Joint estimation of motion and illumination variations for coding of image sequences

H. Nicolas *, J. Konrad **, C. Labit *

* IRISA/INRIA, Campus de Beaulieu 35042 Rennes Cedex, France
Tel: (33) 99 84 71 00, Fax: (33) 99 38 38 32

**INRS-Télécommunications, 16 Place du Commerce
Ile-des-Soeurs, Verdun, Canada, H3E 1H6
Tel: (514) 765-7768, Fax: (514) 761-8501

Abstract

This paper describes a new approach to the problem of motion estimation for the coding of image sequences. The goal is to obtain an efficient description (parametrization) of temporal variations between two successive images in a sequence. To achieve this we propose to use the standard hypothesis of luminance constancy along a motion trajectory simultaneously introducing a polynomial representation of illumination variations. The estimation process consists of two iteratively alternating stages: a region-based estimation of apparent 2D motion parameters and an estimation of 2D illumination variations. Such an approach reduces the residual reconstruction error after motion compensation due to improved estimation of motion parameters.

1 Introduction

In the coding of image sequences, apparent motion permits the reconstruction of an image through motion compensation; only the preceding image and motion information have to be known. The information about pixel movements may be local, as in the case of pel-recursive methods, or global, as in correspondence methods. Usually, it is assumed in motion estimation/analysis that temporal variations between two successive images in a sequence are due to object motion and occlusion effects in the original 3D scene [1],[2]. Under this hypothesis, two similar regions from images at times t and $t+1$ can be matched using motion models, for example through a displacement and a transformation (e.g., zoom).

In practice, the hypothesis of luminance invariance along a motion trajectory rarely holds since there exist other, than object motion, sources that induce temporal variations between two images (e.g., lighting, noise). Thus, although often used, this hypothesis may not lead to reliable motion estimates and consequently may reduce the efficiency of motion compensation and the quality of reconstructed images.

The goal of this study is to re-examine the hypothesis of luminance invariance in motion estimation in the context of image sequence coding. More precisely, the improvement of the reconstructed image quality without a significant increase in bit rate is the target. We hope to achieve this by augmenting the constant luminance model with new parameters that take into account luminance variability of a point along its motion trajectory. We propose a mean-squared reconstruction error as the minimization criterion. The algorithm developed has been tested on various image sequences and shown to improve results when compared with constant luminance model. Here, we show results for one real sequence.

The paper is divided as follows. In Sections 2 and 3 models and estimation method for 2D motion are described. Sections 4 and 5 introduce a model for illumination variations and a method to calculate parameters of such a model. Section 6 presents some simulation results.

2 Models for apparent 2D motion

Two different motion models are used in this study. The first model is the constant motion model ($M1$):

$$\vec{d}_\Phi = \begin{bmatrix} d_x \\ d_y \end{bmatrix} = \begin{bmatrix} t_x \\ t_y \end{bmatrix} \tag{1}$$

where (d_x, d_y) are displacements of a pixel at location $\vec{p} = (x, y)$. This model is characterized by translation parameters (t_x, t_y).

To account better for more complex motion, a model incorporating divergence and rotation is needed (see also [3],[4]). Such a model, called a simplified linear model ($M2$), is expressed as follows:

$$\vec{d}_\Phi = \begin{bmatrix} d_x \\ d_y \end{bmatrix} = \begin{bmatrix} t_x + k.(x - x_g) - \theta.(y - y_g) \\ t_y + \theta.(x - x_g) + k.(y - y_g) \end{bmatrix} \tag{2}$$

where k, θ are divergence and rotation parameters, respectively, and (x_g, y_g) are coordinates of the center of gravity of a region under consideration.

Let Φ be a parameter vector for a given model. Then, $\Phi_{M1} = (t_x, t_y)^t$ and $\Phi_{M2} = (t_x, t_y, k, \theta)^t$ are parameter vectors for models $M1$ and $M2$, respectively.

At this point we would like to stress that the method of estimating the illumination variations is independent of the motion model presented above.

