ADAPTIVE MULTISCALE OPTICAL FLOW ESTIMATION

Jian Li[1]*, Christopher P Benton*[1]*, Stavri G Nikolov*[2]*, Nicholas E Scott-Samuel*[1]

[1]Department of Experimental Psychology, 12A, Priory Road, BS8 1TU, U.K.
[2]Department of Electrical and Electronic Engineering, MVB, BS8 1UB, U.K.
University of Bristol, U.K.

ABSTRACT

The current paper presents a novel adaptive multiscale scheme to estimate optical flow from image sequences. The scheme models estimation uncertainties which are used to reduce the influence of unreliable intermediate estimates on accuracy. The experimental results show that the proposed method provides more accurate estimates for both small and large motions than a standard multiscale scheme in which an increment is added to an intermediate estimate regardless of estimation certainty.

Index Terms— Optical flow, multiscale, pyramid, least squares. uncertainty

1. INTRODUCTION

The current paper presents a novel adaptive multiscale scheme to recover optical flows from image sequences. In a standard multiscale scheme, for example [1], a warped image at a finer pyramid level is produced using estimates from a coarser pyramid level. By using the warped image and video image at the same level, a velocity increment is estimated which is used as a correction to the velocity estimate from a coarser level. In the standard scheme, an increment could be affected by noise at the finer scale and once the increment is erroneously estimated, the scheme is not able to recover from the error [2]. Regarding this problem, Simoncelli modeled cross-scale refinement as a stochastic process and applied the Kalman filtering technique to ensure the optimality of intermediate estimates.

A second problem which is less frequently addressed is the influence of the number of pyramid levels on estimation accuracy, especially for small displacements. In real application, the largest number of pyramid levels should generally be used (within the limit of image size) to cover all possible displacements. However, because of the down-sampling procedure in constructing a Gaussian pyramid [3], a small velocity from the original images could be down-scaled to a tiny velocity at the coarsest scale. In this case, image noise may introduce large error to the coarsest estimate and the error remains in the refinements at finer scales. As we show below, the standard scheme is not able to produce accurate estimate in this case.

The proposed adaptive scheme solves the above problems through improving the accuracy of intermediate estimates at all levels. The scheme assumes a stochastic process for the cross-scale velocity refinement, in which estimation uncertainties are modeled as variances of intermediate estimates obtained from a least squares estimation scheme. By adaptively reducing the variances, superior accuracy can be guaranteed. Our experiments show that the proposed technique produces more accurate estimates than the standard scheme for both small and large displacements. Moreover, the proposed scheme ensures that the use of a large number of pyramid levels does not introduce serious errors to small displacements and the scheme is suitable for a procedure in which both cross-scale and same-scale refinements are adopted.

The work has been funded by the UK MOD Data and Information Fusion Defence Technology Centre.

2. ADAPTIVE MULTISCALE ESTIMATION

2.1. Optical Flow Estimation

If a pixel moves from (x, y, t) to $(x + u, y + v, t + 1)$, we assume:

$$I(x + u, y + v, t + 1) + c = I(x, y, t), \tag{1}$$

$$\nabla I(x + u, y + v, t + 1) = \nabla I(x, y, t), \tag{2}$$

where I denotes image intensity, u and v are velocities in x and y directions, respectively. ∇ is a partial differentiation operator, c is a parameter compensating temporal variation of intensities. Eq.(1) models the constraint on image intensities, in which intensity variation c is allowed [4] while Eq.(2) models the constraint on spatial derivatives which are also assumed to be conserved over time. Applying Taylor expansion to the above models, we can get a linear expression of the unknown parameters $\mathbf{w} = [u, v, c]^T$:

$$\mathbf{A}\mathbf{w} = \begin{bmatrix} I_x & I_y & 1 \\ I_{xx} & I_{xy} & 0 \\ I_{xy} & I_{yy} & 0 \end{bmatrix} \mathbf{w} \approx - \begin{bmatrix} I_t \\ I_{xt} \\ I_{yt} \end{bmatrix} = -\mathbf{b}, \tag{3}$$

