Decoupling Fourier Components of Dynamic Image Sequences: A Theory of Signal Separation, Image Segmentation, and Optical Flow Estimation

David Vernon

Department of Computer Science
National University of Ireland, Maynooth
Ireland

Abstract. This paper presents a new Fourier-based approach to the separation or decoupling of m additive images from a time-sequence of the sum of these images where at least $m-1$ images are translating with distinct and unique velocity. A closed-form solution is presented for the case where $m = 2$. A generalization is then presented which extends the theory to embrace situations where the images are not additive but are, instead, formed by the superposition of an occluding object or objects on an occluded background. That is, the approach is generalized to effect a model-free segmentation of objects undergoing translatory fronto-parallel motion in dynamic image sequences. Object velocities of one pixel per frame are sufficient to guarantee segmentation.
We also show how the technique can be applied on a local basis to compute a dense instantaneous optical flow field for the image sequence, even in relatively featureless regions. The technique is evaluated using Otte's and Nagel's benchmark image sequence, for which ground-truth data is available, and results comparable with the ground-truth flow field are achieved. RMS errors of velocity magnitude and direction are computed and reported.

1 Introduction

Traditional approaches to segmentation typically exploit one of two broad approaches. These are (a) boundary detection, which depends on the detection of spatial intensity discontinuities (using first or second order gradient techniques) and their aggregation into contour-based object descriptions, and (b) region-growing, which depends on the identification of local regions that satisfy some regional similarity predicate (see [1] and [2] for an overview).

Equally, the measurement of object velocity in images normally exploits one of two primary techniques. The first involves the computation of the spatio-temporal gradient, differentiating the (filtered or unfiltered) image sequence with respect to time and subsequently computing the optical flow field (*e.g.* [3]). The second involves the segmentation of the object or feature in question using either region-based gradient (first or second order) filtering and analysis followed

either by the computation of the optical flow field or by identification object correspondence, typically by matching contour or region primitives (*e.g.* [4]). A third, lesser-used, approach exploits the regularity in spatiotemporal-frequency representations of the image, such as the spatiotemporal Fourier Transform Domain, resulting from certain types of image motion, such as fronto-parallel translation [5–13]. Comparisons of the many variations of these approaches and the relationship between them can be found in [14–17].

In this paper, we develop the Fourier analysis of dynamic image sequences and we present a new approach (i) to the separation of additive signals (such as in the case of reflections superimposed on optical images), (ii) to the conventional segmentation of occluding and occluded objects, and (iii) to the estimation of instantaneous velocity (including estimation of optical flow). This is effected by processing the resultant Fourier phasors derived from the FFT of the composite image and by resolving each of them into the individual Fourier components corresponding to each object.

2 Theory

2.1 The Additive Model: Signal Separation

First, we show that it is possible to separate (or recover) m additive images, given only a time sequence of the sum of these images and given the assumption that at least $m-1$ images are translating, each with a distinct and unique velocity. That is, given a composite image $\psi_{t_j}(x,y)$ at time t_j in a temporally-ordered image sequence:

$$\psi_{t_j}(x,y) = \sum_{i=1}^{m} \psi^i_{t_j}(x,y) \tag{1}$$

where $\psi^i_{t_j}(x,y)$ is the (unknown) i^{th} additive component image at time t_j, and assuming

$$\psi^i_{t_j}(x,y) = \psi^i_{t_0}(x - v^i_x j\delta t, y - v^i_y j\delta t) \tag{2}$$

where (v^i_x, v^i_y) is the spatial velocity of the i^{th} component image and δt is the incremental time interval, it is possible to recover, or compute, each individual image $\psi^i_{t_0}(x,y), \forall i$.

This is accomplished by computing the Fourier transform $\mathsf{F}_{t_j}(k_x,k_y)$ of each image in the image sequence

$$\mathsf{F}_{t_j}(k_x,k_y) = \mathcal{F}\left(\psi_{t_j}(x,y)\right) \tag{3}$$

and then by decoupling the resultant (composite) Fourier component at each spatial frequency into the m individual Fourier components $\mathsf{F}^i_{t_0}(k_x,k_y)$ at time t_0. The required individual images are then computed using the inverse Fourier transform

$$\psi^i_{t_0}(x,y) = \mathcal{F}^{-1}\left(\mathsf{F}^i_{t_0}(k_x,k_y)\right) \tag{4}$$

In particular, the composite Fourier component at time t_j is a function of the m decoupled Fourier comonents $\mathsf{F}^i_{t_0}(k_x, k_y), 1 \leq i \leq m$ at time t_0 and of the incremental spatial frequency-dependent and velocity-dependent phase change $\Delta\Phi^i$ which results from the individual image translation where $\Delta\Phi^i$ is given by

$$\Delta\Phi^i = e^{-i(k_x v^i_x \delta t + k_y v^i_y \delta t)} \tag{5}$$

Specifically, we have

$$\mathsf{F}_{t_j}(k_x, k_y) = \sum_{i=1}^{m} \mathsf{F}^i_{t_0}(k_x, k_y)(\Delta\Phi^i)^j \tag{6}$$

If we have $i = m$ distinct individual images, equation 6 implies that we have $2m$ complex unknowns (*i.e.* $\mathsf{F}^i_{t_0}$ and $\Delta\Phi^{i^j}$) and consequently we can solve for these $2m$ unknowns if we have $2m$ constraints. These constraints are derived from equation 6 by making $j = 2m$ observations for F_{t_j} (*i.e.* by using $j = 2m$ composite images in the temporal sequence). That is, for a given spatial frequency (k_x, k_y), we observe the Fourier transform F_{t_j} at time $t_0, t_1, \ldots, t_j, \ldots, t_{2m}$ and solve these $2m$ simultaneous equations of degree $2m - 1$ in complex unknowns $\mathsf{F}^i_{t_0}$ and $\Delta\Phi^i$.

In the simplest non-trivial case, $m = 2$, there are two distinct objects. In this case, and dropping the (k_x, k_y) for the sake of brevity while remembering that we are dealing with complex values defined on a 2-D domain, we have

$$\mathsf{F}_{t_0} = \mathsf{F}^1_{t_0}(\Delta\Phi^1)^0 + \mathsf{F}^2_{t_0}(\Delta\Phi^2)^0 \tag{7}$$

$$\mathsf{F}_{t_1} = \mathsf{F}^1_{t_0}(\Delta\Phi^1)^1 + \mathsf{F}^2_{t_0}(\Delta\Phi^2)^1 \tag{8}$$

$$\mathsf{F}_{t_2} = \mathsf{F}^1_{t_0}(\Delta\Phi^1)^2 + \mathsf{F}^2_{t_0}(\Delta\Phi^2)^2 \tag{9}$$

$$\mathsf{F}_{t_3} = \mathsf{F}^1_{t_0}(\Delta\Phi^1)^3 + \mathsf{F}^2_{t_0}(\Delta\Phi^2)^3 \tag{10}$$

This set of four simultaneous equations has a closed-form solution [18]:

$$\Delta\Phi^1 = \frac{-b + \sqrt{z}}{2a} \tag{11}$$

$$\Delta\Phi^2 = \frac{-b - \sqrt{z}}{2a} \tag{12}$$

$$\mathsf{F}^1_{t_0} = \frac{\mathsf{F}_{t_0}\Delta\Phi^2 - \mathsf{F}_{t_1}}{(\Delta\Phi^2 - \Delta\Phi^1)} \tag{13}$$

$$\mathsf{F}^2_{t_0} = \frac{\mathsf{F}_{t_0}\Delta\Phi^1 - \mathsf{F}_{t_1}}{(\Delta\Phi^1 - \Delta\Phi^2)} \tag{14}$$

where $a = (\mathsf{F}_{t_1})^2 - \mathsf{F}_{t_0}\mathsf{F}_{t_2}$, $b = \mathsf{F}_{t_0}\mathsf{F}_{t_3} - \mathsf{F}_{t_1}\mathsf{F}_{t_2}$, $c = (\mathsf{F}_{t_2})^2 - \mathsf{F}_{t_1}\mathsf{F}_{t_3}$, and $z = b^2 - 4ac = \alpha + i\beta$.

