Mechanisms for Isolating Component Patterns in the Sequential Analysis of Multiple Motion

Peter J. Burt, Rajesh Hingorani and Raymond J. Kolczynski
David Sarnoff Research Center
Subsidiary of SRI International
Princeton NJ, 08543-5300

Abstract

Pyramid techniques are commonly used to provide computational efficiency in the analysis of image motion. But these techniques can play an even more important role in the analysis of multiple motion, where, for example, a transparent pattern moves in front of a differently moving background pattern. The pyramid framework then separates motion components based on their spatial and temporal frequency characteristics so that each can be estimated independently of the others. This property is key to recently proposed selective stabilization algorithms for the sequential analysis of multiple motion and for the detection of moving objects from a moving platform.

In this paper we determine the conditions for component selection. Results can provide important guidance in practical applications of motion analysis.

1 Introduction

Most approaches to image motion analysis adopt a single motion model: it is assumed that within any sufficiently small region an image can be described as a single pattern undergoing uniform motion. This model fails wherever two differently moving patterns appear superimposed in the image, as where a transparent surface or a pattern of light and shadow moves over another surface, or where the analysis region falls on the boundary between a moving object and its background.

Recently several investigators have considered the problem of estimating the motion of two or more differently moving patterns within a given analysis region. Girod and Kuo [1] detect multiple peaks in the cross correlation function between successive image frames. Shizawa and Mase [2] use a direct method to estimate the translation parameters of two patterns simultaneously. We have introduced a sequential approach to the estimation of multiple motion in which an estimator based on a single motion model is applied repeatedly to the image region in order to measure the motion of one pattern component at a time [3],[4].

Success of the sequential approach to multiple motion analysis is due in large part to a "selection property" of the single component motion estimator that is used: when applied within a region that contains two differently moving patterns, this estimator tends to "lock onto" and report just one of these motions. Component selection is due to spatial and temporal filtering characteristics of the motion estimator when implemented within a pyramid structure.

In this paper we identify the mechanisms underlying component selection and show how spatial (pyramid level) and temporal (frame rate) parameters can best be controlled in order to isolate moving components of interest.

2 Problem Statement

Suppose that analysis is to be carried out within a region R that contains two moving patterns and that each is undergoing simple translation. Let P and Q be the patterns and let p and q be their velocities. Then

$$I(x,t) = P(x - pt) + Q(x - qt). \quad (1)$$

The objective of motion analysis is to estimate p and q from the observed image I.

For R sufficiently restricted in space and time, Eq. 1 can be used to model a variety of motion configurations that commonly occur in natural scenes, including transparency and motion at object boundaries [4]. An example of two component motion is shown in Figure 1a. Here a small object Q moves in front of an extended background P. The paths followed by a number of points on the patterns are indicated by...
arrows in the diagram. Object \( Q \) is moving slightly faster than background \( P \).

3 Selective Stabilization

The technique of selective stabilization obtains an estimate of a pattern’s motion by finding an image transformation that effectively cancels that motion within the analysis region \( R \) and over time \( T \). Here we assume the transformation is based on a uniform translation model.

Stabilized Image

Let \( I' \) be a transformed image obtained from the original by compensating for translation \( \mathbf{v} \):

\[
I'(x,t;t_0) = I(x - (t_0 - t)\mathbf{v}, t). \tag{2}
\]

If \( I \) contains a pattern moving at velocity \( \mathbf{v} \) then that pattern is rendered stationary in \( I' \). Figure 1b shows such a transformation applied to the example in Figure 1a with \( \mathbf{v} \approx \mathbf{p} \) to stabilize background pattern \( P \).