where $I_a = \frac{\partial I}{\partial a}$ and $I_{ab} = \frac{\partial}{\partial b}(\frac{\partial I}{\partial a})$ are the first order and the second order derivatives, respectively. By further assuming that velocities and intensity variations remain constant for all pixels in a spatial region Ω containing N pixels, $\mathbf{w}$ can be estimated through a least squares (LS) scheme [5]:

$$\hat{\mathbf{w}} = -\bar{\mathbf{A}}^{-1}\bar{\mathbf{b}}, \tag{4}$$

where $\bar{\mathbf{A}} = \sum_\Omega \mathbf{A}^T\mathbf{A}$ and $\bar{\mathbf{b}} = \sum_\Omega \mathbf{A}^T\mathbf{b}$. According to the properties of a LS estimator [6], $\hat{\mathbf{w}}$ is an unbiased estimate of the true vector $\tilde{\mathbf{w}}$. Furthermore, a variance is associated with $\hat{\mathbf{w}}$ and its unbiased estimate is given as [6]:

$$\mathbf{D} = \frac{\sum_\Omega(-\mathbf{b}-\mathbf{A}\hat{\mathbf{w}})^T(-\mathbf{b}-\mathbf{A}\hat{\mathbf{w}})}{3N-p}(\bar{\mathbf{A}})^{-1}, \tag{5}$$

where $3N$ is the total number of samples used in the LS regression and $p = 3$ is the number of unknown parameters.

In a standard multiscale scheme, the above LS regression is used to produce the initial estimate at the coarsest level and estimates of increments at all finer levels. For a pyramid containing H levels, where the 1^{st} level is the finest level and the H^{th} level is the coarsest level, if we use $\hat{\mathbf{w}}_1$ and $\hat{\mathbf{w}}_2$ to denote the estimates before and after adding an increment, the refinement procedure from a coarser level $h+1$ to a finer level h consists of two steps. The estimate is firstly up-scaled because of the down-sampling applied when constructing pyramids:

$$\hat{\mathbf{w}}_1^h = \mathbf{G}\hat{\mathbf{w}}_2^{h+1}, \tag{6}$$

where $\mathbf{G} \in \mathcal{R}^{3\times3}$ is a diagonal matrix in which $\mathbf{G}_{11} = \mathbf{G}_{22} = 2$ are the up-scaling factors in x and y directions respectively and $\mathbf{G}_{33} = 1$ indicates that the intensity variation c is assumed to be the same for all levels. At level h, the LS scheme is also used to produce an estimate of an increment $\hat{\mathbf{z}}^h$. The final estimate at the level h is then computed as:

$$\hat{\mathbf{w}}_2^h = \hat{\mathbf{w}}_1^h + \hat{\mathbf{z}}^h = \mathbf{G}\hat{\mathbf{w}}_2^{h+1} + \hat{\mathbf{z}}^h. \tag{7}$$

Moreover, if the refinement is performed within the same level, the procedure can be modeled as:

$$\hat{\mathbf{w}}_2^h = \hat{\mathbf{w}}_2^h + \hat{\mathbf{z}}^h. \tag{8}$$

2.2. The Model of Estimation Uncertainty

We measure estimation uncertainty through computing the variance of an estimate. To this end, we firstly model the true values of $\mathbf{w}^h$ at any level h as $\tilde{\mathbf{w}}^h$. The cross-scale relationship between $\tilde{\mathbf{w}}^h$ and $\tilde{\mathbf{w}}^{h+1}$ can then be expressed as:

$$\tilde{\mathbf{w}}^h = \mathbf{G}\tilde{\mathbf{w}}^{h+1}, \tag{9}$$

where $\mathbf{G}$ is defined in Eq.(6).