Note, however, that the assignment of $\Delta\Phi^1$ rather than $\Delta\Phi^2$ in equation 11 and *vice versa* in equation 12 (and, hence the assignment of $\mathsf{F}^1_{t_0}$ and $\mathsf{F}^2_{t_0}$

in equations 13 and 14) is arbitrary and the alternative assignment is equally valid. Consequently, once $\mathsf{F}^1_{t_0}(k_x, k_y)$ and $\mathsf{F}^2_{t_0}(k_x, k_y)$ have been solved for all spatial frequencies k_x, k_y, it is still necessary to sort these sets of $\mathsf{F}^1_{t_0}(k_x, k_y)$ and $\mathsf{F}^2_{t_0}(k_x, k_y)$. That is, we only have at this point two sets of phase changes $\{\Delta\Phi^A\}$ and $\{\Delta\Phi^B\}$ and two corresponding sets of Fourier components $\{\mathsf{F}^A\}$ and $\{\mathsf{F}^B\}$. These sets need to be sorted into two new sets $\{\mathsf{F}^1\}$ and $\{\mathsf{F}^2\}$ corresponding to the two individual component images. In [18], we presented a method of doing this based on the regularity of the incremental phase change $\Delta\Phi$ as a function of frequency (k_x, k_y). Specifically, we have:

$$\begin{aligned}\Delta\Phi(k_x, k_y) &= e^{-i(v_x k_x \delta t + v_y k_y \delta t)} \\ &= e^{i\delta\phi(k_x, k_y)}\end{aligned} \tag{15}$$

For a given image i, (v^i_x, v^i_y) is constant. Thus a given image i will exhibit a phase change $\delta\phi(k_x, k_y)$:

$$\delta\phi(k_x, k_y) = -(v^i_x k_x \delta t + v^i_y k_y \delta t) \tag{16}$$

which will differ for each image i. Since we require $(v^i_x, v^i_y) \neq (v^j_x, v^j_y), i \neq j$, in order to sort the components of the two waves we simply need to identify the two velocities (v^1_x, v^1_y) and (v^2_x, v^2_y) which will, in turn, allow us to identify the corresponding expected phase change for images 1 and 2, respectively. Let these expected phase changes be denoted $\delta\phi^1_e(k_x, k_y)$ and $\delta\phi^2_e(k_x, k_y)$, respectively. Then we assign a component $\mathsf{F}^A(k_x, k_y)$ to image 1, *i.e.* we include it in $\{\mathsf{F}^1\}$, if $\left|\delta\phi^1_e - \delta\phi^A\right| < \left|\delta\phi^2_e - \delta\phi^A\right|$, otherwise we assign it to $\{\mathsf{F}^2\}$; $\mathsf{F}^B(k_x, k_y)$ is assigned to the other image.

It only remains, then, to identify the two velocities (v^1_x, v^1_y) and (v^2_x, v^2_y). We do this using a Hough transform. From equation 16 we have:

$$v_y = \frac{1}{k_y \delta t}\left(\delta\phi(k_x, k_y) - k_x v_x \delta t\right) \tag{17}$$

This equation represents a Hough transform in the two variables (v_x, v_y) which we solve for all $\delta\phi(k_x, k_y), k_x, k_y$, and v_x. Note that

$$\delta\phi(k_x, k_y) = \arctan(\Im\Delta\Phi(k_x, k_y), \Re\Delta\Phi(k_x, k_y)) \tag{18}$$

Local maxima in this Hough transform space represent the velocities which are exhibited by the frequency components. In this case, there are two velocity maxima, one corresponding to image 1 and the other to image 2. The location of these maxima give us (v^1_x, v^1_y) and (v^2_x, v^2_y) and, thus, we can proceed to sort the components.

Note that the Hough transform equation 17 becomes degenerate if $k_y = 0$ in which case we use an alternative re-arrangement as follows:

$$v_x = \frac{1}{k_x \delta t}\left(\delta\phi(k_x, k_y)\right) \tag{19}$$

2.2 The Generalized Occlusion Model: Signal Segmentation

The foregoing theory assumes that the individual images combine additively to form the resultant image. Although there are many situations or applications where this assumption is valid, *e.g.* the superposition of reflections on an image acquired through a transparent medium, there are important situations where it is not. In particular, the common situation where a moving object occludes a (possibly moving) background violates the assumption. Although the foregoing approach does in fact produce a reasonable segmentation of the (occluding) forground objects from the (partially occluded) background, the results are inevitably imperfect. To deal with this, we next present a generalization of the foregoing theory to deal explicitly with the decoupling or segmentation of (non-additive) occluding objects undergoing translatory fronto-parallel motion in dynamic image sequences.

Recall equations 7 to 10. At time t_0, the situation where occlusion obtains is still accurately represented by equation 7, *i.e.* the composite image is the sum of the occluded object (or image) and the background object (or image), assuming the background occluded image signal has a zero value wherever the occluding image is non-zero, *i.e.* assuming that there is nothing behind the occluding object. Of course, this is not correct in most instances and there is a non-zero occluded signal. Whilst this does not invalidate equation 7, provided we allow the assumption, it does then cause a problem with equation 8 since, at time t_1, when the object and/or the background object has moved, there will be a change in the spectral content of the image due to the appearance of visual information in the background which was previously occluded at time t_0 and to the disappearance of visual information which is now occluded at time t_1 (remember that the velocities of the occluded and the occluding object are different). We will call this revealed/hidden signal the *occluded residue* or simply the *residue*. Consequently, to render equation 8 accurate again we must include a new term F^3 to represent the residue. Equation 8 now becomes

$$\mathsf{F}_{t_1} = \mathsf{F}^1_{t_0}\left(\Delta\Phi^1\right)^1 + \mathsf{F}^2_{t_0}\left(\Delta\Phi^2\right)^1 + \mathsf{F}^3 \tag{20}$$

Similarly, at time t_2, we must again add a new residue term F^4 and, in addition, we must alter the phase of the previous residue F^3 to reflect the translation of the background. Consequently, we must alter the phase either by $\Delta\Phi^1$ or by $\Delta\Phi^2$ depending on whether $\mathsf{F}^1_{t_0}$ or $\mathsf{F}^2_{t_0}$ is the background. Since we don't know which is the case at this point, we choose, arbitrarily, $\mathsf{F}^2_{t_0}$ as the background occluded image and $\mathsf{F}^1_{t_0}$ as the foreground occluding image. Later we will address the resolution of this arbitrary choice. Equation 9 then becomes:

$$\mathsf{F}_{t_2} = \mathsf{F}^1_{t_0}\left(\Delta\Phi^1\right)^2 + \mathsf{F}^2_{t_0}\left(\Delta\Phi^2\right)^2 + \mathsf{F}^3\left(\Delta\Phi^2\right)^1 + \mathsf{F}^4 \tag{21}$$

In the same way, equation 10 becomes:

$$\mathsf{F}_{t_3} = \mathsf{F}^1_{t_0}\left(\Delta\Phi^1\right)^3 + \mathsf{F}^2_{t_0}\left(\Delta\Phi^2\right)^3 + \mathsf{F}^3\left(\Delta\Phi^2\right)^2 + \mathsf{F}^4\left(\Delta\Phi^2\right)^1 + \mathsf{F}^5 \tag{22}$$

Now, instead of four complex unknowns as in the additive situation, we now have three additional unknowns, making a total of seven. Consequently, the system is underdetermined. Unfortunately, every time we add a new constraint, we also introduce an additional unknown representing a new residue. In order to make the problem tractable, we must adopt an alternative strategy.

