# Robust Motion Estimation Using Chrominance Information in Colour Image Sequences

Julian Magarey, Anil Kokaram, and Nick Kingsbury

Signal Processing and Communications Laboratory
Cambridge University Engineering Department
Cambridge CB2 1PZ, United Kingdom

**Abstract.** This paper describes a method for incorporating the chrominance information when estimating motion in a colour image sequence. It is based on a Maximum-Likelihood (ML) formulation of the motion estimation problem which assumes homogeneous additive Gaussian noise in each colour component, with known inter-field correlation statistics. It defines a noise-decorrelating colour space transform which provides a simple implementation of the ML formulation. Results for noisy synthesised colour sequences with known motion and noise statistics demonstrate the superiority of the exact ML formulation over straightforward, unweighted three-component estimation, most noticeably in high noise conditions.

## 1 Introduction

Colour video signals consist of both luminance (intensity) and chrominance information. The chrominance has two degrees of freedom, so a full colour signal consists of three *fields* for every frame. There are a variety of representations of these three fields, most commonly defined as some linear transformation of the red-green-blue (RGB) basis which is rooted in the human visual system for analysing colour. For example, the YUV colour space is defined as [2, page 67]

$$\begin{bmatrix} y \\ u \\ v \end{bmatrix} = \begin{bmatrix} 0.3 & 0.6 & 0.1 \\ -0.15 & -0.3 & 0.45 \\ 0.4375 & -0.375 & -0.0625 \end{bmatrix} \begin{bmatrix} r \\ g \\ b \end{bmatrix} \tag{1}$$

where $y$ is the luminance component and the $u$ and $v$ fields contain chrominance information (the matrix entries vary somewhat depending on the visual constants used). Most algorithms for motion estimation of colour sequences have worked in YUV-space and ignored the chrominance components.

It was suggested by Mitiche et al. [7] that chrominance could be used as an additional cue in a multiconstraint method for estimating motion between video frames, though they reported no results which incorporated chrominance. Konrad and Dubois [4] extended their original maximum likelihood (ML) motion estimation framework for scalar signals to encompass three component fields. To do this, they defined a quantity called the *vector* displaced pel difference (DPD) by analogy with the scalar DPD of monochrome estimation. They demonstrated

improvements in the quality of the colour-based motion field over the purely luminance-based field for synthetic and real colour sequences. Their method assumed that the colour fields contribute equally and independently to a single motion estimate field. This is implicitly based in turn on the assumption that the noise in each field is additive Gaussian, uncorrelated and equivariant. In their paper on the use of colour in video resolution enhancement, Tom and Katsaggelos [8] used the same implicit assumption.

However, this cannot be relied on in practice. For example, if noise is uncorrelated in YUV-space, it will be far from uncorrelated in RGB-space. If there is no knowledge of the original image-gathering system and its noise properties, unweighted three-component estimation could go seriously wrong. This paper describes the true ML estimator for vector image sequences which allows for correlated additive noise in the three colour components. The ML formulation may be applied to extend the standard region-based matching and gradient-based algorithms. We then define a noise-decorrelating transform, akin to the Karhunen-Loeve transform, into an "optimal" colour space. If this transform is applied to the three original colour fields, the true ML estimate may be found using the common implicit assumption of equal and independent contributions by the three transformed components.

Our test results, obtained on colour sequences with synthesised motion and correlated additive noise, demonstrate the superiority of the optimal method over estimation based on equal and independent contributions from the each of the red, green, and blue fields. The improvement becomes more noticeable as the amount of noise increases. Our results also suggest that the most efficient strategy would be adaptive, based primarily on luminance and only incorporating chrominance appropriately where required.