3 Estimation of motion

Motion estimation used in this study is based on interleaved iterations of a gradient method and of deterministic relaxation. The estimation process is initialized using motion parameters from the preceding frame. To identify regions over which motion models are defined, quadtree image segmentation is used. Such a segmentation is simple in implementation and requires low bit rate for transmission. Initial segmentation uses square regions of 32×32 pixels. Later in the process, a division of regions that are insufficiently compensated is carried out.

The criterion used to estimate motion is the motion-compensated pixel difference defined for pixel at $\vec{p}$ and time t as follows [5]:

$$DFD(\vec{p}, \vec{d}) = I_{t+1}(\vec{p}) - \tilde{I}_t(\vec{p} - \vec{d}) \tag{3}$$

where $\tilde{I}_t$ represents the luminance value at point $(\vec{p} - \vec{d})$ obtained by bilinear interpolation of the discretized image I_t. In practice this error is rarely zero for two reasons:

- motion estimation is not exact in the sense of the optimization criterion (the global minimum of the error function is not perfectly achieved),
- the global minimum of the error function is not zero (this minimum depends on the motion model and segmentation used).

The first point above expresses the dependence of motion estimation on the minimization algorithm. In various experiments we have observed that the achieved minima are very close to the global minimum of the error function. Thus, one cannot expect to improve the results substantially by just modifying the minimization algorithm. The second point suggests that the error function above may not be suitable for the task of estimating motion under illumination variations. Thus, we modify in the next section this error function by incorporating a more complex model of temporal variations between images. In order to achieve this we propose an illumination variations model.

For more details on the motion estimation method used, please consult [4] and [6] (also see Fig. 1).

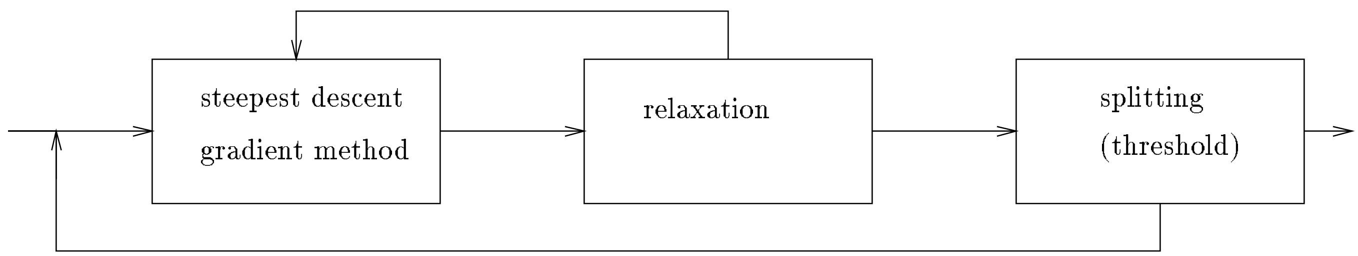

Figure 1: Block diagram of apparent 2D motion estimation.

4 Model of illumination variations

If a displacement estimate $\hat{d}$ for pixel at location $\vec{p}$ is known, the reconstruction error equals $DFD(\vec{p}, \hat{d})$ with DFD defined in (3). The estimate $\hat{d}$ is calculated by a suitable minimization of the squared reconstruction error

$$\hat{d} = \arg\min_{\vec{d}}[DFD^2(\vec{p}, \vec{d})]. \tag{4}$$

We assume that the estimated displacement $\hat{d}$ is sufficiently close to the true motion, and consequently that the reconstruction error $DFD(\vec{p}, \hat{d})$ is due only to illumination variations and perhaps to some occlusion effects. We cannot separate these two effects, as we do not explicitly model occlusions. Note that the vector $\hat{d}$ is often different from a projection of the 3D motion vector onto the image plane [7].