Incremental Motion Estimator

Motion estimation is based on a single component model: it is assumed that the image contains just one pattern within \( R \), and that this pattern moves at a uniform velocity \( \mathbf{v} \). Based on this model the estimate \( \hat{\mathbf{v}} \) is defined as that shift that minimizes the squared difference between successive image frames. For small shifts this satisfies the differential equations (e.g., [5]):

\[
\begin{align*}
\hat{v}_x \int_R \left( \frac{\partial I}{\partial x} \right)^2 + \hat{v}_y \int_R \frac{\partial I}{\partial x} \frac{\partial I}{\partial y} &= - \int_R \frac{\partial I}{\partial x} \frac{\partial I}{\partial t} \\
\hat{v}_y \int_R \frac{\partial I}{\partial x} \frac{\partial I}{\partial y} + \hat{v}_y \int_R \left( \frac{\partial I}{\partial y} \right)^2 &= - \int_R \frac{\partial I}{\partial y} \frac{\partial I}{\partial t}
\end{align*} \tag{3}
\]

In practice these equations are computed for pairs of discretely sampled image frames. Errors in the estimate are reduced systematically through frame alignment in an iterative process that alternates between motion estimation and stabilization steps.

Let \( \hat{\mathbf{v}}_k \) be the estimate obtained at the \( k^{th} \) iteration. In step \( k + 1 \) the image \( I \) is transformed according to Eq. 2 to form \( I_{k+1} \). Equation 3 is then evaluated for \( I_{k+1} \) to obtain an estimate for the incremental motion \( \Delta \hat{\mathbf{v}}_{k+1} \). This is added to the previous estimate to obtain the refined motion estimate: \( \hat{\mathbf{v}}_{k+1} = \hat{\mathbf{v}}_k + \Delta \hat{\mathbf{v}}_{k+1} \). These steps are repeated until the desired accuracy is achieved.

Pyramid Implementation

The above computations are implemented within a Laplacian, band-pass, pyramid structure. (A Gaussian, low-pass pyramid can also be used.) Let \( l_k \) be the \( k^{th} \) pyramid level. \( l_0 \) represents the highest resolution (highest frequency) band of the image, while \( l_1, l_2, \ldots \) represent successively lower bands, spaced at octave intervals.

Let \( \ell_k \) be the pyramid level at which the incremental motion estimator is evaluated at step \( k \). Thus the motion estimate at the \( k^{th} \) iteration, Eq. 3, is based on \( I_{\ell_k} \). The sequence of refinement steps starts at a low resolution level of the pyramid, then moves to progressively higher resolution levels.

Two Selection Modes

Motion estimation through selective stabilization is achieved by alternating between image stabilization and incremental motion estimation steps. As noted above, this procedure has the important property that it will tend to “lock onto” and report the motion of one component even when two are present within the analysis region. Thus in Figure 1b it is pattern \( P \) that has been selected and stabilized.

The estimator exhibits two selection modes: it may pool, or average, the component motions, or it may select one. Generally speaking, pooling occurs when the
computations are preformed at a low resolution pyramid level, while selection occurs when computations are performed at high resolution level.

4 Frequency Domain Interpretation

The tendency of the motion estimator to pool motion components or to select one component depends on the difference between the frame-to-frame displacement of the two components, and on whether that difference can be resolved at the pyramid level at which computations are performed. These relationships are best described in the frequency domain.

Component Spectra

Let the Fourier transform of a function $F(x, t)$ be $F(u, w)$. The Fourier transform of a pattern undergoing constant motion is a tilted plane in the frequency domain (e.g., [0]). For the two component image, Eq. 1, it is described by

$$I(u, w) = P(u)\delta(w + pu) + Q(u)\delta(w + qu). \quad (4)$$

As shown in Figure 2, the two moving patterns become tilted lines in one spatial dimension, with slopes equal to $-p_x$ and $-q_x$ respectively. Here the line corresponding to $P$ is made bolder than that for $Q$ to suggest greater energy, as in the configuration in Figure 1a where $P$ is the extended background and $Q$ is a small object.

Motion stabilization has the effect of reducing the velocities of the component patterns, and thereby rotating their spectra, $P$ and $Q$, towards the $w = 0$ axis. The spectrum of the component that is fully stabilized becomes a horizontal plane while the slope of the spectrum of the second component is reduced to the difference between component velocities.

Selection Band

The computation used in estimating incremental motion, Eq. 1, has the effect of partitioning the spatial/temporal frequency domain into a set of bandpass zones, also shown in Figure 2.