Here we consider the estimate with an increment ($\hat{\mathbf{w}}_2^h$) and the one without an increment ($\hat{\mathbf{w}}_1^h$) as random variables with the same expectation:

$$E(\hat{\mathbf{w}}_2^h) = E(\hat{\mathbf{w}}_1^h) = \tilde{\mathbf{w}}^h, \tag{10}$$

where $E(*)$ denotes the expectation. Correspondingly, an increment $\hat{\mathbf{z}}^h$ is assumed to be a random variable which is independent from $\hat{\mathbf{w}}_1^h$ and has an expectation of $\mathbf{0}$. The assumption for $\hat{\mathbf{w}}_2^h$ is valid because the LS estimator is adopted throughout the whole procedure and any refined estimate is expected to be the same as its true value. Furthermore, because:

$$E(\hat{\mathbf{w}}_1^h) = E(\mathbf{G}\hat{\mathbf{w}}_2^{h+1}) = \mathbf{G}E(\hat{\mathbf{w}}_2^{h+1}) = \mathbf{G}\tilde{\mathbf{w}}^{h+1} = \tilde{\mathbf{w}}^h, \tag{11}$$

the assumption for $\hat{\mathbf{w}}_1^h$ stands.

For each component in the vector $\hat{\mathbf{w}}^h$, the corresponding uncertainty is modeled as the variance of the component. According to Eq.(5), at the coarsest level H we can obtain a variance matrix $\mathbf{D}_2^H$ for the initial estimate $\hat{\mathbf{w}}_2^H$. Because the covariances between components are not considered, we use the notation $\overline{\mathbf{D}}_2^H$ instead, in which the covariances are set to zero. Based on Eq.(6), when up-scaled to a finer level $H-1$, the variance matrix is computed as:

$$\overline{\mathbf{D}}_1^{H-1} = \mathbf{G}\overline{\mathbf{D}}_2^H\mathbf{G}^T. \tag{12}$$

As all matrices are diagonal, we use the following expression:

$$\overline{\mathbf{D}}_1^{H-1} = \mathbf{G}^T\mathbf{G}\overline{\mathbf{D}}_2^H. \tag{13}$$

Similarly, because the LS is used to estimate an increment $\hat{\mathbf{z}}^{H-1}$, we can also obtain its variance $\overline{\mathbf{D}}_z^{H-1}$ according to Eq.(5). As $\hat{\mathbf{w}}_1^{H-1}$ and $\hat{\mathbf{z}}^{H-1}$ are assumed to be independent from each other, the variance of $\hat{\mathbf{w}}_2^{H-1}$ is then computed as:

$$\overline{\mathbf{D}}_2^{H-1} = \overline{\mathbf{D}}_1^{H-1} + \overline{\mathbf{D}}_z^{H-1} = \mathbf{G}^T\mathbf{G}\overline{\mathbf{D}}_2^H + \overline{\mathbf{D}}_z^{H-1}. \tag{14}$$

More generally, the cross-scale variance updating procedure corresponding to the velocity updating procedure shown in Eq.(7) is expressed as:

$$\overline{\mathbf{D}}_2^h = \overline{\mathbf{D}}_1^h + \overline{\mathbf{D}}_z^h = \mathbf{G}^T\mathbf{G}\overline{\mathbf{D}}_2^{h+1} + \overline{\mathbf{D}}_z^h. \tag{15}$$

2.3. The Adaptive Scheme: Rule 1

From Eq.(15), it can be seen that variances are accumulated across scales. In order to increase estimation certainties at the finest scale, it is necessary to reduce the variance shown in Eq.(15) at any level h. As $\hat{\mathbf{z}}^h$ is a major source of errors, we reduce the variance through adaptively setting thresholds to $\hat{\mathbf{z}}^h$. In rule 1, we compute a final estimate $\hat{\mathbf{w}}_f^h$ at the level h as a weighted sum of the estimates with and without adding an increment, i.e. the weighted sum of $\hat{\mathbf{w}}_1^h$ and $\hat{\mathbf{w}}_2^h$ by using their variances:

$$\hat{\mathbf{w}}_f^h = \overline{\mathbf{D}}_2^h(\overline{\mathbf{D}}_1^h + \overline{\mathbf{D}}_2^h)^{-1}\hat{\mathbf{w}}_1^h + \overline{\mathbf{D}}_1^h(\overline{\mathbf{D}}_1^h + \overline{\mathbf{D}}_2^h)^{-1}\hat{\mathbf{w}}_2^h. \tag{16}$$