4.1 Variation of apparent illumination over a region

Given motion model Φ, for each region ϑ_j we can estimate model parameters by executing the following minimization:

$$\hat{\Phi} = \arg\min_{\Phi}[\sum_{\vec{p}\in\vartheta_j} DFD^2(\vec{p}, \vec{d}_\Phi)]. \tag{5}$$

The optimal parameter vector $\hat{\Phi}$ does not give, however, a zero reconstruction error over region ϑ_j

$$\sum_{\vec{p} \in \vartheta_j} DFD^2(\vec{p}, \vec{d}_{\hat{\Phi}}) \neq 0 \tag{6}$$

due to the effects mentioned before. From the above relationships it is clear that modeling of variations of apparent illumination must depend on estimated motion as well as on image segmentation.

4.2 Polynomial approximation of illumination variations

The reconstruction error $DFD(\vec{p}, \vec{d}_{\hat{\Phi}})$ comprises errors due to real illumination variation as well as errors due to occlusion effects and noise. If the noise is assumed to possess Gaussian properties, then other sources of error, e.g., illumination effects, can be considered deterministic and thus can be modeled (at least partially) by a polynomial. Disregarding terms of order higher than 1, the reconstruction error $DFD(\vec{p}, \vec{d}_{\hat{\Phi}})$ for pixel at location $\vec{p}$ can be modeled as follows

$$\gamma(\vec{p}, \Gamma) = \gamma_0 + \gamma_1 \cdot (x - x_g) + \gamma_2 \cdot (y - y_g) + \alpha(\vec{p}), \tag{7}$$

where $\Gamma = (\gamma_0, \gamma_1, \gamma_2)^t$ is the parameter vector describing the illumination variation model. $\alpha(\vec{p})$ is the residual error for pixel at location $\vec{p}$ that depends not only on Γ but also on estimated motion parameters $\hat{\Phi}$ and region ϑ_j under consideration.

4.3 Error criterion

The predicted image can be calculated using parameter vectors Φ and Γ

$$\hat{I}_{t+1}(\vec{p}) = \lfloor \tilde{I}_t(\vec{p} - \vec{d}_{\Phi}) + \gamma(\vec{p}, \Gamma) \rfloor, \tag{8}$$

where $\lfloor x \rfloor$ is defined as follows:

$$\lfloor x \rfloor = \begin{cases} \text{integer part of } (x + \frac{1}{2}), & \text{if } 0 \leq x \leq 255, \\ 0, & \text{if } x < 0, \\ 255, & \text{if } x > 255. \end{cases} \tag{9}$$

In coding, the goal is to obtain the best possible quality of a reconstructed image given a bit rate. Thus, a quality measure or criterion must be proposed in order to carry out optimization. We think that a very reasonable criterion for each pixel at location $\vec{p}$ and time t is the following modified reconstruction error that incorporates the illumination variations model

$$\begin{aligned} E(\vec{p}) &= I_{t+1} - \hat{I}_{t+1} \\ &= I_{t+1} - \lfloor \tilde{I}_t(\vec{p} - \vec{d}_{\Phi}) + \gamma(\vec{p}, \Gamma) \rfloor. \end{aligned} \tag{10}$$

By accumulating pixel errors $E(\vec{p})$ over each region ϑ_j, and then by accumulating region errors over the whole image, mean-squared error can be calculated as follows:

$$MSRE = \frac{1}{N} \sum_{\vartheta_j \in S} \sum_{\vec{p} \in \vartheta_j} (I_{t+1} - \lfloor \tilde{I}_t(\vec{p} - \vec{d}_{\Phi}) + \gamma(\vec{p}, \Gamma) \rfloor)^2. \tag{11}$$

Note that above S represents segmentation of the image into regions and N is the total number of pixels in the image.

From the point of view of motion estimation, accounting for illumination variations causes modification of the reconstruction error on a region by region basis due to the introduction of the term $\gamma(\vec{p}, \Gamma)$. Thus, the minimum value of the error function is reduced or, in the worst case, unchanged. This modification results in:

1. Direct reduction of the reconstruction error due to the introduction of the term γ.
2. Increase in efficiency of motion compensation in regions where impact of illumination variations is substantial.
3. Increase of the number of parameters to be estimated; the estimation process becomes more difficult.