The Laplacian pyramid divides spatial frequencies into octave wide bands. These are shown in the figure bounded by vertical lines. The temporal difference and spatial integration operations in Eq. 3 act as a temporal bandpass filter. It is shown in the appendix that this is bounded by $w = 0$ and $|w| < \frac{1}{2\tau}$, as shown by horizontal lines in the figure. A selection band is formed by the intersection of the temporal pass band and the spatial pass band corresponding to the pyramid level used at step $k$. This band is shown as the shaded zone in Figure 2.

This frequency domain diagram indicates the basic mechanisms that underlie component selection in the selective stabilization algorithm. If only one of the component patterns $P$ or $Q$ has energy in the selection band at a given iteration of the algorithm, as in Figure 2b, then that component will dominate the estimates and will be selected by the computation. On the other hand, if both components have energy in the band, as in Figure 2a, then the motion estimate will be an average of the two component motions.

Coarse-Fine Stabilization

The selective stabilization algorithm causes systematic changes in the frequency domain characteristics of both the computation and the signal. As analysis is performed at progressively higher resolution levels of the pyramid, the selection band moves towards higher spatial frequencies in the diagram. At the same time, image stabilization causes the spectra of the component patterns to be rotated towards $w = 0$. These two processes systematically create conditions in which only one motion component falls in the selection band so that selection can be achieved.

A typical scenario is suggested in Figure 2. At first both $P$ and $Q$ have velocities that are large compared to their difference. These fall in the same selection band when $\ell = 3$, as shown in Figure 2a. The resulting motion estimate is an average of the component velocities. Stabilization tilts the spectra of both components towards zero, and the selection zone is shifted as the computation is moved to pyramid level $\ell = 1$, Figure 2b. Now only $P$ falls within the selection band, so subsequent motion estimates approximate its velocity. Figure 2c shows near full stabilization with $P$ falling near $w = 0$ and $Q$ falling outside the selection band at $\ell = 0$.

Extended Time Base

Once pyramid level $\ell = 0$ has been reached, further refinement, if required, must be achieved in the temporal domain, by extending the temporal sample interval, $\tau$. In practice this is done by performing analysis between non-consecutive frames in an image sequence. Increasing $\tau$ has the effect of shrinking the temporal dimensions of the selection band and shifting it towards $w = 0$.

Critical Velocity Difference

In accordance with the above account we anticipate that selection occurs when the velocity difference
Figure 2: Frequency domain representation of two component motion and the mechanisms underlying component selection.

\[ \text{Average motion obtained at low spatial frequency.} \]

\[ \text{Component selection obtained at medium frequency.} \]

\[ \text{Component stabilization obtained at high frequency.} \]

\[ |q - p| \text{ is sufficiently large, and the frequency band } \ell \text{ is sufficiently high, that the spectra of the two components cannot both fall in the selection band. For a given level } \ell \text{ the lower spatial frequency limit of the selection band is } \frac{u}{2^\ell}. \text{ The upper limit of the temporal band is roughly } w = \frac{v}{2^\ell}. \text{ Thus for a given pyramid level, selection should occur when the velocity difference exceeds a critical value:} \]

\[ |q - p| \geq \frac{|v|}{|u|} = \frac{2^{\ell+1}s}{\ell}. \]  

Or, for a given velocity difference selection should occur when the pyramid level is less than a critical value:

\[ \ell < \log_2 \frac{\tau}{s}|q - p| - 1. \]  

These equations express a qualitative relationship between velocity difference and pyramid level for selection to occur. The critical velocity difference will depend in addition, on the distribution of spectral energy within the selection band.

5 Detecting the Second Motion

Once one component of motion has been found it remains to determine if a second component exists in the analysis region. Again, suppose that the image can be described in terms of two patterns, \( P \) and \( Q \), Eq. 1, and that it is \( P \) that has been stabilized. The first step towards detecting \( Q \) is to remove pattern \( P \) from the image sequence. This is accomplished after stabilization by forming the difference between successive frames of \( I^v (v \approx p) \).

