75–104, 1996.

[8] L. Breiman, W. Meisel, and E. Purcell, "Variable kernel estimates of multivariate densities," *Technometrics*, vol. 19, no. 2, pp. 135–144, 1977.

[9] L. Chen, P. Arambel, and R. Mehra, "Estimation under unknown correlation: Covariance intersection revisited," *IEEE Trans. Automatic Control*, vol. 47, no. 11, pp. 1879–1882, 2002.

[10] C. Chong and S. Mori, "Convex combination and covariance intersection algorithms in distributed fusion," in *Proc. of 4th Intl. Conf. on Information Fusion,* Montreal, Canada, 2001.

[11] D. Comaniciu, "An algorithm for data-driven bandwidth selection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 25, no. 2, pp. 603–619, 2002.

[12] D. Comaniciu, V. Ramesh, and P. Meer, "The variable bandwidth mean shift and data-driven scale selection," in *Proc. 8th Intl. Conf. on Computer Vision,* Vancouver, Canada, volume I, July 2001, pp. 438–445.

[13] G. Farnebäck, "Very high accuracy velocity estimation using orientation tensors, parametric motion, and simultaneous segmentation of the motion field," in *Proc. 8th Intl. Conf. on Computer Vision,* Vancouver, Canada, volume I, 2001, pp. 171–177.

[14] T. Gevers, "Robust histogram construction from color invariants," in *Proc. Intl. Conf. on Computer Vision,* Vancouver, Canada, volume I, July 2001, pp. 615–620.

[15] G. Golub and U. von Matt, "Tikhonov regularization for large scale problems," in *Scientific Computing, G.H. Golub et. al. eds.*, 1997, pp. 3–26.

[16] D. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proceedings of IEEE: Special Issue on Data Fusion*, vol. 85, no. 1, pp. 6–23, 1997.

[17] A. Hoerl and R. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.

[18] S. Julier and J. Uhlmannn, "A non-divergent estimation algorithm in the presence of unknown correlations," in *Proc. American Control Conf.,* Alberqueque, NM, 1997.

[19] M. Kam, X. Zhu, and P. Kalata, "Sensor fusion for mobile robot navigation," *Proceedings of IEEE: Special Issue on Data Fusion*, vol. 85, no. 1, pp. 108–119, 1997.

[20] G. Kamberova, R. Mandelbaum, M. Mintz, and R. Bajcsy, "Decision-theoretic approach to robust fusion of location data," *J. Franklin Institute*, vol. 336, no. 2, pp. 269–284, 1999.

[21] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, pp. 226–239, 1998.

[22] S. Lai and B. Vemuri, "Reliable and efficient computation of optical flow," *Intl. J. of Computer Vision*, vol. 29, no. 2, pp. 87–105, 1998.

[23] X. Li, Y. Zhu, and C. Han, "Unified optimal linear estimation fusion - part i: Unified models and fusion rules," in *Proc. of 3rd Intl. Conf. on Information Fusion,* Paris, *France*, 2000.

[24] M. Liggins, C. Chong, I. Kadar, M. Alford, V. Vannicolla, and S. Thomopoulos, "Distributed fusion architectures and algorithms for target tracking," *Proceedings of IEEE: Special Issue on Data Fusion*, vol. 85, no. 1, pp. 95–107, 1997.

[25] H. Liu, T. Hong, M. Herman, and R. Chellappa, "A general motion model and spatio-temporal filters for computing the optical flow," *Intl. J. of Computer Vision*, vol. 22, no. 2, pp. 141–172, 1997.

[26] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. DARPA Imaging and Understanding Workshop*, 1981, pp. 121–130.

[27] D. W. Marquardt, "Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation," *Technometrics*, vol. 12, no. 3, pp. 591–612, 1970.

[28] E. Memin and P. Perez, "Optical flow estimation and object segmentation with robust techniques," *IEEE Trans. Image Process.*, vol. 7, no. 5, pp. 703–719, 1998.

[29] E. Memin and P. Perez, "Hierarchical estimation and segmentation of dense motion fields," *Intl. J. of Computer Vision*, vol. 46, no. 2, pp. 129–155, 2002.

[30] A. Mitiche and P. Bouthemy, "Computation of image motion: A synopsis of current problems and methods," *Intl. J. of Computer Vision*, vol. 19, no. 1, pp. 29–55, 1996.

[31] H. Nagel, "Image sequence evaluation: 30 years and still going strong," in *Proc. Intl. Conf. on Pattern Recognition,* Barcelona, Spain, volume I, September 2000, pp. 149–158.

[32] E. Simoncelli, "Bayesian multi-scale differential optical flow," in B. Jahne, H. Haussecker, and P. Geissler, editors, *Handbook of Computer Vision and Applications*, 1998, pp. 297–422.

[33] E. Simoncelli, E. Adelson, and D. Heeger, "Probability distributions of optical flow," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition,* Hawaii, June 1991, pp. 310–315.

[34] A. Singh and P. Allen, "Image-flow computation: An estimation-theoretic framework and a unified perspective," *CVGIP: Image Undestanding*, vol. 56, no. 2, pp. 152–177, 1992.

[35] E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision*. Prentice Hall, 1998.

[36] J. Weber and J. Malik, "Robust computation of optical flow in a multi-scale differential framework," *Intl. J. of Computer Vision*, vol. 2, pp. 5–19, 1994.