There are three steps in this strategy. The first step involves a reduction in the number of unknowns, and the second and third steps concern the solution of these unknowns.

Approximation of the Residues If we assume that the occluding object velocity is small then it follows that the residues will be approximately the same and also that they will be small. Thus, we assume that $\mathsf{F}^3(k_x, k_y) \approx \mathsf{F}^4(k_x, k_y) \approx \mathsf{F}^5(k_x, k_y)$. Adding a fifth constraint equation, we now have the following five simultaneous equations:

$$\mathsf{F}_{t_0} = \mathsf{F}^1_{t_0}\left(\Delta\Phi^1\right)^0 + \mathsf{F}^2_{t_0}\left(\Delta\Phi^2\right)^0 \tag{23}$$

$$\mathsf{F}_{t_1} = \mathsf{F}^1_{t_0}\left(\Delta\Phi^1\right)^1 + \mathsf{F}^2_{t_0}\left(\Delta\Phi^2\right)^1 + \mathsf{F}^3 \tag{24}$$

$$\mathsf{F}_{t_2} = \mathsf{F}^1_{t_0}\left(\Delta\Phi^1\right)^2 + \mathsf{F}^2_{t_0}\left(\Delta\Phi^2\right)^2 + \mathsf{F}^3\left(\Delta\Phi^2\right)^1 + \mathsf{F}^3 \tag{25}$$

$$\mathsf{F}_{t_3} = \mathsf{F}^1_{t_0}\left(\Delta\Phi^1\right)^3 + \mathsf{F}^2_{t_0}\left(\Delta\Phi^2\right)^3 + \mathsf{F}^3\left(\Delta\Phi^2\right)^2 + \mathsf{F}^3\left(\Delta\Phi^2\right)^1 + \mathsf{F}^3 \tag{26}$$

$$\mathsf{F}_{t_4} = \mathsf{F}^1_{t_0}\left(\Delta\Phi^1\right)^4 + \mathsf{F}^2_{t_0}\left(\Delta\Phi^2\right)^4 + \mathsf{F}^3\left(\Delta\Phi^2\right)^3 + \mathsf{F}^3\left(\Delta\Phi^2\right)^2 + \mathsf{F}^3\left(\Delta\Phi^2\right)^1 + \mathsf{F}^3 \tag{27}$$

Recall that this system of equations assumes that $\mathsf{F}^2_{t_0}$ corresponds to the occluded background image which is why we use the associated incremental phase change $\Delta\Phi^2$ in the residue terms. Note that if we let the residue equal zero then the equations are those of the purely additive case, *i.e.* equations 7 to 10, as one would expect.

Solution for the Incremental Phase Change We now have a system of five non-trivial simultaneous equations in five complex unknowns. Unfortunately, attempts at finding a well-posed closed-form analytical solution have been unsuccessful so far and an implementation of a gradient-descent numerical solution proved unreliable, principally because of the high-dimensionality of the solution space (it is a 10-D solution-space since there are five complex unknowns).

Instead, in order to solve these equations for $\mathsf{F}^1_{t_0}$, $\mathsf{F}^2_{t_0}$, $\Delta\Phi^1$, $\Delta\Phi^2$, and F^3, we exploit the velocity-dependent and frequency-dependent regularity of $\Delta\Phi^1$ and $\Delta\Phi^2$.

Recall from section 2.1 above that we used this relationship to sort the frequency components into the two distinct sets corresponding to each individual image in the case of additive images. In the occluded case which we are addressing in this paper, we solve for $\Delta\Phi^1$ and $\Delta\Phi^2$ as in the additive case using equations 11 and 12 above, *i.e.* by assuming the residue is zero. However, since

the residue is not zero, the solution for each $\Delta\Phi^1$ and $\Delta\Phi^2$ will include an error $\delta_{\Delta\Phi^1}$ and $\delta_{\Delta\Phi^2}$. Since these errors are not systematic for all spatial frequencies (k_x, k_y), we can estimate the true value of $\Delta\Phi^1$ and $\Delta\Phi^2$ by fitting a plane to the erroneous values $\Delta\Phi^{1'} = \Delta\Phi^1 + \delta_{\Delta\Phi^1}$ and $\Delta\Phi^{2'} = \Delta\Phi^2 + \delta_{\Delta\Phi^2}$. Specifically, we use the values $\Delta\Phi^{1'}(k_x, k_y)$ and $\Delta\Phi^{2'}(k_x, k_y)$ for all spatial frequencies to identify the velocity (v_x^1, v_y^1) of the foreground occluding object and (v_x^2, v_y^2) of the background occluded image as described above. Having computed these velocities, we can then compute an estimate of the true incremental phase change $\Delta\Phi^1(k_x, k_y)$ and $\Delta\Phi^2(k_x, k_y)$ for each Fourier component, *i.e.* at each spatial frequency, from

$$\Delta\Phi^i = e^{-i(k_x v_x^i \delta t + k_y v_y^i \delta t)} \tag{28}$$

and, hence,

$$\Re(\Delta\Phi^i) = \cos(k_x v_x^i \delta t + k_y v_y^i \delta t) \tag{29}$$

$$\Im(\Delta\Phi^i) = \sin(k_x v_x^i \delta t + k_y v_y^i \delta t) \tag{30}$$

Solution for the Fourier Components At this point, we have 'solved' (or, rather, estimated) two of the unknowns $\Delta\Phi^1$ and $\Delta\Phi^2$, and it remains simply to solve for the remaining three $\mathsf{F}_{t_0}^1$, $\mathsf{F}_{t_0}^2$, and F^3. This can be accomplished in a straight-forward manner using equations 23, 24, and 25. These equations have the form

$$a + b - p = 0 \tag{31}$$

$$ad + be + c - q = 0 \tag{32}$$

$$ad^2 + be^2 + c(1 + e) - r = 0 \tag{33}$$

where a, b, and c are unknown. These three simultaneous equations yield the following solutions for a, b, and c:

$$c = \frac{p(de^2 - d^2e) + q(d^2 - e^2) + r(e - d)}{(d^2 - e^2) + (1 + e)(e - d)} \tag{34}$$

$$b = \frac{q - pd - c}{e - d} \tag{35}$$

$$a = p - b \tag{36}$$

However, before we solve for $\mathsf{F}_{t_0}^1$, $\mathsf{F}_{t_0}^2$, and F^3 in this way, we must identify which is the occluding and which is the occluded component as we need to assign the appropriate known value of background phase change $\Delta\Phi^2$ to the term e in equations 34, 35, and 36.

We assume, arbitrarily, that the background velocity is smaller than that of the foreground occluding object. This provides the basis for deciding on which

values of $\mathsf{F}^1_{t_0}$, and $\mathsf{F}^2_{t_0}$ are the occluding and occluded. Specifically,

$$\left.\begin{array}{l} d = \Delta\Phi^1 \\ e = \Delta\Phi^2 \end{array}\right\} |\Delta\Phi^1| > |\Delta\Phi^2|$$

and solve for $a = \mathsf{F}^1_{t_0}$, $b = \mathsf{F}^2_{t_0}$, and $c = \mathsf{F}^3$. On the other hand

$$\left.\begin{array}{l} d = \Delta\Phi^2 \\ e = \Delta\Phi^1 \end{array}\right\} |\Delta\Phi^1| \leq |\Delta\Phi^2|$$

and solve for $a = \mathsf{F}^2_{t_0}$, $b = \mathsf{F}^2_{t_0}$, and $c = \mathsf{F}^3$.