According to Eq.(7), $\hat{\mathbf{w}}_f^h$ can be further written as:

$$\begin{aligned}\hat{\mathbf{w}}_f^h &= \hat{\mathbf{w}}_1^h + \overline{\mathbf{D}}_1^h(\overline{\mathbf{D}}_1^h + \overline{\mathbf{D}}_2^h)^{-1}\hat{\mathbf{z}}^h \\ &= \hat{\mathbf{w}}_1^h + \overline{\mathbf{D}}_1^h(2\overline{\mathbf{D}}_1^h + \overline{\mathbf{D}}_z^h)^{-1}\hat{\mathbf{z}}^h.\end{aligned} \tag{17}$$

Because $\hat{\mathbf{w}}_f^h$ is the final intermediate estimate at any level, we have $\hat{\mathbf{w}}_1^h = \mathbf{G}\hat{\mathbf{w}}_f^{h+1}$ and the cross-scale refinement procedure can be finally expressed as follows according to Eq.(16):

$$\hat{\mathbf{w}}_f^h = \mathbf{G}\hat{\mathbf{w}}_f^{h+1} + \Lambda_1 \hat{\mathbf{z}}^{\,h}, \tag{18}$$

where $\Lambda_1 = \overline{\mathbf{D}}_1^h(2\overline{\mathbf{D}}_1^h + \overline{\mathbf{D}}_z^h)^{-1}$. According to the assumptions shown in Section 2.2, it can be proved that $\hat{\mathbf{w}}_f^h$ is also an unbiased estimate of $\tilde{\mathbf{w}}^h$. The unbiased estimate of the variance of $\hat{\mathbf{w}}_f^h$ is computed as:

$$\overline{\mathbf{D}}_f^h = \overline{\mathbf{D}}_1^h + \Lambda_1^T \Lambda_1 \overline{\mathbf{D}}_z^h = \mathbf{G}^T\mathbf{G}\overline{\mathbf{D}}_f^{h+1} + \Lambda_1^T \Lambda_1 \overline{\mathbf{D}}_z^h. \tag{19}$$

From Eq.(16), it can be seen that the larger $\overline{\mathbf{D}}_2^h$ is, the less $\hat{\mathbf{w}}_2^h$ is considered. Equivalently, according to Eq.(17), the larger $\overline{\mathbf{D}}_z^h$ is, the less $\hat{\mathbf{z}}^h$ contributes to the intermediate estimate. Moreover, from Eq.(17), it is clear that less than half of $\hat{\mathbf{z}}^h$ is considered in each refinement. This is particularly helpful to ensure accuracy when sensor noise is large. A further improvement of accuracy is achieved by performing the refinement within the same scale. Finally, according to Eq.(17) and Eq.(19), it can be seen that in the adaptive scheme, both estimates and variances are updated in the refining procedure.

2.4. The Adaptive Scheme: Rule 2

We modify Eq.(17) to generatethe following adaptive rule:

$$\begin{aligned} \hat{\mathbf{w}}_f^h &= \hat{\mathbf{w}}_1^h + \overline{\mathbf{D}}_1^h(\overline{\mathbf{D}}_1^h + \overline{\mathbf{D}}_z^h)^{-1}\hat{\mathbf{z}}^h \\ &= \mathbf{G}\hat{\mathbf{w}}_f^{h+1} + \Lambda_2 \hat{\mathbf{z}}^h, \end{aligned} \tag{20}$$

where $\Lambda_2 = \overline{\mathbf{D}}_1^h(\overline{\mathbf{D}}_1^h + \overline{\mathbf{D}}_z^h)^{-1}$. In this rule, the upper limit of the proportion of an increment being considered is 1 instead of 0.5 shown in rule 1. The limit occurs when the diagonal components in $\overline{\mathbf{D}}_z^h$ are approaching zeros. The variance of $\hat{\mathbf{w}}_f^h$ in this rule can be computed by replacing Λ_1 by Λ_2 in Eq.(19).

2.5. Same-Scale Updating

The above rules are also suitable for refining velocities within the same scale. This can be achieved by considering the model in Eq.(8) and ignoring the matrix $\mathbf{G}$ in rules 1 and 2.