5 Estimation scheme

The estimation of illumination variations must be incorporated into the motion estimation scheme described in Section 3. Two approaches are possible:

1. joint estimation of motion and illumination variations using a global optimization method,
2. separate estimation of motion and of illumination variations through alternating (interleaved) processing.

We chose the second approach for two reasons. Firstly, fewer unknowns have to be found at each of the two stages, and thus each estimation should be easier to carry out separately. Secondly, illumination variations cannot be evaluated unless a feasible motion field is known. The best we can do is to estimate such a field without accounting for illumination variations. We have observed in experiments without the illumination model that the residual error is small and noise-like in numerous regions. It is substantial only where illumination variations are present (or occlusions, which are not considered here). Thus, we propose to estimate a motion field first without accounting for illumination variations. Then, we suggest to introduce *a posteriori* the phase of illumination variations estimation. Since illumination variations obtained should be significant only in the areas where the luminance invariance assumption has failed, they can be applied only locally at the subsequent motion estimation stage; homogeneous use of illumination variations is not needed. With the compensation of illumination effects, the subsequent motion estimation can be thought of as being applied in the context of the standard invariant intensity assumption. The general estimation diagram is presented in Fig. 2.

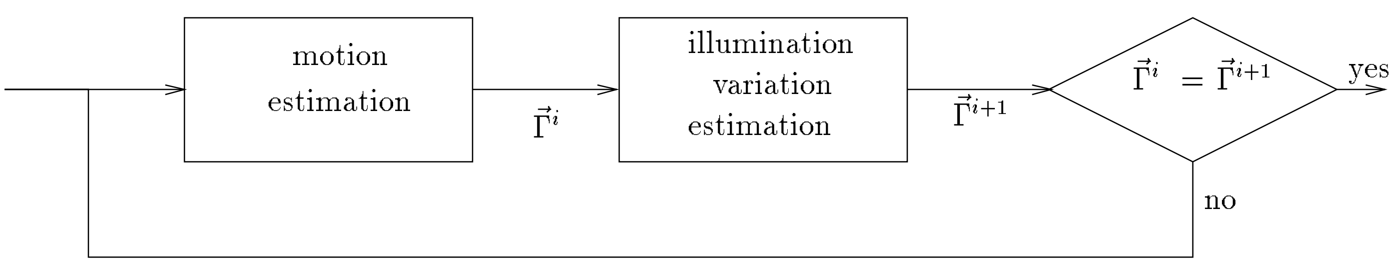

Figure 2: Block diagram for joint estimation of motion and illumination variations.

As shown in Fig. 2, new vector Φ is estimated for each region at each iteration by updating the previously calculated vector Φ. This is executed through the following update:

$$\Phi^{i+1} = \Phi^i + G^i(\Gamma^i) \tag{12}$$

where G^i is the gradient of the cost function (11) with respect to Φ. Once the new Φ has been estimated, and a segmentation has been carried out, the new vector Γ is calculated for each region. To do so, the following iterative update is executed:

$$\Gamma^{i+1} = \Upsilon(\Phi^{i+1}) \tag{13}$$

where Υ is a function to be derived by minimizing the mean-squared error (11) with respect to Γ. The iterative updates (12) and (13) are executed in an alternating fashion. Thus, a new vector Φ is expected to affect an estimate of Γ, and a new vector Γ should improve the new estimate of Φ. The iterative process stops if lack of change in estimated parameters is detected.

Taking the partial derivatives of (11) with respect to γ_0, γ_1, γ_2 and for each region ϑ_j (to obtain a derivable function, we eliminate function $\lfloor . \rfloor$ in equation (11)) , the following relationships can be obtained:

$$\begin{aligned}
&\sum_{\vec{p}\in\vartheta_j} [(DFD(\vec{p},\vec{d}) - \gamma_0) - \gamma_1(x - x_g) - \gamma_2(y - y_g)] = 0,\\
&\sum_{\vec{p}\in\vartheta_j} [(DFD(\vec{p},\vec{d}) - \gamma_0)(x - x_g) - \gamma_1(x - x_g)^2 - \gamma_2(x - x_g)(y - y_g)] = 0,\\
&\sum_{\vec{p}\in\vartheta_j} [(DFD(\vec{p},\vec{d}) - \gamma_0)(y - y_g) - \gamma_2(x - x_g)^2 - \gamma_1(x - x_g)(y - y_g)] = 0.
\end{aligned} \tag{14}$$