Note that equations 34 and 35 are degenerate if $d = e$, that is, if the incremental phase change of the occluded and occluding objects are equal: $\Delta\Phi^1(k_x, k_y) = \Delta\Phi^2(k_x, k_y)$. However, significant errors also arise as $d \rightarrow e$, *i.e.* as $\Delta\Phi^1 \rightarrow \Delta\Phi^2$, since the numerator in equations 34 and 35 involve estimates of d and e rather than exact values. Hence, the relative magnitude of the error inherent in this estimate grows exponentially as $(d - e) \rightarrow 0$; that is, as the denominator approaches zero. To eliminate the influence of these errors in the computation of the residue on the estimate of the occluded and occluding signal b and a, we simply set $c = (c' + \epsilon_c)$ equal to zero as $d \rightarrow e$. Specifically, we let $c = 0$ if

$$|e - d| \leq \tau v_{\max} \tag{37}$$

where τ is some specified toleralance and $v_{\max}$ is the magnitude of the maximum velocity exhibited by either the occluded or the occluding image.

However, we cannot effect the same strategy for b as $d \rightarrow e$ since this would be equivalent to the implementation of an ideal pulse-shaped band-stop filter. Instead, it is necessary to attenuate progressively the band-stop frequencies before computing the inverse FFT of the occluding and background images. The approach adopted in this paper is to attenuate frequencies satisfying equation 37 according to the function

$$\mathsf{G}^{1,2}_{t_0}(k_x, k_y) = \mathsf{F}^{1,2}_{t_0}(k_x, k_y)\left(\sin\left(\frac{e-d}{\tau} \times \frac{\pi}{2}\right)\right)^{2n}, |d - e| \leq \frac{\tau}{2} v_{\max}; \tag{38}$$

$$\mathsf{G}^{1,2}_{t_0}(k_x, k_y) = \mathsf{F}^{1,2}_{t_0}(k_x, k_y), \text{otherwise.} \tag{39}$$

The exponent $n \geq 1$ determines the slope of the filter cutoff. As $n \rightarrow \infty$, the filter approaches an ideal filter. In the following, we use values of $\tau = 0.2$ and $n = 1$ throughout.

2.3 Velocity Estimation: Computation of Instantaneous Optical Flow

We have seen above that the solution of the system of equations 7 – 10 (additive case) and equations 23 – 27 (generalized occluded case) yields not only the decoupled or segmented Fourier components but also the rate of change of phase

and, hence, the velocity of the image translation. Consequently, we can exploit this approach to compute the local instantaneous velocity of a local region by treating it simply as a sub-image or image window. Unfortunately, the image data in such a region will exhibit a change due not only to the signal shift but also the translation of objects into the window and out of the the window. Consequently, there is a change in the spectral content of the window and not just a phase change as is assumed in the model. In order to reduce the impact of this 'edge effect', image data in a region is apodized or weighted as a function of its distance from the region centre. In this paper, a Gaussian weighting function is used and the Gaussian's standard deviation σ chosen such that the weighting at a some distance from the region centre is 50% of that at the region centre, where w is the length of the side of the 2-D region. Results are presented for Gaussian weighting functions of three standard deviations, each representing increased attenuation of image data toward the edge of the image (the three functions provide 50% weighting at $\frac{w}{8}$, $\frac{2w}{8}$, and $\frac{3w}{8}$ from the region centre). In the following, we will denote the three Gaussian functions as $\sigma_{\frac{w}{8}}$, $\sigma_{\frac{2w}{8}}$, and $\sigma_{\frac{3w}{8}}$.

3 Results and Evaluation

3.1 Image Separation and Segmentation

An image scenario which displays significant occlusion was used to test the approach. In each test, there are two objects, the foreground and the background, moving independently of one another with velocities (v_{x_1}, v_{y_1}) and (v_{x_2}, v_{y_2}), respectively. The technique was tested for velocities in the ranges: $(0,0) \leq (v_{x_1}, v_{y_1}) \leq (5,0)$ and $(0,0) \leq (v_{x_2}, v_{y_2}) \leq (5,5)$, in increments of one pixel. Due to lack of space, only quantitative results for $(v_{x_1}, v_{y_1}) = (0,0)$ are reported here although the images in figures 1 and 2 are derived from an image sequence where $(v_{x_1}, v_{y_1}) = (1,0)$. Nonetheless, the quantitative results for $(1,0) \leq (v_{x_1}, v_{y_1}) \leq (5,0)$ are comparable except where $(v_{x_1}, v_{y_1}) \rightarrow (v_{x_2}, v_{y_2})$ in which case the segmentation is unreliable and the RMS error is large.

The test scenario comprises two images, one of a child and the other of a sea-gull with a zero background. Figure 1 shows each individual image at time t_0 and the additive and occluding superposition of the seagull on the child background at time t_0. Figure 2 shows the result of computing (segmenting) the additively-generated composite images based on the sequence at time t_0, t_1, t_2, and t_3 using the original additive model. It also shows the result of segmenting the occlusion-based composite images using the generalized occlusion model.

Table 1 shows the RMS errors between the segmented occluding (foreground) image and the original occluding image for all combinations of image velocities (in the range specified above) for (i) the additive images segmented using the additive model, (ii) the occluded images using the additive model, and (iii) the occluded images using the occlusion model, respectively. Note that the values in this table are based on the RMS error computed using both the non-zero pixel values of foreground object and its surrounding zero pixel values (*i.e.* we take

Table 1. RMS error in reconstruction of gull foreground image computed for the complete image including its zero-valued backdrop; background image (child) velocity = (0, 0) pixels, foreground image (gull) velocity = (v_{x_2}, v_{y_2}) pixels.

RMS Error in Reconstruction of Foreground Image (*i.e.* Occluding Image including Zero Background)								
Type of Image Sequence	Decoupling Model	v_{x_2}	v_{y_2}					
			0	1	2	3	4	5
Additive	Additive	0	–	28.5	27.6	27.4	27.7	27.8
		1	18.5	13.4	12.0	14.0	14.1	14.6
		2	18.0	10.1	11.7	10.7	12.0	11.6
		3	18.2	11.3	10.6	12.0	11.1	12.0
		4	18.1	11.1	11.2	11.9	13.1	13.0
		5	18.2	11.5	12.0	11.9	12.4	14.0
Occlusion	Additive	0	–	50.2	46.9	47.8	49.3	49.9
		1	52.0	51.4	47.5	49.4	53.2	53.0
		2	49.4	48.2	46.4	49.0	51.5	52.0
		3	45.8	50.1	49.2	48.5	50.7	50.9
		4	46.4	48.2	49.1	50.5	51.1	52.1
		5	43.4	48.2	49.8	49.8	53.7	52.2
Occlusion	Occlusion	0	–	44.8	44.9	45.0	45.2	45.4
		1	44.2	46.1	46.2	46.1	46.0	46.2
		2	44.5	46.3	47.0	46.4	46.5	46.6
		3	44.6	45.8	46.6	47.2	47.3	46.9
		4	44.7	45.6	46.4	46.9	47.3	47.5
		5	44.9	45.8	46.7	46.9	47.7	48.0

into consideration the success of the technique in segmenting an occluding object and in reconstructing a zero-valued background).

3.2 Estimation of Optical Flow

Figure 3 demonstrates the results of applying the technique to four images in Otte's and Nagel's ground-truth test sequence [15]. Figure 3 (a) and (b) show images number 40 and 43 in the sequence whilst (c) shows the true optical flow field extracted directly from the Otte's and Nagel's ground-truth data (sampled every ten pixels). Figure 3 (d) shows the optical flow field computed using the technique described in this paper. All of the results shown in this paper were computed with a window size of 64 × 64 pixels. Flow vectors are plotted every ten pixels and their magnitude has been scaled by a factor of four.