3. EVALUATIONS

We examine the performance of the proposed scheme and the standard scheme described in Section 2.1 using artificial sequences with ground truth data: the **Office** sequence [7] and the **Yosemite** sequence [8]. To simulate noisy environments, Gaussian noise is added to the sequences with a predefined Signal Noise Ratio (SNR). To simulate different motion speeds for the **Yosemite** sequence, we estimate the optical flows between the 8^{th} frame and the 9^{th}, 10^{th}, 11^{th}, 12^{th} and 13^{th} frames. The ground truth data for each pair of images are computed by summating all ground truth data between the starting frame and the ending frame. This results in maximum speeds of 5.4, 11, 16.5, 22.1, and 27.8 pixels/frame, respectively. Because ground truth data are only available for the mountain region[1], the sky region is cut from the **Yosemite** sequence. Similarly, for the **Office** sequence, we estimate optical flows between the 10^{th} frame and the 11^{th}, 15^{th}, 20^{th}, 25^{th} and 30^{th} frames, respectively. The correspondent maximum speeds are 1.5, 7.6, 15.5, 23.8 and 32.3 pixels/frame. In the experiments, a 2D separable Gaussian filter with a standard deviation of 1 pixel is used to construct the pyramids. Spatial derivatives are calculated using the 4-tap kernel [8]. No pre-smoothing procedure is adopted within either the spatial or temporal domains. Accuracy is measured using the Mean Angular Error (MAE) [8].

In the first experiment, we examine accuracy for different speeds. We set SNRs to 20 dB and 5 dB, respectively, to simulate the cases where images are slightly and severely affected by noise. Examples of the noisy frames are shown in Fig.1. Here a 5-level pyramid is adopted. The size of the region performing the LS regression is set to 9×9 pixels for all levels. $\hat{\mathbf{w}}_f^h$ is also recursively refined within the same scale (4 times in our experiment). The results are shown in Fig.2. It can be seen that the proposed rules provide more accurate estimates than the standard scheme for different speeds. The results also demonstrate that rule 1 outperforms rule 2 in all conditions. The results indicate that, when using a 5-level pyramid, small motions are more difficult to estimate than large ones. The adoption of the proposed rules, especially rule 1, guarantees better accuracy than the standard scheme for small motions. In the second experiment, we examine the influence of number of pyramid levels on estimation accuracy. Here we use pyramids with different numbers of levels to estimate optical flows between the 10^{th} and the 11^{th} frames in the **Office** sequence and optical flows between the 8^{th} and the 9^{th} frames in the **Yosemite** sequence. The corresponding maximum speeds are 1.5 pixels/frame and 5.4 pixels/frame, respectively. The results are shown in Fig.3. For small motions, the optimal number of levels for a single scheme can be affected by several factors, such as motion structure and noise condition. For example, in the Yosemite sequence with velocities of up to 5.4 pixels/frame, the MAE of rule 1 reaches a minimum with 3 levels when SNR=20 dB and with 4 levels when SNR=5 dB. Such a difference suggests that additional

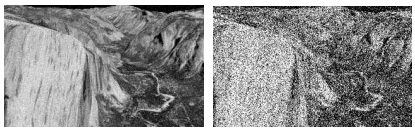
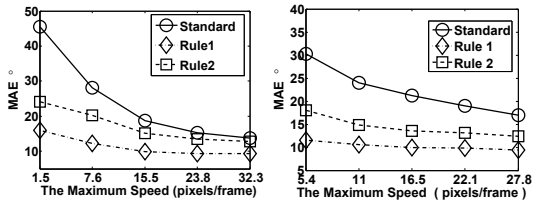
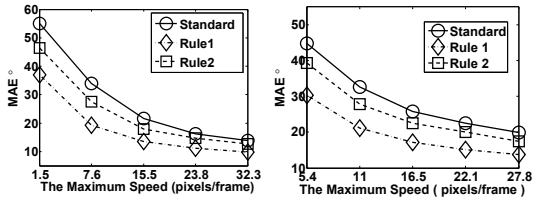


Fig. 1. Noisy images. Left: SNR=20 dB; Right: SNR=5 dB.