Difference Image

Let \( D^v(x, t; t_0, \tau) \) be a difference image formed between image frames separated in time by \( \tau \) in the stabilized image \( I^v \):

\[ D^v(x, t; t_0, \tau) = I^v(x, t; t_0) - I^v(x, t - \tau; t_0). \]  

(Note: \( \tau \) need not be the same here as in the estimation of \( p \). After stabilization a larger time base may be required in detecting \( Q \).)

A temporal difference for an image containing patterns undergoing uniform translation can be expressed in terms of spatial differences of these component patterns. Let \( \Delta P(x; d) \) be a difference of pattern \( P \) with itself shifted by \( d \):

\[ \Delta P(x; d) = P(x) - P(x - d). \]  

(8)

Pattern \( P \) moves at velocity \( p - v \) in \( I^v \) so it is shifted by \( (p - v)\tau \) in \( D^v \). Thus combining Eq. 1, 2, 7, and 8:

\[ D^v(x, t; t_0, \tau) = \Delta P(x - (p - v)t - v t_0; (p - v)\tau) + \Delta Q(x - (q - v)t - v t_0; (q - v)\tau). \]  

(9)

Note that \( \Delta P \) is zero when \( v = p \). Thus when \( P \) is perfectly stabilized, the difference sequence consists just of pattern \( \Delta Q(x; (q - p)\tau) \) moving at velocity \( q - p \).

Estimating Motion \( q \)

A motion estimate for component \( Q \) can be obtained with a second application of the selective stabilization algorithm, this time to \( D^v \). If \( v \approx p \) then pattern \( P \) will be largely removed, and the algorithm will converge to pattern \( \Delta Q \). A difference image with \( P \) removed is shown in Figure 1c.

Once an estimate has been obtained for \( q \) then it will sometimes be desirable to repeat the analysis in
Demonstrations

In this section we present two examples to show the selection property of the motion estimation algorithm.

Component Selection in Transparency

The first example uses an artificially constructed image sequence to simulate transparent motion. Patterns $P$ and $Q$ were constructed separately, then summed to form the first image frame. The patterns were then shifted by controlled amounts $p$ and $q$ and summed again to form the second image, in accordance with Eq. 1. Both component patterns were formed as random noise that was shaped in the spatial frequency domain to have spectral energy that falls inversely with frequency. Such a “1/f” spectrum is typical of natural images.

A sequence of experiments was performed in which the velocity difference $q - p$ was systematically varied, and the selective stabilization algorithm was used to obtain an estimate of motion. Results are shown in Figure 4. Each data point represents the result of one experiment. In all experiments the $z$ components of velocity for both patterns were equal $p = qz$, while the $y$ components were opposite $p = -qy$. The $x$ components were held constant at 5.5 pixels for all experiments, while the $y$ components were increased from 0 to 16 pixels in steps of 0.5 from experiment to experiment. Thus the magnitude of the velocity difference $|q - p|$ is twice the $y$ velocity of $P$, $2p$. (Here we take the spatial and temporal sample distances $s$ and $T$ to be both unity.)

In an initial sequence of experiments the initial velocity estimate, $v_0$, was taken to be zero and the algorithm was run coarse-fine starting at level $l = 3$. At each level the incremental estimator was iterated until a stable velocity estimate was obtained. This estimate was used as the starting point for estimation at the next level. The $y$ components of the resulting velocity estimates at levels $l = 3, 2, 1$ are shown in Figure 4. Note that when the velocity difference is small the estimate is roughly zero, representing an average of the two component velocities. Above a transition the estimates represent one of the other component. A second set of experiments was run with the initial velocity $v_0$ perturbed in order to pick up the second stable point. The transition between average and selection occurs roughly at velocity differences of 4, 8 and 16 for the three levels. This is in good qualitative agreement with Eq. 5.

Moving Object Detection

The technique of selective stabilization can be an effective means for detecting moving objects from a
moving platform [3]. A basic three step detection procedure was suggested schematically in Figure 1. Here we illustrate the technique with an example in which a helicopter is observed from a second helicopter while flying low over hilly terrain. This example extends previously reported results in two respects: the size of the analysis region is controlled so that complex background motions can be handled, and the position of the analysis region is controlled in order to implement a scanning strategy.