Table 2 summarizes the mean magnitude and the mean direction of ground-truth data and the measured velocities; table 3 provides a summary of the RMS and mean errors of the measured velocities.

Note that all of the results presented in this paper are the unprocessed output of the algorithm (apart from interpolation); each velocity vector has been estimated independently and the vector field has not be subjected to median or mean filtering.

Table 2. Summary of mean and standard deviation of the magnitude and direction of the ground-truth data and the measured velocities.

Sequence	Gaussian Weighting	Magnitude		Direction	
		mean	standard deviation	mean	standard deviation
Image Translation Benchmark	Ground Truth	2.828	0.0	0.785	0.0
	$\frac{3w}{8}$	2.428	0.126	0.786	0.054
	$\frac{2w}{8}$	2.759	0.046	0.786	0.009
	$\frac{w}{8}$	2.748	0.063	0.784	0.063
Otte & Nagel Benchmark	Ground Truth	1.573	0.641	0.306	0.202
	$\frac{3w}{8}$	1.307	0.582	0.339	0.411
	$\frac{2w}{8}$	1.307	0.596	0.322	0.366
	$\frac{w}{8}$	1.311	0.616	0.312	0.337

Table 3. Summary of RMS and maximum errors in the measured velocities; errors are defined with respect to ground-truth data.

Sequence	Gaussian Weighting	RMS Error		Maximum Error	
		Magnitude (pixels)	Direction (radians)	Magnitude (pixels)	Direction (radians)
Image Translation Benchmark	$\frac{3w}{8}$	0.419	0.054	0.803	0.231
	$\frac{2w}{8}$	0.083	0.009	0.211	0.049
	$\frac{w}{8}$	0.102	0.049	1.028	0.814
Otte & Nagel Benchmark	$\frac{3w}{8}$	0.479	0.154	5.369	2.494
	$\frac{2w}{8}$	0.492	0.114	2.775	0.745
	$\frac{w}{8}$	0.529	0.124	5.625	1.601

4 Discussion

All of the images shown in figures 1 and 2 have a velocity of (1, 0) and (4, 0) pixels for each image, respectively. These velocities were chosen not because they provide the best results in each case but because they provide a reasonably typical example of a slowly translating background and a foreground translating in the same direction but with a larger velocity.

It is clear from figure 2 that the best segmentation is achieved for the additive case (using the additive model). This is to be expected since the problem is well-posed and exactly determined; any errors are due to the impossibility of solving for some spatial frequencies, specifically those for which the phase change is identical in forground and background and those above a certain limiting range. The RMS errors in the cases of the occluded image sequence are inevitably larger since it is less well-posed and, as we have seen, requires an approximation in the solution. Finally, we note that the error in the computed velocity is consistently in the order of 0.1 to 0.2 pixels.

Concerning the computation of optical flow, we can see that, qualitatively, the approach described produces a dense and reasonably complete flow field (with the flow field associated with 'wider' weighting functions, *i.e.* the Gaussians with standard deviation $\sigma_{\frac{3w}{8}}$ and $\sigma_{\frac{2w}{8}}$ exhibiting somewhat poorer localization of the flow field). More importantly, however, is the accuracy of the technique. Otte and Nagel's [15] benchmark sequence was used to evaluate quantitatively the

performance of the algorithm. This sequence has the major benefit that ground truth optical flow is available (*i.e.* the magnitude and direction of the optical flow of (almost) every point in the image). To compare the optical flow computed with the algorithm presented in this paper and ground-truth, the optical flow was estimated every ten pixels (for the three Gaussian weighting functions) and then a complete optical flow image was produced for both magnitude and direction by interpolating bi-linearly among these points. These were then compared to the ground-truth magnitude and direction images by estimating the RMS error (see table 3). Finally, the mean and standard deviation of the magnitude and direction of the ground-truth flow field and the three computed flow fields are given in table 2.

Referring to these images and tables, a number of points can be noted.

First, it is clear that the main errors occur at the occluding contours and, in particular, at the contour where the two objects are moving with significant velocities (as, for example, is the case with the white block and the large dark block in the foreground). Again, as expected, this error is greater for the wider weighting functions and, because the velocity estimate is based on a larger effective support, the error propagates into a bigger region around the occluding contour.

Second, the mean magnitudes and directions of the three computed flow fields are consistent and do not vary significantly (1.307, 1.307, and 1.311 pixels, and 0.339, 0.322, 0.312 radians, for mean magnitudes and directions, respectively). This compares with the ground-truth mean magnitude and direction values of 1.573 and 0.306, respectively. Clearly, there is a difference in the measured and ground-truth magnitude value. This apparent difference is evident in the vector-field (compare the field in 3 (c) and (d)) and shows up in the RMS error estimate. On average, the error in the estimate of the magnitude of the optical flow field is just over one quarter of a pixel and the error in direction is of the order of 0.02 radians radian or approximately one degree.

As the average magnitude error is significant, for comparison we also applied the technique to a trivial benchmark data set wherein an image is simply translated in the x and y direction by 2 pixels/frame. It should be emphasised that this test is intended to do no more than demonstrate the accuracy and the repeatability of the technique on real data with a known (and trivial) flow-field. Table 2 summarizes the mean magnitude and the mean direction of ground-truth data and the measured velocities; table 3 provides a summary of the RMS and maximum errors of the measured velocities. The chief point to note about these results is that, with the exception of the narrowest apodization window, the correct flow field is computed to within 0.1 pixels (magnitude) and 0.001 radians (direction). As it happens, the magnitude measurement accuracy is presently set at 0.1 pixels/frame as this is the sampling period of the (velocity) Hough transform space used in all of the work described in this paper. It remains to be seen whether or not this can be improved by increasing the decreasing the sampling period.

5 Conclusions

This paper presented and validated the theoretical basis for the use of Fourier techniques in separating additive images and in segmenting images which are formed by an occluding foreground and an occluded background. This theory facilitates the model-free segmentation of moving objects in dynamic image sequences in situations where the object velocity is constant and normal to the principal ray of the image sensor.

The significance of the approach and the main contribution of the work is that it allows model-free segmentation and, importantly, the visual complexity of the foreground and the background is irrelevant since the segmentation is effected independently for each individual spatial frequency in the Fourier domain.

The technique is also applied to the estimation of instantaneous optical flow by decoupling the Fourier transform of a local Gaussian-weighted window centred at every point at which the flow field is to be computed. The results compare very favourably with ground truth optical flow data for the benchmark image sequence used to test the approach. Problems remain at occluding contours because, so far, we have used the additive model for the optical flow estimation. However, since the generalized occlusion model computes the velocities of both objects in the local window around the occluding boundary, we expect to be able to improve this estimate even further in the future and to be able to compute the correct flow field on either side of the velocity discontinuity (*i.e.* the occluding boundary).

Acknowledgements

I would like to acknowledge the helpful comments of Dr. W. F. Lunnon, Department of Computer Science, National University of Ireland, Maynooth.

References

1. D. Vernon, *Machine Vision* Prentice-Hall International, London (1991).
2. D. Vernon and G. Sandini, *Parallel Computer Vision — The* VIS *a* VIS *System*, Ellis Horwood, London (1992).
3. J.H. Duncan and T.-C. Chou, "On the detection and the computation of optical flow", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14(3)**, 346-352 (1992).
4. H. Shariat and K.E. Price, "Motion estimation with more than two frames", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12(5)**, 417-434 (1990).
5. M.P. Cagigal, L. Vega, P. Prieto, "Object movement characterization from low-light-level images", *Optical Engineering*, **33(8)**, 2810-2812 (1994).
6. M.P. Cagigal, L. Vega, P. Prieto, "Movement characterization with the spatiotemporal Fourier transform of low-light-level images", *Applied Optics*, **34(11)**, 1769-1774 (1995).