## 2 The ML Motion Estimator for Colour Signals

We follow the formulation of Konrad and Dubois [4] in setting up the ML estimator for a three-component sequence. The vector $\mathbf{u} = [u_1, u_2, u_3]^T$ represents the true or underlying image, with its three colour components, which would be obtained by a noise-free optical system. The sequence $\{\mathbf{u}_n, n \in \mathcal{Z}\}$ represents the sequence of images sampled at integer time instants. We use the common assumption of intensity conservation, which requires that intensity *in each component* is constant along the motion trajectory defined by the displacement $\hat{\mathbf{d}}(\mathbf{x})$ at pel $\mathbf{x}$:

$$\mathbf{u}_n(\mathbf{x}) = \mathbf{u}_{n-1}(\mathbf{x} - \hat{\mathbf{d}}(\mathbf{x})) \tag{2}$$

Taking into account observation noise, we can rewrite (2) using the *observed* signal sequence $\{\mathbf{g}_n, n \in \mathcal{Z}\}$ as

$$\mathbf{g}_n(\mathbf{x}) - \mathbf{g}_{n-1}(\mathbf{x} - \hat{\mathbf{d}}(\mathbf{x})) = \mathbf{e}_n(\mathbf{x}) \tag{3}$$

where $\mathbf{e}_n$ is the differential noise vector at frame $n$, which has twice the variance of the individual frame noise. Equation (3) may be rewritten as

$$\mathbf{DPD}(\mathbf{x}, \hat{\mathbf{d}}(\mathbf{x})) = \mathbf{e}_n(\mathbf{x}) \tag{4}$$

having replaced its left hand side with the *vector displaced pel difference.*

Our aim is to estimate the translation model parameter $\hat{\mathbf{d}}$ based on the observations $\mathbf{g}_{n-1}$ and $\mathbf{g}_n$. The Maximum Likelihood (ML) estimate is defined as

$$\hat{\mathbf{d}}(\mathbf{x}) = \arg\max \{p(\mathbf{g}_n(\mathbf{x})|\mathbf{d}, \mathbf{g}_n(\mathbf{x}))\} \tag{5}$$

If we assume zero-mean Gaussian noise statistics, we can write the joint probability density function (pdf) of $\mathbf{e}(\mathbf{x})$ as

$$p(\mathbf{e}(\mathbf{x})) \propto \exp\left(-\frac{1}{2}\mathbf{e}^T(\mathbf{x})R^{-1}(\mathbf{x})\mathbf{e}(\mathbf{x})\right) \tag{6}$$

where $R(\mathbf{x})$ is the 3-by-3 covariance matrix characterising the interaction of the three noise components at pel $\mathbf{x}$. The likelihood is given by the vector noise joint pdf. Combining this fact with (4), we can write

$$p(\mathbf{g}_n(\mathbf{x})|\mathbf{g}_{n-1}(\mathbf{x}), \mathbf{d}) \propto \exp\left(-\frac{1}{2}\mathbf{DPD}^T(\mathbf{x}, \mathbf{d})R^{-1}(\mathbf{x})\mathbf{DPD}(\mathbf{x}, \mathbf{d})\right) \tag{7}$$

so the ML estimator becomes

$$\hat{\mathbf{d}}(\mathbf{x}) = \arg\min \{\mathbf{DPD}^T(\mathbf{x}, \mathbf{d})R^{-1}(\mathbf{x})\mathbf{DPD}(\mathbf{x}, \mathbf{d})\} \tag{8}$$

### 2.1 ML Estimation Using Region-based Matching

To obtain a more robust estimate of $\hat{\mathbf{d}}$, an assumption of *constant local flow* over a region of pels $\Omega = \{\mathbf{x}_i, i = 1, \ldots, N\}$ is commonly invoked. This assumption is approximately valid if the motion field is continuous and the regions are not too large. If we define the $3N$-element *displaced region difference* vector as

$$\mathbf{DRD}(\Omega, \mathbf{d}) = \left[\mathbf{DPD}(\mathbf{x}_1, \mathbf{d}) \ldots \mathbf{DPD}(\mathbf{x}_N, \mathbf{d})\right]^T \tag{9}$$

we can estimate $\hat{\mathbf{d}}$ over the region $\Omega$ as

$$\hat{\mathbf{d}}(\Omega) = \arg\min \{\mathbf{DRD}^T(\Omega, \mathbf{d})R_\Omega^{-1}\mathbf{DRD}(\Omega, \mathbf{d})\} \tag{10}$$

where $R_\Omega$ is the $3N$-by-$3N$ noise component covariance matrix over the region $\Omega$.