[1] http://www.cs.brown.edu/people/black/

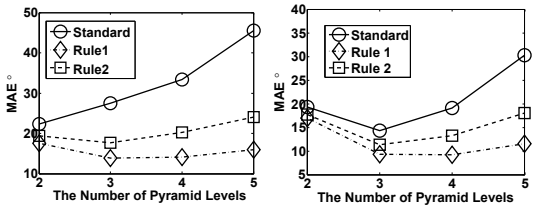


(a) **Office** (SNR=20 dB) (b) **Yosemite** (SNR=20 dB)

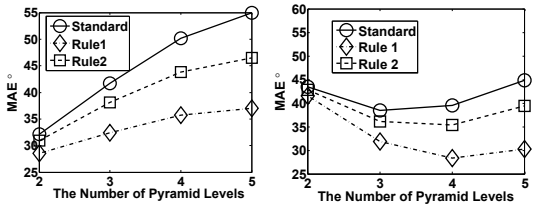


(c) **Office** (SNR=5 dB) (d) **Yosemite** (SNR=5 dB)

Fig. 2. The MAEs for different moving speeds.



(a) **Office** (SNR=20 dB) (b) **Yosemite** (SNR=20 dB)



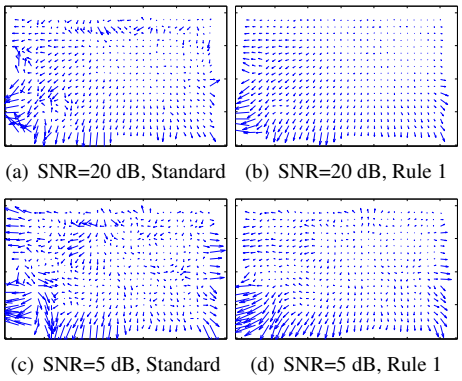
(c) **Office** (SNR=5 dB) (d) **Yosemite** (SNR=5 dB)

Fig. 3. The MAEs for different numbers of pyramid levels

levels can further reduce the influence of noise. However, the slight accuracy reduction when using 5 levels appears to show that using a larger number of pyramid levels can lead to an accumulation of estimation uncertainty. Nevertheless, the results show that when using a large number of pyramid levels, the proposed rules, especially rule 1, provides more accurate estimates for small motions than the standard scheme. Examples of optimal flow estimates are shown in Fig.4.

4. CONCLUSIONS

We have presented a novel adaptive multiscale scheme to improve the accuracy of optical flow estimation. The proposed scheme models estimation uncertainties of intermediate estimates in the cross-scale and same-scale refinements. The uncertainties are used as weights to reduce the influence of unreliable increments on estimation accuracy. Our experiments show that the proposed scheme provides more accurate estimates than the standard scheme for both small and large motions. It also provides more robust estimates with small displacements as the number of pyramid levels increases.



(a) SNR=20 dB, Standard (b) SNR=20 dB, Rule 1

(c) SNR=5 dB, Standard (d) SNR=5 dB, Rule 1

Fig. 4. Optical flow estimates between frame 8 and frame 9 of the **Yosemite** sequence.

5. REFERENCES

[1] F. G. Meyer and P. Bouthemy, "Region-based tracking using affine motion models in long image sequences," *CVGIP: Image Understanding*, vol. 60, no. 2, pp. 119–140, 1994.

[2] E. P. Simoncelli, "Bayesian multi-scale differential optical flow," in *Handbook of Computer Vision and Applications*, vol. 4, pp. 397–422. 1998.

[3] P. J. Burt and E. H. Adelson, "The laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, vol. Com-31, pp. 532–540, 1983.

[4] J. M. Odobez and P. Bouthemy, "Robust multiresolution estimation of parametric motion models," *Journal of Visual Communication and Image Representation*, vol. 6, no. 4, pp. 348–365, 1995.

[5] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," *IJCAI*, pp. 674–679, 1981.

[6] S. Weisberg, *Applied Linear Regression*, Wiley, New York, 1980.

[7] B. McCane, K. Novins, D. Crannitch, and B. Galvin, "On benchmarking optical flow," *Computer Vision and Image Understanding*, vol. 84, no. 1, pp. 126–143, 2001.

[8] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *IJCV*, vol. 12:1, pp. 43–77, 1994.