7. S. A. Mahmoud, M.S. Afifi, and R. J. Green, "Recognition and velocity computation of large moving objects in images", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **36(11)**, 1790-1791 (1988).
8. S. A. Mahmoud, "A new technique for velocity estimation of large moving objects", *IEEE Transactions on Signal Processing*, **39(3)**, 741-743 (1991).
9. S.A. Rajala, A. N. Riddle, and W.E. Snyder, "Application of one-dimensional Fourier transform for tracking moving objects in noisy environments", *Computer Vision, Graphics, and Image Processing*, **21**, 280-293 (1983).
10. D. Vernon, "Phase-Based Measurement of Object Velocity in Image Sequences using the Hough Transform", *Optical Engineering* (1996).
11. D. J. Fleet and A.D. Jepson, "Hierarchical construction of orientation and velocity selective filters", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11(3)**, 315-325 (1989).
12. D. J. Fleet and A.D. Jepson, "Computation of component image velocity from local phase information", *International Journal of Computer Vision*, **5**, 77-104 (1990).
13. D. J. Fleet and A.D. Jepson, "Stability in Phase Information", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **15(12)**, 1253-1268 (1993).
14. J.L. Barron, D.J. Fleet, and S. Beauchemin, "Performance of optical flow techniques", *Int. Journal of Computer Vision*, **12(1)**, 43-77 (1994).
15. M. Otte and H.-H. Nagel, "Optical flow estimation: advances and comparisons", *Lecture Notes in Computer Science*, J.O. Eklundh (Ed.), *Computer Vision - ECCV '94*, Springer-Verlag, Berlin, 51-60 (1994).
16. M. Tistarelli, "Multiple constraints for optical flow", *Lecture Notes in Computer Science*, J.O. Eklundh (Ed.), *Computer Vision - ECCV '94*, Springer-Verlag, Berlin, 61-70 (1994).
17. L. Jacobson and H. Wechsler, "Derivation of optical flow using a spatiotemporal-frequency approach", *Computer Vision, Graphics, and Image Processing*, **38**, 29-65 (1987).
18. D. Vernon, "Decoupling Fourier Components of Dynamic Image Sequences: Theory and Application to Segmentation and Estimation of Optical Flow", Technical Report, Department of Computer Science, National University of Ireland, Maynooth (1997).
19. D. Vernon, "Segmentation in Dynamic Image Sequences by Isolation of Coherent Wave Profiles", Proceedings of the 4th European Conference on Computer Vision, Springer-Verlag, 293-303 (1996).
20. P.V.C. Hough, 'Method and Means for Recognising Complex Patterns' U.S. Patent 3,069,654, (1962).
21. L. Hahn, *Complex Numbers and Geometry*, The Mathmatical Association of America, Washington, D.C. (1994).

7. S. A. Mahmoud, M.S. Afifi, and R. J. Green, "Recognition and velocity computation of large moving objects in images", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **36(11)**, 1790-1791 (1988).
8. S. A. Mahmoud, "A new technique for velocity estimation of large moving objects", *IEEE Transactions on Signal Processing*, **39(3)**, 741-743 (1991).
9. S.A. Rajala, A. N. Riddle, and W.E. Snyder, "Application of one-dimensional Fourier transform for tracking moving objects in noisy environments", *Computer Vision, Graphics, and Image Processing*, **21**, 280-293 (1983).
10. D. Vernon, "Phase-Based Measurement of Object Velocity in Image Sequences using the Hough Transform", *Optical Engineering* (1996).
11. D. J. Fleet and A.D. Jepson, "Hierarchical construction of orientation and velocity selective filters", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11(3)**, 315-325 (1989).
12. D. J. Fleet and A.D. Jepson, "Computation of component image velocity from local phase information", *International Journal of Computer Vision*, **5**, 77-104 (1990).
13. D. J. Fleet and A.D. Jepson, "Stability in Phase Information", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **15(12)**, 1253-1268 (1993).
14. J.L. Barron, D.J. Fleet, and S. Beauchemin, "Performance of optical flow techniques", *Int. Journal of Computer Vision*, **12(1)**, 43-77 (1994).
15. M. Otte and H.-H. Nagel, "Optical flow estimation: advances and comparisons", *Lecture Notes in Computer Science*, J.O. Eklundh (Ed.), *Computer Vision - ECCV '94*, Springer-Verlag, Berlin, 51-60 (1994).
16. M. Tistarelli, "Multiple constraints for optical flow", *Lecture Notes in Computer Science*, J.O. Eklundh (Ed.), *Computer Vision - ECCV '94*, Springer-Verlag, Berlin, 61-70 (1994).
17. L. Jacobson and H. Wechsler, "Derivation of optical flow using a spatiotemporal-frequency approach", *Computer Vision, Graphics, and Image Processing*, **38**, 29-65 (1987).
18. D. Vernon, "Decoupling Fourier Components of Dynamic Image Sequences: Theory and Application to Segmentation and Estimation of Optical Flow", Technical Report, Department of Computer Science, National University of Ireland, Maynooth (1997).
19. D. Vernon, "Segmentation in Dynamic Image Sequences by Isolation of Coherent Wave Profiles", Proceedings of the 4th European Conference on Computer Vision, Springer-Verlag, 293-303 (1996).
20. P.V.C. Hough, 'Method and Means for Recognising Complex Patterns' U.S. Patent 3,069,654, (1962).
21. L. Hahn, *Complex Numbers and Geometry*, The Mathmatical Association of America, Washington, D.C. (1994).