Equation (10) may be simplified by the assumption that there is no noise correlation between different pels in the region $\Omega$. This gives $R_\Omega$ a block diagonal structure, where each block is the 3-by-3 matrix $R$ of component noise covariance of (6) (the $\mathbf{x}$ argument may be dropped if we further assume homogeneity, i.e. position-independence). When $R_\Omega$ has this structure, so too does its inverse, with each block equal to $R^{-1}$. The ML estimator becomes

$$\hat{\mathbf{d}}(\Omega) = \arg\min \left\{\sum_{i=1}^{N} \mathbf{DPD}^T(\mathbf{x}_i, \mathbf{d})R^{-1}\mathbf{DPD}(\mathbf{x}_i, \mathbf{d})\right\} \tag{11}$$

Equation (11) shows how to find $\hat{\mathbf{d}}$ by an exhaustive search over a set of $\mathbf{d}$ candidates. This is the optimal *region-matching* strategy in the presence of correlated component noise.

If, furthermore, the noise in each component is uncorrelated, $R$ becomes diagonal:

$$R = \text{diag}\ (\sigma_k^2, k = 1, 2, 3) \tag{12}$$

In this case (11) becomes

$$\hat{\mathbf{d}}(\Omega) = \arg\min \left\{ \sum_{i=1}^{N} \sum_{k=1}^{3} \frac{1}{\sigma_k^2} DPD_k^2(\mathbf{x}_i, \mathbf{d}) \right\} \tag{13}$$

which now involves minimising the sum, over the region, of the squared (scalar) DPDs of each component, weighted by the inverse of the noise variance. This is the formulation obtained by Konrad and Dubois [4]. In effect the contribution of each component to the estimate is weighted by the SNR of the corresponding difference image.

### 2.2 Gradient-based ML Estimation

An approximate solution to the region-based vector ML estimator (10) may be found by expanding $\mathbf{g}_{n-1}(\mathbf{x}_i - \mathbf{d})$ around $\mathbf{x}_i$ using a first-order Taylor series:

$$\mathbf{g}_{n-1}(\mathbf{x}_i - \mathbf{d}) \approx \mathbf{g}_{n-1}(\mathbf{x}_i) - \left(\nabla \mathbf{g}_{n-1}(\mathbf{x}_i)\right)^T \mathbf{d} \tag{14}$$

where

$$\nabla \mathbf{g}_{n-1}(\mathbf{x}_i) = \begin{bmatrix} \frac{\partial}{\partial x} g_{1,n-1}(\mathbf{x}_i) & \frac{\partial}{\partial y} g_{1,n-1}(\mathbf{x}_i) \\ \frac{\partial}{\partial x} g_{2,n-1}(\mathbf{x}_i) & \frac{\partial}{\partial y} g_{2,n-1}(\mathbf{x}_i) \\ \frac{\partial}{\partial x} g_{3,n-1}(\mathbf{x}_i) & \frac{\partial}{\partial y} g_{3,n-1}(\mathbf{x}_i) \end{bmatrix}^T \tag{15}$$

This approximation allows a closed-form least-squares solution to be found for $\mathbf{d}$, called the *gradient-based* ML estimator:

$$\hat{\mathbf{d}}(\Omega) = (G^T R_\Omega^{-1} G)^{-1} G^T R_\Omega^{-1} \mathbf{z} \tag{16}$$

where

$$G = \left[ \nabla \mathbf{g}_{n-1}(\mathbf{x}_1) \ldots \nabla \mathbf{g}_{n-1}(\mathbf{x}_N) \right]^T \tag{17}$$

$$\text{and } \mathbf{z} = \left[ \mathbf{g}_{n-1}(\mathbf{x}_1) - \mathbf{g}_n(\mathbf{x}_1) \ldots \mathbf{g}_{n-1}(\mathbf{x}_N) - \mathbf{g}_n(\mathbf{x}_N) \right]^T \tag{18}$$