One frame of the test sequence is shown in Figure 5a. The moving helicopter is roughly in the center of the image, but is difficult to pick out in a single image frame. Because the terrain is quite rugged, and parallax is significant, stabilization applied to the full image does not cancel background motion sufficiently to detect the moving object. However, if the same algorithm is performed within a smaller region, as in Figure 5b, detection is possible. Here the computed change energy has been inserted into the original image frame to show where the analysis has been performed. The effect of reducing the size of the analysis region is to simplify the background motion until it can be adequately represented by the (affine) model used in stabilization. At the same time the parameters of analysis conform roughly to the guidelines outlined in the Section V. The analysis region, $R$, is large compared to the object to be detected, and the pyramid level $\ell = 0$ and time interval used in forming the difference, $\tau = 2$, are reasonably well matched to the motion of the object.

Since the stabilization algorithm is not being performed over the full image in this example, it is necessary to move the analysis region systematically over the scene to detect moving objects wherever they may occur.

7 Summary

Important insights into the role of pyramid based processing in the selective stabilization algorithm can be obtained by analyzing properties of the algorithm in the spatial/temporal frequency domain. In particular, the "component selection property" that is key to its use in sequential analysis of multiple motion, can be attributed to band-pass filter characteristics of the incremental motion estimator, Eq. 3. The computations performed in this estimator create a "selection band" bounded in spatial frequency by the pyramid level $\ell$ used in the computation, and in temporal frequency by $w = 0$ and roughly $|w| = \frac{1}{2\tau}$. (The tem-
poral limit applies to image patterns with broad-band spectra. This assumption can be violated by image patterns that are dominated by periodic structures.)

Selection occurs when the spectrum of just one of the component patterns falls in the selection band, while averaging occurs when the spectra of both patterns fall in that band.

In the selective stabilization algorithm, estimation steps are alternated with stabilizations steps. Together these processes systematically shift the selection band and the component pattern spectra until conditions for selection occur, until just one component falls in the selection band.

Appendix: Motion Estimator in the Frequency Domain

The component selection property of the selective stabilization algorithm can be attributed to filter characteristics of the incremental motion estimator, Eq. 3. To simplify analysis consider motion estimation in one spatial dimension. When the differential terms in Eq. 3 are replaced by discrete differences, the velocity estimate becomes

$$\hat{v} = \frac{\sum_{x} \frac{\Delta I}{\Delta x}}{\sum_{y} \frac{\Delta I}{\Delta y}}$$  (A1)

Suppose that the image consists of just one component $P$, moving at constant velocity $v$. Then the temporal difference can be expressed as a spatial difference. For $\Delta x = s$ and $\Delta t = r$, and at time $t = 0$:

$$\hat{v} = \frac{s}{r} \frac{\int R(P(x + \frac{s}{2}) - P(x - \frac{s}{2}))(P(x) - P(x - pr))}{\int R(P(x + \frac{s}{2}) - P(x - \frac{s}{2}))^2}$$  (A2)

This estimate can be stated in terms of integrals over spatial frequency rather than over space. From the power theorem, and the relation $w = pu$:

$$\hat{v} = -s \frac{\int \sin(\pi u) \sin(2\pi w)|P(u)|^2 du}{\int \sin(\pi u)^2 |P(u)|^2 du}.$$  (A3)

If analysis is performed at pyramid level $l$, then the power spectrum $|P_u(u)|^2$ is confined to a corresponding frequency band. The term $\sin(2\pi w)$ may be interpreted roughly as a temporal filter with pass band between $w = 0$ and $|w| = \frac{1}{2}$. Spectral energy sums incoherently for $|w| > \frac{1}{2}$.

Acknowledgements

Many individuals have contributed to the ideas and results presented here. These include P. Anandan, James R. Bergen, Jeffrey Lubin, and Keith Hanna of David Sarnoff Research Center, and Shmuel Peleg of the Hebrew University.

References