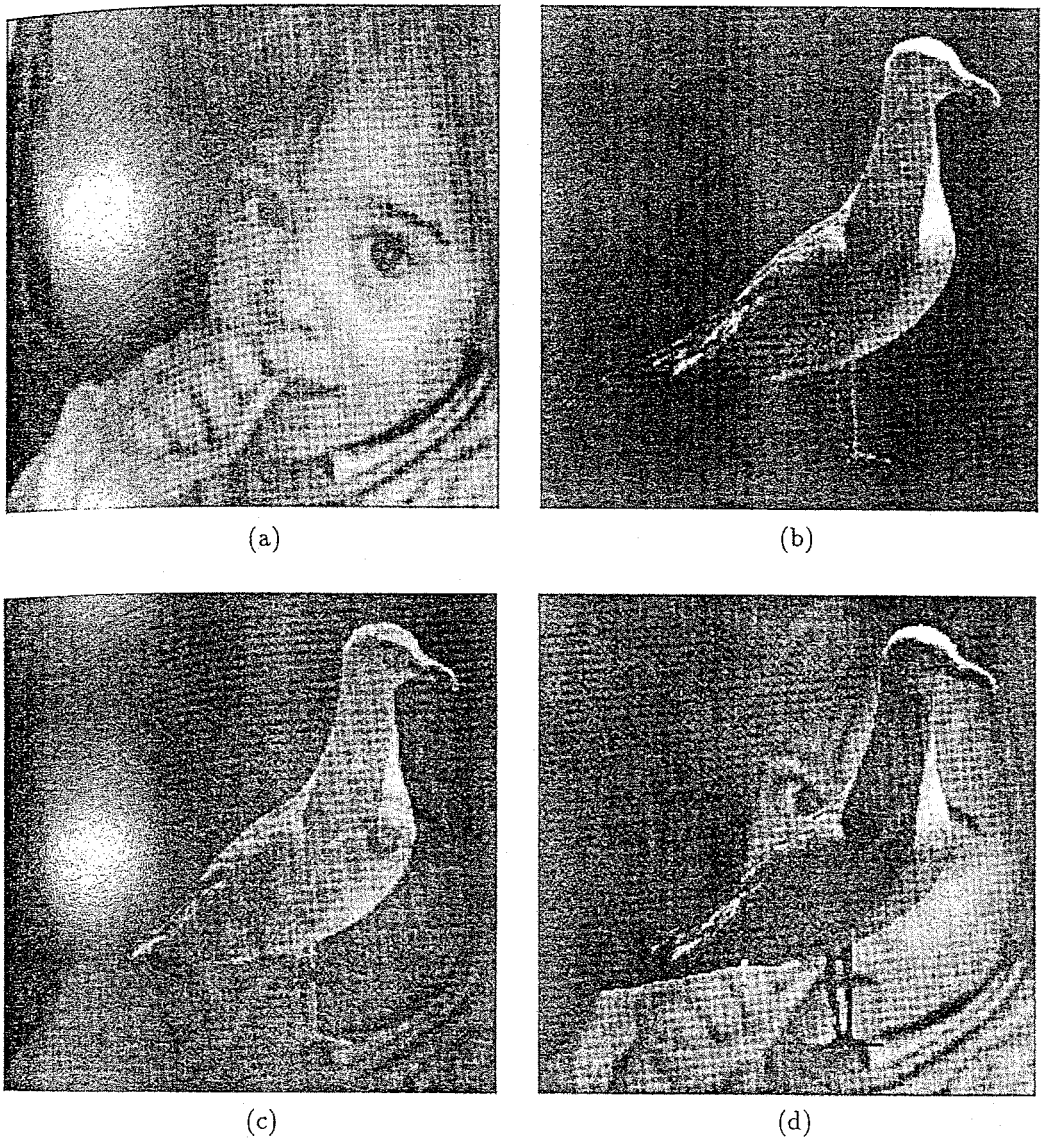


Fig. 1. Image separation and segmentation: (a) and (b) images 1 and 2 at time t_0 translating with velocities (1, 0) and (4, 0) pixels, respectively; (c) and (d) the additive and occluding superposition of the seagull on the child background at time t_0.

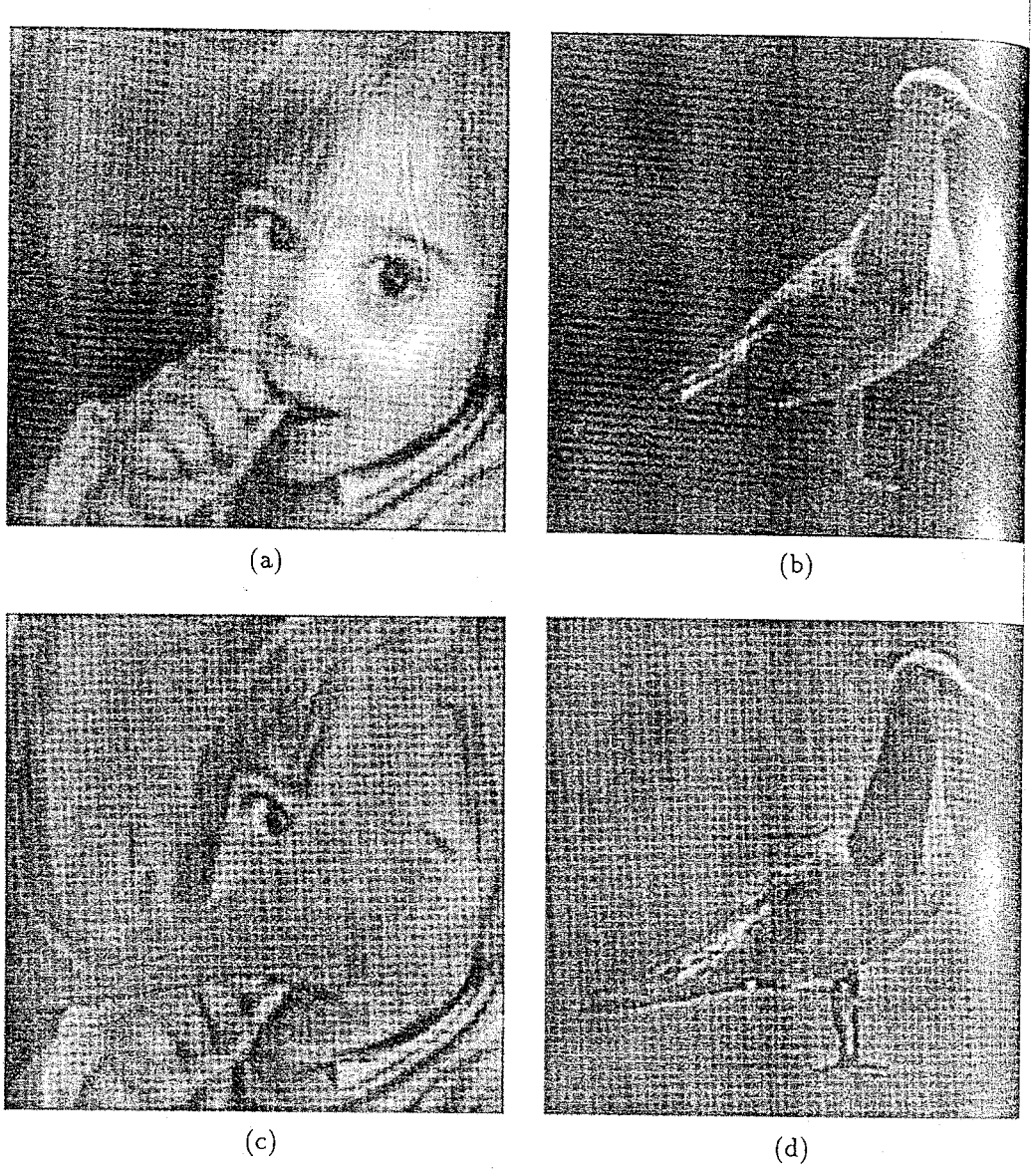


Fig. 2. Image separation and segmentation test scenario: (a) and (b) show the results of decoupling (segmenting) the additively-generated composite images based on the sequence at time t_0, t_1, t_2, and t_3 using the additive model; (e) and (f) show the results of segmenting the occlusion-based composite images using the generalized occlusion model.