As with region-based matching, the assumption that noise at different pels is uncorrelated and homogeneous simplifies the computation in (16). In practice this method is severely limited in its measurement range because of the neglect of higher order terms in (14). The range may be increased by using an iterative approach which uses an equation similar to (16) to compute updates to an initial estimate of $\hat{\mathbf{d}}$ [3].

### 2.3 A Decorrelating Transform

Equation (1) defines the transform from the RGB colour space to the YUV space. A general linear colour space transform on RGB space may be written as

$$\mathbf{g} = C \begin{bmatrix} r \\ g \\ b \end{bmatrix} \tag{19}$$

for an $n$-by-3 matrix $C$. In the new colour space ("C-space"), the inter-component noise correlation matrix becomes

$$R_C = C R_{rgb} C^T \tag{20}$$

Clearly if we can find a matrix $C$ such that $R_C = \sigma^2 I$ for some $\sigma$, the vector ML gradient-based estimator (16) and region-based matching estimator (11) revert to straightforward formulations in which the colour components make equal and independent contributions to the final estimate, as postulated by Konrad and Dubois [4].

To find such a *noise-decorrelating* transform, given $R_{rgb}$ (or $R_{yuv}$, in which case we apply the transform to the YUV fields), we can use singular value decomposition (SVD). Because $R_{rgb}$ is square, symmetric and non-negative definite [2, page 33], it is orthogonally diagonalisable with non-negative eigenvalues:

$$R_{rgb} = V D V^T \tag{21}$$

where $D$ is diagonal with non-negative entries, and $V$ is orthogonal. The number of non-zero eigenvalues is the rank $n$ of $R_{rgb}$. The case $n < 3$ results when at least one component is a linear combination of the others. In this (exceptional) case, when $R_{rgb}$ is non-invertible, (11) and (16) may not be used. The SVD-based method will identify this case and project to a colour space of appropriately reduced dimensionality, thus saving computation. This is done by extracting the invertible $n$-by-$n$ portion $D'$ of $D$ and the corresponding rows $V'$ of $V$. Setting

$$C = (\sqrt{D'})^{-1} V'^T \tag{22}$$

guarantees that $C R_{rgb} C^T = I_n$ as desired. This procedure is similar to the Karhunen-Loeve Transform for compressing images [2, page 163], except it is carried out using noise rather than signal statistics.

## 3 Tests on Synthesised Sequences

The synthesised test sequences were obtained by applying motion fields of three distinct kinds—uniform translation, rotation, and divergence—to the 128-by-128 pel central portions of frame 1 of the "carphone", "foreman", and "suzie" colour sequences respectively.

To add correlated noise, we first found three 128-by-128 uncorrelated, equivariant white Gaussian noise images with variance $\sigma^2$. An invertible 3-by-3 matrix

$M$ was used to transform the noise fields into $M^{-1}$-space; the transformed noise was added to the RGB signal fields. This procedure gave

$$R_{rgb} = \sigma^2 M^{-1}(M^{-1})^T = \sigma^2 \begin{bmatrix} 1.7393 & 0.1871 & -0.1886 \\ 0.1871 & 0.1318 & -0.0742 \\ -0.1886 & -0.0742 & 0.3654 \end{bmatrix} \quad (23)$$

This particular $M$ was chosen to give a clearly non-diagonal $R_{rgb}$ in order to best illustrate the potential improvement from the proposed approach over unweighted RGB-space estimation.

A modified version of the iterative gradient-based algorithm was used. The algorithm includes a stabilising term in the matrix-inversion step of the update (16) to give better convergence performance and robustness during iteration. To further increase the measurement range, the algorithm was implemented *hierarchically*, based on a 4-level Gaussian pyramid decomposition—see [3] for details. The final full-density field was obtained by bilinear interpolation from the field of region vectors.