Since regions over which estimates are calculated are squares, we have $\sum_{\vec{p}\in\vartheta_j}(y - y_g) = \sum_{\vec{p}\in\vartheta_j}(x - x_g) = \sum_{\vec{p}\in\vartheta_j}(x - x_g)(y - y_g) = 0$, and hence the above equations can be rewritten as follows:

$$\begin{aligned}
\gamma_0 &= \frac{1}{N_j}\sum_{\vec{p}\in\vartheta_j} DFD(\vec{p},\vec{d}),\\
\gamma_1 &= \frac{\sum_{\vec{p}\in\vartheta_j} DFD(\vec{p},\vec{d})(x - x_g)}{\sum_{\vec{p}\in\vartheta_j}(x - x_g)^2}\\
\gamma_2 &= \frac{\sum_{\vec{p}\in\vartheta_j} DFD(\vec{p},\vec{d})(y - y_g)}{\sum_{\vec{p}\in\vartheta_j}(y - y_g)^2}
\end{aligned} \tag{15}$$

where N_j is the number of points in region ϑ_j. Note that equations above form a detailed version of equation (13). Corresponding equations for the estimation of motion can be found in [4] and [6].

Improvements due to the new estimate of illumination variations $\Gamma^{(i+1)}$ can be interpreted in two ways:

- immediate reduction of the reconstruction error,
- improved motion estimate; this improvement is obtained in the motion estimation loop (Fig. 2).

6 Results and conclusion

Results presented below have been obtained by applying the proposed method to the real image sequence "campagne" (original image shown in Fig. 3.a). This sequence has been acquired by a camera mounted on a moving car (divergent motion in the whole image). In the left part of the image, another car approaches the camera. Thus, the relative motion is substantial, especially within the car and the sign post. The sequence also contains significant illumination variations within the car at left, and several occlusions and noisy areas within the trees.

The results obtained show that our algorithm is capable of detecting areas with variation of illumination within the car and the trees (Figs. 3.c and 3.d). Consequently, the reconstruction error and the number of regions obtained from quadtree segmentation are reduced (Table 1). Figs. 3.e and 3.f show vector fields obtained for models $M1$ and $M2$, respectively. Figs. 3.g and 3.h show corresponding differences between vector field obtained using illumination variation model and without such a model. It can be seen that the illumination variation model modifies motion estimates, especially in image areas were such variation is present, so that the reconstruction error is reduced.

The above results show usefulness of the proposed illumination variation model incorporated into an image coding algorithm based on motion compensation.

References

[1] WALKER D.R. , RAO K.R. Improved pel-recursive motion compensation. *IEEE Transactions on Communications*, Vol.32, No.10:pp 1128–1134, October 1984.

[2] DUBOIS E. Motion-compensated filtering of time-varying images. *Multidimensional Systems and Signal Processing (invited paper), No. 3*, pages pp. 211–239, 1992.

[3] ADIV G. Determining three-dimensional motion and structure from optical flow generated by several moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol 7:pp 384–401, July 1985.

[4] LABIT C. , NICOLAS H. Compact motion representation based on global features for semantic image sequence coding. *Proc. of VCIP, Boston*, November 1991.

[5] SCHOTT J.P. , Three-dimiensional motion estimation using shading information in multiple frames. Technical report, Massassuchett Institut of Technology, Artificial Intelligence Laboratory, USA, Aout 1989.

[6] NICOLAS H. , LABIT C. Region-based motion estimation using deterministic relaxation schemes for image sequence coding. *Proc. ICASSP, San-Francisco*, pp. III–265–268, March 1992.

[7] VERRI A. , POGGIO T. Motion field and optical flow: qualitative properties. *M.I.T.*, A.I. 917, Décembre 1986.