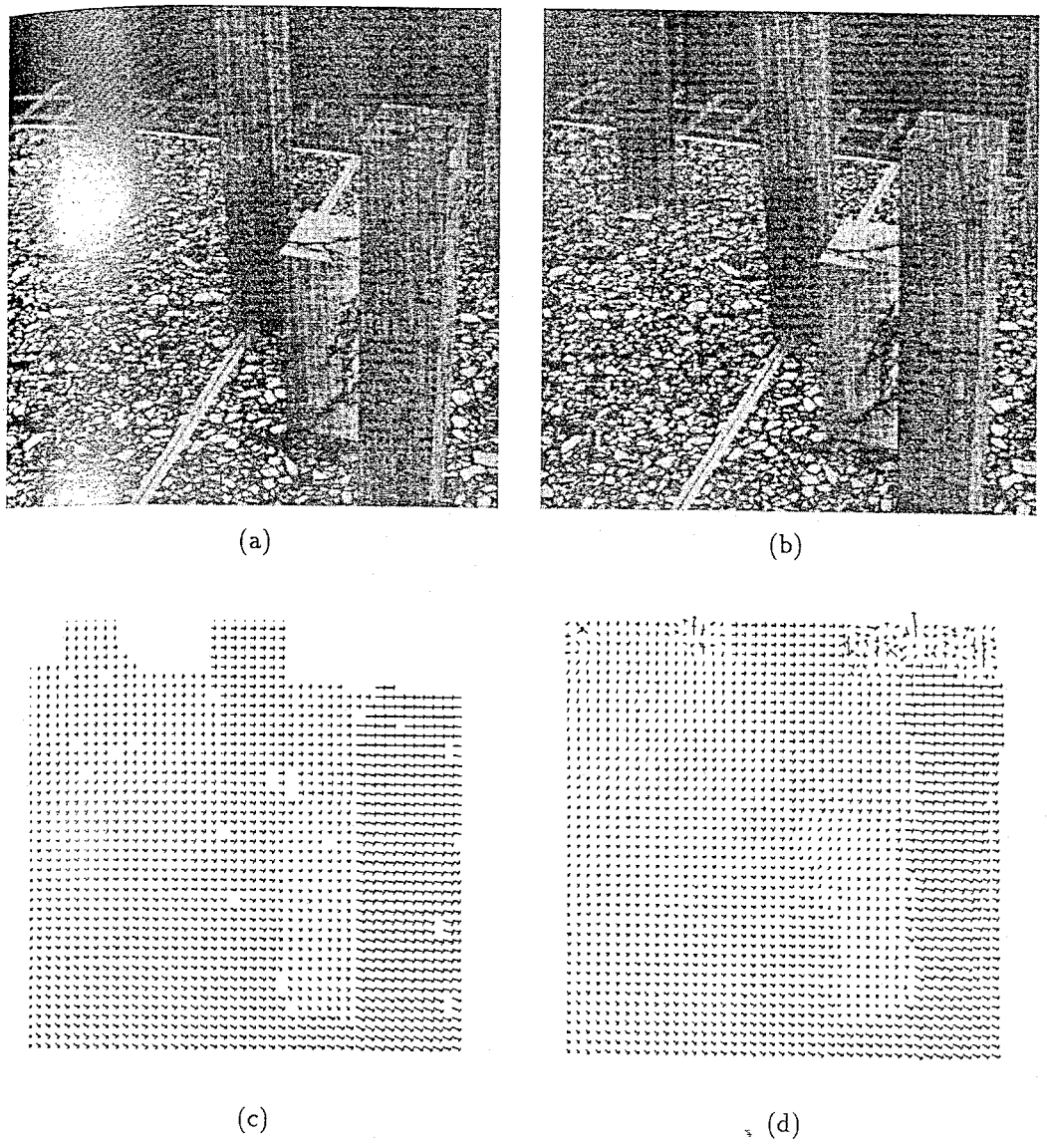


Fig. 3. (a) and (b) Images number 40 and 43 of Otte and Nagel's ground-truth motion sequence. (c) True optical flow field given by Otte and Nagel's ground-truth data (sampled every ten pixels) (d) Optical flow field computed using phase information: Gaussian weighting function with 50% weight at $\frac{w}{4}$ pixels from window centre (w, the window size, equals 64 pixels).

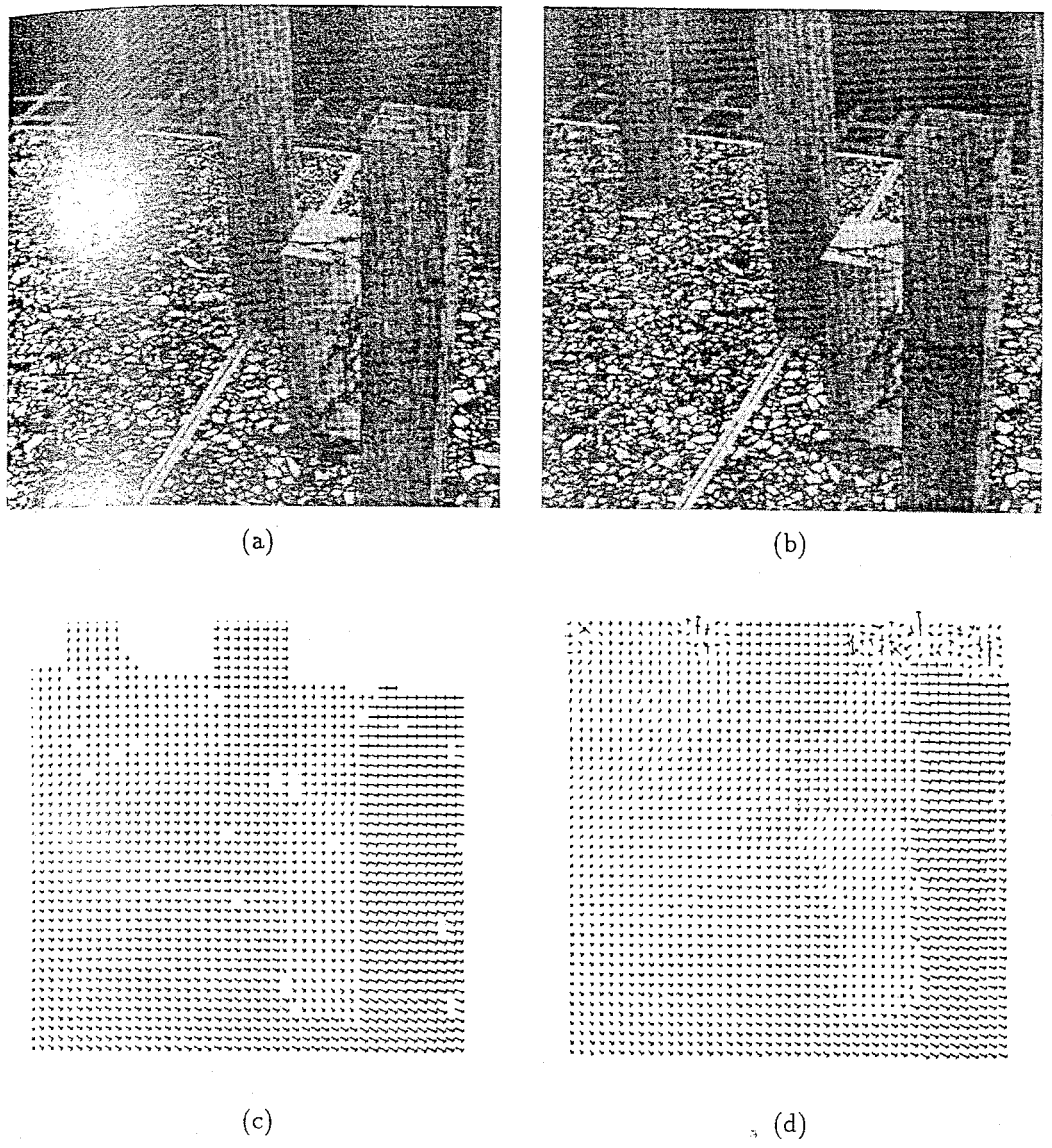


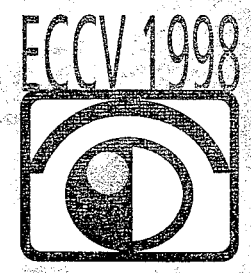
Fig. 3. (a) and (b) Images number 40 and 43 of Otte and Nagel's ground-truth motion sequence. (c) True optical flow field given by Otte and Nagel's ground-truth data (sampled every ten pixels) (d) Optical flow field computed using phase information: Gaussian weighting function with 50% weight at $\frac{w}{4}$ pixels from window centre (w, the window size, equals 64 pixels).

Lecture Notes in Computer Science 1407

Hans Burkhardt
Bernd Neumann (Eds.)

Computer Vision – ECCV'98

5th European Conference
on Computer Vision
Freiburg, Germany, June 1998
Proceedings, Volume II



Springer