To measure the accuracy of the full-density motion field, we used Fleet and Jepson's angular measure of error [1], averaged over the field excluding a strip of width 16 pels around the boundary of the image. This is is akin to a relative measure of error, except it does not give undue weighting to errors in very small motion vectors as relative error would.

For each of the three test sequences, three sets of results were obtained: those using unweighted RGB-space ME; those from luminance-only estimation; and those from vector ML estimation, obtained by first transforming from RGB to the optimal colour space. Figure 1 shows these results, plotted as mean error angle against $\sigma$. In each case the optimal strategy is clearly the most robust to noise. The difference between the RGB and optimal strategies is illustrated in Fig. 2, which shows the lower right portion of the motion fields at $\sigma = 36$ for the rotation sequence, superimposed on images of error angle (darker means greater error.) The improvement using the optimal strategy under conditions of high, correlated noise is clear.

The results also show that at low noise levels, there is little to be gained by using the optimal approach as opposed to RGB-space ME. Furthermore, luminance-only estimation loses little by comparison with the optimal strategy at the lower noise levels. These results have been repeated for the complex-wavelet-domain ME algorithm of Magarey and Kingsbury [5, 6].

Luminance-only ME requires only slightly more than a third of the amount of computation required for full-colour ME. Our results suggest that this strategy provides near-optimal performance except where noise overwhelms luminance contrast. In such cases, chrominance information may be incorporated (according to the ML formulation) to increase the robustness of the estimates. An adaptive strategy, in which luminance is the primary quantity for estimation, with some criterion to indicate where the chrominance information should be incorporated, would provide the best tradeoff between accuracy and efficiency. Our on-going

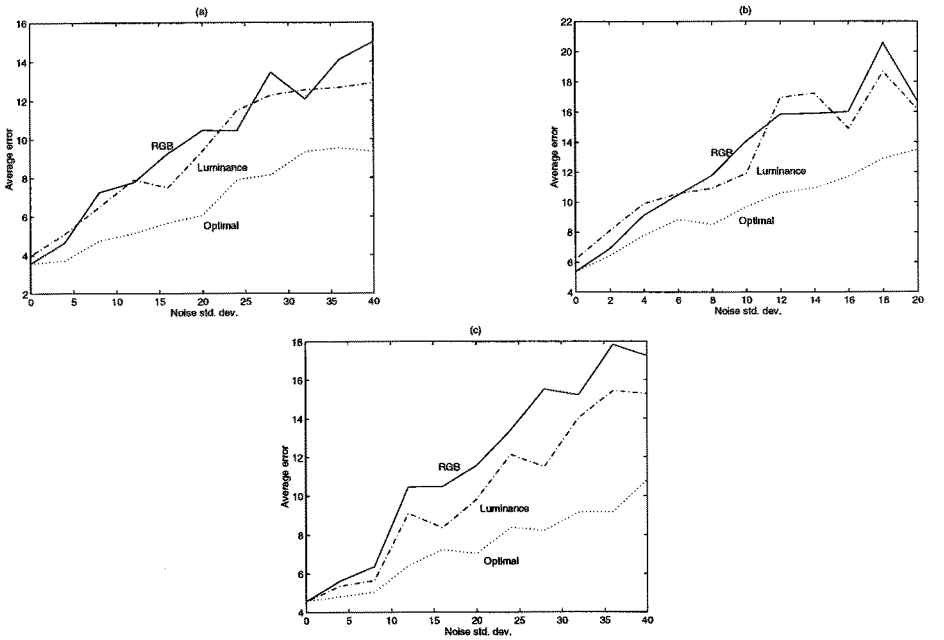


**Fig. 1.** Mean motion field error vs $\sigma$ for RGB, luminance-only, and optimal estimation. (a) Translation sequence. (b) Divergence sequence. (c) Rotation sequence.

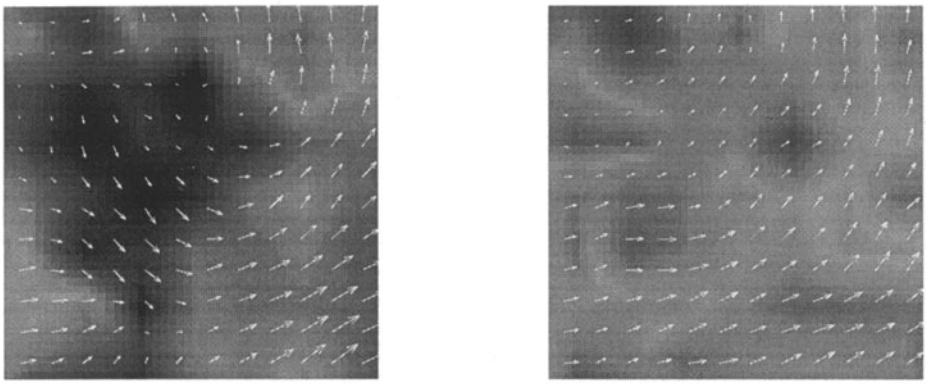

**Fig. 2.** Lower right portion of motion fields superimposed on error angle images (darker means greater error.) Sequence: rotation, with correlated noise $\sigma = 36$. (Left) RGB-estimated field (one estimate per 4 by 4 pels). (Right) Optimally-estimated field (same resolution).

work is aimed at finding a general technique for characterising the noise from typical video sources, and finding a criterion for incorporation of chrominance information.

## 4 Conclusion

In this paper we have shown how to formulate the ML motion estimator for colour image sequences in the presence of correlated, homogeneous Gaussian noise in the three component fields. Vector ML formulations for region-matching and gradient-based approaches were given in terms of the inter-component noise covariance matrix. If the covariance matrix is diagonal, each component contributes independently to the ML estimate. We have shown how to define a linear transformation into a new colour space in which the noise covariance is the identity matrix. In the new, "optimal" colour space, the components may be treated as equal and independent contributors to the ML estimate. The effectiveness of the optimal colour space transformation was demonstrated for a modified gradient-based algorithm applied to three synthesised test sequences containing additive noise with deliberately induced covariance. Our tests also showed that luminance-only estimation performs reasonably well by comparison with the more expensive full-colour approach, particularly at low noise levels. This suggests that the best strategy for a general colour sequence would be adaptive.

## References


1. D.J. Fleet and A.D. Jepson. Computation of component image velocity from local phase information. *Intern. J. Comput. Vis.*, 5:77–104, 1990.
2. A.K. Jain. *Fundamentals of Digital Image Processing.* Prentice-Hall International, 1989.
3. A. C. Kokaram and S. J. Godsill. A system for reconstruction of missing data in image sequences using sampled 3D AR models and MRF motion priors. In *Computer Vision - ECCV '96*, volume II, pages 613–624. Springer Lecture Notes in Computer Science, April 1996.
4. J. Konrad and E. Dubois. Use of colour information in Bayesian estimation of 2-D motion. In *Proc. ICASSP*, pages 2205–2208. IEEE, 1990.
5. J.F.A. Magarey. *Motion estimation using complex wavelets.* PhD thesis, Cambridge University Department of Engineering, 1997.
6. J.F.A. Magarey, A.C. Kokaram, and N.G. Kingsbury. Optimal schemes for motion estimation using colour image sequences. In *Proc. Int. Conf. On Image Processing*, October 1997. *(To appear).*
7. A. Mitiche, Y.F. Wang, and J.K. Aggarwal. Experiments in computing optical flow with the gradient-based, multiconstraint method. *Pattern Recognition*, 20:173–179, 1987.
8. B.C. Tom and A.K. Katsaggelos. Resolution enhancement of colour video. In *Proc. EUSIPCO-96*, pages 145–148, September 1996.